Dynamic Coarse-to-Fine Learning for Oriented Tiny Object Detection

Chang Xu, Jian Ding, Jinwang Wang, Wen Yang*,
Huai Yu, Lei Yu*, Gui-Song Xia

Wuhan University

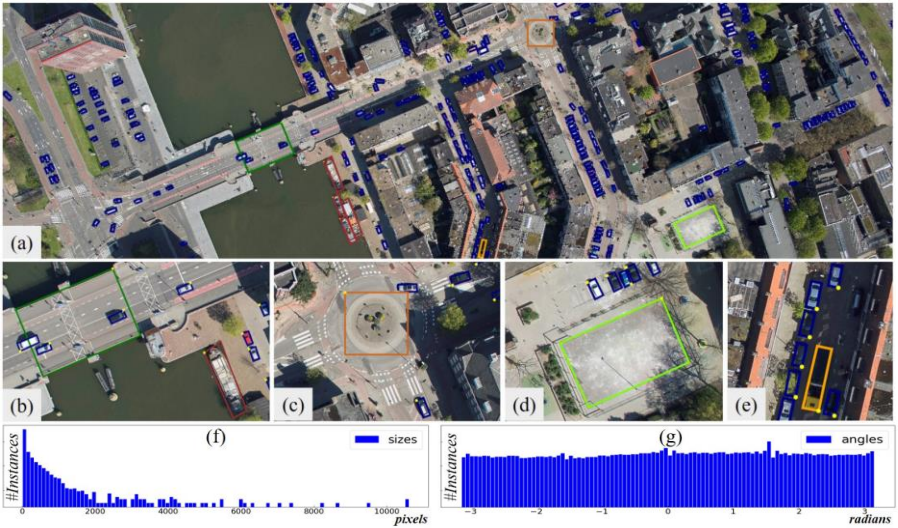
TUE-PM-305

武漢大學
WUHAN UNIVERSITY

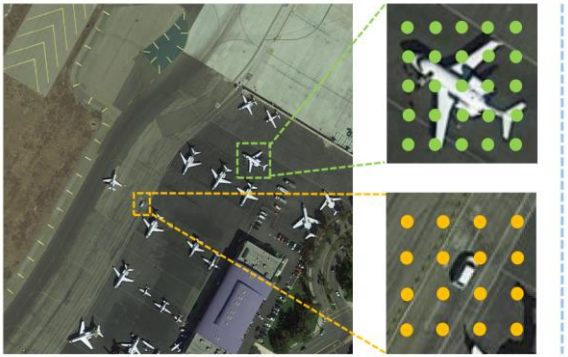
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Motivation

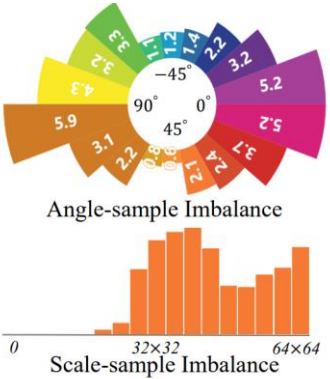
Oriented Tiny Object Detection



Mismatch and Imbalance Issues



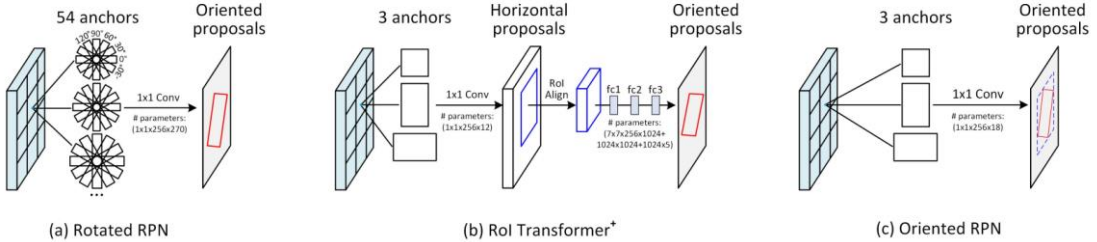
Mismatch



Imbalance

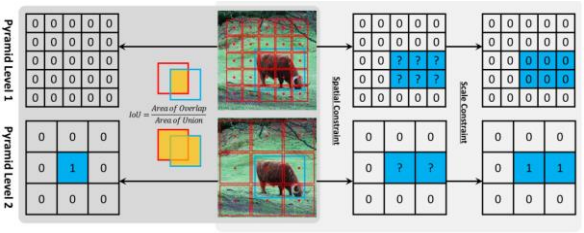
Related Works

Prior for Oriented Objects

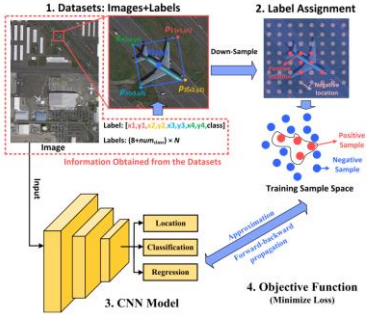


[1] Jianqi Ma, et al. Arbitrary-oriented scene text detection via rotation proposals. IEEE TMM, 2021.
[2] Jian Ding, et al. Learning roi transformer for detecting oriented objects in aerial images, 2019.
[3] Xingxing Xie, et al. Oriented R-CNN for Object Detection. ICCV, 2021.

Label Assignment



ATSS[4]

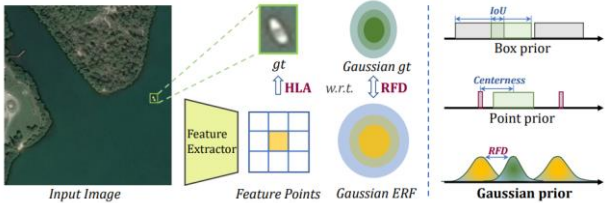


GGHL[5]

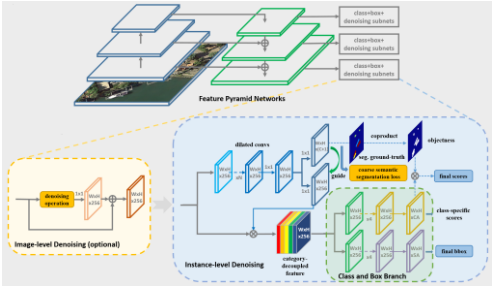
[4] Shifeng Zhang, et al. Bridging the Gap Between Anchor-based and Anchor-free Detection via Adaptive Training Sample Selection. CVPR, 2020.

[5] Zhanchao Huang, et al. A General Gaussian Heatmap Label Assignment for Arbitrary-Oriented Object Detection. IEEE TIP, 2022.

Tiny Object Detection



RFLA[5]

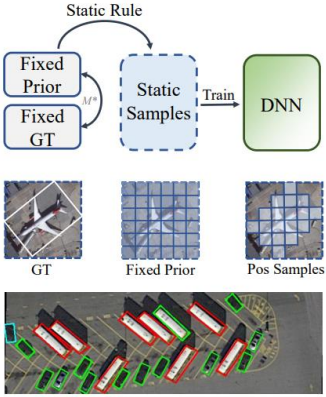


SCRDet++[6]

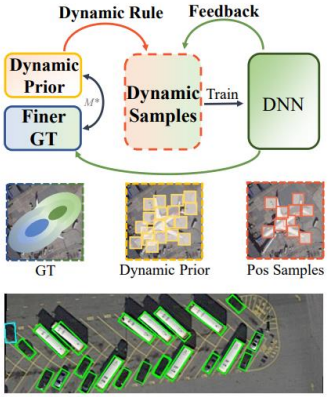
[6] Chang Xu, et al. RFLA: Gaussian Receptive Field based Label Assignment for Tiny Object Detection. ECCV, 2022.
[7] Yang X., et al. Detecting Small, Cluttered and Rotated Objects via Instance-Level Feature Denoising and Rotation Loss Smoothing. TPAMI, 2022

Method

Overview



(a) RetinaNet, FCOS, Rotated RPN



(b) DCFL

Static Learning:

$$\mathcal{L} = \sum_{i=1}^{N_{pos}} \mathcal{L}_{pos}(D_i, G_i) + \sum_{j=1}^{N_{neg}} \mathcal{L}_{neg}(D_j, y_j)$$

Dynamic Learning:

$$\mathcal{L} = \sum_{i=1}^{\tilde{N}_{pos}} \mathcal{L}_{pos}(\tilde{D}_i, \tilde{G}_i) + \sum_{j=1}^{\tilde{N}_{neg}} \mathcal{L}_{neg}(\tilde{D}_j, y_j)$$

Note: ~ denotes dynamic items

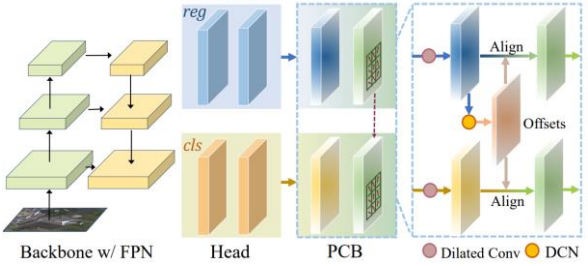
Dynamic Prior

RetinaNet

$$D = \mathrm{DNN}_h(P)$$

DCFL

$$\hat{D} = \mathrm{DNN}_h(\underbrace{\mathrm{DNN}_p(P)}_{\text{Dynamic Prior } \hat{P}})$$



- Component:

A Prior Capturing Block (PCB) with DCN

- Steps:

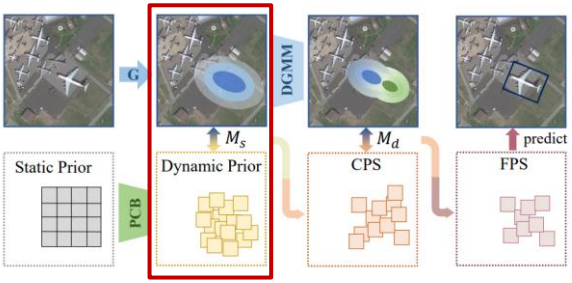
(1) Forward the network before presetting prior, get a set of offsets

(2) Use Gaussian to fit offsets as dynamic prior

Coarse Prior Matching

DCFL

$$\mathrm{GJSD}(\mathcal{N}_p, \mathcal{N}_g) = (1-\alpha)\mathrm{KL}(\mathcal{N}_\alpha, \mathcal{N}_p) + \alpha\mathrm{KL}(\mathcal{N}_\alpha, \mathcal{N}_g)$$



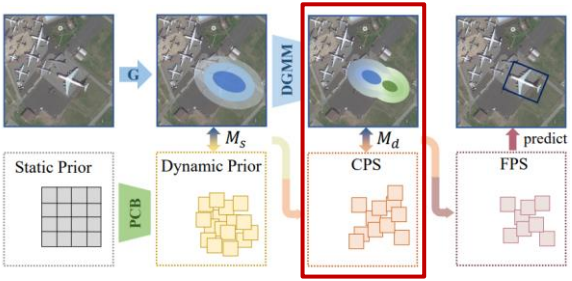
- Component:
Generalized Jensen-Shannon Divergence (GJSD) for coarse assigning

- Step:
Measure the similarity between the dynamic prior and the *gt* with GJSD, assign K samples as Coarse Positive Samples (CPS)

Finer Dynamic Posterior Matching

DCFL

$$DGMM_i(s|x,y) = \sum_{m=1}^{2} w_{i,m}\sqrt{2\pi|\boldsymbol{\Sigma}_{i,m}|}\mathcal{N}_{i,m}(\boldsymbol{\mu}_{i,m}, \boldsymbol{\Sigma}_{i,m})$$



- Component:
Dynamic Gaussian Mixture Model (DGMM) for finer matching

- Steps:
Set negative masks to samples which have $DGMM(s|MPS) < e^{-g}$

Experiments

Datasets and Implementation Details

Datasets:
OBB Datasets: DOTA-v1.0, DOTA-v1.5, DOTA-v2.0, DIOR-R
HBB Datasets: VisDrone2019, MS COCO

Augmentation:
Random Flipping

Platform:
RTX 3090

Default Training Time:
12 epochs: DOTA-v1.0, DOTA-v2.0, VisDrone2019, MS COCO
40 epochs: DOTA-v1.5, DIOR-R

Main Results

Main results on the DOTA-v2.0 OBB task

Method	Backbone	Plane	BD	Bridge	GTF	SV	LV	Ship	TC	BC	ST	SBF	RA	Harbor	SP	HC	CC	Air	Heli	mAP
multi-stage:																				
FR OBB [44]	R50	71.61	47.20	39.28	58.70	35.55	48.88	51.51	78.97	58.36	58.55	36.11	51.73	43.57	55.33	57.07	3.51	52.94	2.79	47.31
FR OBB + Dp	R50	71.55	49.74	40.34	60.40	40.74	50.67	56.58	79.03	58.22	58.24	34.73	51.95	44.33	55.10	53.14	7.21	59.53	6.38	48.77
MR [19]	R50	76.20	49.91	41.61	60.00	41.08	50.77	56.24	78.01	55.85	57.48	36.62	51.67	47.39	55.79	59.06	3.64	60.26	8.95	49.47
HTC* [5]	R50	77.69	47.25	41.15	60.71	41.77	52.79	58.87	78.74	55.22	58.49	38.57	52.48	49.58	56.18	54.09	4.20	66.38	11.92	50.34
RT [10]	R50	71.81	48.39	45.88	64.02	42.09	54.39	59.92	**82.70**	**63.29**	58.71	41.04	52.82	53.32	56.18	57.94	25.71	63.72	8.70	52.81
Oriented R-CNN [56]	R50	77.95	50.29	**46.73**	**65.24**	42.61	54.56	**60.02**	79.08	61.69	59.42	42.26	**56.89**	51.11	56.16	**59.33**	25.81	60.67	9.17	53.28
one-stage:																				
DAL [38]	R50	71.23	38.36	38.60	45.24	35.42	43.75	56.04	70.84	50.87	56.63	20.28	46.53	33.49	47.29	12.15	0.81	25.77	0.00	38.52
SASM [21]	R50	70.30	40.62	37.01	59.03	40.21	45.46	44.60	78.58	49.34	60.73	29.89	46.57	42.95	48.31	28.13	1.82	**76.37**	0.74	44.53
RetinaNet-O [30]	R50	70.63	47.26	39.12	55.02	38.10	40.52	47.16	77.74	56.86	52.12	37.22	51.75	44.15	53.19	51.06	6.58	64.28	7.45	46.68
R^3Det w/ KLD [63]	R50	75.44	50.95	41.16	61.61	41.11	45.76	49.65	78.52	54.97	60.79	42.07	53.20	43.08	49.55	34.09	**36.26**	68.65	0.06	47.26
FCOS-O [51]	R50	74.84	47.53	40.83	57.41	43.89	47.72	55.66	78.61	57.86	63.00	38.02	52.38	41.91	53.24	40.22	7.15	65.51	7.42	48.51
Oriented Rep [26]	R50	73.02	46.68	42.37	63.05	47.06	50.28	58.64	78.84	57.12	**66.77**	35.21	50.76	48.77	51.62	34.23	6.17	64.66	5.87	48.95
ATSS-O [68]	R50	77.46	49.55	42.12	62.61	45.15	48.40	51.70	78.43	59.33	62.65	39.18	52.43	42.92	53.98	42.70	5.91	67.09	10.68	49.57
S^2A-Net [17]	R50	77.84	51.31	43.72	62.59	47.51	50.58	57.86	80.73	59.11	65.32	36.43	52.60	45.36	52.46	40.12	0.00	62.81	11.11	49.86
one-stage:																				
DCFL	R50	75.71	49.40	44.69	63.23	46.48	51.55	55.50	79.30	59.96	65.39	41.86	54.42	47.03	55.72	50.49	11.75	69.01	7.75	51.57
S^2A-Net w/ DCFL	R50	74.79	**53.25**	45.81	**65.46**	46.49	53.23	58.10	**81.51**	60.13	**66.42**	43.24	55.09	50.52	55.58	54.53	5.23	68.73	13.06	52.84
DCFL†	R50	**78.30**	53.03	44.24	60.17	**48.56**	**55.42**	58.66	78.29	60.89	65.93	**43.54**	55.82	**53.33**	**60.00**	54.76	30.90	74.01	**15.60**	**55.08**
DCFL†	ReR101	**79.49**	**55.97**	**50.15**	61.59	**49.00**	**55.33**	**59.31**	81.18	**66.52**	60.06	**52.87**	**56.71**	**57.83**	**58.13**	**60.35**	**35.66**	**78.65**	**13.03**	**57.66**

Main Results

Comparison with one-stage detectors on the DOTA-v1.0

Method	CFA [16]	RetinaNet-O [30]	R^3Det [60]	Oriented Rep [26]	ATSS-O [68]
mAP	69.63	69.79	70.18	71.94	72.29
Method	KLD [63]	S^2A-Net [17]	GGHL [23](3x)	DCFL	DCFL(3x)
mAP	72.76	73.91	73.98	74.26	**75.35**

Comparison with one-stage detectors on the DIOR-R

Method	RetinaNet-O [30]	FR-OBB [44]	RT [10]	AOPG [7]
mAP	57.55	59.54	63.87	64.41
Method	GGHL [23]	Oriented Rep [26]	DCFL	DCFL (ReR101)
mAP	66.48	66.71	66.80	**71.03**

Comparison with one-stage detectors on the DOTA-v1.5

Method	Backbone	SV	Ship	ST	mAP
RetinaNet-O [30]	R50	44.53	73.31	59.96	59.16
FR OBB [19]	R50	51.28	79.37	67.50	62.00
CMR [19]	R50	51.64	79.99	67.58	63.41
RT [10]	R50	52.05	80.72	68.26	65.03
ReDet [18]	ReR50	52.38	80.92	68.64	66.86
DCFL	R50	56.72 (+12.19)	80.87 (+7.56)	75.65 (+15.69)	67.37 (+8.21)
DCFL	ReR101	**57.31** (+12.78)	**86.60** (+13.29)	**76.55** (+16.59)	**70.24** (+11.08)

Results of one-stage object detectors on HBB datasets

Dataset	VisDrone		MS COCO		DOTA-v2.0 HBB	
Method	RetinaNet [30]	DCFL	RetinaNet	DCFL	FCOS [58]	DCFL
$AP_{0.5}$	29.2	32.1	55.4	57.3	55.4	57.4

Ablation Study

Method	CPS	MPS	DGMM	mAP
baseline [30]				51.70
DCFL	✓	✓		53.41
	✓		✓	57.20
	✓	✓	✓	**59.15**

(a) **Individual effectiveness.** CPS, MPS, and DGMM denote Coarse, Medium Sample Candidates and Dynamic Gaussian Mixture Model.

Strategy	Measurement	mAP
All-FPN-layer	Gaussian	50.12
Single-FPN-layer	Gaussian	56.72
Cross-FPN-layer	KLD [63]	57.82
Cross-FPN-layer	GWD [61]	58.55
Cross-FPN-layer	GJSD	**59.15**

(b) **Comparsions of different CPS.** The FPN layer number varies for different strategies of getting the CPS.

DCN	Dilated Conv	DP	mAP
			58.07
✓			58.41
✓	✓		58.65
Separate	✓	✓	58.71
Guiding	✓	✓	**59.15**

(c) **Effects of detailed designs in the PCB.** DP denotes the dynamic prior. Guiding denotes *reg* guides *cls*.

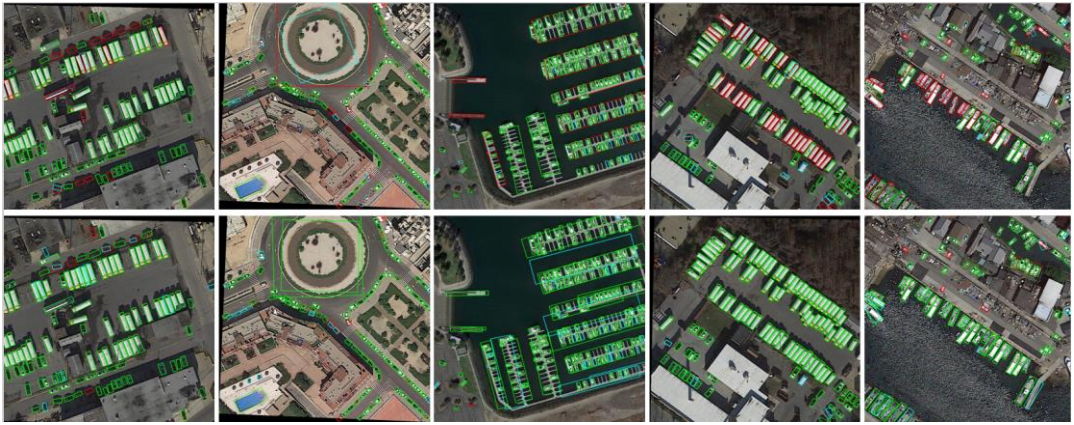
K	24				20			
Q	20	16	12	8	16	12	10	8
mAP	58.31	58.11	58.95	59.06	58.66	58.71	58.92	58.28
K	16				12			
Q	12	10	8	6	10	8	6	4
mAP	**59.15**	58.57	58.97	57.84	58.79	58.25	57.01	57.37

(d) **Effects of parameters K and Q.**

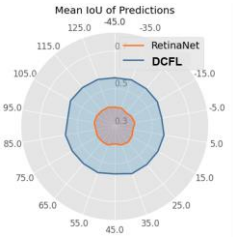
g	1.2	1.0
mAP	57.91	58.20
g	0.8	0.4
mAP	**59.15**	58.95

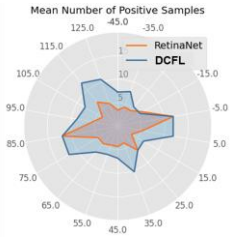
(e) **Effects of parameter g.**

Visual Results

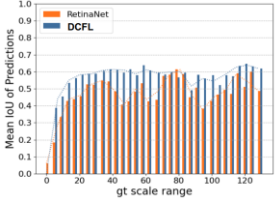


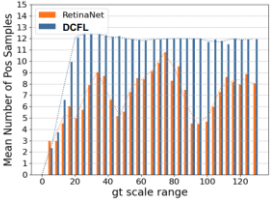
Analysis





Angle Imbalance





Scale Imbalance

Reconciliation of Imbalance Issues

Contributions

- We identify that there exist severe mismatch and imbalance issues in the current learning pipeline for oriented tiny object detection
- We design a Dynamic Coarse-to-Fine Learning (DCFL) scheme for oriented tiny object detection
- Extensive experiments on six datasets show promising results

Codes