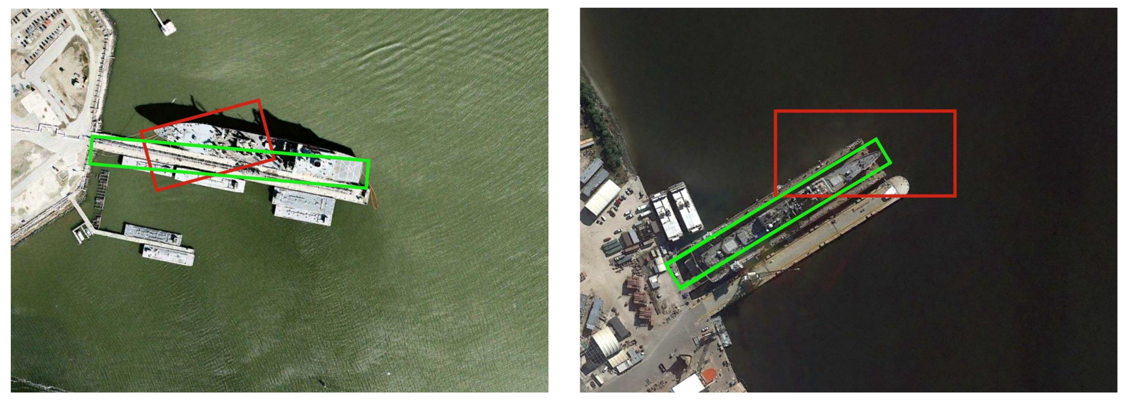
Dynamic Anchor Learning for Arbitrary-Oriented Object Detection

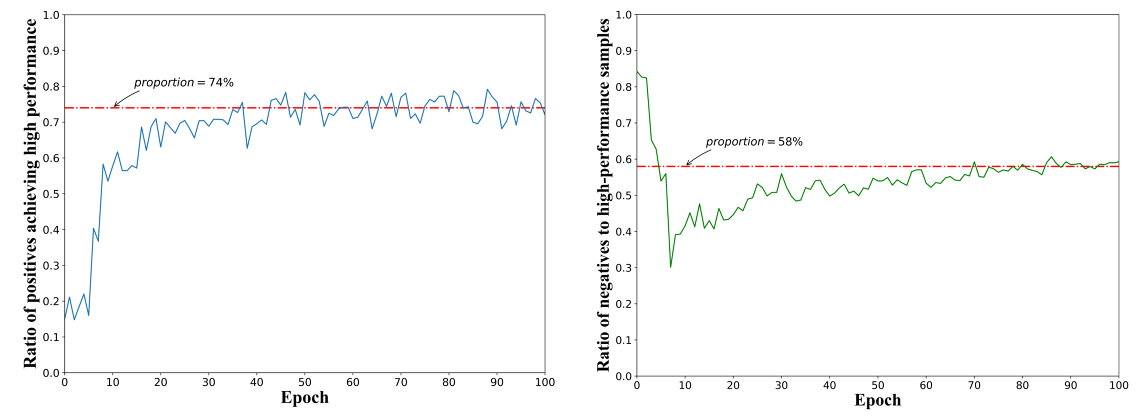
Qi Ming, Zhiqiang Zhou, Lingjuan Miao, Hongwei Zhang, Linhao Li Beijing Institute of Technology

Background



Many current rotation detectors use plenty of anchors with different orientations to achieve spatial alignment with ground truth boxes. Intersection-over-Union (IoU) is then applied to sample the positive and negative candidates for training. However, we observe that the selected positive anchors cannot always ensure accurate detections after regression, while some negative samples can achieve accurate localization. It indicates that the quality assessment of anchors through IoU is not appropriate, and this further leads to inconsistency between classification confidence and localization accuracy.

Aynalysis of regression



Positive samples are supposed to be able to accurately predict the ground-truth objects after training. However, only 74% of the positive anchors can localize GT well after regression (with output IoU higher than 0.5), which illustrates that many false positive samples are introduced.

Surprisingly, there are up to 58% of the high-quality detections (output IoU is higher than 0.5) come from unmatched anchors, which implies that quite a lot of negative anchors (58% in this example) have the potential to achieve accurate localization.

Matching-sensitive loss

Matching-sensitive losss

Definition 1 *For each ground-truth gand its matching degrees* $\boldsymbol{md}$*. Supposing that the maximal matching degree for* g *is* md_{max}*, we have* $\Delta md = 1 - md_{max}$*. Then we can obtainthe matching compensation factor:*

$$\boldsymbol{w} = \boldsymbol{md}_{pos} + \Delta md.$$

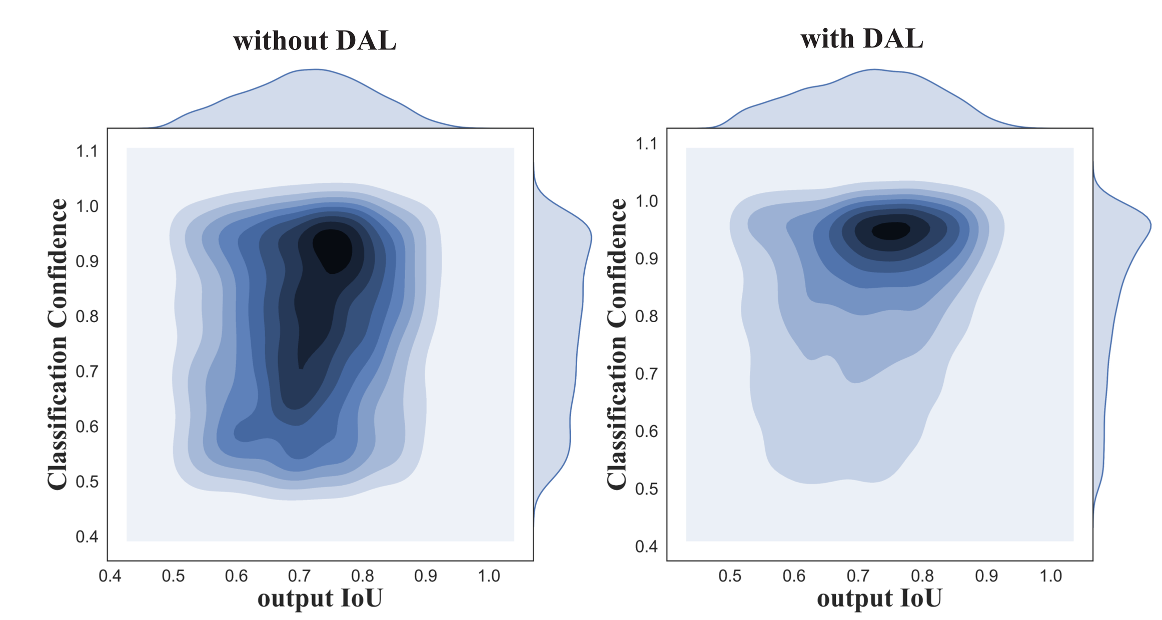
Definition 2 *Let* ψ *and* ψ_p *respectively represent all anchors and the positives.* $FL(\cdot)$ *is focal loss, the matching-sensitive classifaion loss as follows:*

$$L_{cls} = \frac{1}{N}\sum_{i\in\psi} FL\left(p_i, p_i^*\right) + \frac{1}{N_p}\sum_{j\in\psi_p} w_j \cdot FL\left(p_j, p_j^*\right)$$

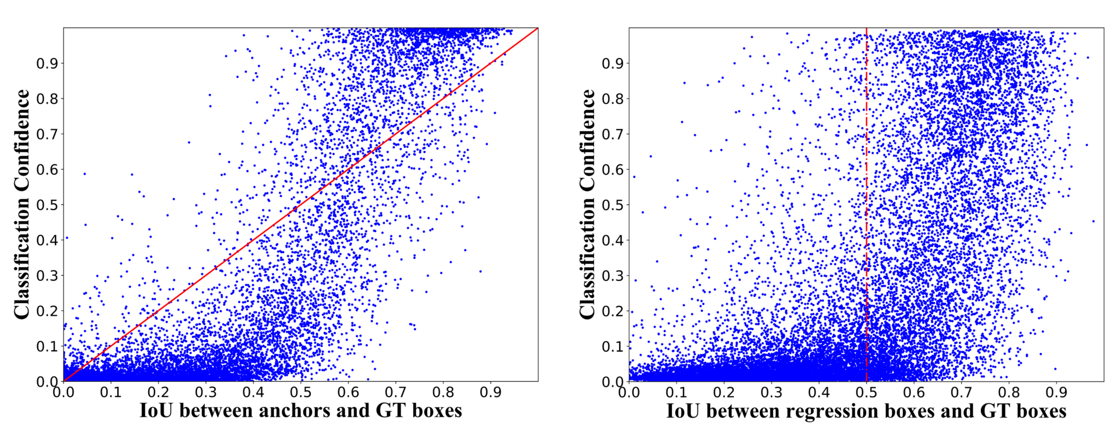
Definition 3 *Let* $L_{smooth_{L_1}}$ *denotes the smooth-*L_1 *loss for regression. The matching-sensitive regression loss as follows:*

$$L_{reg} = \frac{1}{N_p}\sum_{j\in\psi_p} w_j \cdot L_{smooth_{L_1}}\left(\boldsymbol{t_j}, \boldsymbol{t_j^*}\right)$$

Visualization Results



Misaligned Detections



Current IoU-based label assignment methods lead to a positive correlation between the classification confidence and the input IoU. However, the input IoU is not entirely equivalent to the localization performance. Therefore, we cannot distinguish the localization performance of the detection results based on the classification score. The picture on the right further illustrates this point of view: a large number of regression boxes with high output IoU are misjudged as background. It is not conducive to selecting accurate detection results through classification score during inference.

Dynamic Anchor selection

Definition 4 *Let* sa *denotes input IoU.* fa *represents the output IoU.* u *is a penalty term, which is obtained via the IoU variation:*$u = |sa - fa|$*. Then we can obtain a novel metric for label assignment named matching degree(*$\boldsymbol{md}$*) as follows:*

$$md = \alpha \cdot sa + (1 - \alpha) \cdot fa - u^{\gamma}$$

α and γ are hyperparameters used to weight the different items. Label assignmetn with the matching degree helps to control the quality of samples, thereby accelerating network convergence and achieving superior detection performance.

Different from the previous work that utilize output IoU implicitly and indirectly, we focus on how to directly use output IoU to optimize detection. The suppression of interference during regression is vital for high-quality anchor sampling and stable training.

Experimental Results

Comparisons with other label assignment methods

model	Baseline	S^3FD	HAMBox	ATSS	DAL
mAP	80.8	82.2	85.4	86.1	**88.6**

Performance evaluation of different datasets

Datasets	BBox	Models	Backbone	Size	NA	mAP/F_1
HRSC2016	OBB	RetinaNet	ResNet50	416×416	3	80.81
	OBB	R^3Det	ResNet101	800×800	21	89.26
	OBB	DAL	ResNet101	800×800	3	89.77
UCAS-AOD	OBB	Faster R-CNN	ResNet50	800×800	5	88.36
	OBB	RoI Transformer	ResNet101	512×800	5	88.95
	OBB	DAL	ResNet101	800×800	3	89.87
DOTA	OBB	RetinaNet	ResNet50	800×800	3	68.43
	OBB	DAL	ResNet50	800×800	3	71.78
	OBB	S^2A-Net	ResNet50	800×800	1	74.12
	OBB	S^2A-Net + DAL	ResNet50	800×800	1	76.95
VOC2007	HBB	RetinaNet	ResNet50	800×800	3	74.9
	HBB	DAL	ResNet50	800×800	3	76.1
ICDAR2013	HBB	RetinaNet	ResNet50	800×800	3	77.2
	HBB	DAL	ResNet50	800×800	3	81.3

References

[1] Li, H.; Wu, Z.; Zhu, C.; Xiong, C.; Socher, R.; and Davis, L. S. Learning from noisy anchors for one-stage object detection In *CVPR '20*

[2] Zhang, X.; Wan, F.; Liu, C.; Ji, R.; and Ye, Q. Freeanchor: Learning to match anchors for visual object detection In *NIPS '19*

Conclusion

The propose a dynamic anchor learning strategy make the label assignment more reasonable. The matching degree is constructed to comprehensively considers the spatial alignment, feature alignment ability, and regression uncertainty for label assignment. Then dynamic anchor selection and matching-sensitive loss are integrated into the training pipeline to improves the high-precision detection performance and alleviate the gap between classification and regression tasks.

Source Code

The source code of the proposed Dynamic Anchor Learning methods as a plug-in of an evaluation framework is available at : `https://github.com/ming71/DAL`