RankDetNet: Delving into Ranking Constraints for Object Detection Supplementary Material

Ji Liu, Dong Li, Rongzhang Zheng, Lu Tian, Yi Shan
Xilinx Inc., Beijing, China
{jiliu1, dongl, treemann, lutian, yishan}@xilinx.com

1. Overview

In this supplementary material, we present additional experimental results and analysis.

- We present more results of reverse pairs compared between different detection baselines and our method.
- We show additional sample distribution for the FCOS-ATSS baseline and our method.
- We present ablation studies on the bin size in our grouping strategy of global and class-specific ranking losses.
- We present the per-class AP for the 10 classes with the most and fewest GT boxes on COCO, respectively.
- We provide more qualitative results of our method for 2D and 3D object detection.

2. More Analysis of Reserve Pairs

We define three types of reverse pairs for better understanding the effect of our ranking constraints. Table 4 presents the ratios of reverse pairs to all pairs after training compared with different detection baselines and our method. The results show that our RankDetNet can largely reduce the amount of reverse pairs for both anchor-based and anchor-free detectors, e.g., reducing 7% and 6% ratios of F-B and P-N reverse pairs compared with RetinaNet (ResNet-50), respectively. These results demonstrate the effectiveness of our ranking constraints for better optimization.

3. More Analysis of Sample Distribution

Figure 1 shows sample distributions for the FCOS-ATSS baseline and our method. For the foreground or positive samples, our method can generate higher object confidence scores. Compared with the foreground and background distributions, the results show that our method can reduce their overlap. Similarly, the reduced overlap can be observed between the positive and negative sample distributions for a specific class (person). Compared with IoU overlap and object confidence scores for positive samples, the results show that our method obtains more consistent distributions and stronger correlation. These results further validate that our method can help improve optimization for detection from the perspective of sample distribution.

4. Ablation Study on Bin Size

We test different numbers of bins for the global and class-specific losses. Table 1 presents the detailed results. The extreme case of #bin=1 does not work well, and too many bins may cost longer training time and more memory usage. In general, the hyper-parameter of bin size is not sensitive in a wide value range based on our experiments. We set the bin size as 15 in the global ranking loss and 3 for each class in the class-specific ranking loss.

5. Per-class AP on COCO

COCO is a highly imbalanced dataset. The amount of ground-truth bounding boxes for each class varies a lot. Table 2 and Table 3 present the per-class AP for the 10 classes with the most and fewest GT boxes on COCO, respectively. The results show that our method performs better than the RetinaNet baseline on most of imbalanced classes, e.g., +3.5% AP on the person class with 262465 GT boxes and +6.1% on the toaster class with 225 GT boxes.

6. More Qualitative Results

Figure 2 presents more qualitative results of our method for 2D and 3D object detection. Our RankDetNet can generate accurate 2D and 3D bounding boxes for RGB and Bird-Eye-View (BEV) images.

Methods	#Bin for global	#Bin for class-specific	AP
RankDetNet-R50	15	1	37.0
RankDetNet-R50	15	3	37.8
RankDetNet-R50	15	5	37.8
RankDetNet-R50	15	8	37.7
RankDetNet-R50	15	12	37.7
RankDetNet-R50	1	3	33.5
RankDetNet-R50	6	3	37.6
RankDetNet-R50	10	3	37.7
RankDetNet-R50	15	3	37.8
RankDetNet-R50	25	3	37.7

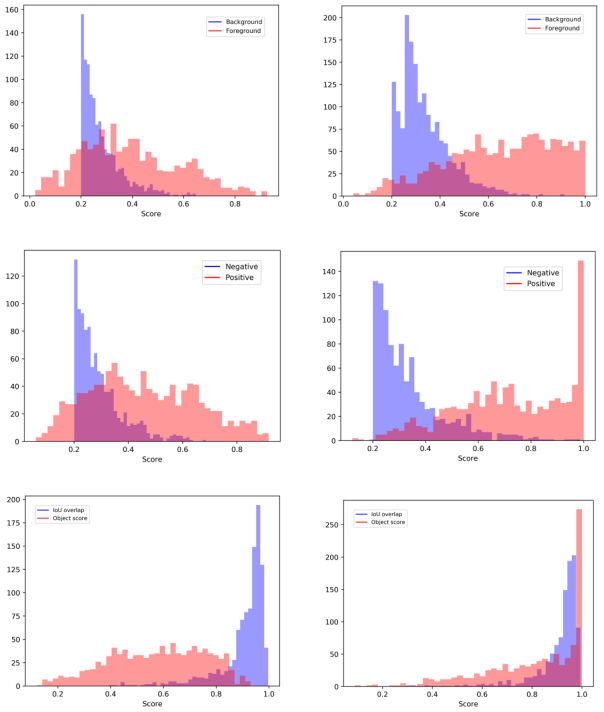
Table 1. Detection performance comparisons (%) on the COCO 2017 validation set with different numbers of bins.

Methods	person	carrot	car	chair	book
#GT boxes	262465	51719	43867	38491	24715
RetinaNet	49.3	19.9	38.9	23.2	11.5
RankDetNet	52.8	19.9	41.6	24.1	13.1
AP gain	3.5	0.0	2.7	0.9	1.6
Methods	bottle	cup	dining table	bowl	traffic light
#GT boxes	24342	20650	15714	14358	12884
RetinaNet	33.8	38.7	24.1	37.3	22.8
RankDetNet	35.1	41.2	26.2	39.5	23.8
AP gain	1.3	2.5	2.1	2.2	1.0

Table 2. Comparisons between our method and RetinaNet on the top 10 classes with the most GT boxes of COCO.

Methods	mouse	stop sign	tooth brush	fire hydrant	microwave
#GT boxes	2262	1983	1954	1865	1673
RetinaNet	56.5	59.6	13.4	61.3	48.3
RankDetNet	56.7	63.5	15.4	62.7	55.0
AP gain	0.2	3.9	2.0	1.4	6.7
Methods	scissors	bear	parking meter	toaster	hair drier
#GT boxes	1481	1294	1285	225	198
RetinaNet	19.0	64.5	46.0	11.5	0.8
RankDetNet	24.3	67.5	45.6	17.6	1.5
AP gain	5.3	3.0	-0.4	6.1	0.7

Table 3. Comparisons between our method and RetinaNet on the last 10 classes with the fewest GT boxes of COCO.



(a) FCOS-ATSS baseline (b) Our method

Figure 1. Comparisons of sample distribution between the FCOS-ATSS baseline and our method. We plot foreground and background distribution (first row), positive and negative distribution for the person class (second row) and distribution of IoU overlap and object confidence score for the person class (third row). We select background / negative samples with score > 0.2 for better visualization.

Methods	Backbone	F-B	P-N	P-P
RetinaNet	ResNet-50	0.33	0.28	0.44
RetinaNet + RankDetNet	ResNet-50	0.26	0.22	0.41
RetinaNet	ResNet-101	0.30	0.26	0.44
RetinaNet + RankDetNet	ResNet-101	0.24	0.21	0.41
RetinaNet	ResNeXt-64×4d-101	0.27	0.23	0.44
RetinaNet + RankDetNet	ResNeXt-64×4d-101	0.20	0.18	0.41
FCOS-ATSS	ResNet-50	0.24	0.18	0.42
FCOS-ATSS + RankDetNet	ResNet-50	0.17	0.13	0.41
FCOS-ATSS	ResNet-50-DCN	0.21	0.16	0.42
FCOS-ATSS + RankDetNet	ResNet-50-DCN	0.13	0.11	0.41
FCOS-ATSS	ResNeXt-64×4d-101-DCN	0.11	0.12	0.43
FCOS-ATSS + RankDetNet	ResNeXt-64×4d-101-DCN	0.09	0.10	0.42

Table 4. Ratios of reverse pairs compared with different detection baselines and our method. F-B: foreground-background reverse pair where foreground has lower score. P-N: positive-negative reverse pair where the positive sample of a specific class has lower score. P-P: positive-positive reverse pair where the one positive sample with larger IoU overlap with GT has lower object confidence than the other.

Figure 2. Examples of qualitative results by our RankDetNet for 2D and 3D object detection.