Table 1: Detection results of Faster R-CNN with varying # proposals

# proposals	AP	
	w/ RPN	w/ CRPN
100	34.8	38.2
300	37.0	40.5
1000	36.8	40.4

Table 2: Detection results of Cascade R-CNN w.r.t. different RPN methods

Method	Proposal method	AP
Cascade R-CNN	RPN	40.4
	CRPN	41.2

Table 3: Proposal results with different # stages

# stages	AR_{1000}	Time (s)
1	66.3	0.04
2	71.7	0.06
3	72.2	0.08

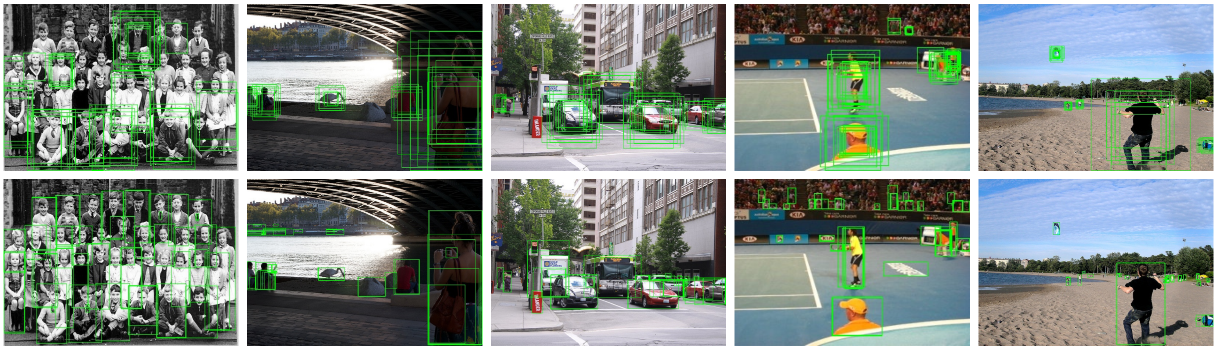

Figure 1: Examples of region proposal results at stage 1 (first row) and stage 2 (second row) of Cascade RPN

We thank the reviewers for their valuable feedback and encouraging comments.

1. Response to Reviewer #1

1.1. Performance with varying the number of proposals. If we understand the question correctly, the reviewer is probably more concerned with AP as opposed to AR of different RPN methods since the AR w.r.t. different number of proposals was presented in the Table 1 of the paper. Here, we report the AP of Faster R-CNN on COCO 2017 `val` with 100, 300, and 1000 proposals in Table 1, showing that Cascade RPN consistently outperforms conventional RPN.

2. Response to Reviewer #2

2.1. Higher AP improvements for large objects. The Cascade RPN does not have preferences for large objects as it does not make any prior assumptions regarding the object scale. Large objects are simply easier to detect than small objects, and the difference in performance gain w.r.t. object scales is also observed in other RPN methods such as the GA-RPN.
2.2. Why IOU loss? The use of IoU loss is simply a design choice. The IoU loss regresses the bounding box as a single unit instead of 4 independent variables considered in L1-smooth loss. The experiment results when using the different losses were performed and reported Table 3 of the paper: IoU loss marginally improves AR_{1000} from 71.5 to 71.7.
2.3. Figure label size and typos. We will revise and fix the label size and typos.

3. Response to Reviewer #3

3.1. Performance with Cascade R-CNN. Table 2 shows preliminary results obtained using Cascade R-CNN which is fine-tuned with precomputed proposals acquired from converged Cascade RPN. Here, the hyper-parameters of the Cascade R-CNN are not modified for fine-tuning. The Cascade RPN improves AP by 0.8 points compared to RPN. This performance could probably be further improved with a better hyper-parameter setting of Cascade R-CNN.
3.2. Performance w.r.t. different number of stages. These experiment results are shown in Table 3. In 3-stage Cascade RPN, an IoU threshold of 0.75 is used for the 3rd stage. The 2-stage Cascade RPN achieves the best trade-off between AR_{1000} and inference time. Increasing # stages to 3 improves AR_{1000} by 0.5 points but results in 33% slower.
3.3. Using multiple anchor shapes. We have not used or tried multiple anchor shapes per location. The main theme of the paper is to avoid heuristically defined hyper-parameters such as aspect-ratios of anchors. However, Cascade RPN can be easily extended to multiple anchor shapes.
3.4. Visualization at different stages. The output proposals of stage 1 and 2 of Cascade RPN are shown respectively in first and second rows of Figure 1. The proposals at stage 2 are more accurate and cover a larger number of objects.
3.5. Motivation of anchor-free metric for the first stage. The motivations are twofold: (1) the anchor-free metric is more relaxed than anchor-based metric since anchor-free metric is defined irrespective of the anchor shape and (2) at first stage, anchors are initialized uniformly over the image, using anchor-based metric may result in poor overlaps between anchors and ground truth boxes. We reported the detailed ablation study on sample metrics in Table 5 in the paper. The best AR_{1000} we were able to obtain when using anchor-based metric is 67.8, where the IoU thresholds of 0.5 and 0.7 were used in the first and second stages. Combining anchor-free and anchor-based metric improves AR_{1000} to 68.6.
3.6. "CRAFT Objects from Images" paper. We thank the reviewer for the constructive suggestion. We will cite the CRAFT method in our paper.