View Author Feedback For Paper

Paper ID	125
Title	RefinedBox: Towards Object Proposal Refinement and Joint Object Detection

	Question	Response
1	Rebuttal	**We would like to thank all the reviewers for your time and detailed comments on our manuscript. We provide clarification on major comments, which will be included in the future version of our paper.** **Novelty: Although using deep nets for proposal reranking is not new, the main contribution comes from the carefully designed network that balances the efficiency and effectiveness of box refinement. Because there are usually thousands of initial proposals, it is not straight-forward to refine them efficiently and effectively. Although the proposed RefinedBox module is simple, the evaluation results have demonstrate its significant improvement in proposal refinement.** **Detection Performance: As shown in introduction, we mainly focus on reducing the number of proposals for each image in this paper. In some research fields, e.g. mining knowledge from huge amounts of unlabeled/weakly-labeled data, the large number of false positives will pose significant challenges not only for computational efficiency but also for system stability. Some algorithms that attempt to reduce the number of proposals for their applications have been proposed, such as multi-label image classification (Wei et al. 2016), pedestrian detection (Paisitkriangkrai, Shen, and van den Hengel 2016), and weakly supervised learning (Wu et al. 2015). Thus, object detection is just one of our applications, and it does not represent all of the work. Our proposed method can cover most of the objects in an image using only several proposals. For example, using only 10 proposals per image, the object detection recall is 88.8% and 78.8% for IoU 0.5 and 0.7, respectively. The generated high-quality proposals will benefit many vision tasks, not only object detection.** **Response to Reviewer 1:** **Thank you for your suggestions and patience again.** **Runtime: We are sorry for the lack of runtime analysis. Using about 6000 initial boxes per image, the runtime of box refinement is about 0.057 second on a single GPU without considering object detection part. The runtime can be further reduced if fewer initial boxes are used. The increased computational load can be ignored when compared with many proposal generation methods, such as MCG (17.46s per image), Selecive Search (1.60s per image), and LPO (0.46s per image). We will add detailed runtime comparison in the future version of this paper.** **Evaluation: Given limited page length, we only show the evaluation on the VOC2007, though we have many more experimental results on VOC2012 and COCO datasets. We will try our best to adjust the paper structure for additional results.** **Response to Reviewer 2:** **We appreciate your very detailed comments on the analysis of the proposed algorithm.**

		System: As we shown in Table 1, EdgeBoxes can be easily replaced by other proposal generation methods (e.g. MCG, SelectiveSearch, and RPN), and the resulting systems can perform similar to the one using EdgeBoxes. VGG16 can also be replaced by other base networks. Since we think object detection is an important application of proposals, we use Fast R-CNN to demonstrate the effectiveness of our refinement. Of course, Fast R-CNN can be easily replaced by other detection modules. Please note that the proposed RefinedBox network is carefully designed and can be connected to other networks to perform high-quality proposal refinement rather than a trivial combination of several components. **Analysis of reranking and box regression: We show detection recall of proposals using IoU 0.7 as below.** **with box regression: 1:0.511, 5:0.710, 10:0.751, 20:0.776, 100:0.804, 300:0.820** **without box regression: 1:0.491, 5:0.660, 10:0.712, 20:0.748, 100:0.789, 300:0.803** **The detection recall with box regression is 2%-5% higher than refinement without box regression. Especially when 5-20 proposals per image are used, the performance gap is large. A few tens of proposals are very important for many vision tasks as described above.** **The performance sharply drops as it approaches 1 IoU: This phenomenon is highly related to the initial boxes. We can see EdgeBoxes suffers a more serious descent. In our experiments, we find RefindBox with MCG outperforms all the competitors as it approaches 1 IoU.** **Comparison with DeepBox: Both reranking and box refinement contribute to the improvement over DeepBox. We will explain why RefinedBox performs much better than DeepBox in the future version.** **Network Architecture: We are sorry for the unclear description. Both RefinedBox and object detection modules use the full depth of the VGG network. We divide the VGG16 network into 5 stages according to the four pooling layers. Thus, Conv5 in Fig.2 consists of conv5_1, conv5_2, and conv5_3. We will provide better description in the next version.** **Response to Reviewer 3:** **Thank you for your positive comments. We will reinterpret Fig.1 and fix the grammatical errors in the final version. More runtime comparison will be included, too. In addition, RefinedBox only need 0.057 second per image when 6000 initial proposals are used. This efficiency makes it convenient for many vision tasks.** **Since the proposal quality of RefinedBox is very high, we think the bottleneck of object detection is not proposal generation but the subsequent classification. More powerful semantic features for proposal classification may be needed. We will investigate this by using more powerful detection modules rather than only Fast R-CNN.**
2	Confidential comments to the Senior Program Committee.	**Thank you very much for your reading.** **As shown in introduction, this paper mainly focuses on reducing the number of proposals for each image. Thus we spend lots of space to show proposal evaluation results. However, some reviewers seem to mainly focus on the application of object detection and ignore the importance of our significant improvement on proposal quality. In some research fields, e.g. mining knowledge from huge amounts of unlabeled/weakly-labeled data, the large number of false positives will pose significant challenges not only for computational efficiency but also for system stability.** **Besides, our proposed network is carefully designed to balance the efficiency and effectiveness of box refinement. Since there are usually thousands of initial proposals, it is not straight-forward to refine them efficiently. Thus, we argue that our algorithm is not a simple combination of existing technics but a novel design.**