

[AAAI-18](#)**Association for the Advancement of Artificial Intelligence 2018**

February 2 - 7, 2018, New Orleans, USA

Reviews For Paper**Paper ID** 125**Title** RefinedBox: Towards Object Proposal Refinement and Joint Object Detection**Masked Reviewer ID:** Assigned_Reviewer_1**Review:**

Question	
[Summary] Please summarize the main claims/contributions of the paper in your own words.	The paper presents a refinement network for improving the quality of object bounding box proposals, which can be jointly trained with the Fast R-CNN model for object detection. The refinement network consists of a convolution, ROI Pooling and fully-connected layer, and re-estimates the objectness scores and locations of any initial bounding box proposal set. The refined proposals are then fed into a Fast-RCNN to generate the detection results. This framework is evaluated on Pascal 2007 dataset. It achieves much better average recalls than several other proposal methods and slightly improves the detection performance.
[Relevance] Is this paper relevant to an AI audience?	Relevant to researchers in subareas only
[Significance] Are the results significant?	Moderately significant
[Novelty] Are the problems or approaches novel?	Somewhat novel or somewhat incremental
[Soundness] Is the paper technically sound?	Technically sound
[Evaluation] Are claims well-supported by theoretical analysis or experimental results?	Somewhat weak
[Clarity] Is the paper well-organized and clearly written?	Satisfactory
[Detailed Comments] Please elaborate on your assessments and provide constructive feedback.	<p>Pros:</p> <ul style="list-style-type: none"> - The method achieves significantly better object recall performance on the Pascal 2007 datasets than previous proposal methods. - The paper is well written and easy to follow. <p>Cons:</p> <ul style="list-style-type: none"> - The idea of refining initial object proposal using additional network module is not new, see for example, SharpMask. The proposed refinement network is essentially a foreground object detection network that uses a simplified Fast RCNN network. The novelty of this work seems a bit incremental. - The analysis of runtime complexity is lacking. The paper did not provide detailed statistics of runtime of the proposed method with comparison to the other proposal baselines. The computation analysis in the paper may not reflect the actual runtime cost. - The evaluation of the method is a bit lacking. Only one relatively small dataset (Pascal 2007)

	<p>is used in the evaluation. More datasets, e.g., Pascal 2012 or MS coco, should be also included. It is unclear if this method overfits the smaller 2007 dataset.</p> <p>- The overall improvement on the final detection performance is a bit disappointing. It is only slightly better than Faster-RCNN although the generated proposals are much better than the RPN's results. It seems the improvement on the proposal generation has very small impact on the final detection outcomes, and hence less relevant.</p>
<p>[QUESTIONS FOR THE AUTHORS]</p> <p>Please provide questions for authors to address during the author feedback period.</p>	See the detailed comments above.
[OVERALL SCORE]	Reject
<p>Please acknowledge that you have read the author rebuttal. If your opinion has changed, please summarize the main reasons below.</p>	<p>The author rebuttal did not address the main concern on lack of other datasets in the evaluation as they did not provide any new results in the reply. Also, it is still unclear to me how useful it is for any downstream vision task: for detection, the improvement is marginal. I would maintain my rating after reading the rebuttal.</p>

Masked Reviewer ID: Assigned_Reviewer_2

Review:

Question	
[Summary] Please summarize the main claims/contributions of the paper in your own words.	The authors represent a network for refining bottom-up object hypotheses. The network is used to rank proposals using convolutional features from a deep feature extractor and perform bounding box regression to improve localisation accuracy. They show a marked improvement over the state of the art.
[Relevance] Is this paper relevant to an AI audience?	Relevant to researchers in subareas only
[Significance] Are the results significant?	Highly significant
[Novelty] Are the problems or approaches novel?	Somewhat novel or somewhat incremental
[Soundness] Is the paper technically sound?	Technically sound
[Evaluation] Are claims well-supported by theoretical analysis or experimental results?	Sufficient
[Clarity] Is the paper well-organized and clearly written?	Good
	This is a very important problem and improvements are really welcome. In general, I find the results impressive and convincing, with a very extensive evaluation. It is hard to argue with that.

<p>[Detailed Comments] Please elaborate on your assessments and provide constructive feedback.</p>	<p>On the other hand, I don't see much novelty or new insight. It is hard to balance the two. The use of a deep net for reranking is not new -- DeepBox is mentioned in the paper. The novel loss function introduced to train this seems rather straight-forward.</p> <p>Bounding box regression with deep nets is also used by other methods (see e.g. some of Xiaogang Wang's work like http://www.ee.cuhk.edu.hk/~xgwang/papers/deepIDNet.pdf). The loss function for minimising this is taken from previous work.</p> <p>The new refined box architecture forms a small part of the overall system, which also incorporates EdgeBox, VGG16, and Fast R-CNN. While system building is really hard and really important, I did not feel that there is much new theoretical insight in this paper.</p> <p>This could be improved by e.g. analysing the contribution of the ranking and box regression parts of the algorithm. It would be interesting to find out why the performance of the system sharply drops as it approaches 1 IoU -- refinement should be helping against this.</p> <p>However, the evaluation is thorough, detailed, and shows a marked improvement over the state of the art.</p>
<p>[QUESTIONS FOR THE AUTHORS] Please provide questions for authors to address during the author feedback period.</p>	<p>1) Did you do an analysis on why RefinedBox performs so much better than DeepBox? Is it due to better reranking or the additional bounding box refinement?</p> <p>2) Under Network Architecture, you first state that image goes through 13 convolutional layers of VGG16, but then you say that your network plugs in at layer 5 (which is also what Fig2 shows). Does this simply mean that the object detection part of the entire system uses the full depth of the VGG network, but that you only need features from level 5?</p>
<p>[OVERALL SCORE]</p>	<p>Marginally below threshold</p>
<p>Please acknowledge that you have read the author rebuttal. If your opinion has changed, please summarize the main reasons below.</p>	<p>I have read the rebuttal. Thank you for the detailed response.</p> <p>My main issue with the paper is that it is rather incremental, and this concern remains.</p> <p>I agree with R1 that evaluation on larger datasets would be very useful and would suggest reporting results on the most challenging dataset if space is limited. The paper would be stronger if similar improvements were shown on a variety of datasets.</p>

Masked Reviewer ID: Assigned_Reviewer_3

Review:

Question	
<p>[Summary] Please summarize the main claims/contributions of the paper in your own words.</p>	<p>This paper describes an "object proposal refinement" component that fits into a general object detection architecture. It uses previous (standard) components for the original object proposal, for the initial convolution layers for image feature extraction, and for the object detection, as shown nicely in Fig. 2. Results show improved object detection with minimal extra computation. Key features of the refinement component are that it reduces the number of proposals to be considered, and improves their ranking.</p>
<p>[Relevance] Is this paper relevant to an AI audience?</p>	<p>Relevant to researchers in subareas only</p>
<p>[Significance] Are the results significant?</p>	<p>Significant</p>
<p>[Novelty] Are the problems or approaches novel?</p>	<p>Novel</p>
<p>[Soundness] Is the paper technically</p>	<p>Technically sound</p>

sound?	
[Evaluation] Are claims well-supported by theoretical analysis or experimental results?	Very convincing
[Clarity] Is the paper well-organized and clearly written?	Good
[Detailed Comments] Please elaborate on your assessments and provide constructive feedback.	<p>Figure 1 should be explained a bit better.</p> <p>The paper has minor grammatical errors.</p> <p>Results are well presented. Some timing results would be helpful to show, quantitatively, how low the overhead of the additional refinement component is.</p> <p>The authors make the claim that the higher detection recall and fewer proposals will benefit the subsequent object detection a lot. From Table 2, this is clearly true for some categories, but the overall mAP isn't that much better than the next best.</p>
[QUESTIONS FOR THE AUTHORS] Please provide questions for authors to address during the author feedback period.	Do the authors have any insights about what it would take to get mAP significantly higher (like to 80 percent)?
[OVERALL SCORE]	Accept
Please acknowledge that you have read the author rebuttal. If your opinion has changed, please summarize the main reasons below.	I've read the author rebuttal and it does not change my opinion of the paper.