

[CVPR 2016](#)**IEEE Conference on Computer Vision and Pattern Recognition 2016**

Las Vegas, USA

Reviews For Paper**Paper ID** 2195**Title** Attention to Scale: Scale-aware Semantic Image Segmentation**Masked Reviewer ID:** Assigned_Reviewer_28**Review:****Question**

Paper Summary	This paper proposes a method to improve fully convolutional neural network on multiple scale semantic segmentation. In stead of using mean-pooling or max-pooling on multiple scale classification outputs, the proposed method linearly combines the multi-scale soft classification using a set of weights from an attention network. The attention network and the classification network are trained at the same time.
Paper Strengths. Please discuss the positive aspects of the paper. Be sure to comment on the paper's novelty, technical correctness, clarity and experimental evaluation. Notice that different papers may need different levels of evaluation: a theoretical paper may need no experiments, while a paper presenting a new approach to a known problem may require thorough comparisons to existing methods. Also, please make sure to	The proposed method is reasonable and straightforward. Experiments show it gives slightly better results than mean-pooling and max-pooling.

<p>justify your comments in great detail. For example, if you think the paper is novel, not only say so, but also explain in detail why you think this is the case.</p>	
<p>Paper Weaknesses. Please discuss the negative aspects of the paper: lack of novelty or clarity, technical errors, insufficient experimental evaluation, etc. Justify your comments in great detail. If you think the paper is not novel, explain why and give a reference to prior work. Keep in mind that novelty can take a number of forms; a paper may be novel in terms of the method, the problem, the theory, analysis for an existing problem, or the empirical evaluation. If you think there is an error in the paper, explain in detail why it is an error. If you think the experimental evaluation is</p>	<p>From our experience, systematically expanding the training dataset so that the example objects include different scales is an effective way to train a multi-scale FCN. This paper uses only three different scales 1, 0.75 and 0.5. Even though objects already have different sizes in Parscal dataset, expanding the data with three scales may still be too few. If we use more scales in training, max or mean pooling results may be even closer to the proposed method.</p> <p>The attention network ideally should output the selection of the best scale for each pixel. So it can be trained independent of the semantic segmentation network. It is not very clear why a joint training is necessary.</p> <p>There is no theoretical guarantee this proposed method can really improve the result. Table 3 shows including the attention model can sometimes make the result worse.</p> <p>The proposed method needs extra supervision in different scales to outperform competing methods. How this extra supervision is implemented is never elaborated in the text.</p> <p>The experiments are on the Pascal voc 2012 dataset. Only three scales are used in the training and testing. The result improves the result only slightly. Max pooling result should be include in Table 3. It is unclear how the conclusion can be generalized to datasets with larger scale ranges, e.g. 0.1-10.</p> <p>The author admits the proposed method does not give better results than those from more recent works.</p>

insufficient, remember that theoretical results/ideas are essential to CVPR and that a theoretical paper need not have experiments. It is <i>*not*</i> okay to reject a paper because it did not outperform other existing algorithms, especially if the theory is novel and interesting. It is not reasonable to ask for comparisons with unpublished, non peer reviewed papers (e.g. ArXiv) or papers published after the CVPR'16 deadline.	
Preliminary Rating. This rating indicates to the area chair, to other reviewers, and to the authors, your current opinion on the paper. Please use 'Borderline' only if the author rebuttal and/or discussion might sway you in either direction.	Borderline
Preliminary Evaluation. Please explain to the AC, your	Even though recognizing the scale is an intuitive way to improve the multiple scale semantic segmentation, the improvement is not encouraging even competing with the simple mean or max pooling. The implementation also seems over complicated. The setting in the experiment is too

<p>fellow reviewers, and the authors your current opinion on the paper. This explanation may include how you weigh the importance of the various strengths and weaknesses you described above in Q1-Q3. Please summarize the key things you would like the authors to include in their rebuttals to facilitate your decision making. There is no need to summarize the paper.</p>	<p>restricted. It is hard to generalize the conclusion.</p>
<p>New exciting ideas. CVPR'16 would like to draw attention to papers that explore highly innovative ideas, novel problems, and/or paradigm shifts in conventional theory and practice. Such papers may not be "complete" in the "traditional" manner in the sense that it may not be possible to have experimental results comparing other related efforts or that they may</p>	<p>No</p>

not have large, publicly available data sets to be used for performance comparison. However, we expect these papers to be visionary by nature. Should this paper be considered under "new exciting ideas"?	
Reproducibility. Could the work be reproduced from the information in the paper? Are all important algorithmic or system details discussed adequately?	How this extra supervision is implemented is never elaborated in the text. It is a key to make the result better than mean or max pooling.
Confidence. Select: "Very Confident" to stress that you are absolutely sure about your conclusions (e.g., you are an expert who works in the paper's area), "Confident" to stress that you are mostly sure about your conclusions (e.g., you are not an expert but can distinguish good work from bad work in that area), and "Not Confident" to	Very Confident

stress that that you feel some doubt about your conclusions. In the latter case, please provide details as confidential comments to PC/AC chairs (point 7.)

Masked Reviewer ID: Assigned_Reviewer_5

Review:

Question

Paper Summary	The paper presents an approach that leverages inputs at multiple resolutions for the task of semantic segmentation. For this purpose, the paper trains the CNN architecture of [5] with shared weights for multi-resolution inputs and proposes an "attention mechanism", which learns weights for each scale and each spatial location in order to combine the corresponding multi-scale features. The method is evaluated on three datasets, showing an improvement over simpler max- and average-pooling alternatives for feature combination.
Paper Strengths. Please discuss the positive aspects of the paper. Be sure to comment on the paper's novelty, technical correctness, clarity and experimental evaluation. Notice that different papers may need different levels of evaluation: a theoretical paper may need no experiments, while a paper presenting a new approach to a known problem	<ul style="list-style-type: none"> - The paper studies one of the hottest topics in the field - The proposed attention model on top of DeepLab is simple yet effective and the overall system can be trained end-to-end - Strong experimental results - As a bonus, the attention model allows diagnostic visualization of the network - The paper is clearly written, technically correct and well organized

<p>may require thorough comparisons to existing methods. Also, please make sure to justify your comments in great detail. For example, if you think the paper is novel, not only say so, but also explain in detail why you think this is the case.</p>	
<p>Paper Weaknesses. Please discuss the negative aspects of the paper: lack of novelty or clarity, technical errors, insufficient experimental evaluation, etc. Justify your comments in great detail. If you think the paper is not novel, explain why and give a reference to prior work. Keep in mind that novelty can take a number of forms; a paper may be novel in terms of the method, the problem, the theory, analysis for an existing problem, or the empirical evaluation. If you think there is an</p>	<ul style="list-style-type: none">- Although the paper validates empirically the method on three datasets, the absolute gains in performance are relatively modest (+1% when training on PASCAL in Table 3. The variant DeepLab-MSc-CRF-COCO-LargeFOV-Attention is missing from the table), indicating that performance is mainly driven by the underlying DeepLab model. The best numbers are also below the current state-of-the-art of [24, 27].- The proposed method is only instantiated on top of DeepLab, and hence it is unclear whether the improvement could generalize to other semantic segmentation systems- The proposed modification to DeepLab is incremental, as it involves only combining features from the last layer. A more interesting extension would be to replace all max-pooling layers in the network by the attention model.

<p>error in the paper, explain in detail why it is an error. If you think the experimental evaluation is insufficient, remember that theoretical results/ideas are essential to CVPR and that a theoretical paper need not have experiments. It is <i>*not*</i> okay to reject a paper because it did not outperform other existing algorithms, especially if the theory is novel and interesting. It is not reasonable to ask for comparisons with unpublished, non peer reviewed papers (e.g. ArXiv) or papers published after the CVPR'16 deadline.</p>	
<p>Preliminary Rating. This rating indicates to the area chair, to other reviewers, and to the authors, your current opinion on the paper. Please use 'Borderline' only if the author rebuttal and/or discussion might sway you in</p>	<p>Weak Accept</p>

either direction.	
<p>Preliminary Evaluation. Please explain to the AC, your fellow reviewers, and the authors your current opinion on the paper. This explanation may include how you weigh the importance of the various strengths and weaknesses you described above in Q1-Q3. Please summarize the key things you would like the authors to include in their rebuttals to facilitate your decision making. There is no need to summarize the paper.</p>	<p>This paper presents a simple yet effective extension to a top performing semantic segmentation system. Appropriate handling of multiple scales is a topic of high interest for the CVPR community in the transition from PASCAL to the more varied and challenging MS-COCO.</p>
<p>New exciting ideas. CVPR'16 would like to draw attention to papers that explore highly innovative ideas, novel problems, and/or paradigm shifts in conventional theory and practice. Such papers may not be "complete" in the "traditional" manner in the sense that it may</p>	<p>Yes</p>

not be possible to have experimental results comparing other related efforts or that they may not have large, publicly available data sets to be used for performance comparison. However, we expect these papers to be visionary by nature. Should this paper be considered under "new exciting ideas"?	
Reproducibility. Could the work be reproduced from the information in the paper? Are all important algorithmic or system details discussed adequately?	The work could hardly be reproduced from the information in the paper, but a public release of the code is promised in P4.L375
Confidence. Select: "Very Confident" to stress that you are absolutely sure about your conclusions (e.g., you are an expert who works in the paper's area), "Confident" to stress that you are mostly sure about your conclusions (e.g.,	Very Confident

you are not an expert but can distinguish good work from bad work in that area), and "Not Confident" to stress that that you feel some doubt about your conclusions. In the latter case, please provide details as confidential comments to PC/AC chairs (point 7.)

Masked Reviewer ID: Assigned_Reviewer_7

Review:

Question

Paper Summary	This paper develops a new model for semantic segmentation with multiscale input images. The image at each scale is analyzed by the same DeepLab network, and their outputs are then fused pixel-wise by weighted average. The key distinction from past work is that, instead of merging features from different scales via average or max pooling over scales (which roughly correspond to constant and uniform weights across space and scale in this model), this work additionally uses an attention model to learn the weights. Specifically, it takes the fc7 features of VGG16 and has two fully convolutional layers, $3 \times 3 \times 512$ then $1 \times 1 \times \text{\#scales}$, producing a per-pixel per-scale weight map. The attention model is learned jointly with the rest of the DCNN during training. Experimental results show that such weights tend to give small additional boost to the performance than the average/max pooling fusion across scales.
Paper Strengths. Please discuss the positive aspects of the paper. Be sure to comment on the paper's novelty, technical correctness, clarity and experimental evaluation.	The paper works on a very interesting topic and is well written. The survey on multiscale integration into skip-net and share-net is nice. There are a good set of experiments on PASCAL and MSCOCO.

<p>Notice that different papers may need different levels of evaluation: a theoretical paper may need no experiments, while a paper presenting a new approach to a known problem may require thorough comparisons to existing methods. Also, please make sure to justify your comments in great detail. For example, if you think the paper is novel, not only say so, but also explain in detail why you think this is the case.</p>	
<p>Paper Weaknesses. Please discuss the negative aspects of the paper: lack of novelty or clarity, technical errors, insufficient experimental evaluation, etc. Justify your comments in great detail. If you think the paper is not novel, explain why and give a reference to prior work. Keep in mind that novelty</p>	<p>However, overall the paper presents an incremental work that offers superficial insight into scale fusion and makes little differentiation on the concept of scale: inner scale, outer scale, semantic scale etc. While there is improvement in performance than several baselines (mainly on PASCAL), the size of improvement is too small to convince the reader that the weighted fusion is a significant contribution. Due to the design of the attention model, as demonstrated in the individual weight map examples, this weight model tends to capture the relative trivial aspect of scale -- region/blob size. It does well on "aero,train, boat" etc, but it is significantly worse on "bike" than the skipnet type of aporoach.</p>

<p>can take a number of forms; a paper may be novel in terms of the method, the problem, the theory, analysis for an existing problem, or the empirical evaluation. If you think there is an error in the paper, explain in detail why it is an error. If you think the experimental evaluation is insufficient, remember that theoretical results/ideas are essential to CVPR and that a theoretical paper need not have experiments. It is <i>*not*</i> okay to reject a paper because it did not outperform other existing algorithms, especially if the theory is novel and interesting. It is not reasonable to ask for comparisons with unpublished, non peer reviewed papers (e.g. ArXiv) or papers published after the CVPR'16 deadline.</p>	
<p>Preliminary Rating. This rating indicates</p>	<p>Weak Reject</p>

to the area chair, to other reviewers, and to the authors, your current opinion on the paper. Please use 'Borderline' only if the author rebuttal and/or discussion might sway you in either direction.	
Preliminary Evaluation. Please explain to the AC, your fellow reviewers, and the authors your current opinion on the paper. This explanation may include how you weigh the importance of the various strengths and weaknesses you described above in Q1-Q3. Please summarize the key things you would like the authors to include in their rebuttals to facilitate your decision making. There is no need to summarize the paper.	I was hoping to like the paper, but was suspicious about the simple weighted feature idea, and finally gave it up upon the small gain :(
New exciting ideas. CVPR'16 would like to draw attention to papers that explore highly	No

innovative ideas, novel problems, and/or paradigm shifts in conventional theory and practice. Such papers may not be "complete" in the "traditional" manner in the sense that it may not be possible to have experimental results comparing other related efforts or that they may not have large, publicly available data sets to be used for performance comparison. However, we expect these papers to be visionary by nature. Should this paper be considered under "new exciting ideas"?	
Reproducibility. Could the work be reproduced from the information in the paper? Are all important algorithmic or system details discussed adequately?	Seems to be.
Confidence. Select: "Very Confident" to	Confident

stress that you are absolutely sure about your conclusions (e.g., you are an expert who works in the paper's area), "Confident" to stress that you are mostly sure about your conclusions (e.g., you are not an expert but can distinguish good work from bad work in that area), and "Not Confident" to stress that that you feel some doubt about your conclusions. In the latter case, please provide details as confidential comments to PC/AC chairs (point 7.)