ACCV 2016
13th Asian Conference on Computer Vision
November 20-24, 2016, Taipei, Taiwan

Reviews For Paper

Paper ID	663
Title	HF-FCN: Hierarchically Fused Fully Convolutional Network for Robust Building Extraction

Masked Reviewer ID: Assigned_Reviewer_1

Review:

Question	
Paper Summary. Please summarize in your own words what the paper is about.	This paper proposes Fully ConvNet based method for building extraction from remote sensing images.
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 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.	They build a Fully-ConvNet based semantic segmentation method for building extraction on remote sensing images.
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 insufficient, remember that theoretical results/ideas are essential to ACCV and that a theoretical paper need not have experiments. It is *not* 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 It is not reasonable to ask for comparisons with unpublished, non peer reviewed papers (e.g. ArXiv) or papers published after the ACCV'16 deadline.	1. The so-called "Building Extraction" task is actually a standard semantic segmentation task, and the method proposed in this paper is a semantic segmentation method. I didn't find any special points in their method on handling remote sensing images compared to common natural images. All those Fully ConvNet methods for semantic segmentation, like FCN (CVPR 2015), DeepLab (ICLR 2015), can be directly applied on this task. 2. No technical contributions. The proposed method mainly explores the middle-layer feature maps for dense prediction, for which they called hierarchically fused FCN. Actually use the middle-layer response is a very common technique for semantic segmentation, e.g, this technique is applied in those well-known methods like HyperColumn (CVPR 2015), skip connections in FCN, DeepLab. 3. The benefits they claim in the paper: "can take whole aerial images as inputs without warping or cropping and output building map directly" This benefit is just from the use of fully ConvNet, very standard method for semantic segmentation. 4. No comparison with the standard baseline FCN and DeepLab. Those standard method can be directly applied in this task. Their method is based on fully ConvNet, but we didn't see any comparison with FCN. I doubt the improvement is just from the use of VGG based Fully ConvNet.
Rating. This rating indicates to the area chair, to other reviewers, and to the authors, your current opinion on	Weak Reject

the paper. Please use 'Borderline' only if the author rebuttal and/or discussion might sway you in either direction.	
Justification of Rating. Please explain to the AC, your fellow reviewers, and the authors your 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.	no technical contribution lack of important baseline comparison
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 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.)	Very Confident

Masked Reviewer ID: Assigned_Reviewer_2

Review:

Question	
Paper Summary. Please summarize in your own words	The paper proposes a method for binary semantic segmentation of aerial images into buildings and background. It is based on a recent idea of fully-convolutional neural networks. The method does not have many hyper-

what the paper is about.	parameters, shows good precision-recall in comparison to prior art, and the inference is much faster.
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 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.	This is a solid application paper. It applies a reasonable technique to solve an important problem from the geospatial science domain. Novelty. The fully-convolution network [Long et al., 2015] seems to be a state-of-the-art semantic segmentation method. The paper builds on it by fusing feature maps from different scales to achieve some degree of scale invariance, which in turn requires solving several technical issues. Technical correctness. The proposed architecture and loss function are relatively simple and described well enough, so the results should be reasonably reproducible (as much as it is possible in deep learning). Related work is described well enough. Experimental evaluation is sufficient to prove that the final method is faster and more accurate than the prior art, but can be improved in other aspects (see below).
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	The description can be improved, and the paper will benefit from additional experiments. It would be interesting to see the performance of different variants of the proposed system. Most importantly, I wonder if the additions to the FCN are the deal-breaker, i.e. would we see the drop in accuracy without the feature map fusion over scales? Also, the paper mentions that tuning the deconvolution layers does not improve the performance. What was the result in that case? The paper will benefit from experiments on more datasets, ideally from different domains. Can this layer fusion be useful for semantic segmentation in general? The paper does not mention how the introduced 1×1 convolution layers are initialized. Is training sensitive to initialization? I assume that the network has been trained end-to-end, i.e. once the weights had been initialized, all the trainable weights were being optimized simultaneously. Can the paper state that clearly if that is correct? Was SGD learning rate multiplier uniform for all weights (except for deconvolutions)? Why is zero-padding introduced only on the first layer? Does not it bias the training set for the convolutions on that layer? Is not it better to pad all the layers but with a smaller border (á la ResNet)? Otherwise, the later feature maps will still be downscaled, but with some noise close to the borders.

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 ACCV and that a theoretical paper need not have experiments. It is *not* 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 ACCV'16 deadline.	Minor issues: Line 156 says that the hand-crafted features are sensitive to input data. I would say the opposite: the learned features may overfit to the training set domain, and may not easily transfer e.g. to different camera/region. Line 219 states that the fusion helps to cope with variation in accuracy and occlusions. While it is obviously helpful with scale, that statement is not obvious and has to be proven / motivated. In line 221, it may be not clear that U1, U2, etc. are the result of fusion, not e.g. VGG feature maps. Please explain that in words. Line 474 claims that the method is not sensitive to border effects caused by patch selection, but is it really true given that the input patch is zero-padded? Besides, there are lots of grammatical mistakes. Here are some of them: * line 020: surpass ← surpasses, * line 021: reduce ← reduces, * line 026: softwares ← software, * line 034: much ← many, * line 038: probability ← probabilistic, * line 042: publicly ← public, * line 051: a ← an, * line 058: a ← an, * line 058: construct ← constructed, * line 064: literatures ← literature, * line 068: 'very close to the plot on the ground or road' ? * line 076: it ← is, * line 116: literatures ← literature, * line 139: from ← in, * line 142: a similar framework ← similar frameworks, * line 145: regions ← region, * line 152: is ← are, * line 154: "It is problematic to recognize a over-segmented region as part of buildings because terrestrial objects have huge variant appearances in real aerial scene. (2)" – several mistakes, * line 159: are not satisfied to ← do not satisfy, * line 164: huge ← hugely, * line 175: into ← by, * lines 198, 200, 204: denote ← denoted, * line 200: to the same size of input image ← to the size of input image, * line 219: is ← will be, * line 248: surprised ← surprising, * line 249: a ← an, * line 293: possibility ← probability? * line 355: comparing ← compare, * line 469: with ← of, * line 474: inconsistant ← inconsistent, * line 474: occurred ← occur, * line 476: is increased ← will increase. In some cases, the hyphenation is weird, e.g. lines 014, 142, 218, 325.
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.	Weak Accept

Justification of Rating. Please explain to the AC, your fellow reviewers, and the authors your 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.	This is a strong application paper. It applies a modern technique to an important practical problem. It should be of interest to at least some part of the community. However, the improvement is quite incremental, until the broader applicability is demonstrated.
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 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.)	Confident

Masked Reviewer ID: Assigned_Reviewer_3

Review:

Question	
Paper Summary. Please summarize in your own words what the paper is about.	Paper focuses on the problem of automatic building extraction from remote sensing images. This problem comes with its hassles. Occlusion,shadows and variety of buildings in the images makes this problem harder for CNN type approaches. To overcome this, they propose a novel CNN structure.
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 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.	The authors propose a novel CNN structure. Applying CNN to building extraction is not a new thing as they mentioned in the paper. But the other methods in the literature lack some aspects. Referenced paper [13] uses patch based framework but it crops the images and breaks the integrity of building images. HF-CNN model preserves the structure and at the same time decreases the computation time due to their integrated CNN structure. Paper is very well organized and clear. They represent enough experimental evalutaion and explanatory figures and show comparisons against the other mentioned CNN based approaches.
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	One weakness I could identify is related work section. Main focus of the paper

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 ACCV and that a theoretical paper need not have experiments. It is *not* 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 ACCV'16 deadline.	is a CNN based model. Related work section should more focus on other CNN based models in detail. Old approaches should not take more attention.
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.	Strong Accept
Justification of Rating. Please explain to the AC, your fellow reviewers, and the authors your 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.	Authors contribute a novel CNN structure for building extraction and they show promising results in comparison to other CNN based models. Overall the idea is well represented and experimented and it shows improvements over other models. The weakness I listed is not important and it does not affect neither novelty nor the success of their method.

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 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.)	Confident