We would like to thank the reviewers for their insightful and constructive comments. We are happy to see an appreciation of our idea and results. We will respond to the comments of each reviewer individually.

TO REVIEWER 1

[Q1.1]...I didn't find any special points in their method on handling remote sensing images compared to common natural images...

[A1.1]We are very sorry for our vague expression. We detail the differences in three aspects. Firstly, building extraction has much higher demand in precision of structure, thus we fuse the edge maps generated from conv1. Secondly, most of buildings have small size in remote sensing image, while Noh et al.[17] noted that FCN(CVPR,2015) has less ability to extract small object. We merge the response of conv layers with a set of small receptive fields by learned weight, our experiments show that this strategy has strong ability to handle tiny rooftops. Thirdly, several large-sized buildings are comprised of several parts with variant appearance, we directly integrate the coarse but strong semantic response (conv4,5) into net output, instead of CRF post-processing (like [12], DeepLab). Our experiment indicates that our architecture is more adaptive to building extraction than CNN+CRF framework [12].

[Q1.2]...use the middle-layer response is a very common technique for semantic segmentation...

[A1.2]Using the middle-layer response is a general idea, but not a special solution. HyperColumn(CVPR 2015) sums the response of middle layers, while we using linear weighting. If we only change the operation, our accuracy drops 0.24%. FCN(CVPR,2015) suffers from the gradient back-propagation problem, our architecture relieves it much. DeepLab needs dense CRF as post-processing, while our network can train end-to-end. In conclusion, we design a new variant of FCN which significantly improves building extraction accuracy and shortens time loss.

[Q1.3]...This benefit is just from the use of fully ConvNet, a very standard method for semantic segmentation...

[A1.3]We emphasize this because related methods [12,13] need cropping.

[Q1.4]...Their method is based on fully ConvNet, but we didn't see any comparison with FCN...

[A1.4]We conducted experiments for comparison with FCN. We will add more detailed discussions in our revision. We reproduced three variants of FCN (CVPR 2015), namely, FCN8s, FCN4 and FCN2s. We finetuned the network in our remote sensing dataset using author’s code. The accuracy of FCN8s, FCN4s, FCN2s are 62.10%, 61.72%, 61.93%, respectively. Our proposed method outperforms them on the building extraction problem. Our network dramatically surpasses Mnih’s [12] which has the same framework with DeepLab, we’ll verify it again.

TO REVIEWER 2

[Q2.1]…see the drop in accuracy without the feature map fusion over scales?

[A2.1]If we merge the outputs of single only scale (conv5 layers), the accuracy drops 11.91%.

[Q2.2]...tuning the deconvolution layers does not improve the performance. What was the result in that case?

[A2.2]The accuracy drops 0.06%. It is almost the same with our current architecture.

[Q2.3]...Can this layer fusion be useful for semantic segmentation in general...

[A2.3]Theoretically, our HF-FCN can be applied to general semantic segmentation problem. We will test our system in PASCAL VOC dataset in the future.

[Q2.4]...Is training sensitive to initialization? …Was SGD learning rate multiplier uniform for all weights...

[A2.4]1\*1 conv layers are initialized as zero. The results are quite similar if using Gaussian distribution or Xavier. The SGD learning rate is divided by 10 with the level increasing.

[Q2.5]…Is not it better to pad all the layers but with a smaller border…

[A2.5]We set zero-padding as 3 for all convolution layers in revised VGGNet. The accuracy drops 1.1%. If we increase the border, the result might be close to ours.

[Q2.6]Minor issues

[A2.6]Thank you very much for your suggestions and detailed comments on the writing. We will address this issue more clearly and correct all the grammar errors.

To REVIEWER 3

Thanks for your strong support.