Dear Prof. Markus Gerke,

We aim to propose an end-to-end framework to semantically segment high-resolution aerial images without post-processing. The network structure is modified based on Pyramid Scene Parsing Network (PSPNet) (Zhao, Shi et al. 2016). It is aimed to segment aerial images into impervious surface, building, low vegetation, tree, car and clutter/background based on RGB data only, using the ISPRS Vaihingen and Potsdam benchmark data sets. The general architecture is shown in Figure 1. It mainly comprises two parts, ResNet101 (He, Zhang et al. 2016) and Pyramid pooling module(Zhao, Shi et al. 2016).The ResNet101 is used to extract features, encoding the input 3-channel 393x393 RGB image to 2048-channel 60x60 feature maps. Pyramid Pooling Module is applied to extract features at multiple scales and upsample the feature maps to learn global contextual information by concatenating the multi-scale features and get the segmentation result. The general framework consists of multiple loss function between different network blocks.

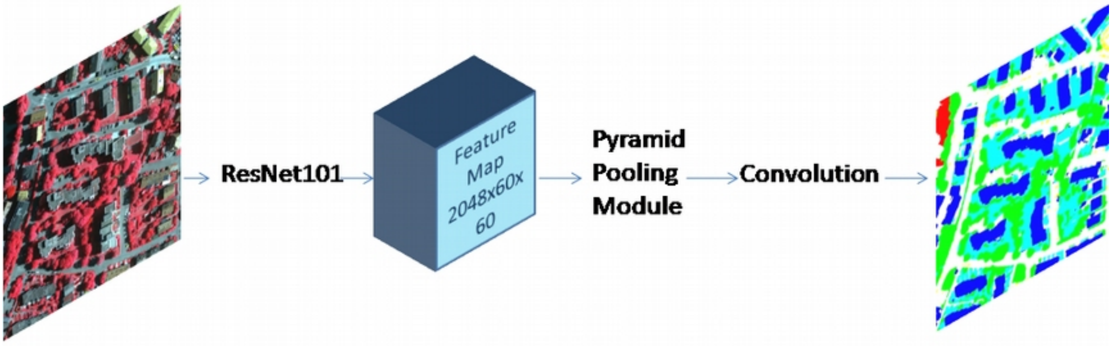

Figure 1. General architecture of proposed network

Looking forward to the evaluation results.

All the best!

Bo Yu
Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China

**References:**

He, K., X. Zhang, S. Ren and J. Sun (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Zhao, H., J. Shi, X. Qi, X. Wang and J. Jia (2016). "Pyramid Scene Parsing Network." arXiv preprint arXiv:1612.01105.

Bo Yu[a] , Haiping Yang[b],

[a]Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China
[b]College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou, P.R.China.