From Contexts to Locality: Ultra-high Resolution Image Segmentation via Locality-aware Contextual Correlation (Supplementary Material)

Qi Li[1],Weixiang Yang[1],Wenxi Liu[1*],Yuanlong Yu[1*],Shengfeng He[2]
[1]*College of Mathematics and Computer Science, Fuzhou University*[*]
[2]*School of Computer Science and Engineering, South China University of Technology*

1. Visualization of Contextual Attention

In Fig. 1 and 2, we demonstrate the effectiveness of contexts, while we show the attention maps (i.e. correlation) of the local, medium, and large contexts with regards to the local patch. As shown, the contexts of various scales guide the model to focus on different parts of the local patch. Accordingly, our model can estimate the corresponding weight maps for feature fusion, which appears to be consistent with the attention maps.

[*]Wenxi Liu and Yuanlong Yu are the corresponding authors.

2. Comparison with State-of-the-arts

We showcase more comparison results from DeepGlobe and Inria Aerial, against the state-of-the-arts methods, GLNet [1], CascadePSP [2], and FCN-8s [3], in Figs. 3, 4, and 5. As observed, our approach outperforms the competing methods in extracting semantic regions of various scales, especially those small and irregular regions.

3. Qualitative Results of Ablation Studies

We demonstrate the results with and without contexts in Fig. 6. Besides, we illustrate the comparison results with

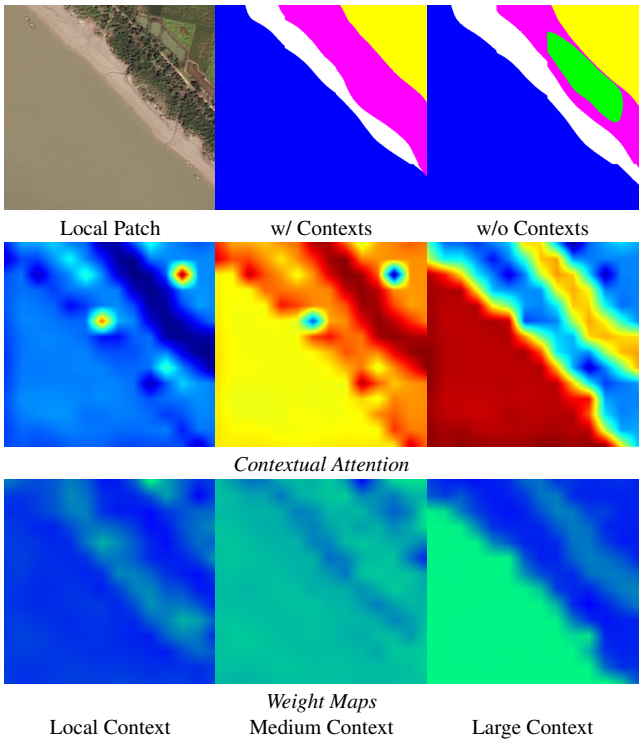


Figure 1. Illustration of an example with and without using contexts, as well as their contextual attention maps that imply the correlation between contexts and local patch and the corresponding estimated weight maps.

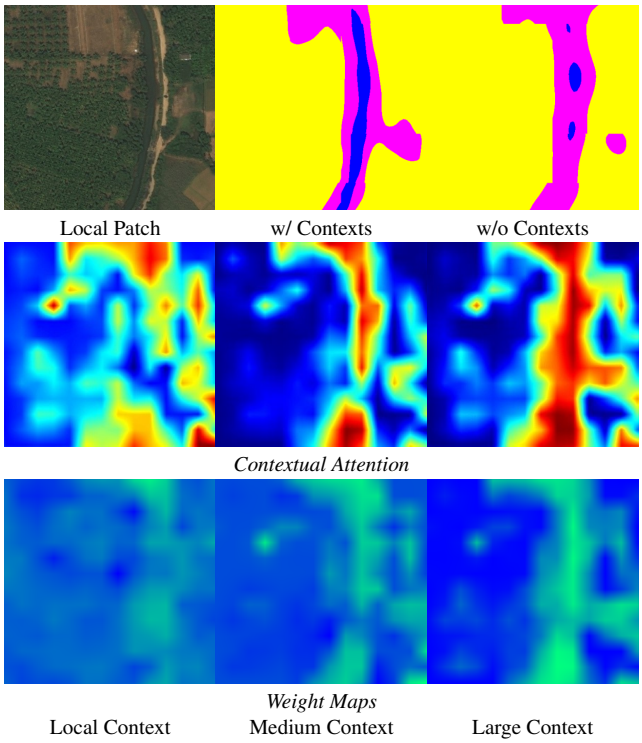


Figure 2. Illustration of an example with and without using contexts, as well as their contextual attention maps that imply the correlation between contexts and local patch and the corresponding estimated weight maps.

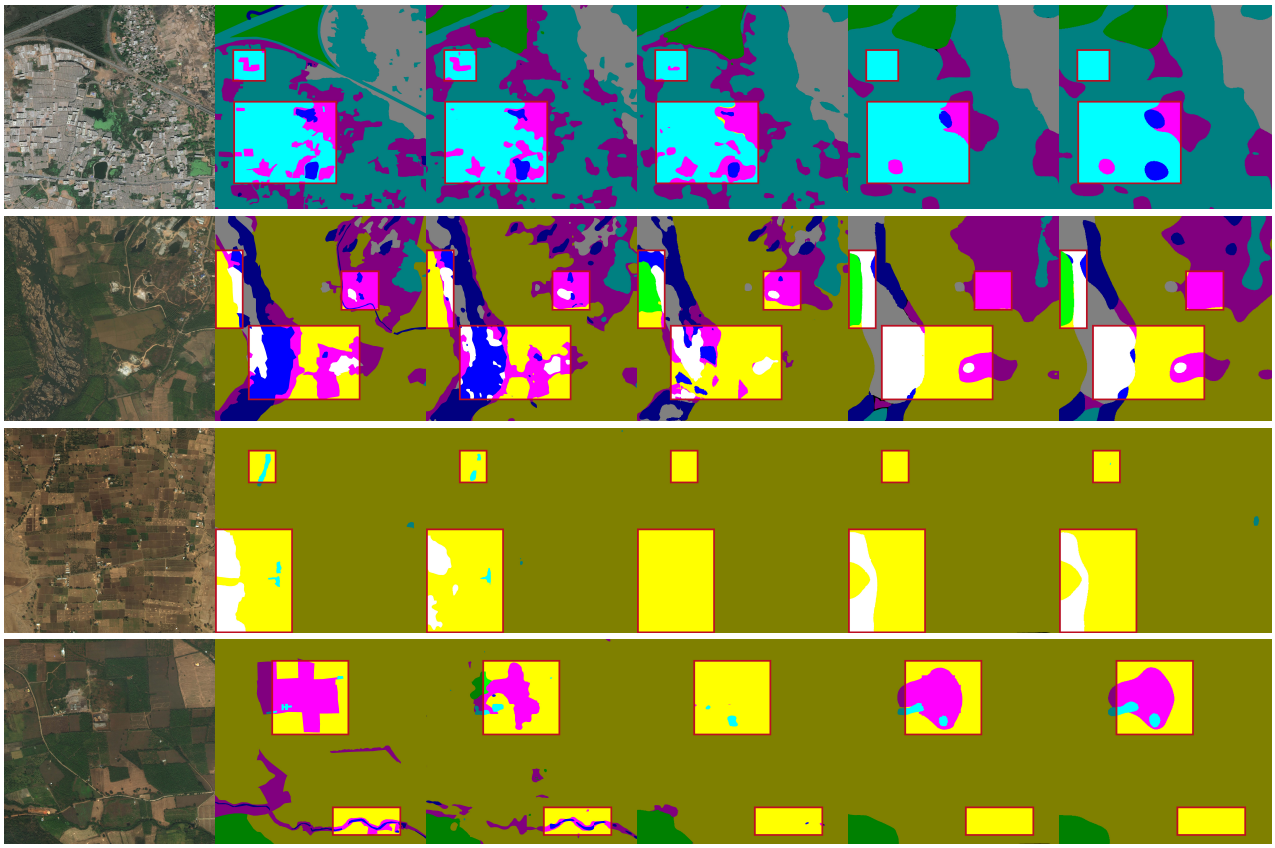


Figure 3. We illustrate several examples of semantic segmentation in ultra-high resolution images, comparing with the state-of-the-arts. In the figures, the semantic masks with varied colors represent different semantic regions. Particularly, cyan represents "urban", yellow represents "agriculture", purple represents "rangeland", green represents "forest", blue represents "water", and white represents "barren".

and without our fusion scheme in Fig. 7. As observed, the proposed schemes indeed improve the model performance.

References

[1] Wuyang Chen, Ziyu Jiang, Zhangyang Wang, Kexin Cui, and Xiaoning Qian. Collaborative global-local networks for memory-efficient segmentation of ultra-high resolution images. In *CVPR*, pages 8924–8933, 2019. 1, 2, 3

[2] Ho Kei Cheng, Jihoon Chung, Yu-Wing Tai, and Chi-Keung Tang. Cascadepsp: Toward class-agnostic and very high-resolution segmentation via global and local refinement. In *CVPR*, pages 8890–8899, 2020. 1, 2, 3, 4

[3] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, pages 3431–3440, 2015. 1, 2, 3, 4

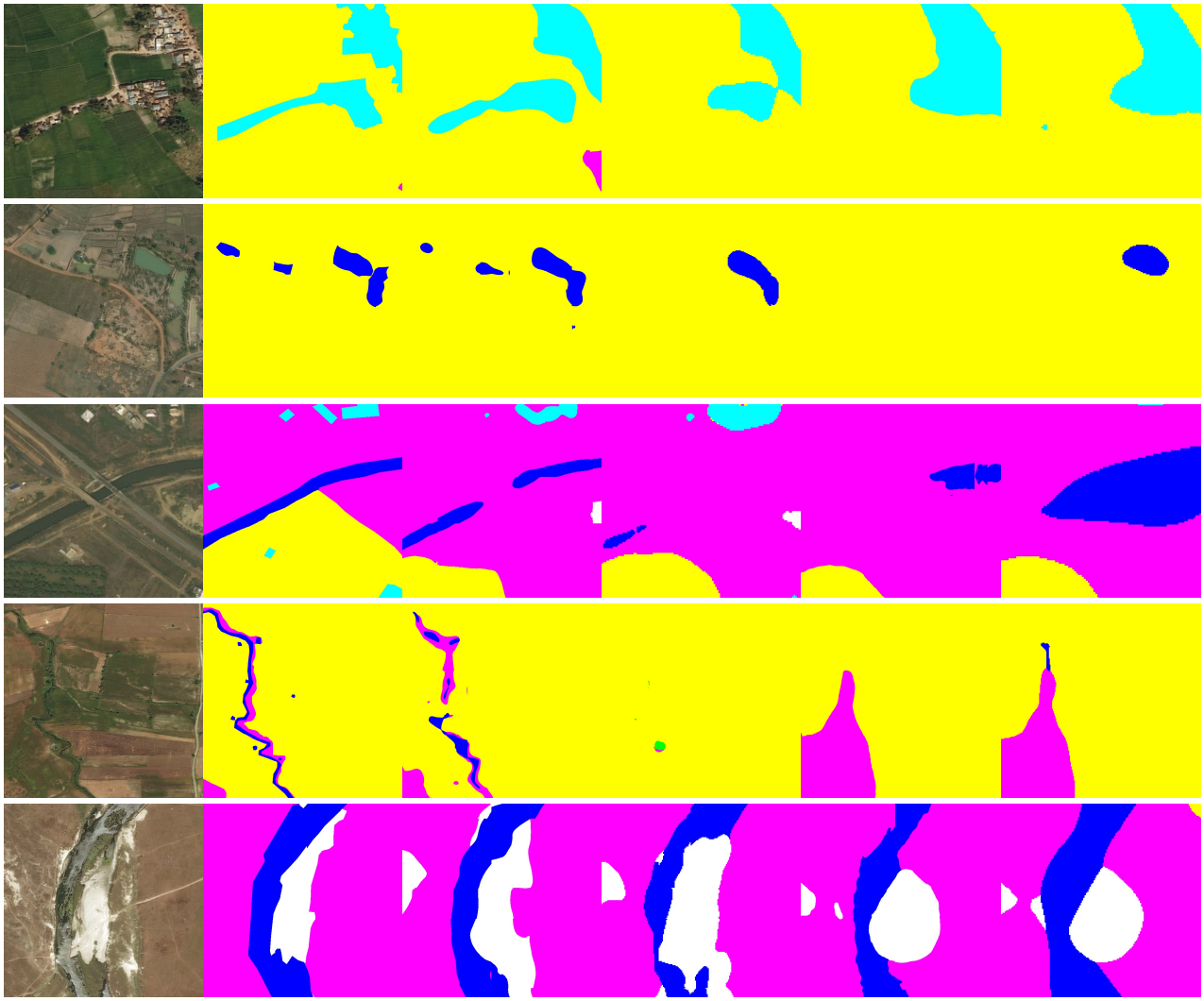


Figure 4. We illustrate several examples of semantic segmentation in ultra-high resolution images, comparing with the state-of-the-arts. In the figures, the semantic masks with varied colors represent different semantic regions. Particularly, cyan represents "urban", yellow represents "agriculture", purple represents "rangeland", green represents "forest", blue represents "water", and white represents "barren".

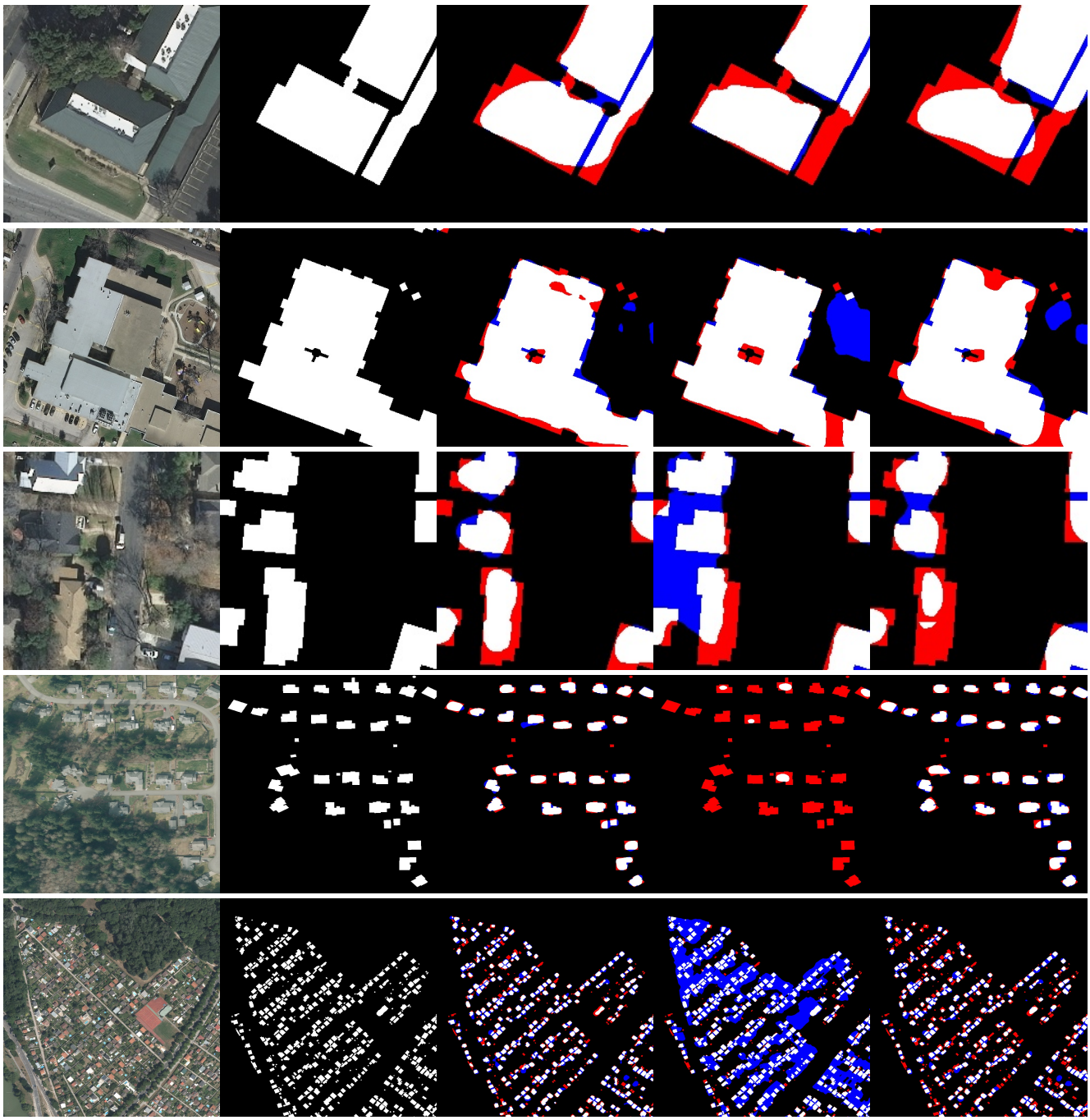
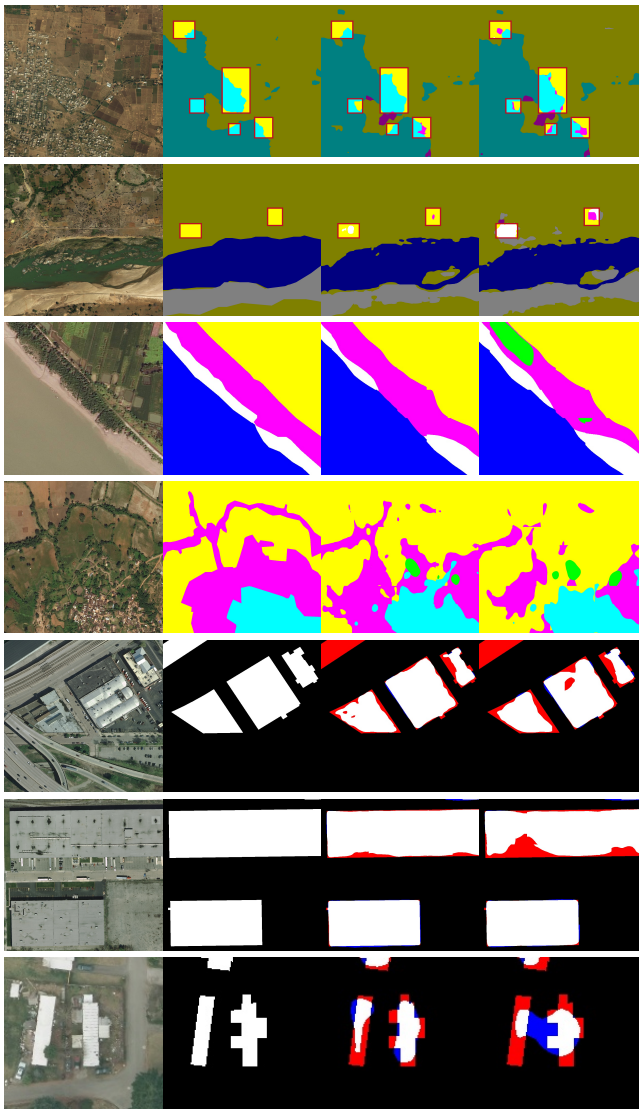
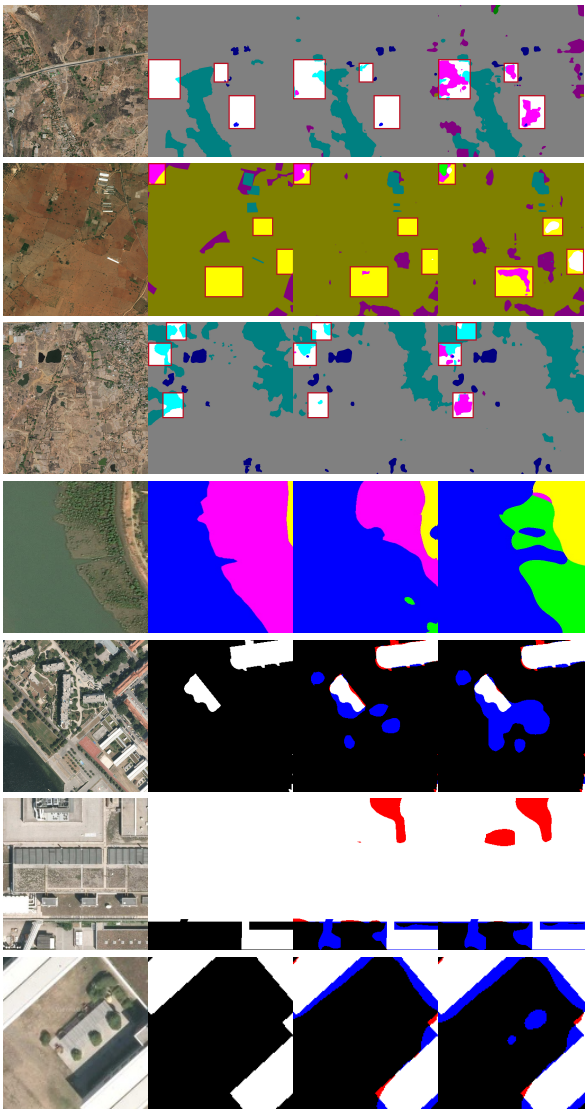


Figure 5. In the figures, the regions with color black represent background and the white regions represent the extracted buildings. Besides, the regions with color red represent False Negative and the blue ones represent False Positive.



Input Image Ground-truth w/ Contexts w/o Contexts

Figure 6. Examples show the efficacy of contexts. On the last three rows, the white regions represent True Positive, the black regions represent True Negative, while the red regions refer to False Negative and the blue ones refer to False Positive.



Input Image Ground-truth w/ Fusion w/o Fusion

Figure 7. Examples show the efficacy of our adaptive fusion. On the last three rows, the white regions represent True Positive, the black regions represent True Negative, while the red regions refer to False Negative and the blue ones refer to False Positive.