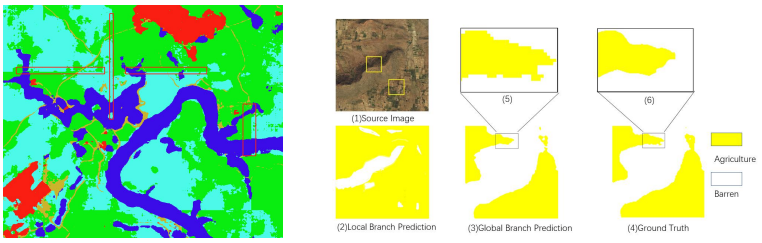
UHRSNet: A Semantic Segmentation Network Specifically for Ultra-High-Resolution Images

Lianlei Shan[1], Minglong Li[1], Xiaobin Li[1], Yang Bai[1], Ke Lv[1], Bin Luo[2], Si-Bao Chen[2], Weiqiang Wang
[1]University of Chinese Academy of Sciences, Beijing 100049, China
[2]MOE Key Lab of Signal Processing and Intelligent Computing
School of Computer Science and Technology, Anhui University, Hefei 230601, China
{shanlianlei18, lixiaobin161,liminglong18, baiyang17}@mails.ucas.ac.cn
{wqwang,luk}@ucas.ac.cn,{luobin,sbchen}@ahu.edu.cn

1. Problem

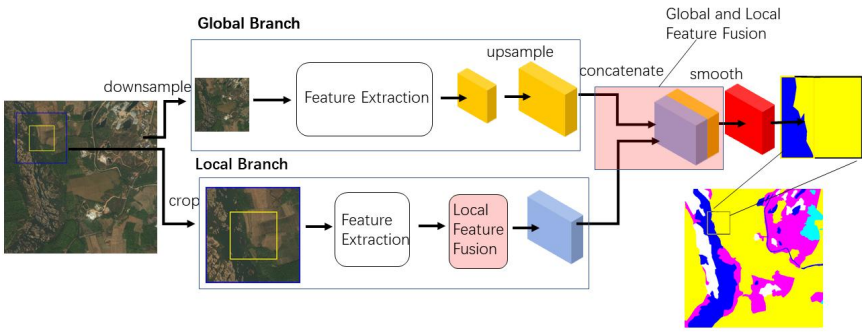


Since UHR images occupy too much memory, they cannot be directly put into GPU for training. Previous methods are cropping images to small patches or downsampling the whole images. Cropping and downsampling cause the loss of contexts and details.

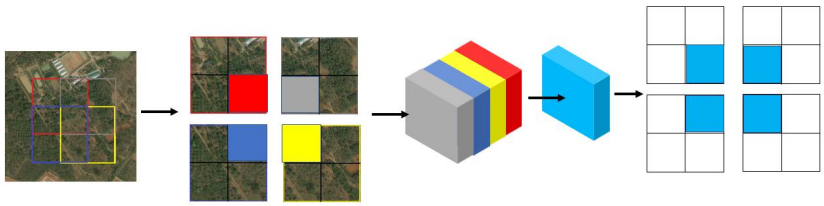
2. Contributions

(i). We improve the global and local feature fusion method, to make it efficient and handy. Improved fusion method can achieve comparable results with only one-time fusion.
(ii). We for the first time propose one fusion approach called local feature fusion, which can make patches get information from surrounding patches. Local feature fusion is different from the previous fusion in essence, which can make results better and more stable.
(iii). We achieve a remarkable improvement with these two fusions on Deepglobe dataset.

3. Methodology



OVerview



Local fusion

4. Results and analysis

TABLE I
COMPARISON OF DEEPGLOBE AND OTHER COMMON SEGMENTATION DATASETS.

Dataset	Image Size	Number of UHR Images
VOC2012 [16]	500 × 500	0
COCO-Stuff [17]	640 × 640	0
CamVid [18]	960 × 720	0
Cityscapes [19]	2048 × 1024	0
DeepGlobe	2448 × 2448	803

TABLE II
EXPERIMENTAL RESULTS:PERFORMANCE ON DEEPGLOBE TEST SET.

Model	Local (mIOU/%)	Global (mIOU/%)
ICNet [8]	35.5	43.0
UNet [9]	37.3	38.4
PSPNet [20]	53.3	56.6
SegNet [4]	60.1	61.2
FCN-8s [3]	63.1	68.0
DeepLabv3+ [21]	63.6	65.5
GLNet [7]	68.9	69.4
UHRSNet (ours)	**71.4**	**72.0**

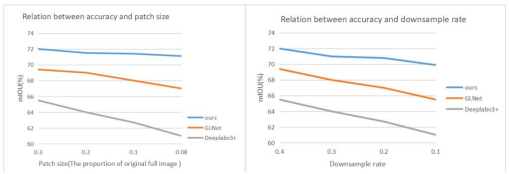


TABLE III
ABLATION STUDY:PERFORMANCE ON DEEPGLOBE TEST SET.

Model	Result (mIOU/%)
only local branch w/o fusion	57.3
only local branch with fusion	63.1
only global branch	66.4
cooperative both branches w/o local fusion	70.9
cooperative both branches with local fusion	72.0

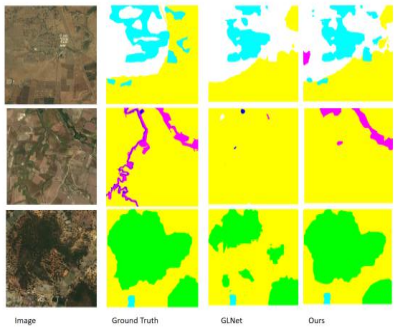


Fig. 6. Examples of segmentation results on Deepglobe. The second row is difficult case, and the third row is simple case and the first row is medium difficult. Our method improves results in the three cases compared with GLNet.

Important Reference

W. Chen, Z. Jiang, Z. Wang, K. Cui, and X. Qian, Collaborative global-local networks for memory-efficient segmentation of ultra-high resolution images, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 8924 8933