Ultra-High Resolution Segmentation with Ultra-Rich Context: A Novel Benchmark

Deyi Ji, Feng Zhao, Hongtao Lu, Mingyuan Tao, Jieping Ye

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Problem Definition and Contribution

Goal

Presenting a novel large-scale dataset for Ultra-High Resolution (UHR) segmentation, namely URUR (Ultra-High Resolution dataset with Ultra-Rich Context) , and a more efficient and effective framework, namely WSDNet.

Motivation

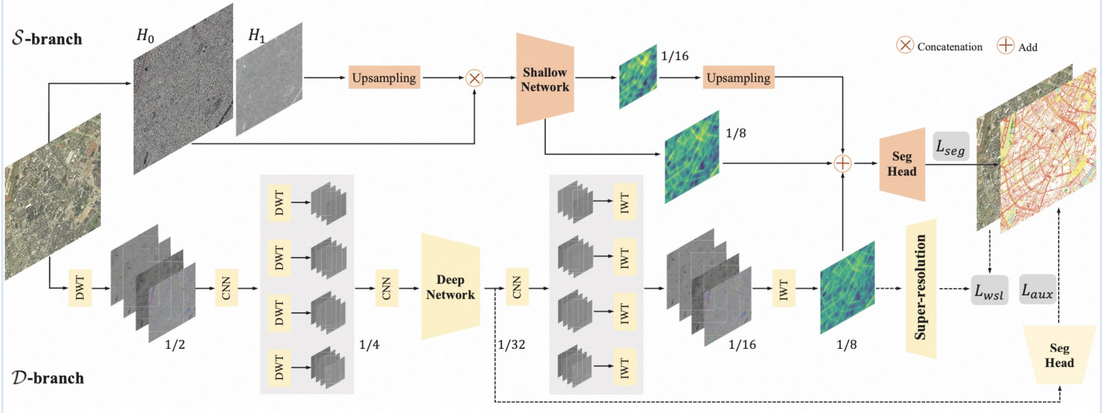
- A large-scale dataset covering a wide range of scenes with full fine-grained dense annotations is urgently needed to facilitate the field of UHR segmentation.
- Proposing a more efficient framework to effectively balance the memory occupation and accuracy when the image resolution grows to ultra-high

Key Contributions

- URUR dataset: a novel large-scale dataset covering a wide range of scenes with full fine-grained dense annotations, which is superior to all the exiting UHR datasets.
- WSDNet is proposed to preserve more spatial details with multi-level DWT-IWT, and a Wavelet Smooth Loss is presented to reconstruct original structured context and texture distribution with the smooth con- strain in frequency domain.
- Statistics and experiments demonstrate the superiority of URUR and WSDNet. WSDNet achieves state-of-the-art balance among accuracy, memory and inference speed on several UHR datasets.

Framework

Overview



Main Idea

- Shallow Branch (upper): the input image is decomposed into two subbands with Laplacian pyramid, which are then concatenated and fed into a shallow network to extract full-scale spatial details.
- Deep Branch (lower): the input image is down-sampled with two-level Discrete Wavelet Transform and then fed into the deep network to harvest high-level category-wise context.
- Multi-level Discrete Wavelet Transform (DWT) and Inverse Discrete Wavelet Transform are naturally integrated to release computation burden while preserve more spatial details in the deep branch. Thus heavy feature fusion modules can be removed for higher inference speed.
- The Wavelet Smooth Loss (WSL) is also de- signed to reconstruct original structured context and texture distribution with the smooth constrain in frequency domain.

Results

Visualization of URUR Dataset

Generic Models	mIoU (%)↑	Acc (%)↑	Mem (M)↓	FPS ↑
PSPNet [39]	32.0	-	5482	1.86
DeepLabv3+ [2]	33.1	-	5508	1.97
STDC [11]	42.0	-	7617	4.31
UHR Models				
GLNet [3]	41.2	71.5	3063	0.04
FCtL [21]	43.1	73.8	4508	0.03
ISDNet [14]	45.8	75.6	4920	6.31
WSDNet	**46.9**	**76.8**	4560	**7.13**