IAF EARTH OBSERVATION SYMPOSIUM (B1)
Earth Observation Data Management Systems (4)

Author: Mr. ZHIHUI ZHENG
National Key Laboratory of Science and Technology on Aerospace Intelligence Control, Beijing Aerospace Automatic Control Institute, China, bitzhengzhihui@163.com

Mr. Xiao Liping
China, xiaoliping2000@126.com
Mr. Bo Tang
Beijing Aerospace Automatic Control Institute, China, twinsbo@sina.com
Mr. Wei Yan
Beijing Aerospace Automatic Control Institute, China, yanwei0623@126.com
Mr. Chenrui Guo
National Key Laboratory of Science and Technology on Aerospace Intelligence Control, Beijing Aerospace Automatic Control Institute, China, g-ch-r@163.com
Mr. li fei
Beijing Aerospace Automatic Control Institute, China, lifeijita@163.com
Mr. SHUO WANG
Beijing Aerospace Automatic Control Institute, China, wangs2012@163.com

REAL-TIME SEMANTIC SEGMENTATION SYSTEM OF URBAN AREAS IN VERY HIGH RESOLUTION REMOTE SENSING IMAGES

Abstract

Semantic segmentation aims to assign land cover labels to each pixel in an image. Semantic segmentation for very high resolution (VHR) images in urban areas, plays a vital role in many important remote sensing applications, such as infrastructure planning, territorial planning and urban change detection, which is a long-standing research problem in remote sensing field. However, many confusing man-made objects and intricate fine-structured objects make it very difficult to obtain both accurate and real time labeling results. Specifically, on one hand, these confusing man-made objects with high intra-class variance and low inter-class variance bring much difficulty for accurate labeling. On the other hand, building deeper and larger convolutional neural networks (CNNs) is a primary trend for solving scene understanding tasks, while the extensive computational burden limits the usage of CNNs in mobile platforms. In this paper, we propose a novel lightweight architecture networks to balance the segmentation accuracy and implementing efficiency, adopting an asymmetric encoder-decoder architecture for real-time semantic segmentation. The core unit of encoder is a novel residual module that leverages skip connections and convolutions with channel split and shuffle. On the feature maps outputted by the encoder, global-to-local contexts are sequentially aggregated for confusing man-made objects recognition. In the decoder, an attention pyramid network (APN) is designed to extract dense features, where the attention mechanism is utilized to estimate semantic label for each pixel. The experimental results show our approach achieves best trade-off on ISPRS 2D Semantic Labeling Challenge dataset in terms of segmentation accuracy and implementing efficiency. The model has less than 1M parameters, and is able to run at over 50 FPS in a single GTX 1080Ti GPU with image size 1024X512. It provides possibility for real-time semantic segmentation of on-board processor on satellite.