BUILDING FOOTPRINT EXTRACTION FROM VERY HIGH RESOLUTION SATELLITE IMAGES: A COMPARISON OF OBJECT-BASED IMAGE ANALYSIS AND DEEP REINFORCEMENT CONVOLUTIONAL NEURAL NETWORKS

Rudolph Joshua Candare (1), Rolyn Daguil (1), Jaymer Jayoma (1)

[1] College of Computing and Information Sciences, Caraga State University, Ampayon, Butuan City, Philippines

Email: rucandare@carsu.edu.ph, rcdaguil@carsu.edu.ph, jmjayoma@carsu.edu.ph

Abstract: The automatic detection and footprint extraction of building from optical remote sensing images is very important for natural disaster reduction and assessment. It can be used for infrastructure damage assessment after disaster events. Traditional methods mainly are semi-automatic methods which require human-computer interaction or rely on purely human interpretation. A quite recent paradigm shift in satellite image analysis, which is the Object-based Image Analysis (OBIA) characterized by segmentation-first then-classification pipeline, has now been overshadowed by astounding developments in deep convolutional neural networks which solves supervised-learning problems in an end-to-end manner. In this work, we compare the performance and accuracy of OBIA with recent instance segmentation techniques based on convolutional neural networks applied to the problem of building extraction. We perform OBIA using the eCognition software and the Deep CNN using the Pytorch framework. We build on the Mask-RCNN architecture for object detection and instance segmentation. However, unlike the pre-trained models of Mask-RCNN which produces very smooth segmentations for most objects, the Deep CNN for building detection and segmentation should output predominantly straight edged polygons. To achieve this requirement, we have added modifications to the supervised Mask-RCNN architecture to include a reinforcement learning concept called actor-critic. Using the actor-critic approach, we let the actor network recurrently predict one instance mask at a time and utilize the gradient from a concurrently trained critic network. We set the state, action, and the reward in order to let the critic model the long-term effects of the current prediction. We perform our experiments on the publicly available xView dataset. The xView dataset are satellite images from WorldView-3 satellites at 0.3m ground sample distance. The original dataset has labels for over 1 million objects across 60 classes in

over 1,400 km^2 of imagery, however, in our study we only utilize the labeled building objects. In this work, we discuss the performance and accuracy differences between OBIA and Deep CNN for building footprint extraction from very high resolution images.

Keywords: Building Footprint Extraction, Actor-Critic Deep CNN, OBIA