S. Ganguly
NCTS# 25415-17 AGU 2016

DeepSAT: A Deep Learning Approach to Tree-cover Delineation in 1-m NAIP Imagery for the Continental United States

High resolution tree cover classification maps are needed to increase the accuracy of current land ecosystem and climate model outputs. Limited studies are in place that demonstrates the state-of-the-art in deriving very high resolution (VHR) tree cover products. In addition, most methods heavily rely on commercial softwares that are difficult to scale given the region of study (e.g. continents to globe). Complexities in present approaches relate to (a) scalability of the algorithm, (b) large image data processing (compute and memory intensive), (c) computational cost, (d) massively parallel architecture, and (e) machine learning automation. In addition, VHR satellite datasets are of the order of terabytes and features extracted from these datasets are of the order of petabytes. In our present study, we have acquired the National Agriculture Imagery Program (NAIP) dataset for the Continental United States at a spatial resolution of 1-m. This data comes as image tiles (a total of quarter million image scenes with ~60 million pixels) and has a total size of ~65 terabytes for a single acquisition. Features extracted from the entire dataset would amount to ~8-10 petabytes. In our proposed approach, we have implemented a novel semi-automated machine learning algorithm rooted on the principles of "deep learning" to delineate the percentage of tree cover. Using the NASA Earth Exchange (NEX) initiative, we have developed an end-to-end architecture by integrating a segmentation module based on Statistical Region Merging, a classification algorithm using Deep Belief Network and a structured prediction algorithm using Conditional Random Fields to integrate the results from the segmentation and classification modules to create per-pixel class labels. The training process is scaled up using the power of GPUs and the prediction is scaled to quarter million NAIP tiles spanning the whole of Continental United States using the NEX HPC supercomputing cluster. An initial pilot over the state of California spanning a total of 11,095 NAIP tiles covering a total geographical area of 163,696 sq. miles has produced true positive rates of around 88 percent for fragmented forests and 74 percent for urban tree cover areas, with false positive rates lower than 2 percent for both landscapes.

Authors

Sangram Ganguly
sangram.ganguly@nasa.gov
- NASA Ames Research Center

Saikat Basu
- Louisiana State University

Ramakrishna R Nemani
rama.nemani@nasa.gov
- NASA Ames Research Center

Supratik Mukhopadhyay
supratik@csc.lsu.edu

- Louisiana State University

Andrew Michaelis
andrew.r.michaelis@nasa.gov

- California State University Monterey Bay

Petr Votava
petr.votava-1@nasa.gov

- California State University Monterey Bay