Machine Learning for Global Land Cover Classification

A Radiant.Earth Expert Workshop to Further Global Development

14-15 June 2018
Omni Shoreham Hotel, Washington, DC

Background:

Radiant.Earth is developing an open source dataset of labeled satellite images, which will be hosted on the Radiant.Earth platform with a Creative Commons (CC) license. This dataset will enable fully-automated and dynamic Land Cover (LC) classification algorithms leading to a living open image library for machine learning. Imagery from Sentinel-2 satellites (at 10 m spatial resolution) will be labeled using deep learning and crowdsourcing to generate a human-verified training dataset.

This work will represent a major innovation in the field and will drive advanced, targeted and timely insights supporting progress in agriculture, food security, urban planning, conservation, and other areas relevant to global development and humanitarian response.

There are several LC class taxonomies developed by different institutions throughout the years. While these definitions are commonly used in different projects, they have discrepancies which makes it hard to combine datasets that use different taxonomies, or conduct cross-evaluation among different LC products. Therefore, there is a need to aggregate these taxonomies and generate a hierarchical inclusive LC taxonomy for global applications. The hierarchical structure will enable LC classification at different spatial scales (corresponding to spatial resolution of different satellite instruments) using the same terminology independent of the instrument.

Furthermore, to foster future applications of this new training dataset, Radiant.Earth is supporting a community standard development on best practices for use of machine learning in LC classification as well as commons for storing LC labels in the imagery metadata.

Purpose:

Radiant.Earth will host a two-day technical meeting to discuss and develop specifications for the global LC labeled training dataset by aggregating inputs from all participants. This meeting will consist of a half-day of presentations by participants sharing their successes and challenges and 1.5 days of working breakout sessions to develop documentation and standards related to three topics..

To achieve the goals of this meeting, three topic-specific working groups will address the following during the breakout sessions:

- *Working Group 1: Land Cover Taxonomy*
  This group focuses on developing the hierarchical LC class taxonomy. Participants will use current taxonomies as a baseline, and develop the globally inclusive LC class taxonomy which has a hierarchical structure.

- *Working Group 2: Machine Learning Algorithms*
  This group will review and document best practices in using ML for LC classification. A non-inclusive list of the topics to be discussed are:
    o How to achieve higher accuracies within each class, and between different classes?
    o What should be the metric for measuring training data diversity?
    o How to improve the quality of training data?
  (Additional topics may emerge from the morning session).

- *Working Group 3: Training data specifications*
  This group will develop specifications for LC labels to be stored in the imagery metadata. The goal would be to use the Spatio-Temporal Asset Catalogue (STAC) for this purpose, and design specifications for labels to be stored in imagery with Cloud-Optimized Geotiff (COG) format.

The following repository on GitHub is dedicated for coordination of the materials related to this meeting:

https://github.com/radiantearth/ml-workshops

Separate repositories for each working group will be established prior to the meeting, so participants can actively engage in the developments throughout the workshop.

Expected Outcomes:
- Generating an inclusive hierarchical taxonomy of LC classes at global scale;
- Defining specifications of the signature library for labeling and metadata storage;
- Reviewing, examining and documenting best practices for using ML with satellite imagery for LC classification; and
- Identifying knowledge gaps.

Participants:
- Approximately 20 attendees representing government, non-profit, academic, and commercial sectors will be invited.

- Participants will be leading experts in land cover mapping, machine learning, land use classification, remote sensing, and/or algorithm development.
- Participants are expected to devote two full days to this effort and actively participate in presentations, discussions, and breakout sessions.
- Participants are encouraged to engage in the follow-up conversations and taxonomy development efforts after the workshop by contributing to the GitHub repositories created for this purpose.