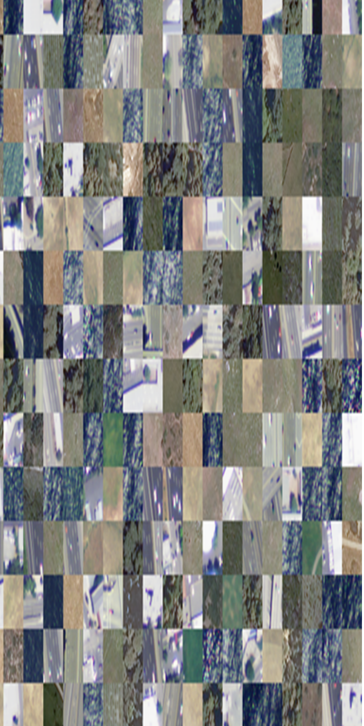


Large-Scale Land Cover Classification with Convolutional Neural Networks

Ekaba Bisong & Yajing Deng
School of Computer Science & Sprott School of Business
Carleton University

Agenda

Introduction

Problem Statement

- The most common approach in land use and land cover mapping is automated classification — throwing huge amounts of digital image data into an image classifier.
- Automated methods of image classification are based on spectral image data and are often plagued by problems of misclassification.

Problem Statement (contd.)

- Spectral reflectance of land surfaces — and more broadly spectral response patterns — measured by remote sensors may be quantitative but they certainly are not absolute.
- They may be distinctive but they are not necessarily unique. In reality, there is often extreme variability of spectral reflectance associated with various land cover types.
- This variability poses major challenges in mapping and analyzing land cover types based solely on their spectral properties.

Problem Statement (contd.)

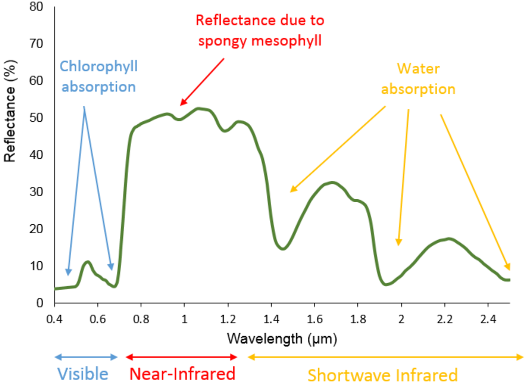


FIGURE – Spectral reflectance curve of vegetation

Project Focus

- This study designs a large-scale convolutional neural network to build a land cover classifier using the SAT-6 airborne dataset made available by the National Agriculture Imagery Program (NAIP).
- We discuss the impacts of incorporating state-of-the-art vision models into Remote Sensing and Geographic Information Systems pipelines to solving spatial problems especially in areas of low economic potential.

Project Focus (contd.)

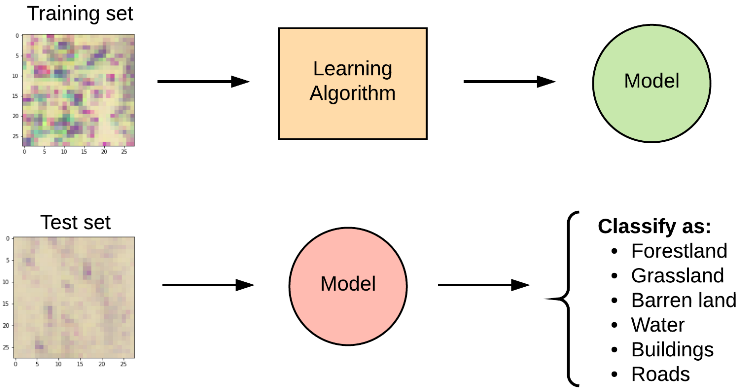


FIGURE – Land cover classification comprises of supervised, unsupervised and object-based techniques for identifying land cover types. The supervised method of land classification is the focus of this work - due to its more robust evaluation mechanisms for determining the performance of the learning model.

Relevant Background

Image Classification/ Object Detection

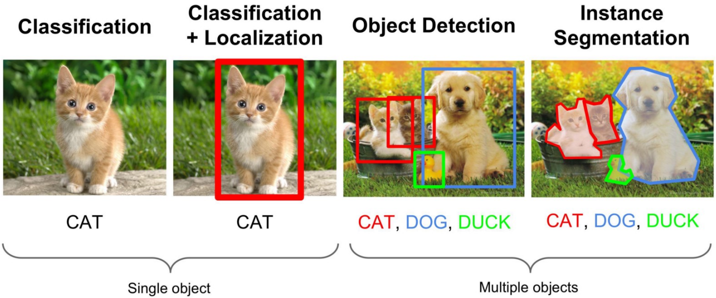


FIGURE – Difference between image classification, object detection and instance segmentation.

Image Classification/ Object Detection (contd).

Classification

- Image classification consists of classifying an image into one of many different categories

Image Classification/ Object Detection (contd).

- Similar to classification, localization finds the location of a single object inside the image.
- It can be combined with classification for not only locating the object but categorizing it into one of many possible categories.

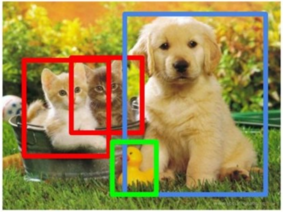
Image Classification/ Object Detection (contd).

Instance Segmentation

- In Instance Segmentation, we would want to not only find objects inside an image, but find a pixel by pixel mask of each of the detected objects.

Image Classification/ Object Detection (contd).

Object Detection

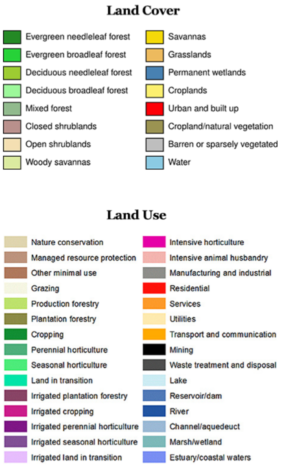


- Putting together localization plus classification we end up with the need for detecting and classifying multiple objects at the same time. Object detection is the problem of finding and classifying a variable number of objects on an image.

Land-Cover/ Land-Use Classification

- **Land Cover & Land Use.** Although the terms land cover and land use are often used interchangeably, their actual meanings are quite distinct.
- Land cover refers to the surface cover on the ground, whether vegetation, urban infrastructure, water, bare soil or other.
- Land use refers to the purpose the land serves, for example, recreation, wildlife habitat, or agriculture.

Land-Cover/ Land-Use Classification (contd.)



- It is important to distinguish this difference between land cover and land use, and the information that can be ascertained from each. The properties measured with remote sensing techniques relate to land cover, from which land use can be inferred, particularly with ancillary data or a priori knowledge.

Land-Cover/ Land-Use Classification (contd.)

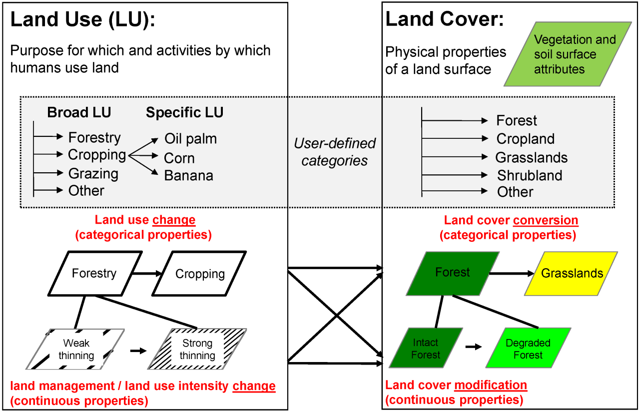


FIGURE – Conceptual sketch and examples of the relations between land use and land cover conversions and modifications.

Land-Cover/ Land-Use Classification (contd.)

- Mapping land cover over time requires an approach that generates consistently accurate maps over time for reliable change detection. Of two basic mapping approaches — computer automated classification and visual image interpretation — one needs to be chosen.
- This work is on automated classification.

GIS/ Remote Sensing

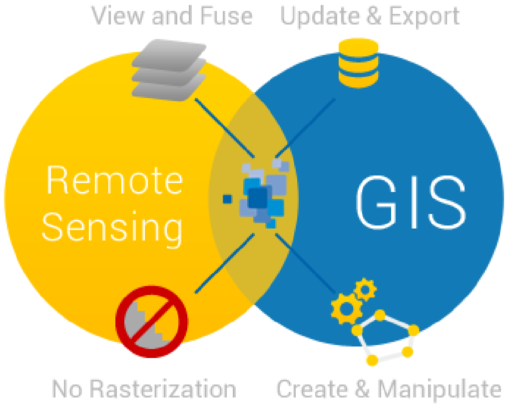


FIGURE – Remote sensing is the art and science of making measurements of the earth using sensors on airplanes or satellites. A geographic information system (GIS) is a computer-based tool for mapping and analyzing feature events on earth.

GIS/ Remote Sensing (contd.)

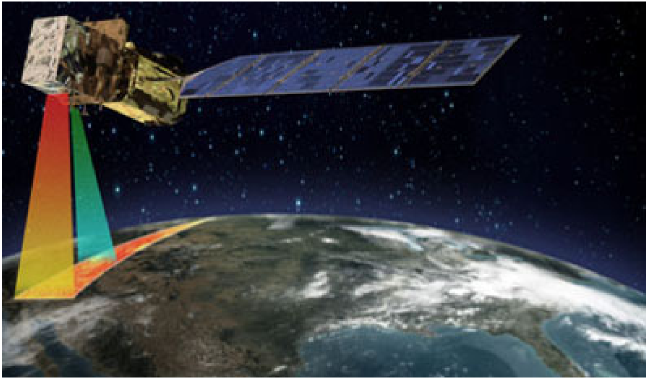


FIGURE – The Landsat 8 satellite images the entire Earth every 16 days in an 8-day offset from Landsat 7. Data collected by the instruments onboard the satellite are available to download at no charge from EarthExplorer, GloVis, or the LandsatLook Viewer within 24 hours of acquisition.

GIS/ Remote Sensing (contd.)

Band	Wavelength (micrometers)	Resolution (meters)
Band 1 – Coastal Aerosol	0.43 – 0.45	30
Band 2 – Blue	0.45 – 0.51	30
Band 3 – Green	0.53 – 0.59	30
Band 4 – Red	0.64 – 0.67	30
Band 5 – Near Infrared (NIR)	0.85 – 0.88	30
Band 6 – SWIR 1	1.57 – 1.65	30
Band 7 – SWIR 2	2.11 – 2.29	30
Band 8 – Panchromatic	0.50 – 0.68	15
Band 9 – Cirrus	1.36 – 1.38	30
Band 10 – Thermal Infrared (TIRS) 1	10.60 – 11.19	100
Band 11 – Thermal Infrared (TIRS) 2	11.50 – 12.51	100

FIGURE – The Landsat 8 spectral bands.

Convolutional Neural Networks

- Convolutional Neural Networks are a powerful artificial neural network technique. These networks preserve the spatial structure of the problem and were developed for object recognition tasks.
- They are popular because people are achieving state-of-the-art results on difficult computer vision problems.

Convolutional Neural Networks

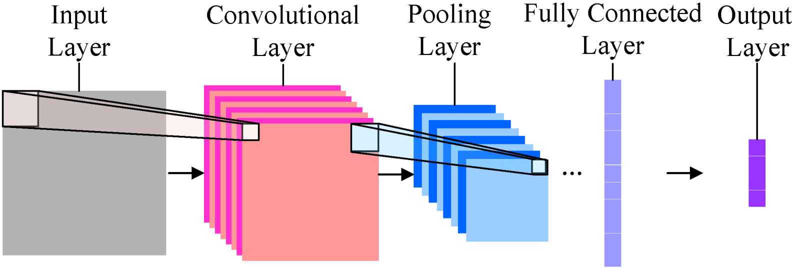


FIGURE – There are three types of layers in a Convolutional Neural Network : (1). Convolutional Layers. (2). Pooling Layers, and (3). Fully-Connected Layers.

The Convolutional Layer

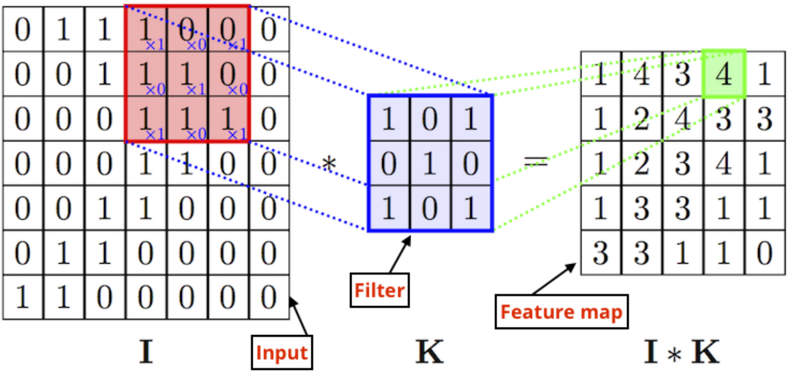


FIGURE – The Convolutional layers are comprised of filters and feature maps. The filters are essentially the neurons of the layer and the feature map is the output of a filter

The Pooling Layer

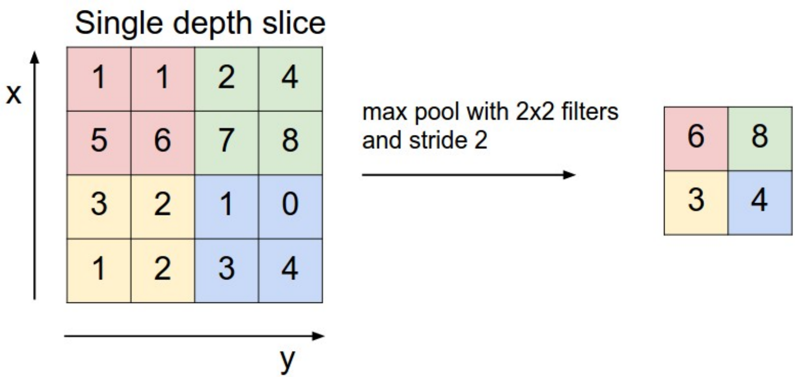


FIGURE – The Pooling layers down-sample the feature map. Pooling layers follow a sequence of one or more conv. layers. They consolidate the learned features. This example is a 2 x 2 MaxPool filter to aggregate the input to the Pooling layer.

The Fully Connected Layer

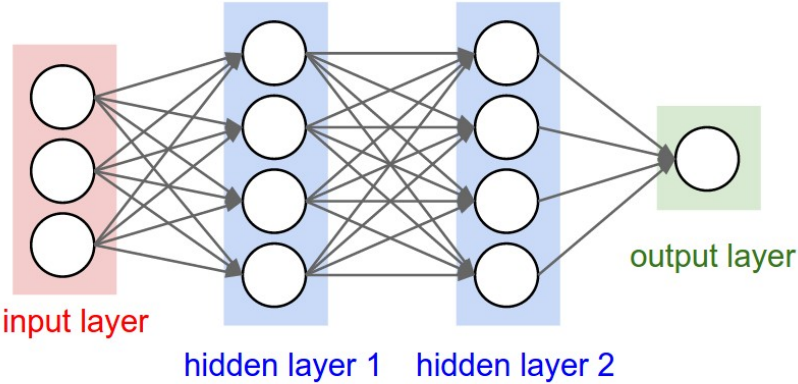


FIGURE – The Fully Connected layer are used at the end of the network after feature extraction and consolidation has been performed by the convolutional and pooling layers. They are used to create final nonlinear combinations of features and for making predictions by the network.

Infrastructure

Google Cloud Platform

- We use the Google Cloud Platform infrastructure for training our large-scale land classification vision system.
- In particular, we take advantage of Google Datalab and Google CloudML for distributed, serverless training.

The Google Cloud Platform Ecosystem

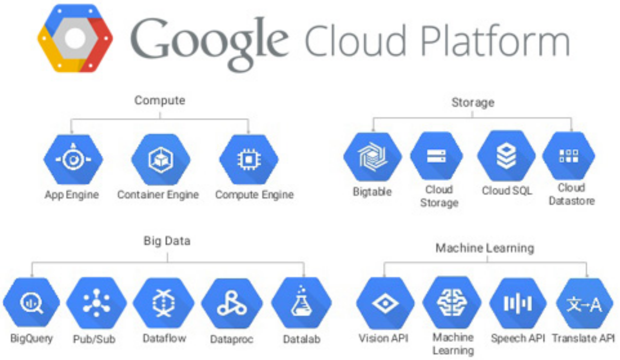


FIGURE – Google Cloud Platform is a suite of cloud computing services that runs on the same infrastructure that Google uses internally for its end-user products, such as Google Search and YouTube.

Google Datalab

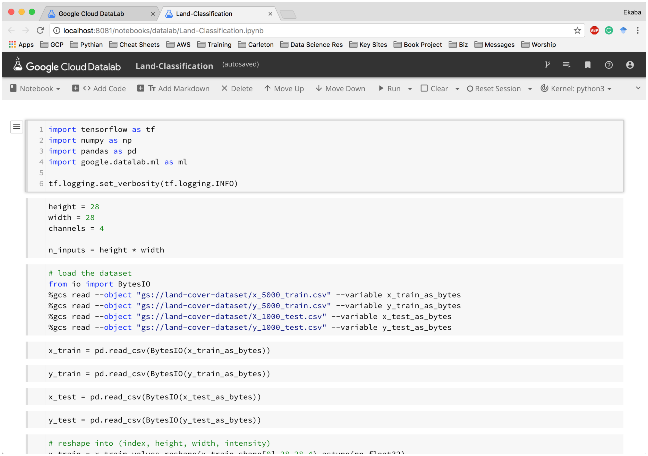


FIGURE – Datalab is Google's managed Jupyter notebook for rapidly prototyping machine learning models.

Cloud Machine Learning

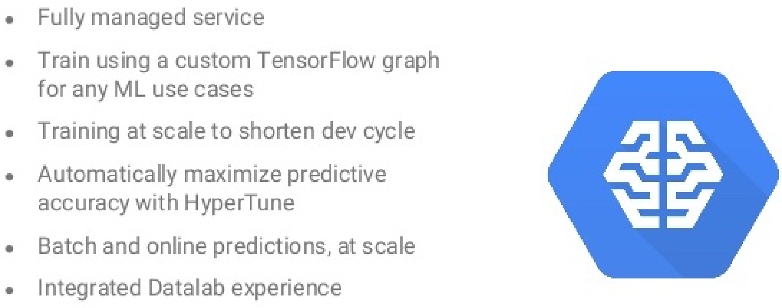


FIGURE – Cloud Machine Learning.

Training Infrastructure

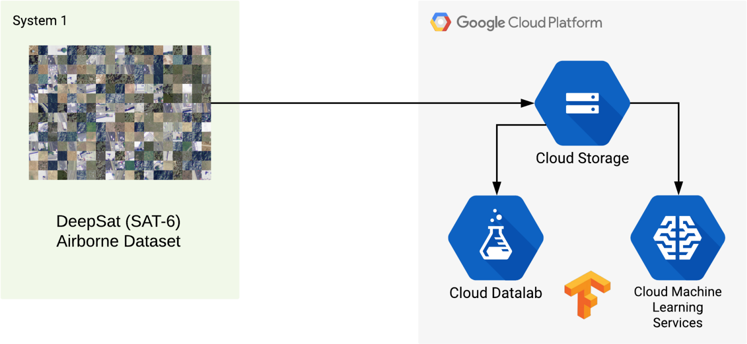


FIGURE – Designed model on Datalab & distributed training with Cloud ML.

Methodology

Dataset : SAT6 Airborne Dataset

- Each sample image is 28x28 pixels and consists of 4 bands - red, green, blue and near infrared.
- Contains 324,000 training and 81,000 test imageries.
- The six classes represent the six broad land covers which include barren land, trees, grassland, roads, buildings and water bodies.

Network Architecture

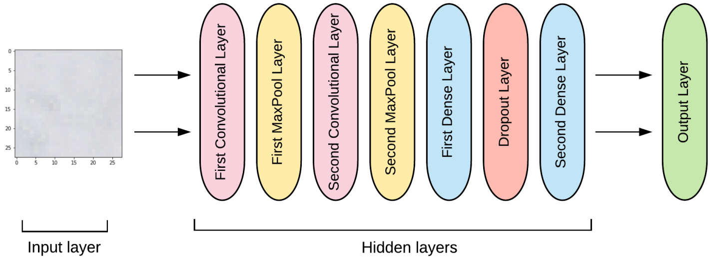


FIGURE – The CNN architecture contains 2 convolutional layers, 2 pooling layers, 2 dense layers, with one dropout layer in-between for regularization.

Model Hyper-parameters

- Conv. Layer #1 : A 32, 5x5 filter (extracting 5x5-pixel subregions), with ReLU activation function.
- Pooling Layer #1 : Max pooling with a 2x2 filter and stride of 2 (which specifies that pooled regions do not overlap).
- Conv. Layer #2 : Applies 64 5x5 filters, with ReLU activation function
- Pooling Layer #2 : Another max pooling with a 2x2 filter and stride of 2.
- Dense Layer #1 : 1,024 neurons, with dropout regularization rate of 0.5 (prob- ability of 0.5 that any given element will be dropped during training).
- Dense Layer #2 (Logits Layer) : 6 neurons, one for each land use class.

The Computational Graph

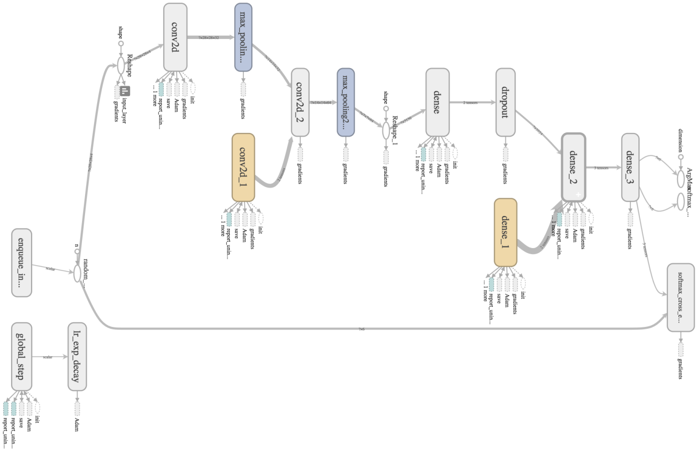


FIGURE – The TensorFlow Computational Graph.

Results

Precision

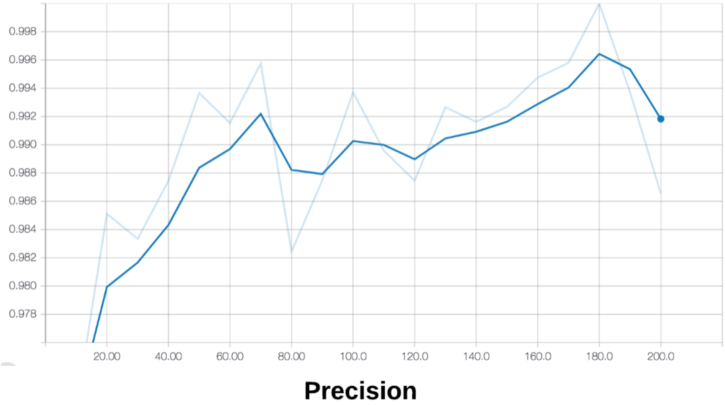


FIGURE – A high precision shows that most of the classified image maps are correct.

Recall

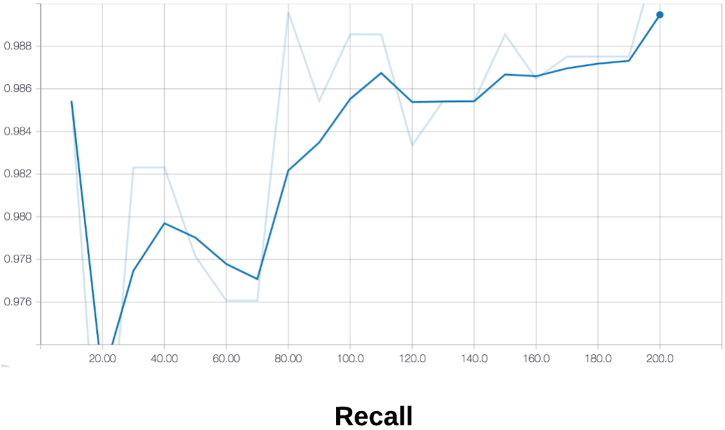


FIGURE – A high recall shows that a large fraction of the image dataset was classified correctly.

Accuracy

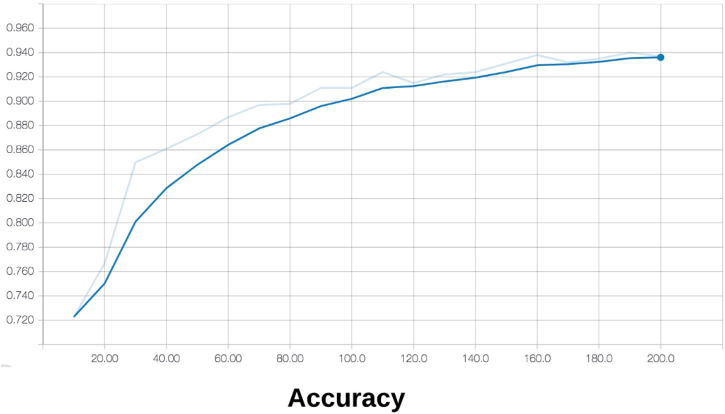


FIGURE – A high accuracy shows the strength of the overall classification performance of the model.

Loss

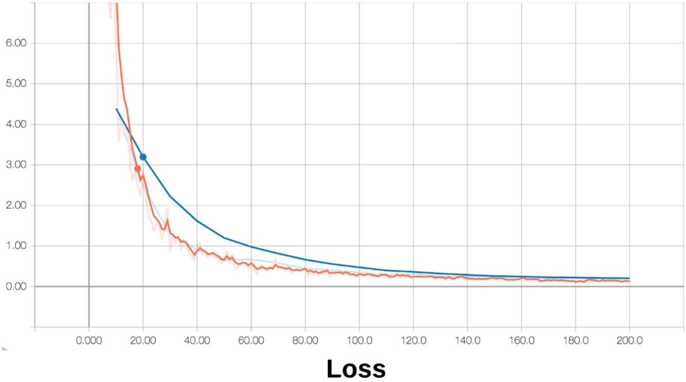
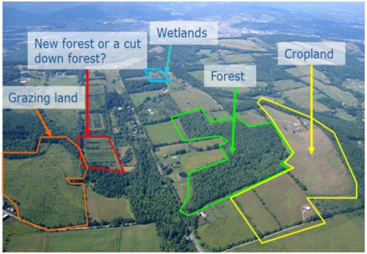


FIGURE – The loss or error function is constantly reducing at every iteration of the model. This is the function we want to minimize. It shows that our models weights or parameters are correctly learning.

Significance of the Study

Significance/ Implication



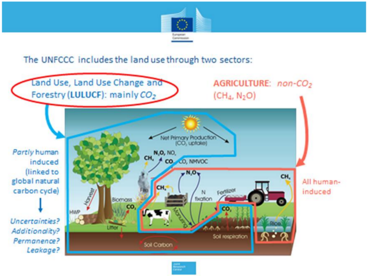


FIGURE – The analysis of remotely sensed data provides critical insights into the evolving human-environment relationship. In particular, the analysis of multispectral imagery is a key driver for estimating greenhouse emissions from Land Use, Land Use Change and Forestry (LU-LUCF).