CS230

Building a Deep Learning Model for Satellite Image Segmentation: Milestone Report

Sage McGinley-Smith
Department of Computer Science
Stanford University
Project Area: Computer Vision
sagems@stanford.edu

1 Introduction

Satellite data has long been used as a way to observe the Earth at a spatial and temporal scale. Earth observation via satellites has been increasingly used across the public and private sectors with agriculture, military, commercial, and other goals for the past several decades. In recent years, the availability of high quality, temporally regular, and public satellite image data has coincided with the rise of Machine Learning and Computer Vision. Advancements in image segmentation technology have been applied to satellite imagery to map landslides and waterways, analyze poverty within cities, and more, as in related literature [1] [2] [3]. Only recently has deep learning been used to segment images [4] [5]. The Mordecai Lab, which I work in at Stanford, is working with researchers at Centro Nacional de Alta Technologie (CeNAT) in Costa Rica to study the impacts of pineapple cultivation on dengue transmission. To study this, they need the data on pineapple plantation distribution from 2019-present. No dataset currently exists for 2020-present. Thus, this project aims to build a deep learning model that can identify pineapple plantations across the entirety of Costa Rica for the years 2020-2024 based on ground-truth data for plantation distribution in 2019. this dataset will allow for the investigation of the impacts of agriculture, land-use change, and climate variation on infectious disease transmission. Additionally, the advancements in resolution of spectral bands in satellite imagery have allowed for multi spectral image segmentation as opposed to traditional RGB image segmentation [6]. This project aims to build off of that work, showing the impact of incorporating multi spectral bands and indices in traditional deep learning image segmentation, outcomes that will help inform future research involving multi spectral imagery.

2 Dataset

Spectral Imagery: The Sentinel-2 Harmonized dataset is a publicly accessible earth observation product provided through the European Space Agency's (ESA) Copernicus program. This dataset includes optical imagery captured by the Sentinel-2A and Sentinel-2B satellites, spanning a temporal coverage from 2015 to the present. The dataset includes Level-2A products, which have undergone atmospheric correction to derive bottom-of-atmosphere reflectance values. This ensures consistency across images and reduces the impact of atmospheric effects like haze or aerosols. The dataset is georeferenced in the WGS 84 / UTM projection, with tiles organized into a Military Grid Reference System (MGRS) grid structure. It is distributed as part of the Copernicus Open Access Hub and made available through Google Earth Engine (GEE).

Although the dataset covers twelve spectral bands, ranging from the visible to the shortwave infrared regions of the electromagnetic spectrum, this project uses only five of those bands. Specifically, this project employs the red green, and blue bands (B4, B3, and B2 in the catalog), as well as the near infrared (B8) and short-wave infrared 1 (B11) bands. Bands B3, B4, B8, and B11 are used to compute spectral indices Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), and Normalized Difference Water Index (NDWI), for detailed land surface analysis. All baseline bands except B11 are at provided a 10m resolution, and B11 was resampled from 20m resolution using nearest-neighbor resampling to have a resolution of 10m. The data was then composited over a three month period to identify cloud-free images, which were then filtered down to four areas of interest. The areas of interest are in the northwest, northeast, central, and southwest region of the country and are pictured below.

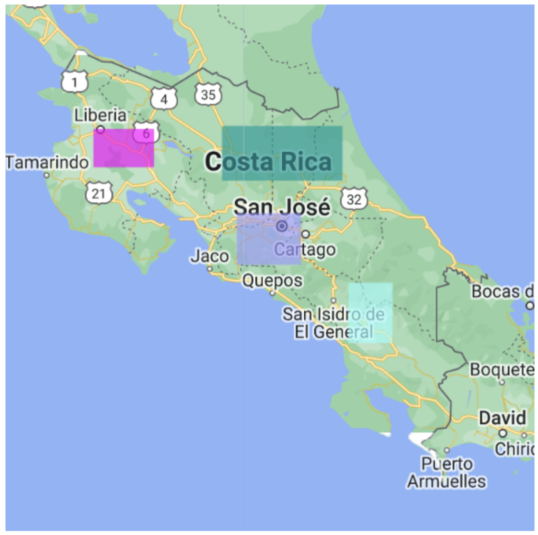

Figure 1: Quadrants for Training

Data Labels: The image masks used for segmentation were sourced from researchers at the Centro Nacional de Alta Technologie (CeNAT), one of Costa Rica's premier research institutions. Through a partnership with the Stanford Mordecai Lab, the CeNAT researchers shared the 2019 ground truth data, which shows the pineapple plantation distribution for the whole country in 2019. The data is a feature collection in Google Earth Engine, which was rasterized and re-projected using bilinear interpolation to have the same spatial projected and resolution as the Sentinel data.

Image Tiling and Storage: After exporting the Sentinel band images and the mask images separately from Google Earth Engine, the rasterio package was used in Google Colab to export the images to a Google Cloud Computing bucket as chips of size 128 x 128. The mask chips were exported and bucketed separately from the sentinel band chips, but use naming conventions that make them easy to match up later during model training.

3 Approach

Linear Baseline and Results: To form a baseline for comparison to deeper learning models, a logistic regression model was trained on a subset of the data. The model selected 25,000 pixels that were labeled as 0 (not plantation) and 1,000 pixels that were labeled as 1 (plantation present) from the southwest quadrant image, which contains several plantations. These ratios reflect the ratio of each class in the quadrant as a whole. the southwest quadrant was chosen because the image is high quality (low cloud cover) and covers both plantation and non-plantation areas. A model was run on both the full spectral band pixels (9 features) as well as simple RGB 3 feature pixels. The baseline results are as follows.

UNet Model and Training Strategy: Although the logistic model has high overall accuracy for both the full-feature and RGB datasets, a closer look at the precision and recall for the model reveals poor performance. These initial results present two hypotheses about the final project results. First, they suggest that using spectral band imagery (i.e. not just RGB features) enhances the functionality of the model. Second, it suggests that the pixel-based, shallow-network nature of logistic regression makes it difficult to have high precision and recall for this segmentation problem. The approach taken by this

```
Accuracy: 0.98
Classification Report:
              precision    recall  f1-score

           0       0.98      0.99      0.99
           1       0.83      0.60      0.70
```

Figure 2: All Features

```
Accuracy: 0.9625
Classification Report:
              precision    recall  f1-score

           0       0.96      1.00      0.98
           1       0.58      0.10      0.16
```

Figure 3: RGB Only

project is thus to build a UNet model for segmentation on entire image chips, aiming to show that a deeper network which considers both pixel values and relative pixel locations in the image (as a CNN does) will significantly outperform the pixel-based regression baseline. Currently, I have written the code to train a basic UNet model in Tensorflow on my data. The model takes images of shape (128, 128, 9), as well as ground truth labels of shape (128, 128, 1) and performs image segmentation on the image tiles. The loss function I plant to use is BCE, as the image segmentation task outputs a single-class mask, and it will effectively measure the difference between the predicted probability from the final sigmoid layer (between 0 and 1) and the actual binary label (0 or 1). My model will also use Adam optimization. As a reminder, the formula for BCE on an image segmentation task is below.

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)\right)$$

where:

- N is the total number of pixels in the image (or batch of images).
- y_i is the ground truth value for pixel i (usually 0 for background or 1 for foreground).
- $\hat{y}_i$ is the predicted probability for pixel i being in the foreground class.

Remaining Steps and Future Directions: The remaining steps for this project are to actually run training on the model that has been set up. This step is pending a meeting with the project mentor, as I want to confirm that I have structured the training correctly and implement any last feedback before I take the costly step of training the model. If there is time, this project also aims to tune a pre-trained model on the RGB bands of the images, providing a bases for comparison between multi-spectral and RGB imagery for CNNs. The final step will be model evaluation, including assessing precision and accuracy across the training areas. If time permits and the models is deemed high-performance enough, the model will be deployed on the unclassified data for 2020 as a proof of usefulness.

4 Code

Resource	**Link**
Data Source (Sentinel)	`https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_HARMONIZED`
Data Processing in GEE	`https://code.earthengine.google.com/645e761366100e67ba81f31b8a6540b0`
Data Tiling	`https://colab.research.google.com/drive/1D4K7t45S01-wwmHZ7dZoH9vogPulEzfH?usp=sharing`
Logistic Model	`https://colab.research.google.com/drive/15PRkwwH_VYkhsdaJLXnFg37ElMhEaYo9?usp=sharing`
Code Skeleton for U-Net Model	`https://colab.research.google.com/drive/1Hh045GIW1zwEXomkEq-9NaHphisHwpDP?usp=sharing`

References

[1] Zhang Y., Wu W., Qin Y., Lin Z., Zhang G., Chen R., Song Y., Lang T., Zhou X., Huangfu W., et al. (2020) Mapping Landslide Hazard Risk Using Random Forest Algorithm in Guixi, Jiangxi, China. *ISPRS International Journal of Geo-Information* **9**(11):695. https://doi.org/10.3390/ijgi9110695.

[2] Felton B.R., O'Neil G.L., Robertson M.-M., Fitch G.M., Goodall J.L. (2019) Using Random Forest Classification and Nationally Available Geospatial Data to Screen for Wetlands over Large Geographic Regions. *Water* **11**(6):1158. https://doi.org/10.3390/w11061158.

[3] [9] Zhao X., Yu B., Liu Y., Chen Z., Li Q., Wang C., Wu J. (2019) Estimation of Poverty Using Random Forest Regression with Multi-Source Data: A Case Study in Bangladesh. *Remote Sensing* **11**(4):375. https://doi.org/10.3390/rs11040375.

[4] Ayushi, Preetpal Kaur Buttar (2024) Satellite Imagery Analysis for Crop Type Segmentation Using U-Net Architecture. *Procedia Computer Science* **235**:3418-3427. https://doi.org/10.1016/j.procs.2024.04.322.

[5] Masolele, R.N., Marcos, D., De Sy, V. et al. (2024) Mapping the diversity of land uses following deforestation across Africa. *Scientific Reports* **14**:1681. https://doi.org/10.1038/s41598-024-52138-9.

[6] Vali A, Comai S, Matteucci M. Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data: A Review (2020). *Remote Sensing* **12(15)**:2495. https://doi.org/10.3390/rs12152495