Estimation of Crop Productivity From Multi-Sensor Fused Satellite Data” - Coding Test Report

Adam Yang | ayang115@terpmail.umd.edu

Design Outline

For this implementation, the problem was cast as a supervised, image-segmentation task. A UNet was trained on input images to predict between corn, soybean, and neither.

Data preprocessing

The data contains an image of size $(5, 5959, 9425)$ with a corresponding image of dimension $(1, 5959, 9425)$ containing pixel labels. For training, the label classes were reassigned to 1 for corn, 2 for soybean, and 0 for neither. The data was then divided into patches of size $(64, 64)$, which were partitioned into training, testing, and validation sets (totalling to 6835 training images, 3418 validation images, and 3418 testing images).

Data visualization

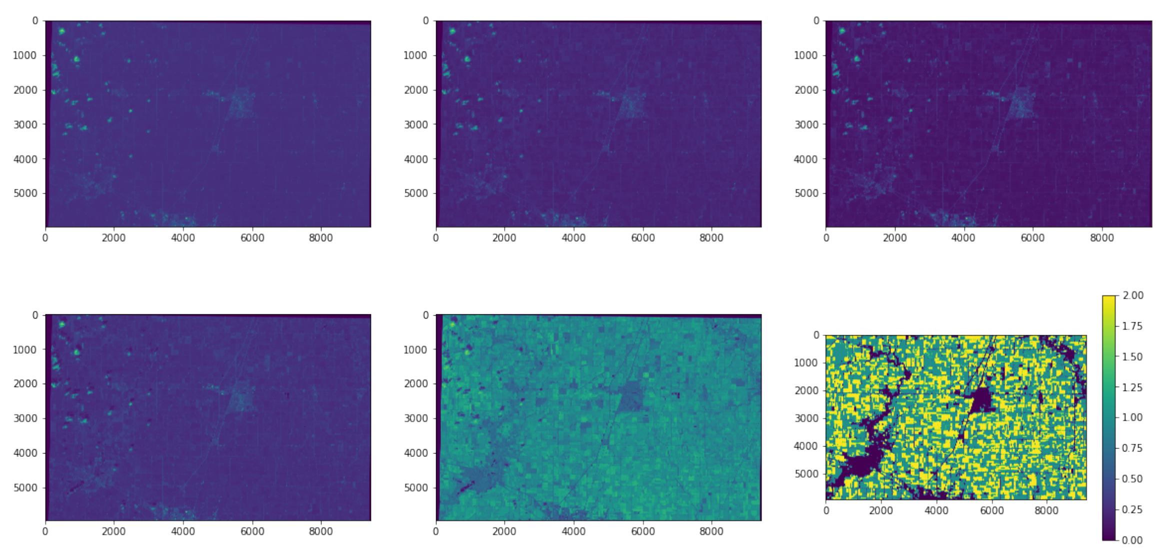

Figure 1: Visualization of training data and class labels

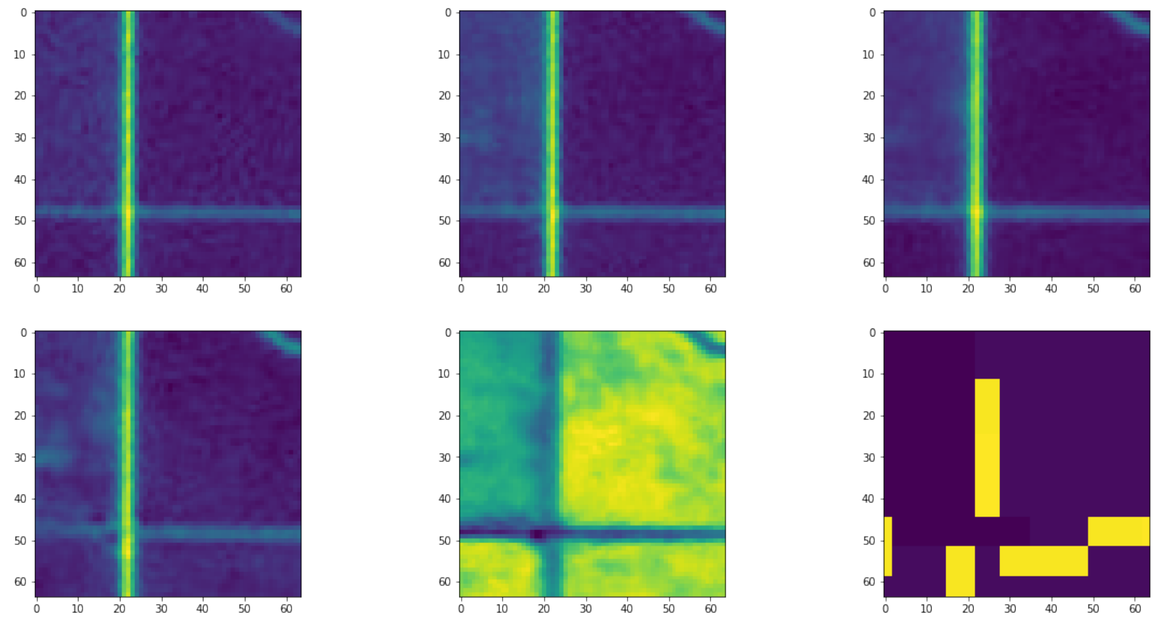

Figure 2: Visualization of a sample patch with data and labels

Model and Training
A UNet was chosen for its ability to capture spatial correlations along with its ability to run inference on images of varying size. A PyTorch implementation of the model was acquired from this repository. The model and training setups are defined as follows:

Model parameters:

1. **In channels**: 5 (1 for each input band)
2. **Classes**: 3 (also the number of output channels, 1 for each class)
3. **Activation**: Softmax (for multi-classification behavior)

Training parameters:

1. **Loss Function**: Cross Entropy Loss
2. **Optimizer**: Adam Optimizer
3. **Hyperparams**
   (a) Learning rate: $1e-4$
   (b) Batch size: 32

(c) Epochs: 30

Cross Entropy Loss was used given the multi-class nature of the task and to improve output probabilities. Adam optimizer was used to dynamically scale learning rates during training. To compute the loss, original labels were one-hot-encoded to an array of size $(3, H, W)$, then compared with model outputs of the same dimension.

Evaluation

The model was evaluated based on the Sørensen-Dice coefficient:

$$d(x, y) = \frac{2 \times (\# \text{ of true labels})}{\# \text{ of true labels} + \# \text{ of false labels}}$$

This metric best represents accuracy across the various class labels, with scores closer to 1 indicating high accuracy. To compute the score, the class with the highest probability at each pixel was selected then compared with ground truth labels.

Results

Loss/Dice curves:

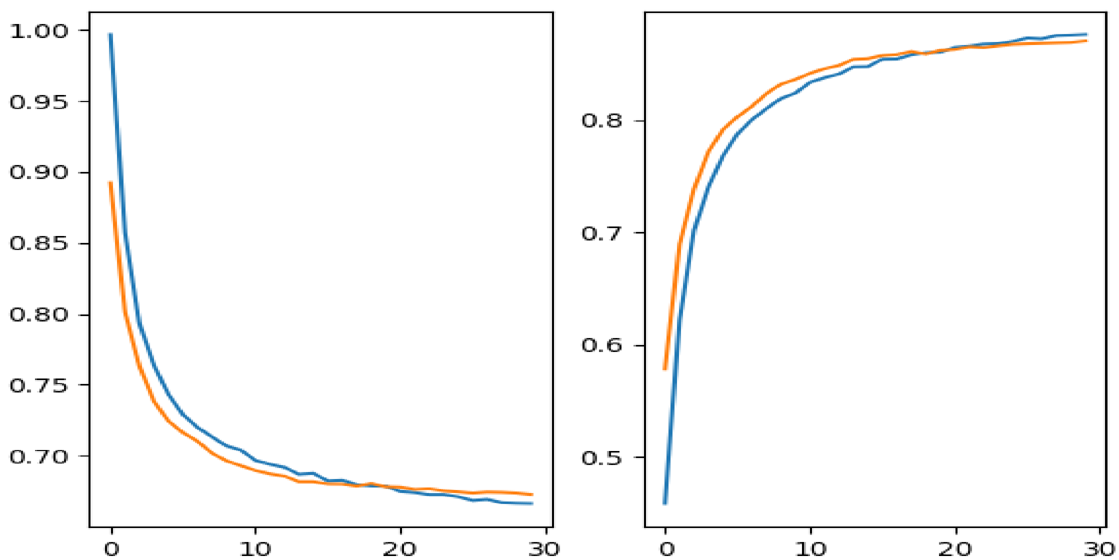

Figure 3: Training (blue) and validation (orange) loss (left) and Dice score (right) curves over 30 epochs

Testing Dice score: 0.87

Inference on South data:

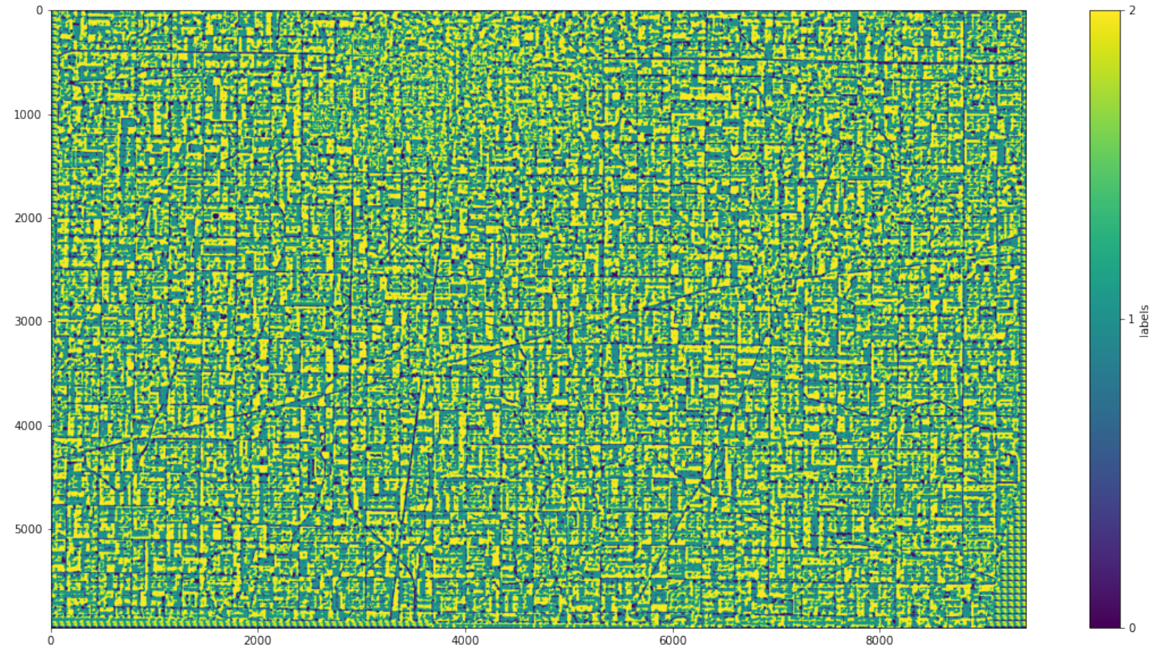

Figure 4: Model output on south image. Original data was cropped such that the dimensions were divisible by 32

Notes and Reflections

Notes on training:

1. Due to limitations in local hardware, Google Colab was used for training (taking roughly 30min for 30 epochs).
2. Training involved random horizontal and vertical flips on input images.
3. Upon visual inspection, the model seems able to identify crop boundaries along with patches of soybean/corn. However, it struggles with boundary artifacts (such as in the bottom left and right corners).
4. Training on various batch sizes (64 and 128) and input sizes (128 and 256) were attempted but exceeded memory limits on cloud GPU.

Possible improvements:

1. Training: providing a model with more diverse sizes for patches may allow for better generalization to fields of different sizes. In addition, adding more bands such as the normalized difference vegetation index or synthetic aperture radar may provide essential information for crop classification.

2. Model selection: transformer based architectures/generative models may prove better at generalizing across geographic regions and crop-types.

Relevant Links

1. **Code**: github

2. **Colab**: notebook

3. **Model**: smp models

4. **Dice implementation**: brain tumor segmentation

5. **UNet**: original paper