TreeNet: Deep U-Net for Image Segmentation

Andrew Kondrich and Isaac Kasevich

Motivation

TreeNet is an image segmentation network that seeks to outline trees from satellite images. The techniques used in this project have previously been shown success in segmenting medical images.

Image Segmentation is the process of partitioning an image into multiple 'segments' (sets of pixels), with the goal of simplifying the representation of the image.

We motivated our implementation, TreeNet, based on the impact of this still relatively novel application of deep learning, ranging from medical image processing to autonomous driving.

Data

Our dataset consists of 25 satellite images downloaded from the dataset of a Kaggle competition, *DSTL Satellite Imagery Feature Detection.* We split our data into 20 training images and 5 test images. Each raw image is 3600x3600 pixels and shows one square kilometer of satellite image data. Below is an example raw image:

Our output ground truth information came in the form of GeoJSON objects, which we transformed into pixel masks. Details discussed in data augmentation section.

References

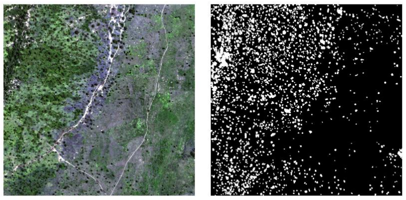
Ronnenberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation.
Castrejon, L., Kundu, K., Urtasun, R., Fidler, S.: Annotating Object Instances with a Polygon-RNN
https://www.kaggle.com/the1owl/html5-wkt-to-svg-3
https://github.com/zhixuhao/unet

Features and Data Augmentation

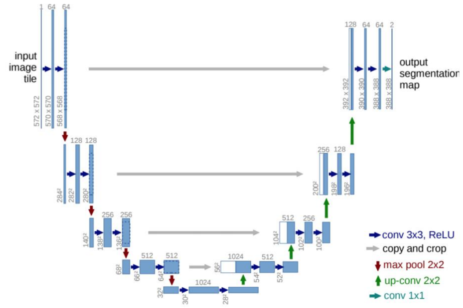
- Used augmented images as inputs, utilizing color shifts, stretching, random cropping and blurring.
- Generated output pixel masks from ground truth GeoJSON objects, which specified pixel coordinates of trees in the image.
- Ultimately trained on 510 images, which is sufficient for the U-Net architecture.

We found that inverted pixel masks (white trees, black everything else) worked best.

Training Data: A sample input, output pair of augmented image and output pixel mask from our training set shown below:



Model Architecture



Dice Loss

$$s_v = \frac{2|\mathbf{a} \cdot \mathbf{b}|}{|\mathbf{a}|^2 + |\mathbf{b}|^2}$$

Where a and b are the predicted and target images, respectively

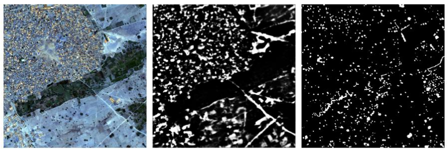
U-Net Architecture

- Model originally proposed by Ronenberger et. al., 2014.
- Successive 'contracting layers' down-sample image, skip connections to 'merge layers', which up-sample image.
- Dice (F1) Loss, shown above.
- Skip connections allow for localization, with high-res feature from early layers combining with abstract features from later layers for context.
- No fully connected layers; 1x1 convolution on output layer.

Results

After training on 13epochs, which took 10 hours, our model achieved -.4339 training loss and -.4379 loss on the validation set. We were satisfied with the results of our model. An example from the test set is shown below:

Input **Predicted** **Ground-Truth**

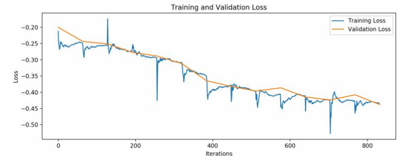


There is a slight distortion, but the model outputs an image visually similar to the ground-truth image.

Discussion

We saw our model's loss decrease drastically over the first few epochs. In early development, our team had issues getting stuck in local optima, where our output images would be empty masks despite low loss.

We ultimately solved this problem by inverting our ground-truth images, and changing our loss function from cross-entropy to dice-loss.



Future Work

Given 9 more months to develop this project, we would first look into more advanced data augmentation techniques to get more mileage out of our initial inputs. We would also hope to work on the second aspect of the Kaggle competition, which was to predict actual GeoJSON locations of trees in the image from the resulting pixel mask.