Forest Species Segmentation in Aerial Imagery of Wasatch Mountains Using 3D Convolutional Neural Network

Joshua Montoya

Motivation and Impact

Northern Utah has a delicate natural environment, small changes in which could result in substantial consequences for Utah residents and native species, as well as perennial migrating species. Forest fires contribute to hazardous air quality and endanger neighborhoods and animals. Mountain forests have a slight effect on local weather and serve as a great protection from premature snowpack melt off. The emergence of an endangered canopy could, therefore, lead to severe consequences on local water supply throughout the year. The canopy also has a symbiotic relationship with many undergrowth plants, which serve as food and shelter for many animal species. Plant roots also help prevent mudslides. Different species face different challenges. Some burn easily. Some are invasive, others vulnerable.

Determining the range of species over an area and tracking change over time can be useful in forest management and in monitoring local effects of climate change, allowing solutions to problematic changes to be found before critical effects occur. Accurately mapping forests over a large area can help forest managers accomplish this. The end goal of this project is to create a map of the Wasatch Range in Northern Utah that segments the forests into color-coded labels based on species.

Approach

The imagery that was used for training, and subsequently segmented, was downloaded from the United States Geological Survey (USGS) EarthExplorer portal: https://earthexplorer.usgs.gov/. The specific source dataset is aerial imagery, taken between June 27 and August 20, 2016, of the Wasatch Mountains in Northern Utah, as part of the National Agriculture Imagery Program (NAIP). This aerial imagery has a resolution of 1 meter, meaning every pixel in each image represents a square area on the ground with each side measuring 1 meter. Images contain 4 bands: a red channel, blue channel, green channel, and near-infrared (NIR) channel. In total, 144 images were downloaded via this portal for segmentation.

Landsat and Sentinel satellite imagery was initially considered for this project. Although the satellite imagery contains many more bands--Sentinel-2 has 13--the resolutions are much lower, with a best resolution of 10 meters. While this may be sufficient for many purposes, the Wasatch Mountains can contain several species of trees within a few feet of each other, and higher resolution imagery is expected to produce better results.

To accomplish the task of segmentation, a deep learning approach using a 3 dimensional (3D) convolutional neural network (CNN) was implemented. 3 of the 144 images were selected for training. Partial ground truth images were created for these to develop a model for segmenting every image in the collected dataset. These ground truth images were labeled manually as a part of this project and accounted for a large portion of the effort. The labeling process was assisted using other imagery of the same area from other periods of the year (autumn and winter), to help differentiate between species. Later, 3 more images were partially labeled exclusively for testing to gauge how well the model would scale up to make predictions on unseen images over the entire range.

The 3 training images were selected based on the point of the year in which they were taken and their location in the range. To elaborate, one was taken in June in the northern part of the range, one in July in the southern part, and one in August in the central part. This was done to create a better representation for all images in the dataset. The 3 test images were similarly chosen.

Wasatch Mountain forests can be primarily divided into four parts: The Bigtooth Maple, Quaking Aspen, Gambel Oak, and several species of evergreen trees. The evergreens consist almost entirely of fir, juniper, and spruce trees. As these would be difficult to differentiate for labeling, all evergreens were labeled as Conifer. There are many other species of tree in the Wasatch Mountains that were excluded from labeling. They were consequently mislabeled as one of the other 4 tree species in the final prediction map; however, this is not a big issue as other tree species combined likely cover well below 1% of the mountain range map area, mostly along the edges of streams and rivers at lower altitudes.

Grasslands and shrub areas were labeled as Other, and areas of little to no vegetation were labeled as None. In total, there are six labels in the training data:

1. Maple
2. Aspen
3. Oak
4. Conifer
5. Other
6. None

As mentioned previously, other NAIP images taken from other periods of the year (particularly autumn images) were used to assist in labeling the ground truth image. These images were not used in training, only for assistance in differentiating species for labeling. The model was trained on the summer images because most aerial images are taken in the summer.

Once the 3 input images were labeled for ground truth, transformations were performed on the images to create other images that were also used for training. The images were flipped horizontally to form one image, and vertically to form another. With these transformed images, a total of 9 images were used for training. Flipping the images is meant to simulate conditions in real imagery. Trees are found on slopes that face many different directions. Also, aerial images are taken at different times, causing shadows to be cast in different directions depending on the time of day.

To increase performance of the CNN model, the input data was augmented further using hyperspectral imaging (HSI) cubes. These "cubes" are of size S*S*4, where S is the manually set cube height and width, and 4 is the cube depth, equal to the number of bands in the input images. The cubes were passed over the entire training image. Each cube contains the data in the image array with a cube centered on each pixel. The image array was zero padded to allow for this. This results in one cube of size S*S*4 per pixel. For a TIF image of size (8000, 4000, 4), this would result in data with dimensions of (8000*4000, S, S, 4, 1). Cubes centered on pixels corresponding to 0 labels in the ground truth image were dropped for more compact input data. After all this processing, the model input data is ready.

The 3D CNN model consists of 4 convolutional layers, a flattening, and several dense and dropout layers. Many of the parameters were chosen through trial and error, but depth was prioritized for the convolutional layers.

After the model was set, segmentation predictions were made for all 144 images in the dataset. These prediction images were compressed for ease of display in a web app. Code was written to extract the bounding coordinate information from the downloaded text files and save it to a new text file. Finally, a web app was created to display this information on a zoomable map. The predicted values are only valid for the mountain areas. Predictions in residential areas along the edges of the map are not accurate.

Results

The training images were split into training and test data, then the model was trained. The resulting predictions on this initial test data were 96% accurate, with a test loss of 0.1049 and recall and precision for each class all being over 90%.

When the model was used to predict on the labeled portions of the unseen test images, accuracy on these images were 84, 74, and 73% accurate. This is a great result for one person with limited resources, and self-labeled the training data. There were consistent issues with maple and aspen trees being labeled as each other and oak and other being labeled as each other. The result would likely be improved by labeling more data, and possibly introducing one or more new training images. Additional augmentations could be helpful as well, such as image rotation and zoom.

The results are acceptable for a preliminary full mapping of the Wasatch Range.

Implementation Details

As mentioned previously, image data was downloaded via USGS EarthExplorer.

Labeling was accomplished using the open-source image editor GIMP.

All code was written in Python. The Tensorflow and Keras libraries were used to develop, run, and test CNN models. EarthPy, SciPy, Matplotlib, and OpenCV were used to read, write, and display image data. Pandas and NumPy were used for working with the data. Scikitlearn was used for creating train-test splits and evaluating model performance. The Seaborn package was used for generating heatmaps of the results. Pillow (PIL) was used for image compression.

Dash was used to create the web app, the dash-leaflet package was used for displaying the map, and Pythonanywhere was used for hosting. The custom domain name www.wasatch-forests.us was purchased for a few dollars.

For ease of labeling, the forests with Oak and Other labels for the task contained many patches with little to no vegetation that should be classified as None. To remove the labels over these patches, an image employing the normalized difference vegetation index (NDVI) for the training image was created. NDVI is calculated using Red and NIR channels and outputs high values for areas with dense vegetation and low values for areas with little to none. This NDVI image was used to remove labels in Oak and Other images where the NDVI was low.

The ground truth images are partial images, as most areas with the species of interest were not labeled, but enough was labeled for training for the purposes of this task. Mostly, large uninterrupted groves of a single species were chosen for labeling, as this was easiest to do accurately, but much smaller segments, and in some cases individual trees, were also labeled.

The function to create cube patches from the data was not mine. This function was repurposed from code in Syam Kakarla's Github folder: https://gist.github.com/syamkakarla98/c76733c48d739a17c1d638450be06f4e, though I later found nearly identical code in a research study conducted around the same time: https://ieeexplore.ieee.org/document/9307220, code here: https://github.com/mahmad00/A-Fast-and-

Compact-3-D-CNN-for-HSIC/blob/master/A_Fast_3D_CNN_for_HSIC_v2.ipynb. I made slight modifications for this project. Other than this code, all code was mine, assisted by official documentation for the libraries used.

This project was conceived prior to finding reference materials, and consequently is not based on any single research paper, though segmenting aerial and satellite imagery based on ground species is nothing new. Many of the implementation ideas for this project came from research and blog posts about the subject. These are listed below.

https://ieeexplore.ieee.org/document/9311828

https://medium.com/geekculture/remote-sensing-deep-learning-for-land-cover-classification-of-satellite-imagery-using-python-6a7b4c4f570f

https://towardsdatascience.com/land-cover-classification-of-satellite-imagery-using-convolutional-neural-networks-91b5bb7fe808