G26

# Truths on the Ground: Melding Satellite Imagery and Machine Learning

Alex Hobeychi    Leo Benac


## Abstract

*Satellite imagery can provide valuable insights on land cover when combined with machine learning techniques. In this paper, we present three ways of working with the ESA[1]'s Sentinel-2 satellite imagery and discuss their real world applications: from measuring damage of the California wildfires to tracking the change of ice cover across time, as well as labelling sections of large raster images into 10 classes. To accomplish these goals, the following approaches were used:*

1. *Unsupervised clustering techniques (K-means, Gaussian Mixture Models) to give quick a overview of wildfire damage with mixed results.*

2. *Traditional Supervised techniques (random forests, logistic regression & nearest centroids) applied on classified pixel values to track ice quantities with decent results.*

3. *Writing and training Res-Nets on Eurosat[1, 2] dataset with 93% accuracy and applying the learned weights to classify patches of unseen large imagery to create colored mosaics with positive results.*


## 1. Introduction

Every day The ESA's Sentinel-2 satellite uploads 1.6TB of Earth observations freely to the public with up to a 10 meter per pixel resolution. Providing lots of data for those looking to apply machine learning to satellite imagery. The applications we have decided to work on are forest fire assessment, Greenland ice cover change overtime, and labelling large 100x100km imagery. The data was acquired using the Sci-Hub-Copernicus API which, when queried with a desired footprint, returns a catalog of Sentinel-2 spectral data to download from. The spectral data was then combined to create interpretable images; for example combining the red, green and blue bands forms an rgb image .

[1]European Space Agency

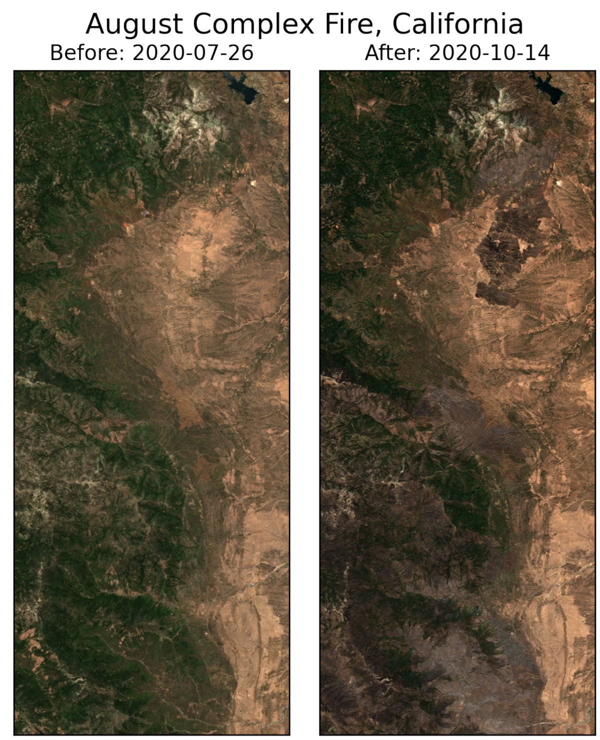


Figure 1. Pre and Post Images of The August Complex Forest Fire in California acquired by combining the red, green and blue spectral bands downloaded using the Sentinel-Sat API.

The data is then used to make conclusions in three different ways. Firstly, we use unsupervised learning through K-means and Gmm to cluster pixel's values of a very specific type of image called NDVI[2] (to be discussed later) in an effort to track cluster changes which, when combined with a human supervisor, could help with tracking the damage on vegetation caused by forest fires with mixed results. Secondly, we created our own labelled rgb pixel data to classify image pixels into 3 classes and through nearest-centroid we tracked ice, ground, and water for 11 time-series images and compared our results with those published by the ESA with good results. Finally, we wrote a 50 layers residual convolutional neural network from scratch using Pytorch,

[2]Normalized difference vegetation index

 G26

trained it on a novel 10 class 27 000 image dataset named Eurosat[1, 2] with 93% validation accuracy, and applied it to massive $10000km^2$ imagery to create colored mosaics of cities and get regional zoning percentages.

# 2. Methodology & Experimental Results

## 2.1. Unsupervised Learning

Spectral bands can be combined in specific ways to amplify certain aspects of the imagery. For example, the normalized difference vegetation index could be leveraged to bring out chlorophyll density. NDVI is calculated by reading the red and near visible infrared into arrays and doing the following calculation:

$$\text{NDVI} = \frac{\text{Near Visible Infrared} - \text{Red}}{\text{Near Visible Infrared} + \text{Red}}$$

The final result is a single band image with values ranging from -1 to 1, where 1 is dense vegetation and -1 is nonexistent vegetation. For the forest images, the initial shapes were (10980, 10980) and after cropping out the non affected region the final shapes were (10980,4552). The data would then be normalized by subtracting the mean and dividing by the standard deviation. The data is then flattened to make it easier to work with a final shape of (49980960,1). Our K-means model was then fitted to the image taken before the fire and these new clusters were then used to find the groupings on both images. By plotting the cost function, we found that 3 clusters would be best for these images (see figure 2).


Figure 2. Cost Function for each of the California images before and after the fires.

We then tuned the K-means model by using 30 initialisations as any more would provide little to no difference in the cluster centers while being much more computationally intensive. The results where then given color mappings by hand and plotted (see figure 3). As this is an unsupervised

| | Green Cluster Percentage |
|---|---|
| Pre-Fire | 34% |
| Post-Fire | 23% |

Table 1. Table Showing the Change in the Green Cluster Relative to the Total Amount of Pixels.

learning technique, no labels where given to these clusters. However, if we compare the results with the original rgb image, the green cluster could give a rough estimate of the change of the vegetation (see table 1). With a loss of 11 percent for the green clusters, it would translate to 5497905 pixel changing clusters, and, as the image resolution is 10m per pixel the results show an approximate loss $549km^2$ of green vegetation.

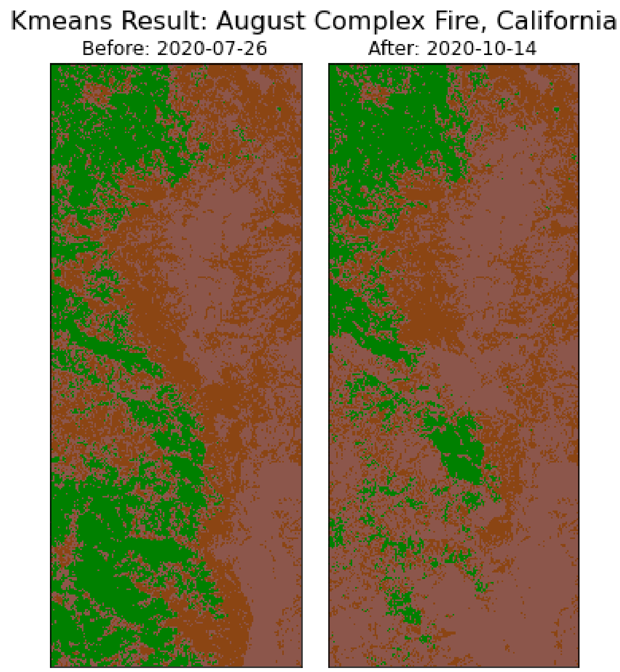


Figure 3. Before and After The California Fires using 3 K-means clusters .

The main limitation of this technique is that there is no way of confirming whether these results fall within a certain error threshold from the ground truth. Furthermore, K-means is a computationally intensive task and training the model takes around 30 minutes for the single band image and would not scale well for larger images. Gmm[3] clustering was also tried with similar results, however, we've omitted it from detailed discussion, as we found that the additional computational cost only yields a result with only a marginal difference.

[3] Gaussian Mixture Models

G26

### 2.2. Traditional Supervised Learning

To apply supervised learning on rgb images, we had to create our own pixel data by downloading imagery, cropping out, and saving desired pixels to create labelled data. We created labelled data so that we can give our classifiers training data with pixels labeled as ice, water, or ground, in order to be able to apply supervised learning algorithms. Therefore, we sought to train a classifier to track the change of ice in the arctic and compared our results to those published by the ESA's data catalog.

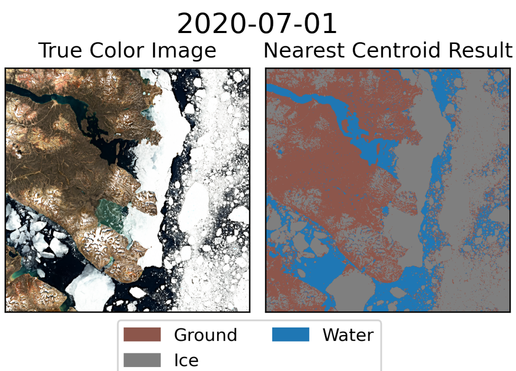


Figure 4. True Color and Classification Result of The Studied Area.

All labelled data were created from different images than those presented in this paper to reduce data bias. We then used the labelled data to train a nearest centroid, random forests, logistic regression, and svm classifier in order to distinguish between the three classes: ice, water and ground. Our results were then compared to the original rgb images and we noticed that the nearest centroid method gave the best result (see figure 4), while svm gave the worst results. Tuning our svm classifier was a major challenge as the default sklearn multi class implementation uses a computationally costly one-vs-one scheme taking 5 to 6 hours to train each model and make predictions. As an alternative, the one-vs-many scheme is much less computationally costly, however, the results were quite poor as this scheme is more sensitive to data imbalance. As the nearest centroid classifier was the most efficient model, we decided to apply it to a time series data set of 11 images of the same area from the month of February to September in order to plot the change of these 3 classes and compare our results to those of that the ESA publishes in their image catalog (see figure 5).

From figure 5, its apparent that on average our results were within 5% percentage points from those published by the European Space Agency leading us to believe that our classifier did well in distinguishing these three classes. Although our results were positive, this technique is not very applicable to cases where labelled classes have a high color variation; for example, classifying forests during the fall or differentiating between residential and industrial areas. In the next section we discuss another approach to carry out such tasks.

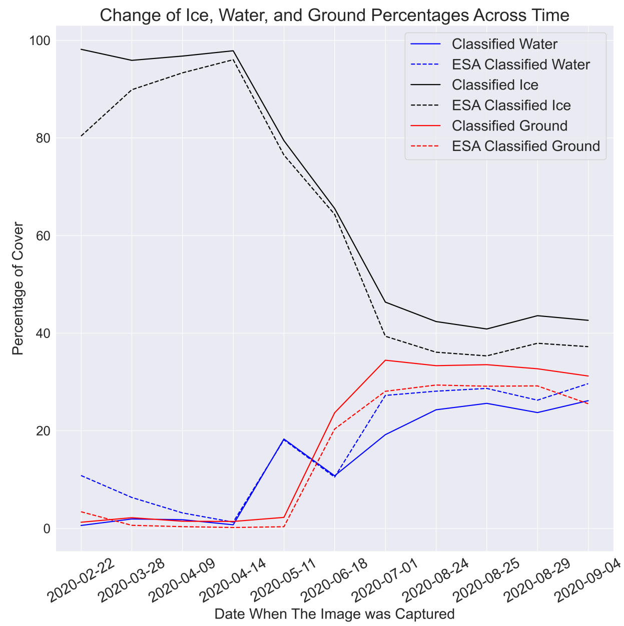


Figure 5. Ice, Water and Ground Variation from the Month of February to September in a $10000km^2$ Area on the Coast of Greenland.

### 2.3. Neural Networks

To classify large Images into 10 classes we decided to write and train our own 50 layers residual neural network using Pytorch. We then applied the network on the Euro-Sat dataset. Euro-Sat is a 10 class 27 000 sample dataset of 64x64 pixel images taken from the Sentinel-2's TCI[4] published images. The network was then trained using a series of 25 epoch sequences where, between each of those sequences, the data was reshuffled, a random vertical-horizontal flip was applied to each of the images, and then split into a 80 percent train-validation split. The images were then normalized by calculating the mean and the standard deviation for each bands for the totality of the dataset, subtracting and dividing each image by those respective calculated values. The loss was calculated using cross-entropy and the Adam optimizer was used to find the weights of the network. To avoid over-fitting, we limited ourselves to using three load-train-save sequences to train the network, and our network ended with an overall 93% accuracy. As the dataset only contains data that has been cropped from European imagery, we sought to deploy the network by cre-

[4]True Color Image

G26

ating colored mosaics of North-American cities. A San-Francisco TCI image was then downloaded, chunked into 29241 64x64 images normalized in the same way as the training data and fed to the network to get our prediction. As each prediction represents the class of a 64x64 image, the prediction was multiplied by 64 along each axis and reshaped to the same dimension of the original. The result was then given a color-mapping for each class (see figure 6). The model is obviously limited by the classes it was trained on; thus, it cannot accurately classify certain data, such as mountainous areas. Zoning percentages, moreover, were calculated using the unique values of the predictions and dividing them by the total amount of pixels (see table 2). 101 and 152 layered residual networks were also written and trained, however, their additional training and testing time encouraged us to restrict the network to 50 layers.

| | | | |
|---|---|---|---|
| Vegetation | 0.51 % | Residentual | 10.52% |
| Forest | 2.53% | River | 3.22% |
| Sealake | 37.11% | Pasture | 15.21% |
| Highway | 1.91% | Industrial | 0.14% |
| AnnualCrop | 12.19% | PermanentCrop | 16.67% |

Table 2. Zoning Percentages for Each Class.

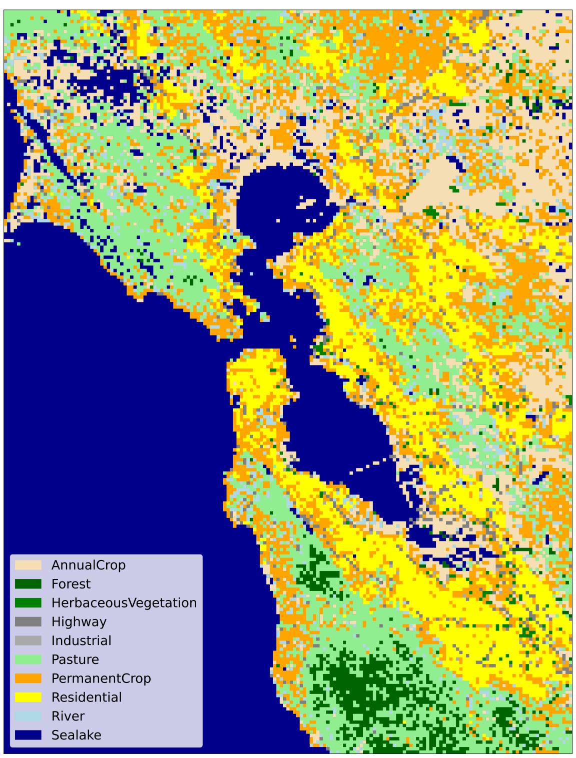


Figure 6. Mosaic of San-Francisco.

## 3. Conclusion

The main goal of our project was to apply machine learning algorithms to Sentinel-2 satellite imagery to make real life ascertainment. It is safe to say that our goal was accomplished. As previously mentioned, our project can be divided into three parts. The first part dealt with showing how clustering algorithms can highlight the negative effects of wildfires on California's vegetation. The second part dealt with how we managed to create data, and train a classification algorithm in order to do a 3 class classification of images, with satisfactory results using nearest centroids; however, we were not successful in tuning a svm model to make predictions as accurate as our nearest centroid. Finally we trained a Res-Net on the Euro-Sat dataset with 93% accuracy, to create a colored mosaic of the city of San Francisco. One regret is that our neural network uses a 64x64 patch approach, rather than a variable size patch such those in image segmentation.

## References

[1] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 204–207, IEEE, 2018. 1, 2

[2] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019. 1, 2