Palak Agarwal and Max Masuda-Farkas
MUSA 650
Professor Guray Erus
Final Project

Detecting Settlements Lacking Electricity with Remote Sensing

Introduction

Problem

Given the rate at which urbanization is occurring around the world, it is vital that new human developments have access to basic amenities such as sanitation, water, and electricity. Although these amenities are easily provided in developed countries, developing countries face enormous challenges to offering these basic resources to their citizens. In this project, we will be using satellite imagery to detect human settlements that have no electricity in order to identify where resources are needed most.

Data

Dataset

The data was a part of this year's remote sensing competition - the detection of settlements without electricity challenge track (Track DSE) of the 2021 IEEE GRSS Data Fusion Contest, organized by the Image Analysis and Data Fusion Technical Committee (IADF TC) of the IEEE Geoscience and Remote Sensing Society (GRSS), Hewlett Packard Enterprise, SolarAid, and Data Science Experts. Participants were required to submit binary classification maps and the performance will be evaluated using F1 score.[1]

The contest dataset is composed of 98 tiles of 800×800 pixels, distributed respectively across the training, validation and test sets as follows: 60, 19, and 19 tiles. Each tile includes 98 channels from the below listed satellite images. As the test and validation sets do not contain labels, we used only the training set and split it as 50 and 10 to create training and testing sets.

Satellite Data

- Sentinel-1 polarimetric SAR dataset
    - 2 channels corresponding to intensity values for VV and VH polarization at a 5×20 m spatial resolution scaled to a 10×10m spatial resolution.

- Sentinel-2 multispectral dataset
    - 12 channels of reflectance data covering VNIR and SWIR ranges at a GSD of 10 m, 20 m, and 60 m. The cirrus band 10 is omitted, as it does not contain ground information.

---

[1] To access the data and further details, visit: https://ieee-dataport.org/competitions/2021-ieee-grss-data-fusion-contest-track-dse#files.

- Landsat 8 multispectral dataset
    - 11 channels of reflectance data covering VNIR, SWIR, and TIR ranges at a GSD of 30m and 100 m, and a Panchromatic band at a GSD of 15m.
- The Suomi Visible Infrared Imaging Radiometer Suite (VIIRS) night time dataset
    - The Day-Night Band (DNB) sensor of the VIIRS (Visible Infrared Imaging Radiometer Suite) provides on 1 channel, the global daily measurements of nocturnal visible and near-infrared (NIR) light at a GSD of 750 m. The VNP46A1 product is a corrected version of the original DNB data, and is at a 500m GSD resolution.

For our models, we elected to use only the Sentinel-2 multispectral dataset. But the code can be easily edited to read and work with any other satellite image by a small tweak in the file path name.

Semantic Labels

The provided training data is split across 60 folders named TileX, X being the tile number. Each folder includes reference information (‘groundTruth.tif’ file) for each tile. The reference file (‘groundTruth.tif’) is 16×16 pixels large, with a resolution of 500m corresponding to the labelling strategy described. The pixel values (1, 2, 3 and 4) correspond to the four following classes:

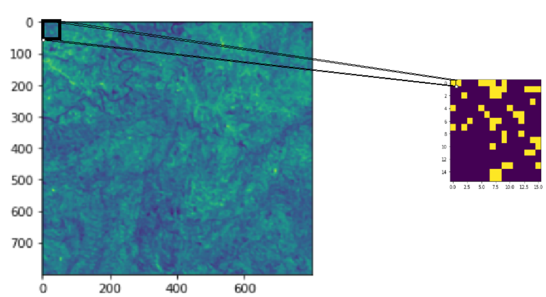
- 1 (ff0000): Human settlements without electricity (Region of Interest)
- 2 (0000ff): No human settlements without electricity
- 3 (ffff00): Human settlements with electricity
- 4 (b266ff): No human settlements with electricity

Data resolution

- The satellite images have been resampled to a Ground Sampling Distance (GSD) of 10 m. Thus each tile corresponds to a 64km2 area. So each image is 800x 800 pixels large
- The label images are 16×16 pixels large, with a resolution of 500m corresponding to the labelling strategy described above.

Relation between Satellite image and labels

The illustrator below shows how the image is connected to its label. Each 50X50 pixel in the satellite image is a label in the 16x16 ground truth image. This is true for all the 12 bands in the Sentinel-2 multispectral dataset. So each satellite image is 800x800 which is reshaped as 256x2500 using the blockshaped function in the code. It is 256 as the truth image is a 16x16 image and 2500 represents the 50x50 pixel value.



Methods

Data Wrangling

Because of the labeling convention of the data described in the previous section, a substantial amount of wrangling was required to prepare the data for modeling. Since each of the 256 labels of the groundTruth image corresponds to a 50x50 pixel grouping in each tile, the data was further separated into 256 discrete groups of 50x50 pixels. Additionally, because all of the 12 bands share the same groundTruth image, every label ultimately corresponds to a total of 12, 50x50 pixels, which amount to 30,000 pixels per label when flattened. All of the data were reshaped and reformatted accordingly.

Statistical Learning

Several traditional, statistical learning models were applied to the data to predict which of the four classes mentioned above are represented in each grouping of pixel values. These models included support vector machine (SVM), random forest, and k-nearest neighbor classifiers. Each model was evaluated by splitting the data into training and testing sets, training each model on the training data, and deriving its accuracy through the testing data. After first running each model with its default parameters, parameters were optimized to improve, however marginally, the accuracy of the model.

Deep Learning

A deep learning model was also applied to the data in the form of a convolutional neural network. Each input of 50x50 pixels for all 12 bands were passed through 8 different 2D convolution layers. From beginning to end, the data was reduced from a shape of (50, 50, 12)—representing the 50x50 pixels and 12 bands—to a one-dimensional array of (5). The model was further optimized by changing the size of filters and maxpooling size. Initially to limit the loss in data, we ran the convolution network twice before maxpooling and reducing the dimension. Also, we played around with the final dense layer in terms of features. While training the model itself, we used different epochs and batch sizes firstly to wrangle the large dataset but also to increase testing accuracy while reducing loss. Finally, the model was fit to the training data in batches of size 36 over 10 epochs, the accuracy was evaluated using the testing data.

Results

Statistical Learning

Each of the statistical learning models returned unfavorable results. The SVM, random forest, and k-nearest neighbor models each returned a testing accuracy of approximately 0.49. Even after the parameters of the SVM model were optimized, the testing accuracy remained unchanged or worsened. Although an accuracy that hovers around 0.5 for a multi-class classification model such as this one is not as poor as for a binary class model, it is still too low to offer a reliable prediction model for the use case here. Another reason for such low accuracy is that each band of Sentinel-2 has very different ranges and varies drastically in value. So using only a subset of the whole dataset, means that the values of the features are not comparable and that leads to lower accuracy as well.

Deep Learning

After optimization, the deep learning model returned robust accuracy results. Although in the absence of optimization accuracy scores initially hovered around values similar to, if only slightly better than, the statistical learning models, once optimized the deep learning model returned an accuracy score of up to 0.75 at the tenth epoch. What helped the model do so well was the convolution network design. Adding multiple layers of convolution before maxpooling and also making sure the feature reduction was not drastic and gradual, increased the models accuracy quite significantly.

Discussion

Statistical Learning

The relatively poor results of the traditional machine learning models are likely best explained by the nature of the particular data here. As a matter of sheer computing power, there were significant constraints on the amount of data that could be inputted into the models such that only the first 20 of the 60 total tiles were used. Because the statistical models required data in a two-dimensional form, even the 20-tile subset represented a data array of 153,600,000 total pixels. Given that the models here were tasked with identifying four discrete classes, a large quantity of data would have been necessary to train adequately each model, and this was simply not possible with limited memory.

Deep Learning

The strong results of the convolutional neural network exemplify the advantages inherent to deep learning models, especially in cases that involve large amounts of data such as the one here. First, unlike the statistical learning models which required that data be flattened into a two-dimensional array and thus could not handle more than a 20-tile subset of the entire dataset, the deep learning model was easily able to work through all 60 tiles since the data were passed through the convolutional layers as is. That distinction alone may have contributed to the stronger results of the deep learning model, as it had significantly more data from which to learn.

Next Steps

Even given the already robust results of the deep learning model, opportunities for further improvement remain. For both the statistical learning and deep learning models, only the 12 bands of the Sentinel-2 multispectral dataset were used. One option for refining the models is to try running them on the datasets from the other available satellite images. As mentioned, this is easily done by changing the relevant file path. A second option is to reduce the number of classes contained in the dataset by relabeling the four classes as a binary, two classes. The principal aim of these models—that is, to detect human settlements that lack electricity—would still be fulfilled if instead of training a model that could detect all four classes, the model could simply be able to distinguish between human settlements without electricity and anything else. Such a refinement not only may improve the deep learning model, but may also make the statistical learning models more viable.