SOIL EROSION REPORT

INTRODUCTION

Soil erosion is really an undesirable process that spoils huge areas of fertile land. The task was to train a model to detect erosion.

SOLUTION

We can divide the solution to the problem into 2 sections “EDA & preprocessing” and “modeling”:

a) EDA & preprocessing:

We had a shapefile with erosion areas and satellite imagery at our disposal. The first thing that comes to mind is converting the data to a **single CRS** is important because of the distortion that different systems have different. Then we should visualize ourselves to see what we are dealing with, a good program here is ArcGIS Pro. What we can notice is that JP2 does not cover all shp polygons. We **should clip the shp to jp2** at this point. Since we have an object detection problem we should dataset. For this we should **split the image into smaller images.** Then **filter out images** that do not contain polygons. With EDA, it would be **useful to know when the photo was taken** if it was sometimes taken during some climate anomaly or environmental tragedy.
In addition, it would be useful, for example, **climate data** that could improve the results of the model. Can also **calculate everything can** that is, for example, various indicators such as NDVI, NDWI, NDSI, EVI etc.
Admittedly, for smarter algorithms it will not be an added value because it is not new data while it can help EDA and maybe sometimes algorithms.

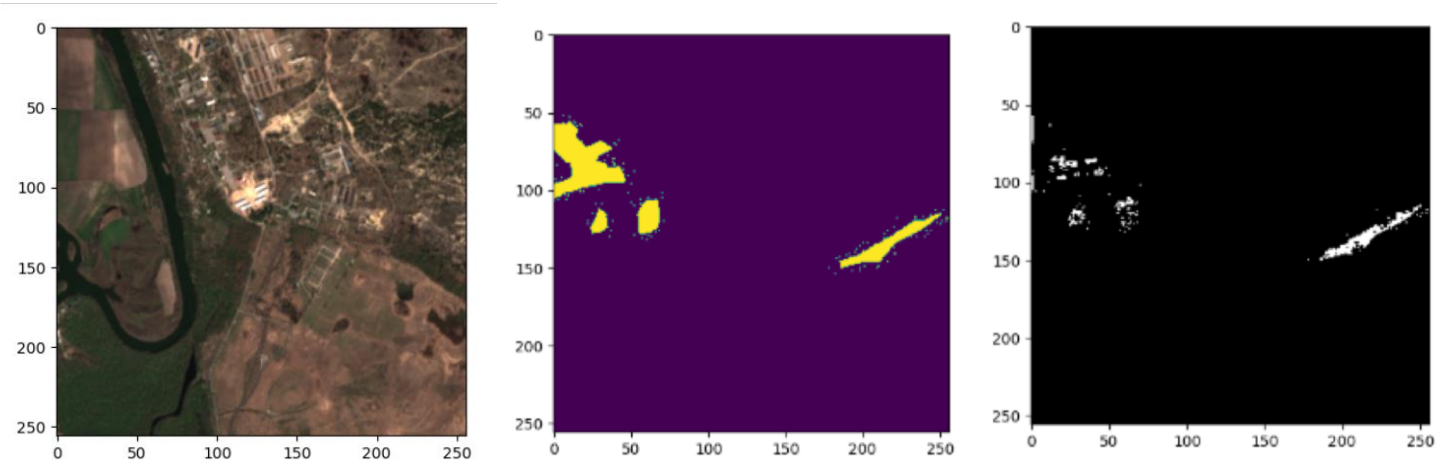
b) Modeling

Since we **don't have too much data** what comes to mind is **traditional machine learning.**
This will most likely be a better solution than creating a neural network due to the fact that it needs a large amount of data. Of course, it is always worth trying NN while with little time and data to execute I would choose **Random Forest** or **SVM**. The plus side of RF is that it shows us what influenced the prediction and that it handles noise well.
In this project, I decided to **extract the data using CNN and then train RF based on the features from CNN**. I haven't used any specfic architecture here while I think you can check here such architectures as SegNet, U-Net or EfficientNet.

After fast training, the results were not satisfactory for the new given while for the trained ones they were. It seems to me that if we made a deeper features extractor and hyper parameters tuning of RF the results could be good even from such a small amount of data.

Results for training data for test data will not show because there were none:

RAW MASK PREDICTED



Sources:

Knowledge from university and self-study of spatial data.
https://medium.datadriveninvestor.com/preparing-aerial-imagery-for-crop-classification-ce05d3601c68

Coding sources:

https://github.com/Justdjent

https://www.youtube.com/watch?v=5ct8Yqkiioo&list=LL&index=21&ab_channel=DigitalSreeni