Project Neural Network

Planet: Understanding the Amazon from Space

Aloukou Eirini BAPT1702

Lymperi Ourania-Anna BAFT1711

Team members

- Aloukou Eirini as Data Scientist. Eirini is a statistician, graduated from Statistics and Insurance Science of University of Piraeus and
- Lymperi Orianna as Business Analyst. Orianna is an Environmental Engineer, graduated from Environmental Engineering School of Technical University of Crete.

Business case

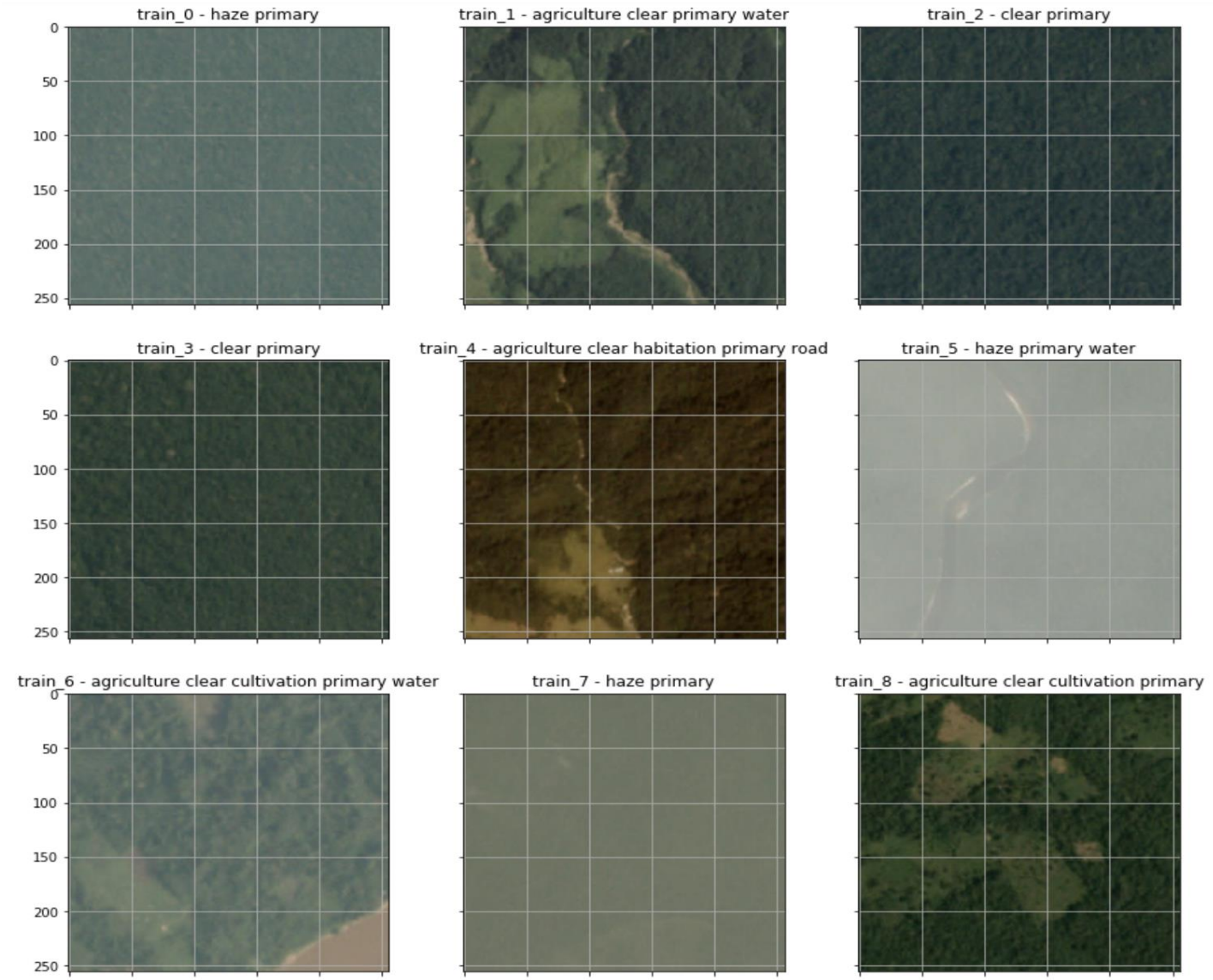
- The Amazon is the world's largest rainforest and the largest river basin on the planet. The region is home to 10 percent of all plant and animal species known on Earth.
- Deforestation in the Amazon Basin accounts for the largest share, contributing to reduced biodiversity, habitat loss, climate change, and other devastating effects in the whole planet.
- In this project, we are challenged to perform a deep learning analysis in order to label satellite image chips with various classes of land cover/land use.
- Resulting algorithms will help the global community to better understand where, how, and why deforestation happens all over the world - and ultimately how to respond.

The data

- 256x256 jpg satellite images (RBG) of 17 categories of land use areas in Amazon
- 40,479 images, as training set and 10% validation

primary	37513
clear	28431
agriculture	12315
road	8071
water	7411
partly_cloudy	7261
cultivation	4547
habitation	3660
haze	2697
cloudy	2089

Satellite images with categories



Pre-processing data

- Reduced the number of different labels into two of the most frequent and important ones for our purpose to analyze, agriculture and road.
- Rescaled our images to smaller dimension (128, 128) using skimage algorithm.
- Gray scaling.
- Normalization of the data.

Hardware

1 PC MacBook Pro

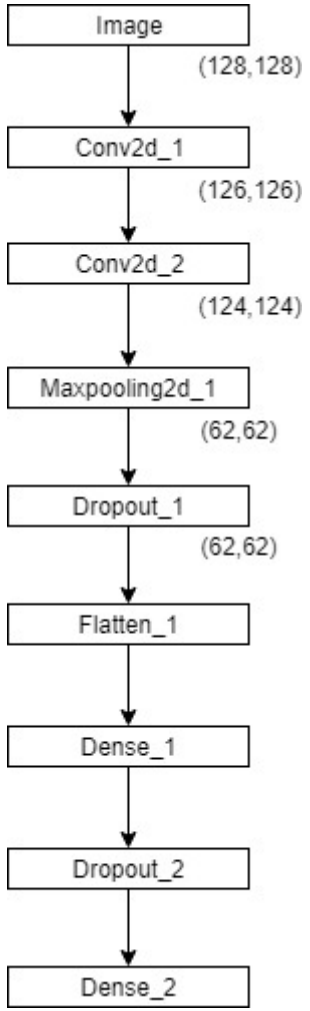
- *Processor 2,3 GHz Intel Core i5*
- *RAM 8 GB 2133 MHz LPDDR3*
- *SSD Storage*

Algorithm

Convolutional Neural Networks (CNN)

Methodology

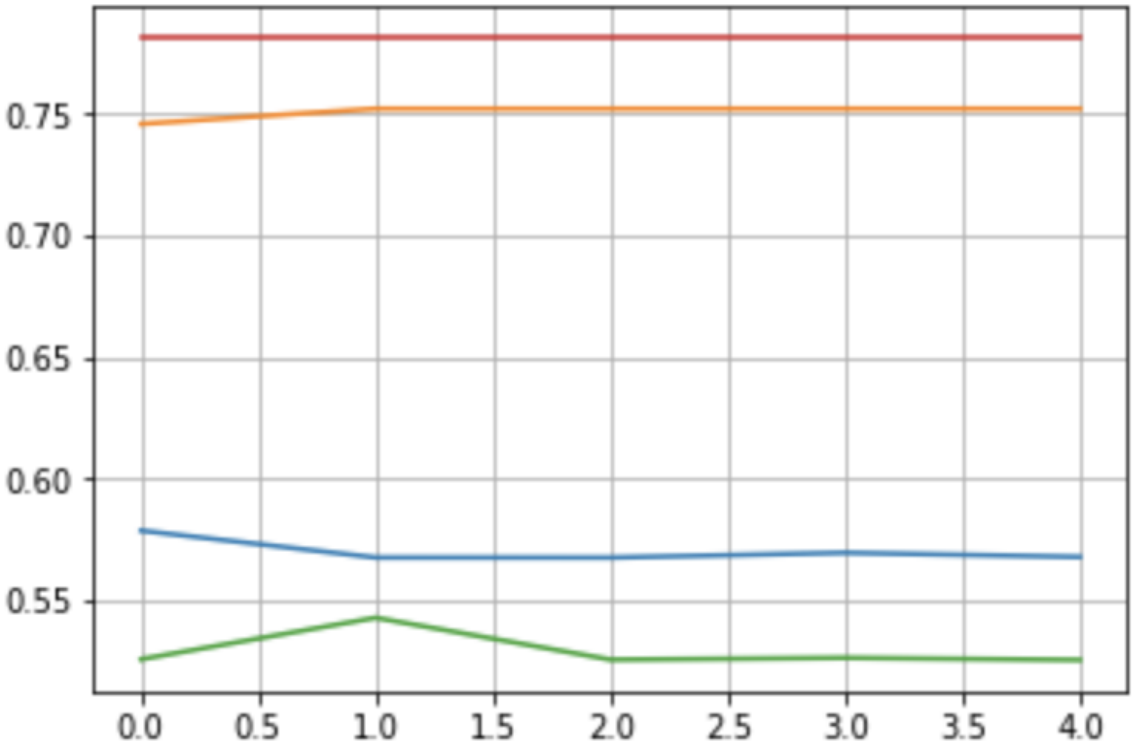
- Conv ReLu Layer with 64 3x3 filters,
- Conv ReLu with 64 filters,
- 2x2 Max Pool, Dropout,
- 0.25 Dropout after max pooling,
- Fully Connected ReLu layer,
- 0.5 Dropout after first dense layer
- Softmax activation



Results

	Accuracy	Loss
train	0.7459	0.5785
validation	**0.7812**	0.5256

Loss function



Conclusions

- We considered **agriculture** and **road** as the most important labels, since they are the strongest indication of a deforested area.
- **Gray scaling** reduces accuracy significantly, since in our dataset the colors play vital role in the recognition of the images.
- The fact that the accuracy could not get higher than **78%**, could be explained by the small amount of data remaining after the extraction of 15 categories.
- Further work could include all categories, tuning hyperparameters further and exploring other types of networks.

THANK YOU!