

**FIRE IN THE AMAZON**  
**GROUP PROPOSAL**  
**Understanding the Amazon with Pytorch**

**Daniela Schloeter**  
**Guillermina Sutter Schneider**

## Introduction

During the last couple of years, the focus of international-development-oriented policies has slowly moved towards the environment. The increasing loss of area forest imposes a threat to biodiversity, contributes to climate change, habitat loss, and has many other devastating effects. Therefore, it is of crucial importance to analyze and understand these areas to be able to take action faster and more effectively. Being able to classify satellite images of the Amazon rainforest would help policy-makers in the region in the design of effective measures to fight deforestation.

## Dataset

The dataset belongs to the Kaggle competition “[Planet: Understanding the Amazon from Space](#).” Satellite image chips with atmospheric conditions and various classes of land cover/land use are provided by [Planet Labs](#) and its Brazilian partner [SCCON](#).

The data is already divided into a train and test set, and contains 40,479 jpg files and 20,522 jpg files of size 256x256, respectively which is has enough data to train a deep learning network. Each image chip has at least one and potentially a combination of more than one of the following labels: agriculture, artisanal mining, bare ground, blooming, blow down, clear, cloudy, conventional mining, cultivation, habitation, haze, partly cloudy, primary, road, selective logging, slash and burn, and water.

## Network structure and framework

Using the scripts from Exam II, we will adapt the CNN built for the exam to this dataset. If we see this CNN structure does not perform as expected, we will evaluate moving to a pretrained network such as ResNet50, VGG19, VGG19\_BN, etc.

We will use pytorch for this project since it is a flexible framework and it is more straight-forward than those others studied in class.

## Reference materials

[Pytorch documentation](#)

[Pytorch Forum](#)

[Towards Data Science](#)

[Medium - Biasly AI](#)

[Kaggle Competition](#)

[Stack Over Flow](#) (for debugging purposes)

[Sample Codes from the class's GitHub repo](#)

## Metrics

F-beta Score: [Based on the sklearn documentation](#), the F-beta score is the weighted harmonic mean of precision and recall, reaching its optimal value at 1 and its worst value at 0. It takes into consideration both the recall and precision metrics. In order for us to compare with the models in the Kaggle competition, we chose this metric with a  $\beta$  equal to 2 to favor recall over precision.

## Schedule for completing the project

