Schneider Electric
Hackathon 2022

Classification of deforestation types from satellite imagery

Team NaN: Nazar Puriy and Ferran Espuña

Problem Statement

01

Many images of deforestation are given.

Also, the images have associated metadata: location and year.

02

Labels are given for only a subset of the images.

These have the following meanings:

0: Plantation
1: Grassland/Shrubland
2: Smallholder Agriculture

03

The goal is to correctly classify the remaining images.

That is, to associate to each of them its label 0, 1 or 2

Handling of the data

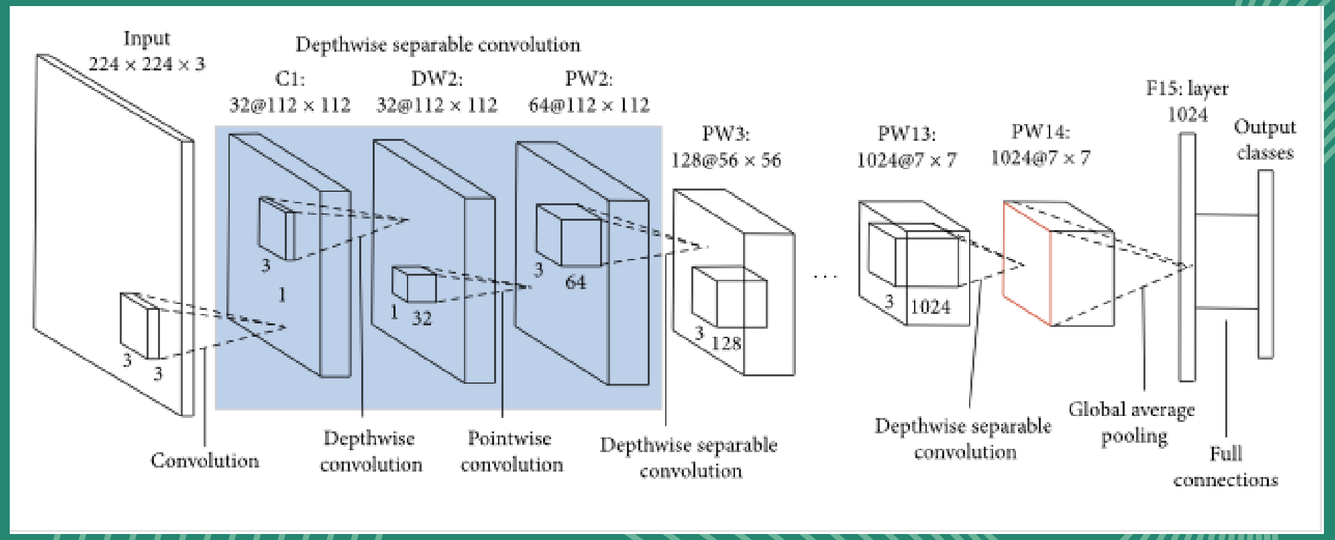
We have used one-hot encoding for the labels in order to make them understandable to a wide variety of machine learning models. This lets us specify a loss function, categorical log-loss, which is the standard for this type of classification problem.

We have done data augmentation on our test dataset to make the models generalize better. As there was quite an imbalance in the data (much fewer examples of type 1), we have augmented these by a factor of 7, while the other 2 classes only by a factor of 2. This improved our accuracy by 0.03-0.04.

In order to be able to try different models and test their accuracy, we have split our train dataset into 80% training data and 20% testing data. Once we've been happy with the chosen model, we have trained it with the full train to make the final predicions on the actual test dataset.

Model we used to predict

- MobileNetV2 is a Deep Neural Network model open sourced by Google in early 2018.
- It's a very powerful model for image classification and segmentation, and also object detection.
- The model is already trained, but we retrained its last layers with our train examples to make it learn how its extracted features map to our labels.
- In our experiments we got about more than 0.82 accuracy. This technique is called transfer learning,



Results

The model quickly fits to the dataset. Experimentally we have seen that in betweeen 10 and 20 epochs the loss fuction of validation set stays the same while the accuracy increases. We used 20 epochs to train the final model.

Hope to get great results!

Contact us!

npuriypu7@alumnes.ub.edu
ferranespuna@gmail.com