



Zero deforestation mission

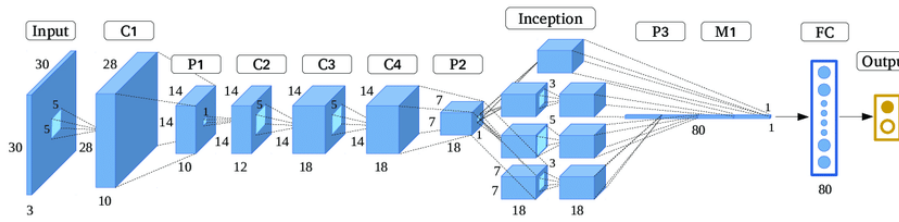
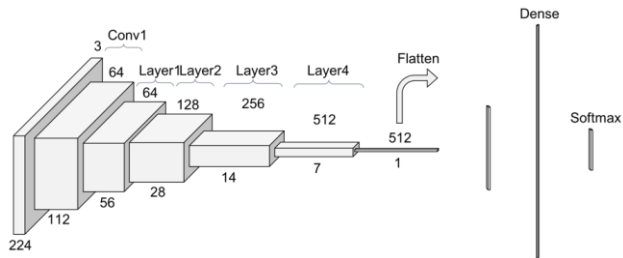
Alejandro Argüello, Pablo Bernabeu, Rubén Cuervo (Table 113)
Schneider Electrics Hackathon 2022 - Data Science
21st May 2022

PROCEDURE

1. Importation of the train and test **datasets and images**. **Standard scale** the longitude, year and latitude features.
2. **Split training data** in training, validation (20% for validation) and test (20% for test) so as to select the model which bests generalizes the problem.
3. Perform **data augmentation**.
4. Try several **CNN architectures** to select the optimal one.
5. Use **CNN labels** as a new feature for the **XGBoost** classifier **together with year, latitude and longitude**
6. **Select hyperparameters** for the XGBoost.
7. Once the optimal value is selected, **train with the whole training dataset** (no validation).
8. **Selecting features** not required performing.
9. **Predict labels** for testing dataset and transform the to **JSON** format

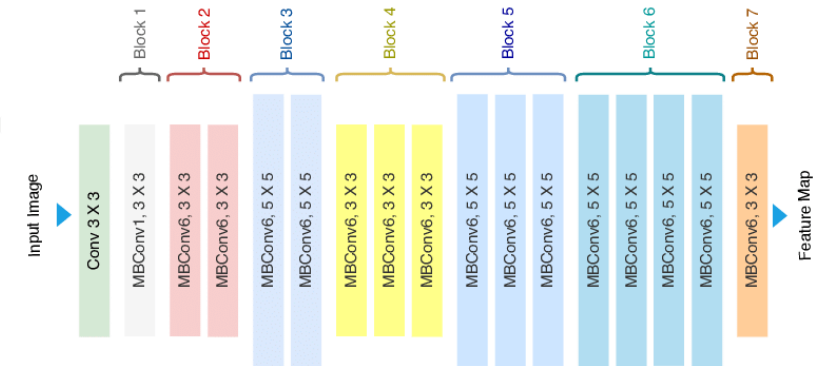
CNN architectures tried

Resnet50

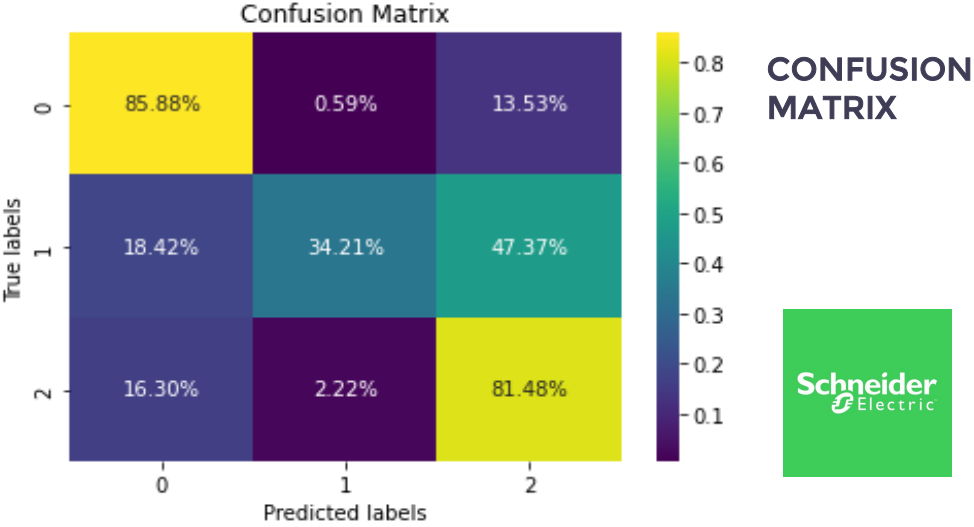
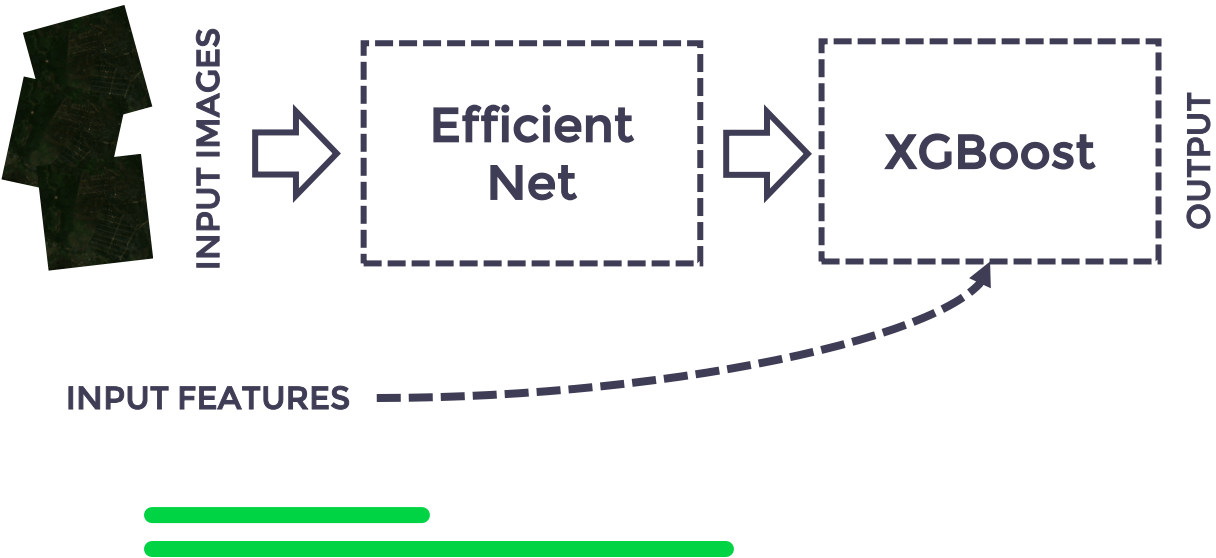


Google Net

Efficient net



RESULTS OBTAINED



CONCLUSIONS



- In 2022, CNN are the gold standard for computer vision problems, and some resources such as FastAI contained the state-of-the-art architectures to solve these complex problems
- Data augmentation (rotations, blurring, flipping, ...) is crucial for problems with scarce data.
- We can combine CNNs with ML algorithms to obtain the best possible outcomes combining image information with features
- Computer Vision can be crucial for helping the environment, and the project proposed by Schneider Electric is a great example.