Satellite Image Classification and Detection Models for Earth Observation

Corentin Perdrizet

September 27, 2024

This document presents key details about the various models I have developed for satellite image classification and detection tasks. Each model is designed for specific objectives, such as image classification, object detection, or image segmentation, using different datasets and neural network architectures. The evaluation of these models includes metrics such as precision and accuracy, along with visual representations like confusion matrices and architecture diagrams, which provide insight into their performance and structure. All necessary information regarding the operation of these models can be found in the project's Readme files.

1 Model 1: Image Classification - Eurosat

1.1 Model Description

This model performs **image classification** using a convolutional neural network (CNN). The task of this model is to take an input satellite image and classify it into one of the following 10 predefined land cover classes:

AnnualCrop **Forest** **HerbaceousVegetation** **Highway** **Industrial**
Pasture **PermanentCrop** **Residential** **River** **SeaLake**

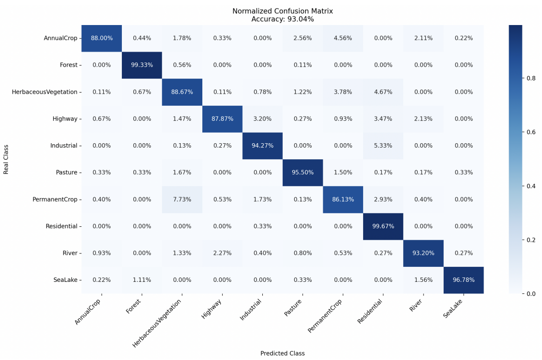
In this classification task, the model processes each input image and associates it with the most likely class based on its features.

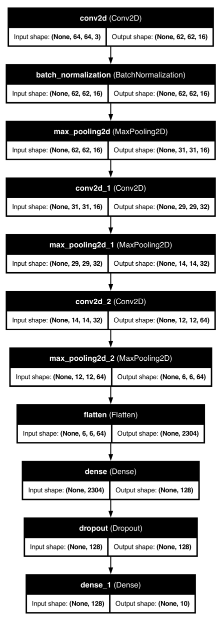
1.2 Useful Information

- **Model Size**: 1.3 MB
- **Dataset**: The model was trained on the **EuroSAT** dataset, which is publicly available at the following link: `https://github.com/phelber/EuroSAT`. The dataset contains 27,000 satellite images from 10 different classes. Each image is 64x64 pixels, with a resolution of approximately 10 meters per pixel.
- **Input**: 64x64 pixels satellite images.

1.3 Evaluation

- **Precision**: The model achieves a classification precision of 93.03% on the validation set.
- Below are the confusion matrix and the detailed architecture of the model.





2 Model 2: Boat Recognition - MASATI

2.1 Model Description

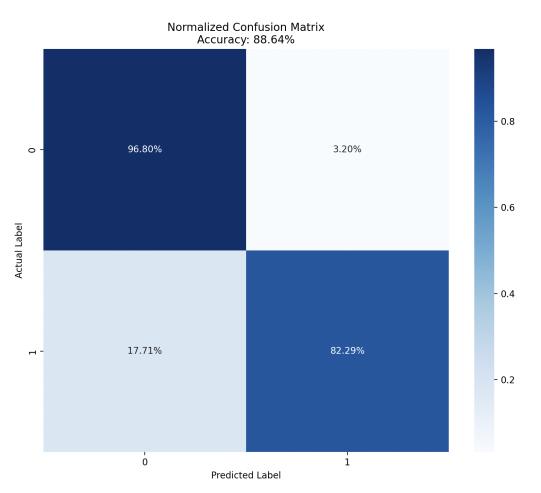
This model performs **binary classification** to detect the presence of boats in satellite images. It is specifically designed to classify whether an image contains a boat (*Boat*) or no boat (*No Boat*). The model processes each input image and outputs a probability for each of the two classes, assigning the image to the class with the highest probability.

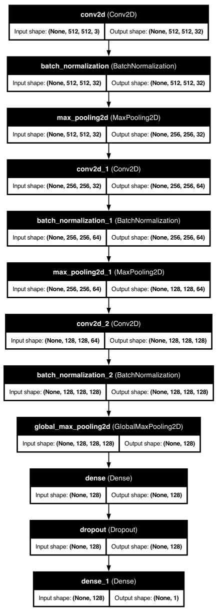
2.2 Useful Information

- **Model Size**: 496 KB
- **Dataset**: The model was trained using the **MASATI** dataset, which contains a variety of 512 x 512 pixels satellite images capturing coastal and maritime areas. This dataset includes images featuring both boats and no-boat scenarios, making it ideal for binary classification tasks. More informations about the dataset is publicly available at the following link: `https://www.iuii.ua.es/datasets/masati/` and can be downloaded at the following link : `https://drive.google.com/file/d/1Oj98Ujhbtw54DRh42GBfDRS-UbOaOfNs/view`. (Private)
- **Input**: 512 x 512 pixels satellite images.

2.3 Evaluation

- **Precision**: The model achieves a classification precision of 88.64% on the validation set.
- Below are the confusion matrix and the detailed architecture of the model.





3 Model 3: Image Segmentation - U-Net

3.1 Model Description

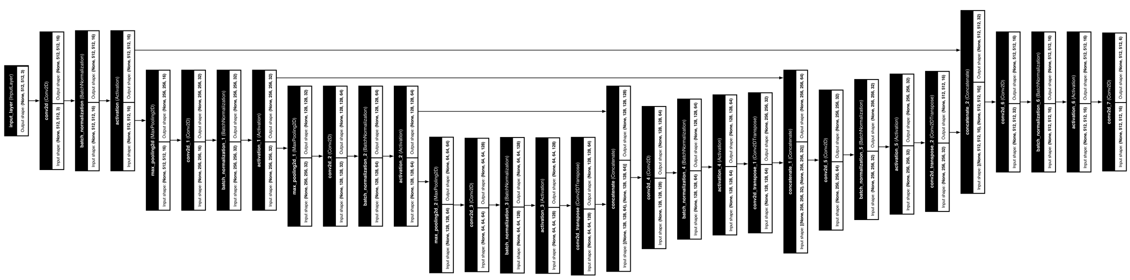
This model performs image segmentation using the U-Net architecture, a powerful model commonly employed in medical imaging for pixel-level classification. However, significant modifications were made to adapt the original U-Net model, which typically weighs several hundred megabytes, for satellite imagery tasks. In this case, the model is trained to recognize and segment various features in satellite images, including buildings, land, roads, vegetation, water, and unlabeled regions. The output is a detailed segmentation map, where each pixel is classified into one of these categories.

3.2 Useful Information

- **Model Size**: 1.1 MB
- **Dataset**: The model was trained using a dataset containing satellite images of differents size and their corresponding masks, all coming from diverse landscape in Dubai. The dataset can be downloaded by following this link :
`https://www.kaggle.com/datasets/humansintheloop/semantic-segmentation-of-aerial-imagery`.
- **Input**: 512 x 512 pixels satellite images.

3.3 Evaluation

- **Accuracy**: The model achieved an accuracy of 80.50%, meaning that 80.50% of the pixels are correctly classified into their respective categories.
- **Comment**: Although the model performs well overall, it faces challenges in accurately defining roads, particularly because the dataset contains images of Dubai, where many roads are made of limestone and appear quite white. This similarity in color and texture between the roads and surrounding landscape elements makes them harder to distinguish. However, the model remains effective for the other classes, such as buildings, vegetation, and water.
- Below are the detailed architecture of the model.



4 Model 4: Wildfire Prediction - Satellite Images

4.1 Model Description

This model performs **binary classification** to predict whether a given satellite image shows an area that is susceptible to wildfires or not. The model analyzes the features of the input satellite image to classify it into one of two categories:

- Susceptible to wildfire
- Not susceptible to wildfire

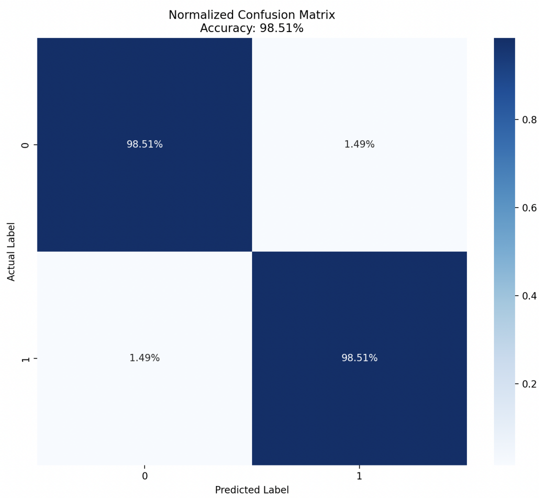
The model aims to assist in early detection and prevention efforts by identifying high-risk areas.

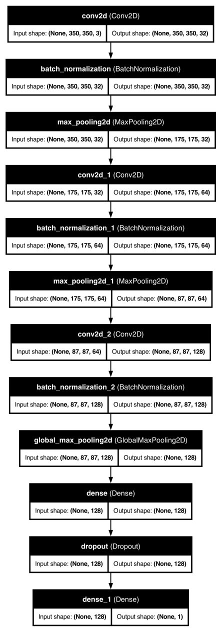
4.2 Useful Information

- **Model Size**: 496 KB
- **Dataset**: The model was trained using the **Wildfire Prediction Dataset**, which consists of satellite images of areas in Canada that have previously experienced wildfires. The dataset can be download by following this link :
`https://www.kaggle.com/datasets/abdelghaniaaba/wildfire-prediction-dataset`.
- **Input**: 350 x 350 pixels satellite images.

4.3 Evaluation

- **Precision**: The model achieved a precision of 98.51%, indicating a high level of accuracy in predicting whether an area is susceptible to wildfires
- Below are the confusion matrix and the detailed architecture of the model.





5 Model 5: Multiclass Classification - BigEarth

5.1 Model Description

This model performs **multilabel classification** using satellite images composed of 12 spectral bands. The input images belong to multiple classes simultaneously, chosen from a set of 43 predefined land cover categories. The task of this model is to predict which classes the input image is most likely to belong to.

The 43 classes are:

Sea and ocean **Mixed forest** **Non-irrigated arable land** **Coniferous forest**
Transitional woodland/shrub **Water bodies** **Pastures**
Land principally occupied by agriculture, with significant areas of natural vegetation
Continuous urban fabric **Complex cultivation patterns** **Agro-forestry areas**
Broad-leaved forest **Inland marshes** **Discontinuous urban fabric** **Natural grassland**
Sclerophyllous vegetation **Annual crops associated with permanent crops** **Peatbogs**
Industrial or commercial units **Water courses** **Vineyards** **Moors and heathland**
Mineral extraction sites **Permanently irrigated land** **Rice fields**
Olive groves **Green urban areas** **Burnt areas** **Sparsely vegetated areas** **Bare rock**
Sport and leisure facilities **Intertidal flats** **Estuaries** **Airports**
Fruit trees and berry plantations **Road and rail networks and associated land**
Coastal lagoons **Construction sites** **Salt marshes** **Beaches, dunes, sands**
Dump sites **Port areas** **Salines** **Other**

The model must handle multiple outputs since each image can belong to several of these classes at the same time.

5.2 Useful Information

- **Model Size**: 2.9 MB
- **Dataset**: The model was trained on the **BigEarth** dataset, which consists of satellite images with 12 bands representing various spectral data. The bands are structured as follows:
  - 2 bands of size 20x20 pixels (B01, B09)
  - 6 bands of size 60x60 pixels (B05, B06, B07, B8A, B11, B12)
  - 4 bands of size 120x120 pixels (B02, B03, B04, B08)

  These bands provide different resolutions and spectral information, making the dataset highly versatile for multilabel classification tasks. The dataset can be download by following this link : `https://tubcloud.tu-berlin.de/s/cPk5mGC3iqs6ogi`. (Private)
- **Input**:
  - Input 1: shape=(20, 20, 2) - 2 channels for B01, B09
  - Input 2: shape=(60, 60, 6) - 6 channels for B05, B06, B07, B8A, B11, B12
  - Input 3: shape=(120, 120, 4) - 4 channels for B02, B03, B04, B08

  The input must therefore be a folder containing all the corresponding bands, each band must end with its number and must be a .tif, more technical informations are provided on the Readme. The model uses three separate neural networks for each input size and concatenates the results to make the final prediction.

5.3 Evaluation

- Accuracy : The model achieved an accuracy of 45%, which means that in multilabel classification tasks, it correctly predicts a subset of the true classes in less than half of the cases on average. This includes correctly identifying both the presence and absence of certain classes (true positives and true negatives), but the relatively low score indicates the model's difficulty in consistently predicting all relevant classes.

- Handling the Dataset: The dataset consists of 12 image bands, each at three different resolutions. To manage this, three separate neural networks are required to process each resolution independently. These networks are then concatenated into a final combined network. While this architecture is necessary for handling the varying band resolutions, it effectively triples the size of the model. To compensate for this, the size of each individual network must be reduced to keep the model lightweight. However, reducing the size of these smaller networks also compromises their performance, leading to a loss in efficiency and a further decline in classification accuracy.

- Challenges of Lightweight Models: Lightweight models are particularly ill-suited for multilabel classification, especially when dealing with a high number of classes like the 43 in this dataset, combined with the challenge of processing image bands of different size. Multiclass classification with a large number of categories and diverse input data like this typically requires more robust, larger models for better performance, making the use of a lightweight architecture a significant limitation in this context.

- Below are the detailed architecture of the model.