Land-use and Land-cover Classification using EuroSAT Dataset

Master in Data Science, FIB
Advanced Machine Learning (AML)

December 15th, 2023

Pau Comas Herrera - Odysseas Kyparissis

Project Proposal

The advent of freely accessible satellite imagery, such as the Sentinel-2 data in the Copernicus program[1], has opened up new avenues for earth observation applications. Land use and land cover classification is a fundamental task in utilizing this wealth of satellite data. In this project, we aim to apply advanced machine learning techniques to the EuroSAT dataset[2], a novel dataset comprising **27,000** labeled and geo-referenced images based on Sentinel-2 satellite imagery, which was presented in the paper titled *EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification*[3]. It is important to mention that the dataset original version has a multi-spectral nature, covering *13* spectral bands. However, the objective of the project is focused on the RGB version of the EuroSAT dataset, with an emphasis on understanding and classifying land use patterns.

The specific dataset was selected due to its openly accessible nature, as well as, its significance in earth observation applications, including agriculture, disaster recovery, climate change monitoring, urban development, and environmental studies. The RGB version of the dataset offers color information crucial for a comprehensive analysis of land use patterns. As mentioned before, the dataset includes **27,000 images** categorized into **10 distinct land use and land cover classes**, including ***Residential, Industrial, Highway, River, Pasture, Herbaceous Vegetation, Annual Crop, Permanent Crop, Forest, and Sea/Lake***. Additionally, the **size** of all the images is **64 × 64** pixels. Here is the dataset's overview:
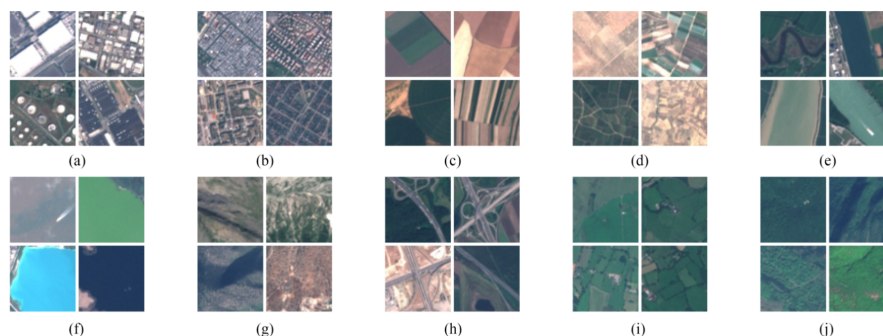
EuroSAT Dataset Overview[4]. (a) Industrial buildings. (b) Residential buildings. (c) Annual crop. (d) Permanent crop. (e) River. (f) Sea and lake. (g) Herbaceous vegetation. (h) Highway. (i) Pasture. (j) Forest.

Moreover, the paper reports an impressive overall classification accuracy of **98.57%** with a fine-tuned ResNet-50[5] CNN using single-band images, showcasing the dataset's potential. The term single-band images refers to the fact that the input is an image consisting of information from a single spectral band on all three input channels.

To conclude, the aim of the project is to contribute to the field of land use and land cover classification by harnessing the potential of the EuroSAT dataset. The focus on the RGB version allows for in-depth exploration of color information in land use patterns, offering insights into the challenges and opportunities specific to this aspect of satellite imagery classification.

---

[1] https://sentiwiki.copernicus.eu/web/copernicus-programme
[2] https://github.com/phelber/eurosat
[3] https://arxiv.org/abs/1709.00029
[4] https://arxiv.org/abs/1709.00029
[5] https://arxiv.org/abs/1512.03385