SSIP 2023

SUMMER SCHOOL ON IMAGE PROCESSING (SSIP) 2023, NOVI SAD

Teodora Salea,
Department of Computer Science,
West University of Timisoara, Romania
https://www.math.uvt.ro

Experiences in Building a Dataset for Semantic Segmentation. Use-case: Crop Type Mapping using Satellite Data

With the increasing volume of collected Earth observation (EO) data, artificial intelligence (AI) methods have become state-of-the-art in processing and analyzing them. However, there is still a lack of high-quality, large-scale EO datasets for training robust networks. This paper presents AgriSen-COG, a large-scale benchmark dataset for crop type mapping based on Sentinel-2 data. AgriSen-COG deals with the challenges of remote sensing (RS) datasets. First, it includes data from five different European countries (Austria, Belgium, Spain, Denmark, and the Netherlands), targeting the problem of domain adaptation. Second, it is multitemporal and multiyear (2019–2020), therefore enabling analysis based on the growth of crops in time and yearly variability. Third, AgriSen-COG includes an anomaly detection preprocessing step, which reduces the amount of mislabeled information.

AgriSen-COG comprises 6,972,485 parcels, making it the most extensive available dataset for crop type mapping. It includes two types of data: pixel-level data and parcel aggregated information. By carrying this out, we target two computer vision (CV) problems: semantic segmentation and classification. To establish the validity of the proposed dataset, we conducted several experiments using state-of-the-art deep-learning models for temporal semantic segmentation with pixel-level data (U-Net and ConvStar networks) and time-series classification with parcel aggregated information (LSTM, Transformer, TempCNN networks). The most popular models (U-Net and LSTM) achieve the best performance in the Belgium region, with a weighted F1 score of 0.956 (U-Net) and 0.918 (LSTM).The proposed data are distributed as a cloud-optimized GeoTIFF (COG), together with a SpatioTemporal Asset Catalog (STAC), which makes AgriSen-COG a findable, accessible, interoperable, and reusable (FAIR) dataset.