A mixed spectral and spatial Convolutional Neural Network for Land Cover Classification using SAR and Optical data

Gopal Singh Phartiyal (1), Nicolas Brodu (2), Dharmendra Singh (1), and Hussein Yahia (2)

(1) Indian Institute of Technology Roorkee, Roorkee, India, (2) GeoStat, INRIA, France

Today, both SAR and optical data are available with good spatial and temporal resolutions. The two data modalities complement each other in many applications. There are numerous approaches to process the two data modalities, separately or combined. Domain or modality specific approaches such as polarimetric decomposition techniques or reflectance based techniques cannot work with the two datasets combined together. Data fusion approaches incur information loss during the process and are highly application specific. Machine learning (ML) approaches can operate on the combined dataset but have their own advantages and disadvantages. There is a need to explore new ML based approaches to achieve higher performance. Convolutional neural networks (CNNs) are young, trending, and promising ML tools in remote sensing applications. CNNs have the capability to learn complex features exclusively from data. Data from the two modalities can thus be brought together and processed with increased performance.

In this paper an attempt is made to analyze CNN capabilities to perform land cover classification using multi-sensor data. SAR data used in this study is L band fully polarimetric PALSAR 2 data with 6 meter spatial resolution. Three basic polarimetric bands, namely, HH, HV, and VV, and four derived bands (polarization signatures) are used. Six multispectral Landsat 8 bands, pan sharpened and resampled at 6 meter spatial resolution, are used as optical data. All 13 features are stacked together and fed as input data to the proposed CNN. The areas selected for study are Haridwar and Roorkee regions of northern India. This study introduces a CNN where convolution is performed both spatially and spectrally. We show how this is an advantage over performing only spatial convolution. Five land cover classes namely, urban, bare soil, water, dense vegetation, and agriculture are considered.

The CNN is trained on more than 1200 ground truth class data points measured directly on the terrain. The classification shows results with good generalization. Comparison with other classifiers such as SVMs shows that the proposed approach provides better classification results in terms of generalization, although the cross-validation accuracy is on the same order. The evaluation of the generalization of the classified image is done using ground truth knowledge on selected subset areas in the study area.