



Deep learning based F-Mask alternative for Sentinel-2 images in polar regions

Thorsten Seehaus¹, **Kamal Gopikrishnan Nambiar**², Veniamin Morgenshtern², Philipp Hochreuther¹, and Matthias Braun¹

¹Institute of Geography, Friedrich-Alexander University Erlangen-Nürnberg, 91054 Erlangen, Germany;

²Chair of Multimedia Communications and Signal Processing, Friedrich-Alexander University Erlangen-Nürnberg, 91054 Erlangen, Germany;

Screening clouds, cloud shadows, and snow is a critical pre-processing step that needs to be performed before any meaningful analysis can be done on satellite image data. The state of the art 'F-Mask' algorithm, which is based on multiple pixel-level threshold tests, segments the image into clear land, cloud, cloud shadow, snow, and water classes. However, we observe that the results of this algorithm are not very accurate in polar and tundra regions. The unavailability of labeled Sentinel-2 training datasets with these classes makes the traditional supervised machine learning techniques difficult to implement. Experiments with large, noisy training data on standard deep learning classification tasks like CIFAR-10 and ImageNet have shown neural networks learn clean labels faster than noisy labels.

We present a multi-level self-learning approach that trains a model to perform semantic segmentation on Sentinel-2 L1C images. We use a large dataset with labels annotated using the F-mask algorithm for the training, and a small human-labeled dataset for validation. The validation dataset contains numerous examples where the F-mask classification would have given incorrect labels. At the first step, a deep neural network with a modified U-Net architecture is trained using a dataset automatically labeled with the F-mask algorithm. The performance on the validation dataset is used to select the best model from the step, which would then be used to generate more training labels from previously unseen data. In each of the subsequent steps, a new model is trained using the labels generated using the model from the previous step. The amount of data used for training increases with each step and the application of techniques like data augmentation and dropout improves the generalization of the trained model. We show that the final model from our approach can outperform its teacher, i.e. F-Mask algorithm.