

Using Deep Learning To Generate Fractional Vegetation Cover From Multispectral Data

Peter Schwind¹, Kevin Kühl¹, David Marshall Ingram¹, Martin Bachmann¹, Uta Heiden¹

¹ German Aerospace Center (DLR), Earth Observation Center (EOC), Weßling, Germany

Keywords: Earth Observation, Fractional Vegetation Cover, Soils, Deep Learning, Hyperspectral

Challenge

The launch of hyperspectral missions such as EnMAP, DESIS, PRISMA and EMIT in recent years has led to substantial improvements in the measurement of soil and vegetation cover world-wide. Within the Earth Observation Center at DLR, an fCover processor has been developed, which extracts and classifies endmembers from hyperspectral data and, following a spectral unmixing step, generates soil and vegetation cover maps (containing fractions between 0-100% for active vegetation, non-photosynthetically active vegetation and bare soil).

Even though the availability of hyperspectral data has increased significantly, the overall world coverage and revisit times are not yet comparable to multispectral missions such as Sentinel-2 or Landsat 8. Ideally, a method should be devised which would exploit both the high spectral resolution of EnMAP data and the high spatial and temporal coverage of S2 data. In this work, such a method, based on deep learning vector regression, is presented.

Methodology

To generate fCover maps based on Sentinel-2 imagery, the following approach is proposed: A deep learning vector regression model was trained using fCover labels generated on EnMAP imagery and corresponding Sentinel-2 imagery as inputs. Several image segmentation/classification methods (U-Net, Res-Net, FSKNet, HybridsN) were adapted and tested to perform the required regression task. Out of these, HybridSN was evaluated more vigorously, since initial tests showed the most promising results. Even though originally designed for hyperspectral image classification, adapting HybridSN for a multispectral vector regression problem is straightforward: The kernel size of the three initial convolution layers has to be reduced in the third dimension to account for the lower amount of input bands and the final Softmax activation has to be replaced by a Linear activation step to obtain a vector with three scalars (see Figure 1).

To train the modified HybridSN algorithm, 40 EnMAP and Sentinel-2 pairs each acquired at approximately the same time were selected. Out of the 40 scene pairs, 37 were used to train the model and 3 were used for validation. After masking out cloudy, urban and water areas, the Sentinel-2 scenes were split into 24,081,056 training patches with a size of 25x25x10, with the corresponding (normalized) label vectors computed by fCover on the EnMAP imagery. The model was subsequently trained for ten epochs on a single NVIDIA GeForce RTX 2080 Ti.

Results

The training over 10 epochs took approximately 20 hours. The number of epochs could probably be reduced in the future as the training (0.015 MSE) and validation losses (0.021 MSE) already reached their minimum points at the second and third epochs, respectively. The weights obtained after the third epoch

were subsequently used for the creation of fCover maps from Sentinel-2 data. An example of such a map, created for an area not included in the training data, can be compared to an fCover map extracted using the fCover processor in Figure 2. Even though the DL based approach in this case somewhat underestimated the photosynthetically active vegetation, the two maps look very similar overall. The MSE between these two scenes is 0.031 and it took 18 minutes to extract the fCover map from a full Sentinel-2 tile (110x110 km²).

In a more detailed analysis of all the individual training images, it could be observed that the MSE ranges from 0.007 to 0.044. Some of the comparatively higher losses might be explained by the temporal distance between the training pairs, while in others it might indicate consistency problems within the used fCover references, which were so far created without absolute ground truth. The overall relatively low MSE for training and validation data however, indicates that the proposed method is well suited for the task of fCover generation from multispectral data.

Outlook for the future

While these first results already look very promising, there is potential for improvement. First of all, the currently used training data is not optimal as it is often not possible to find suitable EnMAP data for the corresponding Sentinel-2 scene. For example, some of the used training pairs have a temporal distance of more than 10 days, which might introduce inconsistencies (e.g. harvested fields) in the model training. The acquisition of more data over time by the relatively young EnMAP mission should improve the quality of these training pairs.

Concerning the method evaluation, so far, the predicted results were only compared to the fCover outputs without an absolute reference. It would also make sense to validate the fCover maps, both derived from EnMAP and Sentinel-2 data, with actual ground truth values.

Finally, optimizations in the used DL model itself should be investigated. While the used HybridSN model with minor adaptations already delivered very robust results, it should be kept in mind that this model was originally developed for a very different purpose (hyperspectral image classification). There are many parameters (band selection, patch size, kernel sizes, layers, etc.) which could be tuned to try to improve accuracy, robustness and runtime of the presented methodology.

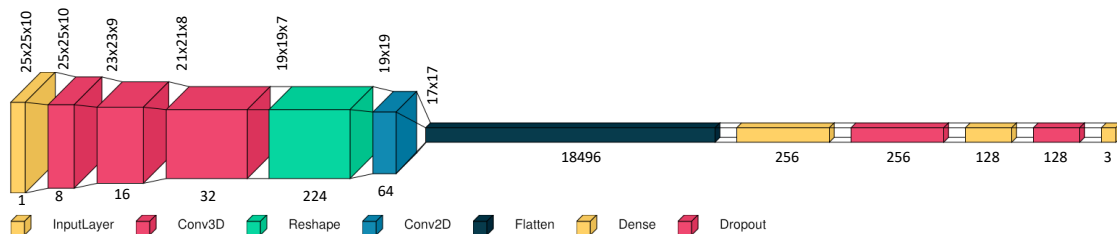


Figure 1 Adapted HybridSpectralNet (HybridSN) Model

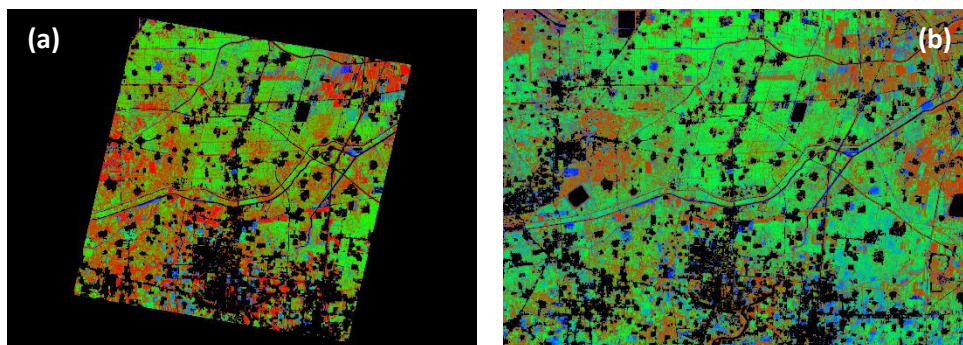


Figure 2 (a) fCover map derived from EnMAP scene using the fCover processor and (b) fCover map derived from Sentinel-2 scene using DL regression for the same region close to Binzhou, China (Legend: red: photosynthetically active vegetation (PV); green: non-photosynthetically active vegetation (NPV); blue: bare soil (BS). The intensity of the colours is related to the cover fractions.)