LARGE-SCALE VEGETATION HEIGHT MAPPING FROM SENTINEL DATA USING DEEP LEARNING

Anders U. Waldeland, Arnt-Børre Salberg,
Øivind D. Trier

Norwegian Computing Center
Department SAMBA
P.O. Box 114 Blindern, N-0314 OSLO

Andreas Vollrath

ESA - European Space Agency
ESRIN Largo G. Galilei 1,
00044 Frascati (RM), Italy

ABSTRACT

The deep learning revolution in computer vision has enabled a potential for creating new value chains for Earth observation that significantly enhances the analysis of satellite data for tasks like land cover mapping, change analysis, and object detection. We demonstrate a deep learning based value chain for the task of mapping vegetation height in the Liwale region in Tanzania using Sentinel-1 and -2 data. As ground truth data we use lidar measurements, which are processed to provide the average vegetation height per Sentinel-2 pixel grid (10 m). We apply the UNet, which is a widely used neural network for segmentation tasks in computer vision, to estimate average vegetation height from the Sentinel data. Preliminary results show that we are able to map the forest extent with high accuracy, with an RMSE of 3.5 m for Sentinel-2 data and 4.6 m for the Sentinel-1 data.

***Index Terms*—** vegetation height, convolutional neural network, Sentinel-1/2

1. INTRODUCTION

Vegetation height may be used to characterize the structure of a forest. It is known to correlate with important biophysical parameters like primary productivity, above-ground biomass, and bio-diversity [1, 2]. In-situ observations are in practice only feasible for a limited number of sample plots and logging sites. Airborne light detection and ranging (lidar) can provide canopy height over ground maps densely and accurately, but the cost and the limited area covered makes it infeasible for large-scale monitoring.

Trier et al. [1] estimated vegetation height from Landsat data covering the Liwale area in Tanzania. A regression model between the average vegetation height computed from the lidar data and the specific leaf area vegetation index computed from the Landsat data, was established. By using all available Landsat acquisitions of the same area within 1 year, and producing a yearly estimate of vegetation height, the estimation error variance was reduced. The variance was further reduced by Kalman filtering the sequence of yearly estimates.

Lang et al. [2] also estimated the vegetation height, but from Sentinel-2 data. Their study areas were Gabon and Switzerland, and their apporach was to train a deep convolutional neural network (CNN) to regress per-pixel vegetation height. Their results showed good qualitative agreement with existing vegetation height maps, and the authors demonstrated that vegetation height maps with 10 m pixel-spacing can be derived at country scale from Sentinel-2 imagery. As stated by Lang et al. [2], single-pixel based prediction of vegetation height at 10m pixel spacing is not suitable due to physical phenomena like shadowing, roughness, and species distribution that extends across neighboring pixels. Deep CNN architectures like UNet are perfectly tailored to account for the spatial context of the problem.

In this study, we explore and compare Sentinel-1 and -2 data to estimate the height of dry tropical vegetation. We consider the same area in as Trier et al. [1], but establishes a regression model using a deep CNN to estimate the vegetation height from the Sentinel data. The network is based on the UNet [3] architecture, working in regression mode, and implemented in the deep learning framework for large-scale processing of Sentinel data proposed by Salberg and Waldeland [4].

2. STUDY AREA AND DATA

2.1. Study area

The study area was Liwale in Tanzania (S9° 54', E37° 38'). It covers 15,867 km^2 of Miombo woodlands with altitudes in the range 150–900 m above sea level. The rainfall pattern in Liwale is bi-modal with a dry season from June to October. A short rainy period usually starts in late November and lasts until January. Normally, there is a dry spell in February followed by a longer wet season that lasts from March until May.

Thanks to ESA - European Space Agency for funding.

2.2. Lidar data

The lidar data were collected along 1.4 km wide and parallel strips in a systematic design at 5 km intervals, i.e. there was a gap of 3.6 km between neighbouring strips. A 113 km (east-west) × 156 km (north-south) area was covered with 34 strips in the east-west direction.

From the lidar data set, average vegetation height was computed on the 10 m Sentinel-2 grid. Each lidar pulse had been labelled with class (“ground” or “other”) and return number. From all “ground” returns, a digital terrain model (DTM) was created at 1 m pixel spacing. From all the first returns, a digital surface model (DSM) was created at the same pixel spacing. By subtracting the DTM from the DSM, a normalized DSM (nDSM) was obtained, which may be used as an estimate of vegetation height. By aggregating this to the 10 m pixel-spacing Sentinel-2 grid, the resulting average vegetation height and fractional forest cover maps are regarded as exact for the purpose of developing a method to estimate forest features from satellite images with 10 m pixel spacing.

2.3. Sentinel-1 and -2 data

In this study training data are selected from the tiles Sentinel-2 tiles T37LCK and T37LDK, validation data from tile T37LDJ and test data from tile T37LCJ. They all overlapped with the lidar data. In total 72 tiles for the period 2016 – 2018 and months April to June were used.

Sentinel-1 data (dual-polarized, ascending direction) covering the Sentinel-2 tiles for the period October 2016 to September 2017 (in total 26 acquisitions) were used for the SAR part. The Sentinel-1 data was processed splitted according to the Sentinel-2 training, validation and test tiles.

3. PRE-PROCESSING

3.1. Sentinel-2

Bands 1, 2, 4, 5, 8, 8A, 9, 10, 11, 12 were selected and used for cloud detection. Clouds are detected using the “Sentinel Hub’s cloud detector for Sentinel-2 imagery”. This cloud detection method is based on a Light Gradient Boosting Machine (LightGBM) [5], which is a gradient boosting framework that uses tree based learning algorithms. Areas in the label images corresponding to clouded pixels were masked with an ignore value in order to prevent that cloud pixels were used to train the network.

For the forest estimation, we applied the Sentinel-2 bands with 10m (bands 2, 3, 4 and 8) and 20m (bands 5, 6, 7, 8A, 11 and 12) resolution. For each tile, the selected bands were calibrated to top-of-the-atmosphere reflectance using the attached metadata.

3.2. Sentinel-1

Normalized radar backscatter for each Sentinel-1 scene is processed to the CEOS conform Radiometrically Terrain Corrected product at 20m, taking into account both geometric as well as radiometric distortions along terrain slopes. In addition, the interferometric coherence was calculated for both polarizations between each consecutive pair of acquisition (12 - 24 days), with the earlier date defined as master and the following data as slave image. For the processing steps that depend on a terrain model, auxiliary SRTM 30m Digital Elevation Model (DEM) had been used [6].

Subsequently, each of the resulting products was then individually stacked in time, and a multi-temporal speckle filter had been applied on each stack in order to reduce noise. Multi-temporal statistics were calculated on the resultant time-series stack resulting in temporal composite layers.

For the backscatter in VV and VH polarisation, the minimum and standard deviation were calculated for each pixel over time. The use of the minimum value should assure that the contrast between temporally constant woody vegetation and other, more dynamic natural land cover types such as grasslands and agricultural fields is enhanced. In addition, SAR backscatter is increased by soil moisture, which decreases the capability to differentiate between different tree cover classes. The minimum value therefore assures that for each pixel the driest conditions are used. Similarly, the standard deviation of the backscatter is inversely related to tree cover. Closed forests with full tree cover exhibit a stable backscatter over time that results in a low standard deviation. The more the canopy opens, effects of soil moisture, and thus seasonality, affects the signal and the standard deviation over time increases.

For the coherence layers in VV and VH polarisation the maximum, standard deviation and average value over time were calculated. The coherence itself contributes information by separating urban from forest environments, which both exhibit higher backscatter values that often are not distinguishable. While large and rigid scattering objects (e.g. rocks, buildings) on the ground feature a high coherence, small unstable objects (e.g. leaves) lead to a decrease in coherence. Therefore, the interferometric coherence is inversely related to tree cover, because closed forests exhibit very low values of coherence. Since this behavior can be assumed temporally stable for forested areas, using the maximum coherence value over the observed time period should enhance the contrast between different tree cover classes similar to the minimum value of the backscatter, while simultaneously excluding temporal dynamic land cover classes as well as urban environments. The rationale for the use of the standard deviation again is to provide the machine learning algorithm with a feature that distinguishes temporal stability of coherence that is assumed for closed forests with more fluctuation of the signal for open forests. The average is considered useful since it drastically reduces noise, which is more prominent in the coherence layers then in the backscatter.

4. DEEP LEARNING FRAMEWORK

The underlying deep learning framework we apply is a pixel-to-pixel mapping network [4], where the network learns a pixel-wise mapping from Sentinel data to a given vegetation height. In this paper we have apply the UNet, which a convolutional neural network that was developed for biomedical image segmentation [3]. The network is based on the fully convolutional network [7] and its architecture was modified and extended to work with fewer training images and to yield more precise segmentation.

The network consists of a contracting path and an expansive path, which gives it the u-shaped architecture. The contracting path is a typical convolutional network that consists of repeated application of convolutions, each followed by a ReLU and a max pooling operation. During the contraction, the spatial information is reduced while feature information is increased. The expansive pathway combines the feature and spatial information through a sequence of up-convolutions and concatenations with high-resolution features from the contracting path.

The deep learning framework is implemented in PyTorch.

4.1. Loss function

Since the UNet performs a regression task, we apply the mean-squared error as loss function.

4.2. Sampling strategy

A problem we often encounter in real-life machine learning problems is imbalanced data. This often leads to biases in the trained machine learning algorithm if it is not accounted for. In the case of vegetation mapping, there is an overweight of samples with low vegetation height. This means that the CNNs are biased towards predicting low vegetation heights because this lead to the lowest loss value in the training. To account for this, we artificially changed the imbalance of the data when computing the loss by ignoring samples such that the height distribution is close to uniform.

4.3. Hyperparameters

The U-Net is learned from 256 × 256 image crops that are randomly sampled from the set of Sentinel training images. The training runs for 15000 iterations with a mini batch size of 16. We used a learning rate decay strategy, where we start with an initial learning rate of 0.0004 and decayed it with a factor of three every 3000 iteration. The model is tested on the validation set every 250th iteration, and the model giving the best validation loss is kept.

5. RESULTS

When predicting tree heights, the predictions are highly correlated with the ground truth data computed from the lidar measurements (Figs. 1 and 2). However, for both the Sentinel-1 and -2 data we tend to underestimate tree heights above 15 m (Fig. 1 and 2). For Sentinel-1 we tend to overestimate the tree height for values around 5 m (Fig. 2). The RMSE for the Sentinel-1 and -2 data were, 3.5 m and 4.6 m, respectively.

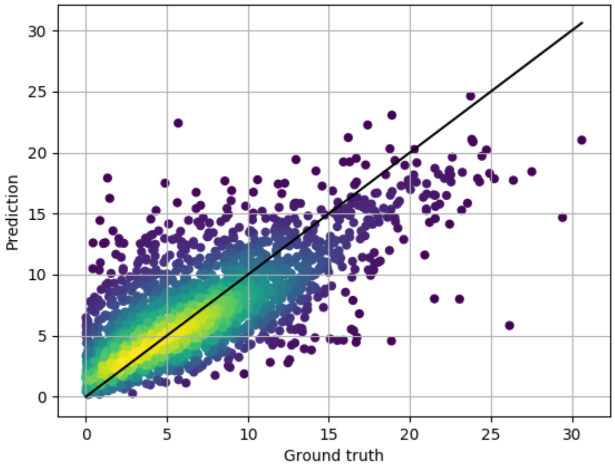

Fig. 1. Predicted tree height versus true tree height for Sentinel-2 data. The mean absolute error is 3.5 m.

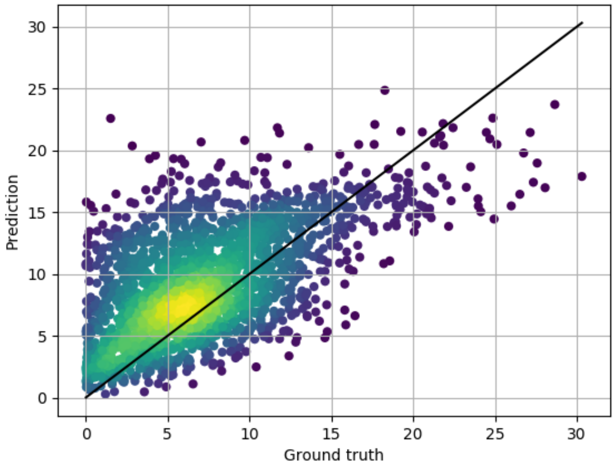

Fig. 2. Predicted tree height versus true tree height for Sentinel-1 data. The RMSE is 4.6 m.

When we evaluate the predicted vegetation height of a test patch, we observed that both the Sentinel-1 and -2 based

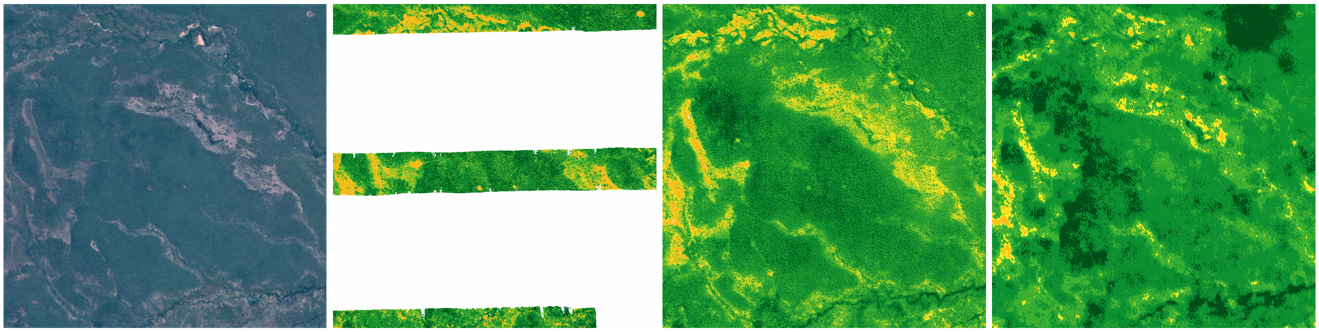

Fig. 3. Test patch results: Left: Sentinel-2 RGB image. Middle/left: Ground truth lidar measurements. Middle/right: Sentinel-2 based predictions. Right: Sentinel-1 based predictions. Yellow areas corresponds to low vegetation, whereas dark green areas correspond to tall vegetation.

predictions are able to capture the structure of the vegetation (Fig. 3). However, there are some differences between the predictions. In particular, for Sentinel-2 we are able to predict zero vegetation height for clearly bare ground areas (Fig. 3). In that respect, the Sentinel-2 predictions appears to be more close to the lidar mearsurements.

Trier et al. [1] also studied the task of estimating the average vegetation height in the Liwale area. However, their results are not directly comparable with ours since their predictions was related to the 30m Landsat pixels. Their test area were also different.

6. CONCLUSIONS

This work has demonstrated that both Sentinel-1 and -2 data may be used to predict the average vegetation height of dry tropical vegetation. For the Sentinel-2 data, we trained a deep CNN using TOA band values from Sentinel-2 images from different dates. For Sentinel-1, we processed the Sentinel-1 timescans into features. By doing this we reduce the amount of data, by simultaneously keeping the temporal dynamics, reduce the noise and standardize the input layers, which are desirable for machine learning analysis.

In this study the predictions based on Sentinel-2 data tend to be slightly better than the Sentinel-1 based predictions. However, we don't have the data to support a firm conclusion on this.

Future work will focus on extending the methodology to a multi-sensor approach, exploit other Sentinel-1 multi-temporal features, and other networks and loss functions.

7. REFERENCES

[1] Ø. D. Trier, A. B. Salberg, J. Haarpaintner, D. Aarsten, T. Gobakken, and E. Næsset, "Multi-sensor forest vegetation height mapping methods for Tanzania," *Eur. J. Remote Sensing*, vol. 5, no. 1, pp. 587–606, 2018.

[2] N. Lang, KK. Schindler, and J. D. Wegner, "Country-wide high-resolution vegetation height mapping with Sentinel-2," *Remote Sensing Environ.*, vol. 232, 2019.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, p. 234–241, 2015.

[4] A. B. Salberg and A. U. Waldeland, "Deep learning based value chain for Sentinel-2 land cover mapping," Poster presentation Living Planet Symposium, Milan, Italy, 2019.

[5] G. Ke, G. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Adv. Neural Inform. Process. Syst. (NIPS 2017), year = 2017, pages = 3149–3157.*

[6] T. Farr, P. Rosen, E. Caro, and R. Crippen, "The shuttle radar topography mission," *Reviews of Geophysics*, vol. 45, pp. 1–33, 2007.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2015, pp. 3431–3440.