Semi-supervised Sentinel-2 tree species detection

Martijn Vermeer[1], Tord K. Sørensen[1], David Völgyes[1], Daniele Fantin[1], Heidrun Miller[2]

[1]Science and Technology AS, Oslo, Norway [2]Allskog SA, Trondheim , Norway

Contact: Daniele Fantin (fantin@stcorp.no)

PDF

ALLSKOG
EID AV SKOGEIERE

Introduction

For forest management the availability of complete and up-to-date forest inventories is essential, with one of the most important parameters being the volumetric tree species distribution. Unfortunately, tree species mapping in Norwegian production forests is a time-consuming and largely manual process, leading to forest inventories that are often incomplete and/or outdated. Indeed, the determination of the tree species distribution is currently performed by a forestry expert, mainly by visual interpretation of aerial imagery and in some cases lidar data. High resolution aerial imagery is available, however campaigns are expensive and therefore infrequent. Satellite imagery, on the other hand, provides dense time series, but has a much lower resolution. The primary goal of the SENTREE project is to automate the classification of Norwegian main production tree species (Norway spruce, Scots pine and Birch) using semantic segmentation networks on a fusion of aerial and satellite data sources.

Experiments

Three major directions are explored during the SENTREE project:

- The utility of combining Sentinel-2 (S2) and aerial images.
- The noise robust training.
- Semi-supervised training.

Aerial and S2 combination: S2 bands and derived indices (NDVI, etc.) were aggregated over time by calculating minimum, maximum, median values, with or without cloud masking. This process resulted in almost 400 possible input features, which were fed into TPOT[1] for model based automatic feature selection. The 20 most relevant features were selected, upsampled and concatenated to the 3 RGB channels from the aerial images. Figure 2 depicts the labels, aerial and aggregated S2 features.

Loss functions

The model is based on a standard U-Net[2] and uses several custom loss functions. Firstly, the pixel level predictions are aggregated on stand level. This aggregation yields a vector of species distribution. The L1 measure of the difference of the label and prediction is used as the primary metric. Secondly, this scalar value is fed into a modified Huber loss function. This function is used to prevent overfit on already converged samples. By modifying the Huber loss, the contribution of the mislabeled stands can be limited.

Finally, the semi-supervised training exploits the nonetheless useful information provided by the unlabeled regions. During this step, two loss functions are implemented: focal-loss on pseudo-labeled pixels, and consistency loss over different augmentations [3].

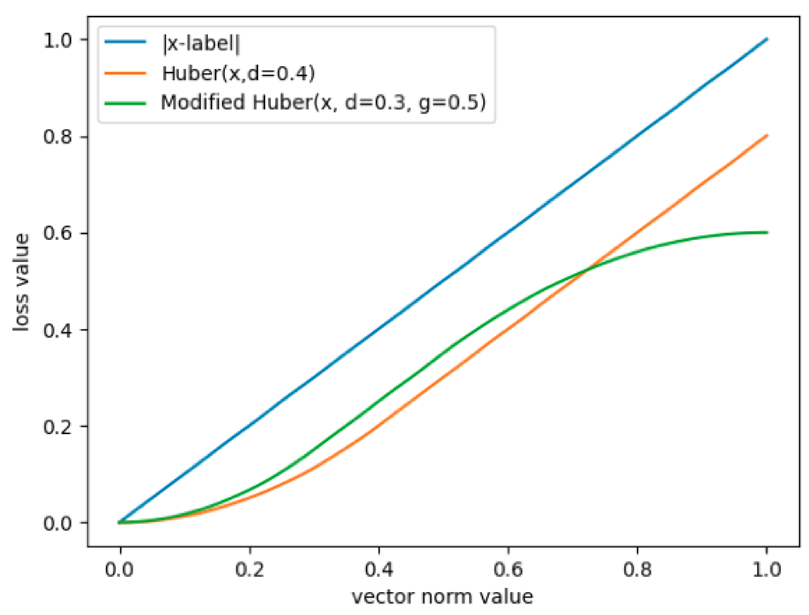

Figure 1: Modified Huber loss with reduced contribution for high values.

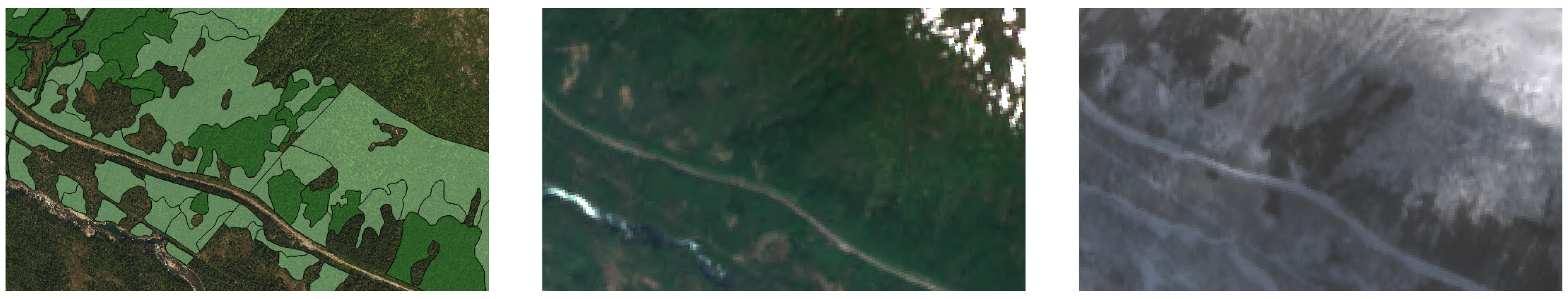

Figure 2: Aerial image with stand labels (left), S2 summer image (middle), S2 winter image (right).
The broadleaf areas are easy to separate from coniferous forest in the winter but not in the summer images.

Model predictions and noise detection

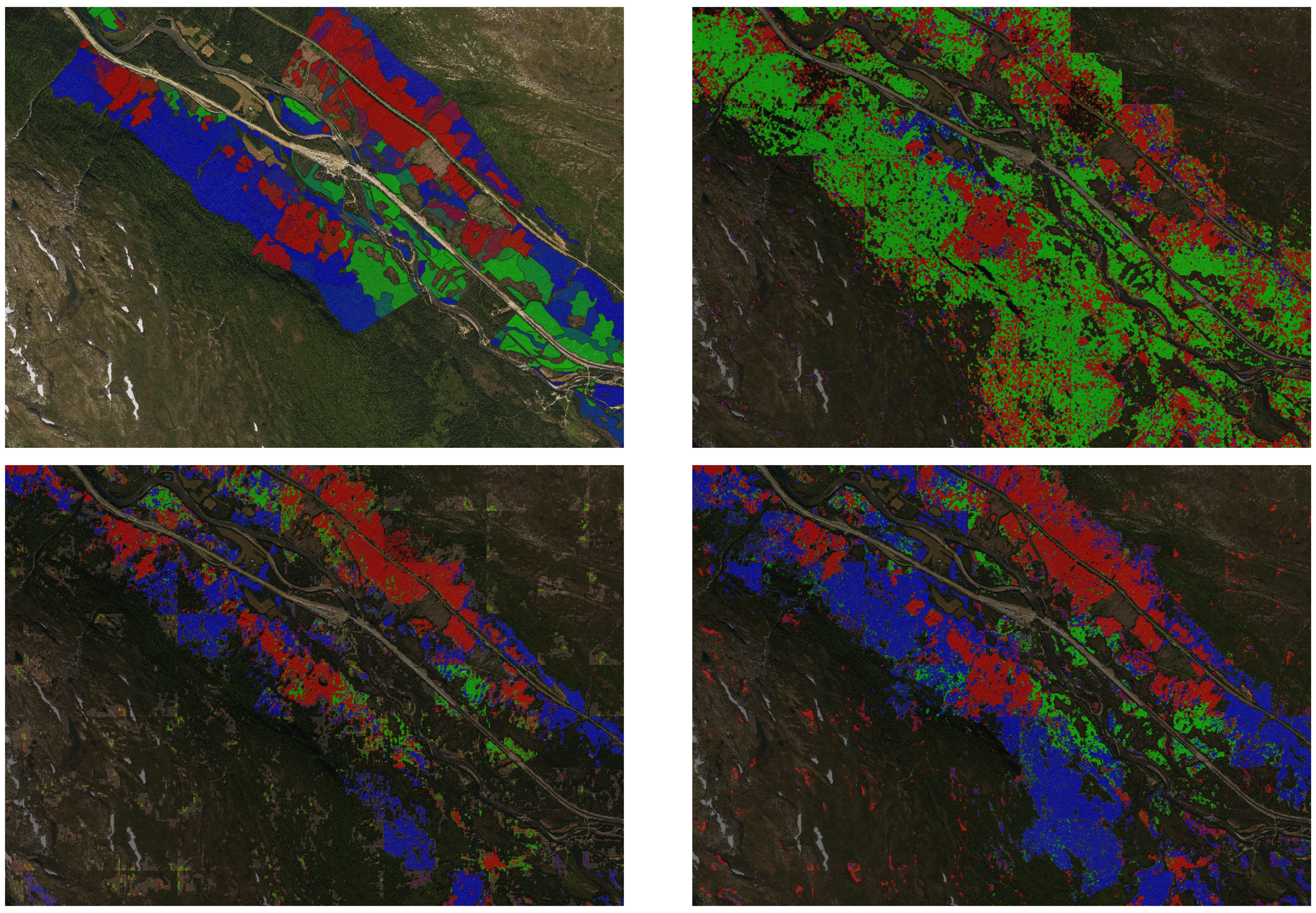

Figure 3: Baseline model prediction for a single test tile in Rana, Norway.
Ground truth (top-left), aerial model prediction (top-right), S2 model prediction (bottom-left), aerial+S2 model prediction (bottom-right).
Norway spruce, Scots pine and birch are visualised in red, green and blue respectively.

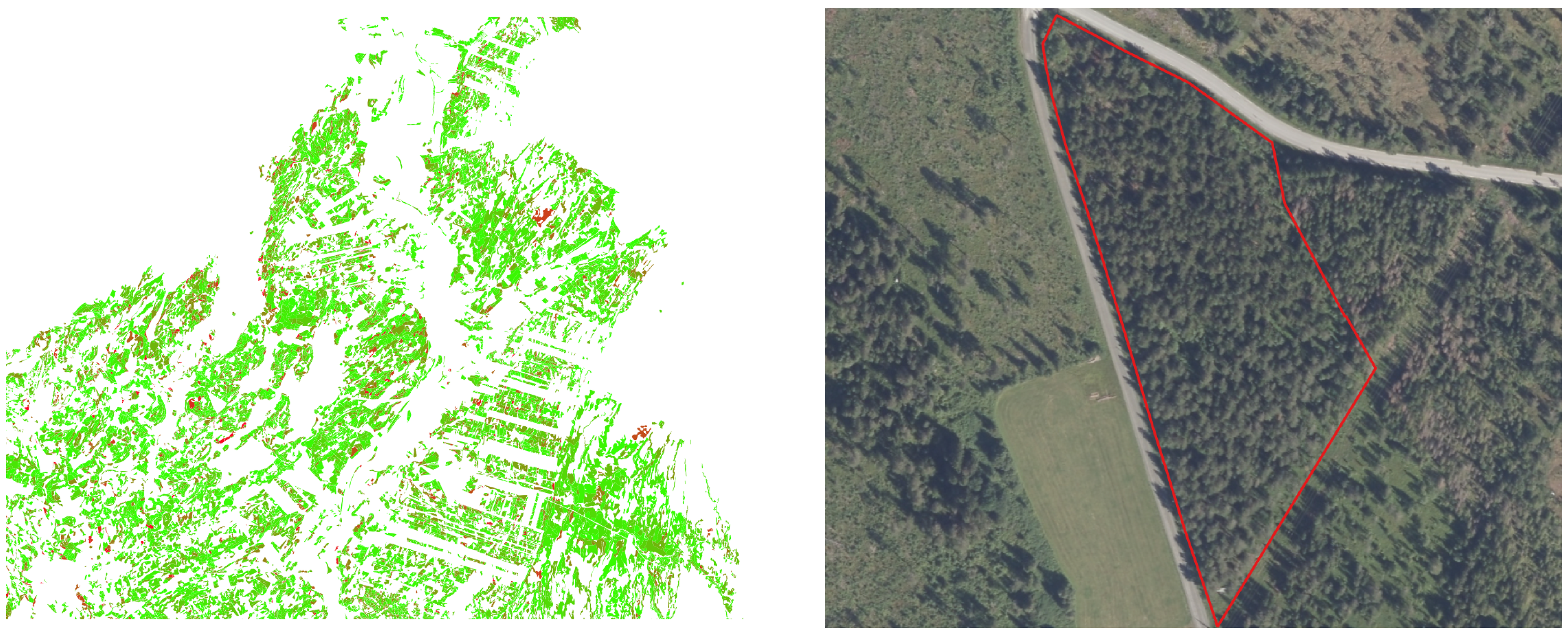

Figure 4: Left: High label-prediction map with mismatches highlighted in red. Right: One of the identified mismatching stands.
The label was Scots Pine while the model predicted Norway Spruce. Human reviewer also identified the stand as Norway Spruce.

Discussion

The baseline models above show that S2 time series data contributes to better generalization. The best model performance is provided by the combined S2-aerial model, followed by aerial-only and finally by S2-only.

Two noise-related approaches were applied: limiting the contribution of incorrect labels through noise robust losses (Figure 1), and detection of mislabeled stands. The latter can be done by identifying regions were labels and predictions strongly mismatch (Figure 4). Furthermore, we foresee using the area under the margin (AUM) [4] approach to inspect the behaviour of the loss curves during training in order to detect noise. The performance of all approaches mentioned above will also be tested with artificial label noise.

The semi-supervised training with consistency loss will be tested before the end of the project by removing labeled data. This training is expected to reduce the data pressure and improve generalization.

Acknowledgements

The project is financed by ESA under contract 4000136015.

References

[1] T. T. Le, W. Fu, and J. H. Moore, ``Scaling tree-based automated machine learning to biomedical big data with a feature set selector,'' *Bioinformatics*, vol. 36, no. 1, pp. 250--256, 2020.

[2] O. Ronneberger, P. Fischer, and T. Brox, ``U-net: Convolutional networks for biomedical image segmentation,'' *CoRR*, vol. abs/1505.04597, 2015.

[3] Q. Xie, Z. Dai, E. Hovy, T. Luong, and Q. Le, ``Unsupervised data augmentation for consistency training,'' *Advances in Neural Information Processing Systems*, vol. 33, pp. 6256--6268, 2020.

[4] G. Pleiss, T. Zhang, E. Elenberg, and K. Q. Weinberger, ``Identifying mislabeled data using the area under the margin ranking,'' *Advances in Neural Information Processing Systems*, vol. 33, pp. 17044--17056, 2020.