# Lidar-based Norwegian tree species detection using deep learning

Martijn Vermeer[1], Jacob Alexander Hay[1], David Völgyes[1], Zsófia Koma[2], Johannes Breidenbach[2], and Daniele Stefano Maria Fantin[1]

[1]Science and Technology AS
[2]Norwegian Institute of Bioeconomy Research (NIBIO), Division of Forest and Forest Resources
Corresponding Author: hay@stcorp.no

## Abstract

**Background**: The mapping of tree species within Norwegian forests is a time-consuming process, involving forest associations relying on manual labeling by experts. The process can involve aerial imagery, personal familiarity, on-scene references, and remote sensing data. The state-of-the-art methods usually use high-resolution aerial imagery with semantic segmentation methods.
**Methods**: We present a deep learning based tree species classification model utilizing only lidar (Light Detection And Ranging) data. The lidar images are segmented into four classes (Norway Spruce, Scots Pine, Birch, background) with a U-Net based network. The model is trained with focal loss over partial weak labels. A major benefit of the approach is that both the lidar imagery and the base map for the labels have free and open access.
**Results**: Our tree species classification model achieves a macro-averaged $F_1$score of 0.70 on an independent validation with National Forest Inventory (NFI) in-situ sample plots. That is close to, but below the performance of aerial, or aerial and lidar combined models.

## 1 Introduction

### 1.1 Motivation

Forest management and biodiversity hotspot identification require detailed, spatially continuous mapping of forests and tree species [1, 2]. These maps are ideally up-to-date and complete. Norway has had a National Forest Inventory since the 1920s [3], and model-assisted national mapping (SR16) is available [4], but often higher resolution maps are required for forest management tasks.

In Norway, the three main species used for production are Norway spruce, Scots pine, and Birch (more generally: broadleaf species). Current methods for mapping the spatial distribution rely heavily on updating previous maps using aerial and lidar sources. This multivariate data is processed by forestry experts, who carry out manual classification. Forest inventories are therefore both time-consuming and expensive. For this reason, maps are often reused between updates, and only actively logged stands will be passed through expert validation, which introduces a possible drift between the map and the actual stands.

To tackle these challenges, deep learning-based semantic segmentation seemed to be an obvious choice. Indeed, semantic segmentation with deep learning has seen increased usage within forest monitoring [5, 6].

We only found studies that are using additional data sources besides lidar, or use higher resolution gridded or point cloud data. Therefore, we performed a feasibility study to establish whether models trained on 1 m resolution, gridded lidar data is comparable to 0.2 m resolution RGB images [7] for Norwegian tree species mapping.

## 2 Background and related work

Machine learning methods used for tree species identification have recently shifted from model-assisted methods [4] and traditional machine learning-based methods [8, 9] towards deep learning [5, 6]. Deep learning is able to utilize contextual and texture information, which opens up opportunities for better utilization of the information available in the very high-resolution imagery data and open lidar data products.

Lidar frequently complements other data sources [4, 10] as the canopy height and tree-crown structure are a strong predictor variable for tree species identification. Other deep learning approaches utilize the point cloud itself for classifying tree species. However, these are rarely applicable across large areas due to the need for intensive data processing and often aim for single-tree classification [11, 12]. Overall, related published deep learning models are not directly comparable to a model that uses rasterized medium-resolution lidar-derived data products such as Digital Surface and Digital Terrain models. The importance of resolution is highlighted in [13], among others.

Proceedings of the 5th Northern Lights Deep Learning Conference (NLDL), PMLR 233, 2024.
© 2024 Martijn Vermeer, Jacob Alexander Hay, David Völgyes, Zsófia Koma, Johannes Breidenbach, & Daniele Stefano Maria Fantin. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

# 3 Materials and Methods

## 3.1 Data sources

We used three main data sources for training the segmentation model:

1. Digital terrain model (DTM) with a resolution of 1 m [14].
2. Digital surface model (DSM) with a resolution of 1 m [14].
3. Norwegian forest resources map (Skogressurskart, SR16 [4]) with a resolution of 16 m.

Input features were derived from the DTM and DSM, and SR16 served as a weak label.

In addition, we used land borders from OpenStreetMap [15] to limit the area only to land and exclude sea and major lakes. For internal validation, we split the data into train and validation regions.

The National Forest Inventory (NFI) was used as test data. The NFI is a continuous inventory system of permanent plots, with $\frac{1}{5}$ of the plots are measured every year. The permanent sample grid covers all land use classes, and if it does not contain trees, the land use class is assigned based on visual interpretation of aerial images. In forests, more than 120 variables are measured in circular plots of 250 $m^2$. For instance, the timber volume of each main tree species (Norway spruce, Scots pine, and Birch) with diameter $> 5$ cm has been recorded. More details can be found in [16].

In this study, we used the plots measured in the 2017-2021 NFI cycle within the 3 km $\times$ 3 km grid. Split plots had been excluded from the analysis. The plots without stocked volume were set as the "background" class, and the dominant tree species classes were determined based on the timber volume.

## 3.2 Lidar pre-processing

The lidar data was downloaded at a resolution of 1 m in gridded surface and elevation models. To capture tree canopies, we generated what is known as a canopy height model, which is the difference between the surface and elevation models:

$$\mathrm{CHM} = \mathrm{DSM} - \mathrm{DTM}$$

## 3.3 Label pre-processing

There are several challenges with matching 1 m resolution lidar data with 16 m resolution label data (SR16).

The first challenge is the time-wise mismatch. Both the SR16 map creation and the lidar data collection campaign took several years. This led to local mismatches where either canopy loss is not yet visible on SR16 but present in the lidar data, e.g. due to construction. The opposite also appeared in the data: SR16 missing areas that are clearly forested in the lidar data. This could be due to a later canopy loss that happened after the lidar campaign but before the SR16 update.

The second problem is the resolution of the data: 16 m label pixels which have a forest border approximately in the center of the pixels will inevitably contain almost 50% error by rounding the pixel into one of the classes. When such label is resampled to 1 m resolution, it could yield a several meters wide band of mislabeled pixels at the forest border due to the coarse resolution of the original label.

Finally, the SR16 map only contains tree species within forested areas, while at 1 m resolution, canopy openings would fall into the "background" class that does not exist in SR16 within forests.

In order to address these issues, we used the following procedure to pre-process the map into training labels. First, all unlabeled pixels were set to the background class. Second, all forest/no-forest border pixels were set to unlabeled, effectively tackling the forest borders. After this step, the forest map was upsampled to 1 m resolution by nearest neighbor upsampling. Finally, an 11 m $\times$ 11 m median filter was applied to the CHM data. Every location where the CHM median value was below 0.3 m was labelled as background.

A second round of labeling was performed after a full training. In this second round, larger areas (100 SR16 pixels (25600 $m^2$) or larger) which were consistently labeled as background and predicted as any kind of forest, or forest label that was consistently predicted as background, got unlabeled. That step is an attempt to remove systematic labeling errors which could severely impact classification performance [17] when the label and the features are different due to e.g. recent clearcuts. This attempt to reduce label noise reduced the total labeled area insignificantly ($< 0.1\%$). A second training was performed with this refined, partially unlabeled data, and only this second training on clean data is reported as result.

## 3.4 Deep Learning model

A standard U-Net network [18] was used. The network parameters are summarized in Table 1.

At inference time, the following extra post-processing steps were applied:

1. The prediction logits were blurred with a Gaussian filter ($\sigma = 1$ px) in order to reduce single pixel noise.
2. The tiles were cropped by removing a 64 pixel wide edge in order to reduce edge artifacts originating from tiling.

**Table 1.** Summary Parameters of a U-Net Network

| Parameter | Value |
|---|---|
| Image size | 2048 × 2048 px |
| Input channels | 2 (DTM, CHM) |
| Output channels | 4 |
| Depth | 5 |
| Initial number of filters | 16 |
| Upsampling method | convolution |
| Padding method | reflection |
| Normalization | instance |

### 3.5 Loss function

Focal loss [19] with $\gamma = 3$ was selected in order to reduce the contribution of easy samples. The softmaxed logits were cut off at a $p \leq 0.1$ level in order to avoid contribution from mislabeled pixels. The classes were weighted inversely with their frequency.

### 3.6 Augmentation

As augmentation of the raw data, standard geometric transformations were used (i.e. rotation and flipping). We did not change the brightness as the lidar values have a direct impact on the interpretation of species.

Another less known augmentation was used, which we call 'CowBatchMix' (see Fig. 1). Originating from the 'CowMix' augmentation [20], this augmentation mixes two batch samples in a cow pattern. The reasoning for its selection was that the forest edges were unlabeled, therefore, the network might have difficulties learning forest – no-forest transitions due to the lack of sharp transitions. Close to edge pixels are hard to classify even in fully supervised settings which could be mitigated by distance maps-based positional weighting [18] when borders are known, or by downweighting already learned samples, see our focal loss [19] choice. By mixing the images with CowMix, artificial edges were introduced with sharp transitions. This incentivizes the network to both learn local features and be precise about transitions. Furthermore, it acts as a regularizer against overfitting.

### 3.7 Study area

The study area was a large area south of Oslo, Norway. The data was split into train (8869 km$^2$) and validation data (541 km$^2$), depicted in Fig. 2. The independent test data was the NFI sample plots. The exact locations are not public.

## 4 Results

The full study area was evaluated using the independent NFI plots. If the plots were mixed, either in the reference or in the prediction, the largest contributor was used as the 'dominant species' or background class.

An inherent limitation of the evaluation comes from the plots. If the plot contains a mix of species, then small prediction errors might be amplified. For instance, a prediction might mismatch by just a few percent, but e.g., 45%-55% reference and 55%-45% for prediction is accounted as a single miss, just like a 0%-100% vs 100%-0% mismatch.

The confusion matrix results with derived metrics (precision, recall, $F_1$ score) are presented in Table 2. In addition, an example region is visualized in Fig. 3 to give a qualitative impression.

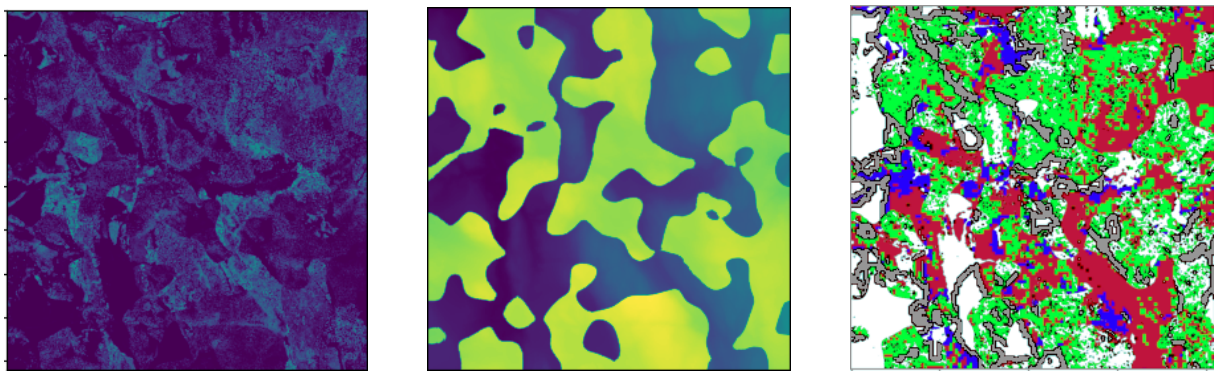
## 5 Discussion

### 5.1 Limitations of the validation

The training data is as independent of the testing data as possible. In Norway, NIBIO produces a 16 m resolution forest resource map (SR16) [4]. This map is model-assisted and generated in an automatic way with traditional machine learning methods. Since we use the SR16 map as input, there is a slight dependency chain between our input data and the test data. To the best of our knowledge, SR16 is used in all known Norwegian forest maps, including private companies' own inventories where they often use SR16 as a base and update their local area. We are not aware of any publicly available Norwegian tree species map that does not rely at least partially on SR16 and therefore indirectly on the NFI plots. This is an inherent limitation, and we did not see any way to overcome this challenge. The difference between our model output and SR16 maps shows that they are structurally different.

### 5.2 The benefits of the lidar source

There are several benefits of using lidar only for tree species detection. First, lidar is an active source; it lacks terrain shadows, which frequently cause problems in satellite or aerial imaging. Similarly, atmospheric effects do not affect it because campaigns are performed in clear sky conditions, and elevation measurements use time-of-flight measurements, which removes the issue with local reflectance (i.e., color of the forest). Finally, another significant benefit is the national coverage with freely available data. Aerial imagery campaigns also exist for Norway [7], but these datasets are expensive.



**(a)** canopy height model **(b)** digital terrain model **(c)** species labels and background

**Figure 1.** Cowmix augmentation: mixing two images in cow patterns introduces, best visible on the elevation data (B), but also noticeable in the two other subfigures. Species are in RGB, background is white, unlabeled is gray.

**Table 2.** Testing results: confusion matrix, per-class recall, precision, $F_1$ score, overall accuracy (OA), and macro-averaged $F_1$ score.

| | | NFI plots | | | | | |
|---|---|---|---|---|---|---|---|
| | | Background | Birch | Scots Pine | Norway Spruce | $\sum$ | Precision |
| Predictions | Background | **352** | 14 | 23 | 23 | 412 | 0.85 |
| | Birch | 9 | **64** | 15 | 36 | 124 | 0.52 |
| | Scots pine | 6 | 17 | **153** | 43 | 219 | 0.70 |
| | Norway spruce | 4 | 26 | 34 | **187** | 251 | 0.75 |
| | $\sum$ | 371 | 121 | 225 | 289 | 1006 | OA: 0.75 |
| | Recall | 0.95 | 0.53 | 0.68 | 0.65 | avg.$F_1$: | |
| | $F_1$ Score | 0.90 | 0.52 | 0.69 | 0.69 | 0.70 | |

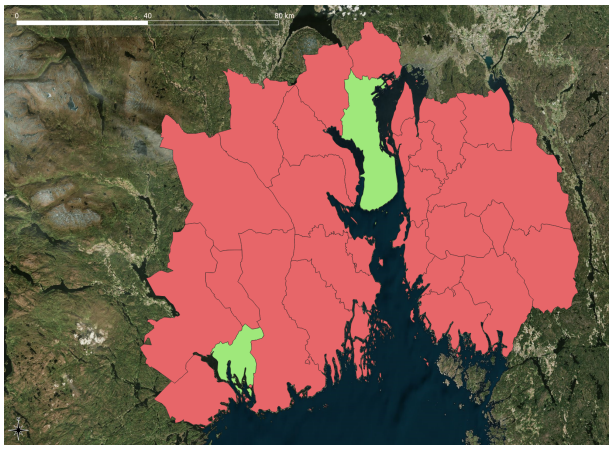


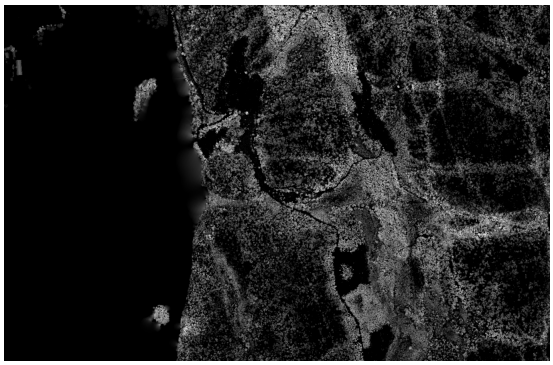
**Figure 2.** Training (red) and validation (green) data split in the study area (Viken, Norway).

## 5.3 The drawbacks of the lidar source

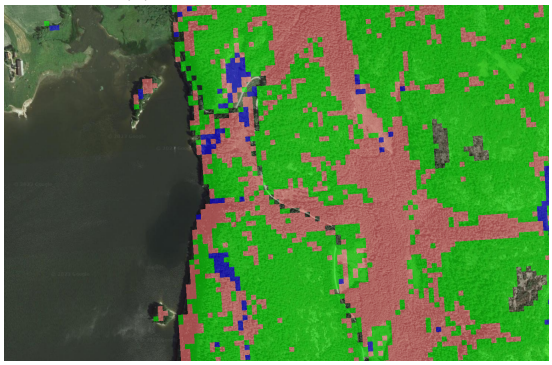
Several drawbacks are also present. First and foremost is the resolution. The national 1 m gridded data is very likely on the edge of feasibility, and seasonal changes of leaves could affect the detection of deciduous species. Moreover, the shape of the tree and the shape of the forest are not independent. In dense forest, the canopy can form a continuous unit where the canopy shape is the result of the interaction between trees, while in sparse forests the shape of individual trees can be more easily recognized. This issue is especially challenging with freshly planted, dense forest where the trees are small and densely packed. These forest structures are not only dependent on the species themselves and the age of the trees but also on the soil, water access, climate, etc. Due to these limitations, we think the model is regionally applicable: it is validated in the same type of area, but if climate, soil, potential species change, then a new regional model would be necessary. While the study area was selected with these issues in mind and tried to incorporate diverse data, verification on a larger scale remains future work.

In addition, lidar cannot capture spectral information. Spectral and hyperspectral information can be very important for tree species classification [21] when the number of species are larger and the species have similar height and crown structure. These regional differences in tree species distribution makes it hard to compare our results to related work. A study [22] performed in the middle of Norway reported macro-averaged $F_1$ score of approximately 0.75 by using semi-supervised deep learning

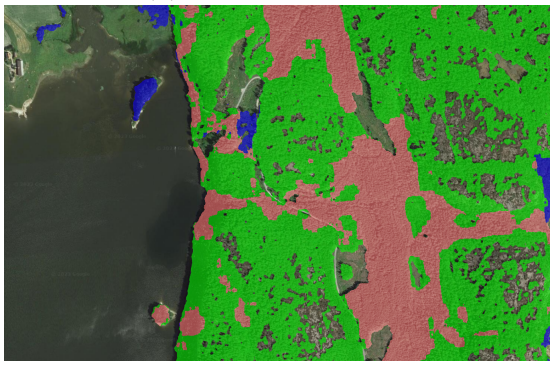
**(a)** Google Maps satellite view



**(b)** Canopy Height Model



**(c)** Satellite view overlaid with unmodified SR16



**(d)** Overlay with our model's prediction

**Figure 3.** Example prediction: the high-resolution prediction follows local edges more closely.

on 20cm resolution aerial imagery by using weak labels. While this study achieved slightly higher score (0.75) than our results (0.70), it used a significantly higher resolution input. On top of that the region is also slightly different, and it was compared to forest stand level majority classes instead of pointwise comparison.

### 5.4 Comparison to SR16

We have compared our lidar-based tree species classification map with the SR16-based tree species map. We have found that some errors from SR16 are independently distributed in the training data and the only-lidar map does not reproduce the same errors. This improvement can be the result of using the shape of the trees in training. However, both methods (SR16 and the one presented in this paper) are sensitive to the density of the forests. Density-related systematic mislabels in SR16 might also be learned by our model. On the other hand, our model is more precise with regards to the forest borders, including but not limited to forest roads and sparse forests with open spaces. While these results are not visible in the numerical comparison, we argue that this increased resolution has great utility for planning (e.g., housing, construction work), where the exact species are less important than the forest borders.

### 5.5 Impact and utility

Using a free data source and achieving comparable results to aerial-based tree species classification has the potential to reduce cost of accurate forest mapping. Some potential application areas:

1. Improving the accuracy of the estimation of forest attributes such as timber volume and biomass.

2. High-resolution mapping of tree species is important for planning sustainable forest management [1].

3. It can provide important information for climate adaptation and mitigation and biodiversity assessment and biodiversity hotspot identification [2].

## 6 Conclusion

Our work suggests that the use of lidar-only methods is not only feasible with regards to detection and segmentation of Norwegian tree species, but it also comes with the benefits of utilizing publically available, open access data. This allows the creation of larger, higher resolution maps, without dependence upon commercial data, at comparatively lower cost, without sacrificing much in terms of performance.

# 7 Future work

A current limitation concerns the study area. Although it encompasses approximately 9400 km$^2$, it is confined to a specific region of Norway. For our internal test, we produced a country-wide tree species map. However, the independent validation was solely focused on the South of Norway region, as detailed in earlier sections. In upcoming work, we anticipate extending the validation from this regional scope to a national one, and possibly to neighboring countries.

A potential direction for future work lies in enhancing prediction accuracy by integrating higher quality data sources. Examples include additional gridded lidar metrics, higher resolution gridded data, or more refined point cloud data.

Subsequent ablation studies might delve into evaluating the significance of particular choices, such as the label noise removal technique or the influence of the CowMix augmentation.

# Acknowledgment

This work was funded by the European Space Agency (ESA) as part of the SENTREE project (ESA Contract No 4000136015/21/I-DT-lr) in the "ESA AO/1-10468/20/I-FvO FUTURE EO-1 EO SCIENCE FOR SOCIETY PERMANENTLY OPEN CALL" program.

The authors thank Tord Kriznik Sørensen for his contributions to the technical infrastructure of the pre-cursor project, and for the discussions about tree species detection.

# References

[1] M. Pecchi, M. Marchi, V. Burton, F. Giannetti, M. Moriondo, I. Bernetti, M. Bindi, and G. Chirici. "Species distribution modelling to support forest management. A literature review". In: *Ecological Modelling* 411 (2019), p. 108817. ISSN: 0304-3800. DOI: 10.1016/j.ecolmodel.2019.108817.

[2] A. Felton, L. Petersson, O. Nilsson, J. Witzell, M. Cleary, A. M. Felton, C. Björkman, Å. O. Sang, M. Jonsell, E. Holmström, et al. "The tree species matters: Biodiversity and ecosystem service implications of replacing Scots pine production stands with Norway spruce". In: *Ambio* 49 (2020), pp. 1035–1049. ISSN: 0044-7447. DOI: 10.1007/s13280-019-01259-x.

[3] J. Breidenbach, L. T. Waser, M. Debella-Gilo, J. Schumacher, J. Rahlf, M. Hauglin, S. Puliti, and R. Astrup. "National mapping and estimation of forest area by dominant tree species using Sentinel-2 data". In: *Canadian Journal of Forest Research* 51.3 (2021), pp. 365–379. ISSN: 0045-5067. DOI: 10.1139/cjfr-2020-0170.

[4] M. Hauglin, J. Rahlf, J. Schumacher, R. Astrup, and J. Breidenbach. "Large scale mapping of forest attributes using heterogeneous sets of airborne laser scanning and National Forest Inventory data". In: *Forest Ecosystems* 8 (2021), pp. 1–15. ISSN: 2197-5620. DOI: 10.1186/s40663-021-00338-4.

[5] S. Nezami, E. Khoramshahi, O. Nevalainen, I. Pölönen, and E. Honkavaara. "Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks". In: *Remote Sensing* 12.7 (2020), p. 1070. DOI: 10.20944/preprints202002.0334.v1.

[6] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz. "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 173 (2021), pp. 24–49. ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2020.12.010.

[7] *Norge I Bilder*. URL: https://norgeibilder.no/.

[8] M. Immitzer, C. Atzberger, and T. Koukal. "Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data". In: *Remote Sensing* 4.9 (2012), pp. 2661–2693. ISSN: 2072-4292. DOI: 10.3390/rs4092661.

[9] F. E. Fassnacht, H. Latifi, K. Stereńczak, A. Modzelewska, M. Lefsky, L. T. Waser, C. Straub, and A. Ghosh. "Review of studies on tree species classification from remotely sensed data". In: *Remote Sensing of Environment* 186 (2016), pp. 64–87. ISSN: 0034-4257. DOI: 10.1016/j.rse.2016.08.013.

[10] A. Ghosh, F. E. Fassnacht, P. Joshi, and B. Koch. "A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales". In: *International Journal of Applied Earth Observation and Geoinformation* 26 (2014), pp. 49–63. ISSN: 1569-8432. DOI: 10.1016/j.jag.2013.05.017.

[11] B. Liu, H. Huang, Y. Su, S. Chen, Z.-y. Li, E. Chen, and X. Tian. "Tree Species Classification Using Ground-Based LiDAR Data by Various Point Cloud Deep Learning Methods". In: *Remote Sensing* 14 (2022), p. 5733. ISSN: 2072-4292. DOI: 10.3390/rs14225733.

[12] S. Mustafic, M. Hirschmugl, R. Perko, and A. Wimmer. "Deep Learning for Improved Individual Tree Detection from Lidar Data". In: *IGARSS – IEEE International Geoscience and Remote Sensing Symposium* (2022), pp. 3516–3519. ISSN: 2153-7003. DOI: 10.1109/IGARSS46834.2022.9884012.

[13] Y. Gan, Q. Wang, and A. Iio. "Tree Crown Detection and Delineation in a Temperate Deciduous Forest from UAV RGB Imagery Using Deep Learning Approaches: Effects of Spatial Resolution and Species Characteristics". In: *Remote Sensing* 15 (2023), p. 778. ISSN: 2072-4292. DOI: 10.3390/rs15030778.

[14] *Nasjonal detaljert høydemodell*. URL: https://www.kartverket.no/geodataarbeid/nasjonal-detaljert-hoydemodell.

[15] OpenStreetMap contributors. *Planet dump retrieved from https://planet.osm.org*. https://www.openstreetmap.org. 2017.

[16] J. Breidenbach, A. Granhus, G. Hylen, R. Eriksen, and R. Astrup. "A century of National Forest Inventory in Norway – informing past, present, and future decisions". In: *Forest Ecosystems* 7 (2020). ISSN: 2197-5620. DOI: 10.1186/s40663-020-00261-0.

[17] A. Maiti, S. J. O. Elberink, and G. Vosselman. "Effect of Label Noise in Semantic Segmentation of High Resolution Aerial Images and Height Data". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (2022). ISSN: 2194-9050. DOI: 10.5194/isprs-annals-v-2-2022-275-2022.

[18] O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[19] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. "Focal loss for dense object detection". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2980–2988. DOI: 10.1109/TPAMI.2018.2858826.

[20] G. French, A. Oliver, and T. Salimans. "Milking CowMask for Semi-Supervised Image Classification". In: *VISIGRAPP*. 2020. DOI: 10.5220/0010773700003124.

[21] M. Dalponte, H. O. Ørka, T. Gobakken, D. Gianelle, and E. Næsset. "Tree Species Classification in Boreal Forests With Hyperspectral Data". In: *IEEE Transactions on Geoscience and Remote Sensing* 51.5 (2013), pp. 2632–2645. ISSN: 0196-2892. DOI: 10.1109/TGRS.2012.2216272.

[22] M. Vermeer, D. Völgyes, T. K. Sørensen, H. Miller, and D. Fantin. "Semi-and weak-supervised learning for Norwegian tree species detection". In: *Proceedings of the Northern Lights Deep Learning Workshop*. Vol. 4. 2023. DOI: 10.7557/18.6817.