TOWARDS GENERAL DEEP-LEARNING-BASED TREE IN-STANCE SEGMENTATION MODELS

Jonathan Henrich
Chairs of Statistics and Econometrics
Faculty of Economics
University of Göttingen
Germany
jonathan.henrich@uni-goettingen.de

Jan van Delden

Institute of Computer Science
University of Göttingen
Germany
jan.vandelden@uni-goettingen.de

ABSTRACT

The segmentation of individual trees from forest point clouds is a crucial task for downstream analyses such as carbon sequestration estimation. Recently, deep-learning-based methods have been proposed which show the potential of learning to segment trees. Since these methods are trained in a supervised way, the question arises how general models can be obtained that are applicable across a wide range of settings. So far, training has been mainly conducted with data from one specific laser scanning type and for specific types of forests. In this work, we train one segmentation model under various conditions, using seven diverse datasets found in literature, to gain insights into the generalization capabilities under domain-shift. Our results suggest that a generalization from coniferous dominated sparse point clouds to deciduous dominated high-resolution point clouds is possible. Conversely, qualitative evidence suggests that generalization from high-resolution to low-resolution point clouds is challenging. This emphasizes the need for forest point clouds with diverse data characteristics for model development. To enrich the available data basis, labeled trees from two previous works were propagated to the complete forest point cloud and are made publicly available at https://doi.org/10.25625/QUTUWU.

1 INTRODUCTION

As global climate change accelerates, driven by anthropogenic activities, the role of forests in carbon sequestration, biodiversity preservation, and regulation of local and global climatic conditions has been brought into sharp focus. To investigate how forests contribute to these environmental aspects, quantifiable data on their structure and development is urgently needed. In this context, technologies that enable the creation of holistic, three-dimensional representations of forests in the form of point clouds play a vital role. Such technologies are terrestrial or mobile laser scanning (TLS, MLS), but also laser scanning via low-flying unmanned aerial vehicles (UAV). Such forest point clouds often need to be segmented into individual trees for further analysis, which is an instance segmentation problem. The most commonly used paradigm for tree segmentation is to first detect tree trunks and then assign the remaining points to individual trees based on hand-crafted features such as distance or local geometry (Trochta et al., 2017; Burt et al., 2019). However, laser scanning characteristics, forest structures, and interactions between trees are diverse. So, defining a fixed set of assignment rules and features that consistently lead to a good segmentation performance is a highly challenging task.

Advances in point cloud processing outside the forest domain show the advantage of performing instance segmentation using deep learning (Vu et al., 2022; Jiang et al., 2020), so that relevant features can be learned in a data-driven way. Only recently, these methods have been applied to the forest domain, yielding promising segmentation results (Xiang et al., 2023; Henrich et al., 2023). Since these methods are trained in a supervised way using specific datasets, a key challenge is to obtain general models that are applicable across a wide range of settings. In this context, an important question is how models generalize to out-of-domain settings. Differences in the training data are, for example, caused by different laser scanning characteristics or forest types. Training a

Table 1: Summary of the characteristics of the forest data used in this work. Number in parantheses of column *n* trees denotes number of trees of at least 10 m height. Summary for NIBIO, CULS, TU_WIEN and SCION is taken from Puliti et al. (2023b).

Name	Country	Reference	n plots	n trees	Annotated area (ha)	Forest type	Sensor
L1W	Germany	Henrich et al. (2023)	1	200 (200)	1.16	temperate deciduous forest	ZEB-Horizon
NIBIO	Norway	Puliti et al. (2023a)	20	575 (482)	1.21	coniferous dominated boreal forest	Riegl miniVUX-1 UAV
CULS	Czech Republic	Kuželka et al. (2020)	3	47 (47)	0.33	coniferous dominated temperate forest	Riegl VUX-1 UAV
TU_WIEN	Austria	Wieser et al. (2017)	1	150 (106)	0.55	deciduous dominated alluvial forest	Riegl VUX-1 UAV
SCION	New Zealand	Unpublished	5	135 (130)	0.33	non-native pure coniferous temperate forest	Riegl MiniVUX-1 UAV
RMIT	Australia	Unpublished	1	223 (92)	0.37	Native dry sclerophyll eucalypt forest	Riegl MiniVUX-1 UAV
LAUTX	Austria	Tockner et al. (2022)	6	514 (354)	0.83	temperate mixed forest	ZEB-Horizon
WYTHAM	England	Calders et al. (2022)	1	877 (608)	1.52	temperate deciduous forest	RIEGL VZ-400

supervised deep learning algorithm requires forest point clouds that come with segmentation labels. Although recent works have acknowledged this need and put considerable effort into making high-quality labeled forest point clouds publicly available (Puliti et al., 2023b; Henrich et al., 2023), the size and diversity of these datasets is still limited. Other works provide segmented trees that have been manually segmented (Tockner et al., 2022) or manually checked for quality assurance (Calders et al., 2022). However, these works do not include the non-tree points in their published data. Only if labels are available for the complete point cloud, it is possible to train a fully deep learning-based segmentation pipeline that does not require separate pre-processing steps.

This work makes two contributions: (1) The existing corpus of labeled forest point clouds is extended by propagating the publicly provided individual tree labels of two previous works (Tockner et al., 2022; Calders et al., 2022) to the complete point clouds. These point clouds are made publicly available. (2) An existing deep-learning-based tree segmentation model (Henrich et al., 2023) is trained with forest point clouds from different settings to provide insights into the generalization capabilities under domain-shift.

2 MATERIALS AND METHODS

2.1 LABELED FOREST DATA

The TreeLearn method can be trained on complete labeled forest point clouds that have a sufficiently high scanning resolution for all parts of a tree. The existing literature was searched for data that fulfils this criterion. First, there is the recently published FOR-instance dataset (Puliti et al., 2023b) in which tree labels and fine-grained semantic labels were manually added to point clouds from existing works. These point clouds have been captured via UAV-laser scanning and consist of diverse forest plots located in Norway (NIBIO), Czech Republic (CULS), Austria (TU_WIEN), New Zealand (SCION) and Australia (RMIT). In another recent work, tree labels for a forest plot located in Germany (L1W) were obtained using the Lidar360 software (GreenValley International, 2022) and then manually corrected. A summary of the characteristics of each dataset can be found in Table 1. More precise information can be found in the respective publications.

Apart from these point clouds, two published datasets were identified that consist of high-quality segmented trees obtained by an automatic segmentation algorithm that were either manually checked (WYTHAM, Calders et al., 2022) or corrected (LAUTx, Tockner et al., 2022) for quality assurance. The respective authors were contacted to obtain the complete unlabeled point clouds. These point clouds additionally contain non-tree points, i.e. belonging to the understory or ground, and non-annotated points, i.e. points that belong to trees but have not been annotated in the published datasets. For example, some parts of the tree crown that are hard to clearly assign to a specific tree might not have been annotated.

To obtain labels for the complete point clouds, the tree labels from the published datasets have to be propagated and the remaining points must be assigned to the classes "non-tree" or "non-annotated". This was done as follows:

1. For each point in the unlabeled forest point cloud, the most common tree label within a 0.1 m radius was assigned.

2. Among the remaining unlabeled points, non-tree points were identified using proximity-based clustering: All points that were within a 0.3 m distance to each other were linked and the largest connected component was labeled as non-tree points. The large grouping radius together with the high resolution of the point clouds ensured that all understory and ground points were added to the non-tree class.

3. The points that were still unlabeled at this stage represent tree points that have not been annotated and were assigned to the non-annotated class. This information can be used to disregard these points during training.

4. Finally, we visually inspected the point clouds to ensure that they were adequately divided into trees, non-tree points and non-annotated points. Remaining errors were manually corrected within a feasible scope. Specifically, one large tree was not segmented in the original labeled data of Calders et al. (2022) which was added, and the tree bases of Tockner et al. (2022) were corrected since they were only roughly segmented in the original labeled data.

For the given datasets, high-quality segmentation labels are only ensured when considering trees larger than 10 m, while assigning the rest as non-trees. In WYTHAM, smaller trees are inconsistently labeled, i.e. sometimes as a tree and sometimes as non-tree. In LAUTX, smaller trees have severe quality limitations. A correction of these mistakes was beyond the scope of this work. Therefore, only trees larger than 10 m were considered here.

2.2 SEGMENTATION METHOD

The model framework used in this study is TreeLearn (Henrich et al., 2023). It employs the widely-used grouping-based paradigm (Qi et al., 2019) for instance segmentation: The point cloud is processed using a 3D-UNet followed by pointwise semantic and offset prediction. The semantic prediction is used to classify points as tree or non-tree. The offset prediction aims to shift each point towards the respective tree base a point belongs to. After applying the predicted offset to each point, tree instances can be identified using density-based clustering. To account for memory limitations, the authors proposed a sliding window approach with subsequent merging of the results.

2.3 EXPERIMENTS

Using the labeled data presented in Section 2.1, TreeLearn was trained in three conditions: (i) In the first condition, only UAV-data was used (NIBIO, CULS, TU_WIEN, SCION). Most of these point clouds come from coniferous dominated forests. (ii) In the second condition, only TLS and MLS data (LAUTX, WYTHAM) were used, which come from mixed or deciduous forests. (iii) Lastly, all data was used for model training. In all three conditions, an area covering roughly 400 trees from WYTHAM was employed as the validation set. The number of trees in the training data in condition (i) and (ii) is roughly equal (765 vs. 762). Test performance was evaluated using L1W, a beech-dominated deciduous forest. Condition (i) assesses the effect of using out-of-domain data during training since the laser scanning characteristics and tree composition are substantially different from L1W. Condition (ii) represents in-domain data. In addition to quantitative test results on L1W, qualitative test results on a low-resolution UAV point cloud (RMIT) are presented.

Table 2: Segmentation results on L1W. Semantic and instance segmentation results in %. Results of TreeLearn trained on Lidar360 labels are taken from Henrich et al. (2023).

	Semantic Seg.	Detection		Instance Seg.
Training Data	Accuracy	FP Predictions	FN Trees	$F1$
MLS (Lidar360 labels)	**99.69**	0	0	93.98
+UAV	99.41	1	0	96.25
+MLS+TLS	99.64	1	0	**97.31**
+UAV+MLS+TLS	99.64	0	0	96.88

In each condition, the initial model weights were set to a publicly available TreeLearn checkpoint that has been obtained by training on large amounts of MLS data with non-corrected labels from a commercial software. From there, fine-tuning was performed for $12\,500$ iterations using the AdamW optimizer (Loshchilov & Hutter, 2017) with a weight decay of 10^{-3} and $\beta = [0.9, 0.999]$. The batch size was set to 2. A cosine learning rate schedule (Loshchilov & Hutter, 2016) with a maximum/minimum learning rate of $1 \times 10^{-3}/5 \times 10^{-5}$ was selected. Training examples were generated by randomly cropping squares of size 35 m by 35 m from the labeled forest point clouds. Only the inner 8 m by 8 m were considered during gradient computation so that the respective tree base of a tree point was ensured to be within the crop.

The performance on the L1W-dataset is evaluated based on the evaluation protocol detailed in Henrich et al. (2023). First, the tree detection performance is measured by the number of false positive and false negative predictions. To assess the semantic segmentation into tree and non-tree points, the accuracy is calculated. Instance segmentation performance is evaluated using the F1-score. It is calculated for each tree separately based on the number of true positive, false positive and false negative points and then averaged across all trees.

3 RESULTS AND DISCUSSION

Segmentation results were computed for the point cloud L1W (MLS, deciduous dominated). Results obtained by using the publicly available TreeLearn checkpoint without further training serve as a baseline (Table 3). Even when fine-tuning the model with out-of-domain data (UAV, coniferous dominated), instance segmentation performance in terms of the F1-score increases substantially from 93.98 % to 96.25 %. In terms of tree detection (one FP) and semantic segmentation (99.41%), the fine-tuned model performs slightly worse than the baseline. However, these aspects are less important compared to instance segmentation performance. For example, FP predictions can be manually discarded or merged into complete trees without much effort if they are not too frequent. When using in-domain data (MLS+TLS, deciduous dominated) for fine-tuning, instance segmentation performance is further increased to 97.31% (see Fig. 1 for a qualitative comparison). Since the number of trees used during training is roughly equal in the in-domain and out-of-domain conditions, the performance gap is most likely due to the domain-shift. However, other factors, such as differences in forest complexity, are hard to quantify and cannot be controlled for. When using all available data for training, test performance decreases slightly (96.88 %) compared to only using in-domain data. In addition to quantitative segmentation results on L1W, all three models were used to obtain qualitative results on RMIT, a low-resolution UAV point cloud (Fig. 2). These results suggest that an adequate segmentation performance on low-resolution UAV data can only be achieved when including it during training. If only MLS+TLS data is used, segmentation quality decreases drastically due to severe cases of merged trees.

4 CONCLUSION

In this paper, we trained the deep-learning-based tree segmentation method TreeLearn with data from various domains and systematically evaluated its test performance. It was shown that a model trained on out-of-domain coniferous dominated UAV point clouds can generalize to deciduous dom-
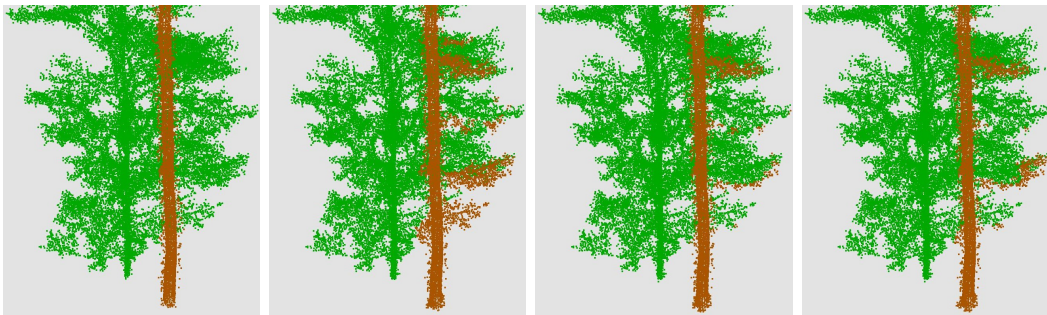
Figure 1: Fine-grained test results on an MLS point cloud (L1W). From left to right the images show (1) the ground truth segmentation and model results obtained by fine-tuning on (2) UAV, (3) MLS+TLS and (4) all data. Results are best when in-domain data is included during training.
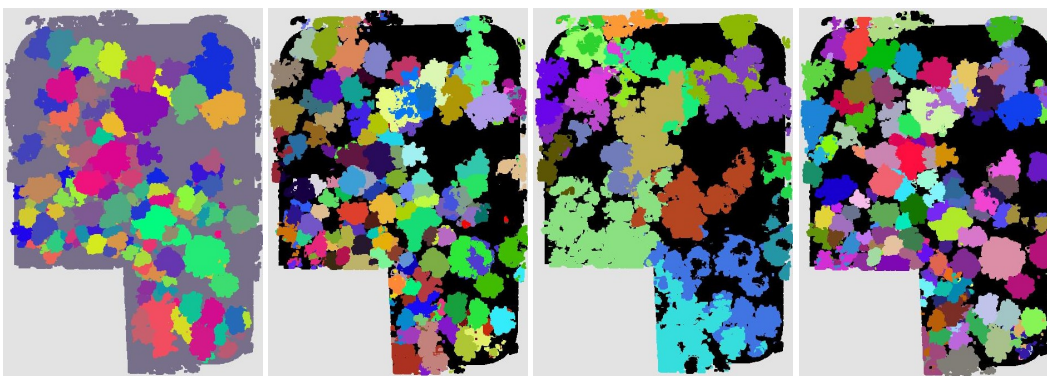

Figure 2: Qualitative test results on a low-resolution UAV-scanned point cloud (RMIT). From left to right the images show (1) the ground truth segmentation and model results obtained by fine-tuning on (2) UAV, (3) MLS+TLS and (4) all data. When only MLS and TLS data is used, segmentation results have severe mistakes. When UAV data is included during training, results are substantially improved.

inated MLS point clouds. Qualitative results indicate that training exclusively with high-resolution data, although improving performance in this domain, leads to poor generalization in low-resolution UAV settings (Fig. 2). Including UAV data in addition to high-resolution data during training alleviates this issue. This emphasizes the importance of a broad training data basis to obtain models that are applicable to a wide range of domains. To enrich the available forest point clouds, labeled tree data from previous works was propagated to the whole forest point cloud and is made publicly available at `https://doi.org/10.25625/QUTUWU`. Going forward, a quantifiable characterization of different forest point clouds should be established to enable a more thorough and systematic comparison between domains. Furthermore, the consequences of stronger domain-shifts in terms of forest structure on model performance should be investigated, for example by using dense tropical forests. Such experiments are crucial to determine what exactly is needed in terms of model development and data provision to obtain powerful and general tree segmentation models.

The results of this study can be regarded as preliminary evidence for the potential of deep learning to obtain general tree segmentation methods. While such methods rely on high-quality labeled forest data, many recent works have acknowledged this need by providing publicly available datasets. Due to the rapid development of deep learning methods and the availability of more and more high-quality labeled data, we expect deep-learning-based tree segmentation to become an increasingly powerful tool. In contrast to traditional segmentation methods, such methods are able to learn segmentation rules for forest point clouds with diverse characteristics in a data-driven way, thus eliminating the need for cumbersome hyperparameter tuning or models designed for specific domains. These features make methods user-friendly and practically applicable given highly diverse forest and point cloud characteristics.

REFERENCES

Andrew Burt, Mathias Disney, and Kim Calders. Extracting individual trees from lidar point clouds using treeseg. *Methods Ecol. Evol.*, 10(3):438–445, 2019. doi: https://doi.org/10.1111/2041-210x.13121.

Kim Calders, Hans Verbeeck, Andrew Burt, Niall Origo, Joanne Nightingale, Yadvinder Malhi, Phil Wilkes, Pasi Raumonen, Robert GH Bunce, and Mathias Disney. Laser scanning reveals potential underestimation of biomass carbon in temperate forest. *Ecological Solutions and Evidence*, 3(4): e12197, 2022.

GreenValley International. Lidar360 point cloud post-processing software, 2022. URL https://greenvalleyintl.com/LiDAR360. Accessed: 26.10.2022.

Jonathan Henrich, Jan van Delden, Dominik Seidel, Thomas Kneib, and Alexander Ecker. Treelearn: A comprehensive deep learning method for segmenting individual trees from forest point clouds. *arXiv preprint arXiv:2309.08471*, 2023.

Li Jiang, Hengshuang Zhao, Shaoshuai Shi, Shu Liu, Chi-Wing Fu, and Jiaya Jia. Pointgroup: Dual-set point grouping for 3d instance segmentation. In *Proc. IEEE/CVF conf. comput. vis. Pattern recognit.*, pp. 4867–4876, 2020. doi: https://doi.org/10.1109/cvpr42600.2020.00492.

Karel Kuželka, Martin Slavík, and Peter Surovỳ. Very high density point clouds from uav laser scanning for automatic tree stem detection and direct diameter measurement. *Remote Sensing*, 12 (8):1236, 2020.

Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts, 2016.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2017.

Stefano Puliti, J Paul McLean, Nicolas Cattaneo, Carolin Fischer, and Rasmus Astrup. Tree height-growth trajectory estimation using uni-temporal uav laser scanning data and deep learning. *Forestry*, 96(1):37–48, 2023a.

Stefano Puliti, Grant Pearse, Peter Surovỳ, Luke Wallace, Markus Hollaus, Maciej Wielgosz, and Rasmus Astrup. For-instance: a uav laser scanning benchmark dataset for semantic and instance segmentation of individual trees. *arXiv prepr. arXiv:2309.01279*, 2023b. doi: https://doi.org/10.1109/ccdc52312.2021.9602282.

Charles R Qi, Or Litany, Kaiming He, and Leonidas J Guibas. Deep hough voting for 3d object detection in point clouds. In *proc. IEEE/CVF Int. Conf. Comput. Vis.*, pp. 9277–9286, 2019. doi: https://doi.org/10.1109/iccv.2019.00937.

Andreas Tockner, Christoph Gollob, Ralf Kraßnitzer, Tim Ritter, and Arne Nothdurft. Automatic tree crown segmentation using dense forest point clouds from personal laser scanning (pls). *Int. J. Appl. Earth Obs. Geoinformation*, 114:103025, 2022. doi: https://doi.org/10.1016/j.jag.2022.103025.

Jan Trochta, Martin Krček, Tomáš Vrška, and Kamil Král. 3D Forest: An application for descriptions of three-dimensional forest structures using terrestrial LiDAR. *PloS one*, 12(5):e0176871, 2017. doi: https://doi.org/10.1371/journal.pone.0176871.

Thang Vu, Kookhoi Kim, Tung M Luu, Thanh Nguyen, and Chang D Yoo. Softgroup for 3d instance segmentation on point clouds. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 2708–2717, 2022. doi: https://doi.org/10.1109/cvpr52688.2022.00273.

Martin Wieser, Gottfried Mandlburger, Markus Hollaus, Johannes Otepka, Philipp Glira, and Norbert Pfeifer. A case study of uas borne laser scanning for measurement of tree stem diameter. *Remote Sensing*, 9(11):1154, 2017.

B Xiang, T Peters, T Kontogianni, F Vetterli, S Puliti, R Astrup, and K Schindler. Towards accurate instance segmentation in large-scale lidar point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10:605–612, 2023. doi: https://doi.org/10.5194/isprs-annals-X-1-W1-2023-605-2023.