THREE-DIMENSIONAL DEEP LEARNING FOR LEAF-WOOD SEGMENTATION OF TROPICAL TREE POINT CLOUDS

W.A.J. Van den Broeck [1] *, L. Terryn[1], W. Cherlet [1], Z.T. Cooper [1], K. Calders [1]

[1] CAVElab, Ghent University, Gent, Belgium - (wouter.vandenbroeck, louise.terryn, wout.cherlet, zane.cooper, kim.calders)@ugent.be

KEY WORDS: Terrestrial laser scanning, virtual forest reconstruction, semantic segmentation, 3D deep learning

ABSTRACT:

Terrestrial laser scanning (TLS) has emerged as a valuable technology for forest monitoring, providing detailed 3D measurements of vegetation structure. However, the semantic understanding of tropical tree point clouds, particularly the separation of woody and non-woody components, remains a challenge. Therefore, this paper addresses the gaps in both (1) data availability and (2) knowledge regarding the potential of 3D deep learning algorithms for leaf-wood segmentation of tropical tree point clouds. First, we contribute a new dataset consisting of 148 tropical tree point clouds with manual leaf-wood annotations. Second, we present initial results using the RandLA-Net 3D deep learning architecture to establish a benchmark on our dataset, achieving a mean intersection over union (mIoU) of 86.8% and overall accuracy of 94.8%. Visual inspection of predictions reveals areas of confusion and indicates applicability across different forest types. Our study demonstrates the potential of 3D deep learning for leaf-wood segmentation in tropical tree point clouds and highlights avenues for future research, including exploring different architectures and investigating the influence of prediction errors on volumetric tree reconstruction.

1. INTRODUCTION

Terrestrial laser scanning (TLS) is being recognised as a key technology in forest monitoring by providing highly detailed in-situ measurements of 3D vegetation structure, and is a particularly valuable tool for (1) nondestructive estimation of aboveground biomass and (2) virtual forest reconstruction to support cal/val activities of remote sensing missions through realistic radiative transfer modelling (Calders et al., 2020). A crucial element in both these applications is to have a semantic understanding of especially the woody component within the TLS point cloud. While deciduous trees are therefore typically scanned in winter to derive leaf-off point clouds, this not an option for evergreen trees such as in tropical forests. Consequentially, tropical tree point clouds require a semantic labelling step to segment the woody points from the point cloud. As manual labelling is extremely laborious and tedious, numerous works have proposed automated methods to tackle this leaf-wood separation (Bai et al., 2023; Krishna Moorthy et al., 2020; Tian and Li, 2022; Vicari et al., 2019; Wang et al., 2020). However, at least two aspects remain insufficiently covered: (1) only a handful of adequate reference datasets for developing and testing leaf-wood segmentation of tropical tree point clouds exist, and (2) only a limited number of studies have explored the potential of applying recent state-of-the-art (SotA) 3D deep learning algorithms for this task (Kaijaluoto et al., 2022; Krisanski et al., 2021; Morel et al., 2020; Windrim and Bryson, 2020). For the latter, prerequisites are that the algorithms should (1) be able to deal with very large point clouds characteristic to TLS forest scans, (2) discriminate between the semantic classes based only on geometric neighbourhood information (i.e. a list of 3D coordinates), and (3) be efficient in terms of execution time and required computing infrastructure. Therefore within this work we aim at contributing to closing both these data and knowledge gaps. Our contribution is twofold:

1. We introduce a new TLS derived tropical tree point cloud dataset, including accompanying manual point-wise leaf-wood annotations.
2. Given our dataset, we present first results using RandLA-Net, a 3D point-wise deep learning network fulfilling all three abovementioned criteria, to set a benchmark on our dataset.

2. MATERIALS & METHODS

2.1 Dataset

We here present our novel dataset, comprising of a total of 148 individual tropical tree point clouds with corresponding manual point-wise semantic labels (i.e., either wood or non-wood).

2.1.1 Study area: The dataset is a combination of three tropical plots in north-eastern Australia (Figure 1): Daintree Rainforest Observatory (DRO), Oliver Creek (OC) and Robson Creek (RC). DRO and RC are supersites from TERN (TERN, 2023a, 2023b), OC and RC are part of the CSIRO rainforest permanent plots (Graham, 2006). The forest type is complex mesophyll and simple notophyll vine forest, with 41 different species occurring within the dataset. A more detailed overview of the plots is given in Table 1.

2.1.2 Data collection All plots were scanned in 2018 with a RIEGL VZ400 at 300kHz. Scans were taken following 10×10 m grids, with both upright and tilt scans at each location. The TLS scans were co-registered by making use of reflectors placed in the plots and using the RiSCAN Pro software. The complete point clouds were subsequently downsampled to 0.02m. The resulting dataset has a totals to $43.6 \cdot 10^6$ points, with the individual trees having on average $3 \cdot 10^5$ points/tree, varying between a maximum of $1.9 \cdot 10^6$, and a minimum of $16.5 \cdot 10^3$ points/tree.

---

* Corresponding author

Table 1. Dataset description.

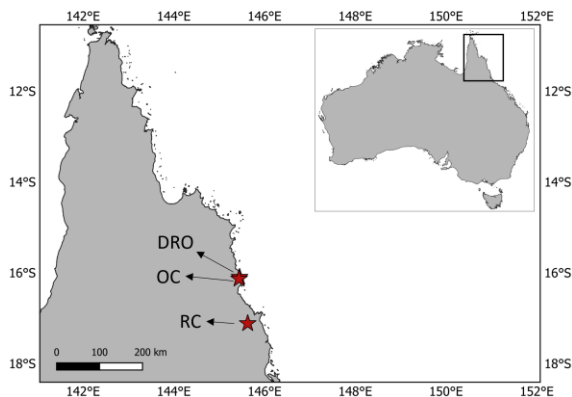|	Daintree Rainforest Observatory	Oliver Creek	Robson Creek
Abbreviation	DRO	OC	RC
Plotsize	1 ha	0.5 ha	1 ha
Trees	59	32	57
Forest type	complex mesophyll vine forest	complex mesophyll vine forest	Simple notophyll vine forest
Lat – long	-16.10; 145.45	-16.14; 145.44	-17.12; 145.63
Elevation	65 m	15 m	700 m
Rainfall	5143 mm	3470 mm	2236 mm

Figure 1. Indication of plot locations in north-east Australia (DRO: Daintree Rainforest Observatory, OC: Oliver Creek, RC: Robson Creek).

2.1.3 Data labelling: Individual trees were manually segmented from the point clouds (Figure 2) and point-wise wood labels were attributed using a semi-automated approach, by combining the output of the algorithms proposed by Krishna Moorthy et al. (2020) and Vicari et al. (2019) and a subsequent rigorous manual correction. Example annotated trees are visualized in Figure 3. Out of all points, 21% was labelled as being woody points.
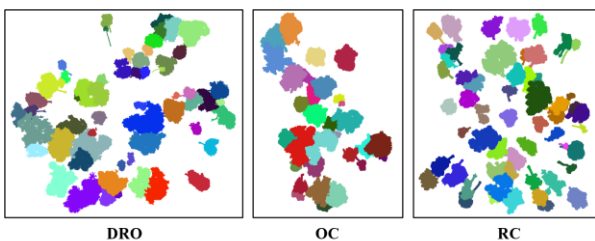

Figure 2. Top view of the three field plots. Each tree is visualised with a unique colour.

Figure 3. Example trees from the DRO plot (top left: *Cardwellia sublimis*; top right: *Dysoxylum papuanum*; bottom left: *Elaeocarpus angustifolius*; bottom right: *Syzygium graveolens*)

2.1.4 Preprocessing: The 148 trees were saved as individual plain text files with (relative) xyz coordinates and corresponding binary labels (0=foliage, 1=woody). To prepare the dataset for training and testing deep learning predictive models, the dataset was further partitioned into a training, validation and test set, using a random 60-20-20 split. Files were saved in separate folders as python numpy files. An overview of the machine-learning-ready dataset is given in Table 2.

Table 2. Pre-processed dataset description.

Partition	Trees	Points	Wood-fraction
Train	89	$25{,}9\ 10^6$	0.22
Validation	29	$10{,}6\ 10^6$	0.17
Test	30	$7{,}2\ 10^6$	0.23

2.2 Leaf-wood segmentation

We here elaborate on using the RandLA-Net 3D deep learning model on our newly presented dataset as an initial exploration of the potential of 3D deep learning algorithms for leaf-wood segmentation of tropical tree point clouds. For the implementation of RandLaNet and the 3D deep learning training and inference pipelines, we use the Open3D-ML library (v.0.17) (Zhou et al., 2018) with pytorch backend (v.1.13) and CUDA support (v.11.6). The hardware used is a 20-core 12th-Gen Intel i7-12800H (2.40 GHz) laptop with a 4Gb NVIDIA RTX A1000 GPU. Code is run within a WSL2 Ubuntu 22.04 kernel with 16 Gb RAM (4 Gb swap).

2.2.1 Deep learning architecture: RandLA-Net, as introduced by Hu et al. (2020), is a lightweight point-based neural architecture proposed for the task of efficient semantic segmentation of large-scale 3D point clouds. The key ideas of their approach are using random sampling for efficient point cloud downsampling across neural layers, combined with a local feature aggregation module to progressively increase the receptive field for each point to preserve geometric details. We used RandLA-Net with 5 layers, 16 nearest neighbours, and taking a fixed 2^{16} points as input. The model takes as input a N × 3 array of 3D coordinates, and outputs a N × 2 array of logits, which can be converted to class probabilities by applying the softmax function. The predicted semantic label is obtained by simply applying the argmax for each point.

2.2.2 Training: The model weights were optimized by minimizing the categorical cross entropy loss between the ground truth labels and predicted labels, as given in Eq. 1.

$$\mathcal{L}_{CCE} = -\frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K} w_k\, y_{k,n} \log \hat{y}_{k,n} \tag{1}$$

where
$y_{k,n}$ = ground truth value of n-th point y ∈ {0, 1}
$y_{k,n}$ = predicted value of n-th point y ∈ [0, 1]
w_k = loss weight for class k
N = total number of points (batch size × sample size)
K = number of classes (2 in this case)

To deal with the class imbalance, the class weights were calculated as the inverse of the class probabilities, estimated from the class frequencies f_k in the training set:

$$w_k = (f_k)^{-1} \tag{2}$$

The model was trained on the 89 trees in the training set for 100 epochs using the Adam optimizer, an exponential learning rate schedule (initial learning rate of 10^{-3} and gamma of 0.9886) and a batch size of 1. For each epoch, to ensure an equal number of model input points, a subset of each point cloud is sampled by randomly selecting a center-point and its 2^{16} - 1 nearest neighbours. In case of a point cloud smaller than 2^{16} points, random points are duplicated. Augmentations were applied including recentering along the three spatial axes, and random

vertical rotations, scaling and noise addition. Training for 100 epochs took 57 minutes.

To prevent overfitting on the training set, the performance on the validation set was computed after each epoch. Validation trees were only recentred and random subsets were taken similarly as described above. The final model was selected as the one showing the highest validation mean intersection over union (mIoU) over the two classes, where the IoU is defined as:

$$IoU = \frac{TP}{TP + FN + FP} \tag{3}$$

where
TP = true positives
TN = true negatives
FP = false positives
FN = false negatives

2.2.3 Inference: Given the trained model, predictions for new (unsubsampled) tree point clouds were obtained using a spatially regular sampling scheme. First, a vector is generated storing random low probabilities between 0 and 0.001 for each point in the point cloud. Second, the point with the lowest probability is selected as centre point. Third, model predictions are computed for the centre point and its 2^{16}-1 nearest neighbours. Fourth, the probabilities are increased with the normalized inverse distance to the centre point. Step 2 - 4 are repeated until the minimum probability is 0.5 for all points. Predictions for already seen points are simply overwritten in subsequent iterations. Inference took 21 minutes for all 30 test trees ($7{,}2\ 10^6$ points).

2.2.4 Evaluation: the performance of the model was evaluated both visually, and quantitatively by computing performance metrics on the test set. First, inference was run for all 30 trees in the test set. Subsequently, (class-wise) metrics could be computed from the confusion matrix by comparing the predictions to the ground truth. The metrics used were the common precision, recall and accuracy:

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

3. RESULTS

3.1 Model training

The evolution of the training and validation loss and mIoU during model training are plotted in Figure 4. It can be seen that the training converges rapidly to a low loss and high mIoU, reaching a training mIoU of ca. 80% after only a single epoch and validation mIoU values of over 80% already after 5 epochs. The differences in training and validation metrics are small, indicating that the model is able to learn the underlying distribution and has generalization potential. Training is fairly stable and there seems no sign of overfitting. However, interesting may be to examine whether this behaviour continues when training for a higher number of epochs.
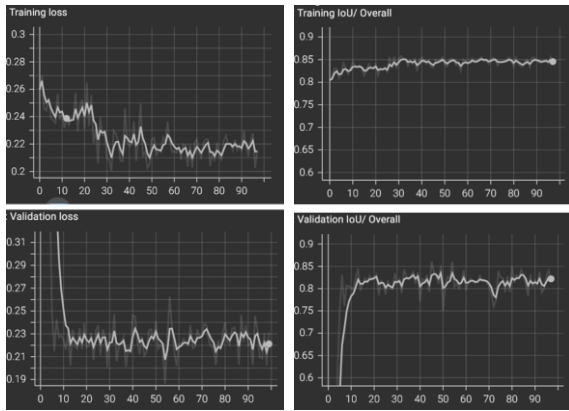
Figure 4. Evolution of the training and validation loss (left) mean intersection over union (IoU) (right) during training for 100 epochs.

3.2 Quantitative evaluation

Table 3 presents the confusion matrix between the ground truth and predictions for all 30 trees in the test set (shown as percentages by dividing by the total number of points in the test set). Of course, as the dataset is unbalanced towards leaf-points, the number of correctly predicted leaf-points is higher (~74%) than the number of correctly predicted woody-points (~21%). Further, it can be inferred that the relative error on the woody points (1.74 / 22.75 = 0.08) is higher than the relative error on the leaf points (3.46 / 77.24 = 0.04. This is also reflected in the class-wise performance metrics, given in Table 4, showing a higher precision, recall and IoU for the leaf-class. Nonetheless, the overall performance is high, reaching a mIoU of 86.8% and accuracy of 94.8%, thus confirming the potential of this paradigm for leaf-wood segmentation of tropical tree point clouds, especially since no hyperparameter search or ablation experiments were conducted.

Table 3. Confusion matrix [%]

		Predicted	
		Wood	Leaf
True	Wood	21.01	1.74
	Leaf	3.46	73.78

Table 4. Performance metrics [%]

	Wood	Leaf
Precision	85.85	97.69
Recall	92.35	95.51
IoU	80.14	93.41
mIoU	86.78	
Accuracy	94.79	

3.3 Qualitative evaluation

Besides quantitative evaluation, we show and examine visual example predictions on the test set. This is especially important as it should be kept in mind that the manual leaf-wood separation of the tree point clouds is not perfect. As such, quantitative performance metrics of 100% are in fact not desired. Figure 4 shows an example of a ground truth segmentation vs. the predicted segmentation, coloured

according to the leaf and wood classes. Overall, the prediction looks reasonable and approximates the ground truth to a high degree. The model works especially well for the stem and main branches. Higher up in the canopy and for the smaller branches some more erroneous predictions appear (see e.g. zoom in Figure 4). As for these points it is also hard for the human interpreter to distinguish wood from leaf, the dataset will likely contain some wrongly classified points, and it is thus reasonable that these regions are harder to learn for the model. Further, to gain insight into the different types of predictions errors, Figure 5 shows example test set predictions coloured according to the TP, TN, FP and FN. Similar observations as mentioned above can be made.
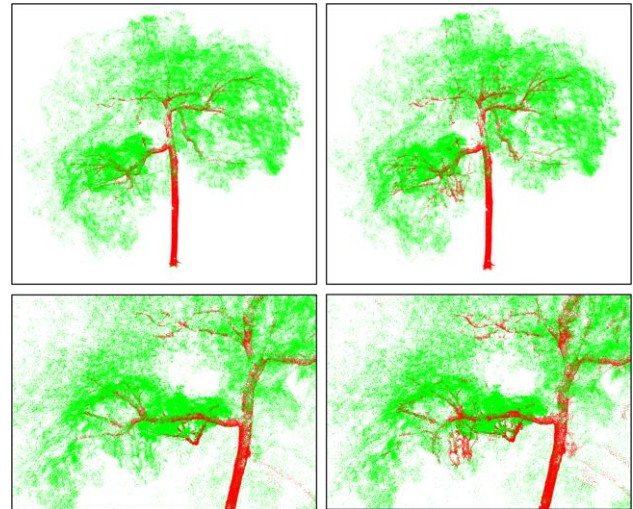

Figure 4. Example of ground truth (left) vs. prediction (right) for a complete tree (top) and zoom (bottom) from the test set (*Syzygium graveolens*, DRO plot)

3.4 Generalization potential

When working with machine learning models, one of the major issues is often its limited predictive capability outside of the training domain. To explore the potential of RandLA-Net trained on our tropical tree dataset for leaf-wood segmentation of trees from another ecosystem, we apply the model on a TLS point cloud of a deciduous leaf-on tree from the well-studied Wytham Woods (Southern UK). The result is visualised in Figure 6. Although we here only performed the exercise for a single tree, it surely shows the potential.
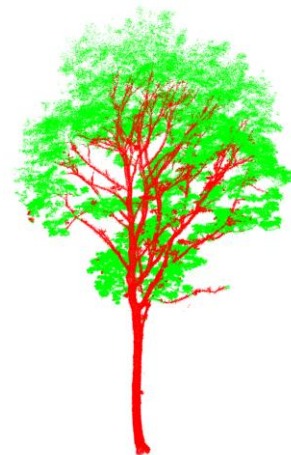

Figure 5. Example prediction on a deciduous leaf-on tree in Wytham Woods using the model trained on tropical trees.
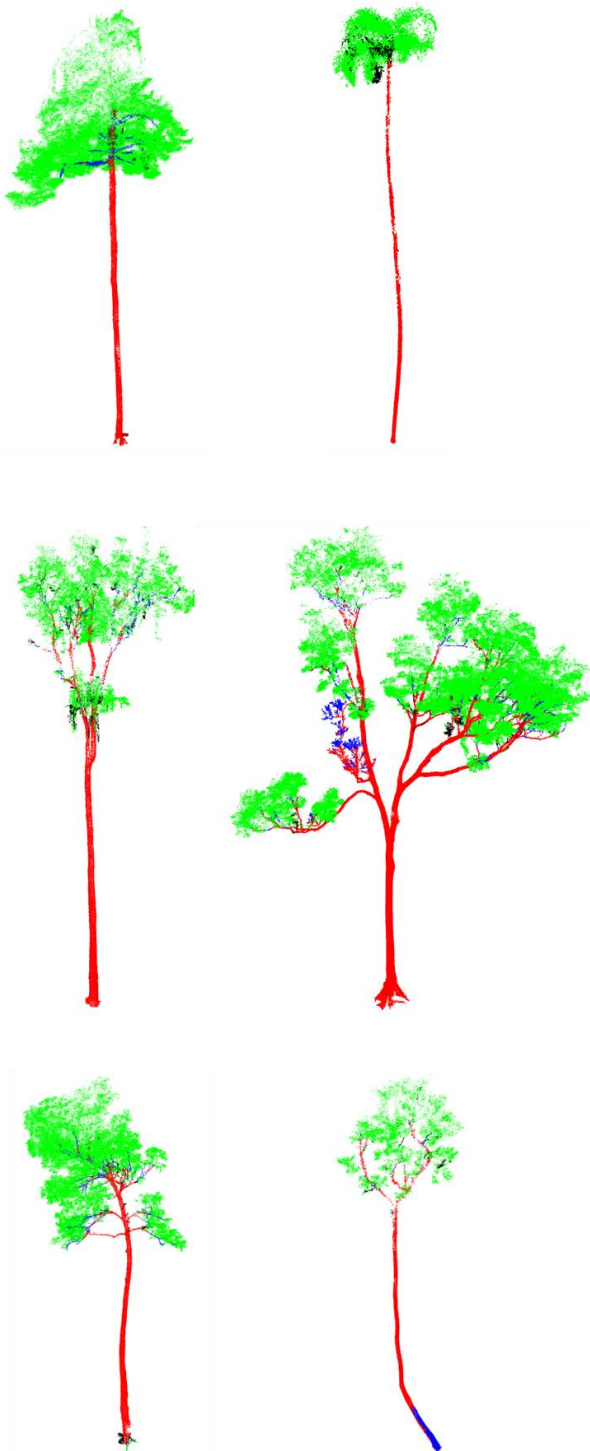
Figure 6. Colour coded test-set predictions (green: leaf correctly predicted as leaf, red: wood correctly predicted as wood, blue: wood wrongly predicted as leaf, black: leaf wrongly predicted as wood)

4. DISCUSSION

Although limited, the preliminary results are promising and give way for some considerations and future research. A first concern with data-driven methods is that they require large amounts of training data, which, especially for point cloud data, is an extremely tedious and time-consuming task. Moreover, unavoidably the annotated data will contain some degree of

error, making it imperative that there are sufficient examples available to find the underlying distribution. However, our observation that there is fast training convergence and a potential for generalisability across tree types given a rather limited training set is encouraging. Second, these learning-based methods solely look for patterns in data and are not bound by any physical constraints. As such, problems with connectivity may occur when using the model outputs for subsequent volumetric tree reconstruction. Therefore, we aim at investigating the influence of prediction errors on the volumetric estimate derived by quantitative structure model (QSM) reconstruction, and its sensitivity to differences in model weights and class loss-weights (influencing the relative importance given to class errors). Here, plotting point clouds coloured according to the prediction probabilities may help in gaining insight into the uncertainties of the model. Furthermore, we intent to experiment with multiple hyperparameters such as using higher batch sizes or higher number of input points (e.g. 2^{18} i.s.o. 2^{16}), and to examine the effect of different augmentations (e.g. horizontal rotations to learn leaning/fallen trees) on the prediction performance. Moreover, other 3D deep learning architectures will be tested and compared to more traditional methods. Last, we plan on testing the 3D deep learning model(s) for leaf-wood segmentation directly on forest point cloud tiles instead of on individually segmented trees.

5. CONCLUSION

In this paper, we present a first exploration of using 3D deep learning for leaf-wood segmentation of tropical tree point clouds. To this end, we introduce a new dataset consisting of 148 individual tropical tree point clouds derived from TLS, with corresponding manual leaf-wood annotations. Preliminary results using the point-based RandLA-Net neural architecture are promising, showing fast and stable training. Quantitative evaluation on a hold-out test set confirms the high performance, with a mIoU of 86.8% and overall accuracy of 94.8%. We further highlight the importance of visual inspection and present prediction examples showing different types of confusion, and the model's ability for generalizability across forest types. Building on these conclusions, we discuss some avenues for planned future work, including comparing traditional and other 3D deep learning models, and investigating the influence of model performance and stability on QSM volume estimates.

REFERENCES

Bai, Y., Durand, J.-B., Forbes, F., Vincent, G., 2023. Semantic segmentation of sparse irregular point clouds for leaf/wood discrimination. *arXiv:2305.16963v1*.

Calders, K., Adams, J., Armston, J., Bartholomeus, H., Bauwens, S., Bentley, L.P., Chave, J., Danson, F.M., Demol, M., Disney, M., Gaulton, R., Krishna Moorthy, S.M., Levick,

S.R., Saarinen, N., Schaaf, C., Stovall, A., Terryn, L., Wilkes, P., Verbeeck, H., 2020. Terrestrial laser scanning in forest ecology: Expanding the horizon. *Remote Sens. Environ.* 251, 112102. https://doi.org/10.1016/j.rse.2020.112102

Graham, A.W. (Ed.), 2006. The CSIRO Rainforest Permanent Plots of North Queensland - Site, Structural, Floristic and Edaphic Descriptions. CSIRO and the Cooperative Research Centre for Tropical Rainforest Ecology and Management.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A., 2020. RandLA-net: Efficient semantic segmentation of large-scale point clouds, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* pp. 11108–11117. https://doi.org/10.1109/cvpr42600.2020.01112

Kaijaluoto, R., Kukko, A., El Issaoui, A., Hyyppä, J., Kaartinen, H., 2022. Semantic segmentation of point cloud data using raw laser scanner measurements and deep neural networks. I*SPRS Open Journal of Photogrammetry and Remote Sensing* 3, 100011. https://doi.org/10.1016/j.ophoto.2021.100011

Krisanski, S., Taskhiri, M.S., Gonzalez Aracil, S., Herries, D., Turner, P., 2021. Sensor Agnostic Semantic Segmentation of Structurally Diverse and Complex Forest Point Clouds Using Deep Learning. *Remote Sensing* 13, 1413. https://doi.org/10.3390/rs13081413

Krishna Moorthy, S.M., Calders, K., Vicari, M.B., Verbeeck, H., 2020. Improved Supervised Learning-Based Approach for Leaf and Wood Classification From LiDAR Point Clouds of Forests. *IEEE Trans. Geosci. Remote Sens.* 58, 3057–3070. https://doi.org/10.1109/TGRS.2019.2947198

Morel, J., Bac, A., Kanai, T., 2020. Segmentation of unbalanced and in-homogeneous point clouds and its application to 3D scanned trees. *Vis. Comput.* 36, 2419–2431. https://doi.org/10.1007/s00371-020-01966-7

TERN, 2023a. Daintree Rainforest SuperSite. URL https://www.tern.org.au/tern-ecosystem-processes/daintree-rainforest-supersite/ (accessed 7.7.23).

TERN, 2023b. Robson Creek Rainforest SuperSite. URL https://www.tern.org.au/tern-ecosystem-processes/robson-creek-rainforest-supersite/ (accessed 7.7.23).

Tian, Z., Li, S., 2022. Graph-Based Leaf–Wood Separation Method for Individual Trees Using Terrestrial Lidar Point Clouds. *IEEE Trans. Geosci. Remote Sens.* 60, 1–11. https://doi.org/10.1109/TGRS.2022.3218603

Vicari, M.B., Disney, M., Wilkes, P., Burt, A., Calders, K., Woodgate, W., 2019. Leaf and wood classification framework for terrestrial LiDAR point clouds. *Methods Ecol. Evol.* 10, 680–694. https://doi.org/10.1111/2041-210x.13144

Wang, D., Momo Takoudjou, S., Casella, E., 2020. LeWoS: A universal leaf-wood classification method to facilitate the 3D modelling of large tropical trees using terrestrial LiDAR. *Methods Ecol. Evol.* 11, 376–389. https://doi.org/10.1111/2041-210x.13342

Windrim, L., Bryson, M., 2020. Detection, Segmentation, and Model Fitting of Individual Tree Stems from Airborne Laser Scanning of Forests Using Deep Learning. *Remote Sensing* 12, 1469. https://doi.org/10.3390/rs12091469

Zhou, Q.-Y., Park, J., Koltun, V., 2018. Open3D: A Modern Library for 3D Data Processing. *arXiv:1801.09847.*