Semantic segmentation of forest point clouds using neural network

Yuchen Bai [1] Jean-Baptiste Durand [2] Grégoire Vincent [3] Florence Forbes [4]

[1] *UGA, France, yuchen.bai@inria.fr*
[2] *CIRAD, France, Jean-Baptiste.Durand@cirad.fr*
[3] *IRD, France, gregoire.vincent@ird.fr*
[4] *INRIA, France, florence.forbes@inria.fr*

Résumé. Depuis des années, la technologie LiDAR (Light Detection And Ranging) s'est imposée comme un outil indispensable pour acquérir des données 3D approfondies dans le domaine de la gestion forestière. Aussi connue sous le nom de scanner laser, cette technologie permet l'acquisition de données sous forme de nuages de points, offrant des détails minutieux sur la structure de la canopée. En particulier, le LiDAR offre la possibilité d'établir des modèles foliaires de forêts avec une précision sans précédent, avec pour motivation le rôle crucial joué par la surface foliaire dans les échanges gazeux entre la végétation et l'atmosphère. La surveillance de la surface foliaire en forêt contribue de manière significative à la compréhension des flux saisonniers dans les forêts tropicales, améliorant ainsi la précision des modèles climatiques pour prédire les impacts du réchauffement.

Divers véhicules sont employés pour l'acquisition de données, avec la numérisation laser terrestre (TLS), qui fournit des données 3D détaillées mais nécessite d'intensives interventions sur le site. La numérisation laser aéroportée (ALS) couvre des zones plus vastes mais avec une densité de points plus faible, source de défis pour l'observation de la végétation de sous-étage, en raison des occlusions causées par la canopée. Les avantages relatifs de la numérisation laser par drone (ULS) se manifestent dans ce contexte, en offrant la possibilité d'une collecte de données dense mais sans nécessité d'intervention sur site. L'enjeu clé de cette étude est l'obtention d'une segmentation sémantique précise pour distinguer les feuilles du bois, exigée en monitoring des forêts pour surveiller les variations de densité foliaire. Ses applications comprennent la prédiction de la séquestration du carbone, la surveillance des maladies et la planification de la récolte. Les méthodes existantes pour les données TLS rencontrent des difficultés lorsqu'elles sont appliquées à l'ULS en raison d'un déséquilibre entre les deux classes et de la dépendance à des informations peu fiables. Afin de résoudre ce problème, nous proposons une nouvelle approche nommée SOUL (Semantic segmentation On ULs) [1], utilisant uniquement les coordonnées des points en entrée du réseau neuronal pour garantir son adaptabilité à différentes localisations et avec divers capteurs.

L'apport de ce travail est triple. Tout d'abord, SOUL est la première approche conçue pour la segmentation sémantique sur des nuages de points ULS dans les forêts tropicales, démontrant une meilleure classification des points de bois. Ensuite, la méthode de prétraitement des données GVD (Geodesic Voxelization Decomposition) relève le défi de l'entraînement des réseaux neuronaux à partir de nuages de points épars dans des environnements forestiers tropicaux. Enfin, nous proposons une fonction de perte ré-équilibrée comme solution polyvalente pour résoudre le problème de déséquilibre des classes dans diverses architectures d'apprentissage profond. Notre étude offre des perspectives prometteuses pour l'application de drones en monitoring des forêts, en repoussant les limitations des méthodologies existantes et en contribuant à une amélioration de la segmentation sémantique dans des environnements forestiers tropicaux.

Mots-clés. Apprentissage profond, segmentation sémantique, déséquilibre de classe, LiDAR, nuage de points

Abstract. In recent decades, LiDAR technology has become an indispensable tool for collecting extensive 3D data in the field of forest inventory. Often known as laser scanning, this technology facilitates the acquisition of point cloud data, providing detailed insights into canopy structure. In particular, LiDAR provides the opportunity to map forest leaf area with unprecedented accuracy, while leaf area has remained an important source of uncertainty affecting models of gas exchanges between the vegetation and the atmosphere. The vigilant monitoring of leaf area contributes significantly to comprehending the seasonal fluxes in tropical forests, thereby refining the precision of climate models in predicting the repercussions of global warming.

Various vehicles are used for data collection, with terrestrial laser scanning (TLS) providing detailed but labor-intensive 3D data. Airborne laser scanning (ALS) covers larger areas but with lower point density, posing challenges for observing understory vegetation due to canopy occlusions. The rise of UAV laser scanning (ULS) addresses these challenges, offering dense data collection without on-site intervention. In fact, forest monitoring requires accurate semantic segmentation to distinguish leaves from wood, which is crucial for monitoring foliage density variations with applications in carbon sequestration, disease monitoring, and harvest planning. Existing methods for TLS data face challenges when applied to ULS due to class imbalance and reliance on unreliable intensity information. To address this, we propose an end-to-end approach named SOUL (Semantic segmentation On ULs) [1] utilizing only point coordinates as input of the neural network for versatility across locations and sensors.

The contributions of this work are three-fold. First, SOUL is the first approach designed for semantic segmentation on ULS point clouds in tropical forests, showcasing superior wood point classification. Second, the GVD (Geodesic Voxelization Decomposition) preprocessing method addresses the challenge of training neural networks from sparse point clouds in tropical forest environments. Third, the proposed rebalanced loss function provides a versatile solution for addressing class imbalance in various deep learning architectures. Our research offers promising insights into the application of ULS for forest monitoring, bridging gaps in existing methodologies and laying the foundation for improved semantic segmentation in challenging tropical forest environments.

Keywords. Deep Learning, Semantic Segmentation, Class Imbalance, LiDAR, Point Cloud

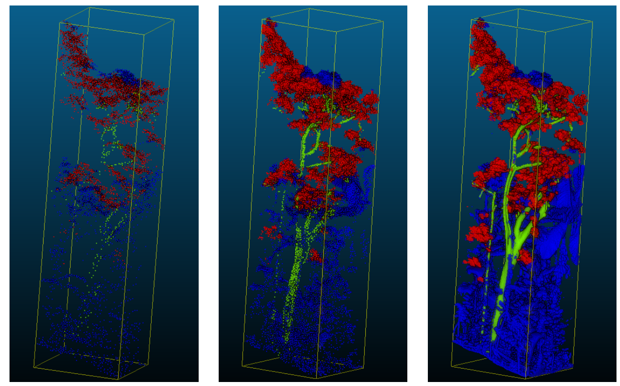
1 Introduction

In the past decades, LiDAR technology has been frequently used to acquire massive 3D data in the field of forest inventory (Vincent et al. [2]; Ullrich & Pfennigbauer [3]). The acquisition of point cloud data by employing LiDAR technology is referred to as laser scanning. The collected point cloud data provides rich details on canopy structure, allowing us to calculate a key variable, leaf area, which controls water efflux and carbon influx. Monitoring leaf area should help in better understanding processes underlying flux seasonality in tropical forests, and is expected to enhance the precision of climate models for predicting the effects of global warming. There are various types of vehicles for data collection, with ground-based equipment and aircraft being the most commonly employed. The former operates a bottom-up scanning called terrestrial laser scanning (TLS), providing highly detailed and accurate 3D data. Scans are often acquired in a grid pattern every 10 m and co-registered into a single point cloud. However, TLS requires human intervention within the forest, which is laborious and limits its extensive implementation. Conversely, airborne laser scanning (ALS) is much faster and can cover much larger areas. Nonetheless, the achieved point density is typically two orders of magnitude smaller due to the combined effect of high flight altitude and fast movement of the sensor. Additionally, occlusions caused by the upper tree canopy make it more difficult to observe the understory vegetation.

In recent years, the development of drone technology and the decreasing cost have led to UAV laser scanning (ULS) becoming one favored option (Brede et al. [4]). It does not require in-situ intervention and each flight can be programmed to cover a few hectares. The acquired data is much denser than ALS (see Figure 1(a) and Figure 1(b)), which provides us with more comprehensive spatial information. Increasing the flight line overlap results in multiple angular sampling, higher point density and mitigates occlusions. Although the data density is still relatively low, compared with TLS, ULS can provide previously unseen overstory details due to the top-down view and overlap flight line. Furthermore, ULS is considered to be more suitable for conducting long-term monitoring of forests than TLS, as it allows predefined flight plans with minimal operator involvement.

Consequently, leaf-wood semantic segmentation in ULS data is required to accurately monitor foliage density variation over space and time. Changes in forest foliage density are indicative of forest functioning and their tracking may have multiple applications for carbon sequestration prediction, forest disease monitoring and harvest planning. Fulfilling these requirements necessitates the development of a robust algorithm that is capable to classify leaf and wood in forest environments. While numerous methods have demonstrated effective results on TLS data, these methods cannot be applied directly to ULS, due in particular to the class imbalance issue: leaf points account for about 95% of the data. Another problem is that many methods rely on the extra information provided by LiDAR devices, such as intensity. In the context of forest monitoring, intensity is not reliable due to frequent pulse fragmentation and variability in natural surface reflectivity (see Vincent et al. [5]). Furthermore, the reflectivity of the vegetation is itself affected by diurnal or seasonal changes in physical conditions, such as water content or leaf orientation (Brede et al. [4]). Therefore, methods relying on intensity information (Wu et al. [6]) may exhibit substantial variations in performance across different locations and even within the same location for different acquisition batches. To address this issue, certain methods (LeWos proposed by Wang et al. [7]; Morel et al. [8]) have good results while exclusively utilizing the spatial coordinates of LiDAR data.

Inspired by the existing methods, we propose a novel end-to-end approach **SOUL** (Semantic segmentation On ULs) based on PointNet++ proposed by Qi et al. [9] to perform semantic segmentation on ULS data. SOUL uses only point coordinates as input, aiming to be applicable to point clouds collected in forests from various locations worldwide and with sensors operating at different wavelengths. The foremost concern to be tackled is the acquisition of labeled ULS data. Since no such data set existed up to now, we gathered a ULS data set comprising 282 trees labeled. This was achieved through semi-automatic segmentation of a coincident TLS point cloud and wood/leaf label transfer to ULS point cloud. Secondly, the complex nature of tropical forests necessitates the adoption of a data pre-partitioning scheme. While certain methods (Krisanski et al. [10]; Wu et al. [11]) employ coarse voxels with overlap, such an approach leads to a fragmented representation and incomplete preservation of the underlying geometric information. The heterogeneous distribution of points within each voxel, including points from different trees and clusters at voxel boundaries, poses difficulties for data standardization. We introduce a novel data preprocessing methodology named geodesic voxelization decomposition (GVD), which leverages geodesic distance as a metric for partitioning the forest data into components and uses the topological features, like intrinsic-extrinsic ratio (IER) (He et al. [12]; Liu et al. [13]), to preserve the underlying geometric features at component level (see Section **??**). The last issue concerns the class imbalance problem during the training stage. To address this issue, we developed a novel loss function named the rebalanced loss, which yielded improved performance compared with the focal loss (Lin et al. [14]) for our specific task. This enhancement resulted in a 23% increase in the ability to recognize wood points, see Table 3.



(a) ALS (b) ULS (c) TLS

Figure 1: Point clouds produced by three scanning modes on the same area (20m × 20m × 42m), illustrate how much the visibility of the understory differs. The colors in the figure correspond to different labels assigned to the points, where red and green indicate leaves and wood, respectively. Blue points are unprocessed, so labeled as unknown.

The contribution of our work is three-fold. First, SOUL is the first approach developed to tackle the challenge of semantic segmentation on tropical forest ULS point clouds. SOUL demonstrates better wood point classification in complex tropical forest environments while exclusively utilizing point coordinates as input. Experiments show that SOUL exhibits promising generalization capabilities, achieving good performance even on data sets from other LiDAR devices, with a particular emphasis on overstory. Secondly, we propose a novel data preprocessing method, GVD, used to pre-partition data and address the difficult challenge of training neural networks from sparse point clouds in tropical forest environments.

Third, we mitigate the issue of imbalanced classes by proposing a new loss function, referred to as rebalanced loss function, which is easy to use and can work as a plug-and-play for various deep learning architectures. The data set (Bai et al. [15]) used in the article is already available in open access at https://zenodo.org/record/8398853 and our code is available at https://github.com/Na1an/phd_mission.

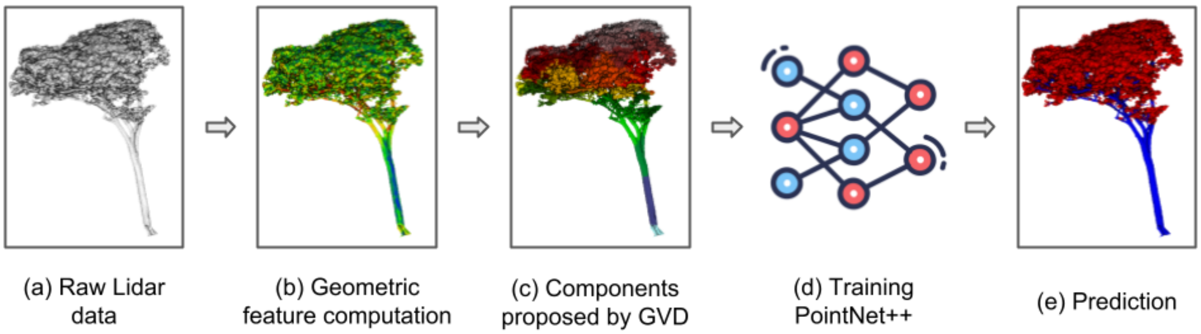


Figure 2: Overview of SOUL. (a) We use only the coordinates of raw LiDAR data as input. (b) Four geometric features linearity, sphericity, verticality, and PCA1 are calculated at three scales using eigenvalues, then standardized. (c) GVD proposes partitioned components and performs data normalization within these components. (d) Training deep neural network. (e) Finally, output are point-wise predictions.

2 Methodology

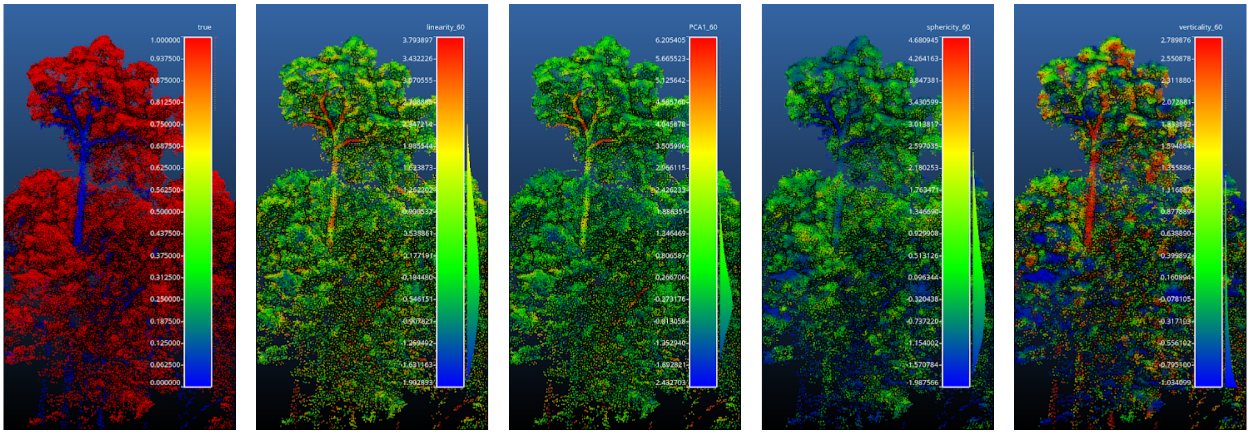
SOUL is based on PointNet++ Multi-Scale Grouping (MSG) [9] with some adaptations. The selection of PointNet++ is not only because of its demonstrated performance in similar tasks (Krisanski et al. [10]; Morel et al. [8]; Windrim & Bryson[11]), but also because of the lower GPU requirements (Choe et al. [16]) compared with transformer-based models developed in recent years, like the method proposed by Zhao et al. [17]. The main idea of SOUL lies in leveraging a geometric approach to extract preliminary features from the raw point cloud, these features are then combined with normalized coordinates into a deep neural network to obtain more abstract features in some high dimensional space [18]. We will introduce our method in detail as follows.

2.1 Data cleaning

Filter out returns below -20 dB, and eliminate noise and ground points.

2.2 Geometric feature computation

At this stage, we introduce four point-wise features: linearity, sphericity, verticality, and PCA1, which are computed at multiple scales of 0.3 m, 0.6 m, and 0.9 m in this task.



(a) True (b) Linearity (c) PCA1 (d) Sphericity (e) Verticality

Figure 3: Four point-wise features introduced at this stage.

2.3 Data pre-partitioning

The geodesic voxelization decomposition (GVD) algorithm is applied to partition the ULS data while preserving the topology of the point cloud. This approach enables the extraction of a set of representative training samples from raw forest data, while preserving the local geometry information in its entirety.

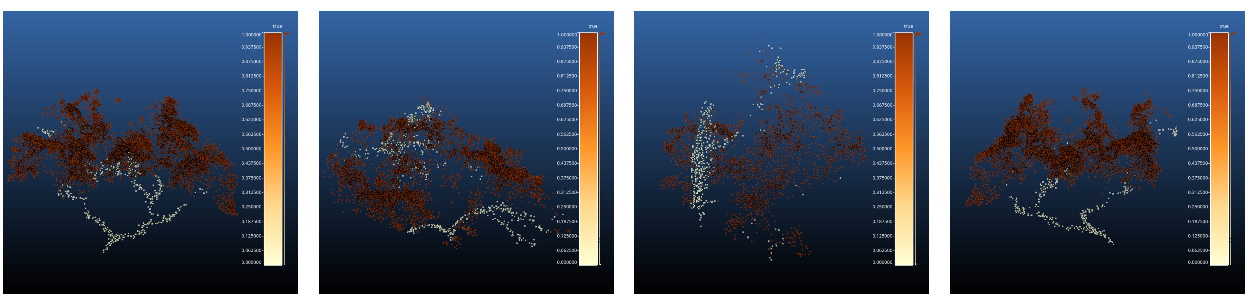


Figure 4: Component examples generated by the GVD algorithm.

2.4 Training neural network

Within the labeled ULS training data set, only 4.4% are wood points. The model is overwhelmed by the predominant features of leaves. Therefore, we propose a rebalanced loss L_R that changes the ratio of data participating to 1:1 at the loss calculation stage by randomly selecting a number of leaf points equal to the number of wood points.

Rebalanced loss, denoted as L_R, is used to address class imbalance issue.

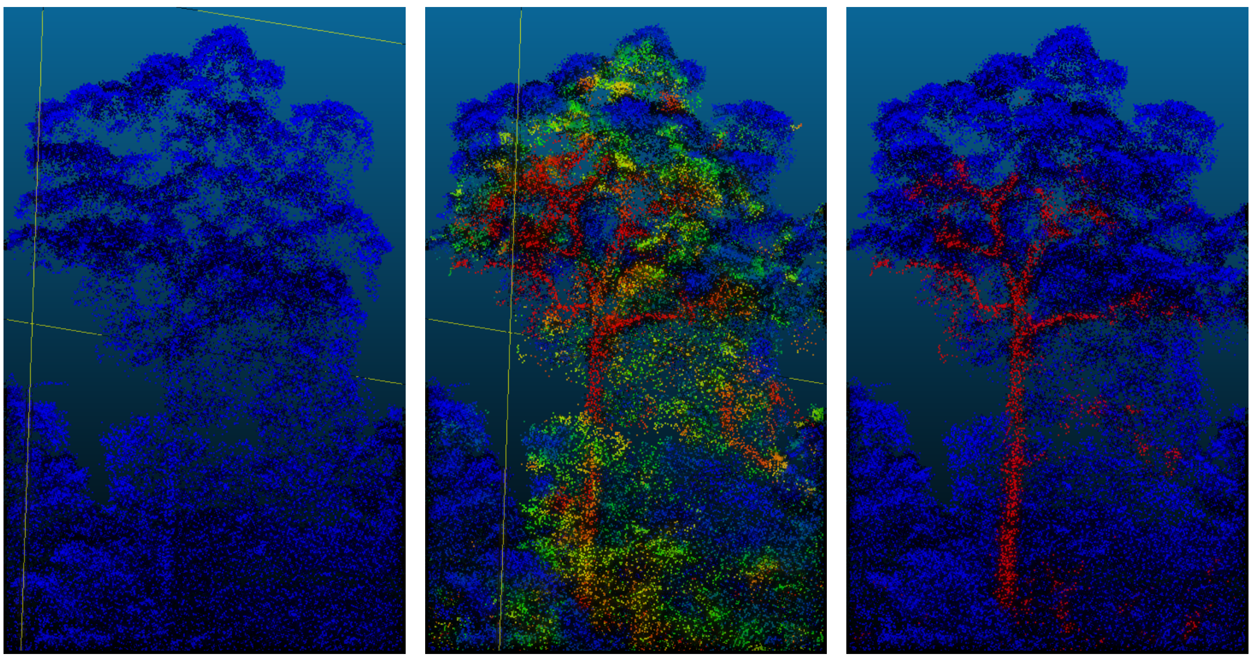
$$L_R(Y_{B_k}) = -\sum y_k \log(\hat{p}_k) + (1 - y_k)\log(1 - \hat{p}_k)\,,\ y_k \in (B'_{k,0} \cup B_{k,1}). \tag{1}$$

where Y_{B_k} specifies the ground truth label of the batch B_k, $\hat{p} \in [0, 1]$ is the model's estimated probability for the label $y = 1$ and $B^{'}_{k,0}$ is defined as

$$B^{'}_{k,0} = \begin{cases} \text{downsampling}(B_{k,0}, |B_{k,1}|)\,, & \text{if } |B_{k,0}| \geq |B_{k,1}| \\ B_{k,0}\,, & \text{otherwise.} \end{cases} \tag{2}$$

3 Results

Comparatively to the prevailing methods employed for forest point clouds, our SOUL approach improves semantic segmentation on ULS forest data by large margins. The issue of class imbalance has been addressed.



(a) FSCT (b) SOUL (c) Ground Truth

Figure 5: Qualitative results on various LiDAR data from different sites.

4 Conclusions

We present SOUL, a novel approach for semantic segmentation in complex forest environments. It outperforms existing methods in the semantic segmentation of ULS tropical forest point clouds and demonstrates high performance metrics on labeled ULS data and generalization capability across various forest data sets. The proposed GVD method is introduced as a spatial split schema to provide refined training samples through pre-partition. One key aspect of SOUL is the use of the rebalanced loss function, which prevents drastic changes in gradients and improves segmentation accuracy. While SOUL shows good performance for different forest types, it may struggle with significantly different trees without retraining. Future work can focus on improving the performance of SOUL on denser forest point clouds to broaden its applications.

References

[1] Y. BAI, J.-B. Durand, G. L. Vincent, and F. Forbes, "Semantic segmentation of sparse irregular point clouds for leaf/wood discrimination," in *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[2] G. Vincent, C. Antin, M. Laurans, J. Heurtebize, S. Durrieu, C. Lavalley, and J. Dauzat, "Mapping plant area index of tropical evergreen forest by airborne laser scanning. A cross-validation study using LAI2200 optical sensor," *Remote Sensing of Environment*, vol. 198, pp. 254–266, 2017.

[3] A. Ullrich and M. Pfennigbauer, "Categorisation of full waveform data provided by laser scanning devices," in *Electro-Optical Remote Sensing, Photonic Technologies, and Applications V* (G. J. Bishop, J. D. Gonglewski, and G. W. Kamerman et al., eds.), vol. 8186, p. 818609, International Society for Optics and Photonics, SPIE, 2011.

[4] B. Brede, H. M. Bartholomeus, N. Barbier, F. Pimont, G. Vincent, and M. Herold, "Peering through the thicket: Effects of UAV LiDAR scanner settings and flight planning on canopy volume discovery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 114, p. 103056, 2022.

[5] G. Vincent, P. Verley, B. Brede, G. Delaitre, E. Maurent, J. Ball, I. Clocher, and N. Barbier, "Multi-sensor airborne lidar requires intercalibration for consistent estimation of light attenuation and plant area density," *Remote Sensing of Environment*, vol. 286, p. 113442, 2023.

Methods	Accuracy	Recall	Precision	Specificity	G-mean	BA[1]
FSCT	0.974	0.977	0.997	0.13	0.356	0.554
FSCT + retrain	0.977	1.0	0.977	0.01	0.1	0.505
LeWos	0.947	0.97	0.975	0.069	0.259	0.520
LeWos (SoD[2])	0.953	0.977	0.975	0.069	0.260	0.523
SOUL (focal loss)	0.942	0.958	0.982	0.395	0.615	0.677
SOUL (rebalanced loss)	0.826	0.884	0.99	**0.631**	**0.744**	**0.757**

[1] BA (Balanced Accuracy) $BA = \frac{1}{2}(Recall + Specificity)$.
[2] SoD (Significant of Difference).

Table 1: Comparison of different methods

[6] B. Wu, G. Zheng, and Y. Chen, "An Improved Convolution Neural Network-Based Model for Classifying Foliage and Woody Components from Terrestrial Laser Scanning Data," *Remote Sensing*, vol. 12, no. 6, 2020.

[7] D. Wang, S. Momo Takoudjou, and E. Casella, "LeWoS: A universal leaf-wood classification method to facilitate the 3D modelling of large tropical trees using terrestrial LiDAR," *Methods in Ecology and Evolution*, vol. 11, no. 3, pp. 376–389, 2020.

[8] J. Morel, A. Bac, and T. Kanai, "Segmentation of Unbalanced and In-Homogeneous Point Clouds and Its Application to 3D Scanned Trees," *Vis. Comput.*, vol. 36, p. 2419–2431, oct 2020.

[9] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, Curran Associates, Inc., 2017.

[10] S. Krisanski, M. S. Taskhiri, S. Gonzalez Aracil, D. Herries, A. Muneri, M. B. Gurung, J. Montgomery, and P. Turner, "Forest Structural Complexity Tool—An Open Source, Fully-Automated Tool for Measuring Forest Point Clouds," *Remote Sensing*, vol. 13, no. 22, 2021.

[11] L. Windrim and M. Bryson, "Detection, Segmentation, and Model Fitting of Individual Tree Stems from Airborne Laser Scanning of Forests Using Deep Learning," *Remote Sensing*, vol. 12, no. 9, 2020.

[12] T. He, H. Huang, L. Yi, Y. Zhou, C. Wu, J. Wang, and S. Soatto, "GeoNet: Deep Geodesic Networks for Point Cloud Analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6881–6890, June 2019.

[13] M. Liu, X. Zhang, and H. Su, "Meshing Point Clouds with Predicted Intrinsic-Extrinsic Ratio Guidance," in *European Conference on Computer Vision (ECCV), Glasgow, UK* (A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, eds.), (Cham), pp. 68–84, Springer International Publishing, August 2020.

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.

[15] Y. Bai, G. Vincent, N. Barbier, and O. Martin-Ducup, "UVA laser scanning labelled las data over tropical moist forest classified as leaf or wood points," Oct 2023.

[16] J. Choe, C. Park, F. Rameau, J. Park, and I. S. Kweon, "PointMixer: MLP-Mixer For Point Cloud Understanding," in *European Conference on Computer Vision (ECCV)*, 2022.

[17] H. Zhao, L. Jiang, J. Jia, P. H. Torr, and V. Koltun, "Point transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16259–16268, 2021.

[18] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.