Canopy to Canopy: Evaluating Model Generalization In 3D Tropical Forest Semantic Segmentation

Brenda Ru Yi SIM*, Sue Han LEE*, Chung Siung CHOO* and Yuen Peng LOH†
* Swinburne University of Technology Sarawak Campus, Malaysia
E-mail: bdsimry@gmail.com, shlee@swinburne.edu.my, cschoo@swinburne.edu.my
† Multimedia University, Malaysia
E-mail: yploh@mmu.edu.my

***Abstract*—Accurate 3D forest semantic segmentation from LiDAR data is fundamental for crucial applications such as biomass estimation and ecological analysis. However, robust performance remains challenging due to inherent forest diversity. The FOR-Instance dataset offers a large-scale benchmark with diverse UAV LiDAR scans from five forest environments. However, there is a notable lack of research assessing how well models trained on FOR-Instance generalize to complex and highly diverse tropical forest environments. As state-of-the-art models in this domain are primarily built upon PointNet-based architectures, this study evaluates the semantic segmentation performance and transferability of three foundational models: PointNet, PointNet++ (SSG), and PointNet++ (MSG). Our results reveal a catastrophic failure across all models when transferred to the unseen tropical Sarawak Forest dataset, highlighting the severe domain shift. Furthermore, models that perform the best in in-distribution (IID) scenarios exhibit significantly degraded performance under out-of-distribution (OOD) conditions, highlighting the need for architectures that better balance generalization and discriminative power. Despite ongoing advances in deep learning, the lack of ecologically diverse training data particularly from tropical forests remains a major bottleneck. This study underscores the pressing need for benchmark datasets that span a broader range of forest types to support the development of more generalizable and robust models for real-world 3D forest analysis.**

I. Introduction

Accurate three-dimensional (3D) semantic segmentation of forests from LiDAR point clouds is essential for ecological monitoring, forest management, and biomass estimation. Substantial research has been conducted on forest classification and semantic segmentation in 3D point clouds, often centered around specific forest types or geographically limited regions. For example, the ForestSemantics dataset [1] focuses on a single boreal forest in Finland, while the Wytham Woods dataset [2] represents temperate woodland in the UK.

However, forest ecosystems vary significantly in composition, structure, and environmental context, resulting in highly diverse point cloud characteristics. In particular, point cloud attributes are influenced not only by the forest type such as boreal, temperate, or tropical but also by factors like canopy density, understory complexity, and terrain variability. Ideally, benchmark datasets would encompass the full spectrum of global forest types, yet in practice, assembling such datasets is hindered by logistical and resource constraints.

The FOR-Instance dataset [3] provides a large-scale benchmark for this task, with diverse UAV LiDAR scans from five distinct environments. While state-of-the-art methods such as ForAINet [4] have demonstrated substantial performance on this dataset, the authors acknowledge the challenge of extending model performance beyond familiar data distributions. In particular, the generalizability of these models to structurally complex and biologically diverse environments, such as tropical forests, remains largely unexplored. This raises an important question: To what extent can models trained on the large-scale FOR-Instance dataset, which encompasses various forest types, effectively generalize to structurally and ecologically different forest ecosystems, such as those found in tropical rainforests? More specifically, how well does the learned forest representation transfer when applied to environments characterized by greater complexity and heterogeneity?

Motivated by this gap, this study investigates the semantic segmentation performance and transferability of three foundational 3D deep learning models: PointNet [5], and PointNet++ (SSG and MSG) [6]. We train these models on the FOR-Instance dataset and evaluate their generalization on an unseen, structurally distinct dataset from the tropical forests of Sarawak. This evaluation aims to quantify the domain gap and offer insights to guide the development of more robust segmentation models for diverse and ecologically complex forest environments.

The contributions of this paper are two-fold: (1) To the best of our knowledge, this is the first study to evaluate the generalizability and transferability of deep learning models trained on a large-scale 3D forest dataset (FOR-Instance) to an unseen, self-collected tropical forest dataset from Sarawak; (2) We provide a comprehensive quantitative and qualitative analysis of three foundational models in 3D point cloud semantic segmentation: PointNet, PointNet++ (SSG), and PointNet++ (MSG), to quantify the performance gap and highlight challenges in domain generalization across ecologically distinct forest environments..

II. Related Work

This section reviews key prior work in deep learning for point cloud semantic segmentation and relevant forest point cloud datasets, focusing on the benchmark used in this study.

979-8-3315-7206-8/25/$31.00

A. Deep Learning Models for Point Clouds

Early deep learning methods processed 3D data indirectly, often by converting it into voxels or multi-view images. A paradigm shift occurred with PointNet [5], which was the first to process raw point sets directly using shared MLPs and symmetric functions. Its primary limitation, a lack of local structural awareness, was addressed by its successor, PointNet++ [6]. This model introduced a hierarchical feature learning approach, enabling it to capture geometric details at multiple scales. As foundational architectures, PointNet and PointNet++ are widely used as benchmarks and backbones, making them highly relevant for studying model transferability. While state-of-the-art methods like ForAINet [4] have achieved strong performance on the FOR-Instance dataset, they also acknowledge the persistent challenge of achieving robust transferability across diverse scenes.

The influence of these foundational models is evident in two primary ways. First, many subsequent works adopt PointNet++ as a direct architectural backbone. For example, VoteNet [7] uses it for feature extraction in 3D object detection, while PointWeb [8] builds its framework directly upon the hierarchical structure of PointNet++ to enhance local feature aggregation. PointNeXt [9] further underscores the power of this backbone by demonstrating that the PointNet++ architecture, when modernized with improved training and scaling strategies, can itself achieve highly competitive performance. Domain-specific models like Forest-PointNet [10] are also direct adaptations tailored for specific challenges like vertical structure segmentation in forests.

Second, in contrast to using it as a monolithic backbone, another line of research integrates PointNet principles as key components within different architectural paradigms, especially Transformers. Point-BERT [11], for instance, employs a mini-PointNet not as its main feature extractor, but as a "tokenizer" to convert local point cloud patches into feature embeddings for its main Transformer architecture. Similarly, models like DensePCR [12] do not use the entire backbone but integrate its core concepts as modules, using a PointNet-style MLP for global features and a PointNet++ inspired method for local features within its own network. In this paper, we focus on 3D forest semantic segmentation by building upon foundational models PointNet and PointNet++, trained on the large-scale FOR-Instance dataset, and evaluate their transferability on an unseen tropical forest dataset from Sarawak.

B. Forest Point Cloud Datasets

Large-scale, well-annotated 3D point cloud datasets for complex natural environments like forests remain less common than for urban areas. While high-quality datasets exist, they often focus on specific forest types or limited geographical areas. For instance, the ForestSemantics dataset [1] provides highly detailed annotations for a single boreal forest in Finland, and the Wytham Woods dataset [2] offers a high-precision scan of a temperate UK forest. Captured with Terrestrial Laser Scanning (TLS), these datasets are invaluable for structural analysis but are limited in their diversity.

In contrast, the FOR-Instance dataset [3] addresses this gap by providing a large-scale benchmark from UAV Laser Scanning (UAV-LS) across five geographically and structurally diverse sites in Norway, the Czech Republic, Austria, New Zealand, and Australia. With 1130 individually annotated trees across 2.79 hectares, it offers unparalleled scale and variety. We selected FOR-Instance for this study precisely because this diversity is crucial for our goal of developing models with strong generalization and transferability, a challenge noted in prior work [3], [4]. The dataset's significant structural variation across different forest types provides an ideal testbed for building a robust and widely applicable segmentation model.

III. Methodology

This study evaluates the performance and transferability of foundational deep learning models for forest semantic segmentation. The following sections detail the experimental setup, including the datasets, model architectures, and evaluation protocols.

A. Dataset and Features

We trained and evaluated models on the FOR-Instance dataset [3], a large-scale benchmark of UAV LiDAR scans from five diverse forest sites. The task was semantic segmentation using the provided labels: Low Vegetation, Terrain, Stem, Live Branches, and Woody Branches. We followed the predefined data splits, further splitting the development set into 20% validation and 80% training sets. We excluded 'Out Points' and 'Unclassified' labels from evaluation, consistent with prior work [4]. For all models, we used a standardized 6-dimensional input of XYZ + RGB coordinates. To handle attribute heterogeneity within the dataset, missing RGB values in any scan were set to zero.

B. Models and Implementation

We evaluated three foundational architectures: PointNet [5], PointNet++ (SSG), and PointNet++ (MSG) [6]. PointNet is a foundational point-based network that consumes unordered point clouds and learns global features using symmetric functions, such as max pooling, for permutation invariance. PointNet++ addresses its limitations by introducing a hierarchical architecture that captures local features at varying scales, enabling a comprehensive understanding of complex geometric structures. Specifically, PointNet++ (SSG) employs a Set Abstraction layer that groups points in local neighborhoods and extracts features in a single-scale manner, building a hierarchical representation. In contrast, PointNet++ (MSG) utilizes Multi-Scale Grouping, concurrently aggregating features from multiple receptive fields within each Set Abstraction layer, capturing diverse and rich contextual information for detailed semantic segmentation.

All models were implemented using PyTorch and trained using the Adam optimizer. PointNet and PointNet++ (SSG/MSG) were trained for 32 epochs with a batch size of 64. The Adam optimizer used an initial learning rate of 0.001, a step size of 10, and an lr decay of 0.7. Cross-Entropy Loss, weighted by inverse class frequency, was used for semantic segmentation to handle class imbalance. Data augmentation included Z-axis rotation to enhance robustness and generalization.

C. Evaluation Metrics

Model performance was quantitatively evaluated using standard semantic segmentation metrics: Overall Accuracy (OA), the proportion of correctly classified points across all classes; and Mean Intersection over Union (mIoU), the average IoU across classes. Per-Class IoU and Accuracy were also computed for insight into performance on individual forest components (Low Vegetation, Terrain, Stem, Live Branches, Woody Branches).

IV. Results

This section reports the quantitative and qualitative segmentation performance of PointNet and PointNet++ models on the FOR-Instance dataset and their transferability to the unseen Sarawak Forest dataset.

A. Quantitative Analysis

Quantitative performance metrics, including Overall Accuracy (OA) and Mean Intersection over Union (mIoU), were computed for each model. Metrics exclude points labeled as 'Out-points' to focus on the main forest components and terrain. The PointNet, PointNet++ (SSG), and PointNet++ (MSG) models were trained and evaluated on the FOR-Instance dataset using XYZ + RGB input features (missing RGB set to 000). Table I summarizes the overall semantic segmentation performance of these three architectures on the FOR-Instance test set.

1) Performance on FOR-Instance Dataset: As shown in Table I, PointNet++ (MSG) achieved the highest overall performance with an mIoU of 0.518 and an OA of 0.888. PointNet and PointNet++ (SSG) exhibited similar overall metrics, with PointNet++ (SSG) showing a slight edge in mIoU (0.484 vs 0.480) while PointNet achieved a slightly higher OA (0.875 vs 0.869). The superior performance of PointNet++ models, particularly the MSG variant, can be attributed to their architectural design, which is better suited to the inherent structural complexity of forest environments. Unlike PointNet, which processes features globally, PointNet++ introduces a hierarchical structure that captures local geometric patterns at multiple scales. This is critical in a forest, where objects of vastly different sizes from large tree stems to fine woody branches and textured terrain coexist. The Multi-Scale Grouping (MSG) module is especially effective, as it allows the model to simultaneously learn features from different neighborhood sizes at each stage. This enables it to distinguish between classes that are spatially intertwined, such as separating low vegetation from the underlying terrain, a task where PointNet's global approach and PointNet++ SSG's fixed-scale analysis are less effective. The MSG architecture's ability to process spatial information across a spectrum of scales directly addresses the natural heterogeneity of forest point clouds.

TABLE I: Overall semantic segmentation performance of PointNet family models on the FOR-Instance test dataset.

Model	mIoU	OA
PointNet	0.480	0.875
PointNet++ (SSG)	0.484	0.869
PointNet++ (MSG)	**0.518**	**0.888**

TABLE II: Per-class semantic segmentation performance on the FOR-Instance test dataset.

Class	**PointNet**		**PointNet++ (SSG)**		**PointNet++ (MSG)**	
	IoU	Acc	IoU	Acc	IoU	Acc
Low-veg.	0.750	0.962	0.740	0.954	**0.772**	**0.969**
Terrain	0.300	0.322	0.274	0.297	**0.386**	**0.405**
Stem	0.429	0.543	0.535	**0.654**	**0.545**	0.629
Live br.	0.919	**0.974**	0.907	0.959	**0.921**	0.970
Woody br.	**0.485**	0.596	0.447	0.643	0.483	**0.655**

To better understand which components contribute to the overall IoU and OA, we examine the per-class performance (IoU and Accuracy) across the five semantic classes in the FOR-Instance dataset, as detailed in Table II. Per-class results (Table II) show that the 'Live branches' class was segmented with consistently high IoU (>0.90) and Accuracy (>0.95) across all three models. The 'Terrain' class was the most challenging, exhibiting the lowest IoU values (0.274 - 0.386). Performance on 'Stem' (0.429 - 0.545 IoU) and 'Woody branches' (0.447 - 0.485 IoU) was moderate. PointNet++ MSG showed the highest IoU for 'Terrain' and 'Low-vegetation', indicating better discrimination of ground and understory layers. PointNet++ SSG achieved a high IoU for 'Stem' (0.535), while PointNet had the highest IoU for 'Woody branches' (0.485), though differences across models for these classes were less pronounced than for 'Terrain'.

2) Model Transferability: Sarawak, located on the island of Borneo in Malaysia, is defined by its archetypal tropical rainforest, the Mixed Dipterocarp Forest (MDF), which is recognized as the richest of the state's major forest types [13]. The structure of this ecosystem is characterized by significant complexity, featuring a deep, dense, and uneven multi-layered canopy that reaches heights of 35 to 55 meters, with emergent trees occasionally exceeding 60 meters [13]. For this study, we collected the Sarawak Forest Dataset, which serves as a unique and valuable benchmark for assessing model transferability to forest environments that are structurally and ecologically distinct from those in the FOR-Instance dataset.

The Sarawak Forest dataset used in this study comprises LiDAR data collected from a 40m x 40m tropical forest plot, providing a realistic testbed for evaluating the robustness and generalization of 3D semantic segmentation models. Importantly, this dataset was not used for training; it was solely used for inference to assess how well models trained on the FOR-Instance dataset generalize to tropical forest environments. This experimental setup offers valuable insights into the extent of the performance gap when transferring models across forest types. Table III presents the overall performance on Sarawak Forest dataset for the PointNet family models.

TABLE III: Comparison of model performance on the FOR-Instance and Sarawak Forest datasets, highlighting the catastrophic performance drop.

Model	Metric	FOR-Inst.[a]	Sarawak[b]	Drop (%)
PointNet	mIoU	0.480	0.069	85.6
	OA	0.875	**0.260**	70.3
PointNet++ (SSG)	mIoU	0.484	**0.079**	83.7
	OA	0.869	0.251	71.1
PointNet++ (MSG)	mIoU	**0.518**	0.060	**88.4**
	OA	**0.888**	0.188	**78.8**

[a] FOR-Inst.: FOR-Instance dataset (in-distribution, IID).
[b] Sarawak: Sarawak Forest dataset (out-of-distribution, OOD).

Evaluation on the Sarawak Forest dataset demonstrated a significant drop in overall performance across all models compared to their results on FOR-Instance, as detailed in Table III, highlighting severe challenges in domain generalization. Overall mIoU dropped dramatically, ranging from 0.060 to 0.079. As Table III reveals, on the Sarawak dataset, PointNet++ SSG achieved the highest overall mIoU (0.079), while PointNet attained the highest Overall Accuracy (OA) at 0.260. Interestingly, although PointNet++ MSG demonstrated strong performance on the FOR-Instance dataset, it exhibited noticeably weaker generalization when evaluated on the Sarawak Forest dataset. This contrast suggests that models which perform well in in-distribution (IID) settings may struggle when applied to out-of-distribution (OOD) environments.

Per-class results on the Sarawak Forest dataset are shown in Table IV. The severe performance drop in the Sarawak Forest dataset is clearly illustrated in Table IV. Most classes showed near-zero IoU and Accuracy for all three models, indicating nearly random classification for low vegetation, terrain, and woody branches. Specifically, Low Vegetation, Terrain, and Woody Branches showed the most significant performance degradation with IoU values below 0.02 across all models. 'Live branches' generally retained the highest per-class IoU on Sarawak among the evaluated classes across all models (ranging from 0.176 to 0.272), although significantly reduced compared to FOR-Instance. PointNet achieved the highest IoU for 'Live branches' (0.272). PointNet++ SSG and MSG showed similar, notably higher IoU for 'Stem' (0.126 and 0.105 respectively) compared to PointNet (0.042). These results underscore the significant challenge of transferring models trained on temperate/boreal forest data (FOR-Instance) to structurally distinct tropical forest environments.

B. Qualitative Analysis

This section presents qualitative segmentation results on specific test scenes to visually illustrate the performance differences between the models and highlight transferability challenges. Visual inspection of the inference results on the Tuwien test scene from the FOR-Instance dataset (Fig. 1) provides qualitative support for the quantitative findings. As shown in Fig. 1a, the PointNet model exhibits a tendency to miss stem points (light green in the ground truth, Fig. 1d) and frequently misclassifies them as live branches (light yellow). In contrast, the PointNet++ models (Fig. 1b and Fig. 1c)

TABLE IV: Per-class semantic segmentation performance on the Sarawak Forest dataset.

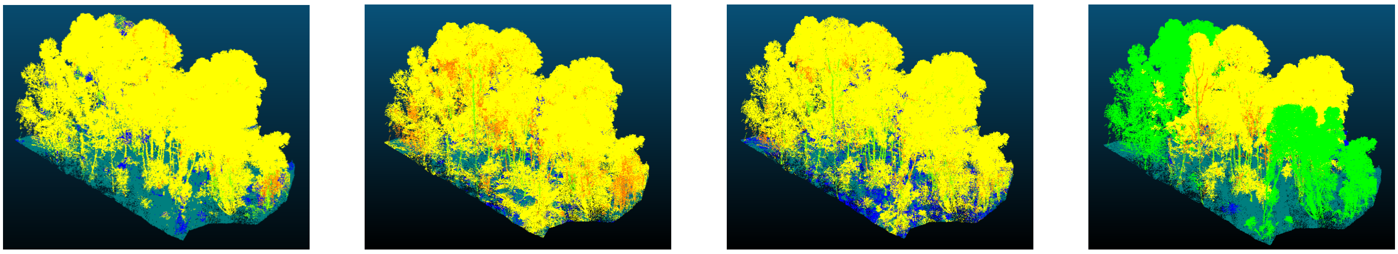
Class	PointNet		PointNet++ (SSG)		PointNet++ (MSG)	
	IoU	Acc	IoU	Acc	IoU	Acc
Low-veg.	**0.008**	**0.008**	0.007	0.007	0.007	0.007
Terrain	**0.020**	**0.027**	0.007	0.007	0.009	0.009
Stem	0.042	0.114	**0.126**	0.682	0.105	**0.703**
Live br.	**0.272**	**0.546**	0.250	0.461	0.176	0.316
Woody br.	0.004	0.064	0.005	0.117	0.005	**0.126**

demonstrate improved performance in identifying stem points. However, the PointNet++ SSG model (Fig. 1b) occasionally misclassifies some stem points and live branches as woody branches (orange). Furthermore, the PointNet segmentation (Fig. 1a) sometimes misclassifies parts of the live branches (light yellow) as lower vegetation (dark blue). It is also important to note that the trees shown in green at the boundary of the ground truth visualization correspond to 'out-points', which were excluded from both training and quantitative evaluation.

The qualitative results on the Sarawak Forest dataset (Fig. 2) provide a visual explanation for the drastic drop in performance observed in the quantitative metrics. The models trained on FOR-Instance largely fail to correctly segment the structurally different tropical forest environment. PointNet (Fig. 2a) demonstrates significant failure, almost entirely missing the stems (light green) and woody branches (orange). Across all models, there is widespread misclassification of lower vegetation (dark green) and terrain (green), which are frequently assigned labels like live branches (yellow) or stem. While the PointNet++ models (Fig. 2b and 2c) show a slightly better ability to identify some stems compared to PointNet, they still struggle significantly with woody branches. These visual examples clearly illustrate the severe domain shift challenge when transferring models from temperate/boreal forests to tropical forests, underscoring the need for domain adaptation techniques or more diverse training data.

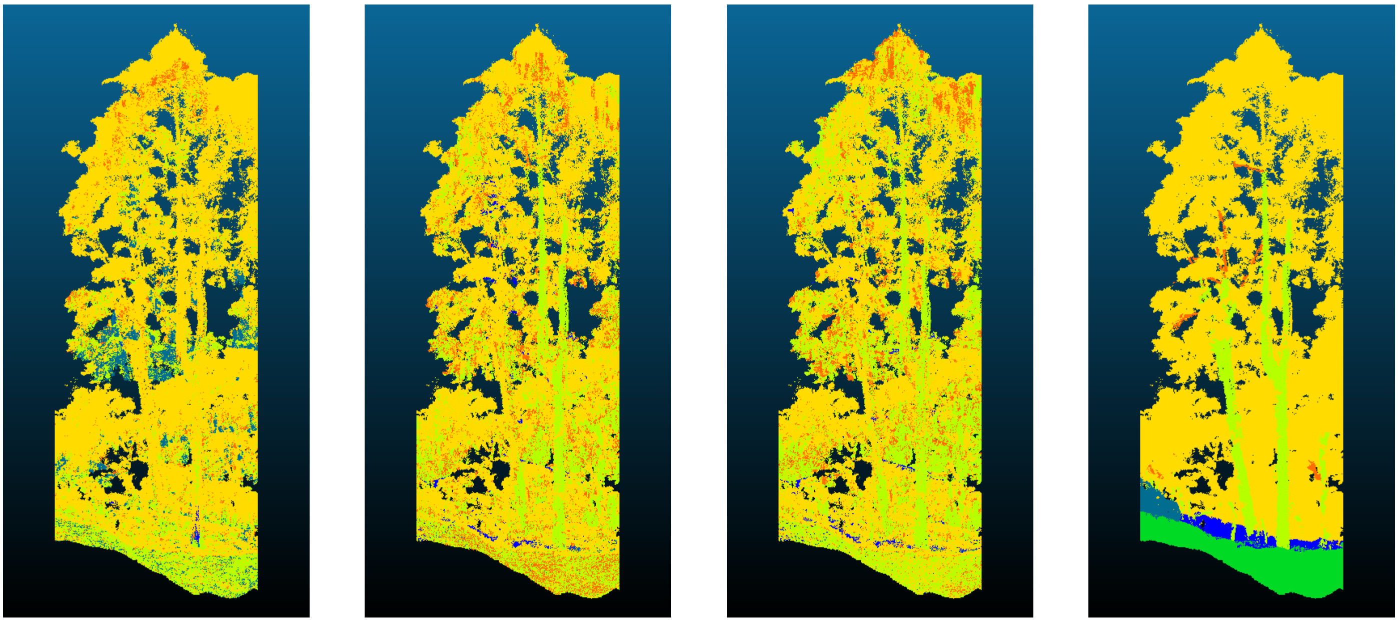
V. Discussion

This study's findings provide a clear quantitative and qualitative assessment of foundational deep learning models on a diverse forest dataset, highlighting both their capabilities and, more critically, their limitations in real-world transferability scenarios. On the in-domain FOR-Instance test set, the superior performance of PointNet++ (MSG) (mIoU 0.518, Table I) underscores the importance of learning local geometric context. Its multi-scale grouping (MSG) architecture is inherently better suited for complex forest scenes containing objects of vastly different scales. The consistently high IoU for 'Live branches' (>0.90, Tables II) suggests this class possesses a dominant and distinct signature that is easily learned. Conversely, 'Terrain' proved to be the most challenging class (IoU 0.274 - 0.386). The Intersection over Union (IoU) metric is highly sensitive to errors in spatial overlap. For 'Terrain' factors like heavy occlusion by low vegetation cause the model to miss large areas of ground, while geometric ambiguity leads it to incorrectly classify low-lying debris as terrain. Both types of



(a) PointNet (b) PointNet++ (SSG) (c) PointNet++ (MSG) (d) Ground Truth

Fig. 1: Qualitative segmentation results on a Tuwien test scene from the FOR-Instance dataset. Subfigures (a), (b), and (c) show model predictions, while (d) shows the corresponding ground truth for comparison.



(a) PointNet (b) PointNet++ (SSG) (c) PointNet++ (MSG) (d) Ground Truth

Fig. 2: Qualitative results on the external Sarawak dataset, demonstrating transferability failure. The models, trained on FOR-Instance, largely fail to segment the structurally different tropical forest.

errors severely reduce the agreement between the predicted and ground truth regions, thus depressing the IoU score, a confusion visually supported by Figure 1.

The most significant finding of this study is the catastrophic failure of all models when transferred to the unseen Sarawak Forest dataset, which represents a severe **domain shift**. The mIoU plummeted from over 0.480 on the FOR-Instance test set to less than 0.080 on the Sarawak Forest data (Table III), a quantitative measure of this **domain gap**. This gap stems from multiple factors. First are the fundamental differences in forest structure, such as the tropical environment's greater complexity and unique elements like lianas. Second, there is a clear discrepancy in data quality; the FOR-Instance dataset consists of relatively complete and dense point clouds, whereas the real-world Sarawak data exhibits significant occlusions, noise, and sparse regions where the LiDAR scanner could not penetrate the dense canopy. This highlights the models' inability to transfer to data collected under different real-world conditions or with different scanner properties, suggesting they are overfitted not just to a specific biome but also to the idealized acquisition characteristics of the training set.

The models, having never seen such structures or data imperfections, failed to generalize. The collapse of the 'Terrain' and 'Low-vegetation' classes is a prime example. Models trained on the predominantly flat ground of FOR-Instance likely learned to associate 'Terrain' with points forming large, locally planar surfaces at consistently low Z-coordinates. When faced with the hilly Sarawak forest environment, this learned prior fails catastrophically. The terrain now exhibits significant local curvature and a wide distribution of Z-values, presenting geometric features in local neighborhoods that the model has never encountered for this class, leading to widespread misclassification. Similarly, the 'Woody branches' class failed because the models likely learned to identify them in the temperate data by their distinct, sparse, and linear geometric signature. In the dense tropical canopy of Sarawak, this signature is lost.

Woody branches are either heavily occluded or their local point neighborhoods are intermixed with dense foliage from 'Live branches' and lianas. This creates a cluttered, ambiguous geometric context that the model, trained on clearer distinctions, overwhelmingly misclassifies as the visually dominant 'Live branches' class. However, this generalization failure was not uniform. While PointNet failed almost completely on 'Stem' (IoU 0.042), PointNet++ models retained a limited ability to identify this class (IoU ˜0.126), suggesting its hierarchical feature learning is slightly more robust to domain shift, as visually corroborated in Figure 2.

Ultimately, this study demonstrates that while architectural improvements like those in PointNet++ offer better performance, they are not a panacea for the domain generalization problem. This work highlights the risks of applying 'off-the-shelf' models to new geographic or ecological domains without rigorous validation. This underscores the need to prioritize domain adaptability in the design of future deep learning models and to develop more comprehensive benchmark datasets that capture the diversity of forest types for effective 3D point cloud analysis in complex natural environments.

VI. Conclusions

This study assessed the generalization capabilities of three foundational deep learning models: PointNet, PointNet++ (SSG), and PointNet++ (MSG) in 3D forest semantic segmentation. While all models achieved reasonable performance on the FOR-Instance dataset, they exhibited a catastrophic performance drop when applied to the structurally complex and ecologically distinct Sarawak tropical forest dataset. Notably, PointNet++ (MSG), which achieved the highest performance on the in-distribution FOR-Instance test set, exhibited a substantial performance decline when evaluated on the out-of-distribution Sarawak Forest dataset, underscoring the need for architectural innovations that better balance generalization and discriminative capability. Moreover, these results emphasize that despite continued advancements in deep learning architectures built upon these foundational models, the lack of ecologically diverse training data remains a critical bottleneck. Hence, expanding benchmark datasets, especially with tropical forests, is crucial to develop models that generalize better for real-world 3D forest analysis.

Acknowledgment

We sincerely thank UAS Mapping Sdn Bhd for providing the Sarawak Forest dataset used in this study. We also acknowledge Darren Chien Hao Goh, Ethan Zhi Jing Ting, Jordan Zhi Haw Hoo, and Hong Wei Sim for their contributions to data annotation, and Abdullahi Hussein Dahir and Munieb Awad Elsheikhidris Abdelrahma for their support in information curation during the early stages of the project.

References

[1] X. Liang, H. Qi, X. Deng, *et al.*, "Forestsemantic: A dataset for semantic learning of forest from close-range sensing," *Geo-Spatial Information Science*, vol. 28, no. 1, pp. 185–211, 2025.

[2] K. Calders, H. Verbeeck, A. Burt, *et al.*, "Laser scanning reveals potential underestimation of biomass carbon in temperate forest," *Ecological Solutions and Evidence*, vol. 3, no. 4, e12197, 2022.

[3] S. Puliti, G. Pearse, P. Surovỳ, *et al.*, "For-instance: A uav laser scanning benchmark dataset for semantic and instance segmentation of individual trees," *arXiv preprint arXiv:2309.01279*, 2023.

[4] B. Xiang, M. Wielgosz, T. Kontogianni, *et al.*, "Automated forest inventory: Analysis of high-density airborne lidar point clouds with 3d deep learning," *Remote Sensing of Environment*, vol. 305, p. 114 078, 2024.

[5] 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*, 2017, pp. 652–660.

[6] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in neural information processing systems*, vol. 30, 2017.

[7] C. R. Qi, O. Litany, K. He, and L. J. Guibas, "Deep hough voting for 3d object detection in point clouds," in *proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9277–9286.

[8] H. Zhao, L. Jiang, C.-W. Fu, and J. Jia, "Pointweb: Enhancing local neighborhood features for point cloud processing," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5565–5573.

[9] G. Qian, Y. Li, H. Peng, *et al.*, "Pointnext: Revisiting pointnet++ with improved training and scaling strategies," *Advances in neural information processing systems*, vol. 35, pp. 23 192–23 204, 2022.

[10] Z. Ma, Y. Dong, J. Zi, F. Xu, and F. Chen, "Forest-pointnet: A deep learning model for vertical structure segmentation in complex forest scenes," *Remote Sensing*, vol. 15, no. 19, p. 4793, 2023.

[11] X. Yu, L. Tang, Y. Rao, T. Huang, J. Zhou, and J. Lu, "Point-bert: Pre-training 3d point cloud transformers with masked point modeling," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 19 313–19 322.

[12] P. Mandikal and V. B. Radhakrishnan, "Dense 3d point cloud reconstruction using a deep pyramid network," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 2019, pp. 1052–1060.

[13] M. Demies, A. K. Sayok, and G. T. Noweg, "Tree diversity, forest structure and species composition in a logged-over mixed dipterocarp forest, bintulu, sarawak, malaysia," 2019.