Deep Learning-Based Semantic Segmentation of LiDAR Point Clouds for Autonomous Driving in Unstructured Environments

Félix Martínez[1] [0009-0004-9184-1194],
David Pascual-Hernández[1] [0000-0002-6549-1694],
Daniel Borja Fernández[1] [0009-0005-3793-3001],
Inmaculada Mora-Jiménez[1] [0000-0003-0735-367X], and
José María Cañas[1] [0000-0003-4179-2211]

RoboticsLabURJC, Universidad Rey Juan Carlos, Spain

Abstract. Autonomous driving in unstructured environments is a challenging task for robotics due to the presence of natural obstacles such as rocks, trees, water, and vegetation, which can hinder navigation. In these scenarios, Unmanned Ground Vehicles rely on robust and reliable perception systems, primarily using cameras and LiDAR sensors for situational awareness. This paper presents an experimental and quantitative analysis of three deep learning-based perception modules, exploring their use for semantic segmentation of raw LiDAR point clouds. We propose a late-fusion approach that incorporates LiDAR intensity, proving its effectiveness in reducing class confusion and improving segmentation performance. All models were trained and tested using the GOOSE dataset, which provides labeled data from unstructured environments. In addition, we analyze the impact of point cloud density and noise on semantic segmentation performance.

Keywords: LiDAR · Semantic Segmentation · Deep Learning · Point Cloud Processing · Intensity LiDAR · Unstructured Environments · Autonomous Driving · PointNet · PointNet++

1 Introduction

Autonomous driving in unstructured environments has emerged as a prominent research area in recent literature, due to its potential to transform fields such as agriculture [6], search and rescue [17], and forest management [11]. While there has been great progress in urban autonomous driving, transferring that progress to off-road navigation is still a challenging task. Unlike structured urban areas, off-road environments lack visual cues to assist navigation such as road markings or traffic signals, and the traversability of different terrains is not clearly defined beforehand. In that way, a robust perception module becomes a critical component for enabling autonomous navigation in unstructured environments.

Perception systems for autonomous navigation typically rely on RGB cameras and LiDAR sensors. While vision-based sensors can potentially provide dense semantic information about the scene, they lack spatial 3D information, which is critical for traversing unstructured environments. On the other hand, LiDAR provides sparser but precise geometric data about the vehicle surroundings. Accurate semantic segmentation of LiDAR, while challenging due to its sparsity, is an invaluable resource for assessing terrain traversability and constructing reliable navigation cost maps.

Data availability is a key factor for the development of perception systems, especially when considering deep learning-based solutions. In this sense, the release of large-scale datasets such as SemanticKITTI [1], NuScenes [3], and Waymo Open Dataset [14] has marked a major breakthrough in the development of LiDAR-based semantic segmentation systems for urban settings. Only recently, datasets like RELLIS-3D [18] and GOOSE [24], while still relatively small in comparison, have begun to bridge the data availability gap between urban and off-road environments. These datasets make it possible to train and evaluate data-driven perception solutions for unstructured scenarios.

In this work, we explore the suitability of deep learning-based approaches for LiDAR semantic segmentation in unstructured off-road environments. Using the GOOSE dataset as a benchmark, we study two foundational lightweight architectures for point cloud processing, namely PointNet [12] and PointNet++ [13]. Furthermore, we study the inclusion of the intensity values provided by the LiDAR sensor. To that end, we propose PointNet++*, a simple yet effective extension of PointNet++ that introduces an MLP module to incorporate the remission intensity signal prior to the segmentation head.

Additionally, in order to better understand the robustness of these models, we conduct a quantitative analysis on the influence of point cloud density and different noise levels in the segmentation accuracy. This is particularly relevant for real-world applications, where hardware limitations and environmental conditions can downgrade the LiDAR signal.

Our contributions can be summarized as follows:

- We benchmark deep learning-based solutions for LiDAR semantic segmentation to assess their suitability in off-road environments.
- We propose PointNet++*, an extension of PointNet++ for leveraging the information provided by the LiDAR intensity signal.
- We study the impact of point cloud density and noise levels in the LiDAR segmentation quality, offering insights into the limitations of the models studied under challenging sensor conditions.

2 Related work

The perception module is a fundamental component of the autonomous navigation stack, with its design depending on both the vehicle type and its intended operational design domain. In the case of Unmanned Ground Vehicles (UGVs), it is often reduced to an assessment of terrain traversability [2], but advanced capabilities such as semantic segmentation might be required for intelligent decision-making [22]. While vision-based segmentation has been extensively studied [10], transferring that information into the 3D space is essential to ensure safe navigation. In that sense, the inclusion of LiDAR sensors in autonomous vehicles has become a popular choice among practitioners. Although the point clouds provided by LiDAR are sparser and harder to interpret than images, deep learning solutions have shown promising success for extracting semantic information [5].

Key to the development of deep learning approaches is the availability of high-quality annotated data. In the case of urban settings, several large-scale datasets are available [1,3,14], but the particularities of autonomous navigation in unstructured environments require datasets specifically tailored for off-road scenarios. In recent years, datasets such as RELLIS-3D [18], GOOSE [24], and WildScenes [19] have emerged, providing an invaluable resource for transferring advances from urban perception to off-road environments.

Numerous deep learning models have been proposed for 3D semantic segmentation. Classic approaches directly process LiDAR data as unstructured point clouds, without any further assumption or spatial discretization. Such is the case of PointNet [12] and PointNet++ [13], which will be further studied in Section 4. These architectures provided a strong foundation for subsequent works that explore improvements such as RandLA-Net's novel sampling strategy [7] and KPConv's kernel point convolution [16]. Other solutions focus on addressing the sparsity problem through different voxelization methods. For instance, MinkUNet [4] voxelizes point clouds and applies 3D convolutional layers following a UNet-like architecture. Meanwhile, Cylinder3D [27] performs voxelization in cylindrical coordinates and SPVCNN [15] follows a hybrid approach. Alternative solutions transform point clouds into graph-like structures [23] or range images [8] before processing. More recently, Transformer-based architectures like PointTransformer [26] and SphereFormer [9] leverage attention mechanisms to further improve context modeling.

Beyond geometry, recent literature demonstrates that the intensity channel of LiDAR sensors provides a strong cue for disambiguating similarly structured objects or surfaces [21], especially when transformation to reflectivity is possible [20]. Much in the spirit of our work, Yan et al. [25] benchmark several learning-based approaches for semantic segmentation using variations of the SemanticKITTI dataset specifically tailored to explore the impact of different noise levels, weather conditions, and LiDAR sensors. However, to the best of our knowledge, our work is the first to explore these topics in the challenging context of off-road perception.

3 Datasets from unstructured environments

The performance and robustness of deep learning models for semantic segmentation are fundamentally dependent on the quality and diversity of the data they are trained on. While autonomous driving in structured urban environments has been the primary focus of research, leading to numerous large-scale datasets such as nuScenes [3], Waymo Open Dataset [14], or SemanticKITTI [1], the great challenge lies in applying these technologies to unstructured environments. These are predominantly natural or rural settings, such as forests, agricultural fields, or trails, which lack predictable infrastructure and present unique perception challenges that have received comparatively less attention.

Datasets in structured environments benefit from the regularity of the scene: well-defined roads, standard signage, buildings with predictable geometries, and a clear distinction between traversable elements and obstacles. In contrast, unstructured environments are characterized by their high variability and geometric complexity. Dense vegetation, irregular topography, the absence of lanes or defined boundaries, severe sensor occlusion, and changing lighting conditions greatly hinder both data acquisition and labeling [18].

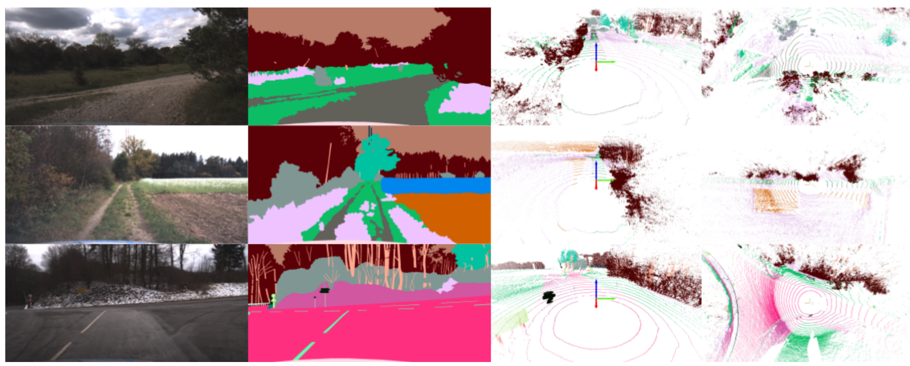

Fig. 1. Illustrative samples of GOOSE dataset.

Manual labeling of 3D point clouds is an inherently time-consuming and resource-intensive task. In unstructured environments, this difficulty is magnified due to the ambiguity of semantic classes (e.g., distinguishing between different types of low vegetation or between traversable ground and irregular terrain) and the complexity of delineating objects with organic and irregular shapes. Despite these challenges, several public datasets have been developed for off-road autonomous navigation.

For our experimental evaluation, we have focused on the German Outdoor and Offroad Dataset (GOOSE) [24]. It is a large-scale dataset specifically designed for perception in a wide variety of unstructured outdoor environments in Germany. The dataset incorporates 10,000 labeled pairs of images and point clouds captured from a ground vehicle across diverse off-road scenarios, including forests, fields, and dirt tracks. As shown in Figure 1, the dataset provides rich, multimodal data, including RGB images of the scene, the corresponding 3D LiDAR point clouds, and the pixel-wise 2D semantic segmentation ground truth. GOOSE provides its own detailed ontology with 64 semantic classes, which for our evaluation purposes were grouped into the main categories relevant to our study. The scale, variety of environments, and quality of the annotations make this dataset a crucial benchmark for developing and testing robust perception models for off-road navigation.

4 Deep learning models for LiDAR semantic segmentation

Processing 3D point clouds with deep neural networks presents a unique challenge compared to structured data like images. Point clouds are unordered sets of points in space, meaning their representation is invariant to permutations. Furthermore, the number of points can vary, and they must be robust to geometric transformations such as rotation and translation. To address these challenges, specialized architectures have been developed. In this work, we focus on two foundational models that pioneered the direct processing of point clouds: PointNet and PointNet++.

4.1 PointNet

PointNet [12] was a groundbreaking architecture, being one of the first to directly consume raw point clouds without converting them to intermediate regular formats like voxels or 2D images. The key insight of PointNet is to learn a spatial encoding for each point individually and then aggregate all individual point features into a global representation.

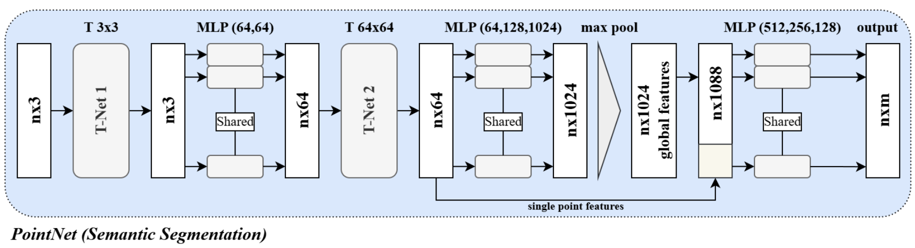

Fig. 2. Block diagram of the PointNet architecture.

To achieve permutation invariance, PointNet uses a simple and powerful strategy:

- **Shared Multi-Layer Perceptrons (MLPs):** Each point in the cloud is processed independently and identically through a series of shared MLPs. This ensures that the network learns to extract the same type of features for every point, regardless of its position in the input order.
- **Symmetric Function:** A symmetric function, specifically max pooling, is applied across the feature dimension of all points. This function aggregates the learned features from all points into a single global feature vector that describes the entire shape. This aggregation is inherently invariant to the order of the points.
- **Transformation Networks (T-Nets):** To ensure invariance to geometric transformations, PointNet incorporates small neural networks called T-Nets. These networks predict an affine transformation matrix that is applied to the input points and their features, canonicalizing the data before feature extraction.

For semantic segmentation, where a label is required for each point, PointNet concatenates the global feature vector with the per-point features learned before the max pooling layer. This combination provides both global context and local information for each point, allowing the network to make a final per-point prediction.

4.2 PointNet++

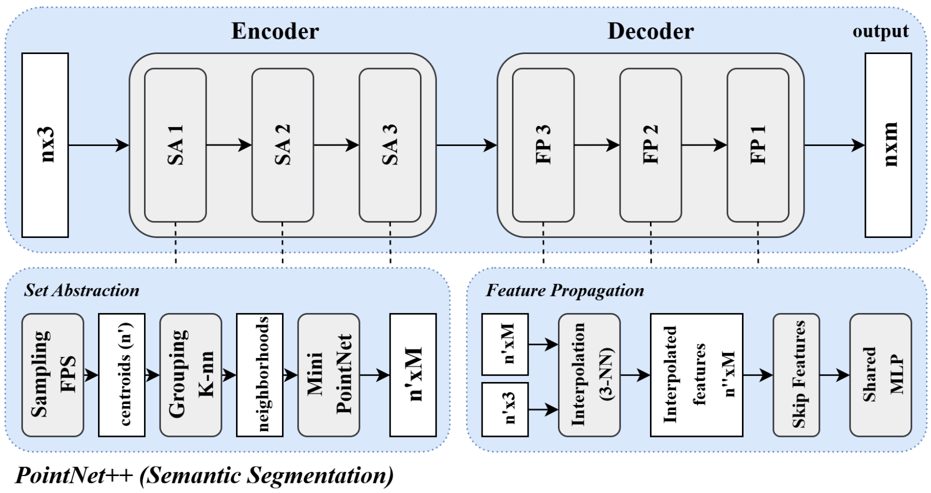

Fig. 3. Block diagram of the PointNet++ architecture.

While PointNet was revolutionary, it did not explicitly capture the local geometric structure induced by the metric space of the points. It treated each point independently before global aggregation, missing fine-grained patterns in local regions. PointNet++ [13] was proposed to address this limitation by introducing a hierarchical feature learning approach that captures features at multiple scales.

PointNet++'s architecture is built upon a series of *Set Abstraction (SA)* modules, which progressively abstract a larger and larger region of the point cloud into a higher-dimensional feature vector. Each SA module consists of three key layers:

- **Sampling Layer:** A subset of points is selected from the input set, defining the centroids of local regions. The Farthest Point Sampling (FPS) algorithm is used to ensure the centroids cover the entire point cloud, providing a better receptive field than random sampling.
- **Grouping Layer:** For each centroid, a local region is defined by finding all points within a certain radius (k-nearest neighbors). These points form a local neighborhood.
- **PointNet Layer:** A mini-PointNet is used to process each local region, extracting a higher-level feature vector that summarizes the geometric pattern within that neighborhood.

By stacking these SA modules, PointNet++ creates an encoding hierarchy. At each level, the number of points is reduced, but the dimensionality of their feature representation is increased. This process is analogous to the convolutional and pooling layers in a CNN, which reduce spatial resolution while increasing the number of feature channels.

For dense prediction tasks like semantic segmentation, the features must be propagated back to the original, full-resolution point cloud. PointNet++ achieves this through a hierarchical feature propagation architecture that works as a decoder. Feature propagation is done level by level, using interpolation and skip connections:

- **Interpolation:** In the decoder stage, features learned at higher levels (where there are fewer points and the information is more global) are transferred back to lower, denser levels. This transfer is done by interpolating the features of the nearest neighboring points, assigning higher weights to closer points (inverse distance weighted interpolation).
- **Skip Connections:** After interpolation, these features are combined with the ones stored from the encoder at the same hierarchical level. This allows the network to recover fine local details that may have been lost during hierarchical sampling, while also leveraging the global context captured at coarser levels. As a result, the model can make per-point decisions using both broad contextual and precise local information.
- **Shared MLP:** After this concatenation, the combined features are passed through a "unit PointNet" (a few shared MLP layers) to update the per-point features. This process is repeated until the features have been propagated to all original points.

This hierarchical structure with its encoding and decoding phases allows the model to learn robust and detailed features at multiple scales, leading to superior performance in complex scene understanding tasks.

4.3 PointNet++*

The standard PointNet++ architecture primarily operates on the spatial coordinates (XYZ) of the point cloud. However, LiDAR sensors provide additional valuable information, such as remission intensity. This value, which measures the return strength of the laser pulse, can offer crucial clues about the material properties of a surface, helping to distinguish between objects with similar geometry but different materials.

To leverage this information, we propose a simple yet effective modification to the PointNet++ backbone, which we denote as PointNet++*. As illustrated in Figure 4, the modification is applied at the final stage of the network, after the decoder has propagated the learned hierarchical features back to the original points. It is important to note that the feature values output by the decoder and the raw remission intensity values have very different scales. The decoder's features are approximately in the range [0, 3.2], whereas intensity can be much

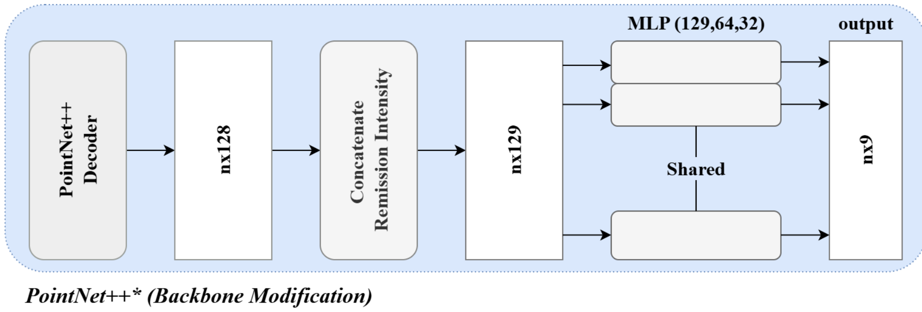

Fig. 4. Proposed modification to the PointNet++ backbone.

larger. To prevent the intensity values from dominating the feature space and to ensure a stable training process, we first normalize the remission intensity using Z-score normalization. The mean and standard deviation required for this normalization are computed exclusively from the GOOSE training set to perform experiments correctly. These are the proposed modifications:

- The normalized remission intensity value for each point is then concatenated to its corresponding feature tensor. This augments the feature space, resulting in a new tensor of size $N \times 129$.
- This combined feature tensor, which now contains both geometric context and material information at a comparable scale, is passed through a final shared MLP for per-point classification.

This late-fusion approach allows the network to first learn complex spatial hierarchies and then use the intensity information as a powerful final discriminator to refine its predictions.

5 Experimental evaluation

In this section, we present a detailed quantitative and qualitative analysis of the performance of the proposed models. To prepare the data, we first preprocessed the point clouds by cropping them within a 25-meter radius centered on the vehicle. From this cropped region, we randomly subsampled 16,384 points to create uniform inputs for training, validation, and testing. To introduce variability and promote generalization, this random subsampling was performed independently at each iteration during training. In contrast, for validation and testing, a fixed subset of subsampled points was used consistently to ensure fair comparisons across all experiments.

Furthermore, the original semantic classes from the GOOSE dataset were consolidated into a set of nine final categories to better align with the intended application. Specifically, classes such as `human` and `animal` were grouped into the broader `Object` category (representing generic obstacles), while the `sky` class was merged into the `Void` category. The final set of evaluation classes

Table 1. Intersection over Union (IoU %) per class (see the corresponding column) and model (row) when considering point clouds not used for training. The mean of all IoU percentages (mIoU %) is presented in the rightmost column.

Model	Construction	Object	Road	Sign	Terrain	DVeg.	NDVeg.	Vehicle	Void	mIoU
PointNet	23.62	11.08	45.62	28.39	31.74	33.46	28.37	10.64	79.60	**32.50**
PointNet++	69.90	23.16	27.66	40.15	59.01	66.69	86.45	72.63	82.78	**58.71**
PointNet++*	68.32	25.24	31.30	42.77	59.26	70.15	88.91	74.67	82.79	**60.38**

includes: `Construction`, `Object`, `Road`, `Sign`, `Terrain`, `Drivable vegetation`, `Non drivable vegetation`, `Vehicle`, and `Void`.

Each architecture is evaluated on point clouds not considered during training. We considered key performance metrics, in particular *Intersection over Union (IoU)* and *Recall*, obtained as

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \qquad \mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{1}$$

where, for a specific class, TP (true positives) is the number of correctly predicted positive points, FP (false positives) refer to points incorrectly identified as positive by the model, and FN (false negatives) are instances incorrectly identified as negative. Thus, IoU measures the spatial overlap between the predicted segmentation and the ground truth, while *Recall* quantifies the model's ability to detect actual positives (higher recall means fewer positive instances are missed). To complement these figures of merit, we also present the corresponding confusion matrices to analyze strengths and limitations across classes, since class representation is imbalanced. Finally, we examine how point cloud density and sensor noise affect model robustness.

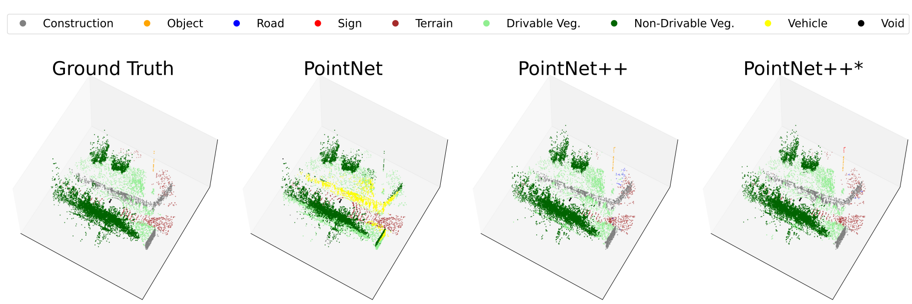

Fig. 5. Qualitative comparison of semantic segmentation results on a GOOSE test point cloud not used for model training. Class labels are shown at the top.

5.1 Quality assessment of PointNet

The PointNet model was evaluated to establish a performance baseline. The quantitative results in Table 1 and Table 2 show that the model struggles with complex classes. This is visually confirmed in Figure 5, where the PointNet

Table 2. Recall (%) per class (see the corresponding column) and model (row) when considering point clouds not used for training. The mean of all Recall percentages (mRecall %) is in the rightmost column.

Model	Construction	Object	Road	Sign	Terrain	DVeg.	NDVeg.	Vehicle	Void	mRecall
PointNet	47.36	18.25	70.39	43.18	45.78	51.63	47.46	23.03	87.79	**48.32**
PointNet++	78.83	40.14	34.05	60.65	73.73	80.63	93.87	78.13	98.07	**70.90**
PointNet++*	86.02	32.10	59.88	59.39	68.95	87.06	93.35	84.91	83.74	**72.82**

segmentation shows significant failures, misclassifying large portions of the scene, such as labeling construction as a vehicle. The confusion matrix in Figure 6 further reveals these issues, with significant confusion between classes, indicating a difficulty for the model to capture the distinct local features necessary to differentiate between geometrically similar objects.

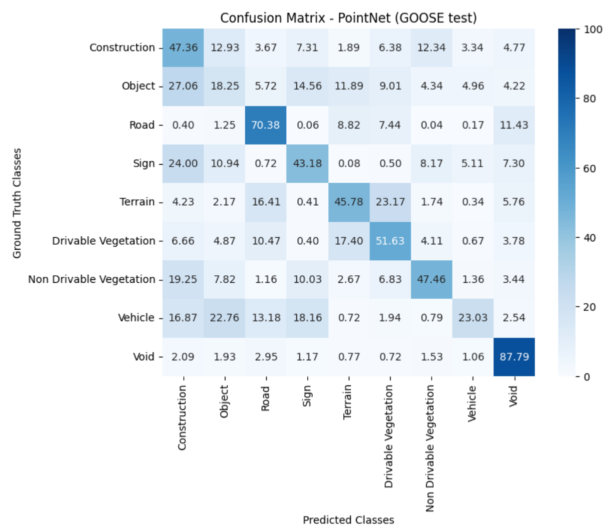

Fig. 6. Confusion matrix for PointNet model.

5.2 Quality assessment of PointNet++

To address PointNet's limitations, the standard PointNet++ model was implemented. This hierarchical model demonstrates a significant performance leap, boosting the overall mIoU to 58.71%. The qualitative results in Figure 5 show a much more coherent and accurate interpretation of the environment, correctly identifying the terrain, constructions, and vegetation boundaries. The confusion matrix in Figure 7 supports this, showing a much cleaner diagonal, which confirms that the hierarchical feature learning approach effectively captures local geometric patterns. However, some confusion persists; for example, `Object` is still misclassified as `Non Drivable Vegetation` (24.33%).

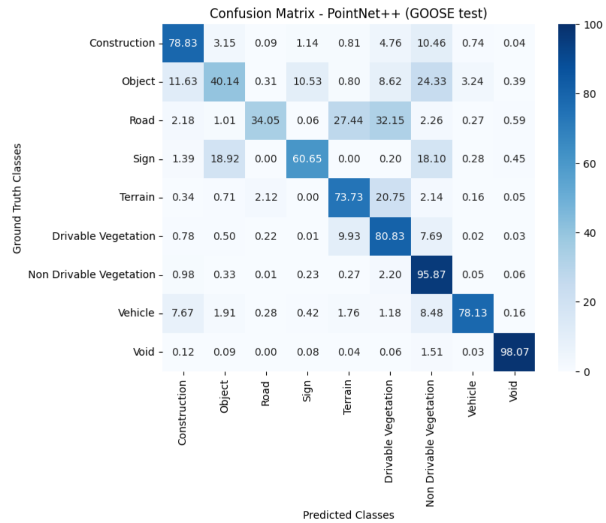

Fig. 7. Confusion matrix for PointNet++ model.

5.3 Quality assessment of PointNet++*

The proposed PointNet++* variant, which incorporates LiDAR intensity data, further refines the predictions, achieving the best overall mIoU of 60.38%. Visually, as seen in Figure 5, the segmentation is very similar to the standard PointNet++, maintaining a coherent scene interpretation while refining small details. The confusion matrix in Figure 8 shows a notable improvement in recall for key classes, with `Vehicle` reaching 84.91%. This suggests that intensity information provides powerful discriminative features, helping the model to better resolve ambiguities between classes with similar geometric profiles.

5.4 Density effect

Point cloud density is a critical factor in real-world applications. To evaluate the robustness of our best model, PointNet++*, we subjected it to different levels of subsampling, from the original 16k points down to 2k points. Figure 9 shows the confusion matrices for each density level. As expected, the model's accuracy degrades as the point cloud becomes sparser. At 10k and 8k points, the model maintains high performance. However, a significant drop is observed at 6k points, where confusion between classes like `Terrain`, `Road`, and different types of vegetation increases. At 2k points, the model struggles to identify most classes correctly, with widespread confusion making the segmentation unreliable. This analysis shows that while the model is robust to moderate density reductions, its performance is highly dependent on a sufficient number of points to capture detailed geometric features.

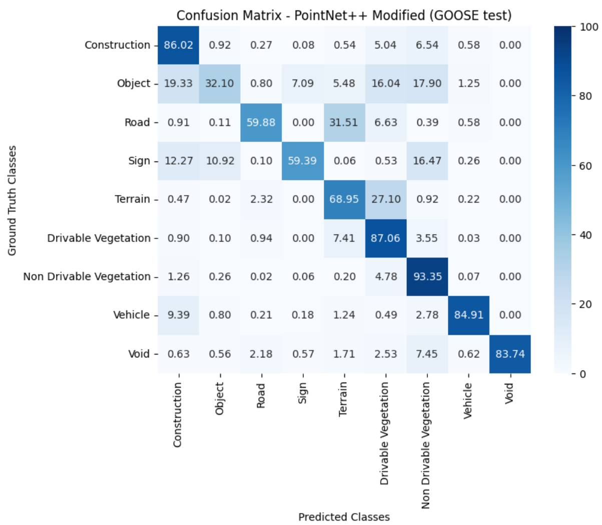

Fig. 8. Confusion matrix for PointNet++* model.

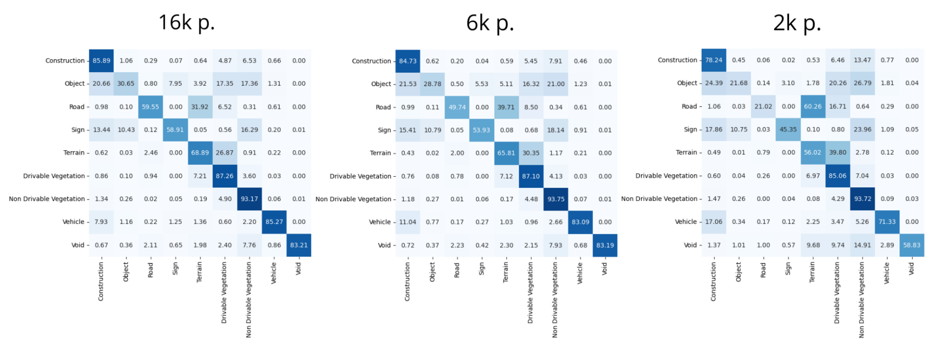

Fig. 9. Impact of Point Cloud Downsampling on Semantic Segmentation (PointNet++*).

5.5 Noise effect

Besides point density, sensor noise is another key challenge. We simulated thermal noise by adding Gaussian noise with an increasing standard deviation (σ) to the point coordinates. Figure 10 illustrates the performance of the PointNet++* model under noise levels of $\sigma = 0.01$, $\sigma = 0.04$, and $\sigma = 0.08$. The model shows remarkable robustness at $\sigma = 0.01$, with minimal changes in the confusion matrix obtained when evaluating original data. At $\sigma = 0.04$, performance begins to degrade, particularly for the `Sign` class. At a high noise level of $\sigma = 0.08$, the model's performance drops sharply, with widespread misclassifications, especially for classes that are not structurally dominant, highlighting the model's reduced ability to extract meaningful features under high noise.

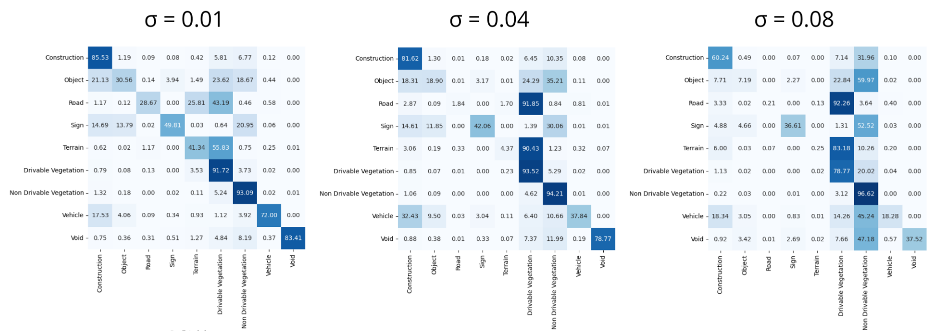

Fig. 10. Impact of Gaussian Noise on Semantic Segmentation (PointNet++*).

6 Conclusions

This work explores the use of deep learning architectures for the semantic segmentation of LiDAR point clouds in unstructured environments, a fundamental challenge for off-road autonomous navigation. Through a comprehensive experimental evaluation using the GOOSE dataset, we benchmarked two baseline models, PointNet and PointNet++, and a novel variant named PointNet++* has been proposed to incorporate LiDAR intensity data.

The experimental results conclusively demonstrate that hierarchical architectures outperform approaches that process points globally. PointNet, although pioneering, showed significant limitations for capturing the fine-grained local structures necessary for accurate segmentation in complex scenes. This is reflected in the low values for mIoU and mRecall, 32.50% and 48.32%, respectively. PointNet++ increased mIoU performance to 58.71% and mRecall to 70.90%, by capturing multi-scale geometric features. Our proposal, PointNet++*, further improved both figures of merit, reaching 60.38% mIoU and 72.82% mRecall, which confirms that fusing additional information such as LiDAR intensity can reduce class confusion.

The robustness analysis, a critical aspect for real-world applications, revealed the operational limits of these models. It was observed that PointNet++* is resilient to moderate reductions in point density, maintaining good performance down to 8k points. Below this, performance decreases considerably, being insufficient for reliable segmentation with only 2k points. Similarly, the model showed high tolerance to low levels of simulated thermal noise, but its performance degraded notably under high noise levels, mainly affecting the ability to distinguish small objects and classes with poorly defined boundaries.

In summary, hierarchical models like PointNet++ are well-suited for semantic segmentation in off-road settings, and multimodal feature integration such as LiDAR intensity enhances prediction performance in unstructured environments. For future work, we will explore more advanced architectures, such as those based on transformers, and investigate on more sophisticated intensity fusion strategies beyond the current late-fusion approach. Enhancing data augmentation techniques could also improve robustness to sparse and noisy inputs.

Finally, the models trained on GOOSE will be tested on other datasets such as the RELLIS-3D to assess generalization by leveraging its distinct environments and sensor configurations in data capture.

Acknowledgments. This work is partially supported by the GAIA project 'Gestión integral para la prevención, extinción y reforestación debido a incendios forestales', funded by the Spanish Agencies AEI and CDTI under Grant PLEC2023-010303 (TRANSMISIONES 2023). It has also been funded by iROBOCITY 2030-CM project (Ref TEC-2024/TEC-62), from Programa de Actividades de I+D entre grupos de investigación de la Comunidad de Madrid en Tecnologías 2024.

References

1. Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., Gall, J.: SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9297–9307 (2019)
2. Borges, P.V., Peynot, T., Liang, S., Arain, B., Wildie, M., Minareci, M.G., Lichman, S., Samvedi, G., Sa, I., Hudson, N., et al.: A survey on terrain traversability analysis for autonomous ground vehicles: Methods, sensors, and challenges. Field Robotics **2**, 1567–1627 (2022)
3. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, X., Baldan, G., Beijbom, O.: nuscenes: A multimodal dataset for autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11621–11631 (2020)
4. Choy, C., Gwak, J., Savarese, S.: 4d spatio-temporal convnets: Minkowski convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3075–3084 (2019)
5. Gao, B., Pan, Y., Li, C., Geng, S., Zhao, H.: Are we hungry for 3d lidar data for semantic segmentation? a survey of datasets and methods. IEEE Transactions on Intelligent Transportation Systems **23**(7), 6063–6081 (2021)
6. Hossain, M.S., Rahman, M., Rahman, A., Kabir, M.M., Mridha, M., Huang, J., Shin, J.: Automatic navigation and self-driving technology in agricultural machinery: A state-of-the-art systematic review. IEEE Access (2025)
7. Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A.: Randla-net: Efficient semantic segmentation of large-scale point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11108–11117 (2020)
8. Jhaldiyal, A., Chaudhary, N.: Semantic segmentation of 3d lidar data using deep learning: a review of projection-based methods. Applied Intelligence **53**(6), 6844–6855 (2023)
9. Lai, X., Chen, Y., Lu, F., Liu, J., Jia, J.: Spherical transformer for lidar-based 3d recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17545–17555 (2023)
10. Muhammad, K., Hussain, T., Ullah, H., Del Ser, J., Rezaei, M., Kumar, N., Hijji, M., Bellavista, P., de Albuquerque, V.H.C.: Vision-based semantic segmentation in scene understanding for autonomous driving: Recent achievements, challenges, and outlooks. IEEE Transactions on Intelligent Transportation Systems **23**(12), 22694–22715 (2022)

11. Oliveira, L.F., Moreira, A.P., Silva, M.F.: Advances in forest robotics: A state-of-the-art survey. Robotics **10**(2), 53 (2021)
12. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. arXiv preprint arXiv:1612.00593 (2016)
13. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. arXiv preprint arXiv:1706.02413 (2017)
14. Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, D.A., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., Vasudevan, V., Han, W., Ngiam, J.: Scalability in perception for autonomous driving: Waymo open dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2446–2454 (2020)
15. Tang, H., Liu, Z., Zhao, S., Lin, Y., Lin, J., Wang, H., Han, S.: Searching efficient 3d architectures with sparse point-voxel convolution. In: European Conference on Computer Vision. pp. 685–702. Springer (2020)
16. Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J.: Kpconv: Flexible and deformable convolution for point clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6411–6420 (2019)
17. Tong, K., Hu, Y., Dikic, B., Solmaz, S., Fraundorfer, F., Watzenig, D.: Robots saving lives: A literature review about search and rescue (sar) in harsh environments. In: IEEE Intelligent Vehicles Symposium. pp. 953–960. IEEE (2024)
18. Varga, D., Witting, C., Stiller, C.: Rellis-3d dataset: A multi-modal dataset for off-road robotics. IEEE Robotics and Automation Letters **6**(3), 5437–5444 (2021)
19. Vidanapathirana, K., Knights, J., Hausler, S., Cox, M., Ramezani, M., Jooste, J., Griffiths, E., Mohamed, S., Sridharan, S., Fookes, C., et al.: Wildscenes: A benchmark for 2d and 3d semantic segmentation in large-scale natural environments. The International Journal of Robotics Research **44**(4), 532–549 (2025)
20. Viswanath, K., Jiang, P., Saripalli, S.: Reflectivity is all you need!: Advancing lidar semantic segmentation. arXiv preprint arXiv:2403.13188 (2024)
21. Viswanath, K., Jiang, P., Sujit, P., Saripalli, S.: Off-road lidar intensity based semantic segmentation. In: Intl Symposium on Experimental Robotics. pp. 608–617. Springer (2023)
22. Viswanath, K., Singh, K., Jiang, P., Sujit, P., Saripalli, S.: Offseg: A semantic segmentation framework for off-road driving. In: IEEE Intl Conf on Automation Science and Engineering. pp. 354–359. IEEE (2021)
23. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph cnn for learning on point clouds. ACM Transactions on Graphics **38**(5), 1–12 (2019)
24. Wenge, R., Zallner, M., von der Burchard, A., Stiller, C.: GOOSE: A german outdoor and offroad dataset for unstructured environments. In: 2023 IEEE International Conference on Robotics and Automation. pp. 6348–6354. IEEE (2023)
25. Yan, X., Zheng, C., Xue, Y., Li, Z., Cui, S., Dai, D.: Benchmarking the robustness of lidar semantic segmentation models. International Journal of Computer Vision **132**(7), 2674–2697 (2024)
26. Zhao, H., Jiang, L., Jia, J., Torr, P.H., Koltun, V.: Point transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 16259–16268 (2021)
27. Zhu, X., Zhou, H., Wang, T., Hong, F., Ma, Y., Li, W., Li, H., Lin, D.: Cylindrical and asymmetrical 3d convolution networks for lidar segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9939–9948 (2021)