Understanding Cross-Domain Robustness in LiDAR Semantic Segmentation

Yewon Song
Automotive Engineering
Hanyang University
Seoul, Korea
swy1155@hanyang.ac.kr

Sumin Lee
Automotive Engineering
Hanyang University
Seoul, Korea
maroona@hanyang.ac.kr

Soonmin Hwang*
Automotive Engineering
Hanyang University
Seoul, Korea
soonminh@hanyang.ac.kr

***Abstract*—Real-world deployment of perception models requires generalization beyond the environments encountered during training. However, collecting and annotating data that cover all possible conditions is infeasible. Consequently, models often suffer from performance degradation when applied to new domains, due to factors such as differences in beam configurations, sensor noise, and environmental conditions. Addressing these cross-dataset domain shifts is therefore essential for ensuring robustness and generalization. In this work, we evaluate perception model across domains and study strategies that help alleviate performance degradation.**

***Index Terms*—LiDAR Semantic Segmentation, Autonomous Driving, Domain Adaptation, Point Clouds, Deep Learning**

I. Introduction

In autonomous driving, perception models are crucial for understanding and interacting with the environment. Recent progress in deep learning has advanced 3D perception tasks including detection, segmentation, and scene understanding. While these models achieve strong results on benchmarks with training-like sensor setups and environments, their performance often degrades sharply in real-world scenarios with different sensors or environments.

One of the key challenges in real-world deployment is domain shift, which refers to the discrepancy between the distribution of training data and that of the deployment environment. In the context of LiDAR-based perception, such domain shifts commonly arise from differences in beam configurations across sensors, variations in sensor noise characteristics, and environmental factors such as weather, lighting, or scene structure. As collecting and annotating data that covers all possible conditions is infeasible, models are inevitably forced to operate outside their learned distribution, which frequently leads to severe drops in accuracy and reliability.

Addressing these cross-dataset performance gaps is crucial for building robust and generalizable perception systems. A large body of work has explored methods for unsupervised domain adaptation (UDA), self-training, feature alignment, and data augmentation strategies to mitigate these issues. While many of these approaches report promising results, their effectiveness often varies depending on the specific characteristics of the source and target domains. This variability highlights the need for systematic evaluation across diverse settings.

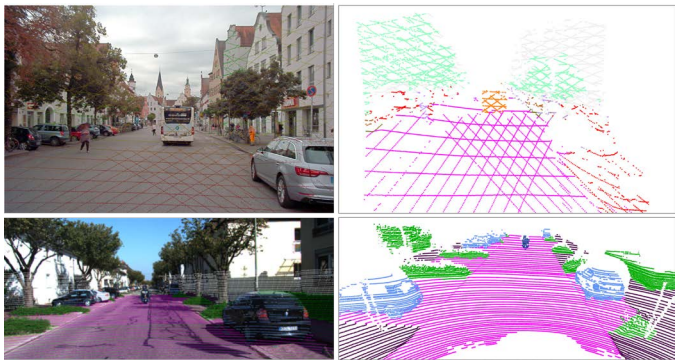

Fig. 1: Examples from different driving datasets: A2D2 with 16-channel LiDAR (top) and SemanticKITTI with 64-channel LiDAR (bottom)

In this work, we focus on evaluating the perception model across diverse domains and highlight the challenges that arise when models are deployed in unseen environments. A key issue in this setting is the significant performance degradation caused by variations in sensor configurations, beam patterns, environmental conditions, and data distributions. Our goal is to systematically assess how existing adaptation strategies perform under such shifts, identifying their ability in narrowing the performance gap. We present experimental results across the domain pair and emulate these sensor characteristics during training and evaluate their impact on model performance.

II. Task: LiDAR Semantic Segmentation (LSS)

Recent advances in LSS can be categorized into point-based, voxel-based, and projection-based approaches. Early works such as [1], [2] introduced point-based processing with permutation-invariant architectures, followed by [3], [4], which improved local context modeling using convolutional kernels and attention mechanisms. However, their limited ability to capture local geometric context and the high computational cost on large-scale scenes restricted scalability. Voxel-based methods address this by discretizing 3D space into grids and applying sparse convolutions. [5] showed the effectiveness of sparse CNNs, while [6], [7] improved efficiency and accuracy with cylindrical voxelization and point-voxel hybrids.

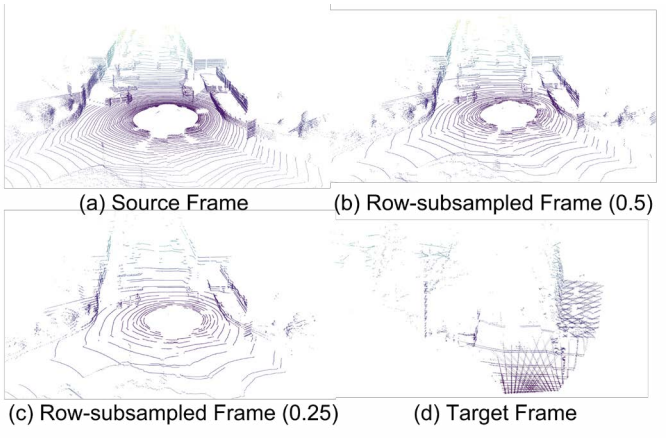

Fig. 2: Comparison of (a) an original SemanticKITTI LiDAR scan with 64 beams; (b–c) scans produced by row-subsampling the range image by uniformly dropping rows with probabilities 0.5 and 0.25, respectively; and (d) a A2D2 LiDAR scan with 16 beams. Row-subsampling effectively simulates scans captured with fewer laser beams.

In parallel, projection-based methods transform point clouds into structured 2D representations, such as range images or BEV maps, enabling the use of mature 2D CNN architectures. Notable works [8]–[10] achieve fast inference but suffer from information loss due to occlusions and discretization. More recently, hybrid strategies combining different paradigms have been proposed to balance efficiency, and scalability in LSS. Representative works include PMF [11] and its efficient extension EPMF [12], which integrate camera-perspective RGB features with point-level LiDAR representations via perspective projection to achieve improved robustness across varying LiDAR configurations.

III. Approach: Sensor-level subsampling

To address the discrepancy introduced by heterogeneous LiDAR sensors, we apply a ray subsampling strategy that adjusts the number of beams to match the angular resolution of the target domain. This simple preprocessing step reduces the distribution shift and stabilizes the subsequent adaptation process. As shown in Figure 2, the subsampling of the rows from the source frame produces scans that resemble those of the sensors with the target frame, for example when transferring from a 64-beam HDL-64E to a 16-beam VLP-16.

Beyond aligning sensor characteristics, this strategy directly mitigates the mismatch in angular resolution that often leads to feature sparsity and degraded predictions when models are applied to unseen sensors. Moreover, by adjusting the subsampling patterns during training, the model is exposed to a broader range of beam configurations, which acts as a lightweight data augmentation and improves robustness. Similar strategies, such as beam dropout and structured subsampling, have also been explored in prior UDA studies, further supporting the effectiveness of sensor-level adaptation in bridging cross-sensor domain gaps.

TABLE I: Performance Gap Across Domains.

		mIoU (%)
Source / Target	SemanticKITTI (64-ch)	A2D2 (16-ch)
SemanticKITTI	74.95	16.27
A2D2	22.16	35.50

Notes. KITTI denotes *SemanticKITTI*; A2D2 denotes *Audi A2D2*. "Source" = no adaptation.

IV. Experiments

A. Datasets

We focus on [13] and [14], which differ in LiDAR configurations. This difference makes them suitable for studying domain shift and adaptation.

SemanticKITTI [13] is a large-scale benchmark for 3D semantic segmentation, derived from the KITTI odometry sequences. It provides point-wise labels for over 4.5 billion points across 22 outdoor driving sequences captured with a Velodyne HDL-64E (64-channel) LiDAR, making it a standard dataset for evaluating LiDAR-based segmentation models in autonomous driving.

A2D2 [14] is a multi-modal benchmark featuring synchronized data from multiple cameras and five Velodyne VLP-16 (16-channel) LiDARs. Semantic and instance-level labels are annotated on 2D images, which can be projected onto the corresponding 3D point clouds. This image-based annotation scheme provides consistent supervision across modalities and supports the development and evaluation of multi-modal fusion approaches for 2D/3D semantic scene understanding.

B. Settings

Label mapping. Since [13] and [14] define different class taxonomies, we restrict our experiments to the ten categories common to both datasets, following [15]. We select 10 shared classes between the 2 datasets by merging or ignoring them (see Tab. 5). The 10 final classes are car, truck, bike, person, road, parking, sidewalk, building, nature, other-objects. We adopt mean Intersection-over-Union (mIoU) as the evaluation metric.

Subsampling. To investigate the effect of sensor heterogeneity, we augment the SemanticKITTI training set by emulating different beam configurations. Specifically, 30% of the training sequences are downsampled to 32-channel scans and another 30% are downsampled to 16-channel scans, while the remaining 40% retain the original 64-channel resolution. The subsampling follows the vertical beam layout of the target sensors, and for each ray, only the nearest return is preserved using a z-buffer strategy. This mixed-resolution training set exposes the model to heterogeneous beam configurations and enables more robust learning under cross-domain deployment.

TABLE II: Quantitative results per class
(Source: Semantic KITTI)

Target	Car	Bicycle	Motorcycle	Truck	Other vehicle	Pedestrian	Drivable Surf.	Sidewalk	Terrain	Vegetation	mIoU (%)
SemanticKITTI	94.35	67.88	47.29	58.94	68.32	67.06	95.86	82.66	75.98	91.15	74.95
A2D2	43.88	39.42	0.00	12.88	0.00	24.82	1.10	0.85	0.22	39.44	16.27
A2D2 (+Subsampling)	**44.10**	38.89	**1.04**	**13.06**	**0.12**	**27.00**	0.09	**1.16**	**0.51**	39.29	**16.53**

Notes. Source = training dataset, Target = evaluation dataset.

C. Results

Table I reports the domain adaptation performance across SemanticKITTI(SK) and A2D2. When training and evaluating within the same dataset (e.g., SK→SK or A2D2→A2D2), the model achieves relatively high performance without adaptation. However, transferring directly across datasets (SK→A2D2 or A2D2→SK) leads to a sharp performance drop due to distributional discrepancies. Applying the proposed strategy improves mIoU over the source-only baseline, demonstrating its effectiveness in mitigating domain shift between heterogeneous LiDAR datasets.

Table II provides per-class IoU scores on SemanticKITTI and A2D2. The results highlight that well-represented classes such as *Car* and *Pedestrian* achieve high IoU, while long-tail categories such as *Motorcycle* and *Sidewalk* remain more challenging. The class-wise breakdown reveals dataset-specific biases and indicates that domain adaptation methods need to account for both frequent and underrepresented categories to achieve balanced performance across classes.

V. Conclusion

In this work, we examined the performance degradation arising from domain discrepancies in LiDAR-based semantic segmentation. While our experiments applied subsampling strategies to the training data, part of the observed improvement may stem from their effect as a data augmentation technique rather than from explicit domain adaptation. As a next step, we plan to explore test-time adaptation approaches that can operate in a plug-and-play manner, enabling models to adapt to new domains without additional training cost or labeled data.

VI. Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.RS-2025-00409492).

References

[1] 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)*, pp. 652–660, 2017.

[2] 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)*, pp. 5099–5108, 2017.

[3] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas, "Kpconv: Flexible and deformable convolution for point clouds," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 6411–6420, 2019.

[4] H. Zhao, L. Jiang, J. Jia, P. H. Torr, and V. Koltun, "Point transformer," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[5] C. Choy, J. Gwak, and S. Savarese, "4d spatio-temporal convnets: Minkowski convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3075–3084, 2019.

[6] X. Zhu, H. Zhou, T. Wang, F. Hong, Y. Ma, W. Li, and D. Lin, "Cylindrical and asymmetrical 3d convolution networks for lidar segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9939–9948, 2021.

[7] H. Tang, Z. Liu, S. Zhao, Y. Lin, J. Lin, H. Wang, and S. Han, "Searching efficient 3d architectures with sparse point-voxel convolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1921–1930, 2020.

[8] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "Rangenet++: Fast and accurate lidar semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 0–0, 2019.

[9] T. Cortinhal, C. Tzelepis, and E. E. Aksoy, "Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds for autonomous driving," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 415–422, 2020.

[10] Y. Zhang, Z. Zhou, P. David, X. Yue, Z. Xi, B. Gong, and H. Foroosh, "Polarnet: An efficient polar representation for 3d segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9601–9610, 2020.

[11] Z. Zhuang, R. Li, K. Jia, Q. Wang, Y. Li, and M. Tan, "Perception-aware multi-sensor fusion for 3d lidar semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 16280–16290, 2021.

[12] M. Tan, Z. Zhuang, S. Chen, R. Li, K. Jia, Q. Wang, and Y. Li, "Epmf: Efficient perception-aware multi-sensor fusion for 3d semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 8258–8273, 2024.

[13] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "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.

[14] J. Geyer, Y. Kassahun, M. Mahmudi, X. Ricou, R. Durgesh, A. S. Chung, L. Hauswald, V. H. Pham, M. Mühlegg, S. Dorn, *et al.*, "A2d2: Audi autonomous driving dataset," *arXiv preprint arXiv:2004.06320*, 2020.

[15] L. Yi, B. Gong, and T. Funkhouser, "Complete & label: A domain adaptation approach to semantic segmentation of lidar point clouds," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15363–15373, 2021.