A. OmniLiDAR details

A.1. Dataset sampling distribution

As mentioned in the main manuscript, the datasets within OmniLiDAR vary drastically in size. For instance, Argoverse 2 Lidar [11] contains nearly 12 million raw scans, whereas smaller datasets like View-of-Delft [9] or SemanticKITTI [2] contain significantly fewer. To prevent the model from biasing toward the largest data collections and to ensure balanced cross-sensor generalization, we employ a custom weighted sampling strategy during training.

We group the datasets into specific sampling buckets and assign a fixed target probability to each. The distribution is designed to prioritize our primary evaluation domains while capping the influence of massive collections:

- **Primary datasets (60% total):** The datasets corresponding to our main evaluation benchmarks are heavily prioritized. We allocate $20\,\%$ to `nuScenes_main`, $20\,\%$ to `WaymoPerception_main`, and $20\,\%$ to `KITTI_main`.
  - For `KITTI_main`, we merge SemanticKITTI [2] and KITTI-360 [6] into a single group, as they share highly similar environments and the exact same hardware configuration (Velodyne HDL-64E).
  - For `WaymoPerception_main`, we strictly route scans from the primary roof-mounted panoramic sensor (`TOP`) to this group [10]. The shorter-range secondary sensors are routed to a generic fallback group.
- **Massive datasets (15% total):** To prevent domination during training, we cap the largest datasets, *i.e.* Argoverse 2 Lidar, ONCE, and Zenseact Open Dataset (ZOD), at $5\,\%$ each.
- **Other datasets (22.5% total):** We assign $5\,\%$ each to AevaScenes [8], Lyft [5], PandaSet [12], and TruckScenes [4]. Because View-of-Delft [9] is a relatively small dataset, we assign it a lower sampling weight of $2.5\,\%$.
- **Fallback group (2.5% total):** An `else` group is allocated the remaining $2.5\,\%$. This group catches Waymo's non-TOP secondary sensors [10] and any other unspecified scan data.

By utilizing these predefined probabilities, our dataloader ensures that each epoch (defined as $20\,000$ LiDAR samples) exposes the network to a consistently diverse, domain-balanced batch of geometries.

A.2. Overview of unique LiDAR sensors

In Table 1, we detail the 15 distinct LiDAR hardware models aggregated within the OmniLiDAR dataset. This hardware diversity is fundamental to achieving TerraSeg's domain-agnostic generalization. The table illustrates a broad spectrum of sensor modalities, ranging from classic high-resolution spinning LiDARs (*e.g.* Velodyne HDL-64E, Hesai Pandar64) to specialized configurations such as roof-mounted panoramic sensors, near-field short-range LiDARs, and Frequency-Modulated Continuous Wave (FMCW) sensors (Aeva). By exposing our self-supervised model to this broad variation in scan patterns, point densities, and vertical fields of view during training, we encourage the network to learn universal geometric priors rather than overfitting to sensor-specific artifacts.

A.3. OmniLiDAR edge cases

During the curation of the OmniLiDAR dataset, we standardize and process raw scans to ensure consistency across the 12 datasets. To maintain a robust data loading pipeline, we addressed the following dataset-specific edge cases:

1. Missing returns in Argoverse 2 Lidar: Occasionally, the secondary LiDAR sensors in this dataset yield an empty point cloud. Because our data loading protocol expects continuous files for the curated $0.2\,\mathrm{Hz}$ sequences, we generate placeholder empty files for these instances. During training, if the dataloader encounters an empty file, it dynamically resamples to retrieve a valid point cloud. Out of the $240\,000$ processed files for this dataset, $236\,904$ contain valid sensor returns, which is accurately reflected in the aggregate statistics.

2. Frame selection in Zenseact Open Dataset: Zenseact provides curated frames structured as two-second snippets containing past, key, and future scans. To align this with our single-scan, per-frame processing strategy, we exclusively extract the central key scan from each frame. Because we subsequently downsample the overall sequences to $0.2\,\mathrm{Hz}$ (one scan every 5 seconds), we naturally avoid any temporal overlap or redundancy between the 2-second snippets.

B. Qualitative results

B.1. Success cases

In this section, we present qualitative success cases of TerraSeg-B. Figures 1, 2, and 3 illustrate the segmentation performance on the nuScenes [3], SemanticKITTI [2], and Waymo Perception [10] validation splits, respectively. Each figure is arranged in two rows: the top row displays the ground truth annotations, while the bottom row shows the predictions generated by our domain-agnostic TerraSeg model. These results demonstrate TerraSeg's robustness in delineating complex ground geometries and adapting to varying terrain flatness across vastly different sensor configurations without relying on manual annotations.

B.2. Failure cases

While TerraSeg exhibits strong generalization capabilities, it occasionally encounters challenges in highly ambiguous geometric contexts. Figures 4, 5, and 6 highlight specific failure cases across the three primary benchmarks. As with

the success cases, the top row provides the ground truth, and the bottom row illustrates the model's predictions. The typical failure modes are the over-segmentation of dense, low-lying vegetation and atypical road debris that closely mimic the continuous geometric profile of the ground surface.

References

[1] Mina Alibeigi, William Ljungbergh, Adam Tonderski, Georg Hess, Adam Lilja, Carl Lindström, Daria Motorniuk, Junsheng Fu, Jenny Widahl, and Christoffer Petersson. Zenseact open dataset: A large-scale and diverse multimodal dataset for autonomous driving. In *International Conference on Computer Vision (ICCV)*, pages 20178–20188, 2023. 3

[2] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Juergen Gall. SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences. In *International Conference on Computer Vision (ICCV)*, pages 9297–9307, 2019. 1, 3, 4, 6

[3] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuScenes: A multimodal dataset for autonomous driving. In *Computer Vision and Pattern Recognition (CVPR)*, pages 11621–11631, 2020. 1, 3, 4, 5

[4] Felix Fent, Fabian Kuttenreich, Florian Ruch, Farija Rizwin, Stefan Juergens, Lorenz Lechermann, Christian Nissler, Andrea Perl, Ulrich Voll, Min Yan, et al. MANTruckScenes: A multimodal dataset for autonomous trucking in diverse conditions. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 62062–62082, 2024. 1, 3

[5] Ramesh Kesten, Muhammad Usman, John Houston, Tapan Pandya, Karthik Nadhamuni, André Ferreira, Ming Yuan, Brian Low, Ashesh Jain, Peter Ondruška, et al. Lyft Level 5 AV Dataset 2019, 2019. 1, 3

[6] Yiyi Liao, Jun Xie, and Andreas Geiger. KITTI-360: A Novel Dataset and Benchmarks for Urban Scene Understanding in 2D and 3D. *Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 45(3):3292–3310, 2022. 1, 3

[7] Jiageng Mao, Minzhe Niu, Chenhan Jiang, Hanxue Liang, Jingheng Chen, Xiaodan Liang, Yamin Li, Chaoqiang Ye, Wei Zhang, Zhenguo Li, Jie Yu, Hang Xu, and Chunjing Xu. One Million Scenes for Autonomous Driving: ONCE Dataset. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 3

[8] Gautham Narayan Narasimhan, Heethesh Vhavle, Kumar Bhargav Vishvanatha, and James Reuther. AevaScenes: A Dataset and Benchmark for FMCW LiDAR Perception, 2025. 1, 3

[9] Andras Palffy, Ewoud Pool, Srimannarayana Baratam, Julian FP Kooij, and Dariu M Gavrila. Multi-class Road User Detection with 3+1D Radar in the View-of-Delft Dataset. *Robotics and Automation Letters (RA-L)*, 7(2):4961–4968, 2022. 1, 3

[10] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Sheng Zhao, Shuyang Cheng, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in Perception for Autonomous Driving: Waymo Open Dataset. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2446–2454, 2020. 1, 3, 5, 6

[11] Benjamin Wilson, William Qi, Tanmay Agarwal, John Lambert, Jagjeet Singh, Siddhesh Khandelwal, Bowen Pan, Ratnesh Kumar, Andrew Hartnett, Jhony Kaesemodel Pontes, Deva Ramanan, Peter Carr, and James Hays. Argoverse 2: Next Generation Datasets for Self-Driving Perception and Forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 1, 3

[12] Pengchuan Xiao, Zhenlei Shao, Steven Hao, Zishuo Zhang, Xiaolin Chai, Judy Jiao, Zesong Li, Jian Wu, Kai Sun, Kun Jiang, Yunlong Wang, and Diange Yang. PandaSet: Advanced Sensor Suite Dataset for Autonomous Driving. In *International Intelligent Transportation Systems Conference (ITSC)*, pages 3095–3101, 2021. 1, 3

Table 1. **Hardware diversity in OmniLiDAR.** A detailed breakdown of the 15 unique LiDAR sensor models utilized across the datasets. The datasets cover a wide spectrum of sensor configurations, including classic spinning LiDARs, solid-state variants, and FMCW sensors, ensuring robust cross-sensor generalization during training.

Sensor Model	Configuration	Rate	Dataset(s)	Description
Velodyne HDL-64E	64 channels	10 Hz	SemanticKITTI [2], KITTI-360 [6]	Classic high-resolution spinning LiDAR with full 360° FoV.
Velodyne HDL-64 S3	64 channels	10 Hz	View-of-Delft [9]	Newer HDL-64 generation with improved optics and timing.
Velodyne HDL-32E	32 channels	20 Hz	nuScenes [3]	Compact spinning LiDAR with higher temporal resolution.
Velodyne VLP-32C	32 channels	10 Hz	Argoverse 2 Lidar [11]	Automotive-grade spinning LiDAR; dual-mounted in Argoverse 2.
Velodyne VLS-128	128 channels	10 Hz	Zenseact Open [1]	High-density long-range automotive spinning LiDAR.
Velodyne VLP-16	16 channels	10 Hz	Zenseact Open [1]	Short-range peripheral LiDARs for side and rear coverage.
Waymo Mid-Range (TOP)	64 channels	10 Hz	Waymo Perception [10]	Roof-mounted panoramic LiDAR developed by Waymo.
Waymo Short-Range	4-5 channels	10 Hz	Waymo Perception [10]	Near-field LiDARs (front, sides, rear) for close-range perception.
Hesai Pandar64	64 channels	10 Hz	PandaSet [12], TruckScenes [4]	Long-range automotive-grade spinning LiDAR.
Hesai PandarGT	64 channels	10 Hz	PandaSet [12]	Research-grade Pandar variant with higher angular accuracy.
Ouster OS0 Rev 7	64 channels	10 Hz	TruckScenes [4]	Wide-vertical-FoV LiDAR optimized for short-range perception.
Aeva Wide-FoV (FMCW)	110° FoV	10 Hz	AevaScenes [8]	FMCW LiDAR with wide angular coverage and velocity sensing.
Aeva Narrow-FoV (FMCW)	35° FoV	10 Hz	AevaScenes [8]	FMCW LiDAR optimized for long-range, high-precision sensing.
ONCE Proprietary	40 channels	10 Hz	ONCE [7]	Medium-resolution automotive spinning LiDAR.
Lyft Proprietary	Unknown	Unknown	Lyft Level 5 [5]	Undocumented hardware; conservatively counted as one unique model.

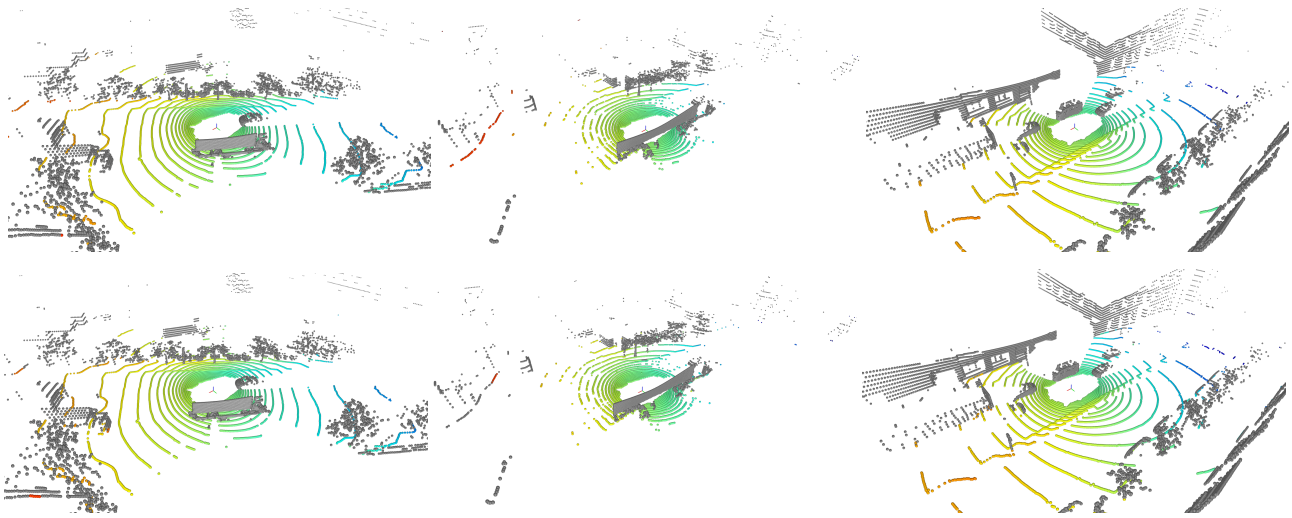

Figure 1. **Qualitative success cases on nuScenes dataset [3].** The top row shows ground truth labels, while the bottom row displays TerraSeg predictions. Ground points are color-coded by elevation; non-ground points are rendered in gray. From left to right, columns correspond to scan IDs 1787, 3238, and 3479.

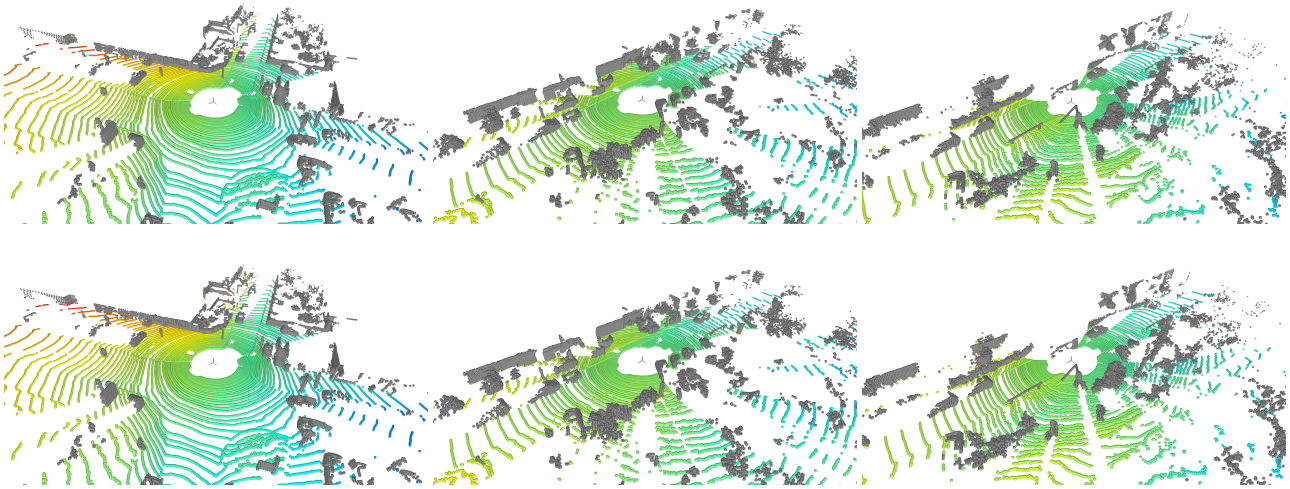

Figure 2. **Qualitative success cases on SemanticKITTI dataset [2].** The top row shows ground truth labels, while the bottom row displays TerraSeg predictions. Ground points are color-coded by elevation; non-ground points are rendered in gray. From left to right, columns correspond to scan IDs 4006, 2714, and 2816.

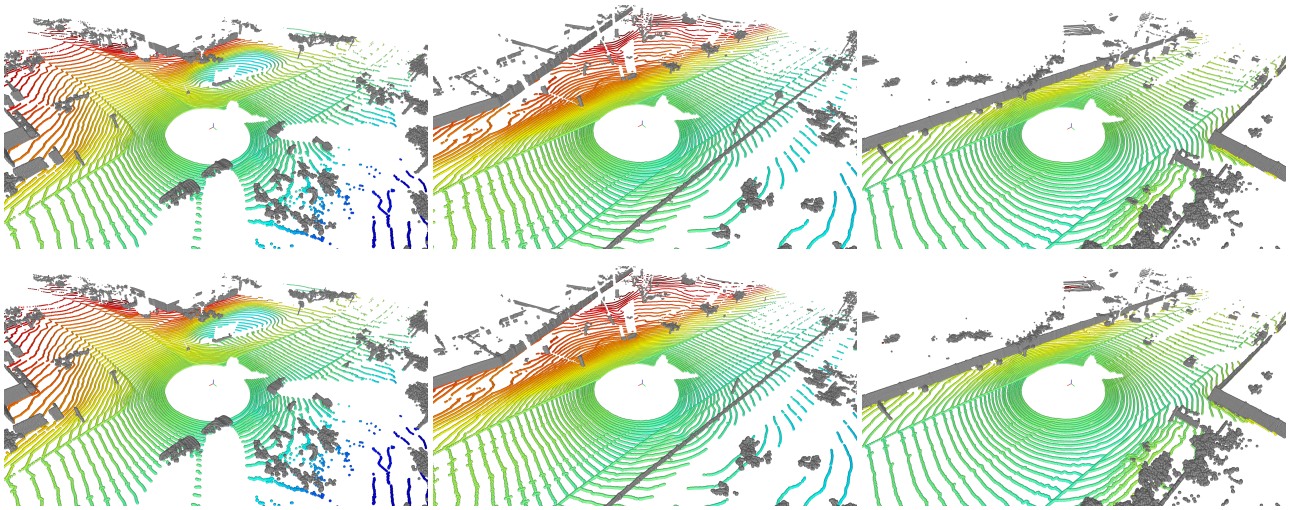

Figure 3. **Qualitative success cases on Waymo Perception dataset [10].** The top row shows ground truth labels, while the bottom row displays TerraSeg predictions. Ground points are color-coded by elevation; non-ground points are rendered in gray. From left to right, columns correspond to scan IDs 3659, 3045, and 3505.

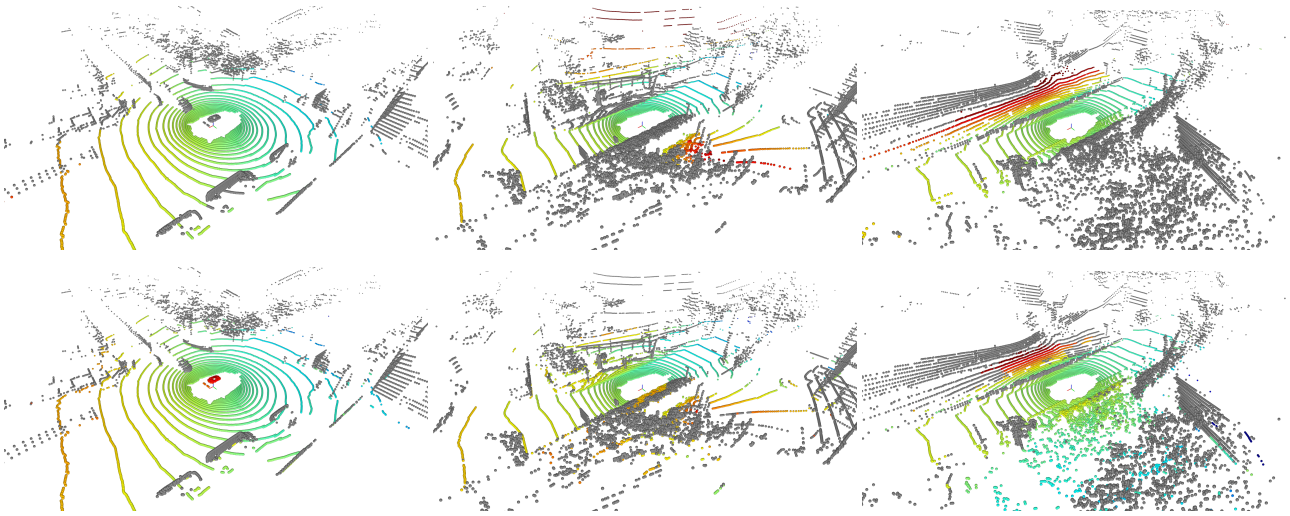

Figure 4. **Qualitative failure cases on nuScenes dataset [3].** The top row shows ground truth labels, while the bottom row displays TerraSeg predictions. Ground points are color-coded by elevation; non-ground points are rendered in gray. From left to right, columns correspond to scan IDs 2124, 4848, and 62.

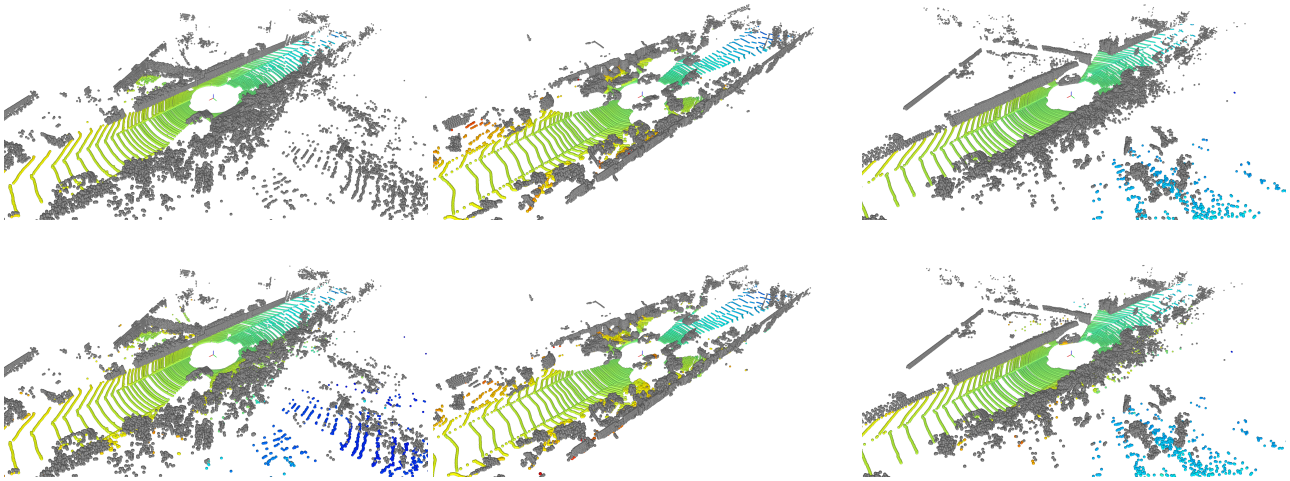

Figure 5. **Qualitative failure cases on SemanticKITTI dataset [2].** The top row shows ground truth labels, while the bottom row displays TerraSeg predictions. Ground points are color-coded by elevation; non-ground points are rendered in gray. From left to right, columns correspond to scan IDs 3300, 947, and 3173.

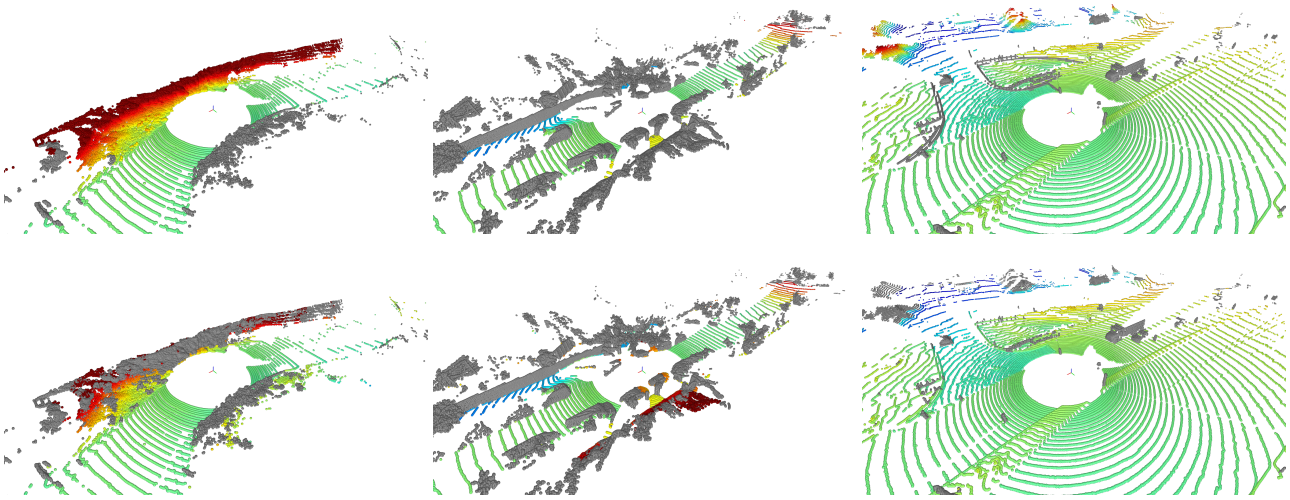

Figure 6. **Qualitative failure cases on Waymo Perception dataset [10].** The top row shows ground truth labels, while the bottom row displays TerraSeg predictions. Ground points are color-coded by elevation; non-ground points are rendered in gray. From left to right, columns correspond to scan IDs 2513, 4755, and 2024.