Robust 3D Object Detection under Sensor Placement Variability

Junlong Shen, Shuangzhi Li, and Xingyu Li
Department of Electrical and Computer Engineering, University of Alberta
Edmonton, Alberta, Canada
{junlong6, shuangzh, xingyu}@ualberta.ca

Abstract

Robust 3D object detection is crucial for autonomous driving, yet existing models often experience significant performance degradation when sensor placements differ across vehicles. This sensitivity limits the scalability of perception systems to real-world platforms with diverse configurations. To address this challenge, we present an enhanced LiDAR-only detection framework built upon the BEVFusion model. The proposed approach integrates three complementary strategies: (1) temporal sequence enhancement, which aggregates consecutive LiDAR sweeps to enrich spatial density and temporal continuity; (2) placement-mixed training, which exposes the model to diverse sensor configurations during training to improve cross-placement generalization; and (3) test-time augmentation (TTA), which performs multiple inference passes with geometric flipping and multi-sweep loading to stabilize predictions. These components are seamlessly incorporated into the BEVFusion LiDAR branch without requiring major architectural modifications. Experimental results demonstrate consistent improvements in detection accuracy and robustness under varying sensor placements, highlighting the potential of the proposed framework for practical deployment in autonomous driving.

1. Introduction

Accurate and reliable 3D object detection forms the cornerstone of perception in autonomous driving [1, 3, 9, 10, 16, 38]. LiDAR-based detectors have demonstrated remarkable performance due to their ability to capture dense and geometrically precise 3D structures [4, 15, 18, 22–24, 31, 34]. Nevertheless, most existing models are developed under the implicit assumption that LiDAR sensors share consistent placements and perfect calibration across vehicles [39, 41, 45]. This assumption is rarely satisfied in large-scale, real-world deployments [2, 32, 38].

In practice, sensor setups may differ because of variations in vehicle design, hardware layout, manufacturing tolerances, or retrofitting practices [6, 29, 30, 37]. Even small changes in LiDAR mounting height, orientation, or field of view can introduce substantial shifts in the captured geometry, leading to notable degradation in detection accuracy and cross-vehicle transferability [12, 27, 28].

This issue reveals a critical research gap. While extensive efforts have improved accuracy under standardized sensor configurations, the robustness of LiDAR-based detection under placement variability remains significantly underexplored [5, 25, 26, 28, 35, 36, 41]. In large-scale autonomous fleets, this lack of robustness poses practical challenges: a model optimized for one sensor layout may fail when deployed on another platform with differing extrinsics or coverage, requiring costly recalibration or retraining [33, 42]. Bridging this gap is therefore essential for realizing scalable perception systems that maintain consistent performance across heterogeneous platforms [3, 13, 14, 16, 30, 40, 43, 44]. To achieve this, a robust detector must compensate for partial observations, adapt to viewpoint-induced distortions, and remain resilient under both geometric and domain-level variations [27].

In this work, we propose an enhanced LiDAR-only detection framework built upon the BEVFusion [33] architecture, specifically designed to improve resilience against sensor placement variability. Our approach incorporates three complementary strategies that address temporal, spatial, and inference-level robustness. First, we introduce *temporal sequence enhancement*, which aggregates multiple consecutive LiDAR sweeps to capture motion continuity and enrich geometric context beyond single-frame observations. Second, we adopt *placement-mixed training*, where the model is exposed to LiDAR data from multiple sensor configurations during training, enabling it to learn geometry-invariant features that generalize across diverse placements [28]. Third, we employ *test-time augmentation (TTA)*, which performs multiple inference passes with geometric flips and multi-sweep fusion, followed by prediction aggregation to improve stability and confidence under unseen configurations. These strategies are seamlessly integrated into the BEVFusion LiDAR branch without altering its backbone design, ensuring compatibility and training efficiency.

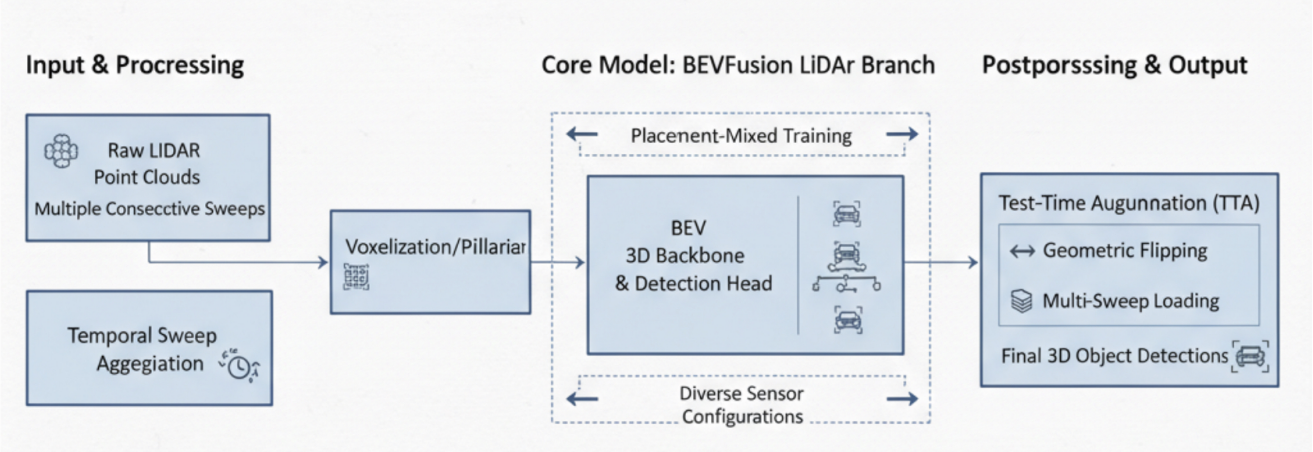

Figure 1. Towards Generalizable 3D Object Detection Across Sensor Placements Method.

Comprehensive experiments validate the effectiveness of the proposed framework. The model achieves consistent improvements in both detection accuracy and robustness across varying LiDAR placements, confirming that combining temporal cues, placement diversity, and inference-time augmentation substantially enhances cross-platform generalization. Beyond empirical gains, our findings highlight an important insight: improving perception robustness does not necessarily require new architectures but rather a principled integration of data-centric strategies that account for real-world variability in sensing conditions.

2. Methodology

2.1. Base Model: BEVFusion (LiDAR Branch)

Our approach is built upon the LiDAR branch of BEVFusion [33], a state-of-the-art framework for bird's-eye view (BEV) perception. BEVFusion projects 3D point clouds into a voxelized or pillar-based representation, which is then processed by a backbone network to extract spatial features. These features are subsequently aggregated into a BEV representation, enabling efficient reasoning over large-scale driving scenes.

Although BEVFusion [33] is typically designed for multi-sensor fusion (*e.g.*, LiDAR and camera), this work focuses exclusively on the LiDAR branch to establish a strong baseline for 3D object detection under variable sensor placements [27]. The baseline BEVFusion pipeline consists of three main stages:

- **Voxelization/Pillarization:** Raw LiDAR point clouds are discretized into fixed-size voxel or pillar grids for efficient processing.
- **Feature Extraction:** A 3D backbone network encodes the voxelized inputs to extract geometric and semantic features.
- **BEV Projection and Detection Head:** Extracted features are projected into BEV space and passed through a detection head to produce final 3D bounding box predictions.

This design enables BEVFusion [33] to fully leverage the geometric fidelity of LiDAR data while providing an efficient BEV representation for downstream detection tasks.

2.2. Temporal Sequence Enhancement

To address the sparsity and occlusion inherent in single-frame point clouds, we extend BEVFusion [33] with temporal information. Specifically, multiple consecutive LiDAR sweeps from the RoboSense 2025 Track 3 dataset [27] are aligned to a common reference frame before voxelization. By aggregating temporal point clouds, the model captures richer spatial and motion cues, improving feature density and robustness against transient occlusions. We adopt a sequence-based preprocessing stage that merges historical sweeps prior to voxelization, allowing the backbone to operate on temporally enriched inputs without modifying its original architecture. The resulting temporal features enhance object continuity across frames and reduce the impact of incomplete or noisy observations.

2.3. Placement-Mixed Training

To explicitly improve robustness to sensor placement variability, we introduce *placement-mixed training*. During training, point clouds collected from different sensor configurations are combined into a unified dataset split. This ensures that the model encounters a diverse range of viewpoints, densities, and extrinsic parameters, encouraging it to learn placement-invariant geometric features.

By simulating variability during training, the model becomes better equipped to handle unseen configurations during deployment, thereby improving generalization across heterogeneous LiDAR setups.

Table 1. Ablation study.

Method	Fixed Placement mAP	Variable Placement mAP
BEVFusion (LiDAR Baseline)	87.9	63.0
+ Temporal Enhancement	89.2	67.0
+ Placement-Mixed Training	86.6	69.5
+ TTA	88.3	66.6
Full Method	90.4	**74.6**

2.4. Test-Time Augmentation (TTA)

To further enhance robustness during inference, we employ a test-time augmentation (TTA) pipeline that performs multiple forward passes on augmented versions of the input point cloud. Two augmentation strategies are used:

- **Geometric Flipping:**
  Horizontal flips are applied to the input scene, and predictions are averaged to mitigate directional bias in detection results.
- **Multi-Sweep Loading:**
  Using the `LoadPointsFromMultiSweeps` module in MMDetection3D [8], additional historical sweeps are sampled at inference time to enrich spatial density and improve detection confidence.

The predictions from all augmented inputs are fused using non-maximum suppression (NMS) and averaging, providing smoother and more reliable final outputs.

2.5. Integrated Framework

The final framework integrates all three strategies within the BEVFusion [33] LiDAR branch. Temporal sweep aggregation enriches spatial inputs, placement-mixed training improves cross-placement generalization, and TTA stabilizes predictions during inference. These enhancements require minimal changes to the original BEVFusion [33] design, maintaining lightweight computation and ease of deployment.

Collectively, the proposed methodology improves detection accuracy and robustness under variable sensor placements, offering a practical and scalable solution for reliable 3D perception in autonomous driving.

3. Experiments

3.1. Dataset

We use the official data provided by the *RoboSense Challenge 2025* [21] held at IROS 2025. This competition builds upon the legacy of the *RoboDepth Challenge 2023* [17, 19] at ICRA 2023 and the *RoboDrive Challenge 2024* [20, 44] at ICRA 2024, continuing the collective effort to advance robust and scalable robot perception. Each track in this competition is grounded on an established benchmark designed for evaluating real-world robustness and generalization [7, 11, 27, 30, 43]. Specifically, this task is built upon the **Place3D** dataset [27] in **Track 3**, which provides a standardized foundation for benchmarking performance under challenging conditions such as cross-domain shifts, sensor variability, and multi-modal alignment.

3.2. Implementation Details

We implemented our method using the MMDetection3D [8] framework, adopting the official BEVFusion [33] LiDAR-only configuration as the baseline. All experiments were conducted with the official training settings provided by RoboSense 2025, including optimizer parameters, learning rate schedules, and data preprocessing pipelines. Our modifications were integrated seamlessly into this setup without altering the backbone architecture.

The model was trained on GPUs with mixed precision enabled to accelerate training and reduce memory consumption. Temporal sequence aggregation was implemented by aligning multiple consecutive sweeps before voxelization, while placement-mixed training was achieved by combining datasets from different sensor configurations into the training split. For inference, we incorporated our TTA pipeline with geometric flipping and multi-sweep loading.

3.3. Evaluation Protocol

We followed the official RoboSense Track3 evaluation metrics, focusing primarily on 3D mean Average Precision (mAP). Models were evaluated under both fixed and variable sensor placements to assess generalization capability.

3.4. Comparative Results & Ablation Study

We first compare our enhanced BEVFusion model against the baseline LiDAR-only BEVFusion configuration. As shown in Tab. 1, our method consistently outperformed the baseline under both fixed and altered sensor placements, demonstrating significant improvements in robustness. In particular, placement-mixed training contributed to stronger generalization, while temporal sequence enhancement improved accuracy in scenarios with sparse observations.

To quantify the individual contributions of each component, we performed ablation experiments:

- Temporal Sequence Only: Enhanced detection of small and partially occluded objects.

- Placement-Mixed Training Only: Improved robustness when transferring between sensor setups.
- TTA Only: Reduced variance in predictions, stabilizing results.
- Full Method: Combined benefits of all strategies, yielding the highest overall performance.

The ablation confirms that each component provides complementary benefits, with the combination delivering the most robust detection performance.

4. Conclusion

In this work, we presented a LiDAR-only 3D object detection framework designed to enhance robustness under variable sensor placements. By integrating temporal sequence enhancement, placement-mixed training, and test-time augmentation into the BEVFusion LiDAR branch, our method addresses both data sparsity and sensor variability challenges. Experimental results demonstrate that each component contributes positively to overall performance, with the combined approach achieving the highest accuracy and stability across different sensor configurations. The proposed strategies require minimal architectural modifications, making them practical for real-world deployment. Overall, our approach highlights the importance of leveraging temporal information, diverse training exposure, and inference-time augmentation to build scalable and reliable autonomous driving perception systems.

References

[1] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences. In *IEEE/CVF International Conference on Computer Vision*, pages 9297–9307, 2019.

[2] Hengwei Bian et al. Dynamiccity: Large-scale 4D occupancy generation from dynamic scenes. In *International Conference on Learning Representations*, 2025.

[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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11621–11631, 2020.

[4] Li Chen, Penghao Wu, Kashyap Chitta, Bernhard Jaeger, Andreas Geiger, and Hongyang Li. End-to-end autonomous driving: Challenges and frontiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[5] Runnan Chen et al. Towards label-free scene understanding by vision foundation models. In *Advances in Neural Information Processing Systems*, pages 75896–75910, 2023.

[6] Yukang Chen, Jianhui Liu, Xiangyu Zhang, Xiaojuan Qi, and Jiaya Jia. Voxelnext: Fully sparse voxelnet for 3d object detection and tracking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 21674–21683, 2023.

[7] Meng Chu, Zhedong Zheng, Wei Ji, Tingyu Wang, and Tat-Seng Chua. Towards natural language-guided drones: GeoText-1652 benchmark with spatial relation matching. In *European Conference on Computer Vision*, pages 213–231, 2024.

[8] MMDetection3D Contributors. MMDetection3D: OpenMMLab next-generation platform for general 3D object detection. https://github.com/open-mmlab/mmdetection3d, 2020.

[9] Whye Kit Fong, Rohit Mohan, Juana Valeria Hurtado, Lubing Zhou, Holger Caesar, Oscar Beijbom, and Abhinav Valada. Panoptic nuScenes: A large-scale benchmark for LiDAR panoptic segmentation and tracking. *IEEE Robotics and Automation Letters*, 7(2):3795–3802, 2022.

[10] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012.

[11] Zeying Gong, Tianshuai Hu, Ronghe Qiu, and Junwei Liang. From cognition to precognition: A future-aware framework for social navigation. In *IEEE International Conference on Robotics and Automation*, pages 9122–9129, 2025.

[12] Xiaoshuai Hao, Mengchuan Wei, Yifan Yang, et al. Is your HD map constructor reliable under sensor corruptions? In *Advances in Neural Information Processing Systems*, pages 22441–22482, 2024.

[13] Xiaoshuai Hao, Guanqun Liu, Yuting Zhao, et al. Msc-bench: Benchmarking and analyzing multi-sensor corruption for driving perception. *arXiv preprint arXiv:2501.01037*, 2025.

[14] Xiaoshuai Hao et al. Safemap: Robust HD map construction from incomplete observations. In *International Conference on Machine Learning*, pages 22091–22102. PMLR, 2025.

[15] Lingdong Kong, Youquan Liu, Runnan Chen, Yuexin Ma, Xinge Zhu, Yikang Li, Yuenan Hou, Yu Qiao, and Ziwei Liu. Rethinking range view representation for LiDAR segmentation. In *IEEE/CVF International Conference on Computer Vision*, pages 228–240, 2023.

[16] Lingdong Kong, Youquan Liu, Xin Li, Runnan Chen, Wenwei Zhang, Jiawei Ren, Liang Pan, Kai Chen, and Ziwei Liu. Robo3D: Towards robust and reliable 3D perception against corruptions. In *IEEE/CVF International Conference on Computer Vision*, pages 19994–20006, 2023.

[17] Lingdong Kong, Yaru Niu, Shaoyuan Xie, Hanjiang Hu, Lai Xing Ng, Benoit Cottereau, Liangjun Zhang, Hesheng Wang, Wei Tsang Ooi, Ruijie Zhu, Ziyang Song, Li Liu, Tianzhu Zhang, Jun Yu, Mohan Jing, Pengwei Li, Xiaohua Qi, Cheng Jin, Yingfeng Chen, Jie Hou, Jie Zhang, Zhen Kan, Qiang Lin, Liang Peng, Minglei Li, Di Xu, Changpeng Yang, Yuanqi Yao, Gang Wu, Jian Kuai, Xianming Liu, Junjun Jiang, Jiamian Huang, Baojun Li, Jiale Chen, Shuang Zhang, Sun Ao, Zhenyu Li, Runze Chen, Haiyong Luo, Fang Zhao, and Jingze Yu. The RoboDepth challenge: Methods and advancements towards robust depth estimation. *arXiv preprint arXiv:2307.15061*, 2023.

[18] Lingdong Kong, Jiawei Ren, Liang Pan, and Ziwei Liu. Lasermix for semi-supervised LiDAR semantic segmentation.

In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21705–21715, 2023.

[19] Lingdong Kong, Shaoyuan Xie, Hanjiang Hu, Lai Xing Ng, Benoit R. Cottereau, and Wei Tsang Ooi. RoboDepth: Robust out-of-distribution depth estimation under corruptions. In *Advances in Neural Information Processing Systems*, pages 21298–21342, 2023.

[20] Lingdong Kong, Shaoyuan Xie, Hanjiang Hu, Yaru Niu, Wei Tsang Ooi, Benoit R. Cottereau, Lai Xing Ng, Yuexin Ma, Wenwei Zhang, Liang Pan, Kai Chen, Ziwei Liu, Weichao Qiu, Wei Zhang, Xu Cao, Hao Lu, Ying-Cong Chen, Caixin Kang, Xinning Zhou, Chengyang Ying, Wentao Shang, Xingxing Wei, Yinpeng Dong, Bo Yang, Shengyin Jiang, Zeliang Ma, Dengyi Ji, Haiwen Li, Xingliang Huang, Yu Tian, Genghua Kou, Fan Jia, Yingfei Liu, Tiancai Wang, Ying Li, Xiaoshuai Hao, Yifan Yang, Hui Zhang, Mengchuan Wei, Yi Zhou, Haimei Zhao, Jing Zhang, Jinke Li, Xiao He, Xiaoqiang Cheng, Bingyang Zhang, Lirong Zhao, Dianlei Ding, Fangsheng Liu, Yixiang Yan, Hongming Wang, Nanfei Ye, Lun Luo, Yubo Tian, Yiwei Zuo, Zhe Cao, Yi Ren, Yunfan Li, Wenjie Liu, Xun Wu, Yifan Mao, Ming Li, Jian Liu, Jiayang Liu, Zihan Qin, Cunxi Chu, Jialei Xu, Wenbo Zhao, Junjun Jiang, Xianming Liu, Ziyan Wang, Chiwei Li, Shilong Li, Chendong Yuan, Songyue Yang, Wentao Liu, Peng Chen, Bin Zhou, Yubo Wang, Chi Zhang, Jianhang Sun, Hai Chen, Xiao Yang, Lizhong Wang, Dongyi Fu, Yongchun Lin, Huitong Yang, Haoang Li, Yadan Luo, Xianjing Cheng, and Yong Xu. The RoboDrive challenge: Drive anytime anywhere in any condition. *arXiv preprint arXiv:2405.08816*, 2024.

[21] Lingdong Kong, Shaoyuan Xie, Zeying Gong, Ye Li, Meng Chu, Ao Liang, Yuhao Dong, Tianshuai Hu, Ronghe Qiu, Rong Li, Hanjiang Hu, Dongyue Lu, Wei Yin, Wenhao Ding, Linfeng Li, Hang Song, Wenwei Zhang, Yuexin Ma, Junwei Liang, Zhedong Zheng, Lai Xing Ng, Benoit R. Cottereau, Wei Tsang Ooi, Ziwei Liu, Zhanpeng Zhang, Weichao Qiu, Wei Zhang, Ji Ao, Jiangpeng Zheng, Siyu Wang, Guang Yang, Zihao Zhang, Yu Zhong, Enzhu Gao, Xinhan Zheng, Xueting Wang, Shouming Li, Yunkai Gao, Siming Lan, Mingfei Han, Xing Hu, Dusan Malic, Christian Fruhwirth-Reisinger, Alexander Prutsch, Wei Lin, Samuel Schulter, Horst Possegger, Linfeng Li, Jian Zhao, Zepeng Yang, Yuhang Song, Bojun Lin, Tianle Zhang, Yuchen Yuan, Chi Zhang, Xuelong Li, Youngseok Kim, Sihwan Hwang, Hyeonjun Jeong, Aodi Wu, Xubo Luo, Erjia Xiao, Lingfeng Zhang, Yingbo Tang, Hao Cheng, Renjing Xu, Wenbo Ding, Lei Zhou, Long Chen, Hangjun Ye, Xiaoshuai Hao, Shuangzhi Li, Junlong Shen, Xingyu Li, Hao Ruan, Jinliang Lin, Zhiming Luo, Yu Zang, Cheng Wang, Hanshi Wang, Xijie Gong, Yixiang Yang, Qianli Ma, Zhipeng Zhang, Wenxiang Shi, Jingmeng Zhou, Weijun Zeng, Kexin Xu, Yuchen Zhang, Haoxiang Fu, Ruibin Hu, Yanbiao Ma, Xiyan Feng, Wenbo Zhang, Lu Zhang, Yunzhi Zhuge, Huchuan Lu, You He, Seungjun Yu, Junsung Park, Youngsun Lim, Hyunjung Shim, Faduo Liang, Zihang Wang, Yiming Peng, Guanyu Zong, Xu Li, Binghao Wang, Hao Wei, Yongxin Ma, Yunke Shi, Shuaipeng Liu, Dong Kong, Yongchun Lin, Huitong Yang, Liang Lei, Haoang Li, Xinliang Zhang, Zhiyong Wang, Xiaofeng Wang, Yuxia Fu, Yadan Luo, Djamahl Etchegaray, Yang Li, Congfei Li, Yuxiang Sun, Wenkai Zhu, Wang Xu, Linru Li, Longjie Liao, Jun Yan, Benwu Wang, Xueliang Ren, Xiaoyu Yue, Jixian Zheng, Jinfeng Wu, Shurui Qin, Wei Cong, and Yao He. The RoboSense challenge: Sense anything, navigate anywhere, adapt across platforms. https://robosense2025.github.io, 2025.

[22] Lingdong Kong, Xiang Xu, Youquan Liu, Jun Cen, Runnan Chen, Wenwei Zhang, Liang Pan, Kai Chen, and Ziwei Liu. LargeAD: Large-scale cross-sensor data pretraining for autonomous driving. *arXiv preprint arXiv:2501.04005*, 2025.

[23] Lingdong Kong, Xiang Xu, Jiawei Ren, et al. Multi-modal data-efficient 3D scene understanding for autonomous driving. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(5):3748–3765, 2025.

[24] Lingdong Kong, Wesley Yang, Jianbiao Mei, Youquan Liu, Ao Liang, Dekai Zhu, Dongyue Lu, Wei Yin, Xiaotao Hu, Mingkai Jia, Junyuan Deng, Kaiwen Zhang, Yang Wu, Tianyi Yan, Shenyuan Gao, Song Wang, Linfeng Li, Liang Pan, Yong Liu, Jianke Zhu, Wei Tsang Ooi, Steven C. H. Hoi, and Ziwei Liu. 3D and 4D world modeling: A survey. *arXiv preprint arXiv:2509.07996*, 2025.

[25] Rong Li, Yuhao Dong, Tianshuai Hu, Ao Liang, et al. 3eed: Ground everything everywhere in 3D. *arXiv preprint arXiv:2511.01755*, 2025.

[26] Rong Li et al. Seeground: See and ground for zero-shot open-vocabulary 3D visual grounding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3707–3717, 2025.

[27] Ye Li, Lingdong Kong, Hanjiang Hu, Xiaohao Xu, and Xiaonan Huang. Is your LiDAR placement optimized for 3D scene understanding? In *Advances in Neural Information Processing Systems*, pages 34980–35017, 2024.

[28] Ye Li et al. Optimizing LiDAR placements for robust driving perception in adverse conditions. *arXiv preprint arXiv:2403.17009*, 2024.

[29] Ao Liang et al. LiDARCrafter: Dynamic 4D world modeling from LiDAR sequences. *arXiv preprint arXiv:2508.03692*, 2025.

[30] Ao Liang et al. Perspective-invariant 3D object detection. In *IEEE/CVF International Conference on Computer Vision*, pages 27725–27738, 2025.

[31] Youquan Liu et al. Segment any point cloud sequences by distilling vision foundation models. In *Advances in Neural Information Processing Systems*, pages 37193–37229, 2023.

[32] Youquan Liu et al. Uniseg: A unified multi-modal LiDAR segmentation network and the openpcseg codebase. In *IEEE/CVF International Conference on Computer Vision*, pages 21662–21673, 2023.

[33] Zhijian Liu, Haotian Tang, Alexander Amini, Xinyu Yang, Huizi Mao, Daniela Rus, and Song Han. Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation. *arXiv preprint arXiv:2205.13542*, 2022.

[34] Andres Milioto, Ignacio Vizzo, Jens Behley, and Cyrill Stachniss. Rangenet++: Fast and accurate LiDAR semantic segmentation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4213–4220, 2019.

[35] Lucas Nunes, Rodrigo Marcuzzi, Xieyuanli Chen, Jens Behley, and Cyrill Stachniss. SegContrast: 3D Point Cloud

Feature Representation Learning through Self-supervised Segment Discrimination. *IEEE Robotics and Automation Letters*, 7(2):2116–2123, 2022.

[36] David Schinagl, Georg Krispel, Horst Possegger, Peter M. Roth, and Horst Bischof. OccAM's Laser: Occlusion-Based Attribution Maps for 3D Object Detectors on LiDAR Data. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1141–1150, 2022.

[37] Jiahao Sun, Chunmei Qing, Xiang Xu, et al. An empirical study of training state-of-the-art LiDAR segmentation models. *arXiv preprint arXiv:2405.14870*, 2024.

[38] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2446–2454, 2020.

[39] Li Wang, Xinyu Zhang, Ziying Song, Jiangfeng Bi, Guoxin Zhang, Haiyue Wei, Liyao Tang, Lei Yang, Jun Li, Caiyan Jia, et al. Multi-modal 3d object detection in autonomous driving: A survey and taxonomy. *IEEE Transactions on Intelligent Vehicles*, 8(7):3781–3798, 2023.

[40] Xuzhi Wang, Xinran Wu, Song Wang, et al. Monocular semantic scene completion via masked recurrent networks. In *IEEE/CVF International Conference on Computer Vision*, pages 24811–24822, 2025.

[41] Xuzhi Wang et al. Nuc-net: Non-uniform cylindrical partition network for efficient LiDAR semantic segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(9):9090–9104, 2025.

[42] Shaoyuan Xie, Lingdong Kong, Wenwei Zhang, Jiawei Ren, Liang Pan, Kai Chen, and Ziwei Liu. RoboBEV: Towards robust bird's eye view perception under corruptions. *arXiv preprint arXiv:2304.06719*, 2023.

[43] Shaoyuan Xie, Lingdong Kong, Yuhao Dong, Chonghao Sima, Wenwei Zhang, Qi Alfred Chen, Ziwei Liu, and Liang Pan. Are VLMs ready for autonomous driving? an empirical study from the reliability, data, and metric perspectives. In *IEEE/CVF International Conference on Computer Vision*, pages 6585–6597, 2025.

[44] Shaoyuan Xie, Lingdong Kong, Wenwei Zhang, Jiawei Ren, Liang Pan, Kai Chen, and Ziwei Liu. Benchmarking and improving bird's eye view perception robustness in autonomous driving. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(5):3878–3894, 2025.

[45] Xiang Xu et al. Frnet: Frustum-range networks for scalable LiDAR segmentation. *IEEE Transactions on Image Processing*, 34:2173–2186, 2025.