Towards Cross-Platform Generalization: Domain Adaptive 3D Detection with Augmentation and Pseudo-Labeling

Xiyan Feng[1], Wenbo Zhang[1], Lu Zhang[1], Yunzhi Zhuge[1], Huchuan Lu[1], You He[2]
[1]Dalian University of Technology
[2]Shenzhen International Graduate School, Tsinghua University

{fxy,zwbo}@mail.dlut.edu.cn, {zhangluu,zgyz,lhchuan}@dlut.edu.cn, heyou@tsinghua.edu.cn

Abstract

This technical report presents the award-winning solution to the Cross-Platform 3D Object Detection track of the RoboSense 2025 Challenge. Our method is built upon PVRCNN++, an efficient and strong 3D object detection framework that combines point-based and voxel-based representations for robust geometric perception. To enhance cross-platform generalization, we introduce two key improvements. First, we design tailored data augmentation strategies that explicitly account for domain discrepancies arising from heterogeneous LiDAR viewpoints and scanning heights. Second, we incorporate a self-training pipeline with pseudo-label refinement, enabling the model to better adapt to unlabeled target domains and progressively reduce the distribution gap. With these enhancements, our approach demonstrated strong robustness across diverse platforms and achieved competitive results in the challenge. Specifically, it attained a 3D AP of 62.67% for the Car category on the Phase 1 target domain, and 58.76% and 49.81% for the Car and Pedestrian categories, respectively, on the Phase 2 target domain. These results underscore the effectiveness of our strategy in addressing cross-platform domain shifts and improving real-world deployability of LiDAR-based detectors.

1. Introduction

With the rapid development of autonomous driving technologies, LiDAR-based 3D object detection has emerged as a core perception task, supporting downstream modules such as localization, motion forecasting, path planning, and obstacle avoidance [1–11].

Modern detectors leverage a variety of feature representations, including point-based networks [12–16], voxel-based architectures [17–19], and hybrid point–voxel frameworks [20–24]. These models have achieved impressive results on large-scale benchmarks such as KITTI [25], nuScenes [26, 27], and Waymo [28], demonstrating strong geometric reasoning capabilities under well-calibrated, fixed LiDAR setups [29–35].

Recently, the application scope of 3D object detection has expanded beyond traditional automotive platforms to include drones, quadruped robots, and other emerging robotic agents [36–39]. However, detectors trained exclusively on vehicle-mounted LiDAR data tend to generalize poorly when deployed on new platforms due to substantial cross-platform domain gaps [39, 40]. These gaps arise from differences in platform motion patterns, sensor mounting positions, and operational environments [9, 41–48]. For example, drones operate at elevated viewpoints with downward-looking LiDAR scans, while quadruped robots capture data from lower angles and exhibit frequent orientation changes. Such variations lead to pronounced shifts in point density, geometric coverage, and spatial distributions [49], causing severe performance degradation when source-trained detectors are applied to unseen target platforms [22, 50–54].

Addressing cross-platform domain shifts is therefore essential for enabling scalable and reliable deployment of 3D perception systems across heterogeneous robots [37, 55]. To this end, we develop a cross-platform adaptation framework that extends the powerful PVRCNN++ detector [22] with two targeted modules. First, we introduce the *Cross-Platform Jitter Alignment (CJA)* augmentation strategy, which simulates platform-specific viewpoint perturbations during training. By modeling geometric jitter consistent with motion and height variations across platforms, CJA narrows the distribution gap and promotes the learning of viewpoint-invariant features. Second, we incorporate the ST3D self-training paradigm [56] to generate high-quality pseudo-labels on unlabeled target-domain scans. Through iterative refinement of pseudo-labels and detector predictions, ST3D enables progressive adaptation and improves target-domain robustness without additional annotations.

Together, these components preserve the strong baseline performance of PVRCNN++ while significantly enhanc-

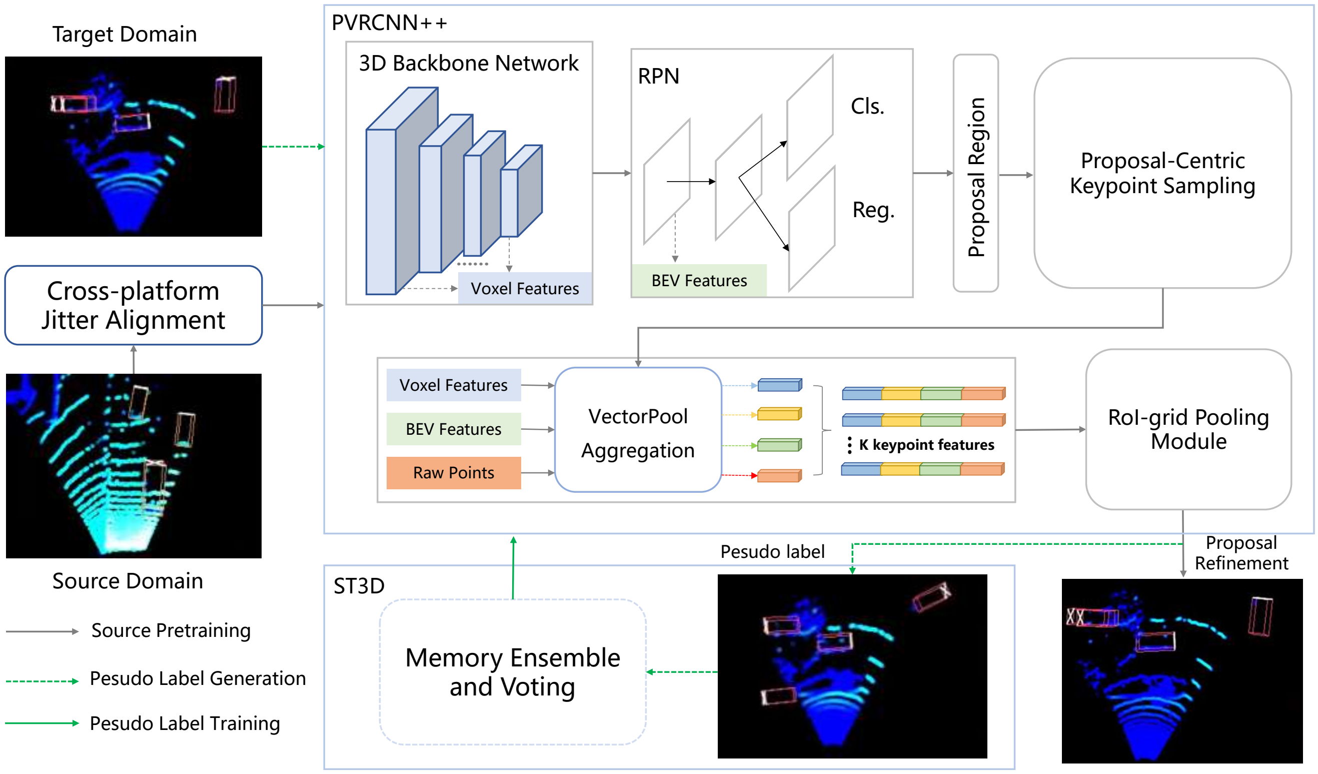

Figure 1. Overall pipeline of the proposed Cross-Platform 3D Detector. Source domain point clouds are processed by PVRCNN++ to produce detection results with labels, while target domain data generates pseudo-labels through the same detector. The ST3D module refines these pseudo-labels, and the CJA module operates exclusively on the source domain to enhance robustness by simulating platform-specific viewpoint jitter. Both source labels and refined target pseudo-labels are then utilized to iteratively update the model, enabling progressive adaptation across platforms without additional annotations.

ing its generalization ability across vehicles, drones, and quadrupeds. The proposed solution offers a practical and effective pathway toward robust cross-platform LiDAR-based 3D object detection in real-world robotic applications.

2. Methodology

2.1. Cross-domain 3D object detection paradigm

3D Detector. We adopt an advanced two-stage 3D object detection framework based on PVRCNN++, which seamlessly integrates the advantages of both point-based and voxel-based detection paradigms. This framework achieves efficient feature extraction from 3D scenes through voxelization while preserving the fine-grained geometric information inherent in raw point clouds, thereby striking a balance between accuracy and efficiency.

In the first stage, PVRCNN++ transforms irregular point cloud data into structured voxel grids via Voxel Feature Encoding. A 3D sparse convolutional network then extracts multi-scale voxel features, which are used to generate high-quality proposals with preliminary classification and bounding box regression. We observed that the CenterHead, used for proposal generation, relies on heatmap peaks from the BEV feature map to localize bounding box centers. However, under large viewing angles, the point cloud density of small objects deteriorates significantly, causing heatmap peaks to frequently localize to ground surfaces or empty voxels. This phenomenon leads to center drift and degraded recall rates. To combat this issue, we replace the CenterHead with an AnchorHead that incorporates predefined anchors of diverse scales and orientations. This modification provides enhanced geometric priors to the model, significantly improving proposal quality while ensuring greater stability and efficiency during training.

In the second stage, representative keypoints are first extracted from the global scene, and the Voxel Set Abstraction (VSA) module aggregates rich contextual voxel features onto these keypoints. RoI-grid pooling is subsequently applied to perform structured grid point sampling within each 3D proposal region, where high-quality VSA features are aggregated. Finally, a proposal refinement network processes these features to produce refined bounding box predictions.

Unsupervised Domain Adaptation. Cross-platform point clouds exhibit significant geometric distribution disparities due to viewpoint variations. For instance, different mobile platforms may introduce distinct LiDAR vibration patterns during carrier motion, and substantial variations in platform elevation lead to completely different absolute coordinates for identical objects across platforms. Consequently, models trained solely on the source domain often suffer severe performance degradation or even failure when directly applied to target domains.

To address this challenge, we employ ST3D, an unsupervised domain adaptation method built on a self-training strategy. ST3D iteratively generates pseudo-labels for unlabeled target domain data and uses them to fine-tune the model pre-trained on the source domain. This iterative process encourages the model to learn domain-invariant feature representations, thereby enhancing robustness across heterogeneous sensor setups and environments.

2.2. Data Augmentation

In cross-platform 3D detection tasks, different mobile platforms induce distinct jitter characteristics in their mounted LiDAR sensors due to variations in motion patterns and mechanical structures. The source domain vehicle platform typically operates on flat road surfaces, where changes in pitch and roll angles are relatively small. As a result, the model trained on such data learns feature representations constrained to a narrow range of viewpoints and motion priors, limiting its ability to generalize to novel motion patterns of target platforms.

Inspired by [37], we implement the Cross-platform Jitter Alignment (CJA) augmentation technique to explicitly compensate for jitter distribution discrepancies across platforms. By introducing controlled pose perturbations during source domain pre-training, CJA encourages the detector to learn feature representations invariant to platform-specific jitter, thereby enhancing the model's cross-platform generalization capability. Concretely, for each training sample, we uniformly sample pitch increments $\Delta\theta$ and roll increments $\Delta\phi$ from a predefined jitter angle range. A composite rotation matrix $\mathbf{R}(\Delta\phi, \Delta\theta)$ is then applied to the entire point cloud scene.

To preserve annotation consistency, the centers of all bounding boxes undergo the same transformation. The dimensions and orientation angles of the boxes remain fixed, while only their spatial positions are adjusted to align with the rotated point cloud. This process maintains geometric consistency of the annotations while effectively simulating viewpoint changes induced by platform motion in real-world LiDAR data.

2.3. Training framework

We adopt a systematic two-stage training framework to address domain adaptation in cross-platform 3D detection. This framework first establishes a strong foundational detection capability on the source domain, followed by fine-tuning on the target domain using a self-training strategy. This phased design ensures good initial performance and facilitates effective adaptation to the data distribution of the target domain.

In stage 1, the objective is to build a robust baseline detector that provides high-quality initial weights for subsequent domain adaptation. To achieve this objective, we train the PVRCNN++ detector on the source domain data using a standard supervised learning paradigm, which facilitates the learning of discriminative features from annotated LiDAR point clouds.

In stage 2, building upon the pre-trained weights from the first stage, we employ the ST3D self-training method for domain adaptation fine-tuning on unlabeled target domain data. Through an iterative pseudo-label optimization process, the model gradually reduces the inter-domain discrepancy, enabling it to adapt to the data distribution characteristics of the target domain.We adopt differentiated threshold design to ensure that high-quality pseudo-labels can be generated in all categories.

3. Experiments

3.1. Dataset

We use the official data provided by the *RoboSense Challenge 2025* [58] held at IROS 2025. This competition builds upon the legacy of the *RoboDepth Challenge 2023* [59, 60] at ICRA 2023 and the *RoboDrive Challenge 2024* [61, 62] 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 [37, 41, 63–65]. Specifically, this task is built upon the **Pi3DET** benchmark [37] in **Track 5**, which studies cross-platform LiDAR-based 3D object detection across vehicle, drone, and quadruped platforms through viewpoint normalization and unified pre-training.

3.2. Experimental Setup

Our experiments were conducted on the Track5 dataset, following the official challenge protocol for data preparation. The dataset consists of three subsets: source domain data, Phase 1 target domain data, and Phase 2 target domain data. These subsets contain point cloud-image pairs collected by LiDAR and camera systems mounted on vehicles, drones, and quadrupeds, respectively. Point cloud annotations are available exclusively for the source domain. The training consists of two stages: model pre-training is performed solely on the source domain data; the subsequent self-training stage generates pseudo-labels on the target domain data and uses them for training, with evaluation conducted on the target domain sets.

Detector	CJA	ST3D	**Car AP@0.5**	Car AP@0.7	**Ped. AP@0.5**	Ped. AP@0.25	Score
PointRCNN[12]	-	-	46.29	25.71	41.17	44.74	43.73
	✓	-	46.04	26.83	40.97	44.90	43.50
	✓	✓	46.46	26.45	27.92	31.34	37.19
VoxelRCNN[17]	-	-	26.95	10.88	28.44	33.24	27.70
	✓	-	40.52	22.38	45.18	49.34	42.85
	✓	✓	45.43	26.09	48.03	51.95	46.73
PDV[57]	-	-	26.12	11.17	26.27	30.27	26.20
	✓	-	43.39	24.31	46.11	50.84	44.75
	✓	✓	44.53	25.23	47.01	51.86	46.28
PVRCNN++[22]	-	-	29.44	9.98	14.94	18.16	22.19
	✓	-	43.94	23.96	46.83	52.28	45.39
	✓	✓	54.72	29.41	48.25	54.76	51.48
PVRCNN++*	✓	✓	**58.79**	**30.89**	**49.81**	**55.27**	**54.29**

Table 1. Performance comparison of different 3D detection frameworks under cross-platform adaptation from vehicle to quadruped robot platforms. Symbol * denotes replacing the RPNHead with AnchorHead. All scores are given in percentage (%). We report the 3D Average Precision (AP) for Cars at IoU thresholds of 0.5 and 0.7, and for Pedestrians at thresholds of 0.25 and 0.5.

3.3. Implementation Details

Our model is implemented based on the OpenPCDet codebase[66] using PyTorch. Both the pre-training and self-training stages are conducted on 4 NVIDIA GTX 4090 GPUs, with a batch size of 4 per GPU. We adopt the AdamW optimizer with OneCycle learning rate policy. The initial learning rate is set to 0.01 for pre-training and 1.5×10^{-3} for self-training . The self-training stage runs for 5 epochs, with pseudo-labels updated every 4 epochs. For Phase 1 data, the confidence threshold is set to 0.7 and the negative sample threshold to 0.2. For Phase 2 data, the confidence thresholds are 0.85 for the Car class and 0.55 for the Pedestrian class, with a uniform negative sample threshold of 0.20.

3.4. Ablation Study

We conduct comprehensive ablation studies to evaluate the effectiveness of our proposed model and its compatibility with different detector architectures. As shown in Table 1, the baseline PVRCNN++ achieves 29.44% Car AP@0.5 and 14.94% Pedestrian AP@0.5. Incorporating CJA augmentation improves these metrics to 43.94% and 46.83% respectively, demonstrating its effectiveness in enhancing the model's robustness against platform-specific jitter through geometric alignment. Further combining with ST3D self-training elevates performance to 54.72% Car AP@0.5 and 48.25% Pedestrian AP@0.5, reflecting its capability to progressively adapt the model to target domain characteristics through iterative pseudo-label refinement. Finally, replacing the original RPN head with our AnchorHead modification establishes the optimal model, reaching 58.79% Car AP@0.5 and 49.81% Pedestrian AP@0.5, demonstrating the critical role of geometric priors in enhancing proposal quality for cross-platform detection.

Beyond evaluating our primary framework, we further analyze the generalizability of CJA and ST3D across different types of detectors. The results show that both components significantly benefit voxel-based and point-voxel hybrid methods while showing limited effectiveness on pure point-based architectures. Specifically, both VoxelRCNN and PDV demonstrate substantial performance gains through the application of CJA and ST3D modules, with approximately 20% improvement in Car AP@0.5 and around 18% gain in Pedestrian AP@0.5. In contrast, PointRCNN remains largely unaffected by CJA and even experiences performance degradation in pedestrian detection when applying ST3D. This architectural sensitivity indicates that our approach is particularly suitable for detectors utilizing voxel representations.

4. Conclusion

In order to enhance cross-platform detection accuracy, we improved our competition framework by implementing a cross-platform 3D detection system built upon the PVRCNN++ detector pre-trained on source domain data. This framework incorporates the CJA data augmentation technique to explicitly mitigate geometric distribution discrepancies across platforms, and is further enhanced by the ST3D self-training paradigm that generates high-quality pseudo-labels for effective domain adaptation. Experimental results demonstrate that our improvements achieve remarkable performance gains in cross-platform scenarios.

References

[1] Jiageng Mao, Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. 3d object detection for autonomous driving: A review and new outlooks. *arXiv preprint arXiv:2206.09474*, 1(1):1, 2022. 1

[2] Duarte Fernandes, António Silva, Rafael Névoa, Cláudia Simões, Dibet Gonzalez, Miguel Guevara, Paulo Novais, João Monteiro, and Pedro Melo-Pinto. Point-cloud based 3d object detection and classification methods for self-driving applications: A survey and taxonomy. *Information Fusion*, 68:161–191, 2021.

[3] Georgios Zamanakos, Lazaros Tsochatzidis, Angelos Amanatiadis, and Ioannis Pratikakis. A comprehensive survey of lidar-based 3d object detection methods with deep learning for autonomous driving. *Computers & Graphics*, page 153–181, Oct 2021.

[4] Eduardo Arnold, Omar Y. Al-Jarrah, Mehrdad Dianati, Saber Fallah, David Oxtoby, and Alex Mouzakitis. A survey on 3d object detection methods for autonomous driving applications. *IEEE Transactions on Intelligent Transportation Systems*, page 3782–3795, Oct 2019.

[5] Rui Qian, Xin Lai, and Xirong Li. 3d object detection for autonomous driving: A survey. *Pattern Recognition*, page 108796, Oct 2022.

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

[7] 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.

[8] Youquan Liu et al. Segment any point cloud sequences by distilling vision foundation models. In *Advances in Neural Information Processing Systems*, volume 36, pages 37193–37229, 2023.

[9] 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. 1

[10] 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.

[11] Xiang Xu, Lingdong Kong, Hui Shuai, Liang Pan, Ziwei Liu, and Qingshan Liu. LiMoE: Mixture of LiDAR representation learners from automotive scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27368–27379, 2025. 1

[12] Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. Pointrcnn: 3d object proposal generation and detection from point cloud. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2019. 1, 4

[13] Zetong Yang, Yanan Sun, Shu Liu, and Jiaya Jia. 3dssd: Point-based 3d single stage object detector. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2020.

[14] Xuran Pan, Zhuofan Xia, Shiji Song, Li Erran Li, and Gao Huang. 3d object detection with pointformer. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2021.

[15] 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.

[16] Ao Liang, Youquan Liu, Yu Yang, et al. LiDARCrafter: Dynamic 4D world modeling from LiDAR sequences. *arXiv preprint arXiv:2508.03692*, 2025. 1

[17] Jiajun Deng, Shaoshuai Shi, Peiwei Li, Wengang Zhou, Yanyong Zhang, and Houqiang Li. Voxel r-cnn: Towards high performance voxel-based 3d object detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, page 1201–1209, Sep 2022. 1, 4

[18] Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. *Sensors*, page 3337, Oct 2018.

[19] Fangzhou Hong, Lingdong Kong, Hui Zhou, Xinge Zhu, Hongsheng Li, and Ziwei Liu. Unified 3D and 4D panoptic segmentation via dynamic shifting networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5):3480–3495, 2024. 1

[20] Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2020. 1

[21] 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.

[22] Shaoshuai Shi, Li Jiang, Jiajun Deng, Zhe Wang, Chaoxu Guo, Jianping Shi, Xiaogang Wang, and Hongsheng Li. Pv-rcnn++: Point-voxel feature set abstraction with local vector representation for 3d object detection. *International Journal of Computer Vision*, page 531–551, Feb 2023. 1, 4

[23] Zhichao Li, Feng Wang, and Naiyan Wang. Lidar r-cnn: An efficient and universal 3d object detector. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2021.

[24] 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. 1

[25] 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. 1

[26] 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. 1

[27] 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. 1

[28] 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. 1

[29] 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. 1

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

[31] Youquan Liu et al. La La LiDAR: Large-scale layout generation from LiDAR data. *arXiv preprint arXiv:2508.03691*, 2025.

[32] Dekai Zhu, Yixuan Hu, Youquan Liu, et al. Spiral: Semantic-aware progressive LiDAR scene generation and understanding. *arXiv preprint arXiv:2505.22643*, 2025.

[33] 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.

[34] 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.

[35] 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. 1

[36] Kenneth Chaney, Fernando Cladera, Ziyun Wang, Anthony Bisulco, M Ani Hsieh, Christopher Korpela, Vijay Kumar, Camillo J Taylor, and Kostas Daniilidis. M3ED: Multi-robot, multi-sensor, multi-environment event dataset. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4016–4023, 2023. 1

[37] Ao Liang et al. Perspective-invariant 3D object detection. In *IEEE/CVF International Conference on Computer Vision*, pages 27725–27738, 2025. 1, 3

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

[39] Lingdong Kong, Dongyue Lu, Xiang Xu, Lai Xing Ng, Wei Tsang Ooi, and Benoit R. Cottereau. EventFly: Event camera perception from ground to the sky. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1472–1484, 2025. 1

[40] Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, page 1–20, 2022. 1

[41] 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*, volume 37, pages 34980–35017, 2024. 1, 3

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

[43] Xiaoshuai Hao, Mengchuan Wei, Yifan Yang, et al. Is your HD map constructor reliable under sensor corruptions? In *Advances in Neural Information Processing Systems*, volume 37, pages 22441–22482, 2024.

[44] 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.

[45] 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.

[46] 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.

[47] Runnan Chen et al. CLIP2Scene: Towards label-efficient 3D scene understanding by CLIP. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7020–7030, 2023.

[48] Runnan Chen et al. Towards label-free scene understanding by vision foundation models. In *Advances in Neural Information Processing Systems*, volume 36, pages 75896–75910, 2023. 1

[49] Ye Li et al. Optimizing LiDAR placements for robust driving perception in adverse conditions. *arXiv preprint arXiv:2403.17009*, 2024. 1

[50] Lingdong Kong, Niamul Quader, and Venice Erin Liong. ConDA: Unsupervised domain adaptation for LiDAR segmentation via regularized domain concatenation. In *IEEE International Conference on Robotics and Automation*, pages 9338–9345, 2023. 1

[51] Maximilian Jaritz, Tuan-Hung Vu, Raoul de Charette, Emilie Wirbel, and Patrick Pérez. xMUDA: Cross-modal unsupervised domain adaptation for 3D semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12605–12614, 2020.

[52] Xiang Xu et al. Beyond one shot, beyond one perspective: Cross-view and long-horizon distillation for better LiDAR representations. In *IEEE/CVF International Conference on Computer Vision*, pages 25506–25518, 2025.

[53] Jingyi Xu, Weidong Yang, Lingdong Kong, Youquan Liu, Qingyuan Zhou, Rui Zhang, Zhijun Li, Wen-Ming Chen, and Ben Fei. Visual foundation models boost cross-modal unsupervised domain adaptation for 3d semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 26(11):20287–20301, 2025.

[54] Xidong Peng, Runnan Chen, Feng Qiao, et al. Learning to adapt SAM for segmenting cross-domain point clouds. In *European Conference on Computer Vision*, pages 54–71. Springer, 2024. 1

[55] 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. 1

[56] Jihan Yang, Shaoshuai Shi, Zhe Wang, Hongsheng Li, and Xiaojuan Qi. St3d: Self-training for unsupervised domain adaptation on 3d object detection. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2021. 1

[57] Jordan S. K. Hu, Tianshu Kuai, and Steven L. Waslander. Point density-aware voxels for lidar 3d object detection, 2022. 4

[58] 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. 3

[59] 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. 3

[60] 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*, volume 36, pages 21298–21342, 2023. 3

[61] 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. 3

[62] 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. 3

[63] 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. 3

[64] 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.

[65] 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. 3

[66] OpenPCDet Development Team. Openpcdet: An open-source toolbox for 3d object detection from point clouds. `https://github.com/open-mmlab/OpenPCDet`, 2020. 4