Unsupervised Domain Adaptation for 3D Object Detection via Adversarial Learning

Shurui Qin[1], Gan Sun[1*], Yao He[1], Wei Cong[1,2]
[1] School of Automation Science and Engineering, South China University of Technology, Guangzhou, 510640, China. [†]
[2] Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, 110016, China.
ausrqin@mail.scut.edu.cn, {sungan1412, heyao0293, congwei45}@gmail.com

Abstract

This technical report presents our solution for Track 5 of the 2025 RoboSense Challenge on Cross-Platform 3D Object Detection. Although 3D object detection is widely used in autonomous driving and robotics, most existing datasets are collected from vehicle-mounted LiDAR sensors, resulting in a noticeable domain gap when models are deployed on robot or drone platforms. This mismatch in viewpoint, motion pattern, and sensor height often leads to significant performance degradation. To address this challenge, we develop ***UADA3D****, an unsupervised adversarial domain adaptation framework built upon IA-SSD, a point-based single-stage 3D detector. UADA3D aims to reduce cross-platform discrepancies through domain-consistent feature alignment without requiring target-domain labels. Experimental results on the challenge benchmarks demonstrate that our method achieves effective cross-platform generalization and validates its suitability for real-world multi-platform 3D perception.*

1. Introduction

Track 5 of the RoboSense Challenge 2025 focuses on the problem of *Cross-Platform 3D Object Detection*, a setting where a detector trained on labeled vehicle-mounted LiDAR data must generalize to unlabeled data collected from drone- and quadruped-mounted LiDAR sensors [1–4]. This task is constructed on top of the Pi3DET dataset [5], which provides synchronized multi-sensor recordings from three distinct robotic platforms. The challenge setting closely mirrors real-world deployment conditions, where perception systems must operate reliably across diverse sensor configurations without requiring costly re-annotation efforts for every new platform [6–11].

Cross-platform generalization is inherently difficult due to large geometric and semantic discrepancies between different platforms [12–16]. Vehicle-mounted LiDAR typically captures ground-level urban scenes with consistent viewpoints, while drone-mounted sensors observe environments from elevated, oblique perspectives [5, 17]. In contrast, quadruped platforms operate at low heights with frequent pitch and roll variations, producing highly dynamic point cloud patterns. These differences lead to domain gaps in point density, occlusion patterns, object appearance, and spatial distributions [18–24]. As a result, a detector optimized for vehicle data often fails to transfer effectively to drone or quadruped domains, motivating the need for robust domain adaptation strategies [25–30].

This challenge is further motivated by the broader landscape of LiDAR-based 3D object detection [31]. Although major datasets such as KITTI [2], Waymo [32], nuScenes [3], LiDAR-CS [33], and Argoverse [34] have propelled rapid advances in 3D detection, they remain limited to fixed sensor configurations and platform types. Detectors trained on one dataset often perform poorly when applied to another due to differences in sensor type, LiDAR scan pattern, sampling rate, environmental conditions, and annotation protocols [35, 36]. These discrepancies highlight the critical importance of *Domain Adaptation (DA)*, which aims to transfer a model from a labeled source domain to an unlabeled or sparsely labeled target domain.

Unsupervised Domain Adaptation (UDA) [37, 38] is particularly relevant for multi-platform robotics, as labeling 3D bounding boxes on every new platform is prohibitively expensive and time-consuming [5]. UDA seeks to bridge the domain gap without relying on target labels. Existing methods primarily follow two major paradigms: (1) self-training with pseudo-label refinement, where progressively improved target-domain pseudo labels guide adaptation, and (2) adversarial feature alignment, where detection features are encouraged to be domain-invariant. Although effective

[*]This team is supervised by Prof. Gan Sun.
[†]This work is supported by NSFC (62273333), and the Fundamental Research Funds for the Central Universities.

in traditional domain shifts, these methods have not been extensively explored in the context of cross-platform adaptation, where viewpoint and motion differences introduce more complex, nonlinear discrepancies.

The Pi3DET dataset provides an excellent testbed for this setting. It contains over 51,000 frames of multi-modal data (LiDAR, camera, event camera, IMU) recorded from vehicle, drone, and quadruped platforms. Compared with standard domain shift benchmarks, Pi3DET introduces more substantial geometric variations, platform-specific motion patterns, and environment differences, making it highly suitable for evaluating methods designed to achieve robust cross-platform perception.

In this report, we explore an unsupervised adversarial adaptation approach aimed at mitigating the cross-platform domain gap. By designing a framework that aligns detection features between vehicular and non-vehicular platforms, our method seeks to enhance robustness without relying on target-domain annotations. The results on the RoboSense Challenge benchmarks illustrate the potential of UDA-based solutions to enable reliable and scalable multi-platform 3D object detection.

2. Related Work

2.1. 3D LiDAR-based Object Detection

LiDAR-based 3D object detection has been extensively explored in autonomous driving and robotics, with methods differing primarily in how they represent and process point clouds. Voxel-based approaches such as VoxelNet [39], VoxelNeXt [40], PV-RCNN [41], PV-RCNN++ [42], and SECOND [43] convert point clouds into regular voxel grids and extract features via dense or sparse 3D convolutions. Pillar-based methods like PointPillars [44] simplify voxels into vertical pillars to improve computational efficiency, while point-based detectors such as PointNet [45], PointNet++ [46], and IA-SSD [47] operate directly on raw points to preserve fine-grained geometric structures. Hybrid point–voxel systems, including PV-RCNN variants, aim to combine the advantages of both paradigms.

From the backbone perspective, early works typically rely on dense 3D or 2D convolutional architectures (*e.g.*, VoxelNet, PointPillars, CenterPoint [48]), whereas more recent efforts employ sparse convolutional backbones for improved efficiency and scalability, as seen in HEDNet [49] and SAFDNet [50]. These advancements have enabled strong performance across multiple autonomous driving benchmarks; however, most methods assume fixed sensor configurations and struggle under domain shifts introduced by heterogeneous robotic platforms.

2.2. Unsupervised Domain Adaptation for 3D Detection

Domain adaptation (DA) aims to transfer models trained on a labeled source domain to an unlabeled or partially labeled target domain, reducing performance degradation caused by domain discrepancies. In LiDAR-based 3D object detection, three major categories of UDA methods have emerged. Pseudo-labeling and self-training approaches, such as ST3D [51], ST3D++ [52], MS3D [53], MS3D++ [54], and REDB [14], iteratively refine pseudo-labels generated by a source-trained teacher model. These methods incorporate denoising strategies [55–58], multi-detector fusion, and cross-domain consistency checks to improve pseudo-label reliability and reduce error accumulation.

Adaptive network frameworks, such as Uni3DA [59], explore unified domain-agnostic architectures designed to operate across heterogeneous autonomous driving datasets. These approaches attempt to learn representations that are inherently robust to variations in sensor type, geographic region, and environmental conditions [10, 60–66].

Adversarial learning-based UDA methods aim to align features across domains through discriminator-guided optimization. Representative works include STAL3D [24] and UADA3D [67], which focus on aligning class-conditional features to achieve domain-invariant representations. Unlike pseudo-labeling frameworks that rely on a teacher model, adversarial approaches directly enforce distribution alignment and are particularly well-suited for cross-platform shifts where geometric discrepancies are large.

Most existing UDA studies address adaptation across autonomous driving datasets (*e.g.*, Waymo $\rightarrow$ nuScenes). In contrast, the cross-platform setting considered in the RoboSense Challenge—vehicle $\rightarrow$ drone or quadruped—introduces more substantial viewpoint, motion, and height differences. Methods such as UADA3D are therefore especially relevant, as they target robust feature alignment without depending on target-domain annotations or pre-trained teachers.

3. Approach

The official baseline for this challenge combines PV-RCNN (a voxel-based 3D object detector) with ST3D++ (a pseudo-label-based self-training method for domain adaptation). However, this method has two critical limitations: the training process of PV-RCNN is computationally expensive on resource-constrained platforms, and ST3D++ is less effective for tasks involving significant sensor angle discrepancies, as its pseudo-labeling is prone to errors in such scenarios.

To overcome these issues, we replace PV-RCNN with IA-SSD (a more efficient point-based detector), and substitute ST3D++ with UADA3D (an adversarial learning-based adaptation framework). As illustrated in Figure 1, this integrated

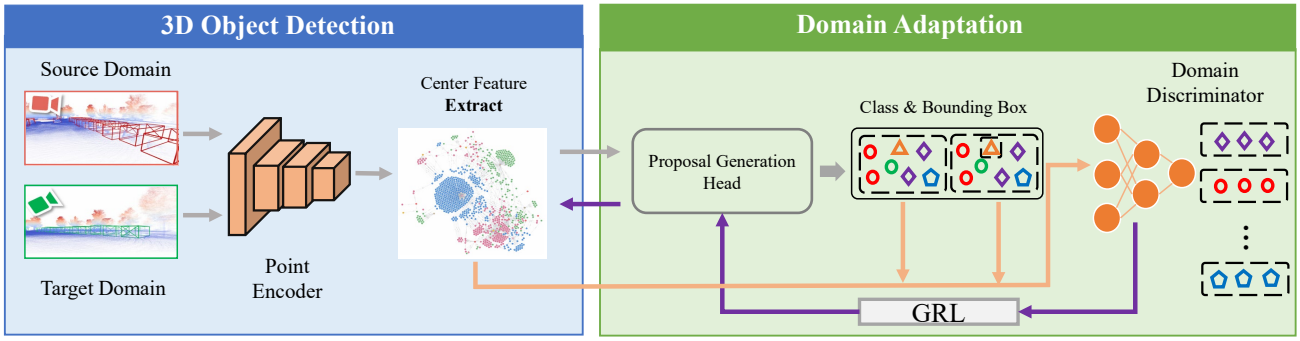

Figure 1. Overview of our method with two parts: 3D object detection and domain adaptation. The orange line indicates that the center feature, bounding box, and predicted class are fed together into the discriminator for domain discrimination. The loss of discriminator is then backpropagated through the purple line via the Gradient Reversal Layer (GRL) to the feature extraction layer and the detection head.

approach is designed for greater efficiency and robustness to cross-platform domain shifts.

3.1. 3D Object Detection

IA-SSD is a single-stage, point-based 3D object detection framework that achieves superior cross-scene adaptability through instance-aware downsampling and feature masking. Unlike voxel-based methods such as PV-RCNN, IA-SSD [47] processes raw point clouds directly. To address sensor pose variations across platforms, IA-SSD incorporates vertical offset correction and enables multi-dataset transfer by adjusting the number of sampled input points. Its feature masking mechanism could extract object-centric representations from predicted 3D bounding boxes while suppressing background noise. These characteristics make IA-SSD particularly well-suited for cross-platform adaptation tasks, which is why we adopt it as our core detection component.

Strengths of IA-SSD rely on two key mechanisms: class-aware sampling (to retain semantically relevant points) and centroid-aware sampling (to weight centroid-proximal points). As a point-based detector, it operates directly on raw point clouds, avoiding voxelization artifacts, reduces computational overhead, combines high efficiency, and supports easy end-to-end training. These advantages enable its robust adaptation across scenarios such as vehicle→Drone, vehicle→Quadruped platform transitions.

3.2. Domain Adaptation Method

Unsupervised Domain Adaptation (UDA) mitigates the intrinsic limitations of self-training architecture of ST3D [51], which relies on iterative pseudo-label transfer for knowledge propagation but suffers from heavy dependence on pseudo-label quality. In sparse LiDAR and large domain gap scenarios, ST3D is susceptible to pseudo-label noise accumulation and bounding box deviation, driven mainly by incomplete point cloud features.

By contrast, UADA3D [67] adopts a dependency-free adversarial learning framework, dispensing with pre-trained models and pseudo-labels. After adversarial interplay between a feature extractor and class-wise domain discriminators, it directly learns domain-invariant features. Augmented with feature masking and vertical offset correction, UADA3D retains core object features in sparse point clouds and aligns cross-scenario class distributions, thereby achieving superior performance across sparse-to-dense, dense-to-sparse, and cross-platform LiDAR adaptation tasks.

Source and target point clouds are first downsampled to a uniform number of points and encoded into point-wise features. These features are progressively downsampled and transformed into high-dimensional semantic representations through a series of Set Abstraction (SA) modules. Subsequently, a Vote Layer aggregates features from object surfaces toward instance centers, generating center features that serve as discriminative representations for domain adaptation. The center features are then passed through fully connected layers to produce bounding box predictions, object classification, and foreground-background segmentation, where the detection loss is computed exclusively on labeled source data. Finally, instance-aware feature masking is applied to ensure that only object-relevant features are utilized for domain discrimination.

Suppose that the masked feature is f, the predicted bounding box is b, the discriminator as d and t as the label of the domain. Then the target domain is 1 and 0 for the source domain, y as the corresponding instance confidence score. The loss function of the domain discrimination L_{dis}, which uses the least-squares loss function, is defined as:

$$L_{dis} = \frac{1}{n}\sum_{i=1}^{n}[y_{i,k} \odot (d_{\theta,k}(f,b) - t)^2], \tag{1}$$

where k is the number of the instance classes, which is 2 in the task, including car and pedestrian. $d_{\theta,k}(f,b)$ is

symboled as f and b are concatenated and then input into the discriminator, θ is the parameter of the discriminator. Next, the L_{dis} will be sent via GRL to the feature extractor and Proposal generation head.

Suppose that the parameters of the feature extractor, proposal generation head, and discriminator are θ_{fea}, θ_{prop} and θ_{dis} respectively, and the optimal parameter are θ^*_{fea}, θ^*_{prop} and θ^*_{dis} respectively. The optimization process is defined as:

$$\begin{aligned} \theta^*_{dis} &= \arg\min_{\theta_{dis}} L_{dis} \\ (\theta^*_{fea}, \theta^*_{prop}) &= \arg\min_{\theta_{fea}, \theta_{prop}} L_{det} - \lambda L_{dis}, \end{aligned} \tag{2}$$

where λ is a hyperparameter, which can be a constant or an adaptive value. The domain adaptation equation can be defined as:

$$\lambda = \alpha(\frac{2}{1+\exp(-\gamma p)} - 1), \quad \alpha \in [0, 1] \tag{3}$$

where $\gamma = 10$ and p is the training progress linearly changing from 0 to 1. According to the UADA3D [67], the α to 0.1 and directly setting λ to 0.1 will achieve a similar average performance.

3.3. Collaborative Framework

In conventional UDA schemes for 3D object detection, object detectors and domain adaptation functions are utilized as separate modules. This separation introduces two key issues. Firstly, voxel-based and point-based feature representations are mismatched, necessitating costly conversion modules that not only add overhead but also induce feature loss. Secondly, staged training decouples the optimization of detection performance and cross-domain generalization, which hinders synergistic gains and ultimately degrades cross-domain detection accuracy.

To address this bottleneck, an integrated synergistic framework is proposed via combining the IA-SSD detector and UADA3D domain adaptation. We leverage the point-based nature of IA-SSD at the feature level, where its point cloud features can be extracted via multiple Set Abstraction (SA) modules while preserving the critical geometric and semantic information of objects. These features can be directly fed into the class-wise domain discriminator of UADA3D without format conversion, thus eliminating adaptation-related losses. At the parameter optimization level, the domain loss of UADA3D L_c is backpropagated to the feature extractor and detection head of the IA-SSD via a Gradient Reversal Layer (GRL), which establishes a "detection loss L_{det}–domain loss L_c" joint optimization mechanism. During the training phase, the model minimizes the 3D detection loss L_{det}, which includes losses for bounding box regression, object classification, and foreground segmentation to ensure baseline accuracy. Meanwhile, UADA3D narrows the feature distribution gap between the source and target domains, thereby avoiding performance degradation in cross-platform migration scenarios (*e.g.*, vehicle→Drone/vehicle→quadruped). The core advantage of this framework lies in the deep synergy between features and parameters: it retains the adaptability of the IA-SSD to sparse point clouds and mitigates cross-platform domain discrepancies via UADA3D, ultimately enabling the efficient deployment of multi-platform perception systems.

4. Experiments

4.1. Dataset

We use the official data provided by the *RoboSense Challenge 2025* [68] held at IROS 2025. This competition builds upon the legacy of the *RoboDepth Challenge 2023* [69, 70] at ICRA 2023 and the *RoboDrive Challenge 2024* [71, 72] 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 [5, 73–76]. Specifically, this task is built upon the **Pi3DET** benchmark [5] 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.

4.2. Implementation Details

Our experiments follow the RoboSense 2025 Challenge configuration. We use the labeled vehicle platform subset of the Pi3DET dataset as the source domain. For Phase 1, the unlabeled drone platform subset serves as the target domain, with detection limited to a single class (car). For Phase 2, the unlabeled quadruped platform subset is used as the target domain, requiring detection of two classes (car and pedestrian). All experiments are implemented using the OpenPCDet v0.5 toolbox [77] and PyTorch 2.1 framework, and are conducted on a single NVIDIA RTX 4070 Super GPU (12GB memory) running CUDA 11.8 on Ubuntu 22.04.

4.3. Experimental Setups

We evaluate three detector-adaptation combinations: PV-RCNN with ST3D++ (baseline), IA-SSD with ST3D++, and IA-SSD with UADA3D. All methods follow a two-stage training protocol. Firstly, IA-SSD is pre-trained on the source domain by using the Adam optimizer with a learning rate of 0.01 and a batch size of 16 for 60 epochs. Subsequently, for adaptation with ST3D++ or UADA3D, the learning rate is adjusted to 0.09, and the batch size to 12, with training continued for 40 epochs. The GRL coefficient λ in UADA3D is fixed at 0.1. All training employs a cosine annealing learning rate schedule. Evaluation is conducted on the official competition server. For data preprocessing, we

Method	CAR-AP@50	CAR-AP@70
PVRCNN + ST3D++(baseline)	45.3005	25.8341
IASSD + ST3D++	51.3025	30.0155
Ours	57.6	34.65

Table 1. Results of Phase 1 mission, vehicle to drone adaptation.

Method	CAR-AP@50	CAR-AP@70	Ped-AP@0.25	Ped-AP@0.50
PVRCNN + ST3D++(baseline)	28.53	12.81	47.41	41.49
Ours	36.59	17.4	60.21	51.31

Table 2. Results of Phase 2 mission, vehicle to quadruped adaptation.

uniformly downsample both source and target point clouds to 16,384 points to meet fixed-size input requirement of IA-SSD. Following the Pi3DET-NET[5] training guidelines, we disable random object scaling (ROS) augmentation, as it has been shown to degrade performance.

4.4. Comparative Study

As show in Table 1, for the same UDA framework ST3D++, change the 3D object detector from PV-RCNN to IA-SSD can make a great progress. We hypothesize that this is because PV-RCNN is more memory-intensive. Under the same GPU memory constraints, IA-SSD allows for a larger batch size, which we believe is the primary reason for its superior performance. For the same 3D object detector IA-SSD,change the UDA framework from ST3D++ to UADA3D can also make a great progress; it may because ST3D++ is a pseudo-label and self-training-based method, and ST3D++ depends on the quality of the pretrained teacher model. Different platforms have a great difference in point geometry distribution.

As the results shown in Table 2, for Phase 2, IA-SSD demonstrates strong performance in pedestrian detection, while UADA3D achieves significant improvements for both car and pedestrian classes.

5. Conclusion

For Track 5 of the 2025 RoboSense Challenge, we introduce a method that combines IA-SSD, a point-based single-stage 3D object detector, with UADA3D, an Unsupervised Adversarial Domain Adaptation framework. Our presented results demonstrate that an adversarial learning-based UDA method holds significant potential for the cross-platform 3D object detection task. Future work could explore the migration of the framework to sparse-convolution-based networks, as well as Domain Adaptation with multi-model inputs.

References

[1] 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.

[2] 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.

[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] 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.

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

[6] 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.

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

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] Zhuoxiao Chen, Yadan Luo, Zheng Wang, Mahsa Baktashmotlagh, and Zi Huang. Revisiting Domain-Adaptive 3D Object Detection by Reliable, Diverse and Class-balanced Pseudo-Labeling. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3691–3703, Paris, France, October 2023. IEEE.

[15] 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.

[16] 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.

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

[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] Youquan Liu et al. Multi-space alignments towards universal LiDAR segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14648–14661, 2024.

[20] Jingyi Xu et al. Visual foundation models boost cross-modal unsupervised domain adaptation for 3d semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 26(11):20287–20301, 2025.

[21] 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.

[22] Xiang Xu et al. 4D contrastive superflows are dense 3D representation learners. In *European Conference on Computer Vision*, pages 58–80, 2024.

[23] Fangzhou Hong et al. Unified 3D and 4D panoptic segmentation via dynamic shifting networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5):3480–3495, 2024.

[24] Yanan Zhang, Chao Zhou, and Di Huang. STAL3D: Unsupervised Domain Adaptation for 3D Object Detection via Collaborating Self-Training and Adversarial Learning. *IEEE Transactions on Intelligent Vehicles*, 9:7339–7350, November 2024.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] Pengfei Wei et al. Unsupervised video domain adaptation for action recognition: a disentanglement perspective. In *Advances in Neural Information Processing Systems*, volume 36, pages 17623–17642, 2023.

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

[32] 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.

[33] Jin Fang, Dingfu Zhou, Jingjing Zhao, Chenming Wu, Chulin Tang, Cheng-Zhong Xu, and Liangjun Zhang. LiDAR-CS Dataset: LiDAR Point Cloud Dataset with Cross-Sensors for 3D Object Detection, March 2024.

[34] 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, January 2023.

[35] Zhanwei Zhang, Minghao Chen, Shuai Xiao, Liang Peng, Hengjia Li, Binbin Lin, Ping Li, Wenxiao Wang, Boxi Wu, and Deng Cai. Pseudo Label Refinery for Unsupervised Domain Adaptation on Cross-dataset 3D Object Detection, April 2024.

[36] Lingdong Kong et al. LargeAD: Large-scale cross-sensor data pretraining for autonomous driving. *arXiv preprint arXiv:2501.04005*, 2025.

[37] Jiahua Dong, Zhen Fang, Anjin Liu, Gan Sun, and Tongliang Liu. Confident anchor-induced multi-source free domain adaptation. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 2848–2860. Curran Associates, Inc., 2021.

[38] Jiahua Dong, Yang Cong, Gan Sun, Zhen Fang, and Zhengming Ding. Where and how to transfer: Knowledge aggregation-induced transferability perception for unsupervised domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(3):1664–1681, 2024.

[39] Yin Zhou and Oncel Tuzel. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection, November 2017.

[40] Yukang Chen, Jianhui Liu, Xiangyu Zhang, Xiaojuan Qi, and Jiaya Jia. VoxelNeXt: Fully Sparse VoxelNet for 3D Object Detection and Tracking, March 2023.

[41] 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, April 2021.

[42] 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, November 2022.

[43] Yan Yan, Yuxing Mao, and Bo Li. SECOND: Sparsely Embedded Convolutional Detection. *Sensors*, 18:3337, October 2018.

[44] Alex H. Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. PointPillars: Fast Encoders for Object Detection from Point Clouds, May 2019.

[45] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation, 2017.

[46] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space, 2017.

[47] Yifan Zhang, Qingyong Hu, Guoquan Xu, Yanxin Ma, Jianwei Wan, and Yulan Guo. Not All Points Are Equal: Learning Highly Efficient Point-based Detectors for 3D LiDAR Point Clouds. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18931–18940, New Orleans, LA, USA, June 2022. IEEE.

[48] Tianwei Yin, Xingyi Zhou, and Philipp Krähenbühl. Center-based 3D Object Detection and Tracking, January 2021.

[49] Gang Zhang, Junnan Chen, Guohuan Gao, Jianmin Li, and Xiaolin Hu. HEDNet: A Hierarchical Encoder-Decoder Network for 3D Object Detection in Point Clouds, October 2023.

[50] Gang Zhang, Junnan Chen, Guohuan Gao, Jianmin Li, Si Liu, and Xiaolin Hu. SAFDNet: A Simple and Effective Network for Fully Sparse 3D Object Detection, September 2024.

[51] Jihan Yang, Shaoshuai Shi, Zhe Wang, Hongsheng Li, and Xiaojuan Qi. ST3D: Self-training for Unsupervised Domain Adaptation on 3D Object Detection, March 2021.

[52] Jihan Yang, Shaoshuai Shi, Zhe Wang, Hongsheng Li, and Xiaojuan Qi. ST3D++: Denoised Self-training for Unsupervised Domain Adaptation on 3D Object Detection, August 2021.

[53] Darren Tsai, Julie Stephany Berrio, Mao Shan, Eduardo Nebot, and Stewart Worrall. Ms3d: Leveraging multiple detectors for unsupervised domain adaptation in 3d object detection, 2023.

[54] Darren Tsai, Julie Stephany Berrio, Mao Shan, Eduardo Nebot, and Stewart Worrall. Ms3d++: Ensemble of experts for multi-source unsupervised domain adaptation in 3d object detection. *IEEE Transactions on Intelligent Vehicles*, 10(3):1999–2014, March 2025.

[55] Zhengfeng Lai, Noranart Vesdapunt, Ning Zhou, Jun Wu, Cong Phuoc Huynh, Xuelu Li, Kah Kuen Fu, and Chen-Nee Chuah. PADCLIP: Pseudo-labeling with Adaptive Debiasing in CLIP for Unsupervised Domain Adaptation. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 16109–16119, Paris, France, October 2023. IEEE.

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

[57] 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.

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

[59] Yu Ren, Yang Cong, Jiahua Dong, and Gan Sun. Uni3da: Universal 3d domain adaptation for object recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1):379–392, 2023.

[60] 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.

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

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

[63] 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.

[64] Lingdong Kong, Xiang Xu, Jun Cen, Wenwei Zhang, Liang Pan, Kai Chen, and Ziwei Liu. Calib3D: Calibrating model preferences for reliable 3D scene understanding. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1965–1978, 2025.

[65] 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.

[66] 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.

[67] Maciej K. Wozniak, Mattias Hansson, Marko Thiel, and Patric Jensfelt. UADA3D: Unsupervised Adversarial Domain Adaptation for 3D Object Detection with Sparse LiDAR and Large Domain Gaps, October 2024.

[68] 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.

[69] 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.

[70] 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.

[71] 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.

[72] 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.

[73] 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.

[74] 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.

[75] 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.

[76] 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.

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