

Using Temporal Information and Mixing-Based Data Augmentations for Robust HD Map Construction

Xiaoshuai Hao¹, Yifan Yang¹, Hui Zhang¹, Mengchuan Wei¹, Yi Zhou¹, Haimei Zhao², Jing Zhang²

¹Samsung R&D Institute China–Beijing

²The University of Sydney

{xshuai.hao, yifan.yang, hui123.zhang, mc.wei, yi0813.zhou}@samsung.com,

{h.zhao, jing.zhang1}@sydney.edu.au

Abstract

In this report, we describe the technical details of our submission to the 2024 RoboDrive Challenge-Robust Map Segmentation Track. The Robust Map Segmentation track focuses on the segmentation of complex driving scene elements in BEV maps under varied driving conditions. Semantic map segmentation provides abundant and precise static environmental information crucial for autonomous driving systems' planning and navigation. While current methods excel in ideal circumstances, e.g., clear daytime conditions and fully functional sensors, their resilience to real-world challenges like adverse weather and sensor failures remains unclear, raising concerns about system safety. In this paper, we explored several methods to improve the robustness of the map segmentation task. The details are as follows: 1) Robustness analysis of utilizing temporal information; 2) Robustness analysis of utilizing different backbones; and 3) Data Augmentation to boost corruption robustness. Based on the evaluation results, we draw several important findings including 1) The temporal fusion module is effective in improving the robustness of the map segmentation model; 2) A strong backbone is effective for improving the corruption robustness; and 3) Some data augmentation methods are effective in improving the robustness of map segmentation models. These novel findings allowed us to achieve promising results in the 2024 RoboDrive Challenge-Robust Map Segmentation Track.

1. Introduction

In the rapidly evolving domain of autonomous driving, the accuracy and resilience of perception systems are paramount [1–3]. Recent advancements, particularly in bird's eye view

(BEV) representations and LiDAR sensing technologies, have significantly improved in-vehicle 3D scene perception [4, 5]. However, the robustness of 3D scene perception methods under varied and challenging conditions — integral to ensuring safe operations — has been insufficiently assessed [6]. To fill this gap, The ICRA 2024 RoboDrive Challenge, seeks to push the frontiers of robust autonomous driving perception. RoboDrive [7] is one of the first benchmarks that targeted probing the Out-of-Distribution (OOD) robustness of state-of-the-art autonomous driving perception models, centered around two mainstream topics: common corruptions and sensor failures.

Track 2 - Robust Map Segmentation requires contestants to use advanced machine learning algorithms to perform accurate map segmentation on high-resolution bird's-eye views. This task includes a detailed analysis of various urban geographic features, such as segmented lanes, sidewalks, green spaces, etc. In addition, this track also tests contestants' image segmentation capabilities under different lighting, weather conditions, and noise conditions.

2. Related Work

2.1. Semantic Map Construction

Semantic map construction is a prominent and extensively researched area within the field of autonomous driving. HDMaNet [1] first introduces the problem of local semantic map learning, which aims to construct semantic maps online from the observations of LiDAR sensors and cameras [8, 9]. It also proposes a learning method to build BEV features from sensory input and predicts vectorized map elements. BEVSegFormer [10] proposes the multi-camera deformable attention to transform image-view features to BEV representations for semantic map construction. Very recently, BEVerse [4] incorporates the semantic map construction as part of the multitask framework and uses vanilla convolutional layers for segmentation prediction. MBFusion [11] proposes a multi-modal BEV feature fusion approach for the

Technical Report of the [2024 RoboDrive Challenge](#).

Track 2: Robust Map Segmentation.

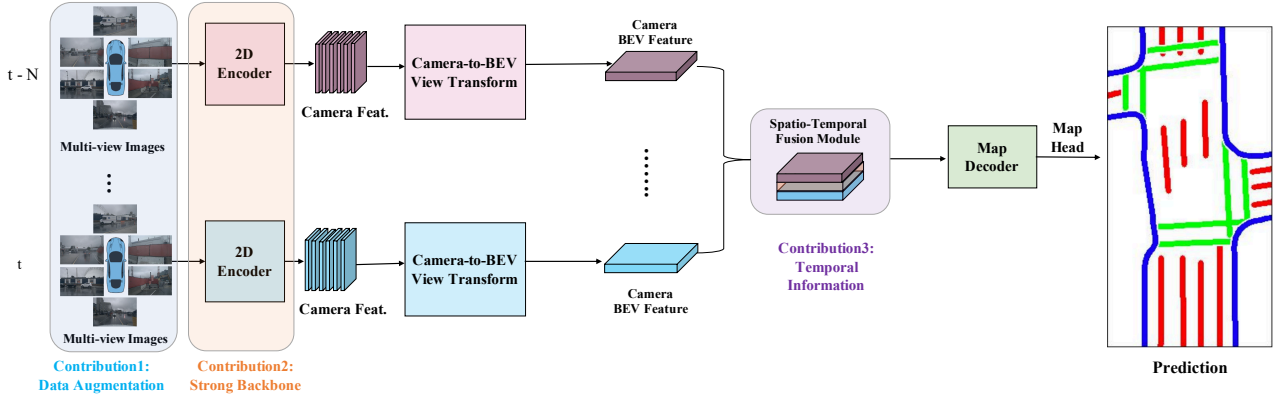


Figure 1. An overview of Our method. We explored several methods to improve the robustness of the baseline map segmentation model.

HD map construction task. In addition, high-definition maps (HD maps) have received widespread attention, which can provide fine-grained information about road scenes and play a vital role in autonomous vehicles.

2.2. Robustness Under Natural Corruptions

Recently, assessing autonomous driving perception model robustness under natural corruptions has burgeoned as a pivotal research domain [12–16]. The natural corruption robustness often refers to the capability of a conventionally trained model maintaining satisfactory performance under natural distribution shifts. ModelNet-C [17] design a taxonomy of common 3D corruptions and identify the atomic corruptions to understand classifiers’ robustness. Recently, the corruption robustness of BEV perception tasks has been widely studied. RoboDepth [14, 17] establishes a robustness benchmark for monocular depth estimation under corruptions. RoboBEV [15] introduces a comprehensive benchmark for evaluating the robustness under natural corruptions of four BEV perception algorithms, such as 3D object detection [18, 19], semantic segmentation [4, 20], depth estimation [21], and semantic occupancy prediction [22, 23]. [12] systematically designs 27 types of common corruptions in 3D object detection for both LiDAR and camera sensors. Robo3D [24] presents a comprehensive benchmark for probing the robustness of 3D detectors and segmentors [25–27] under out-of-distribution scenarios against natural corruptions that occur in real-world environments. In this technical report, we study the robustness of semantic map segmentation models under sensor corruption.

3. Approach

3.1. Baseline Model

We first establish a baseline of semantic map segmentation based on the state-of-the-art BEVerse [4] model. BEVerse takes M surrounding camera images from N timestamps

and the corresponding ego-motions and camera parameters as input. With multi-task inference, the outputs include the 3D bounding boxes and semantic map for the present frame, and the future instance segmentation and motion for the following T frames. The BEVerse framework consists of four sub-modules that are sequentially applied, including the image-view encoder, the view transformer, the spatio-temporal BEV encoder, and the multi-task decoders. These modules are described in detail in BEVerse [4].

In this paper, we explored several methods to improve the robustness of the baseline map segmentation model. The details are as follows: 1) Robustness analysis of utilizing temporal information; 2) Robustness analysis of utilizing different backbones; and 3) Data Augmentation to boost corruption robustness. The overall framework of our method is illustrated in Fig. 1. The experimental results are presented in detail in Sec. 4.

4. Experiments

4.1. Datasets

This work follows the protocol in the 2024 RoboDrive Challenge [7] when preparing the training and test data. Specifically, the model is trained on the official *train* split of the nuScenes dataset [28] and tested on the held-out competition evaluation sets. The evaluation data was created following RoboDepth [29–31], RoboBEV [6, 15, 32], and Robo3D [24, 33]. The corruption types are mainly from three sources, namely camera corruptions, camera failures, and LiDAR failures. For more details, please refer to the corresponding GitHub repositories.

4.2. Implementation Details

The model is implemented based on the MMDetection3D codebase [34]. We implemented BEVerse [4] as the baseline model for Track 2. We train the model with 4 NVIDIA RTX A6000 GPUs. For training, the AdamW [35] optimizer is

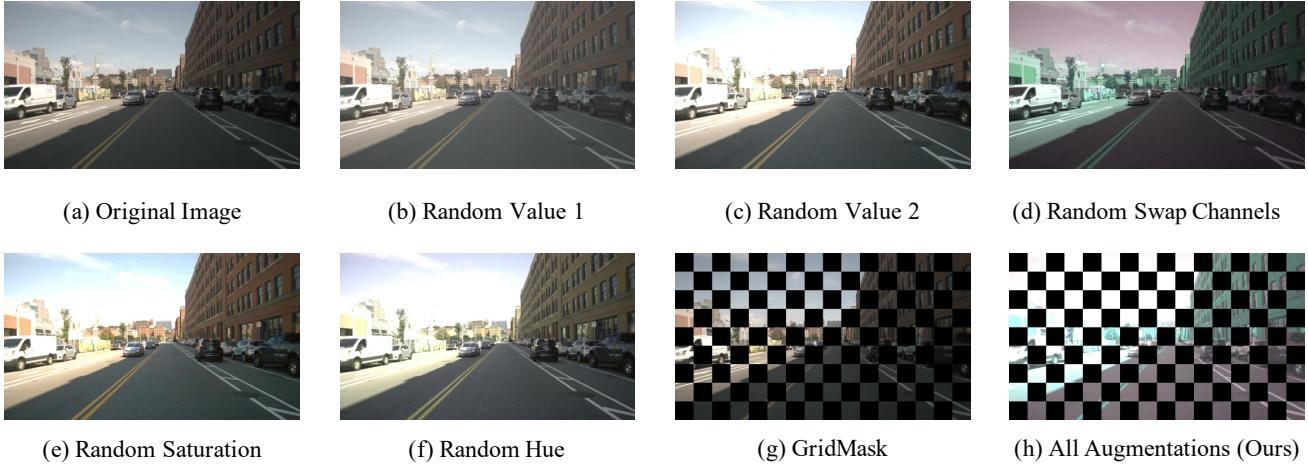


Figure 2. Examples of augmentation methods used in our method.

Table 1. Augmentation methods. **Data augmentations used in our experiments are marked in gray.**

Name	Description	Range
<i>GridMask</i>	Regularly drop the image information by multiplying with a grid mask. The grid mask has the size with image and its values are 0/1	
<i>Random Hue</i>	Convert color form BGR to HSV and add a random value to image hue channel	$[-18, 18]$
<i>Random Saturation</i>	Convert color form BGR to HSV and scale image saturation channel with rate <i>magnitude</i>	$[0.5, 1.5]$
<i>Random swap channels</i>	Randomly sawp the image RGB channels	
<i>Random value 1</i>	Add a random value to all pixels	$[-32, 32]$
<i>Random value 2</i>	Scale all pixels with rate <i>magnitude</i>	$[0.5, 1.5]$
<i>Translate</i>	Translate the image in the horizontal and vertical direction with rate <i>magnitude</i>	$[-0.3, 0, 3]$
<i>Scale</i>	Zoom in or Zoom out the image with rate <i>magnitude</i> and select the center of the scaled image	$[-0.5, 0.5]$
<i>Rotate</i>	Rotate the image <i>magnitude</i> degrees	$[-30^\circ, 30^\circ]$
<i>Shear</i>	Shear the image along the horizontal or vertical axis with rate <i>magnitude</i>	$[-0.3, 0.3]$
<i>Invert</i>	Invert the pixels of the image	
<i>Solarize</i>	Invert all pixels above a threshold value of <i>magnitude</i>	$[0, 256]$
<i>Equalize</i>	Equalize the image histogram	
<i>Color balance</i>	Adjust the color balance of the image. A <i>magnitude</i> = 0 gives a black and white image, while <i>magnitude</i> = 1 gives the original image	$[0.1, 1.9]$
<i>Auto contrast</i>	Maximize the contrast of the image	
<i>Cutout</i>	Set a random square patch of side-length <i>magnitude</i> pixels to gray	$[0, 20]$

utilized, with initial learning rate as $2e-4$, weight decay as 0.01, and gradient clip as 35. The model is trained for 20 epochs with CBGS [36]. For the learning schedule, we apply the one-cycle policy [37] with the peak learning rate as $1e-3$. Moreover, for semantic map construction, the ranges are $[-30.0m, 30.0m]$ for X-axis and $[-15.0m, 15.0m]$ for Y-axis, with the interval as 0.15m.

4.3. Metrics

Following [1, 4], the semantic classes for map construction include lane dividers, pedestrian crossings, and lane boundaries. For quantitative evaluation, we compute the intersection-over-union (IoU) for each class between the predicted and ground-truth maps. The mean IoU (mIoU) is computed as the primary evaluation metric.

4.4. Ablation Study

In this section, we analyze and discuss the impact of different configurations on the robustness of the Map Segmentation.

Robustness analysis of utilizing different backbones.

In this setting, we comprehensively investigate several backbones and show the results in Tab. 2. We construct three versions of BEVerse, namely BEVerse-Tiny, BEVerse-Small and BEVerse-Base. Specifically, we use Swin-Tiny, Swin-Small, and Swin-Base Transformer as the backbone, respectively. The experimental results demonstrate that the BEVerse-Base (using a Swin-base Transformer backbone) has significantly improved the robustness. For example, compared with BEVerse-Tiny, the BEVerse-Base achieves 7.77 absolute gains on the mIoU metric on the phase 1 test set. This shows the effectiveness of the strong backbone in improving the corruption robustness of the semantic map segmentation model.

Robustness analysis of utilizing temporal information.

We investigate the impact of utilizing temporal information module [4] on the robustness of semantic map segmentation and show the results in Tab. 3. Two model variants, BEVerse-Base with and without the temporal fusion module, are examined. The experimental results demonstrate that the temporal fusion module can significantly improve the robustness. It can be observed that BEVerse-Base (with temporal fusion) achieves 2.4 absolute gains on the mIoU metric on the phase 1 test set. This shows the effectiveness

Table 2. Robustness analysis of different backbones on phase 1 test set.

Model	miou	brigh	dark	fog	frost	snow	contr	defoc	glass	motio	zoom	elast	gauss	impul	shot	iso	pixel	jpeg
BEVerse-Tiny	16.31	16.5	10.7	36.4	0.9	4.4	3.8	14.7	27.9	16.9	10.2	39.3	6.1	4.3	6.5	5.6	42.0	32.3
BEVerse-Small	17.33	23.4	13.5	39.8	3.7	7.2	5.7	19.4	29.6	21.2	10.5	40.6	3.3	1.8	2.8	3.2	40.3	31.9
BEVerse-Base	24.08	23.7	23.7	47.1	3.8	11.5	7.0	26.0	39.0	32.0	15.5	41.9	15.0	10.1	14.3	10.9	48.7	41.2

Table 3. Robustness analysis of temporal module on phase 1 test set.

Model	miou	brigh	dark	fog	frost	snow	contr	defoc	glass	motio	zoom	elast	gauss	impul	shot	iso	pixel	jpeg
BEVerse-Base (w/o temporal)	21.68	20.4	21.2	40.9	3.4	11.0	4.4	24.4	36.4	28.3	14.9	39.3	13.2	7.6	12.1	9.0	47.1	37.3
BEVerse-Base (w/ temporal)	24.08	23.7	23.7	47.1	3.8	11.5	7.0	26.0	39.0	32.0	15.5	41.9	15.0	10.1	14.3	10.9	48.7	41.2

Table 4. Robustness analysis of data augmentations on phase 2 test set.

Model	miou	brigh	dark	fog	frost	snow	contr	defoc	glass	motio	zoom	elast	gauss	impul	shot	iso	pixel	jpeg
BEVerse-Base (w/o aug)	23.24	18.2	28.5	24.0	5.4	12.6	5.9	31.9	34.1	23.0	21.4	45.8	14.6	8.8	11.1	24.1	45.5	42.6
BEVerse-Tiny (w/ aug)	29.76	42.4	28.7	30.4	9.1	8.6	13.8	21.6	43.1	24.8	19.6	75.2	23.5	14.3	17.7	31.4	68.2	44.5
BEVerse-Base (w/ aug)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 5. Comparisons with state-of-the-art methods.

Method	miou	brigh	dark	fog	frost	snow	contr	defoc	glass	motio	zoom	elast	gauss	impul	shot	iso	pixel	jpeg
Huangxl0719	48.75	54.6	71.1	28.6	23.1	54.5	54.6	64.8	51.2	44.7	21.8	52.1	58.5	46.1	37.2	64.2	55.2	54.9
CrazyFriday	34.54	42.7	40.4	29.8	25.0	41.7	17.1	42.4	34.8	33.2	22.4	48.0	33.7	23.6	26.0	38.0	45.8	45.6
yf20221012 (Ours)	29.76	42.4	28.7	30.4	9.1	8.6	13.8	21.6	43.1	24.8	19.6	75.2	23.5	14.3	17.7	31.4	68.2	44.5
RoboDrivers	15.67	21.4	14.1	19.2	6.8	3.2	3.7	18.9	27.9	9.2	17.6	44.6	5.2	2.1	2.4	9.6	36.8	28.0

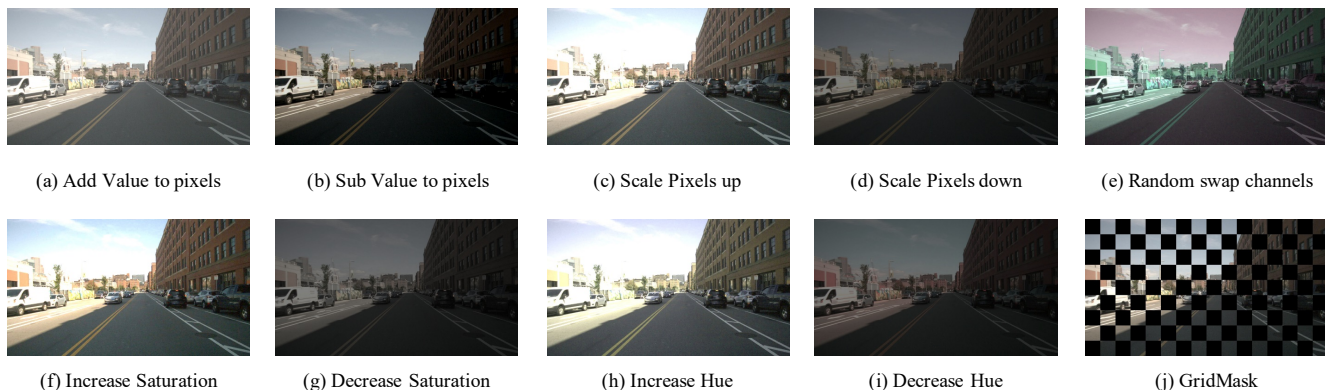


Figure 3. Illustration of different augmentation methods.

of the temporal fusion module in improving the corruption robustness of the semantic map segmentation model.

Data Augmentation to boost corruption robustness In this setting, we investigate the effect of various data augmentation techniques on the robustness of map segmentation models under natural corruptions. Specifically, we use several image data augmentation methods, i.e., GridMask, Random Hue, Random Saturation, Random Swap Channels and Random value. Tab. 1 lists the augmentation methods and their descriptions and ranges we used to augment data. We choose BEVerse-Tiny (w/ temporal) as the base model, and the results are shown in Tab. 4. It can be seen that image data augmentation methods moderately improve the

model’s robustness performance. For example, compared with BEVerse-Base, the BEVerse-Tiny (with Aug) achieves 6.52 absolute gains on the mIoU metric on the phase 2 test set. This shows the effectiveness of the image augmentation methods in improving the corruption robustness of the map segmentation model. **Note that** data augmentations used in our experiments are marked in gray. In addition, we have also listed some other data augmentation methods and hope to further explore them in the future, such as Translate, Scale, Rotate, Shear, etc.

4.5. Main Results

For the main experiments, we use BEVerse-tiny with the temporal module as the base model and use some data augmentation, such as GridMask, Random Hue, Random Saturation, Random Swap Channels and Random value. Tab. 5 shows the overall performance of our method. In a nutshell, our method shows significant superiority over other methods, indicating the benefit of the temporal module and data augmentations.

Note that due to limited time, we did not conduct related experiments on BEVerse-Base. We believe that experimenting with strong backbone models will further improve robustness performance.

5. Conclusion

In this paper, we explored several methods to improve the robustness of the map segmentation task. By conducting large-scale experiments, we draw some important findings, as summarized below:

- By analyzing the impact of different configuration options on corruption robustness, we find that the recipes for a robust map segmentation model may include utilizing a temporal fusion module and strong backbone (*e.g.*, Swin-Base Transformer).
- Some data augmentation methods are effective in improving the robustness of map segmentation models. Nonetheless, further investigation into more advanced augmentation methods is warranted for future research.

These novel findings allowed us to achieve promising results in the 2024 RoboDrive Challenge - Robust Map Segmentation Track and offer insights for designing more reliable map segmentation models.

References

- [1] Qi Li, Yue Wang, Yilun Wang, and Hang Zhao. Hdmapnet: An online hd map construction and evaluation framework. In *IEEE International Conference on Robotics and Automation*, pages 4628–4634, 2022.
- [2] Runnan Chen, Youquan Liu, Lingdong Kong, Xinge Zhu, Yuexin Ma, Yikang Li, Yuenan Hou, Yu Qiao, and Wenping Wang. Clip2scene: Towards label-efficient 3d scene understanding by clip. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7020–7030, 2023.
- [3] Youquan Liu, Lingdong Kong, Jun Cen, Runnan Chen, Wenwei Zhang, Liang Pan, Kai Chen, and Ziwei Liu. Segment any point cloud sequences by distilling vision foundation models. In *Advances in Neural Information Processing Systems*, volume 36, 2024.
- [4] Yunpeng Zhang, Zheng Zhu, Wenzhao Zheng, Junjie Huang, Guan Huang, Jie Zhou, and Jiwen Lu. Beverage: Unified perception and prediction in birds-eye-view for vision-centric autonomous driving. *arXiv preprint arXiv:2205.09743*, 2022.
- [5] Haimei Zhao, Qiming Zhang, Shanshan Zhao, Zhe Chen, Jing Zhang, and Dacheng Tao. Simdistill: Simulated multi-modal distillation for bev 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 7460–7468, 2024.
- [6] 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. *arXiv preprint arXiv:2405.17426*, 2024.
- [7] Lingdong Kong, Shaoyuan Xie, Hanjiang Hu, Yaru Niu, Wei Tsang Ooi, Benoit R. Cottureau, 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, Ziyang 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.
- [8] 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.
- [9] Lingdong Kong, Xiang Xu, Jiawei Ren, Wenwei Zhang, Liang Pan, Kai Chen, Wei Tsang Ooi, and Ziwei Liu. Multi-modal data-efficient 3d scene understanding for autonomous driving. *arXiv preprint arXiv:2405.05258*, 2024.
- [10] Lang Peng, Zhirong Chen, Zhangjie Fu, Pengpeng Liang, and Erkang Cheng. Bevsegformer: Bird’s eye view semantic segmentation from arbitrary camera rigs. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5924–5932, 2023.
- [11] Xiaoshuai Hao, Hui Zhang, Yifan Yang, Yi Zhou, Sangil Jung, Seung-In Park, and ByungIn Yoo. Mbfusion: A new multi-modal bev feature fusion method for hd map construction. In *IEEE International Conference on Robotics and Automation*, 2024.
- [12] Yinpeng Dong, Caixin Kang, Jinlai Zhang, Zijian Zhu, Yikai Wang, Xiao Yang, Hang Su, Xingxing Wei, and Jun Zhu. Benchmarking robustness of 3d object detection to common corruptions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1022–1032, 2023.
- [13] Zijian Zhu, Yichi Zhang, Hai Chen, Yinpeng Dong, Shu Zhao, Wenbo Ding, Jiachen Zhong, and Shibao Zheng. Understanding the robustness of 3d object detection with bird’s-eye-view representations in autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21600–21610, 2023.
- [14] Lingdong Kong, Yaru Niu, Shaoyuan Xie, Hanjiang Hu, Lai Xing Ng, Benoit R Cottureau, Ding Zhao, Liangjun Zhang, Hesheng Wang, Wei Tsang Ooi, et al. The robodepth challenge: Methods and advancements towards robust depth estimation. *arXiv preprint arXiv:2307.15061*, 2023.
- [15] 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.
- [16] Haimei Zhao, Jing Zhang, Zhuo Chen, Shanshan Zhao, and Dacheng Tao. Unimix: Towards domain adaptive and generalizable lidar semantic segmentation in adverse weather. *arXiv preprint arXiv:2404.05145*, 2024.
- [17] Jiawei Ren, Liang Pan, and Ziwei Liu. Benchmarking and analyzing point cloud classification under corruptions. In *International Conference on Machine Learning*, pages 18559–18575, 2022.
- [18] Zhiqi Li, Wenhai Wang, Hongyang Li, Enze Xie, Chonghao Sima, Tong Lu, Yu Qiao, and Jifeng Dai. Bevformer: Learning bird’s-eye-view representation from multi-camera images via spatiotemporal transformers. In *European Conference on Computer Vision*, pages 1–18, 2022.
- [19] Zhijian Liu, Haotian Tang, Alexander Amini, Xinyu Yang, Huiyi Mao, Daniela L Rus, and Song Han. Bevfusion: Multi-task multi-sensor fusion with unified bird’s eye view representation. In *IEEE International Conference on Robotics and Automation*, pages 2774–2781, 2023.
- [20] Brady Zhou and Philipp Krähenbühl. Cross-view transformers for real-time map-view semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13760–13769, 2022.
- [21] Yi Wei, Linqing Zhao, Wenzhao Zheng, Zheng Zhu, Yongming Rao, Guan Huang, Jiwen Lu, and Jie Zhou. Surround-depth: Entangling surrounding views for self-supervised

- multi-camera depth estimation. In *Conference on Robot Learning*, pages 539–549, 2023.
- [22] Yi Wei, Linqing Zhao, Wenzhao Zheng, Zheng Zhu, Jie Zhou, and Jiwen Lu. Surroundocc: Multi-camera 3d occupancy prediction for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21729–21740, 2023.
- [23] Yuanhui Huang, Wenzhao Zheng, Yunpeng Zhang, Jie Zhou, and Jiwen Lu. Tri-perspective view for vision-based 3d semantic occupancy prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9223–9232, 2023.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] Jiahao Sun, Xiang Xu, Lingdong Kong, Youquan Liu, Li Li, Chenming Zhu, Jingwei Zhang, Zeqi Xiao, Runnan Chen, Tai Wang, Wenwei Zhang, Kai Chen, and Chunmei Qing. An empirical study of training state-of-the-art lidar segmentation models. *arXiv preprint arXiv:2405.14870*, 2024.
- [28] 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 11618–11628, 2020.
- [29] Lingdong Kong, Shaoyuan Xie, Hanjiang Hu, Lai Xing Ng, Benoit Cottureau, and Wei Tsang Ooi. Robodepth: Robust out-of-distribution depth estimation under corruptions. *Advances in Neural Information Processing Systems*, 36, 2024.
- [30] Lingdong Kong, Shaoyuan Xie, Hanjiang Hu, Benoit Cottureau, Lai Xing Ng, and Wei Tsang Ooi. The robodepth benchmark for robust out-of-distribution depth estimation under corruptions. <https://github.com/ldkong1205/RoboDepth>, 2023.
- [31] Lingdong Kong, Yaru Niu, Shaoyuan Xie, Hanjiang Hu, Lai Xing Ng, Benoit Cottureau, Ding Zhao, 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.
- [32] Shaoyuan Xie, Lingdong Kong, Wenwei Zhang, Jiawei Ren, Liang Pan, Kai Chen, and Ziwei Liu. The robobev benchmark for robust bird’s eye view detection under common corruption and domain shift. <https://github.com/Daniel-xsy/RoboBEV>, 2023.
- [33] Lingdong Kong, Youquan Liu, Xin Li, Runnan Chen, Wenwei Zhang, Jiawei Ren, Liang Pan, Kai Chen, and Ziwei Liu. The robo3d benchmark for robust and reliable 3d perception. <https://github.com/ldkong1205/Robo3D>, 2023.
- [34] MMDetection3D Contributors. MMDetection3D: OpenMMLab next-generation platform for general 3D object detection. <https://github.com/open-mmlab/mmdetection3d>, 2020.
- [35] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- [36] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. Objects as points. *arXiv preprint arXiv:1904.07850*, 2019.
- [37] Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. *Sensors*, 18(10):3337, 2018.