MSC-Bench: Benchmarking and Analyzing Multi-Sensor Corruption for Driving Perception

Xiaoshuai Hao[1] Guanqun Liu[2] Yuting Zhao[3] Yuheng Ji[3] Mengchuan Wei[4] Haimei Zhao[5]
Lingdong Kong[6] Rong Yin[7] Yu Liu[8]
Beijing Academy of Artificial Intelligence[1] IQIYI[2] Institute of Automation, CAS[3] Samsung[4]
The University of Sydney[5] National University of Singapore[6]
Institute of Information Engineering, CAS[7] Hefei University of Technology[8]

***Abstract*—Multi-sensor fusion models play a crucial role in autonomous driving perception, particularly in tasks like 3D object detection and HD map construction. These models provide essential and comprehensive static environmental information for autonomous driving systems. While camera-LiDAR fusion methods have shown promising results by integrating data from both modalities, they often depend on complete sensor inputs. This reliance can lead to low robustness and potential failures when sensors are corrupted or missing, raising significant safety concerns. To tackle this challenge, we introduce the Multi-Sensor Corruption Benchmark (MSC-Bench), the first comprehensive benchmark aimed at evaluating the robustness of multi-sensor autonomous driving perception models against various sensor corruptions. Our benchmark includes 16 combinations of corruption types that disrupt both camera and LiDAR inputs, either individually or concurrently. Extensive evaluations of six 3D object detection models and four HD map construction models reveal substantial performance degradation under adverse weather conditions and sensor failures, underscoring critical safety issues. The benchmark toolkit and affiliated code and model checkpoints have been made publicly accessible[1].**

***Index Terms*—Autonomous Driving, Perception Robustness, 3D Object Detection, HD Map Construction, Multi-Sensor Corruption**

I. Introduction

The perception system is a critical component of autonomous vehicles, serving as the foundation for interaction between the vehicle and its driving environment. The system's performance—encompassing both accuracy and robustness—fundamentally influences the decision-making processes of autonomous vehicles. Notably, the robustness of perception algorithms is essential for the practical deployment of these vehicles, directly impacting the safety of future transportation systems for the general public.

Recently, researchers have developed fusion-based perception methods that integrate outputs from multiple sensors to enhance overall capabilities, leading to significant performance improvements across various tasks. For example, multi-sensor fusion approaches for 3D object detection and HD map construction have demonstrated superior accuracy compared to single-sensor methods that rely solely on cameras or LiDAR. However, these performance evaluations are typically conducted on clean datasets without any corruption, creating a gap in our understanding of the robustness of fusion-based perception methods under adverse conditions.

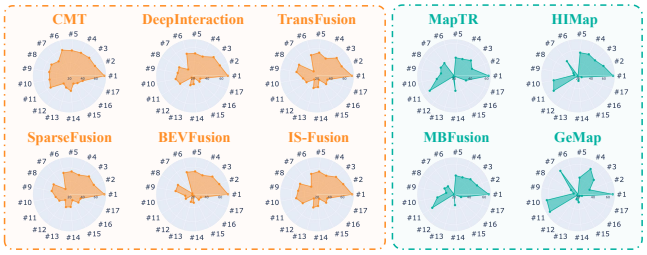

Fig. 1. Radar charts display the performance of state-of-the-art multi-sensor 3D object detection models (left) and HD map construction models (right) under the Multi-Sensor Corruption Benchmark (MSC-Bench). We present NDS scores for 3D object detection methods and mAP scores for map construction methods across each corruption type and severity level. **MSC-Bench:** #1 `Clean`, #2 `Motion Blur`, #3 `Temporal Misalignment`, #4 `Spatial Misalignment`, #5 `Fog`, #6 `Snow`, #7 `Camera Crash`, #8 `Frame Lost`, #9 `Cross Sensor`, #10 `Cross Talk`, #11 `Incomplete Echo`, #12 `Camera Crash & Cross Sensor`, #13 `Camera Crash & Cross Talk`, #14 `Camera Crash & Incomplete Echo`, #15 `Frame Lost & Cross Sensor`, #16 `Frame Lost & Cross Talk` and #17 `Frame Lost & Incomplete Echo`. The radius of each chart is normalized based on the `Clean` score. The larger the area coverage, the better the overall robustness.

The robustness of perception algorithms refers to their performance in adverse conditions, including challenging driving environments, complex scenarios, and sensor failures. Unlike single-sensor algorithms, multi-sensor perception systems face a broader range of issues, such as misalignment and synchronization problems. Additionally, adverse conditions like fog or snow can affect sensors differently. Understanding how these factors impact multi-sensor performance and whether sensor fusion can mitigate these effects is essential and requires thorough investigation.

In this paper, we introduce 16 types of corruption specific to multi-sensor perception algorithms and evaluate the robustness of fusion-based methods across two autonomous driving tasks: six 3D object detection methods and four HD map construction methods. Results, as shown in Fig. 1, reveal significant performance discrepancies between "clean" and corrupted datasets. Key findings include: 1) Camera-LiDAR fusion methods achieve strong performance by leveraging complementary information but often rely on complete sensor data, making them vulnerable to disruptions. 2) In 3D

[1]https://msc-bench.github.io/

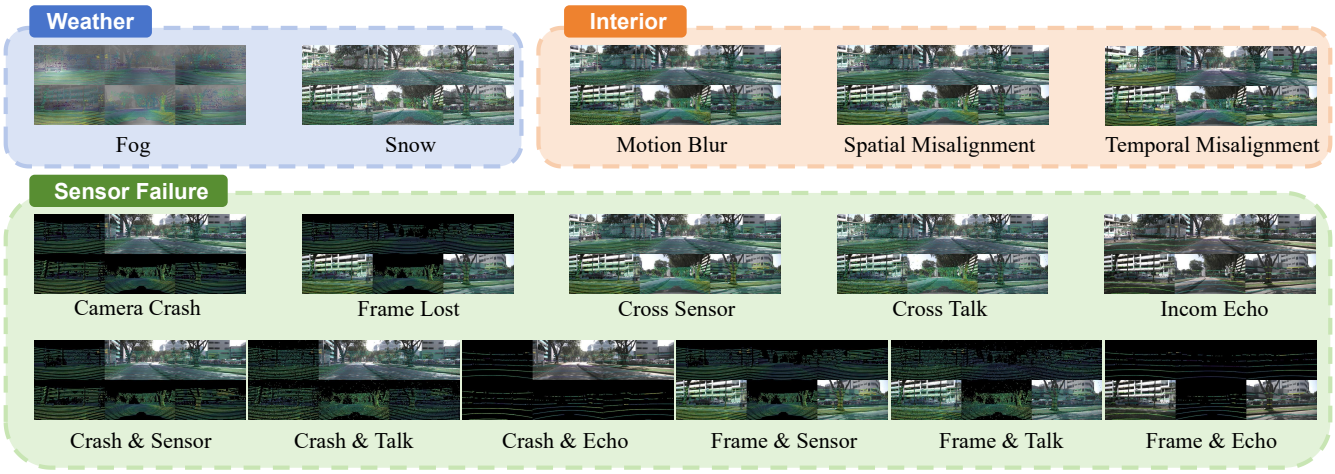

Fig. 2. **Overview of the MSC-Bench.** Definitions of the multi-sensor corruptions in MSC-Bench. Our benchmark encompasses a total of 16 corruption types for multi-modal perception models, which can be categorized into weather, interior, and sensor failure scenarios.

object detection, sensor failures and misalignments degrade performance, particularly under simultaneous disruptions like Frame Lost & Cross Sensor and Camera Crash & Cross Sensor, indicating a lack of adequate domain transfer and generalization. 3) In HD map construction, adverse weather, especially snow corruption, poses the greatest challenge by obscuring roads and reducing LiDAR reflectance, while Frame Lost & Cross Sensor conditions further emphasize the detrimental effects of dual-source information loss. In summary, both 3D object detection and HD map construction models are highly susceptible to sensor corruption, particularly from dual-source disruptions; future fusion models should enhance robustness against LiDAR variations and adapt to partial or missing camera data to improve reliability in real-world scenarios. Through extensive benchmark studies, we further unveil crucial factors for enhancing the reliability of multi-sensor perception models against sensor corruption. The key contributions of this work are three-fold:

- We introduce *Multi-Sensor Corruption Benchmark (MSC-Bench)*, making the first attempt to comprehensively benchmark and evaluate the robustness of multi-sensor autonomous driving perception models against various sensor corruptions.
- We analyze six 3D object detection models and four HD map construction models using MSC-Bench, offering valuable insights into design choices that enhance the robustness of multi-modal models.
- We will provide our data generation source code and benchmark, allowing for the reproducibility of the results presented in this study, which will serve as a valuable contribution to the field.

II. Related Work

Multi-Sensor 3D Object Detection The 3D object detection task focuses on identifying and localizing objects in three-dimensional space by predicting their 3D bounding boxes and categories using data from sensors like LiDAR and cameras, which is crucial for applications such as autonomous driving and robotics. While early methods relied on single sensors, the release of extensive autonomous driving datasets has spurred research into multi-sensor fusion for enhanced accuracy. Recent approaches include BEVFusion [1], which extracts features from both cameras and LiDAR using a Bird's-Eye View (BEV) space; DeepInteraction [2], which facilitates interactions between modality-specific representations; and TransFusion [3], which uses a transformer-based mechanism for adaptive fusion. Other methods, such as SparseFusion [4], utilize parallel detectors for instance-level fusion, while CMT [5] incorporates a Coordinates Encoding Module for position-aware features. IS-Fusion [6] further improves detection by integrating scene-level and instance-level fusion to enhance feature collaboration.

Multi-Sensor HD Map Construction The HD map construction task involves creating high-resolution maps that provide detailed vectorized representations of geometric and semantic information, such as lane boundaries and road structures, which are essential for accurate localization and path planning in autonomous driving. Recent camera-LiDAR fusion methods [7]–[10] leverage the semantic richness of camera data and the geometric precision of LiDAR. BEV-level fusion, which combines inputs from both sensors into a shared BEV space, has gained attention [1] for effectively integrating complementary features. However, existing methods often depend on complete sensor data, making them less robust to missing or corrupted information, which can lead to significant performance degradation. This paper focuses on evaluating the robustness of multi-modal HD map construction.

Driving Perception Robustness Researchers have recently focused on the robustness of various autonomous driving perception tasks. Studies like RoBoBEV [11] evaluate the robustness of BEV perception tasks, while others aim to develop more resilient models and strategies. Robo3D [12] benchmarks LiDAR-based semantic segmentation and 3D object detection under sensor failures. Zhu et al. [13] assess the natural and adversarial robustness of BEV models, introducing a 3D consistent patch attack for spatiotemporal realism. PointDR [14] and UniMix [15] propose domain-adaptive methods for

TABLE I

CORRUPTION METHODS OVERVIEW: TYPES, MODALITIES, DESCRIPTIONS, AND CONFIGURATIONS OF THREE SEVERITY LEVELS OF CORRUPTION.

Corruption	Modality	Description	Level 1	Level 2	Level 3
Camera Crash	C	Dropping view images	2	4	5
Frame Lost	C	Dropping temporal frames	2/6	4/6	5/6
Cross Sensor	L	Cross-sensor data by the number of beams to drop	8	16	20
Crosstalk	L	Light impulse interference by adjusting the percentag	0.03	0.07	0.12
Incomplete Echo	L	Incomplete LiDAR readings by adjusting the drop ratio	0.75	0.85	0.95
Temporal Misalignment	LC	Frozen frame applied with probability p	0.2	0.4	0.6
Spatial Misalignment	LC	Extrinsic misalignment in degrees applied with probability p	1°, 0.2	2°, 0.4	3°, 0.6
Motion Blur	LC	Jitter noise from a Gaussian distribution with σ_t	0.06	0.10	0.13
Fog	LC	Approximated visibility in meters	300 m	150 m	50 m
Snow	LC	Approximated snowfall intensity in mm/h	5 mm/h	35 mm/h	70 mm/h

enhancing 3D semantic segmentation in adverse conditions. MapBench [16] and Multi-corrupt [17] offer benchmarks for evaluating the robustness of HD map construction and 3D object detection, respectively. In contrast to previous work, we present a more comprehensive benchmark that incorporates multi-sensor corruptions for fusion-based autonomous driving perception models, covering both HD map construction and 3D object detection tasks.

III. BENCHMARKING MULTI-SENSOR CORRUPTION

A. Multi-Sensor Corruption Definition

The Multi-Sensor Corruption Benchmark (MSC-Bench) includes 16 corruption types, categorized into weather, interior, and sensor failure scenarios (see Fig. 2). It is constructed by corrupting the *val* set of nuScenes [18]. Definitions of the corruption types can be found in Tab. I, with additional details provided below.

- **Camera Crash**: Simulates continuous loss of images from certain viewpoints due to camera failure. Determine the level of corruption based on the number of dropped cameras. Note that this type of corruption applies only to camera sensors, while the LiDAR sensor remains clean.
- **Frame Lost**: Represents random frame loss to assess the model's resilience to intermittent data loss, with the corruption level determined by the probability of frame dropping. Note that this type of corruption applies only to camera sensors, while the LiDAR sensor remains clean.
- **Cross Sensor**: Arises due to the large variety of LiDAR sensor configurations (e.g., beam number, field-of-view, and sampling frequency). Determine the level of corruption based on the number of beams dropped. Note that this type of corruption applies only to LiDAR sensors, while the camera sensor remains clean.
- **Crosstalk**: Creates noisy points within the mid-range areas between two (or multiple) sensors, simulating interference. Determine the level of corruption by adjusting the percentage of light impulse interference. Note that this type of corruption applies only to LiDAR sensors, while the camera sensor remains clean.
- **Incomplete Echo**: Represents incomplete LiDAR readings in some scan echoes. The level of corruption is determined by adjusting the drop ratio of these readings. Note that this type of corruption applies only to LiDAR sensors, while the camera sensor remains clean.
- **Fog**: We use a fog simulator [19] to simulate LiDAR fog corruption. To maintain scene consistency between images and point clouds, we adapt the LiDAR fog parameters for the image fog generation process.
- **Snow**: We use a snow simulator [20] that models snow particles as opaque spheres and computes the reflection properties of wet surfaces, enabling us to corrupt the point cloud and image data based on snowfall levels.
- **Motion Blur**: To replicate intense motion, vibrations, and the rolling shutter effect, we introduce jitter noise from a Gaussian distribution with a standard deviation of σ_t into both point cloud and image data.
- **Spatial Misalignment**: We introduce translation and rotation misalignment, creating a spatial offset between point cloud and camera inputs. We adjust the rotation angle and the proportion of affected data based on the severity level.
- **Temporal Misalignment**: Timestamps from modalities like LiDAR and cameras are not always perfectly synchronized, so we introduce temporal misalignment to both the camera and point cloud data.
- **Camera Crash** & **Cross Sensor**: For the camera sensor, we apply Camera Crash corruption, while for the LiDAR sensor, we use Cross Sensor corruption.
- **Camera Crash** & **Cross Talk**: For the camera sensor, we use Camera Crash corruption, and for the LiDAR sensor, we apply Crosstalk corruption.
- **Camera Crash** & **Incomplete Echo**: For the camera sensor, we apply Camera Crash corruption, and for the LiDAR sensor, we use Incomplete Echo corruption.
- **Frame Lost** & **Cross Sensor**: For the camera sensor, we apply Frame Lost corruption, and for the LiDAR sensor, we use Cross Sensor corruption.
- **Frame Lost** & **Cross Talk**: For the camera sensor, we use Frame Lost corruption, and for the LiDAR sensor, we use Cross Talk corruption.
- **Frame Lost** & **Incomplete Echo**: For the camera sensor, we apply Frame Lost corruption, and for the LiDAR sensor, we use Incomplete Echo corruption.

TABLE II

BENCHMARKING 3D OBJECT DETECTION MODELS. WE REPORT DETAILED INFORMATION ON THE METHODS GROUPED BY [1] INPUT MODALITY, [2] BACKBONE, AND [3] INPUT IMAGE SIZE. "L" AND "C" REPRESENT LIDAR AND CAMERA, RESPECTIVELY. 'SWIN-T", "R50", "VOV-99", AND "SEC" ARE SHORT FOR SWIN-TRANSFORMER, RESNET50, VOVNET, AND SECOND. WE REPORT NUSCENES DETECTION SCORE (NDS) AND MEAN AVERAGE PRECISION (MAP) ON THE OFFICIAL NUSCENES VALIDATION SET.

Method	Venue	Modal	Backbone	Image Size	NDS↑	mAP↑	mRS↑	mRRS↑
BEVFusion [1]	ICRA'23	C & L	Swin-T & SEC	704×256	71.44	68.72	54.88	0.00
SparseFusion [4]	ICCV'23	C & L	Swin-T & SEC	800×448	73.15	71.02	60.11	17.01
TransFusion [3]	CVPR'22	C & L	R50 & SEC	800×448	70.84	66.72	60.12	12.30
DeepInteraction [2]	NIPS'22	C & L	R50 & SEC	800×448	69.09	68.72	59.01	6.93
CMT [5]	ICCV'23	C & L	VoV-99 & SEC	1600×640	72.90	70.30	67.17	32.93
Is-Fusion [6]	CVPR'24	C & L	Swin-T & SEC	1056×384	74.00	72.8	62.10	22.42

TABLE III

BENCHMARKING HD MAP CONSTRUCTORS. WE REPORT DETAILED INFORMATION ON THE METHODS GROUPED BY [1] INPUT MODALITY, [2] BEV ENCODER, [3] BACKBONE, AND [4] TRAINING EPOCHS. "L" AND "C" REPRESENT LIDAR AND CAMERA, RESPECTIVELY. "EFFI-B0," "R50," "PP," AND "SEC" REFER TO EFFICIENTNET-B0, RESNET50, POINTPILLARS, AND SECOND. AP DENOTES PERFORMANCE ON THE CLEAN NUSCENES *val* SET. THE SUBSCRIPTS $b.$, $p.$, AND $d.$ DENOTE *boundary*, *pedestrian crossing*, AND *divider*, RESPECTIVELY.

Method	Venue	Modal	Backbone	Epoch	$AP_{p.}$↑	$AP_{d.}$↑	$AP_{b.}$↑	mAP↑	mRS↑	mRSS↑
MapTR [10]	ICLR'23	C & L	R50 & SEC	24	55.9	62.3	69.3	62.5	55.91	0.00
MBFusion [21]	ICRA'24	C & L	R50 & SEC	24	61.6	64.4	72.5	66.1	50.83	−4.46
GeMap [22]	ECCV'24	C & L	R50 & SEC	24	66.3	62.2	71.1	66.5	55.25	4.77
HIMap [23]	CVPR'24	C & L	R50 & SEC	24	71.0	72.4	79.4	74.3	50.29	4.27

B. Robustness Evaluation Metrics

To compare the robustness of different 3D object detectors and HD map constructors in multi-modal corrupted scenarios, we introduce two robustness evaluation metrics.

Resilience Score (RS) We define RS as the relative robustness indicator for measuring how much accuracy a model can retain when evaluated on the corruption sets, which are calculated as follows:

$$\mathrm{RS}_i = \frac{\sum_{l=1}^{3} \mathrm{Acc}_{i,l}}{3 \times \mathrm{Acc}^{\mathrm{clean}}}, \quad \mathrm{mRS} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{RS}_i, \tag{1}$$

where $\mathrm{Acc}_{i,l}$ denotes the task-specific accuracy scores, with NDS (NuScenes Detection Score) for 3D object detection and mAP (mean Average Precision) for HD map construction, on corruption type i at severity level l. N is the total number of corruption types, and $\mathrm{Acc}^{\mathrm{clean}}$ denotes the accuracy score on the "clean" evaluation set. mRS (mean Resilience Score) represents the average score, providing an overall measure of the model's robustness across all types of corruption.

Relative Resilience Score (RRS) We define RRS as the critical metric for comparing the relative robustness of candidate models with the baseline model and mRRS as an overall metric to indicate the relative resilience score. The RRS and mRRS scores are calculated as follows:

$$\mathrm{RRS}_i = \frac{\sum_{l=1}^{3} \mathrm{Acc}_{i,l}}{\sum_{l=1}^{3} \mathrm{Acc}_{i,l}^{\mathrm{base}}} - 1, \quad \mathrm{mRRS} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{RRS}_i, \tag{2}$$

where $\mathrm{Acc}_{i,l}^{\mathrm{base}}$ denotes the accuracy score of the baseline model.

IV. EXPERIMENTS AND ANALYSIS

A. Benchmarking Multi-Sensor 3D Object Detection

Candidate models Our MSC-Bench includes a total of six multi-sensor 3D object detection models: CMT [5], DeepInteraction [2], TransFusion [3], SparseFusion [4], BEVFusion [1] and Is-Fusion [6]. We present the basic information for these models in Tab. II, including input modality, backbone, image size, and performance on the official nuScenes validation set.

3D Object Detection Benchmarking Analysis We present the overall robustness benchmarking results, including mRS and mRRS, for the six multi-sensor candidate models in Tab. II. The table shows that model robustness under corruption does not strongly correlate with performance on the clear validation set. For example, while Is-Fusion achieves the highest NDS and mAP scores, its mRS and $mRRS$ scores are below expectations. In contrast, CMT exhibits excellent robustness, achieving the highest robustness scores.

To analyze the models' robustness across different corruption types, we present the Resilience Scores for 16 corruption types in Tab. IV (top) and illustrate robustness performance across varying severity levels in Fig. 3. The data shows that sensor failure and misalignment-related corruptions, such as Cross Sensor, Camera Crash & Cross Sensor, and Frame Lost & Cross Sensor, significantly impact model performance. In contrast, individual sensor failures like Camera Crash, Frame Lost, and Incomplete Echo have minimal effects on robustness. However, when these failures occur simultaneously, as seen in Camera Crash & Incomplete Echo and Frame Lost & Incomplete Echo, model robustness is substantially compromised.

From Fig. 3, most corruption types lead to a linear decline in model robustness as severity increases. However, the robustness degradation from Camera Crash and Frame Lost is relatively minor, showing a distinct pattern compared to other corruptions. Notably, for Temporal Misalignment and Fog, robustness remains stable at severity levels 1 and 2 but drops dramatically at level 3 as severity intensifies.

Fig. 5 shows the relative resilience scores of different models based on BEVFusion. DeepInteraction underperforms the base model in eight corruption types, while only CMT surpasses the base model across all corruption types, achieving the best performance in 12 of them.

B. Benchmarking Multi-Sensor HD Map Construction

Candidate models Our MSC-Bench includes four multi-sensor HD map constructors: MapTR [10], HIMap [23], MBFusion [21], and GeMap [22]. We present the basic information for these models in Tab. III, including input modality, backbone, training epochs, and performance on the official nuScenes validation set.

HD map construction Benchmarking Analysis Tab. III presents the overall robustness performance of the four multi-sensor HD map construction models, measured by mRS and mRRS. MapTR and GeMap achieve comparable scores, outperforming MBFusion and HIMap. Performance on specific corruption types is detailed in Tab. IV (bottom), highlighting

TABLE IV

ROBUSTNESS BENCHMARK OF STATE-OF-THE-ART MULTI-MODAL METHODS UNDER MULTI-SENSOR CORRUPTIONS. FOR THE 3D OBJECT DETECTION TASK, WE USE NDS AS THE METRIC. ADDITIONALLY, WE USE MAP AS THE METRIC FOR THE HD MAP CONSTRUCTION TASK.

	Model	Motion Blur	Temporal Mis.	Spatial Mis.	Fog	Snow	Camera Crash	Frame Lost	Cross Sensor	Cross Talk	Incomplete Echo	Camera Crash, Cross Sensor	Camera Crash, Cross Talk	Camera Crash, Incomplete Echo	Frame Lost, Cross Sensor	Frame Lost, Cross Talk	Frame Lost, Incomplete Echo	mRS↑
3D Object Detection	CMT [5]	84.25	83.05	80.91	80.44	83.15	71.69	70.35	65.21	69.73	73.84	47.71	52.97	58.93	45.78	49.98	56.78	67.17
	DeepInteraction [2]	87.64	85.69	73.02	75.03	75.87	61.97	60.68	45.36	64.76	59.76	33.84	48.94	46.00	33.80	46.80	44.97	59.00
	TransFusion [3]	82.50	82.88	68.15	76.15	72.58	71.54	71.03	39.82	57.34	53.53	37.90	54.39	51.59	37.68	53.61	51.16	60.12
	SparseFusion [4]	81.61	82.25	71.93	74.79	73.42	66.28	65.12	41.26	56.69	52.21	42.27	54.75	53.52	41.10	52.50	52.10	60.11
	BEVFusion [1]	85.92	82.01	71.66	75.26	75.16	64.47	64.48	30.25	44.14	45.29	30.25	44.14	45.30	30.25	44.15	45.29	54.88
	Is-Fusion [6]	86.82	81.61	71.12	75.46	71.97	69.22	67.75	47.40	72.05	62.62	38.11	55.87	51.84	37.70	53.36	50.74	62.10
HD Map Construction	MapTR [10]	70.00	76.94	69.05	67.94	19.55	62.56	58.08	63.34	66.40	88.16	33.28	36.32	61.76	30.56	33.28	57.28	55.91
	HIMap [23]	83.77	74.93	77.31	75.56	23.79	38.09	35.26	65.28	78.47	86.41	19.25	27.59	37.55	19.38	27.19	34.86	50.29
	MBFusion [21]	79.69	74.96	68.19	67.97	23.57	52.60	46.25	50.26	64.13	81.56	25.47	30.97	51.92	22.68	27.67	45.47	50.83
	GeMap [22]	55.08	72.99	86.87	63.63	19.58	46.44	38.98	89.02	93.83	96.52	32.37	40.51	46.01	29.05	34.58	38.58	55.25

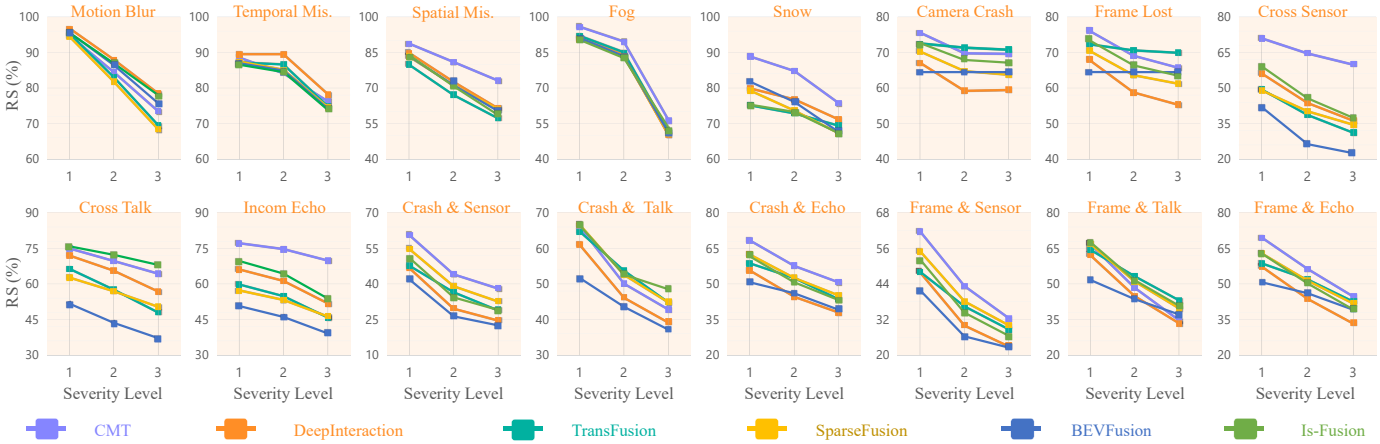

Fig. 3. Robustness against all corruption types and severity levels in 3D object detection tasks is evaluated through the Resilience Score (RS), calculated using the NDS score for varying severity levels.

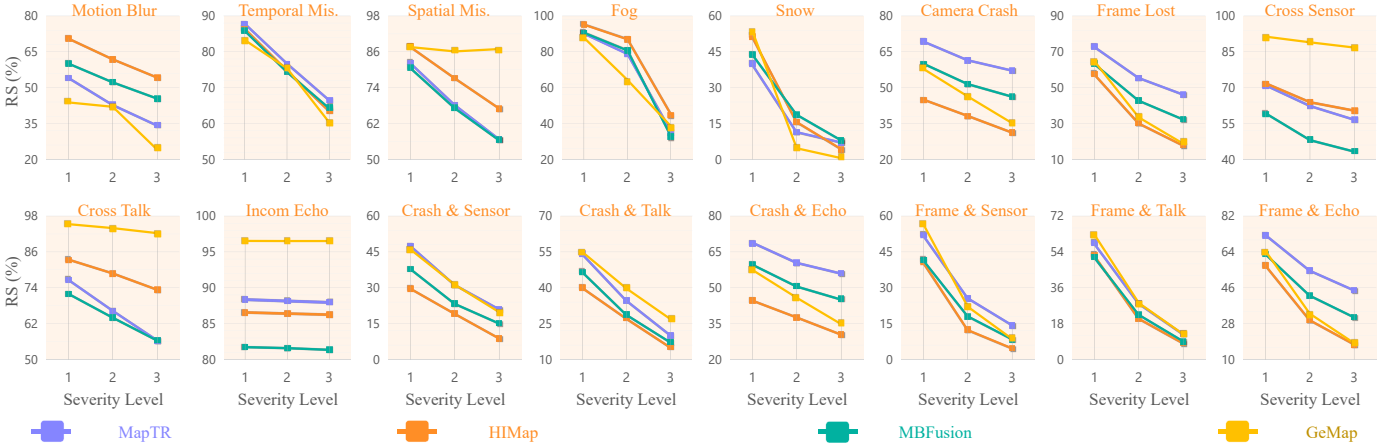

Fig. 4. Robustness against all corruption types and severity levels in HD map construction tasks is assessed using the Resilience Score (RS), calculated based on the mAP score for different severity levels.

that Snow severely impacts all models, reducing mAP to a range of 19 to 24. Additionally, combinations of sensor corruptions, such as Camera Crash & Cross Sensor, Camera Crash & Cross Talk, Frame Lost & Cross Sensor, and Frame Lost & Cross Talk, lead to significant performance degradation.

Fig. 4 illustrates how the performance of the four models changes as corruption severity increases, with most models showing a linear decline. Notably, variations in the severity of Incomplete Echo have a negligible impact on all four models. Among the models, GeMap and MapTR achieve the best results in eight corruption types, demonstrating very similar performance on the aforementioned combination corruptions. Fig. 6 shows the relative resilience scores of other HD map construction models compared to MapTR. GeMap, HIMap, and MBFusion underperform the base model in 5, 8, and 10 types of corruption, respectively. This disparity highlights the varying levels of robustness among these models, emphasizing the need for effective strategies to enhance resilience against diverse types of sensor corruption.

V. CONCLUSION

In this paper, we introduced the Multi-Sensor Corruption Benchmark (MSC-Bench) to assess the robustness of multi-sensor autonomous driving perception models under 16 types of corruption. Our analysis of six 3D object detection models and four HD map construction models revealed significant performance discrepancies between clean and corrupted datasets, highlighting vulnerabilities to sensor disruptions,

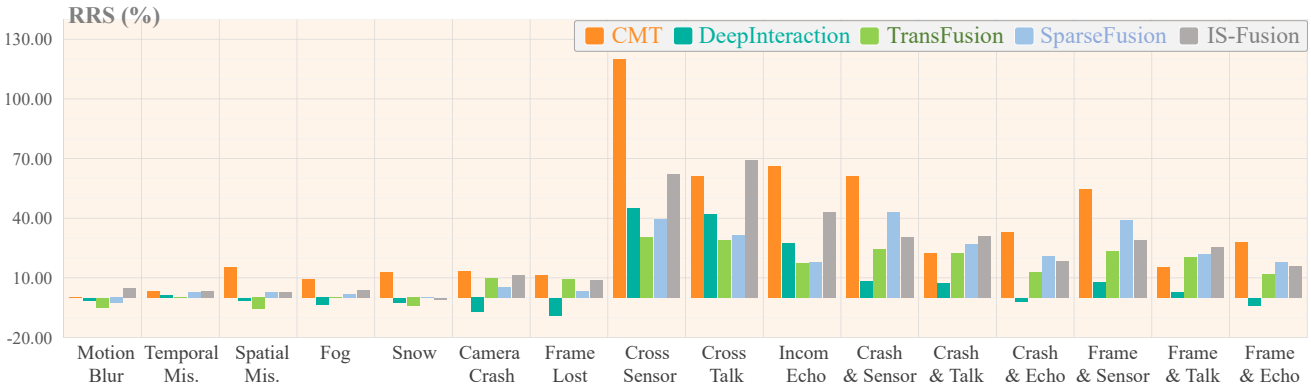

Fig. 5. Relative robustness visualization. Relative Resilience Score (RRS) computed with NDS using BEVFusion [1] as baseline.

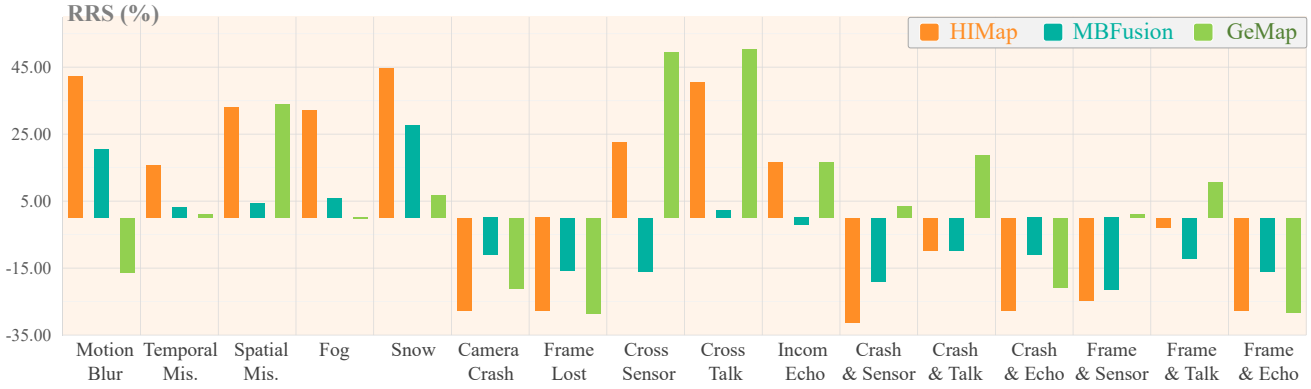

Fig. 6. Relative robustness visualization. Relative Resilience Score (RRS) computed with mAP using MapTR [10] as the baseline.

particularly dual-source failures like Frame Lost & Cross Sensor. While camera-LiDAR fusion methods demonstrated strong performance, they struggled with incomplete sensor data. Additionally, adverse weather conditions, especially snow, severely impacted HD map construction by obscuring critical elements. These findings underscore the need for more resilient fusion models that can effectively handle partial or missing sensor data and misalignment.

References

[1] Zhijian Liu, Tang, et al., "Bevfusion: Multi-task multi-sensor fusion with unified bird's eye view representation," in *ICRA*, 2023, pp. 2774–2781.

[2] Zeyu Yang, Jiaqi Chen, et al., "Deepinteraction: 3d object detection via modality interaction," 2022, pp. 1992–2005.

[3] Xuyang Bai, Zeyu Hu, et al., "Transfusion: Robust lidar-camera fusion for 3d object detection with transformers," in *CVPR*, 2022, pp. 1080–1089.

[4] Yichen Xie, Chenfeng Xu, et al., "Sparsefusion: Fusing multi-modal sparse representations for multi-sensor 3d object detection," in *CVPR*, 2023, pp. 17591–17602.

[5] Junjie Yan, Yingfei Liu, et al., "Cross modal transformer via coordinates encoding for 3d object dectection," *arXiv preprint arXiv:2301.01283*, 2023.

[6] Junbo Yin, Jianbing Shen, et al., "Is-fusion: Instance-scene collaborative fusion for multimodal 3d object detection," in *CVPR*, 2024, pp. 14905–14915.

[7] Qi Li, Yue Wang, et al., "Hdmapnet: An online hd map construction and evaluation framework," in *ICRA*, 2022, pp. 4628–4634.

[8] Xiaoshuai Hao, Ruikai Li, et al., "Mapdistill: Boosting efficient camera-based hd map construction via camera-lidar fusion model distillation," in *ECCV*, 2025, pp. 166–183.

[9] Yicheng Liu, Tianyuan Yuan, et al., "Vectormapnet: End-to-end vectorized hd map learning," in *ICML*, 2023, pp. 22352–22369.

[10] Bencheng Liao, Shaoyu Chen, et al., "Maptr: Structured modeling and learning for online vectorized hd map construction," in *ICLR*, 2023.

[11] Shaoyuan Xie, Lingdong Kong, et al., "Robobev: Towards robust bird's eye view perception under corruptions," *arXiv preprint arXiv:2304.06719*, 2023.

[12] Lingdong Kong, Youquan Liu, et al., "Robo3d: Towards robust and reliable 3d perception against corruptions," in *ICCV*, 2023, pp. 19994–20006.

[13] Zijian Zhu, Yichi Zhang, et al., "Understanding the robustness of 3d object detection with bird's-eye-view representations in autonomous driving," in *CVPR*, 2023, pp. 21600–21610.

[14] Aoran Xiao, Jiaxing Huang, et al., "3d semantic segmentation in the wild: Learning generalized models for adverse-condition point clouds," in *CVPR*, 2023, pp. 9382–9392.

[15] Haimei Zhao, Jing Zhang, et al., "Unimix: Towards domain adaptive and generalizable lidar semantic segmentation in adverse weather," in *CVPR*, 2024, pp. 14781–14791.

[16] Xiaoshuai Hao, Mengchuan Wei, et al., "Is your hd map constructor reliable under sensor corruptions?," *NeurIPS*, 2024.

[17] Till Beemelmanns, Quan Zhang, et al., "Multicorrupt: A multi-modal robustness dataset and benchmark of lidar-camera fusion for 3d object detection," in *IIV*, 2024, pp. 3255–3261.

[18] Holger Caesar, Varun Bankiti, and ohters, "nuscenes: A multimodal dataset for autonomous driving," in *CVPR*, 2020, pp. 11618–11628.

[19] Martin Hahner, Christos Sakaridis, et al., "Fog simulation on real lidar point clouds for 3d object detection in adverse weather," in *ICCV*, 2021, pp. 15283–15292.

[20] Martin Hahner, Christos Sakaridis, et al., "Lidar snowfall simulation for robust 3d object detection," in *CVPR*, 2022, pp. 16364–16374.

[21] Xiaoshuai Hao, Hui Zhang, et al., "Mbfusion: A new multi-modal bev feature fusion method for hd map construction," in *ICRA*, 2024, pp. 15922–15928.

[22] Zhixin Zhang, Yiyuan Zhang, et al., "Online vectorized hd map construction using geometry," in *ECCV*, 2024, pp. 73–90.

[23] Yi Zhou, Zhang, et al., "Himap: Hybrid representation learning for end-to-end vectorized hd map construction," in *CVPR*, 2024, pp. 15396–15406.