Is Your HD Map Constructor Reliable under Sensor Corruptions?

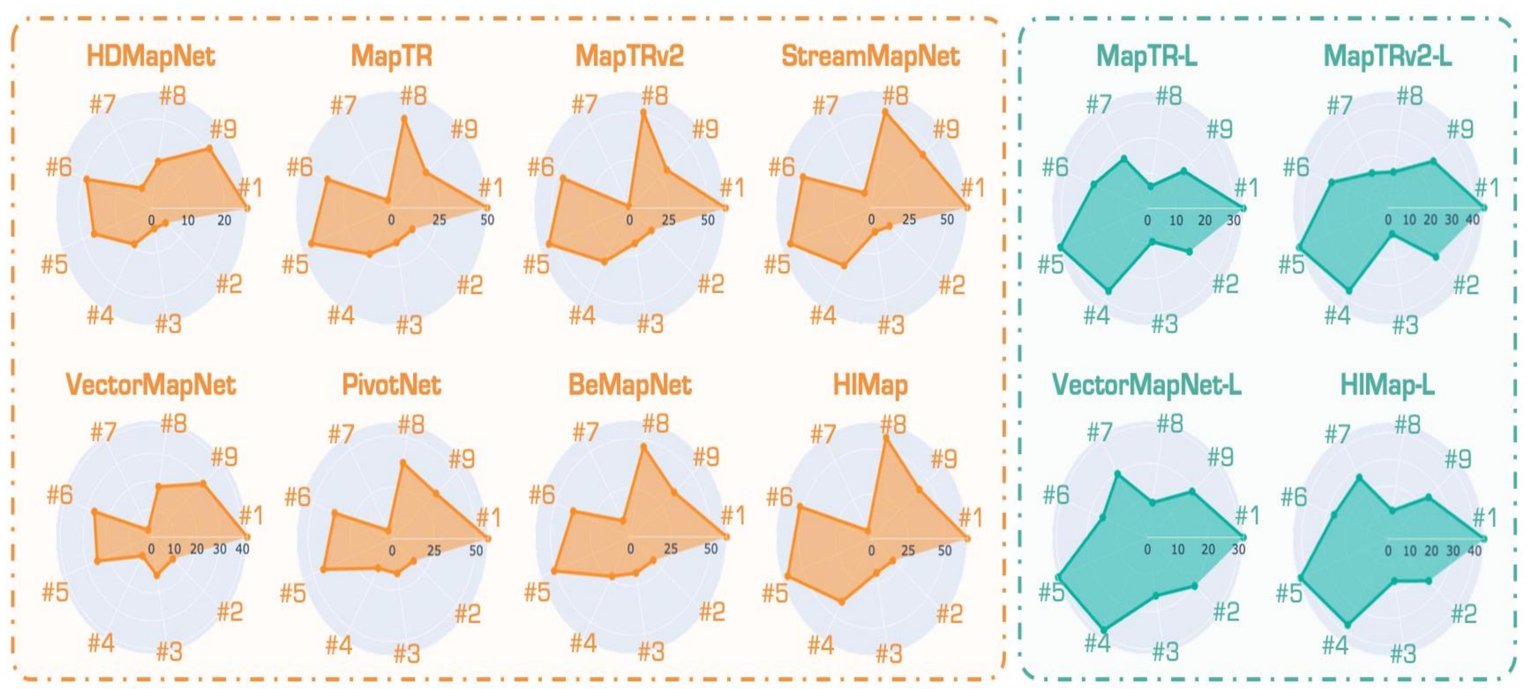
Xiaoshuai Hao, Mengchuan Wei, Yifan Yang, Haimei Zhao, Hui Zhang, Yi Zhou, Qiang Wang, Weiming Li, Lingdong Kong *, Jing Zhang *

SAMSUNG
NUS National University of Singapore
The University of Sydney
Neural Information Processing Systems

Motivation & Contribution

TL;DR

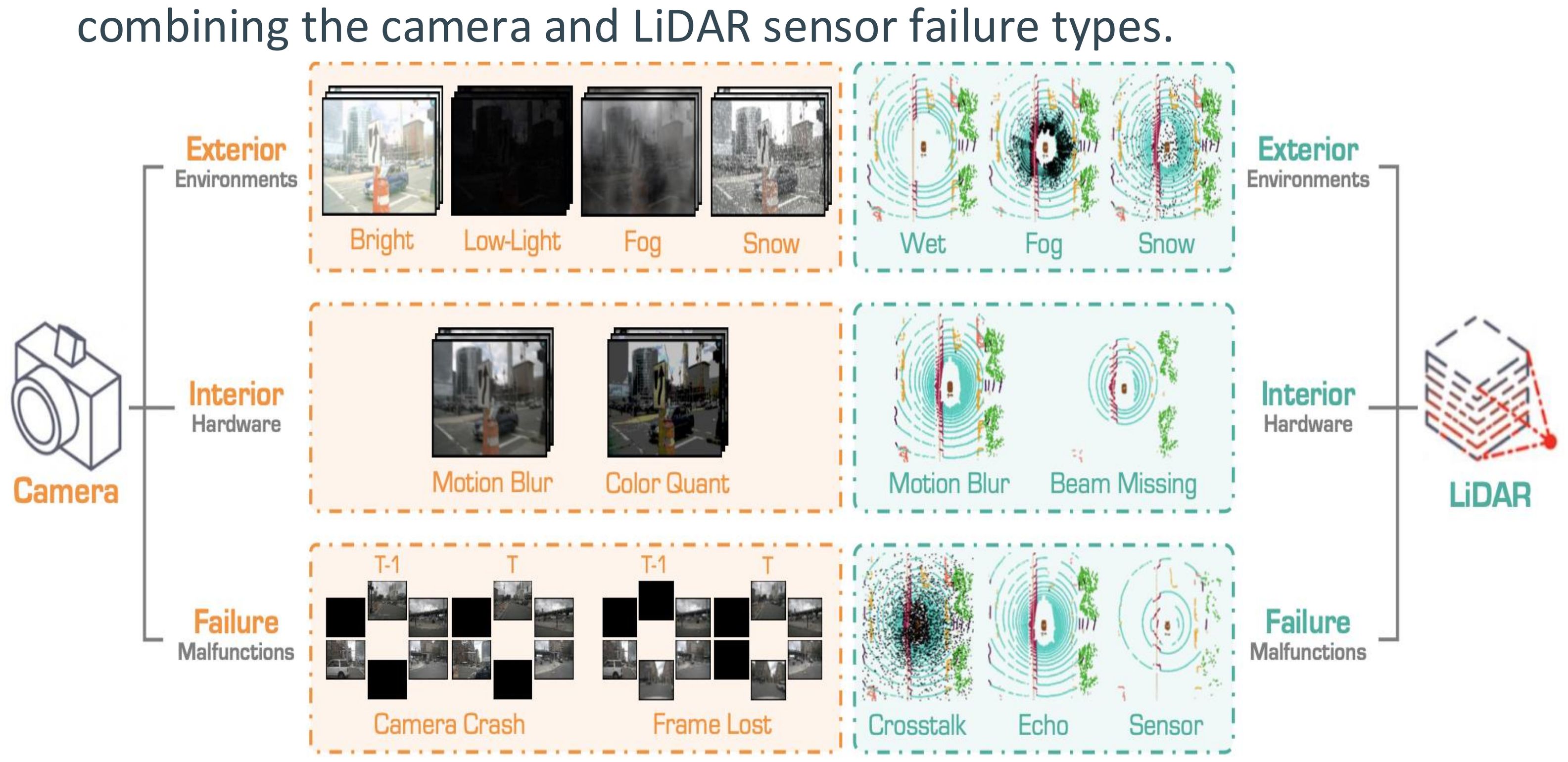
- We introduce MapBench, making the first attempt to comprehensively benchmark and evaluate the robustness of HD map construction models against various sensor corruptions.
- We extensively benchmark a total of 31 state-of-the-art HD map constructors and their variants under three configurations: camera-only, LiDAR-only, and camera-LiDAR fusion. This involves studying their robustness to 8 types of camera corruptions, 8 types of LiDAR corruptions, and 13 types of camera-LiDAR corruption combinations for each configuration.



- We identify effective strategies for enhancing robustness, including innovative approaches that leverage advanced data augmentation and architectural techniques. Our findings reveal strategies that significantly improve performance and robustness, underscoring the importance of tailored solutions to address specific challenges in HD map construction.

MapBench: Benchmarking HD Map Construction Robustness

- Definitions of the Camera and LiDAR sensor corruptions in MapBench. Our benchmark encompasses a total of 16 corruption types for HD map construction, which can be categorized into exterior, interior, and sensor failure scenarios. Besides, we define 13 multi-sensor corruptions by combining the camera and LiDAR sensor failure types.



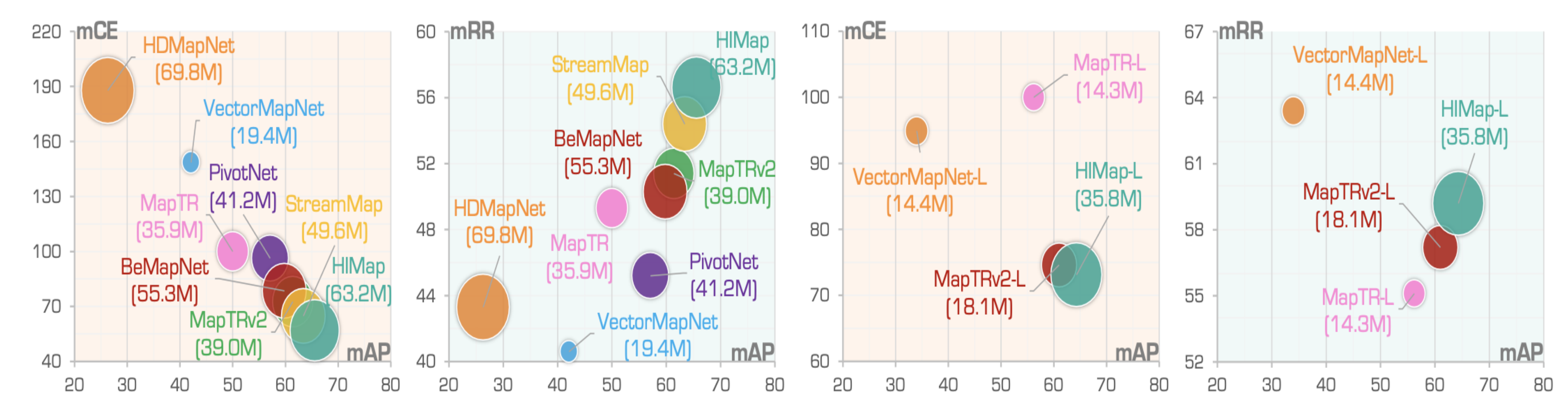
Experimental Analysis

Benchmarking HD map constructors

- We report the basic information of different models in Tab. 1.

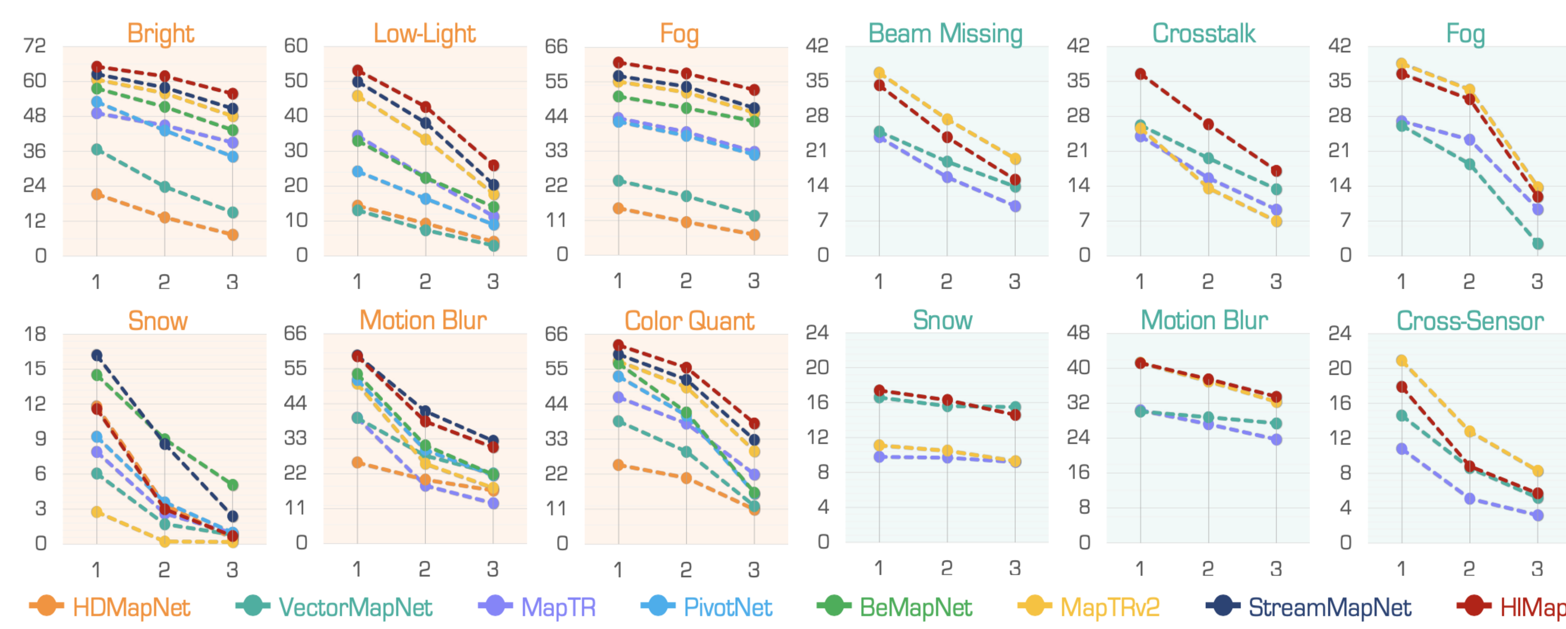
Method	Venue	Modal	BEV Encoder	Backbone	Epoch	$AP_{p.}$↑	$AP_{d.}$↑	$AP_{b.}$↑	mAP↑	mRR↑	mCE↓
HDMapNet [35]	ICRA'22	C	NVT	Effi-B0	30	14.4	21.7	33.0	23.0	43.3	187.8
VectorMapNet [43]	ICML'23	C	IPM	R50	110	36.1	47.3	39.3	40.9	40.6	148.5
PivotNet [11]	ICCV'23	C	PersFormer	R50	30	53.8	58.8	59.6	57.4	45.2	96.3
BeMapNet [50]	CVPR'23	C	IPM-PE	R50	30	57.7	62.3	59.4	59.8	50.3	78.5
MapTR [41]	ICLR'23	C	GKT	R50	24	46.3	51.5	53.1	50.3	49.3	100.0
MapTRv2 [42]	arXiv'23	C	BEVPool	R50	24	59.8	62.4	62.4	61.5	51.4	72.6
StreamMapNet [73]	WACV'24	C	BEVFormer	R50	30	61.7	66.3	62.1	63.4	54.4	64.8
HIMap [77]	CVPR'24	C	BEVFormer	R50	24	62.2	66.5	67.9	65.5	56.6	56.9
VectorMapNet [43]	ICML'23	L	-	PP	110	25.7	37.6	38.6	34.0	63.4	94.9
MapTR [41]	ICLR'23	L	-	SEC	24	48.5	53.7	64.7	55.6	55.1	100.0
MapTRv2 [42]	arXiv'23	L	-	SEC	24	56.6	58.1	69.8	61.5	57.2	74.6
HIMap [77]	CVPR'24	L	-	SEC	24	54.8	64.7	73.5	64.3	59.2	73.1
MapTR [41]	ICLR'23	C & L	GKT	R50 & SEC	24	55.9	62.3	69.3	62.5	57.1	100.0
HIMap [77]	CVPR'24	C & L	BEVFormer	R50 & SEC	24	71.0	72.4	79.4	74.3	41.7	110.6

- The correlations of accuracy (mAP) and robustness (mCE/mRR) for the Camera (a) and (b) and LiDAR (c) and (d) models.

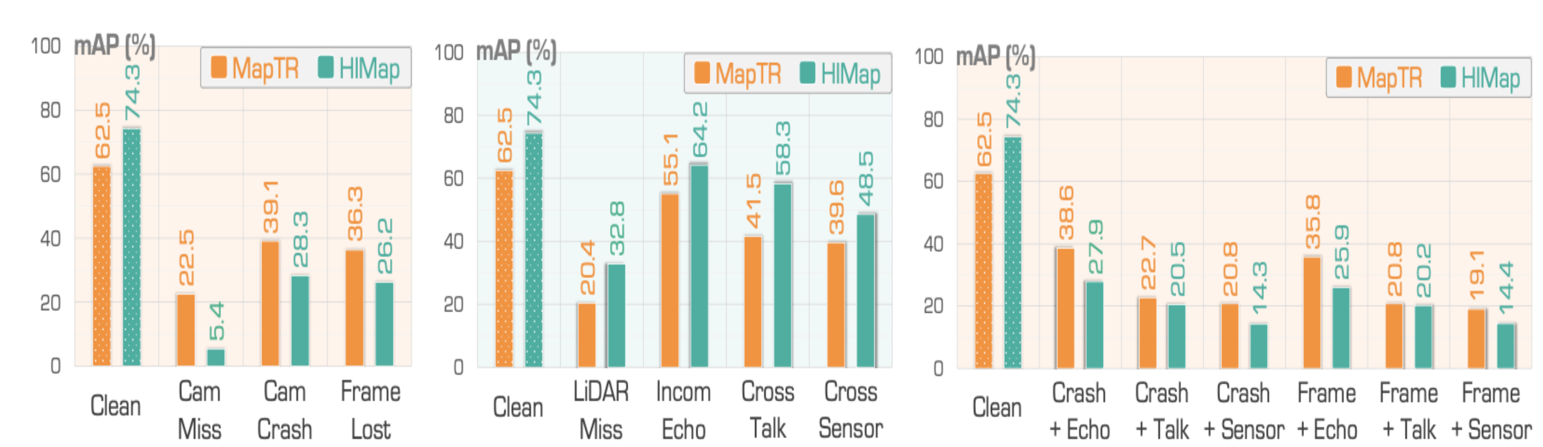


(a) mAP *vs.* mCE (b) mAP *vs.* mRR (c) mAP *vs.* mCE (d) mAP *vs.* mRR

- The mAP metrics of state-of-the-art HD map constructors under each of the three severity levels (Esay, Moderate, and Hard) in different Camera and LiDAR sensor corruption scenarios.



- Camera-LiDAR Fusion Benchmarking Results



(a) Corrupt-C + Clean-L (b) Corrupt-L + Clean-C (c) Corrupt-C + Corrupt-L

Observation & Discussion

Ablation on the use of BEV encoders

Method	Encode	$AP_{p.}$	$AP_{d.}$	$AP_{b.}$	mAP	mRR	mCE
MapTR ○	BEVFormer	43.7	49.8	52.6	48.7	49.3	100.0
MapTR ○	BEVPool	44.9	51.9	53.5	50.1	48.1	99.3
MapTR ●	GKT	**46.3**	**51.5**	**53.1**	**50.3**	**49.3**	**97.2**

Ablation on the use of temporal fusion

Method	Temp	$AP_{p.}$	$AP_{d.}$	$AP_{b.}$	mAP	mRR	mCE
StreamMap ○	✗	17.2	22.6	31.6	23.8	47.1	100.0
StreamMap ●	✓	**21.4**	**27.4**	**35.2**	**28.0**	**55.5**	**85.9**

Ablation on the use of backbone nets

Method	Back	$AP_{p.}$	$AP_{d.}$	$AP_{b.}$	mAP	mRR	mCE
PivotNet ○	R50	53.8	58.8	59.6	57.4	45.2	100.0
PivotNet ○	Effi-B0	53.9	59.7	61.0	58.2	49.9	87.4
PivotNet ●	SwinT	**58.7**	**63.8**	**64.9**	**62.5**	**50.8**	**77.8**
BeMapNet ○	R50	57.7	62.3	59.4	59.8	50.3	100.0
BeMapNet ○	Effi-B0	56.0	62.2	59.0	59.1	53.9	94.0
BeMapNet ●	SwinT	**61.3**	**64.4**	**61.6**	**62.5**	**57.9**	**75.9**

Ablation on different training epochs

Method	Epoch	$AP_{p.}$	$AP_{d.}$	$AP_{b.}$	mAP	mRR	mCE
MapTR ○	24	46.3	51.5	53.1	50.3	49.3	100.0
MapTR ●	110	**56.2**	**59.8**	**60.1**	**58.7**	**49.3**	**80.9**
PivotNet ○	30	58.7	63.8	64.9	62.5	50.8	100.0
PivotNet ●	110	**62.6**	**68.0**	**69.7**	**66.8**	**49.9**	**90.2**
BeMapNet ○	30	61.3	64.4	61.6	62.5	57.9	100.0
BeMapNet ●	110	**64.6**	**68.9**	**67.5**	**67.0**	**56.7**	**89.2**

Efficacy of Camera-based data augmentation techniques

Method	$AP_{p.}$	$AP_{d.}$	$AP_{b.}$	mAP	mRR	mCE
None	45.6	50.1	52.3	49.3	41.1	100.0
Rotate [37]	44.6	50.5	54.0	49.7	38.1	105.1
Flip [37]	44.7	53.0	53.4	50.4	38.7	102.5
PhotoMetric [33]	**46.3**	**51.5**	**53.1**	**50.3**	**49.3**	**84.5**

Efficacy of LiDAR-based data augmentation techniques

Method	$AP_{p.}$	$AP_{d.}$	$AP_{b.}$	mAP	mRR	mCE
None	26.6	31.7	41.8	33.4	55.1	100.0
Dropout [9]	28.4	31.0	42.5	33.9	56.9	98.9
RTS-LiDAR [23]	28.3	32.7	44.1	35.0	57.0	94.0
PolarMix [61]	**30.1**	**33.0**	**46.1**	**36.4**	**55.2**	**93.5**