TUE-AM-097

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Benchmarking Robustness of 3D Object Detection to Common Corruptions in Autonomous Driving

Yinpeng Dong[1,5], Caixin Kang[2], Jinlai Zhang[3], Zijian Zhu[4], Yikai Wang[1], Xiao Yang[1], Hang Su[1,6], Xingxing Wei[2], Jun Zhu[1,5,6]

[1]Dept. of Comp. Sci. and Tech., Institute for AI, Tsinghua-Bosch Joint ML Center, Tsinghua University

[2]Institute of Artificial Intelligence, Beihang University [3]Guangxi University [4]Shanghai Jiao Tong University

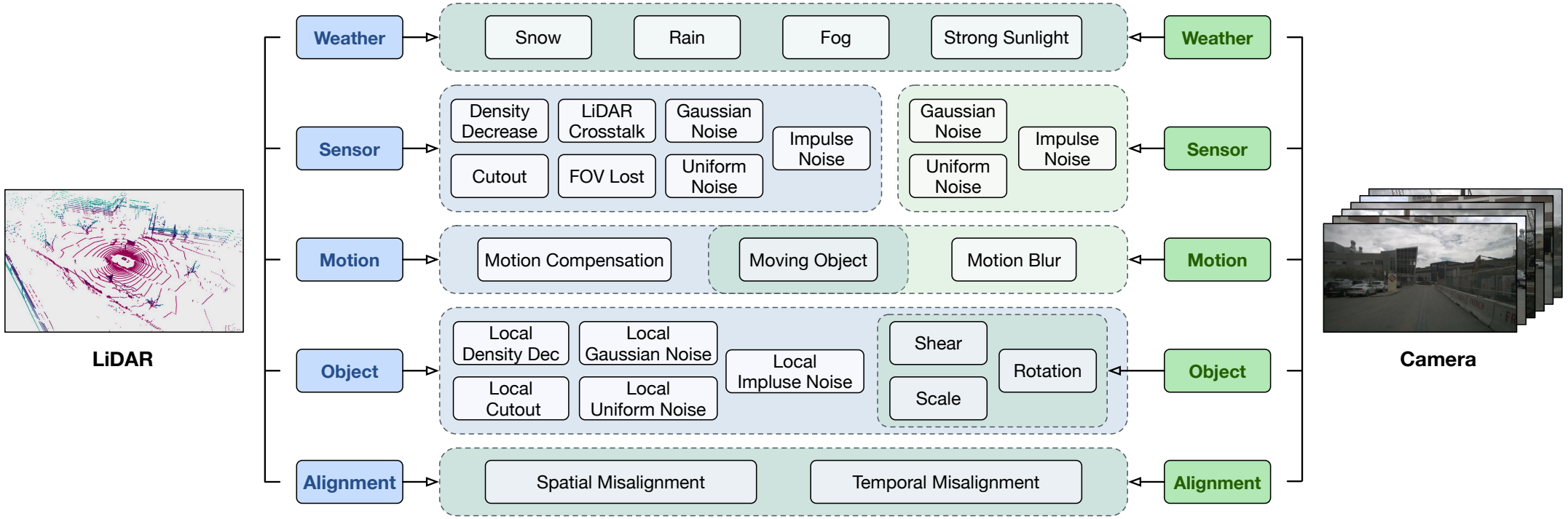
[5]RealAI [6]Peng Cheng Laboratory, Pazhou Laboratory (Huangpu)

Project Page: *https://autodrive-corruption.github.io*

Github: *https://github.com/thu-ml/3D_Corruptions_AD*

Overview

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

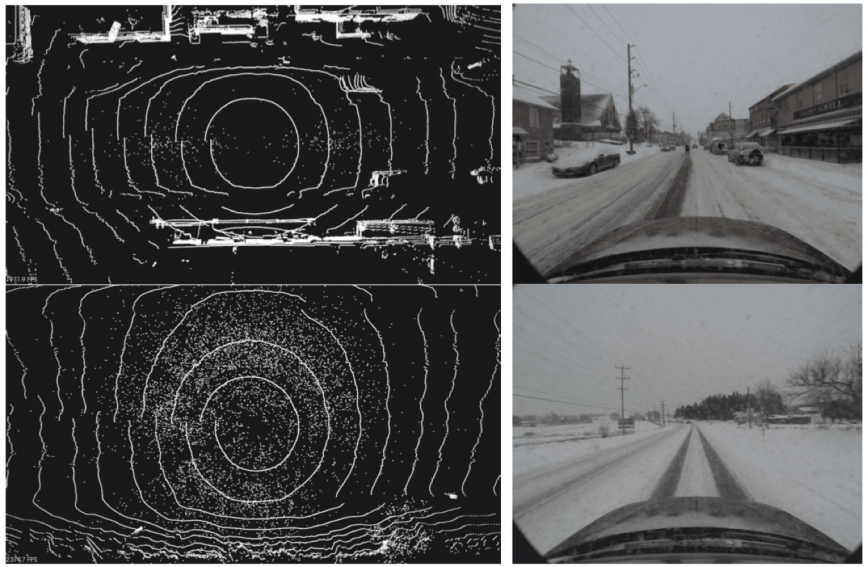


We build corruption robustness benchmarks of **27 corruption types** for 3D object detection in autonomous driving.

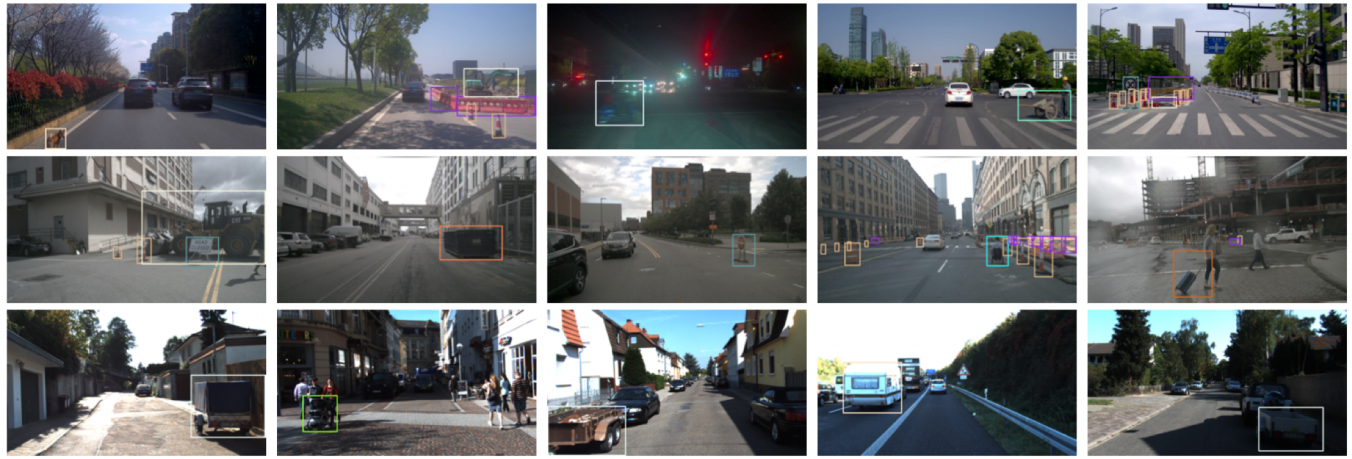
Background

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Autonomous driving may encounter **real-world corruptions** caused by adverse weathers, sensor noises, uncommon objects, etc., leading to inferior performance and causing safety problems. The existing datasets are **not comprehensive** enough due to **high collection costs of rare data**.



Adverse weather (Pitropov et al., 2020)



Corner cases (Li et al., 2022)

Background

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

There are existing benchmarks to evaluate the corruption robustness on image classification and point cloud recognition. But they do not consider the **real-world scenarios** in autonomous driving.

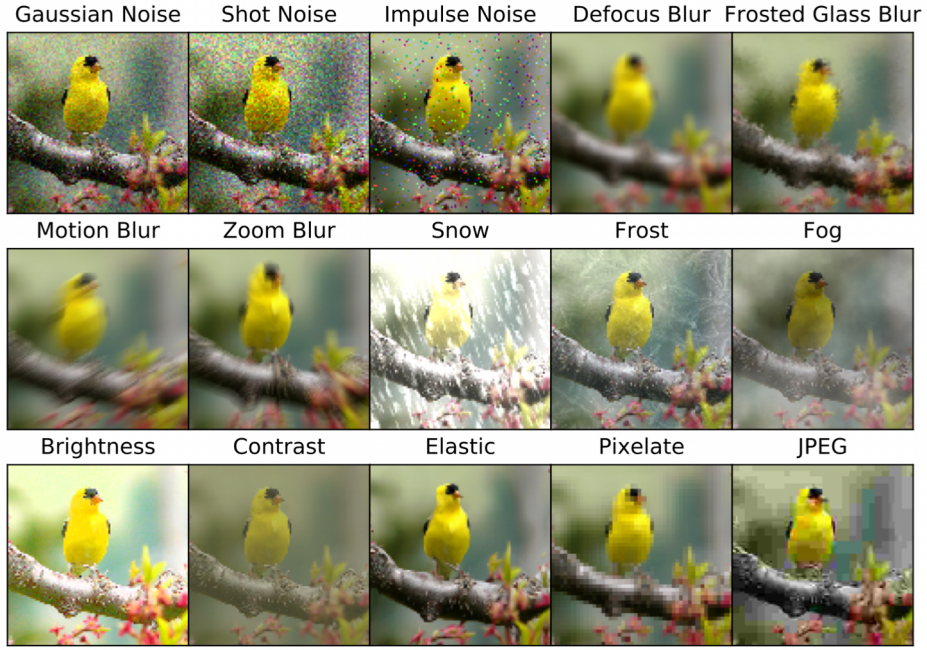
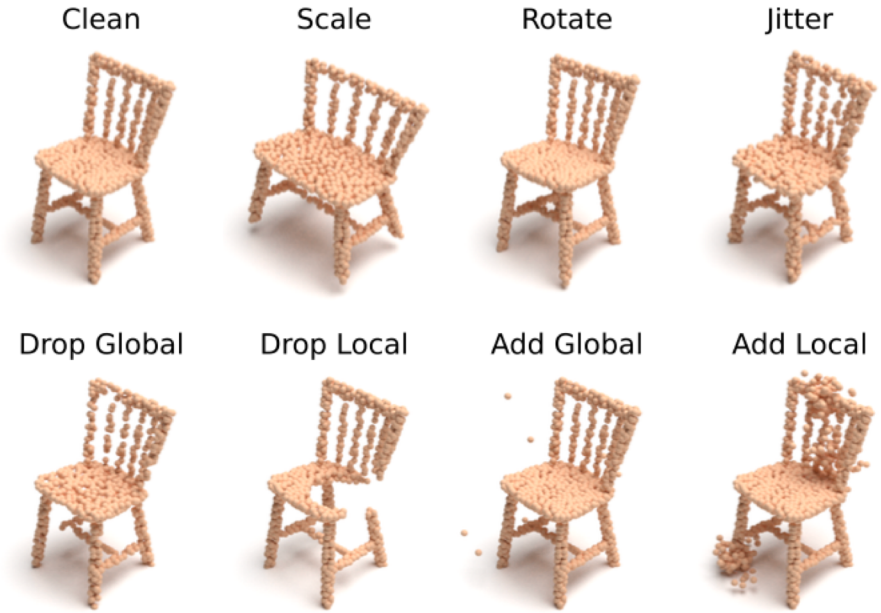


Image corruptions (Hendrycks and Dietterich, 2019)



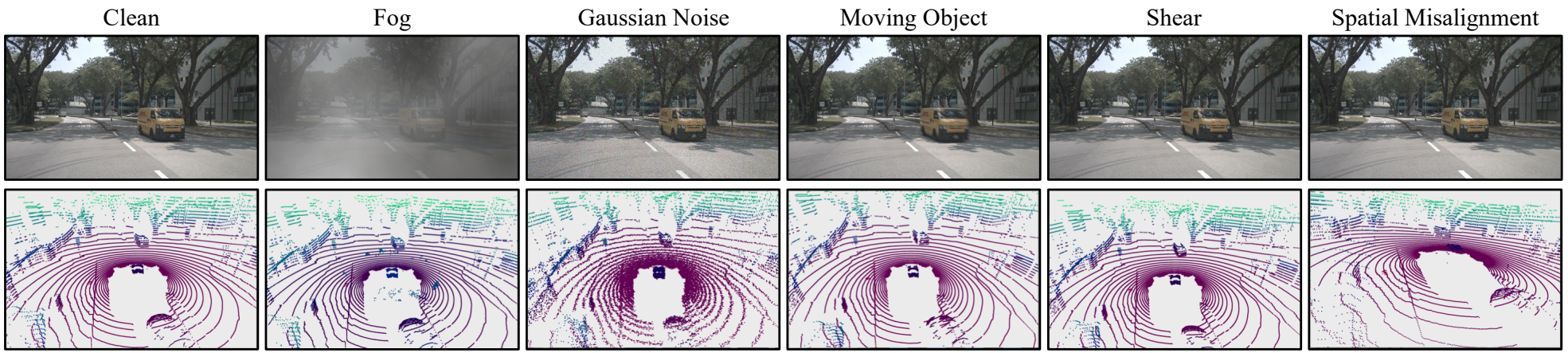
Point cloud corruptions (Ren et al., 2022)

Corruptions

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

We design 27 types of common corruptions for both LiDAR and camera inputs considering real-world driving scenarios.

- Weather-level: *Snow, Rain, Fog, Strong sunlight*
- *Sensor-level: Density decrease, LiDAR crosstalk, FOV lost, etc.*
- *Motion-level: Motion compensation, Moving objects, Motion blur*
- *Object-level: Local density decrease, Shear, Scale, Rotation, etc.*
- *Alignment-level: Spatial misalignment, Temporal misalignment*



Benchmarks

- KITTI-C:

$$\mathrm{AP}_{\mathrm{cor}} = \frac{1}{|C|}\sum_{c\in C}\frac{1}{5}\sum_{s=1}^{5}\mathrm{AP}_{c,s}\,; \quad \mathrm{RCE} = \frac{\mathrm{AP}_{\mathrm{clean}} - \mathrm{AP}_{\mathrm{cor}}}{\mathrm{AP}_{\mathrm{clean}}}$$

  - $\mathrm{AP}_{c,s}$ denotes the performance under corruption c at severity level s.
  - $\mathrm{AP}_{\mathrm{clean}}$ denotes the performance on clean dataset.
- nuScenes-C: we measure $\mathrm{mAP}_{\mathrm{cor}}$ and $\mathrm{NDS}_{\mathrm{cor}}$.
- Waymo-C: we measure $\mathrm{mAP}_{\mathrm{cor}}$ and $\mathrm{mAPH}_{\mathrm{cor}}$.

Benchmarks

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Model	Modality	Representation	Detection
SECOND [60]	LiDAR-only	voxel-based	one-stage
PointPillars [29]	LiDAR-only	voxel-based	one-stage
PointRCNN [48]	LiDAR-only	point-based	two-stage
Part-A^2 [49]	LiDAR-only	voxel-based	two-stage
PV-RCNN [47]	LiDAR-only	point-voxel-based	two-stage
3DSSD [61]	LiDAR-only	point-based	one-stage
SMOKE [36]	camera-only	monocular	one-stage
PGD [55]	camera-only	monocular	one-stage
ImVoxelNet [45]	camera-only	monocular	one-stage
EPNet [26]	fusion	point-level	two-stage
Focals Conv [13]	fusion	point-level	two-stage

(a) Evaluated models on KITTI-C.

Model	Modality	Representation	Detection
PointPillars [29]	LiDAR-only	voxel-based	one-stage
SSN [70]	LiDAR-only	voxel-based	one-stage
CenterPoint [62]	LiDAR-only	voxel-based	two-stage
FCOS3D [56]	camera-only	monocular	one-stage
PGD [55]	camera-only	monocular	one-stage
DETR3D [58]	camera-only	multi-view	query-based
BEVFormer [33]	camera-only	multi-view	query-based
FUTR3D [12]	fusion	proposal-level	query-based
TransFusion [2]	fusion	proposal-level	query-based
BEVFusion [35]	fusion	unified	query-based

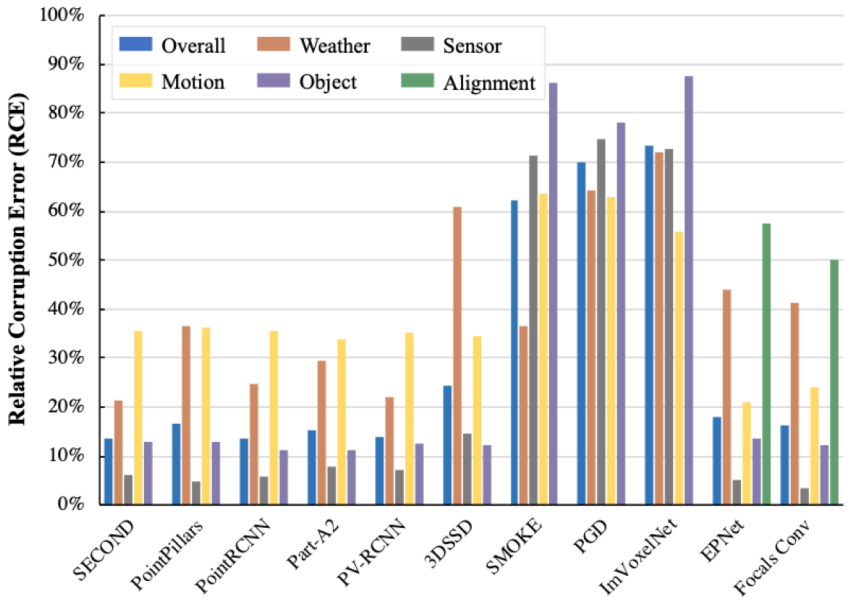
(b) Evaluated models on nuScenes-C.

We select representative 3D object detection models including LiDAR-only, camera-only, and LiDAR-camera fusion models.

Evaluation on KITTI-C

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Corruption		LiDAR-only						Camera-only			LC Fusion	
		SECOND	PointPillars	PointRCNN	Part-A^2	PV-RCNN	3DSSD	SMOKE	PGD	ImVoxelNet	EPNet	Focals Conv
None (AP_{clean})		81.59	78.41	80.57	82.45	**84.39**	80.03	7.09	8.10	**11.49**	82.72	**85.88**
Weather	Snow	**52.34**	36.47	50.36	42.70	**52.35**	27.12	2.47	0.63	0.22	34.58	34.77
	Rain	**52.55**	36.18	51.27	41.63	51.58	26.28	3.94	3.06	1.24	36.27	41.30
	Fog	74.10	64.28	72.14	71.61	**79.47**	45.89	5.63	0.87	1.34	44.35	44.55
	Sunlight	78.32	62.28	62.78	76.45	79.91	26.09	6.00	7.07	10.08	69.65	**80.97**
Sensor	Density	80.18	76.49	80.35	80.53	82.79	77.65	-	-	-	82.09	**84.95**
	Cutout	73.59	70.28	73.94	76.08	76.09	73.05	-	-	-	76.10	**78.06**
	Crosstalk	80.24	70.85	71.53	79.95	82.34	46.49	-	-	-	82.10	**85.82**
	Gaussian (L)	64.90	74.68	61.20	60.73	65.11	59.14	-	-	-	60.88	**82.14**
	Uniform (L)	79.18	77.31	76.39	77.77	81.16	74.91	-	-	-	79.24	**85.81**
	Impulse (L)	81.43	78.17	79.78	80.80	82.81	78.28	-	-	-	81.63	**85.01**
	Gaussian (C)	-	-	-	-	-	-	1.56	1.71	2.43	80.64	**80.97**
	Uniform (C)	-	-	-	-	-	-	2.67	3.29	4.85	81.61	**83.38**
	Impulse (C)	-	-	-	-	-	-	1.83	1.14	2.13	**81.18**	80.83
Motion	Moving Obj.	52.69	50.15	50.54	54.62	54.60	52.47	1.67	2.64	5.93	**55.78**	49.14
	Motion Blur	-	-	-	-	-	-	3.51	3.36	4.19	74.71	**81.08**
Object	Local Density	75.10	69.56	74.24	79.57	77.63	77.96	-	-	-	76.73	**80.84**
	Local Cutout	68.29	61.80	67.94	75.06	72.29	73.22	-	-	-	69.92	**76.64**
	Local Gaussian	72.31	76.58	69.82	77.44	70.44	75.11	-	-	-	75.76	**82.02**
	Local Uniform	80.17	78.04	77.67	80.77	82.09	78.64	-	-	-	81.71	**84.69**
	Local Impulse	81.56	78.43	80.26	82.25	84.03	79.53	-	-	-	82.21	**85.78**
	Shear	41.64	39.63	39.80	37.08	**47.72**	26.56	1.68	2.99	1.33	41.43	45.77
	Scale	73.11	70.29	71.50	75.90	**76.81**	75.02	0.13	0.15	0.33	69.05	69.48
	Rotation	76.84	72.70	75.57	77.50	**79.93**	76.98	1.11	2.14	2.57	74.62	77.76
Alignment	Spatial	-	-	-	-	-	-	-	-	-	35.14	43.01
Average (AP_{cor})		70.45	65.48	67.74	69.92	**72.59**	60.55	2.68	2.42	**3.05**	67.81	**71.87**

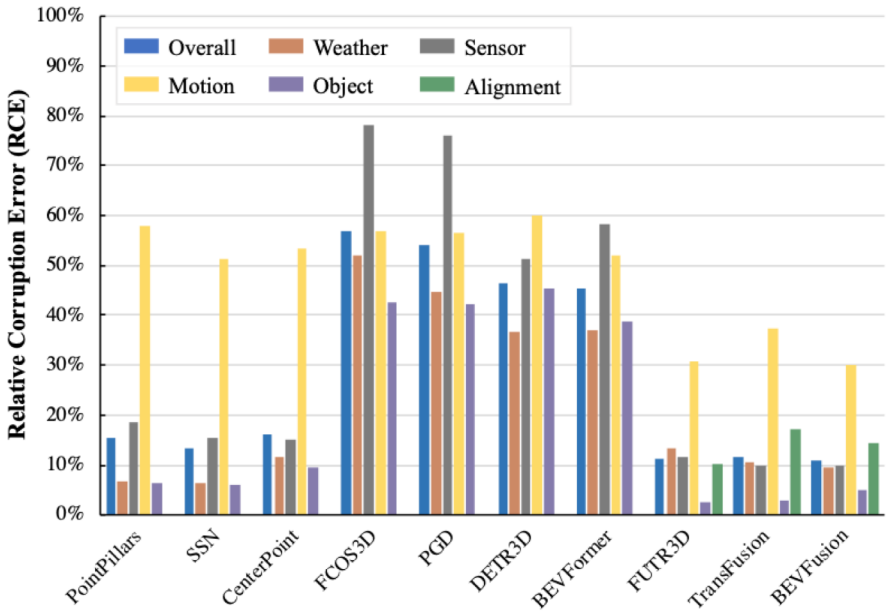


- Weather-level and motion-level corruptions affect the performance most.
- Fusion models have better performance under LiDAR corruptions, but have worst performance under LiDAR-camera corruptions.
- There may exist a trade-off between corruption robustness and efficiency.

Evaluation on nuScenes-C

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Corruption		LiDAR-only			Camera-only				LC Fusion		
		PointPillars	SSN	CenterPoint	FCOS3D	PGD	DETR3D	BEVFormer	FUTR3D	TransFusion	BEVFusion
None (mAP_{clean})		27.69	46.65	**59.28**	23.86	23.19	34.71	**41.65**	64.17	66.38	**68.45**
Weather	Snow	27.57	46.38	55.90	2.01	2.30	5.08	5.73	52.73	**63.30**	62.84
	Rain	27.71	46.50	56.08	13.00	13.51	20.39	24.97	58.40	65.35	**66.13**
	Fog	24.49	41.64	43.78	13.53	12.83	27.89	32.76	53.19	53.67	**54.10**
	Sunlight	23.71	40.28	54.20	17.20	22.77	34.66	41.68	57.70	55.14	**64.42**
Sensor	Density	27.27	46.14	58.60	-	-	-	-	63.72	65.77	**67.79**
	Cutout	24.14	40.95	56.28	-	-	-	-	62.25	63.66	**66.18**
	Crosstalk	25.92	44.08	56.64	-	-	-	-	62.66	64.67	**67.32**
	FOV Lost	8.87	15.40	20.84	-	-	-	-	26.32	24.63	**27.17**
	Gaussian (L)	19.41	39.16	45.79	-	-	-	-	58.94	55.10	**60.64**
	Uniform (L)	25.60	45.00	56.12	-	-	-	-	63.21	64.72	**66.81**
	Impulse (L)	26.44	45.58	57.67	-	-	-	-	63.43	65.51	**67.54**
	Gaussian (C)	-	-	-	3.96	4.33	14.86	15.04	54.96	**64.52**	64.44
	Uniform (C)	-	-	-	8.12	8.48	21.49	23.00	57.61	65.26	**65.81**
	Impulse (C)	-	-	-	3.55	3.78	14.32	13.99	55.16	**64.37**	64.30
Motion	Compensation	3.85	10.39	11.02	-	-	-	-	**31.87**	9.01	27.57
	Moving Obj.	19.38	35.11	44.30	10.36	10.47	16.63	20.22	45.43	51.01	**51.63**
	Motion Blur	-	-	-	10.19	9.64	11.06	19.79	55.99	64.39	**64.74**
Object	Local Density	26.70	45.42	57.55	-	-	-	-	63.60	65.65	**67.42**
	Local Cutout	17.97	32.16	48.36	-	-	-	-	61.85	63.33	**63.41**
	Local Gaussian	25.93	43.71	51.13	-	-	-	-	62.94	63.76	**64.34**
	Local Uniform	27.69	46.87	57.87	-	-	-	-	64.09	66.20	**67.58**
	Local Impulse	27.67	46.88	58.49	-	-	-	-	64.02	66.29	**67.91**
	Shear	26.34	43.28	49.57	17.20	16.66	17.46	24.71	55.42	**62.32**	60.72
	Scale	27.29	45.98	51.13	6.75	6.57	12.02	17.64	56.79	64.13	**64.57**
	Rotation	27.80	46.93	54.68	17.21	16.84	27.28	33.97	59.64	63.36	**65.13**
Alignment	Spatial	-	-	-	-	-	-	-	63.77	66.22	**68.39**
	Temporal	-	-	-	-	-	-	-	**51.43**	43.65	49.02
Average (mAP_{cor})		23.42	40.37	**49.81**	10.26	10.68	18.60	**22.79**	56.99	58.73	**61.03**



- Motion-level corruptions affect the performance most.
- Camera-only models are more vulnerable under corruptions.
- There is a trade-off of corruption robustness of fusion models under camera and LiDAR corruptions, since different models have varying reliance on modalities.

Thanks