A. More Details of Common Corruptions

In this section, we provide more technical details about the 27 common corruptions in 3D object detection studied in this paper.

A.1. Implementation Details

First, we introduce the implementation details and hyperparameters of the 27 common corruptions in the three benchmarks—KITTI-C, nuScenes-C, and Waymo-C. Note that we have five severities for each corruption, thus we introduce the corresponding hyperparameter configuration of each severity.

Snow. For **LiDAR**, we adopt the method proposed in [8] to simulate snow on clean data for nuScenes-C, and use LISA [11] for KITTI-C and Waymo-C. Following the definition in [15], we set the *snowfall rate* as $\{$0.20, 0.73, 1.5625, 3.125, 7.29$\}$ under the five severities for both LISA [11] and [8]. For **Camera**, we use the imgaug library [10] to implement it, and use the pre-defined severities $\{$1, 2, 3, 4, 5$\}$ to simulate different intensity of snow. To keep the consistency with the STF dataset [2], we add a 30%-opacity gray mask layer, and reduce the brightness by 30%.

Rain. For **LiDAR**, we use the LISA rain simulation method proposed in [8] for KITTI-C, nuScenes-C, and Waymo-C. Following the real-world rainfall rates defined in [19] and [16], we set the parameter of *rainfall rate* as $\{$0.20, 0.73, 1.5625, 3.125, 7.29$\}$ in LISA to simulate rain intensity across light rain, moderate rain and heavy rain. For **Camera**, we set the parameter of *rainfall density* as $\{$0.01, 0.06, 0.10, 0.15, 0.20$\}$ in RainLayer in imgaug library [10] to simulate different severity of rain. Besides, we also add a 30%-opacity gray mask layer, and reduce the brightness by 30%.

Fog. For **LiDAR**, we use the method proposed in [9] for all three benchmarks. The parameter of α in [9] can represent meteorological optical range in real foggy weather. Following the settings in their paper, we set α to $\{$0.005, 0.01, 0.02, 0.03, 0.06$\}$ for different severities of fog. For **Camera**, we use imgaug library [10] to implement it, and use the predefined severities $\{$1, 2, 3, 4, 5$\}$ to simulate different intensity of fog. Besides, we also add a $\{$10%, 20%, 30%, 40%, 50%$\}$-opacity gray mask layer.

Strong Sunlight. For **LiDAR**, following the observations in [5], we simulate it by adding 2m Gaussian noises to points. We use the ratio of $\{$1%, 2%, 3%, 4%, 5%$\}$ noisy points to define the severity. For **Camera**, we set $\{$30, 40, 50, 60, 70$\}$-pixel size sun in automold library [18] for strong sunlight simulation.

Density Decrease. We randomly delete $\{$6%, 12%, 18%, 24%, 30%$\}$ of points in one frame of LiDAR.

Cutout. We randomly remove $\{$2, 3, 5, 7, 10$\}$ groups of point cloud, where the number of points within each group is $\frac{N}{50}$, and each group is within a ball in the Euclidean space, where N is the total numbers of point in one frame of LiDAR.

LiDAR Crosstalk. Following [3], we select a subset of points with the ratio of $\{$0.4%, 0.8%, 1.2%, 1.6%, 2%$\}$ to add 3m Gaussian noises.

FOV Lost. Five groups of FOV lost are selected, the reserved angle range is $\{$(-105, 105), (-90, 90), (-75, 75), (-60, 60), (-45, 45)$\}$.

Gaussian Noise. For **LiDAR**, we add Gaussian noises to all points with the severities of $\{$0.02m, 0.04m, 0.06m, 0.08m, 0.10m$\}$. For **Camera**, we use imgaug library [10] to implement it, and use the predefined severities $\{$1, 2, 3, 4, 5$\}$ to simulate different intensity of *Gaussian Noise*.

Uniform Noise. For **LiDAR**, we add uniform noises to all points with the severities of $\{$0.02m, 0.04m, 0.06m, 0.08m, 0.10m$\}$. For **Camera**, we add the uniform noise of $\pm\{$0.08, 0.12, 0.18, 0.26, 0.38$\}$ to the image, thus to simulate different intensity of *Uniform Noise*.

Impulse Noise. For **LiDAR**, we select the number of points in $\{\frac{N}{30}, \frac{N}{25}, \frac{N}{20}, \frac{N}{15}, \frac{N}{10}\}$ to add impulse noise and represent the severities, where N is the total numbers of point in one frame of LiDAR. For **Camera**, we use imgaug library [10] to implement it, and use the predefined severities $\{$1, 2, 3, 4, 5$\}$ to simulate different intensity of *Impulse Noise*.

Motion Compensation. We add Gaussian noises to the rotation and translation matrices of the vehicle's ego pose. The noises are $\{$0.02, 0.04, 0.06, 0.08, 0.10$\}$ for the rotation matrix and $\{$0.002, 0.004, 0.006, 0.008, 0.010$\}$ for the translation matrix.

Moving Object. For **LiDAR**, we first divide a 3D bounding box to three parts, and then move the second part forward with $\frac{c}{2}$, and move the third part forward with c, where c is $\{$0.2, 0.3, 0.4, 0.5, 0.6$\}$. For **Camera**, we first use ground-truth 3D bounding boxes to select the object regions and use imgaug library [10] with zoom factor at $\{$2, 3, 4, 5, 6$\}$ in these regions.

Motion Blur. We use imgaug library [10] and set the zoom factor to $\{$2, 3, 4, 5, 6$\}$ to implement it.

Local Density Decrease. We randomly delete 75% of points within a group, the group number are $\{$1, 2, 3, 4, 5$\}$ and each group has 10% points of a LiDAR frame.

Local Cutout. Similar to *cutout*, we randomly remove $\{$30%, 40%, 50%, 60%, 70%$\}$ of points that within a ball in the Euclidean space, all points are within the objects' 3D bounding boxes.

Local Gaussian Noise. Similar to the sensor-level Gaussian noise, we add Gaussian noises to the points within the objects' 3D bounding boxes. The noises are $\{$0.02m, 0.04m, 0.06m, 0.08m, 0.10.$\}$.

Local Uniform Noise. Similar to the sensor-level uniform noise, we add uniform noises to points within the objects' 3D bounding boxes with the severities of $\{$0.02m,

	Corruption Types					Datasets			3D Object Detection Models			
	Weather	Sensor	Motion	Object	Alignment	KITTI	nuScenes	Waymo	LiDAR-only	Camera-only	Fusion	#Models
Li et al. [13]	✓	✓	✗	✓	✗	✓	✗	✗	✓	✗	✗	7
Yu et al. [21]	✗	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	10
Ours	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**24**

Table A.1. Comparison between our work and two related works [13, 21] in terms of corruption types, datasets, and evaluation models. Our benchmark is more comprehensive in all aspects.

0.04m, 0.06m, 0.08m, 0.10m}.

Local Impluse Noise. Similar to the sensor-level impluse noise, we select the number of points in $\{\frac{N_{bbox}}{30}, \frac{N_{bbox}}{25}, \frac{N_{bbox}}{20}, \frac{N_{bbox}}{15}, \frac{N_{bbox}}{10}\}$ within the objects' 3D bounding boxes to add impulse noises and represent the severities, where N_{bbox} is the total number of points within the 3D bounding box.

Shear. For **LiDAR**, we use the shear transformation for the point cloud within the objects' 3D bounding boxes. Let $\mathcal{X}$ represent the point cloud within a 3D bounding box, the transformation can represent as:

$$\mathcal{X}_t = \mathcal{X} \begin{bmatrix} 1 & 0 & d \\ e & 1 & f \\ g & 0 & 1 \end{bmatrix}, \tag{A.1}$$

where d, e, f, g are selected from the uniform distribution bounded by ±{(0.0, 0.1), (0.05, 0.15), (0.1, 0.2), (0.15, (0.20, 0.30)}. For **Camera**, we obtain the 3D bounding box from annotation, do the same shear transformation as in LiDAR points on 8 bounding box corners in the 3D space. Let $\mathcal{X}_c$ as 8 corners, the transformation can be:

$$\mathcal{X}_{ct} = \mathcal{X}_c \begin{bmatrix} 1 & 0 & d \\ e & 1 & f \\ g & 0 & 1 \end{bmatrix}. \tag{A.2}$$

Then we project the corners before and after transformation to images. The projected corners are used as control points to do Thin Plate Spline (TPS) on images. The severities are the same with LiDAR.

Scale. For **LiDAR**, we use the scale transformation for the point cloud within the objects' 3D bounding boxes. Let $\mathcal{X}$ represent the point cloud within 3D bounding box, the transformation can represent as:

$$\mathcal{X}_t = \mathcal{X} \begin{bmatrix} a & b & c \end{bmatrix}, \tag{A.3}$$

where a, b, c are selected from ±{0.04, 0.08, 0.12, 0.16, 0.20}. For **Camera**, we obtain the 3D bounding box from annotation, do the same scaling transformation on 8 bounding box corners in 3D space. Let $\mathcal{X}_c$ as 8 corners, the transformation can be:

$$\mathcal{X}_{ct} = \mathcal{X}_c \begin{bmatrix} a & b & c \end{bmatrix}. \tag{A.4}$$

Then we project the corners before and after transformation to images. The projected corners are used as control points to do TPS on images. The severities are the same with LiDAR.

Rotation. For **LiDAR**, we rotate each 3D bounding box along z axis with angles sampled from the uniform distribution of ±{(0, 2), (3, 4), (5, 6), (7, 8), (9, 10)}. For **Camera**, we obtain the 3D bounding box from annotation, do the same rotation transformation on 8 bounding box corners in 3D space. Then we project the corners before and after transformation to images. The projected corners are used as control points to do TPS on images. The severities are the same with LiDAR.

Spatial Misalignment. We add Gaussian noises to the calibration matrices between **LiDAR** and **Camera**. Specifically, the noises are {0.02, 0.04, 0.06, 0.08, 0.10} for the rotation matrix and {0.002, 0.004, 0.006, 0.008, 0.010} for the translation matrix.

Temporal Misalignment. For **LiDAR**, the stucked frames are {2, 4, 6, 8, 10}. For **Camera**, the stucked frames also are {2, 4, 6, 8, 10}.

A.2. Comparison with Related Work

As we mentioned in Sec. 2.2, there are two concurrent works [13, 21], which also study the corruption robustness of 3D object detection in autonomous driving. Compared with them, our benchmark is more comprehensive in terms of corruption types, evaluated datasets, and studied 3D object detection models, as shown in Table A.2. Notably, they did not consider motion-level corruptions and we for the first time study motion-level corruptions in a comprehensive robustness benchmark.

A.3. Visualization

We show the full visualization of all 27 corruptions in Fig. A.1. Note that an input (image or point cloud) may not be modified under a corruption, thus we mark it by the black box. For input that has been modified under the corruption, we mark it by the red box.

Since we have 24 corruptions with 5 severities, the KITTI-C dataset is 120× larger than the KITTI validation set, requiring more than 750G storage space. nuScenes-C and Waymo-C are even much bigger than KITTI-C. We will plan to release the full benchmarks.

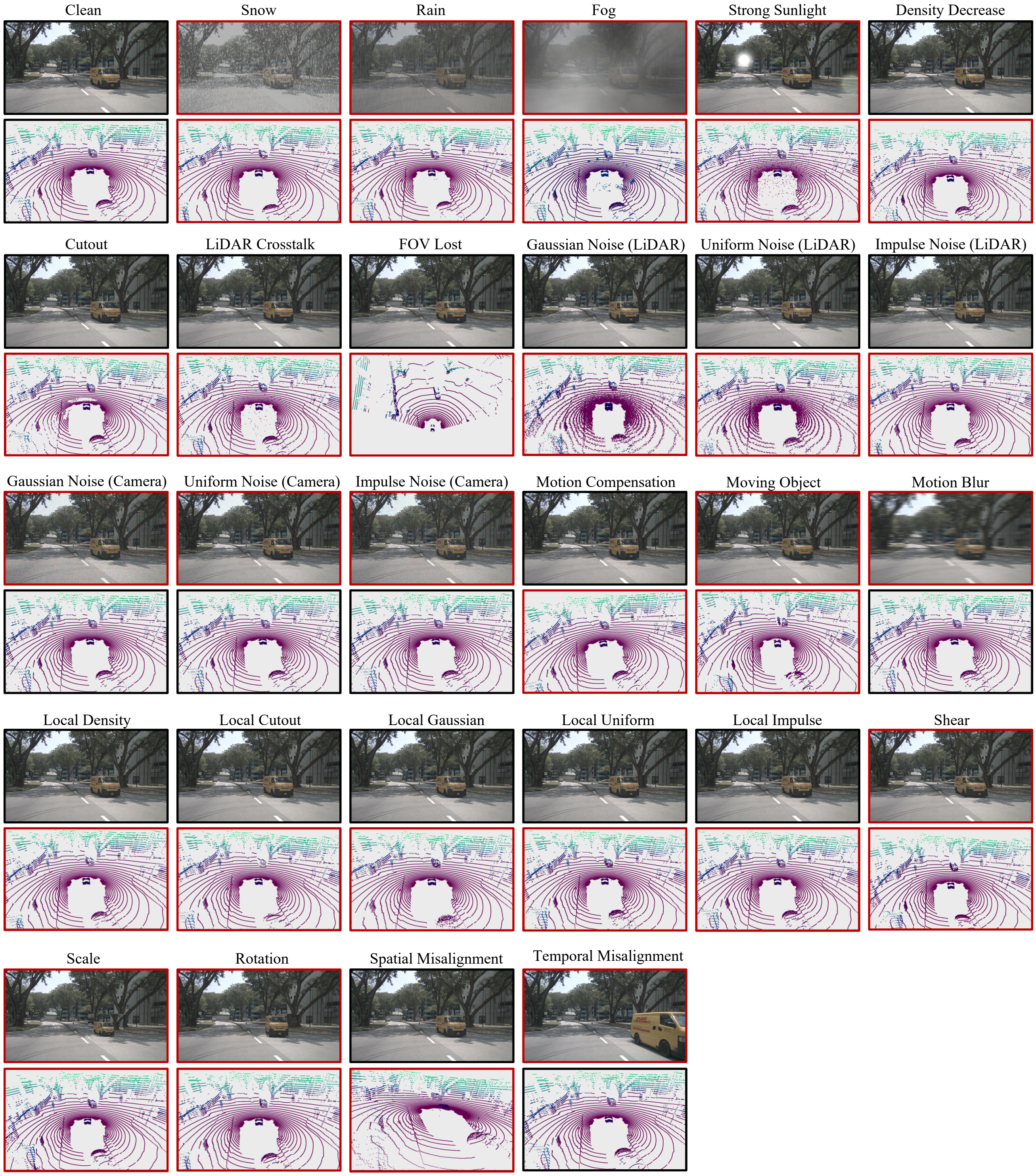

Figure A.1. Full visualization results of all corruptions in our benchmark (best viewed when zoomed in). The images or point clouds in red boxes are modified under the corresponding corruption, while the images or point clouds in black boxes are kept unchanged.

A.4. Naturalness of Common Corruptions

Quantitative analysis. In Sec. 3, we have discussed the gap between synthetic and real-world corruptions. Here

Corruption		LiDAR-only			Camera-only				LC Fusion		
		PointPillars	SSN	CenterPoint	FCOS3D	PGD	DETR3D	BEVFormer	FUTR3D	TransFusion	BEVFusion
Synthetic	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
Real	Sunny	27.60	47.01	56.00	23.88	23.09	34.33	41.02	63.67	66.24	67.20
	Rainy	27.12	44.23	54.20	23.18	22.32	36.25	43.95	65.73	66.62	68.71
	Day	27.41	46.70	56.69	24.15	23.53	34.99	41.88	64.18	66.37	67.50
	Night	18.74	24.48	30.98	12.13	11.15	16.04	21.21	38.44	41.56	39.47

Table A.2. Comparison of model performance under synthetic weathers and real-world dataset of different conditions.

we further examine the performance of 3D object detection models under adverse weathers crafted by synthetic methods or collected in the real dataset. The nuScenes [4] dataset has provided the annotations for *Day*, *Night*, *Sunny*, and *Rainy*. Thus we show the performance of 10 3D object detection models introduced in Sec. 4.2 under both synthetic and real-world conditions. Table A.2 shows the results. Specifically, the model performance under synthetic and real rain weather is largely consistent. The only exception is for camera-only models, where the gap is relatively large. This is due to the difficulty of synthesizing more realistic images under adverse weathers. However, the relative performance of different models is consistent. The results can prove the validity of using our corruption benchmarks for evaluating the robustness of 3D object detection models.

Data quality check. As pointed out by one of the reviewers, data quality check is an important aspect of our benchmark. Actually, we did data quality checks when we simulated the corruptions. We ensured that the objects are detectable for humans by appropriately adjusting the hyper-parameters (i.e., severity) of each corruption, as detailed in Appendix A.1. The only exceptions are Cutout and FOV Lost, which may drop objects in the point clouds. We think that a potential solution is to discard the ground-truth objects if they are invisible. However, we found that the evaluations are hard to perform and compare with each other since fusion models have the ability to detect those objects based on accurate camera inputs. Therefore, we tend to keep the original evaluation results (different from what we promise in the rebuttal) and will further consider this problem.

B. Additional Results on KITTI-C

In addition to the experimental results in Sec. 5.1, we further provide more results on KITTI-C for other classes and difficulties. We show the corruption robustness of 11 3D object detectors on the car class at easy and hard difficulties in Table B.1 and Table B.2, respectively. The results are highly consistent with those based on the car class at moderate difficulty in Table 3. For the other two classes (*i.e.*, pedestrian, cyclist), there are only 6 models that can predict these two classes, including SECOND, PointPillars, PointRCNN, PV-RCNN, SMOKE, and PGD. We show the corruption robustness of these 6 3D object detectors on the pedestrian and cyclist classes at the moderate diffucilty in Table B.3 and Table B.4, respectively. Fig. B.1 further shows the model performance under different severities of each corruption. It can be seen that for most corruptions, model performance drops along with the increasing severity.

C. Additional Results on nuScenes-C

We further provide the results on nuScenes-C under the NDS metric in Table C.1. The findings are consistent across both the mAP and NDS metrics. We similarly provide the curves of model performance along with severity of each corruption in Fig. C.1.

D. Results on Waymo-C

We evaluate the corruption robustness of PointPillars [12], BEVFormer [14], and TransFusion [1] on Waymo-C in Table D.1. Since we do not have enough models for more comprehensive comparison, we can only draw the conclusion that the LiDAR-camera fusion model TransFusion demonstrates better performance than the other models. We would continuously evaluate more 3D object detection models on Waymo-C in future.

E. Data Augmentation as Potential Defense

In this section, we explore data augmentation for improving the robustness of 3D object detection models under common corruptions. We adopt the PA-AUG and Dropout [7] methods and PointCutMix-R [24] method for LiDAR point cloud augmentation. For the camera modality, we use two famous image data augmentations, which are Mixup [23] and CutMix [22].

For LiDAR-only models. We perform experiments on SECOND [20] and PV-RCNN [17] due to their superior robustness among all LiDAR-only models. The results are shown in Table E.1. These augmentations do not improve performance consistently. The Dropout augmentation only improves the corruption robustness of SECOND by 0.84.

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)		90.53	87.75	91.65	91.68	**92.10**	91.07	10.42	12.72	**17.85**	**92.29**	92.00
Weather	Snow	73.05	55.99	71.93	57.56	**73.06**	42.76	3.68	0.86	0.30	48.03	53.80
	Rain	**73.31**	55.17	70.79	55.77	72.37	40.39	5.66	4.85	1.77	50.93	61.44
	Fog	85.58	74.27	85.01	79.74	**89.21**	61.12	8.06	1.32	2.37	64.83	68.03
	Sunlight	88.05	67.42	64.90	84.25	87.27	21.59	8.75	10.94	15.72	81.77	**90.03**
Sensor	Density	90.45	86.86	91.33	90.69	91.98	90.63	-	-	-	91.89	**92.14**
	Cutout	81.75	78.90	83.33	**86.13**	83.40	85.06	-	-	-	84.17	83.84
	Crosstalk	89.63	78.51	77.38	88.58	90.52	44.35	-	-	-	91.30	**92.01**
	Gaussian (L)	73.21	86.24	74.28	65.68	74.61	69.99	-	-	-	66.99	**88.56**
	Uniform (L)	89.50	87.49	89.48	86.64	90.65	87.83	-	-	-	89.70	**91.77**
	Impulse (L)	90.70	87.75	90.80	90.88	91.91	90.04	-	-	-	91.44	**92.10**
	Gaussian (C)	-	-	-	-	-	-	2.09	2.83	3.74	**91.62**	89.51
	Uniform (C)	-	-	-	-	-	-	3.81	5.45	7.66	**91.95**	91.20
	Impulse (C)	-	-	-	-	-	-	2.57	1.97	3.35	**91.68**	89.90
Motion	Moving Obj.	62.64	58.49	59.29	64.40	63.36	62.48	2.69	4.57	9.63	**66.32**	54.57
	Motion Blur	-	-	-	-	-	-	5.39	5.91	6.75	89.65	**91.56**
Object	Local Density	87.74	82.90	88.37	90.30	89.60	**90.33**	-	-	-	89.40	89.60
	Local Cutout	81.29	75.22	83.30	**87.92**	84.38	87.69	-	-	-	82.40	85.55
	Local Gaussian	82.05	87.69	82.44	87.49	77.89	87.82	-	-	-	85.72	**89.78**
	Local Uniform	90.11	87.83	89.30	91.22	90.63	90.57	-	-	-	91.32	**91.88**
	Local Impulse	90.58	87.84	90.60	91.82	91.91	90.89	-	-	-	91.67	**92.02**
	Shear	47.80	45.06	45.52	37.86	**52.39**	32.54	2.41	4.46	1.72	45.23	48.90
	Scale	81.84	80.57	81.41	86.80	85.14	**87.31**	0.12	0.14	0.39	80.53	78.82
	Rotation	87.39	83.61	87.09	88.38	**89.29**	88.71	1.43	3.19	3.68	86.70	87.02
Alignment	Spatial	-	-	-	-	-	-	-	-	-	42.23	**51.21**
Average (AP_cor)		81.40	76.20	79.29	79.58	**82.61**	71.16	3.89	3.87	**4.76**	78.64	**81.05**

Table B.1. The benchmarking results of 11 3D object detectors on **KITTI-C** based on the car class at easy difficulty.

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)		78.57	75.19	78.06	80.22	**82.49**	78.23	5.57	6.20	**9.20**	80.16	**83.36**
Weather	Snow	**48.62**	32.96	45.41	40.03	48.62	23.15	1.92	0.44	0.20	32.39	30.41
	Rain	**48.79**	32.65	45.78	39.09	48.20	22.56	3.16	2.26	0.99	34.69	35.71
	Fog	68.93	58.19	68.05	68.39	**75.05**	41.21	4.56	0.63	1.03	38.12	39.50
	Sunlight	74.62	58.69	61.11	73.55	78.02	24.70	4.91	5.42	8.24	66.43	**78.06**
Sensor	Density	77.04	72.85	77.58	78.33	81.15	74.56	-	-	-	79.77	**82.38**
	Cutout	70.79	67.32	71.57	73.91	74.60	70.52	-	-	-	73.95	**76.69**
	Crosstalk	76.92	67.51	69.41	77.26	80.98	43.67	-	-	-	79.54	**83.22**
	Gaussian (L)	61.09	71.12	56.73	58.71	62.70	55.61	-	-	-	56.88	**77.15**
	Uniform (L)	75.61	74.09	72.25	75.02	78.93	71.77	-	-	-	75.92	**81.62**
	Impulse (L)	78.33	74.65	76.88	78.78	81.79	75.37	-	-	-	79.14	**83.28**
	Gaussian (C)	-	-	-	-	-	-	1.18	1.26	1.96	78.20	**79.01**
	Uniform (C)	-	-	-	-	-	-	2.19	2.46	3.90	79.14	**81.39**
	Impulse (C)	-	-	-	-	-	-	1.52	0.82	1.71	78.51	**78.87**
Motion	Moving Obj.	48.02	45.47	46.23	**53.06**	50.75	50.86	1.40	1.97	4.63	50.97	45.34
	Motion Blur							2.95	2.44	3.32	72.49	**77.75**
Object	Local Density	71.45	65.70	71.09	**77.58**	75.39	75.05	-	-	-	74.36	77.30
	Local Cutout	63.25	56.69	63.50	**72.86**	68.58	70.73	-	-	-	66.53	72.40
	Local Gaussian	68.16	73.11	65.65	75.32	68.03	72.84	-	-	-	72.71	**78.52**
	Local Uniform	76.67	74.68	74.37	78.47	80.17	76.31	-	-	-	78.85	**81.99**
	Local Impulse	78.47	75.18	77.38	79.98	82.33	76.91	-	-	-	79.79	**83.20**
	Shear	39.99	38.11	38.12	37.12	**47.06**	24.87	1.39	2.31	1.18	40.62	44.25
	Scale	70.03	67.22	68.55	73.74	**74.89**	72.56	0.12	0.15	0.32	65.68	66.65
	Rotation	73.24	69.24	72.32	75.33	**78.02**	74.35	0.84	1.67	2.18	71.91	75.25
Alignment	Spatial	-	-	-	-	-	-	-	-	-	33.94	**41.06**
Average (AP_cor)		66.84	61.86	64.31	67.71	**70.28**	57.77	2.18	1.82	**2.47**	65.02	**68.79**

Table B.2. The benchmarking results of 11 3D object detectors on **KITTI-C** based on the car class at hard difficulty.

But these augmentations drop the robustness of PV-RCNN by more than 4.68. The reason is that these augmentations degrade model performance on clean data. Since the corruption robustness is highly correlated with clean perfor-

Corruption		**LiDAR-only**				**Camera-only**	
		SECOND	PointPillars	PointRCNN	PV-RCNN	SMOKE	PGD
None (AP_{clean})		51.14	51.41	54.40	**54.49**	**3.19**	1.27
Weather	Snow	49.68	49.07	**55.73**	53.01	1.11	0.14
	Rain	50.34	49.23	**56.08**	54.98	2.68	0.74
	Fog	**3.10**	0.05	0.14	0.67	2.71	0.22
	Sunlight	**49.63**	29.34	33.49	42.19	2.35	1.15
Sensor	Density	50.67	50.08	54.84	**55.59**	-	-
	Cutout	44.92	44.94	**49.38**	48.05	-	-
	Crosstalk	**50.28**	38.15	43.02	48.20	-	-
	Gaussian (L)	24.82	**40.00**	25.89	26.32	-	-
	Uniform (L)	41.37	**49.54**	44.24	45.58	-	-
	Impulse (L)	50.33	51.22	50.19	**52.39**	-	-
	Gaussian (C)	-	-	-	-	**0.79**	0.22
	Uniform (C)	-	-	-	-	**1.62**	0.58
	Impulse (C)	-	-	-	-	**0.99**	0.10
Motion	Moving Obj.	3.57	3.30	**4.86**	4.80	0.69	0.59
	Motion Blur	-	-	-	-	**1.19**	0.82
Object	Local Density	37.30	33.94	**45.11**	37.74	-	-
	Local Cutout	21.35	23.71	19.99	**23.96**	-	-
	Local Gaussian	27.49	**43.60**	28.54	29.11	-	-
	Local Uniform	44.63	**51.94**	46.17	47.83	-	-
	Local Impulse	50.76	52.20	52.40	**53.20**	-	-
	Shear	35.91	38.31	38.52	**38.70**	0.93	0.41
	Scale	46.00	46.11	**51.30**	50.26	0.18	0.09
	Rotation	50.83	51.05	54.10	**54.49**	1.78	0.74
Average (AP_{cor})		38.58	39.25	39.68	**40.37**	1.42	0.48

Table B.3. The benchmarking results of 6 3D object detectors on **KITTI-C** based on the pedestrian class at moderate difficulty.

Corruption		**LiDAR-only**				**Camera-only**	
		SECOND	PointPillars	PointRCNN	PV-RCNN	SMOKE	PGD
None (AP_{clean})		66.74	62.81	**71.00**	70.38	0.25	**0.86**
Weather	Snow	51.35	44.15	**57.88**	55.56	0.19	0.02
	Rain	51.49	44.65	**58.64**	56.19	0.15	0.21
	Fog	**10.91**	2.77	4.29	4.31	0.30	0.03
	Sunlight	61.12	45.05	60.33	**61.58**	0.40	0.41
Sensor	Density	63.00	60.60	**69.66**	67.76	-	-
	Cutout	59.03	55.80	**63.46**	62.28	-	-
	Crosstalk	64.02	53.52	65.25	**67.67**	-	-
	Gaussian (L)	48.03	52.62	**54.08**	47.53	-	-
	Uniform (L)	62.56	60.58	**66.77**	66.40	-	-
	Impulse (L)	64.34	62.28	**70.13**	68.69	-	-
	Gaussian (C)	-	-	-	-	0.04	**0.09**
	Uniform (C)	-	-	-	-	0.17	**0.21**
	Impulse (C)	-	-	-	-	**0.07**	0.02
Motion	Moving Obj.	21.54	21.04	23.88	**28.77**	0.07	0.04
	Motion Blur	-	-	-	-	**0.17**	0.08
Object	Local Density	47.26	36.98	**52.49**	49.76	-	-
	Local Cutout	24.59	20.47	25.93	**27.01**	-	-
	Local Gaussian	53.61	58.94	**60.81**	56.39	-	-
	Local Uniform	63.18	61.58	**68.80**	68.30	-	-
	Local Impulse	65.11	62.79	**70.80**	69.91	-	-
	Shear	57.09	56.40	**64.42**	60.83	0.10	0.14
	Scale	64.02	60.46	67.31	**68.30**	0.08	0.03
	Rotation	64.23	62.75	**69.67**	69.39	0.08	0.16
Average (AP_{cor})		52.45	48.60	**56.56**	55.61	0.15	0.12

Table B.4. The benchmarking results of 6 3D object detectors on **KITTI-C** based on the cyclist class at moderate difficulty.

mance, the effectiveness of these augmentations is limited.

For LiDAR-camera fusion models. We choose Focals Conv [6] as the target to study the effectiveness of data augmentation techniques. The multi-modal data augmentation is still an open question in computer vision community [25], especially in 3D object detection. Here, we explore the synergistic data augmentation of camera modalities and LiDAR modalities. Specifically, we choose three point cloud

Corruption		**LiDAR-only**			**Camera-only**				**LC Fusion**		
		PointPillars	SSN	CenterPoint	FCOS3D	PGD	DETR3D	BEVFormer	FUTR3D	TransFusion	BEVFusion
None (NDS$_{\text{clean}}$)		46.86	58.24	**67.33**	34.69	35.04	42.23	**51.74**	68.05	69.82	**71.40**
Weather	Snow	46.67	58.07	64.92	8.57	9.83	15.53	15.61	61.52	68.29	**68.33**
	Rain	46.79	58.16	64.98	26.31	26.96	31.60	38.82	64.67	69.40	**70.14**
	Fog	44.91	55.42	58.11	26.05	25.83	37.26	45.42	61.20	62.62	**62.73**
	Sunlight	44.57	54.59	64.41	29.34	34.77	42.20	51.70	63.61	61.36	**68.95**
Sensor	Density	46.62	57.93	66.84	-	-	-	-	67.58	69.42	**71.01**
	Cutout	44.74	55.06	65.73	-	-	-	-	66.91	68.30	**70.09**
	Crosstalk	45.93	56.72	65.83	-	-	-	-	67.17	68.83	**70.72**
	FOV Lost	35.69	41.61	47.07	-	-	-	-	45.66	47.89	**48.65**
	Gaussian (L)	40.62	53.24	58.08	-	-	-	-	64.10	62.32	**65.99**
	Uniform (L)	45.44	57.03	65.22	-	-	-	-	67.28	68.68	**70.18**
	Impulse (L)	46.21	57.42	66.22	-	-	-	-	67.47	69.06	**70.63**
	Gaussian (C)	-	-	-	11.16	12.73	26.38	29.60	62.92	68.94	**69.35**
	Uniform (C)	-	-	-	19.55	20.63	32.13	37.57	64.43	69.33	**70.06**
	Impulse (C)	-	-	-	11.71	12.07	26.03	29.24	63.07	68.89	**69.25**
Motion	Compensation	20.64	27.93	27.71	-	-	-	-	**39.62**	25.69	36.76
	Moving Obj.	39.23	49.19	55.45	23.57	24.33	28.17	34.59	56.41	**60.03**	59.42
	Motion Blur	-	-	-	23.04	23.50	23.49	29.17	63.44	68.85	**69.38**
Object	Local Density	46.27	57.63	66.22	-	-	-	-	67.62	69.34	**70.77**
	Local Cutout	39.37	48.64	60.40	-	-	-	-	66.45	67.97	**68.11**
	Local Gaussian	45.31	56.41	61.27	-	-	-	-	66.85	67.96	**68.32**
	Local Uniform	46.87	58.42	66.22	-	-	-	-	67.92	69.67	**70.68**
	Local Impulse	46.93	58.41	66.70	-	-	-	-	67.89	69.64	**70.93**
	Shear	45.34	55.44	54.02	29.34	40.65	28.74	38.77	61.15	**66.43**	62.95
	Scale	46.58	57.85	61.27	21.68	21.41	25.48	32.81	62.00	**67.81**	66.00
	Rotation	46.78	58.18	58.19	29.38	29.82	36.39	45.45	63.67	**67.42**	66.31
Alignment	Spatial	-	-	-	-	-	-	-	67.75	69.72	**71.35**
	Temporal	-	-	-	-	-	-	-	**57.91**	54.23	56.52
Average (NDS$_{\text{cor}}$)		43.41	53.97	**60.23**	21.64	22.63	29.45	**35.73**	62.82	64.74	**66.06**

Table C.1. The benchmarking results of 10 3D object detectors on **nuScenes-C** under the NDS metric.

augmentations and two image augmentations for LiDAR-camera fusion models. The results are shown in Table E.2. It can be seen that the combination of data augmentations of both modalities degrades the performance a lot. Therefore, it remains an open problem of improving the corruption robustness of 3D object detectors, especially LiDAR-camera fusion models.

References

[1] Xuyang Bai, Zeyu Hu, Xinge Zhu, Qingqiu Huang, Yilun Chen, Hongbo Fu, and Chiew-Lan Tai. Transfusion: Robust lidar-camera fusion for 3d object detection with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1090–1099, 2022. 4

[2] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11682–11692, 2020. 1

[3] Lara Brinon-Arranz, Tiana Rakotovao, Thierry Creuzet, Cem Karaoguz, and Oussama El-Hamzaoui. A methodology for analyzing the impact of crosstalk on lidar measurements. In *2021 IEEE Sensors*, pages 1–4, 2021. 1

[4] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11621–11631, 2020. 4

[5] Alexander Carballo, Jacob Lambert, Abraham Monrroy, David Wong, Patiphon Narksri, Yuki Kitsukawa, Eijiro Takeuchi, Shinpei Kato, and Kazuya Takeda. Libre: The multiple 3d lidar dataset. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 1094–1101, 2020. 1

[6] Yukang Chen, Yanwei Li, Xiangyu Zhang, Jian Sun, and Jiaya Jia. Focal sparse convolutional networks for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5428–5437, 2022. 6

[7] Jaeseok Choi, Yeji Song, and Nojun Kwak. Part-aware data augmentation for 3d object detection in point cloud. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3391–3397, 2021. 4

[8] Martin Hahner, Christos Sakaridis, Mario Bijelic, Felix Heide, Fisher Yu, Dengxin Dai, and Luc Van Gool. Lidar snowfall simulation for robust 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16364–16374, 2022. 1

[9] Martin Hahner, Christos Sakaridis, Dengxin Dai, and Luc

Corruption		PointPillars	BEVFormer	TransFusion
None ($\mathrm{L2/mAPH_{clean}}$)		59.13	22.96	**60.16**
Weather	Snow	33.39	5.98	**34.98**
	Rain	34.85	12.61	**37.20**
	Fog	0.92	**18.50**	0.81
	Sunlight	30.00	23.50	**47.10**
Sensor	Density	58.33	-	**58.95**
	Cutout	54.14	-	**55.40**
	Crosstalk	43.04	-	**57.06**
	FOV Lost	26.16	-	**26.28**
	Gaussian (L)	**52.27**	-	36.68
	Uniform (L)	**57.85**	-	56.54
	Impulse (L)	58.76	-	**59.49**
	Gaussian (C)	-	14.64	**59.98**
	Uniform (C)	-	16.00	**60.01**
	Impulse (C)	-	14.12	**59.95**
Motion	Moving Obj.	40.02	12.99	**41.55**
	Motion Blur	-	7.98	**59.29**
Object	Local Density	58.45	-	**59.39**
	Local Cutout	56.74	-	**58.11**
	Local Gaussian	57.88	-	**58.23**
	Local Uniform	58.77	-	**59.68**
	Local Impulse	58.85	-	**59.95**
	Shear	49.61	6.93	**52.73**
	Scale	54.25	1.89	**55.62**
	Rotation	54.89	10.36	**58.04**
Alignment	Spatial	-	-	**59.66**
Average ($\mathrm{L2/mAPH_{cor}}$)		46.96	12.13	**50.91**

Table D.1. The benchmarking results of 3 3D object detectors on **Waymo-C**. We show the performance under each corruption and the overall corruption robustness mAPcor averaged over all corruption types.

Augmentation	Second	PV-RCNN
None	70.45	72.59
PA-AUG	70.63	65.93
Dropout	71.29	67.91
PointCutMix-R	68.97	64.43

Table E.1. The corruption robustness ($\mathrm{AP_{cor}}$) of SECOND and PV-RCNN with different data augmentations on **KITTI-C**.

Image Aug.	Point Cloud Aug.	Focals Conv
None	None	71.87
Mixup	PA-AUG	55.32
Mixup	Dropout	25.07
Mixup	PointCutMix-R	48.12
CutMix	PA-AUG	30.01
CutMix	Dropout	53.90
CutMix	PointCutMix-R	28.83

Table E.2. The corruption robustness ($\mathrm{AP_{cor}}$) of Focals Conv with different data augmentations on **KITTI-C**.

Van Gool. Fog simulation on real lidar point clouds for 3d object detection in adverse weather. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR)*, pages 15283–15292, 2021. 1

[10] Alexander B. Jung, Kentaro Wada, Jon Crall, Satoshi Tanaka, Jake Graving, Christoph Reinders, Sarthak Yadav, Joy Banerjee, Gábor Vecsei, Adam Kraft, Zheng Rui, Jirka Borovec, Christian Vallentin, Semen Zhydenko, Kilian Pfeiffer, Ben Cook, Ismael Fernández, François-Michel De Rainville, Chi-Hung Weng, Abner Ayala-Acevedo, Raphael Meudec, Matias Laporte, et al. imgaug. https://github.com/aleju/imgaug, 2020. Online; accessed 01-Feb-2020. 1

[11] Velat Kilic, Deepti Hegde, Vishwanath Sindagi, A Brinton Cooper, Mark A Foster, and Vishal M Patel. Lidar light scattering augmentation (lisa): Physics-based simulation of adverse weather conditions for 3d object detection. *arXiv preprint arXiv:2107.07004*, 2021. 1

[12] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12697–12705, 2019. 4

[13] Shuangzhi Li, Zhijie Wang, Felix Juefei-Xu, Qing Guo, Xingyu Li, and Lei Ma. Common corruption robustness of point cloud detectors: Benchmark and enhancement. *arXiv preprint arXiv:2210.05896*, 2022. 2

[14] Zhiqi Li, Wenhai Wang, Hongyang Li, Enze Xie, Chonghao Sima, Tong Lu, Qiao Yu, and Jifeng Dai. Bevformer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers. In *European Conference on Computer Vision (ECCV)*, 2022. 4

[15] FEDERAL METEOROLOGICAL HANDBOOK No. Surface weather observations and reports. *US Department of Commerce/National Oceanic and Atmospheric Administration*, 2005. 1

[16] Meteorological Service of Canada. *MANOBS-Manual of Surface Weather Observations*. 2015. 1

[17] 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. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10529–10538, 2020. 4

[18] UjjwalSaxena. Project title. https://github.com/UjjwalSaxena/Automold--Road-Augmentation-Library, 2018. 1

[19] Wikipedia. Rain — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Rain&oldid=1108688066, 2022. [Online; accessed 13-September-2022]. 1

[20] Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. *Sensors*, 18(10):3337, 2018. 4

[21] Kaicheng Yu, Tang Tao, Hongwei Xie, Zhiwei Lin, Zhongwei Wu, Zhongyu Xia, Tingting Liang, Haiyang Sun, Jiong Deng, Dayang Hao, et al. Benchmarking the robustness of lidar-camera fusion for 3d object detection. *arXiv preprint arXiv:2205.14951*, 2022. 2

[22] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6023–6032, 2019. 4

[23] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations (ICLR)*, 2018. 4

[24] Jinlai Zhang, Lyujie Chen, Bo Ouyang, Binbin Liu, Jihong Zhu, Yujin Chen, Yanmei Meng, and Danfeng Wu. Pointcutmix: Regularization strategy for point cloud classification. *Neurocomputing*, 505:58–67, 2022. 4

[25] Yanan Zhang, Jiaxin Chen, and Di Huang. Cat-det: Contrastively augmented transformer for multi-modal 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 908–917, 2022. 6

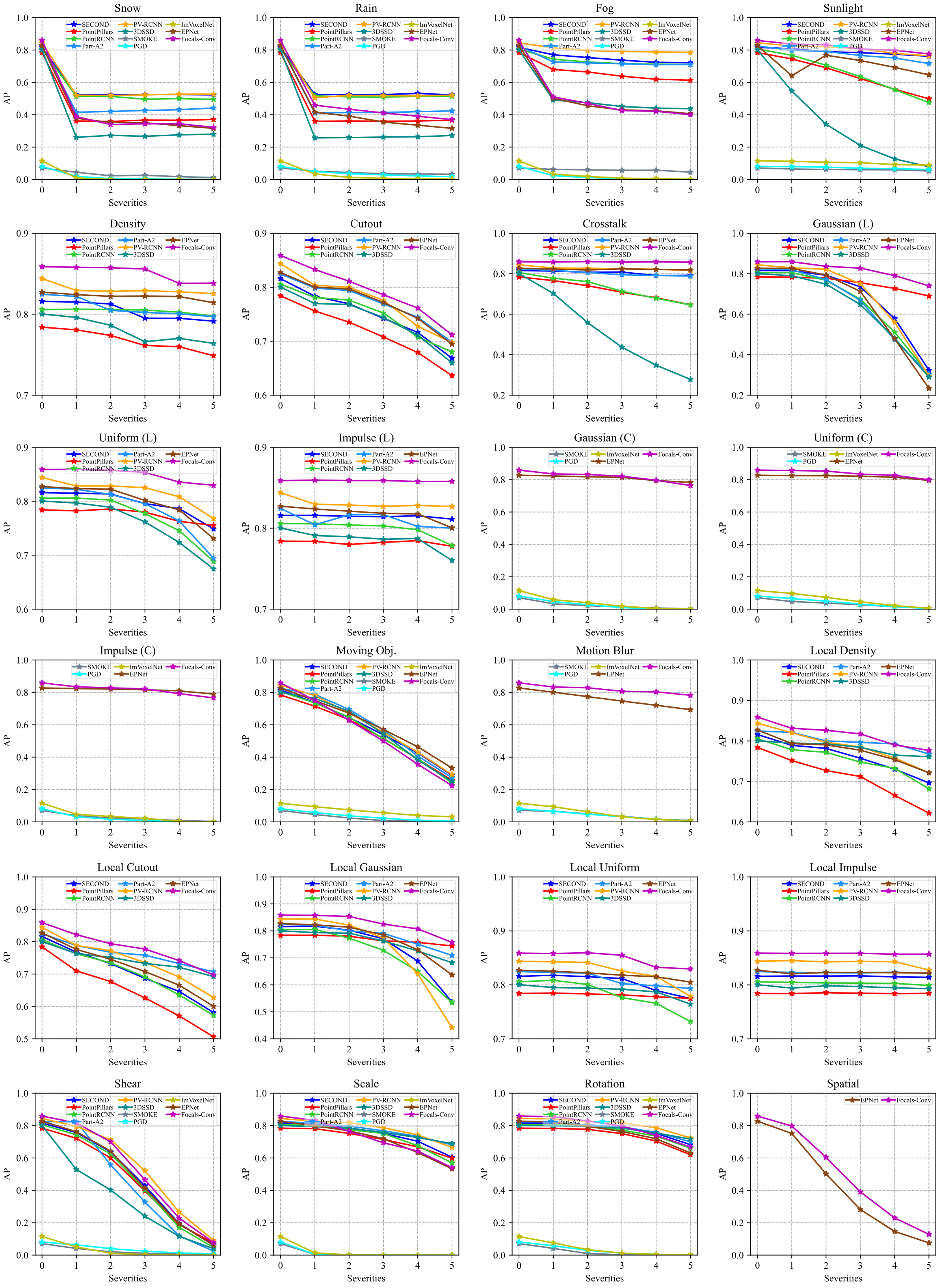

Figure B.1. Model performance w.r.t. severity of each corruption on **KITTI-C**. The results are evaluated based on the car class at moderate difficulty.

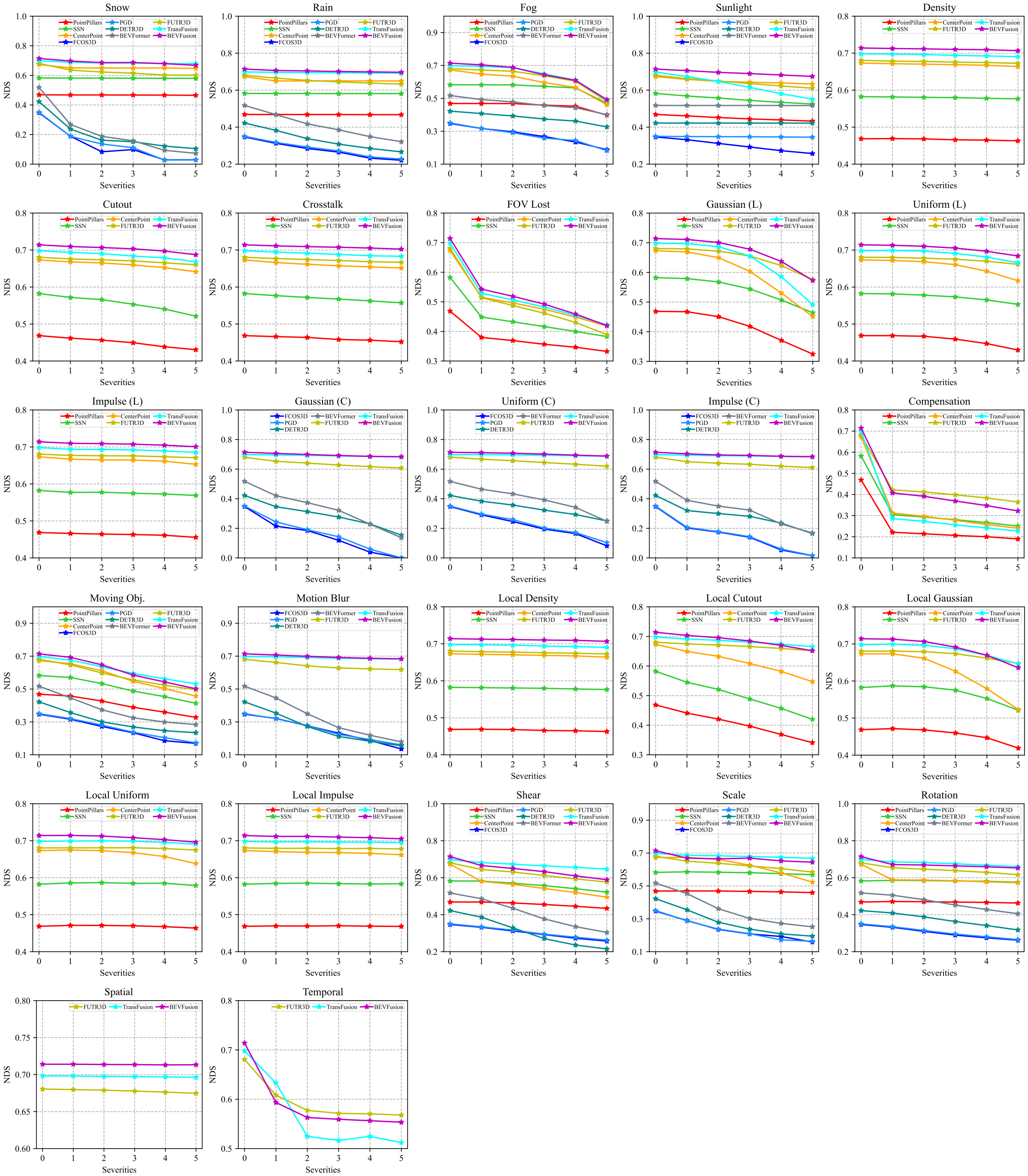

Figure C.1. Model performance w.r.t. severity of each corruption on **nuScenes-C** under the NDS metric.