LEARNING VOXEL-WISE OBSERVATION RELIABILITY FOR ROBUST LIDAR 3D OBJECT DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Voxel-based 3D detectors achieve a favorable accuracy-efficiency trade-off, but remain vulnerable to weather- and sensor-induced corruptions. A key limitation is that existing voxel-based pipelines typically treat all occupied voxels as equally reliable, even though corruption can introduce spurious or perturbed observations with substantially different reliability. Once encoded, unreliable voxel features can be repeatedly propagated through sparse convolutions, contaminating otherwise reliable geometric representations. To address this issue, we propose VORL, a **V**oxel-wise **O**bservation **R**eliability **L**earning framework for robust LiDAR 3D object detection. VORL estimates a continuous reliability score for each occupied voxel from its feature representation and lightweight observation cues, and introduces Reliability-aware Sparse Convolution (RaSC) to modulate source-voxel contributions before neighborhood aggregation. In this way, unreliable feature propagation is suppressed before feature mixing occurs, while reliable geometric evidence is largely preserved. Since voxel-wise reliability annotations are unavailable, we derive corruption-aware pseudo reliability targets from clean-to-corrupted correspondences available during corruption synthesis. Experiments on KITTI-C and nuScenes-C demonstrate consistent robustness improvements on diverse voxel-based detectors. On KITTI-C, VORL improves corrupted AP by 2.62-6.78 points across three baseline detectors while maintaining clean performance. More importantly, VORL can be seamlessly integrated into existing voxel-based detectors, while introducing only marginal computational and memory overhead.

1 INTRODUCTION

LiDAR-based 3D object detection is a fundamental perception task for autonomous driving and mobile robotics, where accurate geometric understanding of the surrounding environment is essential. Among existing paradigms, voxel-based detectors have been widely adopted due to their favorable trade-off between detection accuracy and computational efficiency. By discretizing irregular point clouds into structured voxel representations, they enable efficient geometric feature learning with sparse 3D convolutions.

Despite their strong performance under clean sensing conditions, voxel-based detectors remain vulnerable to weather- and sensor-induced corruptions. Adverse weather and sensor noise can introduce spurious returns or perturb valid measurements, substantially degrading detection performance. As shown in Fig. 1(a), four representative voxel-based detectors suffer pronounced performance drops on KITTI-C. The consistent degradation across different architectures indicates that robustness to corrupted LiDAR observations remains a common challenge for current voxel-based detection pipelines.

A key factor behind this vulnerability is the reliability of voxelized observations. After voxelization, occupied voxels are typically encoded and processed as valid inputs to the detector. However, *occupancy does not necessarily imply reliability*. Corruption may introduce spurious or displaced returns while leaving nearby valid surface observations largely intact, leading to substantial variation in observation reliability across neighboring voxels. Therefore, some occupied voxels may retain reliable geometric evidence, whereas others can be dominated by weather-induced scattering, abnormal reflections, or perturbed measurements. However, existing voxel-based pipelines do not explicitly model such voxel-wise differences in observation reliability.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

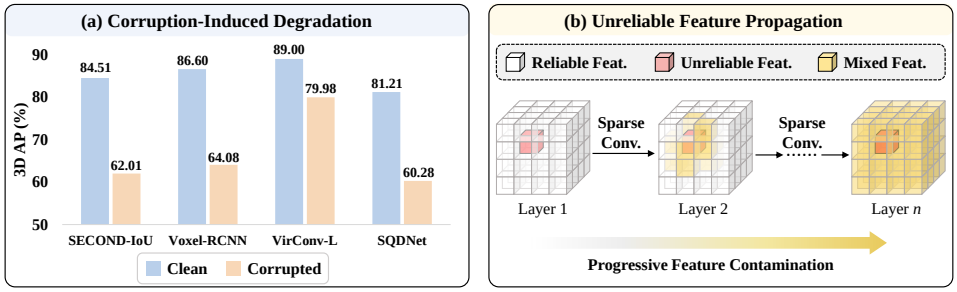


Figure 1: Corruption-induced performance degradation and unreliable feature propagation in voxel-based 3D detection. (a) Representative voxel-based detectors suffer significant performance degradation on KITTI-C under corruptions at severity level 5. (b) Standard sparse convolution indiscriminately propagates unreliable voxel features across layers, progressively contaminating otherwise reliable representations.

More importantly, the influence of unreliable observations does not remain localized after feature encoding. Standard sparse convolution aggregates features from neighboring active voxels without explicitly considering their observation reliability. As shown in Fig. 1(b), an unreliable source voxel can inject corrupted information into neighboring representations during aggregation. Once mixed with reliable features, the resulting representation may participate in subsequent convolutions and further influence surrounding voxels as the receptive field expands. Consequently, localized observation corruption can progressively propagate through the sparse backbone and interfere with otherwise reliable geometric representations. This suggests that corruption robustness in voxel-based detection is not only an input-quality problem, but also a feature-propagation problem.

Existing approaches to robust LiDAR perception mainly improve robustness through corruption-aware training Kong et al. (2023), input purification Cai et al. (2024); Sun et al. (2026), additional sensing or temporal context Song et al. (2024); Huang et al. (2025); Chae et al. (2025), and test-time adaptation Chen et al. (2025; 2024). While effective, these methods primarily operate at the input, frame, modality, or model level. They do not explicitly regulate how voxel-wise observations contribute to sparse feature aggregation according to their reliability. As a result, unreliable voxel features may still enter the backbone and propagate together with reliable geometric evidence.

Motivated by this observation, we propose VORL, a **V**oxel-wise **O**bservation **R**eliability **L**earning framework that makes sparse feature propagation explicitly reliability-aware. VORL predicts a continuous reliability score for each occupied voxel from its initial voxel feature and complementary observation cues, such as point support, range, intensity, local geometry, and neighborhood consistency. We further introduce Reliability-aware Sparse Convolution (RaSC), which uses the predicted reliability to modulate source-voxel contributions before neighborhood aggregation. In this way, unreliable feature propagation is suppressed before feature mixing occurs, while reliable geometric evidence is largely preserved. Since voxel-wise reliability annotations are unavailable, we construct corruption-aware pseudo reliability targets from clean-to-corrupted correspondences available during training-time corruption synthesis. VORL can be integrated into different voxel-based detectors with minimal architectural modification, while introducing only marginal computational and memory overhead. The main contributions of this work are summarized as follows:

- We identify voxel-wise observation reliability as an important yet underexplored factor in the corruption robustness of voxel-based LiDAR detectors, and highlight how unreliable voxel features can progressively propagate through sparse convolutional aggregation.
- We propose VORL, a lightweight framework that explicitly learns voxel-wise observation reliability and incorporates it into feature propagation through Reliability-aware Sparse Convolution. To enable reliability learning without manual annotations, we further develop corruption-aware pseudo supervision from clean-to-corrupted correspondences available during corruption synthesis.
- Experiments on KITTI-C and nuScenes-C demonstrate that VORL consistently improves corruption robustness across diverse voxel-based detectors while maintaining clean detection performance and introducing only marginal computational and memory overhead.

2 Related Work

LiDAR-based 3D object detection. Existing approaches can be broadly categorized into projection-based, point-based, and voxel-based methods. Projection-based methods Fan et al. (2021); Tian et al. (2022); Liang et al. (2021); Bai et al. (2023) transform irregular point clouds into structured 2D representations, such as range-view images or bird's-eye-view (BEV) maps, and then leverage mature 2D architectures for feature extraction and detection. Point-based methods Shi et al. (2019); Liu et al. (2023) operate directly on raw points and learn fine-grained geometric representations through point-wise feature encoding and local neighborhood aggregation. While preserving detailed point-level information, such methods can become computationally demanding when processing large-scale point clouds. Voxel-based methods discretize 3D space into structured cells and perform feature learning on voxelized representations, offering a favorable balance between detection accuracy and computational efficiency. VoxelNet Zhou & Tuzel (2018) introduces end-to-end voxel feature learning, while SECOND Yan et al. (2018) significantly improves efficiency through sparse 3D convolutions. Subsequent methods enhance voxel-based detection through stronger proposal refinement Deng et al. (2021), richer voxel representations Wu et al. (2023), and center-based prediction Yin et al. (2021). More recent approaches further improve efficiency and representation learning through fully sparse architectures Chen et al. (2023); Liu et al. (2025) or stronger long-range context modeling with sparse Transformers and state-space models Wang et al. (2023); Zhang et al. (2024). Despite remarkable progress in detection accuracy and efficiency, these methods are primarily developed and optimized under clean sensing conditions.

Robust 3D perception under common corruptions. Robust 3D perception is essential for reliable deployment under adverse sensing conditions. Dong et al. Dong et al. (2023) systematically evaluated 3D object detectors under weather- and sensor-induced corruptions on KITTI-C, nuScenes-C, and Waymo-C, revealing substantial performance degradation beyond clean conditions. Robo3D Kong et al. (2023) further studies robustness across diverse corruption types for both 3D detection and segmentation. Beyond common corruption benchmarks, Eskandar et al. Eskandar (2024) investigated the generalization of LiDAR-based detectors under weather, sensor, and geographic domain shifts. These studies highlight the vulnerability of current 3D perception systems to degraded sensing conditions and have motivated a growing body of work on corruption robustness.

Existing robustness approaches can be broadly grouped into corruption-aware training, multimodal or temporal enhancement, test-time adaptation, and input purification. Corruption-aware training, such as Robo3D Kong et al. (2023), improve corruption robustness through corruption-aware augmentation and training strategies. Multimodal methods compensate for degraded LiDAR observations using complementary sensing or semantic information. RoboFusion Song et al. (2024) leverages visual foundation-model features, while LiDAR-radar fusion methods Huang et al. (2025); Chae et al. (2025) exploit the complementary sensing characteristics of 4D radar under adverse conditions. Temporal approaches Li et al. (2026) exploit historical observations to compensate for degraded information in the current frame. Test-time adaptation methods Niu et al. (2023); Chen et al. (2025; 2024); Yang et al. (2026) adapt model parameters or feature statistics to distribution shifts encountered at inference time, whereas input purification methods such as LiDARPure Cai et al. (2024) and LTDNet Sun et al. (2026) attempt to restore corrupted point clouds before detection. While effective, these methods primarily improve robustness at the input, frame, modality, or model level. They do not explicitly distinguish the reliability of individual voxelized observations or regulate how unreliable voxel features participate in sparse convolutional aggregation. Consequently, corruption-induced unreliable features can still propagate through the backbone together with reliable geometric evidence. In contrast, VORL explicitly models voxel-wise observation reliability and introduces it directly into sparse feature propagation, suppressing unreliable source contributions before neighborhood aggregation. This design improves corruption robustness without requiring input restoration, online adaptation, or modification of the pretrained detector parameters.

3 Method

3.1 Overview

Voxel-based 3D object detectors discretize irregular LiDAR point clouds into structured voxel representations and employ sparse 3D convolutional backbones for geometric feature learning. However,

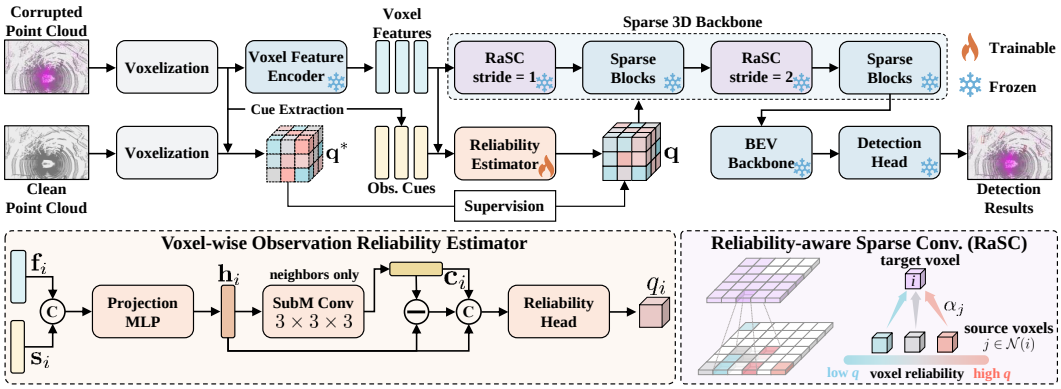


Figure 2: Overview of VORL. Given a corrupted point cloud, VORL estimates voxel-wise observation reliability from voxel features and complementary observation cues, and uses the predicted reliability to regulate feature propagation in the sparse 3D backbone through Reliability-aware Sparse Convolution. During training, clean-to-corrupted correspondences are used only to construct pseudo reliability targets for supervising the reliability estimator, while the pretrained detector remains frozen. Ⓒ denotes feature concatenation, while ⊖ denotes element-wise subtraction.

under corrupted sensing conditions, occupied voxels can exhibit markedly different observation reliability. Some voxels may retain stable geometric evidence from physical surfaces, whereas others can be dominated by noisy returns, abnormal reflections, displaced measurements, or weather-induced scattering. Nevertheless, standard voxel-based pipelines propagate features from all occupied voxels without explicitly accounting for such reliability differences. As a result, unreliable voxel features can be mixed into neighboring representations during sparse convolution and progressively propagate through the backbone, interfering with otherwise reliable geometric evidence and ultimately degrading detection performance.

To address this issue, we propose VORL, a voxel-wise observation reliability learning framework that makes sparse feature propagation explicitly reliability-aware. As illustrated in Fig. 2, VORL predicts a continuous reliability score for each occupied voxel from its voxel feature and complementary observation cues. The predicted reliability is then incorporated into the sparse backbone through Reliability-aware Sparse Convolution (RaSC). Rather than treating all active source voxels equally, RaSC modulates each source-voxel contribution according to its reliability before neighborhood aggregation, thereby suppressing unreliable feature propagation while largely preserving reliable geometric evidence. Since voxel-wise observation reliability is not annotated in existing 3D detection datasets, we construct corruption-aware pseudo reliability targets from clean-to-corrupted correspondences available during corruption synthesis. These targets are used to supervise the reliability estimator during training. At inference, VORL predicts reliability directly from the observed point cloud and requires neither clean references nor corruption labels. Moreover, the pretrained detector remains frozen and only a lightweight reliability module is introduced, allowing VORL to be incorporated into different voxel-based detectors with minimal architectural modification and negligible computational and memory overhead.

3.2 Learning Voxel-wise Observation Reliability

Our goal is to estimate a continuous reliability score $q_i \in (0, 1)$ for each occupied voxel, measuring the reliability of its underlying LiDAR observations under corrupted sensing conditions. Given an input LiDAR point cloud, let $\mathcal{V} = \{v_i\}_{i=1}^{N}$ denote the set of occupied voxels. For each voxel v_i, we denote by $\mathcal{P}_i^{\text{obs}}$ all points initially assigned to the voxel. For computational efficiency, when a voxel contains more than $N_{\max}$ points, at most $N_{\max}$ points are retained for voxel feature encoding, yielding $\widehat{\mathcal{P}}_i \subseteq \mathcal{P}_i^{\text{obs}}$. The corresponding voxel feature is computed as $\mathbf{f}_i = \text{VFE}(\widehat{\mathcal{P}}_i)$.

While $\mathbf{f}_i$ provides an effective representation for object detection, it is computed from the retained point subset $\widehat{\mathcal{P}}_i$. Observation reliability also depends on low-level sensing characteristics that are not explicitly encoded by $\mathbf{f}_i$. Therefore, we complement $\mathbf{f}_i$ with a lightweight observation descrip-

tor computed from $\mathcal{P}_i^{\text{obs}}$. The descriptor captures five complementary properties, including point support, sensing range, return intensity, intra-voxel geometry, and local observation consistency.

We first characterize the point support of voxel v_i by $\rho_i = |\mathcal{P}_i^{\text{obs}}|/N_{\max}$. We leave ρ_i unclipped so that a value larger than one retains information about point density beyond the VFE retention limit. We further encode the sensing range using the normalized radial distance $\bar{r}_i = \text{clip}\Big(\sqrt{x_i^2 + y_i^2}/r_{\max}, 0, 1\Big)$, where (x_i, y_i, z_i) denotes the voxel center and $r_{\max}$ is the maximum radial distance within the detection range.

For return intensity, let I_p denote the normalized intensity of point $p \in \mathcal{P}_i^{\text{obs}}$. We summarize the intensity distribution within each voxel using its mean, standard deviation, and maximum, denoted by μ_i^I, σ_i^I, and $I_i^{\max}$, respectively. The intra-voxel geometry is characterized by the normalized coordinate variance

$$\boldsymbol{\nu}_i^{xyz} = 4\left[\text{Var}_i(x)/\delta_x^2, \text{Var}_i(y)/\delta_y^2, \text{Var}_i(z)/\delta_z^2\right], \tag{1}$$

where $\text{Var}_i(\cdot)$ is computed over $\mathcal{P}_i^{\text{obs}}$ and $(\delta_x, \delta_y, \delta_z)$ denotes the voxel size. We further capture the local observation consistency of voxel v_i using its $3 \times 3 \times 3$ grid neighborhood $\mathcal{G}(i)$. For each location $k \in \mathcal{G}(i)$, we set $\widetilde{\rho}_k = \rho_k$ if the corresponding voxel is occupied, and $\widetilde{\rho}_k = 0$ otherwise. We then define the neighborhood occupancy ratio o_i and point-support discrepancy $\Delta\rho_i$ as

$$o_i = \frac{1}{|\mathcal{G}(i)|}\sum_{k\in\mathcal{G}(i)} \mathbb{1}[k \text{ is occupied}], \quad \Delta\rho_i = \left|\rho_i - \frac{1}{|\mathcal{G}(i)|}\sum_{k\in\mathcal{G}(i)} \widetilde{\rho}_k\right|. \tag{2}$$

The resulting 10-dimensional observation descriptor is

$$\mathbf{s}_i = [\rho_i, \bar{r}_i, \mu_i^I, \sigma_i^I, I_i^{\max}, \boldsymbol{\nu}_i^{xyz}, o_i, \Delta\rho_i] \in \mathbb{R}^{10}. \tag{3}$$

Although $\mathbf{s}_i$ captures explicit sensing statistics and density-level consistency, it does not fully characterize feature-level compatibility with the surrounding geometry. Therefore, we concatenate $\mathbf{f}_i$ and $\mathbf{s}_i$ and project them into a local embedding $\mathbf{h}_i = \phi([\mathbf{f}_i, \mathbf{s}_i])$. A $3 \times 3 \times 3$ submanifold sparse convolution then aggregates neighboring embeddings to produce a contextual representation $\mathbf{c}_i$.

Finally, we combine the local embedding, its spatial context, and their discrepancy to predict voxel reliability:

$$\mathbf{r}_i = [\mathbf{h}_i, \mathbf{c}_i, |\mathbf{h}_i - \mathbf{c}_i|], \quad q_i = \exp[-\text{Softplus}\left(\text{MLP}(\mathbf{r}_i)\right)]. \tag{4}$$

A larger q_i indicates a more reliable voxel observation. The resulting reliability scores are subsequently incorporated into RaSC to regulate sparse feature propagation.

3.3 Reliability-aware Sparse Convolution

After obtaining estimated voxel-wise observation reliability, we incorporate it directly into sparse feature aggregation. For a target voxel v_i, standard sparse convolution aggregates features from its active source voxels $\mathcal{N}(i)$ as:

$$\mathbf{f}_i^{\text{out}} = \sum_{j\in\mathcal{N}(i)} \mathbf{W}_{\boldsymbol{\delta}_{ij}} \mathbf{f}_j^{\text{in}}, \tag{5}$$

where $\boldsymbol{\delta}_{ij}$ denotes the relative offset from v_i to source voxel v_j, and $\mathbf{W}_{\boldsymbol{\delta}_{ij}}$ is the convolution kernel associated with this offset. Although this operation accounts for spatial structure, it treats all active source features identically with respect to observation reliability. Consequently, features originating from unreliable observations can be aggregated and propagated together with reliable geometric evidence.

To explicitly regulate this process, we introduce Reliability-aware Sparse Convolution (RaSC), which modulates each source-voxel contribution before neighborhood aggregation:

$$\mathbf{f}_i^{\text{out}} = \sum_{j\in\mathcal{N}(i)} \alpha_j \mathbf{W}_{\boldsymbol{\delta}_{ij}} \mathbf{f}_j^{\text{in}}, \tag{6}$$

where $\alpha_j \in (0, 1]$ is a scalar modulation coefficient derived from the predicted reliability q_j of source voxel v_j. A straightforward choice is $\alpha_j = q_j$, which directly attenuates low-reliability

features. However, aggressively scaling features by the predicted reliability may also suppress useful geometric evidence when the reliability estimate is imperfect. Therefore, we introduce a nonlinear reliability mapping that suppresses unreliable source features while avoiding excessive attenuation of reliable ones:

$$\alpha_j = \frac{q_j}{q_j + (1 - q_j)^2}. \tag{7}$$

The mapping is continuous and parameter-free, with $\alpha_j > q_j$ for $0 < q_j < 1$, resulting in milder attenuation than direct reliability weighting. Moreover, $1 - \alpha_j = \frac{(1-q_j)^2}{q_j+(1-q_j)^2}$ vanishes quadratically as q_j approaches one. Consequently, unreliable features are strongly suppressed, while reliable geometric evidence is largely preserved.

The key distinction between RaSC and post-hoc feature reweighting lies in *when* reliability is introduced. Post-hoc reweighting operates after neighboring features have already been aggregated, at which point information from reliable and unreliable sources has been mixed. In contrast, RaSC modulates each source contribution *before* aggregation, preventing unreliable features from imposing their full influence on neighboring representations. By intervening directly at the feature propagation stage, RaSC suppresses the spread of unreliable information while preserving the sparse convolutional structure of the original detector.

3.4 Corruption-aware Pseudo Reliability Supervision

Voxel-wise observation reliability is not directly annotated in existing 3D object detection datasets. However, clean-to-corrupted point correspondences and point-level corruption metadata are available during synthetic corruption generation. We exploit these signals to construct voxel-wise pseudo reliability targets that reflect the observable degradation induced by different corruption mechanisms. Since only the retained points $\widehat{\mathcal{P}}_i$ are used for voxel feature encoding, the pseudo target for voxel v_i is defined over this set.

For corruptions with explicit point-level corruption status, including snow, rain, fog, sunlight, and crosstalk, we define voxel reliability according to the fraction of retained observations that remain reliable. Snow, rain, and fog introduce weather-induced fake returns, whereas sunlight and crosstalk displace existing returns. Let $\mathbf{p}_{ij}$ denote the j-th retained point in voxel v_i, and $m_{ij} \in \{0, 1\}$ indicate whether $\mathbf{p}_{ij}$ is unreliable, where $m_{ij} = 1$ denotes a fake or displaced return. The pseudo reliability target is defined as

$$q_i^* = 1 - \frac{1}{|\widehat{\mathcal{P}}_i|} \sum_j m_{ij}. \tag{8}$$

Thus, q_i^* measures the proportion of retained observations that remain reliable within the voxel, assigning lower reliability to voxels containing a larger fraction of corrupted returns.

For Gaussian, uniform, and impulse noise, corruption instead manifests as continuous perturbations of existing returns. A binary corruption indicator cannot distinguish mild perturbations from severe ones, so we define reliability according to the magnitude of the induced displacement. For each retained point, let $(\Delta x_{ij}, \Delta y_{ij}, \Delta z_{ij}, \Delta I_{ij})$ denote its difference from the corresponding clean return. We normalize the displacement by the voxel resolution and input intensity range:

$$d_{ij} = \sqrt{\frac{1}{4}\left[\left(\frac{\Delta x_{ij}}{\delta_x}\right)^2 + \left(\frac{\Delta y_{ij}}{\delta_y}\right)^2 + \left(\frac{\Delta z_{ij}}{\delta_z}\right)^2 + \left(\frac{\Delta I_{ij}}{I_{\max}}\right)^2\right]}, \tag{9}$$

where $(\delta_x, \delta_y, \delta_z)$ denotes the voxel size and $I_{\max}$ denotes the maximum intensity. The voxel-wise pseudo reliability is then obtained by averaging the corresponding point-wise reliability:

$$q_i^* = \frac{1}{|\widehat{\mathcal{P}}_i|} \sum_j \frac{1}{1 + d_{ij}}. \tag{10}$$

We supervise the predicted reliability q_i using binary cross-entropy with the continuous pseudo target q_i^*. Because different corruption domains may contain different numbers of occupied voxels, directly averaging the loss over all voxels would bias training toward domains with larger voxel

counts. Therefore, we first average the reliability loss within each corruption domain and then average across the domains represented in the current batch:

$$\mathcal{L}_{\text{rel}} = \frac{1}{|\mathcal{D}_{\mathcal{B}}|} \sum_{d \in \mathcal{D}_{\mathcal{B}}} \frac{1}{|\mathcal{V}_d|} \sum_{i \in \mathcal{V}_d} \ell_i, \tag{11}$$

where $\mathcal{V}_d$ denotes the set of occupied voxels from corruption domain d, and $\mathcal{D}_{\mathcal{B}}$ denotes the set of corruption domains represented in the current batch. This domain-wise averaging prevents domains with larger voxel counts from dominating reliability learning.

3.5 Training and Inference

During training, we randomly apply a corruption to each training point cloud and use the resulting clean-to-corrupted correspondences and corruption metadata to construct the pseudo reliability targets q_i^*. Starting from a detector pretrained on clean data, we freeze all detector parameters and optimize only the reliability estimator using $\mathcal{L}_{\text{rel}}$. The clean point clouds and corruption metadata are used solely to generate supervision and are not required as inputs to the reliability estimator.

At inference, all training-only information, including clean references, corruption metadata, and pseudo targets, is discarded. VORL estimates voxel-wise observation reliability directly from the observed point cloud and uses the predicted reliability to regulate sparse feature propagation through RaSC within the frozen detector backbone.

4 Experiment

4.1 Experimental Settings

Datasets. We evaluate VORL on two widely used corruption benchmarks, including KITTI-C and nuScenes-C. These benchmarks are constructed by applying synthetic corruptions to the validation sets of KITTI Geiger et al. (2012) and nuScenes Caesar et al. (2020), respectively. In this work, we focus on weather- and sensor-related corruptions that directly degrade LiDAR observations.

Evaluation metrics. We follow the standard evaluation protocols for existing 3D detection datasets. For KITTI-C, we report average precision (AP) at an IoU threshold of 0.7 for the Car category. For nuScenes-C, we adopt mean average precision (mAP) and the nuScenes Detection Score (NDS).

Implementation details. We implement VORL using the open sourced 3D object detection framework *OpenPCDet* and follow the standard training configurations of the corresponding baseline detectors. For VORL training, we initialize from detectors pretrained on clean data, freeze all detector parameters, and optimize only the reliability estimator. We train the reliability estimator using AdamW with a learning rate of 3×10^{-4} and a batch size of 16 for 8 epochs on KITTI-C and 3 epochs on nuScenes-C. RaSC is applied to the input stage and the second stage of the sparse convolution backbone network. The trained reliability model is shared across all corruption types and requires neither corruption labels nor clean references at inference. All experiments were conducted on a computer equipped with two NVIDIA GeForce RTX 4090 GPUs.

4.2 Main Results

Results on the KITTI-C dataset. Table 1 reports the 3D detection results on KITTI-C at severity level 5. VORL consistently improves corruption robustness across all three voxel-based detectors and significantly outperforms the compared test-time adaptation methods such as MOS in terms of overall AP_{cor}. Specifically, VORL improves AP_{cor} by 4.39, 2.62, and 6.78 points for Voxel-RCNN, VirConv-L, and SQDNet, respectively. The gains are observed across all eight corruption types, with particularly pronounced improvements under challenging corruptions such as sunlight, Gaussian noise, and fog. These gains are achieved while largely preserving detection performance on clean data, with clean AP remaining within 0.21 points of the corresponding baseline for each detector. More importantly, VORL also introduces only marginal inference overhead, maintaining near-baseline FPS and GPU memory consumption.

Method	Clean	Weather-related				Sensor-related				AP_{cor} ↑	FPS ↑	Mem. (GB) ↓
		Snow	Rain	Fog	Sunlight	Crosstalk	Gaussian	Uniform	Impulse			
Voxel-RCNN Deng et al. (2021)	**86.60**	50.20	45.83	71.26	76.01	83.13	31.83	72.30	82.05	64.08	**38.69**	**2.05**
+ TENT Wang et al. (2021)	82.98	47.23	43.60	72.21	71.66	78.52	31.52	73.06	81.19	62.37	23.43	7.54
+ SAR Niu et al. (2023)	85.80	48.57	46.04	73.36	76.10	80.41	37.92	**76.79**	81.96	65.14	13.58	8.71
+ MOS Chen et al. (2025)	85.17	47.84	41.74	73.26	79.99	75.36	**39.86**	71.87	79.08	63.63	1.19	10.09
+ VORL (Ours)	86.40	**55.09**	**50.53**	**75.75**	**84.62**	**85.69**	36.94	76.37	**82.76**	**68.47**	37.71	2.16
Δ Gain	-0.20	**+4.89**	**+4.70**	**+4.49**	**+8.61**	**+2.56**	**+5.11**	**+4.07**	**+0.71**	**+4.39**	–	–
VirConv-L Wu et al. (2023)	**89.00**	76.73	75.19	82.49	86.68	88.80	70.86	72.96	86.16	79.98	**18.83**	**2.47**
+ TENT Wang et al. (2021)	87.12	75.25	73.38	83.72	85.29	88.04	71.68	73.65	84.32	79.42	18.17	2.97
+ SAR Niu et al. (2023)	88.41	**78.91**	**77.11**	82.83	86.83	88.21	74.67	77.87	85.87	81.54	11.92	4.82
+ MOS Chen et al. (2025)	88.79	78.58	75.69	80.02	86.27	87.01	**80.90**	**81.29**	85.33	81.89	0.77	10.16
+ VORL (Ours)	88.92	77.08	76.08	**87.19**	**87.06**	**88.92**	77.67	80.66	**86.17**	**82.60**	17.70	2.55
Δ Gain	-0.08	**+0.35**	**+0.89**	**+4.70**	**+0.38**	**+0.12**	**+6.81**	**+7.70**	**+0.01**	**+2.62**	–	–
SQDNet Mo et al. (2024)	81.21	59.70	57.80	61.23	67.55	72.80	36.99	52.91	73.28	60.28	**16.11**	**9.41**
+ TENT Wang et al. (2021)	80.03	53.71	57.91	47.31	67.86	73.21	35.17	51.65	72.18	57.38	9.75	21.78
+ SAR Niu et al. (2023)	**81.63**	60.45	58.27	61.89	66.45	79.26	37.00	53.48	73.04	61.23	6.01	21.81
+ MOS Chen et al. (2025)	80.78	59.31	57.04	60.69	69.28	73.44	**54.20**	57.27	75.40	63.33	1.13	15.48
+ VORL (Ours)	81.00	**63.02**	**60.97**	**71.52**	**79.46**	**80.81**	45.06	**60.12**	**75.54**	**67.06**	15.46	9.98
Δ Gain	-0.21	**+3.32**	**+3.17**	**+10.29**	**+11.91**	**+8.01**	**+8.07**	**+7.21**	**+2.26**	**+6.78**	–	–

Table 1: Robustness comparison on the Car category of KITTI-C at severity level 5. Δ Gain denotes the improvement of VORL over the corresponding clean-pretrained baseline.

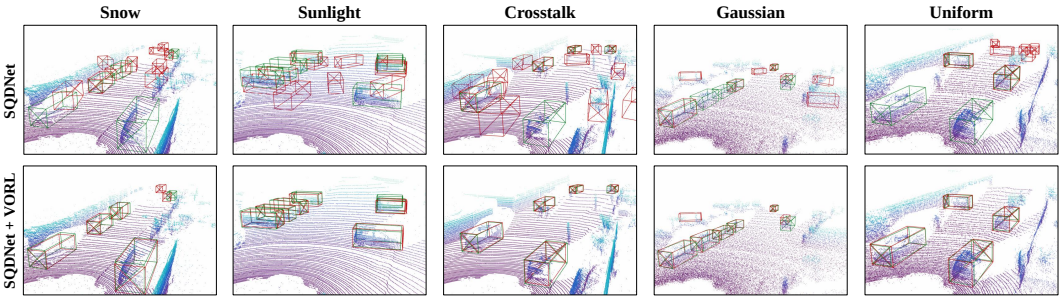


Figure 3: Qualitative comparison of SQDNet and its VORL variant on KITTI-C under snow, sunlight, crosstalk, Gaussian noise and uniform noise at severity level 5. The ground-truth boxes and predicted boxes are shown in green and red, respectively.

Qualitative results. Figure 3 compares SQDNet and its VORL variant on KITTI-C under representative weather- and sensor-related corruptions. We observe that incorporating VORL can suppress spurious detections while preserving valid objects, leading to cleaner and more stable predictions.

Results on the nuScenes-C dataset. Table 2 presents the 3D detection results on nuScenes-C at severity level 5. We find that VORL is especially effective for corruptions that create strongly abnormal or spatially inconsistent returns. The largest improvements are observed under crosstalk, where mAP increases by 30.02 and 32.64 points for TransFusion-L and VoxelNeXt, respectively. Significant gains are also achieved under several weather-related corruptions. In particular, applying VORL to FSHNet improves mAP by 10.41, 6.77, and 12.75 under snow, rain, and fog, respectively, while VoxelNeXt gains 14.92 points under sunlight. By contrast, improvements under Gaussian, uniform, and impulse noise are relatively smaller. Notably, clean-data performance remains largely unaffected as KITTI-C, with mAP changes within 0.08 points across all detectors.

4.3 Ablation Studies

Reliability estimator. Table 3 analyzes the three components of the reliability estimator. Using only voxel features achieves 65.43 AP_{cor}, while using only observation cues yields lower robustness. Combining the two improves AP_{cor} to 66.32, showing that explicit observation statistics provide complementary information beyond voxel representations. Adding local context further raises AP_{cor} to 67.06, validating the benefit of neighborhood information for voxel-wise reliability estimation.

Reliability utilization. Table 4 investigates how voxel reliability should be incorporated into sparse feature propagation. Hard masking, which removes source voxels with $q < 0.5$, yields only a modest improvement over standard sparse convolution. Post-hoc reweighting further decreases AP_{cor} from 60.28 to 59.37, suggesting that reweighting features after neighborhood aggregation cannot

Method	Clean	Weather-related								Sensor-related							
		Snow		Rain		Fog		Sunlight		Crosstalk		Gaussian		Uniform		Impulse	
	mAP	mAP	NDS	mAP	NDS	mAP	NDS	mAP	NDS	mAP	NDS	mAP	NDS	mAP	NDS	mAP	NDS
TransFusion-L Bai et al. (2022)	**64.58**	48.09	59.77	54.84	63.79	20.74	42.41	0.49	11.04	4.29	24.43	28.05	45.74	24.81	43.77	56.04	63.76
+ TENT Wang et al. (2021)	63.73	48.94	60.41	54.07	63.26	1.95	21.17	2.24	7.56	12.11	34.78	27.47	45.03	22.42	42.52	42.33	49.58
+ SAR Niu et al. (2023)	64.19	50.70	61.07	55.36	63.95	**23.90**	**43.94**	7.30	**29.13**	34.02	50.27	**29.42**	**47.26**	**26.26**	**44.73**	55.29	62.13
+ MOS Chen et al. (2025)	55.13	50.27	60.45	53.14	62.62	21.35	42.89	**8.67**	25.30	3.67	24.95	26.39	45.51	25.68	44.31	55.85	63.52
+ VORL (Ours)	64.50	**52.43**	**61.97**	**57.13**	**65.01**	23.56	43.54	6.01	28.87	**34.31**	**52.07**	28.12	45.90	26.17	44.67	**56.13**	**63.84**
Δ Gain	-0.08	**+4.34**	**+2.20**	**+2.29**	**+1.22**	**+2.82**	**+1.13**	**+5.52**	**+17.83**	**+30.02**	**+27.64**	**+0.07**	**+0.16**	**+1.36**	**+0.90**	**+0.09**	**+0.08**
VoxelNeXt Chen et al. (2023)	**60.51**	46.93	59.04	52.49	62.15	21.97	44.85	1.95	19.62	7.31	26.94	24.48	42.96	25.22	43.82	48.27	59.08
+ TENT Wang et al. (2021)	59.66	46.97	59.43	52.29	61.81	22.05	44.74	4.38	25.91	11.17	34.60	23.26	41.53	26.08	44.38	48.45	59.17
+ SAR Niu et al. (2023)	59.77	46.89	58.90	51.84	61.59	22.04	44.91	2.21	20.32	9.86	32.85	**25.15**	**43.74**	**27.89**	**45.93**	**49.74**	**60.31**
+ MOS Chen et al. (2025)	57.12	47.03	59.21	53.19	62.85	21.15	44.68	7.34	17.29	19.43	34.79	22.06	40.51	24.85	42.62	48.91	59.23
+ VORL (Ours)	60.45	**50.40**	**60.97**	**54.59**	**63.27**	**25.17**	**46.62**	**16.87**	**39.31**	**39.95**	**54.57**	24.63	43.27	26.39	44.65	48.60	59.37
Δ Gain	-0.06	**+3.47**	**+1.93**	**+2.10**	**+1.12**	**+3.20**	**+1.77**	**+14.92**	**+19.69**	**+32.64**	**+27.63**	**+0.15**	**+0.31**	**+1.17**	**+0.83**	**+0.33**	**+0.29**
FSHNet Liu et al. (2025)	**68.57**	48.85	59.56	56.45	64.46	17.03	35.81	2.51	19.17	6.23	25.39	27.20	44.81	14.86	35.52	58.32	65.22
+ TENT Wang et al. (2021)	65.23	30.30	46.85	31.20	48.88	9.17	26.40	3.21	20.69	7.95	27.37	26.10	43.64	15.28	35.86	57.53	64.72
+ SAR Niu et al. (2023)	68.33	48.74	59.04	56.31	64.09	16.67	35.10	4.54	21.27	9.56	28.77	26.90	44.47	14.82	35.19	57.73	64.78
+ MOS Chen et al. (2025)	67.14	48.57	59.06	56.91	64.32	17.76	35.94	5.25	22.74	12.64	31.45	25.99	43.97	14.93	35.25	58.06	65.11
+ VORL (Ours)	68.51	**59.26**	**65.85**	**63.22**	**68.72**	**29.78**	**47.34**	**7.63**	**25.52**	**15.37**	**34.23**	**27.32**	**44.90**	**15.46**	**36.19**	**58.50**	**65.43**
Δ Gain	-0.06	**+10.41**	**+6.29**	**+6.77**	**+4.26**	**+12.75**	**+11.53**	**+5.12**	**+6.35**	**+9.14**	**+8.84**	**+0.12**	**+0.09**	**+0.60**	**+0.67**	**+0.18**	**+0.21**

Table 2: Robustness comparison on nuScenes-C at severity level 5. We report mAP and NDS across all object classes under each corruption.

Voxel feat.	Obs. cues	Context	Clean ↑	AP_{cor} ↑
			81.21	60.28
✓			80.98	65.43
	✓		81.29	58.51
✓	✓		**81.45**	66.32
✓	✓	✓	81.00	**67.06**

Table 3: Ablation of the reliability estimator on KITTI-C with SQDNet.

Strategy	Clean ↑	AP_{cor} ↑
Standard SpConv	**81.21**	60.28
Hard masking	80.97	61.77
Post-hoc reweighting	80.87	59.37
Linear pre-aggregation	80.53	66.50
RaSC (Ours)	81.00	**67.06**

Table 4: Ablation of reliability utilization strategies on KITTI-C with SQDNet.

effectively mitigate the influence of unreliable observations because their features have already been mixed with reliable ones. In contrast, incorporating reliability before aggregation is substantially more effective. Direct weighting with q improves AP_{cor} to 66.50, but decreases clean AP by 0.68 points. RaSC further raises AP_{cor} to 67.06 while reducing the clean AP drop to only 0.21 points, showing a better balance between suppressing unreliable features and preserving useful information.

Pseudo reliability target. Table 5 evaluates different designs of pseudo reliability supervision. Binary targets are overly coarse, as they ignore the proportion and magnitude of corrupted observations, resulting in only 58.87 AP_{cor}. Continuous targets are substantially more effective, with count-based and magnitude-based formulations achieving 64.88 and 63.00 AP_{cor}, respectively. Their performance varies across corruption mechanisms, indicating that a single formulation is insufficient to characterize different types of observation degradation. By matching the target formulation to the underlying corruption mechanism, our corruption-aware target achieves the best AP_{cor} of 67.06.

Pseudo target	Clean	Weather-related				Sensor-related				AP_{cor} ↑
		Snow	Rain	Fog	Sunlight	Crosstalk	Gaussian	Uniform	Impulse	
Baseline	**81.21**	59.70	57.80	61.23	67.55	72.80	36.99	52.91	73.28	60.28
Binary	80.94	46.48	43.63	71.81	73.55	77.97	37.88	53.70	65.91	58.87
Count-based	80.97	62.05	59.68	**72.64**	78.34	80.62	41.71	56.87	67.14	64.88
Magnitude-based	81.01	50.53	48.89	70.44	74.86	78.45	**46.42**	58.96	75.44	63.00
Corruption-aware (Ours)	81.00	**63.02**	**60.97**	71.52	**79.46**	**80.81**	45.06	**60.12**	**75.54**	**67.06**

Table 5: Ablation of pseudo reliability targets on KITTI-C with SQDNet.

5 Conclusion

In this paper, we propose VORL, a voxel-wise observation reliability learning framework for improving the robustness of voxel-based 3D object detectors under corrupted sensing conditions. VORL explicitly estimates the reliability of each occupied voxel and incorporates the predicted reliability into sparse feature propagation. We further introduce Reliability-aware Sparse Convolution, which modulates source-voxel contributions before neighborhood aggregation to suppress the propagation of unreliable features while largely preserving reliable geometric evidence. To enable reliability learning without manual annotations, we develop corruption-aware pseudo reliability supervision that derives soft voxel-level targets from clean-to-corrupted correspondences available during training. Experiments on KITTI-C and nuScenes-C demonstrate consistent robustness improvements across different detectors and corruption types. Moreover, VORL can be integrated into various voxel-based detectors with minimal additional computational and memory overhead.

AI USE STATEMENT

In this work, we used generative AI tools for language polishing and improving the clarity and readability of the manuscript. We have not used generative AI tools for research idea, methodology design, and result analysis. All AI-assisted content was carefully reviewed and revised by the authors, who take responsibility for the final manuscript of this work.

REFERENCES

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)*, pp. 1090–1099, 2022.

Yeqi Bai, Ben Fei, Youquan Liu, Tao Ma, Yuenan Hou, Botian Shi, and Yikang Li. Rangeperception: Taming lidar range view for efficient and accurate 3D object detection. *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 36:79557–79569, 2023.

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 Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11621–11631, 2020.

Mumuxin Cai, Xupeng Wang, Ferdous Sohel, and Hang Lei. Diffusion models-based purification for common corruptions on robust 3d object detection. *Sensors*, 24(16):5440, 2024.

Yujeong Chae, Heejun Park, Hyeonseong Kim, and Kuk-Jin Yoon. Doppler-aware lidar-radar fusion for weather-robust 3d detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27197–27208. IEEE, 2025.

Yukang Chen, Jianhui Liu, Xiangyu Zhang, Xiaojuan Qi, and Jiaya Jia. Voxelnext: Fully sparse voxelnet for 3d object detection and tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21674–21683. IEEE, 2023.

Zhuoxiao Chen, Zixin Wang, Yadan Luo, Sen Wang, and Zi Huang. Dpo: dual-perturbation optimization for test-time adaptation in 3d object detection. In *Proceedings of the ACM International Conference on Multimedia (ACM MM)*, pp. 4138–4147, 2024.

Zhuoxiao Chen, Junjie Meng, Mahsa Baktashmotlagh, Yonggang Zhang, Zi Huang, and Yadan Luo. Mos: Model synergy for test-time adaptation on lidar-based 3d object detection. In *Proceedings of the International Conference on Learning Representations (ICLR)*, volume 2025, pp. 92784–92808, 2025.

Jiajun Deng, Shaoshuai Shi, Peiwei Li, Wengang Zhou, Yanyong Zhang, and Houqiang Li. Voxel r-cnn: Towards high performance voxel-based 3D object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 35, pp. 1201–1209, 2021.

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 (CVPR)*, pp. 1022–1032, 2023.

George Eskandar. An empirical study of the generalization ability of lidar 3d object detectors to unseen domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23815–23825. IEEE, 2024.

Lue Fan, Xuan Xiong, Feng Wang, Naiyan Wang, and Zhaoxiang Zhang. Rangedet: In defense of range view for lidar-based 3D object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2918–2927, 2021.

Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3354–3361. IEEE, 2012.

Xun Huang, Ziyu Xu, Hai Wu, Jinlong Wang, Qiming Xia, Yan Xia, Jonathan Li, Kyle Gao, Chenglu Wen, and Cheng Wang. L4dr: Lidar-4dradar fusion for weather-robust 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 39, pp. 3806–3814, 2025.

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 *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 19937–19949. IEEE, 2023.

Xiaoyu Li, Peidong Li, Xian Wu, Long Shi, Dedong Liu, Yitao Wu, Jiajia Fu, Dixiao Cui, Lijun Zhao, and Lining Sun. Rethinking the spatio-temporal alignment of end-to-end 3d perception. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 40, pp. 6513–6520, 2026.

Zhidong Liang, Zehan Zhang, Ming Zhang, Xian Zhao, and Shiliang Pu. Rangeioudet: Range image based real-time 3D object detector optimized by intersection over union. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

Hao Liu, Yanni Ma, Hanyun Wang, Chaobo Zhang, and Yulan Guo. Anchorpoint: Query design for transformer-based 3D object detection and tracking. *IEEE Transactions on Intelligent Transportation Systems*, 24(10):10988–11000, 2023.

Shuai Liu, Mingyue Cui, Boyang Li, Quanmin Liang, Tinghe Hong, Kai Huang, and Yunxiao Shan. Fshnet: Fully sparse hybrid network for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8900–8909. IEEE, 2025.

Yujian Mo, Yan Wu, Junqiao Zhao, Zhenjie Hou, Weiquan Huang, Yinghao Hu, Jijun Wang, and Jun Yan. Sparse Query Dense: Enhancing 3D object detection with pseudo points. In *Proceedings of the ACM International Conference on Multimedia (ACM MM)*, pp. 409–418, 2024.

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Zhiquan Wen, Yaofo Chen, Peilin Zhao, and Mingkui Tan. Towards stable test-time adaptation in dynamic wild world. In *International Conference on Learning Representations (ICLR)*, 2023.

Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. PointRCNN: 3D object proposal generation and detection from point cloud. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–779, 2019.

Ziying Song, Guoxing Zhang, Lin Liu, Lei Yang, Shaoqing Xu, Caiyan Jia, Feiyang Jia, and Li Wang. Robofusion: Towards robust multi-modal 3d object detection via sam. *arXiv preprint arXiv:2401.03907*, 2024.

Chenghao Sun, Pengpeng Sun, and Xiangmo Zhao. Weather-robust lidar perception: Point cloud restoration from adverse weather. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 40, pp. 9162–9170, 2026.

Zhi Tian, Xiangxiang Chu, Xiaoming Wang, Xiaolin Wei, and Chunhua Shen. Fully convolutional one-stage 3D object detection on LiDAR range images. *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 35:34899–34911, 2022.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021. URL https://openreview.net/forum?id=uXl3bZLkr3c.

Haiyang Wang, Chen Shi, Shaoshuai Shi, Meng Lei, Sen Wang, Di He, Bernt Schiele, and Liwei Wang. Dsvt: Dynamic sparse voxel transformer with rotated sets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13520–13529. IEEE, 2023.

Hai Wu, Chenglu Wen, Shaoshuai Shi, Xin Li, and Cheng Wang. Virtual sparse convolution for multimodal 3D object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21653–21662, 2023.

Yan Yan, Yuxing Mao, and Bo Li. SECOND: Sparsely embedded convolutional detection. *Sensors*, 18(10):3337, 2018.

Huitong Yang, Zhuoxiao Chen, Fengyi Zhang, Zi Huang, and Yadan Luo. Codemerge: Codebook-guided model merging for robust test-time adaptation in autonomous driving. *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 38:49096–49123, 2026.

Tianwei Yin, Xingyi Zhou, and Philipp Krahenbuhl. Center-based 3D object detection and tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11784–11793, 2021.

Guowen Zhang, Lue Fan, Chenhang He, Zhen Lei, Zhaoxiang Zhang, and Lei Zhang. Voxel mamba: Group-free state space models for point cloud based 3d object detection. *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 37:81489–81509, 2024.

Yin Zhou and Oncel Tuzel. VoxelNet: End-to-end learning for point cloud based 3D object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4490–4499, 2018.

A APPENDIX

This section provides supplementary analyses and qualitative results for VORL. We present additional ablation studies on different design choices, further evaluations of generalization, and more qualitative visualizations under diverse corruption conditions. These results complement the main experiments and provide additional insights into the behavior of VORL.

A.1 MORE EXPERIMENTAL RESULTS

Stage of using RaSC. Table 6 studies the placement of RaSC within the sparse convolutional backbone. Introducing RaSC at any individual stage consistently improves corruption robustness over the standard backbone, demonstrating the effectiveness of reliability-aware feature modulation under corrupted sensing conditions. Combining the input and second stages improves AP_{cor} to 67.35 while largely preserving clean performance. In contrast, applying RaSC to all three stages provides no additional benefit. These results suggest that selective placement is more effective than densely applying RaSC throughout the backbone.

RaSC Placement			Clean ↑	AP_{cor} ↑
Input stage	first stage	second stage		
			84.51	62.01
✓			84.43	65.70
	✓		84.40	64.39
		✓	84.39	66.25
✓	✓		84.37	65.43
	✓	✓	83.98	66.55
✓		✓	84.38	**67.35**
✓	✓	✓	83.86	66.93

Table 6: Ablation of RaSC placement on KITTI-C with SECOND-IoU.

Effect of observation cues. Table 7 evaluates the contribution of the proposed observation cues to voxel reliability estimation. When using only voxel features, our method achieves a Pearson correlation of 0.689 and an AP_{cor} of 65.85. Adding any individual cue improves both Pearson correlation value and AP_{cor}, with neighborhood consistency providing the strongest standalone performance. Combining all five cues further increases the Pearson correlation to 0.783 and AP_{cor} to 67.06. The leave-one-out results show consistent degradation when any cue is removed, with removing neighborhood consistency causing the largest AP_{cor} drop. These results show that the proposed cues provide complementary observation-level evidence beyond learned voxel features, leading to more accurate reliability estimation and improved corruption robustness.

Support	Range	Intensity	Geometry	Neighbor.	Pearson r ↑	AP_{cor} ↑
					0.689	65.85
✓					0.721	66.28
	✓				0.715	66.17
		✓			0.730	66.33
			✓		0.728	66.37
				✓	0.743	66.44
	✓	✓	✓	✓	0.771	66.92
✓		✓	✓	✓	0.776	66.99
✓	✓		✓	✓	0.769	66.85
✓	✓	✓		✓	0.774	66.89
✓	✓	✓	✓		0.769	66.82
✓	✓	✓	✓	✓	**0.783**	**67.06**

Table 7: Ablation of observation cues on KITTI-C with SQDNet. The first row uses voxel features only, without observation cues. Pearson r measures the correlation between predicted reliability and pseudo reliability targets. Best results are shown in bold.

Generalization across corruption severities. We further investigate whether a reliability estimator trained at a single corruption severity can generalize to different corruption levels. Specifically,

VORL is trained only at severity level 5 and directly evaluated at severity level 1-4 without any retraining or adaptation. As shown in Table 8, VORL consistently improves SQDNet across all evaluated severity levels, with AP_{cor} gains ranging from 2.54 to 7.01 points. These results suggest that VORL does not merely fit a fixed corruption severity level, but learns voxel-wise reliability patterns that transfer across different severity levels.

Severity	Method	Weather-related				Sensor-related				AP_{cor} ↑
		Snow	Rain	Fog	Sunlight	Crosstalk	Gaussian	Uniform	Impulse	
1	SQDNet	68.02	67.61	68.36	76.95	79.98	80.40	74.87	78.81	74.38
	+ VORL	70.60	69.56	77.19	81.01	80.98	80.56	76.19	79.25	76.92
2	SQDNet	66.92	65.82	67.71	75.18	78.07	78.12	67.74	77.08	72.08
	+ VORL	68.43	66.81	76.39	80.43	81.00	78.78	72.06	78.35	75.28
3	SQDNet	63.73	62.86	65.91	72.70	73.75	69.87	57.97	76.26	67.88
	+ VORL	65.99	64.93	74.70	80.26	80.67	72.27	68.55	77.64	73.13
4	SQDNet	61.65	60.56	63.59	69.76	71.50	52.54	54.58	74.25	63.55
	+ VORL	64.09	63.00	73.55	80.18	81.10	61.65	64.15	76.72	70.56

Table 8: Cross-severity generalization on KITTI-C with SQDNet. The reliability estimator is trained only at severity level 5 and directly evaluated at lower severity levels 1-4 without retraining.

Generalization across weather- and sensor-related corruptions. We evaluate whether the learned voxel-wise reliability transfers across different corruption categories. Specifically, we train the reliability estimator exclusively on sensor-related corruptions and evaluate it on weather-related corruptions, and vice versa, without retraining or adaptation. As shown in Table 9 and Table 10, sensor-to-weather transfer improves the average AP from 61.57 to 63.86, while weather-to-sensor transfer improves it from 59.00 to 60.40. Although the gains vary across different corruption types, the performance improvements in both transfer directions suggesting that VORL captures transferable observation-reliability patterns shared across different corruption categories.

Method	Test: Weather				Avg. ↑
	Snow	Rain	Fog	Sunlight	
SQDNet	59.70	57.80	**61.23**	67.55	61.57
+ VORL	**61.11**	**59.18**	58.60	**76.55**	**63.86**

Table 9: Sensor-to-weather generalization on KITTI-C with SQDNet.

Method	Test: Sensor				Avg. ↑
	Crosstalk	Gaussian	Uniform	Impulse	
SQDNet	72.80	**36.99**	**52.91**	73.28	59.00
+ VORL	**79.18**	35.32	52.89	**74.20**	**60.40**

Table 10: Weather-to-sensor generalization on KITTI-C with SQDNet.

Cross-dataset generalization. We further examine whether voxel-wise observation reliability transfers across datasets without target-domain supervision. To reduce effects from detector architecture and voxelization, we use CenterPoint-Voxel Yin et al. (2021) detectors on KITTI and nuScenes with the same point representation $\{x, y, z, \text{intensity}\}$ and voxel size $[0.1, 0.1, 0.2]$ m. The reliability estimator is trained only on nuScenes-C and directly transferred to KITTI-C without fine-tuning. As shown in Table 11, cross-dataset VORL improves AP_{cor} from 56.65 to 59.18 while largely preserving clean performance. Improvements are observed across all eight corruption types, with particularly clear gains under sunlight, snow, and crosstalk. Although cross-dataset transfer remains below the in-domain result of 65.37 AP_{cor}, the consistent gains show that the learned voxel-wise reliability captures observation-quality patterns that remain transferable across different LiDAR datasets.

Method	Clean	Weather-related				Sensor				AP_{cor} ↑
		Snow	Rain	Fog	Sunlight	Crosstalk	Gaussian	Uniform	Impulse	
CenterPoint-Voxel	**77.49**	54.67	61.98	66.37	39.84	62.10	34.32	60.70	73.26	56.65
+ VORL (In-domain)	77.47	**64.47**	**67.25**	**73.00**	**68.92**	**74.80**	**36.00**	**64.81**	**73.71**	**65.37**
+ VORL (Cross-dataset)	77.44	58.17	63.41	68.14	46.34	65.53	35.34	62.84	73.65	59.18

Table 11: Cross-dataset generalization from nuScenes-C to KITTI-C with matched CenterPoint-Voxel detectors. The cross-dataset estimator is trained only on nuScenes-C and directly evaluated on KITTI-C without target-domain reliability supervision or fine-tuning. In-domain denotes VORL trained on KITTI-C.

A.2 Additional Qualitative Results

We provide additional qualitative comparisons on KITTI-C and nuScenes-C to complement the representative examples shown in the main paper. Figures 4 and 5 present detection results under all eight weather- and sensor-related corruptions at severity level 5. Across diverse corruption conditions, VORL produces cleaner and more stable predictions by suppressing spurious detections while preserving valid objects. These qualitative results are consistent with the quantitative improvements reported in the main paper and further illustrate the robustness of reliability-aware feature propagation across different datasets and corruption types.

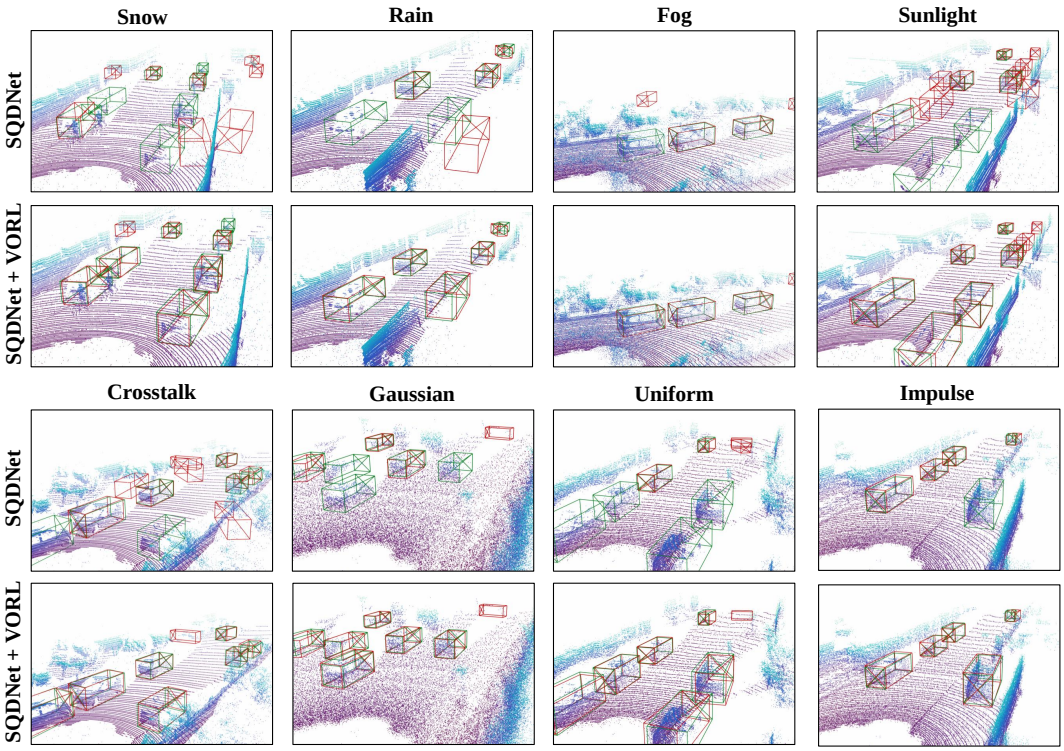


Figure 4: Qualitative comparison of SQDNet and its VORL variant on KITTI-C under all weather- and sensor-related corruptions at severity level 5. The ground-truth boxes and predicted boxes are shown in green and red, respectively.

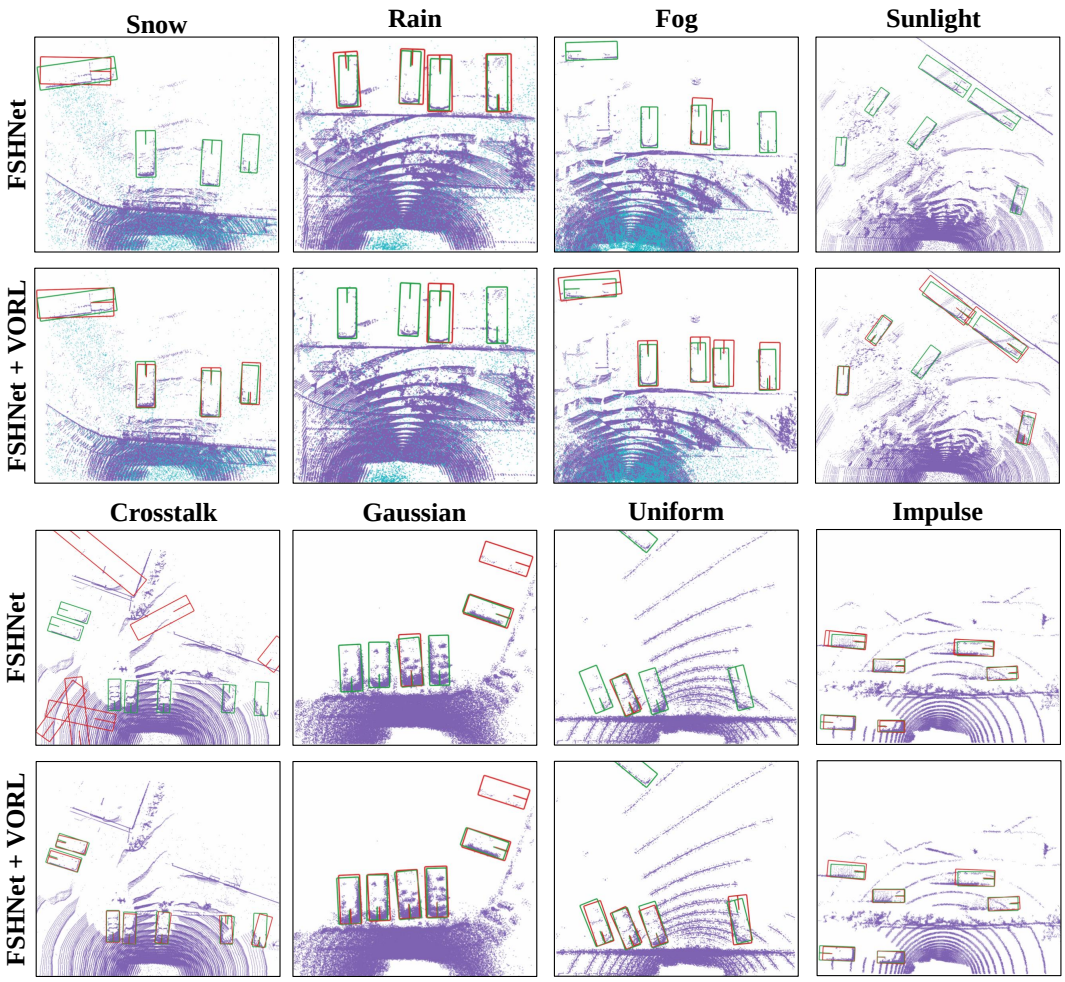


Figure 5: Qualitative comparison of FSHNet and its VORL variant on nuScenes-C under all weather- and sensor-related corruptions at severity level 5. The ground-truth boxes and predicted boxes are shown in green and red, respectively.