LEARNING QUALITY-CONDITIONED FEATURE DISPLACEMENTS FOR LIDAR OUT-OF-DISTRIBUTION DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

LiDAR out-of-distribution (OOD) detection is essential for reliable open-world perception, where unexpected observations beyond the training distribution may appear during real-world deployment. Existing LiDAR OOD methods learn in-distribution feature structures without explicitly modeling the systematic representation changes induced by varying observation quality. Consequently, quality-induced representation deviations of in-distribution observations may be either incorrectly suppressed or absorbed as uncontrolled intra-class variation, obscuring the distinction between sensing-induced variation and genuine OOD deviations. To address this issue, we propose **QuaD**, a framework that learns quality-conditioned feature displacements to explicitly characterize admissible representation variation under different observation qualities. Our method establishes canonical class references from reliable high-quality observations and predicts class-dependent, quality-conditioned feature displacements to explain sensing-induced representation shifts. The predicted displacements are constrained during training such that quality-induced shifts can be accommodated while displacement-corrected representations remain consistent with the canonical class structure. At inference, semantic, contrastive, and displacement evidence are interpreted according to their class- and quality-conditioned in-distribution distributions, enabling more reliable distinction between admissible sensing-induced variation and genuine OOD deviations. Extensive experiments on multiple LiDAR OOD benchmarks demonstrate that QuaD consistently achieves state-of-the-art results, with up to a 108.3% improvement in point-level AP on SemanticKITTI-OoD over the previous best method.

1 INTRODUCTION

Autonomous vehicles and robotic systems depend on robust perception to operate in complex real-world environments. LiDAR semantic segmentation Li et al. (2026a); Xu et al. (2025b) supports three-dimensional scene understanding by assigning semantic labels to individual points, and recent methods have achieved strong performance under the assumption of a predefined label space. However, real-world deployment can expose perception systems to unknown objects outside the training distribution Scheirer et al. (2012); Sodano et al. (2024). These out-of-distribution (OOD) observations may be misclassified as known categories, resulting in unreliable predictions Blum et al. (2019); Yang et al. (2024). Reliable LiDAR OOD detection is therefore essential for extending closed-set segmentation toward open-world perception Nekrasov et al. (2025).

Despite recent advances in OOD detection Grcić et al. (2022); Liu et al. (2023); Nayal et al. (2025), anomaly identification in LiDAR point clouds remains challenging due to the unique characteristics of three-dimensional sensing Kösel et al. (2024). Unlike images that provide dense appearance cues, LiDAR scans consist of sparse and irregular point measurements whose number and spatial distribution vary substantially across objects and sensing conditions, often resulting in incomplete and highly non-uniform object representations Wang et al. (2021); Hu et al. (2022); Yang et al. (2025). These characteristics make it difficult to establish reliable boundaries between known and unknown observations in real-world LiDAR scenes.

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.

Existing LiDAR OOD detection approaches mainly follow two strategies for identifying observations beyond known semantic categories. Some methods Cen et al. (2022); Xu et al. (2025a); Li et al. (2026c) introduce auxiliary OOD supervision by constructing surrogate unknown samples and learning explicit ID–OOD separation boundaries. These approaches generate pseudo-OOD observations through geometric transformations, object insertion, or local perturbations, and optimize anomaly-aware objectives using the synthesized samples Li et al. (2026b). Although auxiliary OOD supervision provides direct guidance for anomaly discrimination, its effectiveness depends on how well the constructed unknown samples approximate the diversity and characteristics of real-world OOD objects. Alternatively, some methods Oh et al. (2026); Mosco et al. (2026) avoid explicit unknown synthesis and instead identify anomalies according to the intrinsic structure of in-distribution (ID) observations. These approaches model known-class representations or semantic feature distributions and detect unknown observations according to their deviations from the learned ID structures. Such ID-based modeling provides a more general solution by avoiding assumptions about the unknown distribution during training.

However, existing ID-based LiDAR OOD methods mainly model the structure of in-distribution representations without explicitly accounting for how observation quality systematically affects them. In LiDAR perception, factors such as sensing range and local point sparsity can induce systematic feature shifts even among observations from the same semantic category Kim et al. (2023); Zhao et al. (2025). Consequently, deviations from learned ID structures may reflect either legitimate sensing-induced variation or genuine OOD behavior, making reliable anomaly discrimination difficult without explicitly modeling quality-dependent representation changes.

To address this issue, we propose QuaD, a quality-conditioned framework that explicitly models admissible feature variation induced by different observation qualities. Instead of treating all deviations from class structures as potential anomalies, we decompose in-distribution representations into a canonical class component and a quality-dependent displacement component. Specifically, reliable high-quality observations are used to establish canonical class references, while a continuous quality-conditioned displacement function models how representations shift with observation quality. This formulation accommodates legitimate quality-induced variation while preventing it from being absorbed as uncontrolled intra-class dispersion. Based on the learned quality-conditioned representation structure, we further develop a quality-conditioned anomaly estimation strategy. Semantic, contrastive, and displacement evidence are interpreted according to their expected in-distribution behavior under the corresponding semantic category and observation quality, enabling anomalous deviations to be distinguished from admissible sensing-induced variation.

The main contributions of this work are summarized as follows:

- We propose **QuaD**, a quality-conditioned framework for LiDAR OOD detection that explicitly models quality-induced representation variation relative to canonical class structures. By learning class-dependent, quality-conditioned feature displacements, our method distinguishes admissible sensing-induced variation from genuine OOD deviations and improves anomaly discrimination.
- We develop a quality-conditioned anomaly estimation strategy that interprets semantic, contrastive, and displacement evidence according to their expected in-distribution behavior under the corresponding semantic category and observation quality. This conditional evidence modeling enables more reliable identification of unknown observations across different observation qualities.
- Extensive experiments on multiple LiDAR OOD benchmarks demonstrate that QuaD achieves state-of-the-art results with consistent gains across diverse datasets and evaluation settings.

2 Related Work

Point Cloud Semantic Segmentation. Point cloud semantic segmentation has achieved substantial progress in recent years with the development of deep learning-based approaches Thomas et al. (2019); Zhu et al. (2021). Early methods mainly focus on learning discriminative point-wise representations from irregular point clouds. Point-based methods Qi et al. (2017a;b); Wang et al. (2019); Wu et al. (2019b) directly process raw points and exploit local geometric relationships to capture

fine-grained structures. Voxel-based methods Choy et al. (2019); Zhu et al. (2021); Zhao et al. (2022) transform point clouds into sparse volumetric representations and leverage efficient convolution operations for large-scale scene understanding. Meanwhile, range-view approaches Wu et al. (2018; 2019a); Xu et al. (2020); Kong et al. (2023) utilize the organized structure of LiDAR scans to achieve efficient semantic prediction. More recently, sparse convolutional networks and transformer-based architectures Cheng et al. (2021); Ando et al. (2023) have further improved segmentation performance by capturing larger contextual information and long-range dependencies. These methods achieve strong performance on large-scale LiDAR benchmarks and support subsequent 3D perception tasks, but their closed-set assumption limits the identification of unknown observations beyond predefined semantic categories.

OOD Detection. OOD detection aims to identify observations that deviate from the training distribution and recognize samples beyond the known semantic space Huang & Li (2021); Yang et al. (2021); Cao & Zhang (2022); Gao et al. (2023). Initial studies mainly focused on image classification, using confidence, entropy, and energy-based scores to quantify test-time uncertainty Hendrycks & Gimpel (2017); Liang et al. (2017); Malinin & Gales (2018); Liu et al. (2020). With the growing demand for reliable perception, OOD detection has extended to dense prediction, leading to pixel-wise anomaly segmentation benchmarks and methods Lis et al. (2019); Di Biase et al. (2021); Chan et al. (2021a); Grcić et al. (2022); Rai et al. (2023). These approaches initially relied on pixel-level classification outputs, while later methods adopted advanced segmentation architectures, including transformers, to improve anomaly localization and semantic modeling Chan et al. (2021b); Jung et al. (2021); Nayal et al. (2023). More recent studies further model the intrinsic structure of in-distribution data through class-dependent feature distributions, prototypes, or learned manifolds Ming et al. (2022); Lu et al. (2024); Zhang et al. (2024); Fang et al. (2024); Ganguly et al. (2025); Liu & Qin (2025), providing a more general basis for open-world recognition without relying on predefined unknown samples.

LiDAR OOD Detection. Existing LiDAR OOD methods mainly follow two directions: auxiliary OOD supervision and in-distribution representation modeling. Auxiliary-OOD approaches construct surrogate unknown samples to learn explicit ID–OOD separation. REAL Cen et al. (2022) and LiON Xu et al. (2025a) generate pseudo-OOD observations from transformed or synthesized point clouds, while REL Li et al. (2026c) and NDP Li et al. (2026b) further improve this paradigm through tailored anomaly objectives and uncertainty modeling. However, their effectiveness depends on how well the constructed unknowns reflect real-world OOD diversity. Alternatively, ID-based methods avoid explicit unknown synthesis and detect anomalies by modeling known-class structures. DOODLE Oh et al. (2026) identifies anomalies through diffusion-based feature reconstruction, whereas LIDO Mosco et al. (2026) jointly exploits semantic and contrastive representations. Although these approaches improve the modeling of in-distribution feature structures, the influence of varying sensing conditions on in-distribution representations remains insufficiently explored.

3 PRELIMINARIES

Problem Statement. A LiDAR scan is represented as a set of N points $\mathcal{X} = \{\boldsymbol{x}_i\}_{i=1}^{N}$ with the known semantic label space $\mathcal{Y}_{\text{in}} = \{1, \dots, C\}$. LiDAR anomaly segmentation predicts the semantic category of each in-distribution point while assigning a point-wise anomaly score to identify unknown observations. For each point, $(\hat{y}_i, S_i) = f_\theta(\mathcal{X})_i$, where $\hat{y}_i \in \mathcal{Y}_{\text{in}}$ and S_i denotes its anomaly score. The ID/OOD decision is defined as

$$G(\boldsymbol{x}_i) = \begin{cases} \text{ID}, & S_i \le \gamma, \\ \text{OOD}, & S_i > \gamma, \end{cases} \tag{1}$$

where f_θ denotes the segmentation model and γ is the decision threshold.

Motivation. Under the above decision rule, sensing-induced representation shifts can increase the deviation of a low-quality in-distribution observation from its class structure. Let $\boldsymbol{z}_i$ denote its feature representation and $\boldsymbol{P}_c$ a reference representation of class c. A larger deviation, i.e., $d(\boldsymbol{z}_i, \boldsymbol{P}_c)\uparrow$, may consequently increase the anomaly response S_i and even lead to $S_i > \gamma$. A straightforward way to accommodate such quality-dependent variation is to introduce multiple quality-specific prototypes $\{\boldsymbol{P}_c^{(1)}, \dots, \boldsymbol{P}_c^{(K)}\}$. However, this requires discretizing the quality space and estimating a

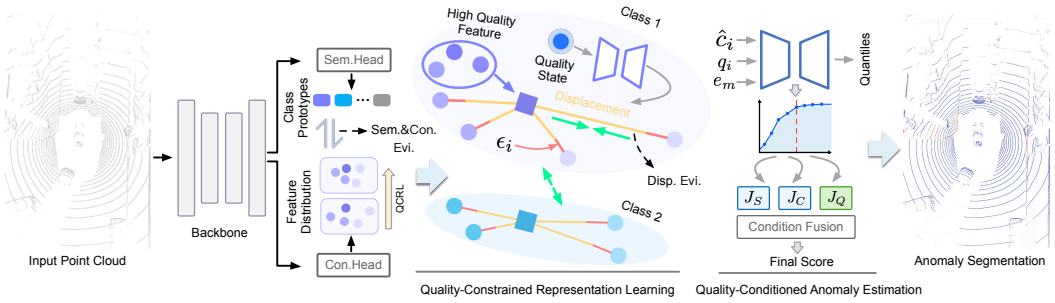

Figure 1: Overview of the proposed QuaD framework. A shared backbone extracts point-wise features for semantic and contrastive prediction. Reliable high-quality observations establish canonical class anchors, while quality-conditioned displacements model sensing-induced feature shifts in the contrastive space. Semantic, contrastive, and displacement evidence are then interpreted by a class- and quality-conditioned quantile model and combined into the final point-wise OOD score.

reliable prototype for each partition, leading to a trade-off between quality resolution and prototype reliability. Instead, we retain a shared canonical class representation and conceptually decompose an in-distribution representation into a canonical component, a quality-dependent displacement, and residual variation:

$$\boldsymbol{z}_i = \boldsymbol{a}_c + \boldsymbol{\Delta}_c(\boldsymbol{q}_i) + \boldsymbol{\epsilon}_i, \tag{2}$$

where $\boldsymbol{a}_c$ denotes the canonical class representation, $\boldsymbol{\Delta}_c(\boldsymbol{q}_i)$ captures the class-dependent feature displacement associated with observation quality $\boldsymbol{q}_i$, and $\boldsymbol{\epsilon}_i$ accounts for residual variation not explained by the quality-dependent displacement. This decomposition provides a continuous characterization of quality-induced feature variation relative to a shared canonical structure, without partitioning the quality space into discrete class references.

4 Methodology

As illustrated in Figure 1, we build upon a semantic–contrastive dual-head framework Mosco et al. (2026), where the semantic branch captures class-discriminative information and prediction uncertainty, while the contrastive branch models in-distribution feature structure for anomaly discrimination. To account for sensing-induced representation variation, reliable high-quality observations define canonical class anchors, from which class-dependent, quality-conditioned displacements model how representations shift under different observation qualities. The learned displacement structure further provides displacement evidence that complements the semantic and contrastive cues. These evidence responses are interpreted according to their semantic category and observation-quality state and subsequently integrated into the final point-wise anomaly score.

4.1 Quality-Constrained Representation Learning

For each LiDAR point $\boldsymbol{x}_i$, sensing quality is characterized by its range and local voxel occupancy. Denoting its spatial coordinate by $\boldsymbol{p}_i \in \mathbb{R}^3$, we define $r_i = \|\boldsymbol{p}_i\|_2$ and let ρ_i denote the number of occupied sparse voxels within the spatial cell of size ϵ_ρ containing $\boldsymbol{p}_i$. After normalization, the quality score and continuous quality state are defined as

$$q_i = \alpha \bar{r}_i + (1-\alpha)(1-\bar{\rho}_i), \quad \boldsymbol{q}_i = [\bar{r}_i,\, 1-\bar{\rho}_i]^\top, \tag{3}$$

where $\boldsymbol{q}_i \in \mathbb{R}^2$ retains the normalized range and sparsity information, and a larger q_i indicates stronger sensing degradation. According to q_i and thresholds τ_{H} and τ_{L}, in-distribution observations are partitioned into high-, medium-, and low-quality sets $\mathcal{H}$, $\mathcal{M}$, and $\mathcal{L}$, respectively, while $\boldsymbol{q}_i$ is retained to characterize continuous quality-dependent representation changes.

The contrastive representation of $\boldsymbol{x}_i$ is denoted by $\boldsymbol{z}_i^{\mathrm{con}} \in \mathbb{R}^d$, with $\hat{\boldsymbol{z}}_i^{\mathrm{con}} = \boldsymbol{z}_i^{\mathrm{con}} / \|\boldsymbol{z}_i^{\mathrm{con}}\|_2$. For each semantic class c, correctly classified observations in $\mathcal{H}$ are used to maintain a canonical class anchor $\boldsymbol{a}_c \in \mathbb{R}^d$ through exponential moving averaging:

$$\boldsymbol{a}_c^{(t)} = \mathrm{Norm}\Big(\mu \boldsymbol{a}_c^{(t-1)} + (1-\mu)\bar{\boldsymbol{z}}_{c,\mathcal{H}}^{\mathrm{con},(t)}\Big), \tag{4}$$

where $\mathrm{Norm}(\boldsymbol{v}) = \boldsymbol{v}/\|\boldsymbol{v}\|_2$, $\tilde{\boldsymbol{z}}_{c,\mathcal{H}}^{\mathrm{con},(t)} \in \mathbb{R}^d$ denotes the mean normalized contrastive representation of correctly classified high-quality observations of class c in the current batch, and μ is the momentum coefficient. The resulting anchor represents the canonical class structure under favorable sensing conditions.

To characterize sensing-induced representation changes, we model a class-dependent, quality-conditioned displacement $\boldsymbol{\Delta}_c(\boldsymbol{q}_i) \in \mathbb{R}^d$ as

$$\boldsymbol{\Delta}_c(\boldsymbol{q}_i) = s\,\boldsymbol{B}_c^\top \boldsymbol{\phi}(\boldsymbol{q}_i), \quad \boldsymbol{\phi}(\boldsymbol{q}_i) = \tanh(\boldsymbol{W}_2\,\mathrm{SiLU}(\boldsymbol{W}_1\boldsymbol{q}_i))\,, \tag{5}$$

where $\boldsymbol{\phi}(\boldsymbol{q}_i) \in \mathbb{R}^{d_q}$ denotes the bounded quality coefficients, $\boldsymbol{B}_c \in \mathbb{R}^{d_q \times d}$ contains class-specific low-dimensional displacement directions, $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ parameterize the quality mapping, and s controls the displacement scale. Combining the displacement with the corresponding anchor yields the quality-conditioned reference $\boldsymbol{p}_c(\boldsymbol{q}_i) \in \mathbb{R}^d$:

$$\boldsymbol{p}_c(\boldsymbol{q}_i) = \mathrm{Norm}(\mathrm{sg}(\boldsymbol{a}_c) + \boldsymbol{\Delta}_c(\boldsymbol{q}_i))\,, \tag{6}$$

where $\mathrm{sg}(\cdot)$ denotes the stop-gradient operation. Unlike assigning independent prototypes to discrete quality levels, $\boldsymbol{p}_c(\boldsymbol{q}_i)$ is continuously derived from the same canonical class anchor according to the current observation quality.

Quality-degraded in-distribution observations are optimized from two complementary perspectives. Their representations are encouraged to follow the corresponding quality-conditioned reference, while removing the estimated displacement is required to recover the canonical class structure:

$$\mathcal{L}_{\mathrm{comp}} = \frac{1}{|\mathcal{D}|}\sum_{i\in\mathcal{D}} [d_{\cos}(\hat{\boldsymbol{z}}_i^{\mathrm{con}}, \boldsymbol{p}_{y_i}(\boldsymbol{q}_i)) + \lambda_{\mathrm{can}} d_{\cos}(\mathrm{Norm}[\hat{\boldsymbol{z}}_i^{\mathrm{con}} - \boldsymbol{\Delta}_{y_i}(\boldsymbol{q}_i)]\,, \mathrm{sg}(\boldsymbol{a}_{y_i}))]\,, \tag{7}$$

where $d_{\cos}(\boldsymbol{u},\boldsymbol{v}) = 1 - \frac{\boldsymbol{u}^\top\boldsymbol{v}}{\|\boldsymbol{u}\|_2\|\boldsymbol{v}\|_2}$ denotes the cosine distance, and $\mathcal{D}$ denotes a subset of degraded in-distribution observations selected from $\mathcal{M}\cup\mathcal{L}$ according to their deviation from the corresponding quality-conditioned reference. The first term aligns degraded representations with their quality-conditioned references, while the second enforces recovery toward the canonical class structure after removing the predicted displacement.

Additional regularization prevents arbitrary representation deviations from being absorbed by the displacement model. Since high-quality observations define the canonical class structure, their predicted displacements are encouraged to remain close to zero, while a global magnitude regularizer limits excessive shifts over the in-distribution space. The resulting quality-constrained representation learning objective is formulated as

$$\mathcal{L}_{\mathrm{QCRL}} = \eta_t \left[\mathcal{L}_{\mathrm{comp}} + \lambda_0 \mathbb{E}_{i\in\mathcal{H}}\left[\|\boldsymbol{\Delta}_{y_i}(\boldsymbol{q}_i)\|_2^2\right] + \lambda_{\mathrm{mag}}\mathbb{E}_{i\in\mathcal{I}}\left[\|\boldsymbol{\Delta}_{y_i}(\boldsymbol{q}_i)\|_2^2\right]\right], \tag{8}$$

where $\mathcal{I}$ denotes the set of valid in-distribution observations, λ_0 and λ_{mag} control the two displacement regularizers, and η_t progressively activates the quality constraint during training. Together with the semantic and contrastive branch objectives of the adopted dual-head framework, the representation network is optimized as

$$\mathcal{L}_{\mathrm{rep}} = \mathcal{L}_{\mathrm{shead}} + \mathcal{L}_{\mathrm{chead}} + \lambda_{\mathrm{QCRL}}\mathcal{L}_{\mathrm{QCRL}}, \tag{9}$$

where λ_{QCRL} controls the contribution of the proposed quality-constrained representation learning objective.

4.2 Quality-Conditioned Anomaly Estimation

The learned class-dependent displacement $\boldsymbol{\Delta}_c(\boldsymbol{q})$ models the systematic representation shift induced by sensing quality for in-distribution observations. The resulting quality-conditioned structure provides an additional cue for distinguishing admissible feature variation from atypical deviations. Semantic and contrastive evidence have already shown strong effectiveness for OOD detection in previous work Mosco et al. (2026). Accordingly, semantic and contrastive cues remain the primary evidence for anomaly estimation, while the quality-conditioned displacement evidence is used to further confirm the resulting anomaly response. The semantic, contrastive, and quality-conditioned displacement evidence are defined as

$$\begin{cases} e_{\mathrm{S}}(\boldsymbol{x}_i) = \big(1 - \max\limits_c \dfrac{\boldsymbol{f}_i^\top \boldsymbol{P}_{\mathrm{S},c}}{\|\boldsymbol{f}_i\|_2\|\boldsymbol{P}_{\mathrm{S},c}\|_2}\big)\big(-\dfrac{1}{\log C}\sum\limits_{c=1}^{C} p_{i,c}\log p_{i,c}\big), \\ e_{\mathrm{C}}(\boldsymbol{x}_i) = \big[1 - \dfrac{\|\boldsymbol{z}_i^{\mathrm{con}}\|_2^2}{r_o}\big]_+, \quad e_{\mathrm{Q}}(\boldsymbol{x}_i) = \dfrac{1}{2}\left(1 - \dfrac{(\boldsymbol{z}_i^{\mathrm{con}})^\top \boldsymbol{p}_{\hat{c}_i}(\boldsymbol{q}_i)}{\|\boldsymbol{z}_i^{\mathrm{con}}\|_2\|\boldsymbol{p}_{\hat{c}_i}(\boldsymbol{q}_i)\|_2}\right). \end{cases} \tag{10}$$

Here, $\boldsymbol{f}_i$, $\boldsymbol{P}_{\mathrm{S},c}$, $p_{i,c}$, and r_o denote the semantic representation, class-c prototype, prediction probability, and objectosphere radius, respectively. We use $\hat{c}_i = \arg\max_c \frac{\boldsymbol{f}_i^\top \boldsymbol{P}_{\mathrm{S},c}}{\|\boldsymbol{f}_i\|_2 \|\boldsymbol{P}_{\mathrm{S},c}\|_2}$ as the semantic category for conditional modeling. The semantic evidence is further normalized within each scan before conditional distribution modeling.

Rather than interpreting these evidence values by their absolute magnitudes, we model their expected distributions under clean in-distribution observations. For each evidence source $m \in \{\mathrm{S}, \mathrm{C}, \mathrm{Q}\}$, we define $F_m(e \mid c, \boldsymbol{q}) = \Pr(E_m \le e \mid c, \boldsymbol{q})$, conditioned on the semantic category and observation quality. A shared quantile model, additionally conditioned on the evidence type, predicts a set of ordered quantiles $\{Q_m(\xi_k \mid c, \boldsymbol{q})\}_{k=1}^{K}$. The model is fitted using only clean in-distribution observations with a tail-weighted pinball objective:

$$\mathcal{L}_{\mathrm{QCAE}} = \sum_{m \in \{\mathrm{S},\mathrm{C},\mathrm{Q}\}} \beta_m \mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K} w_k \, \ell_{\xi_k}\left(e_m - Q_m(\xi_k \mid c, \boldsymbol{q})\right)\right], \tag{11}$$

where $\ell_\xi(u) = \max\{\xi u, (\xi - 1)u\}$ denotes the pinball loss and $w_k \propto 1 + \kappa \xi_k^2$ assigns greater emphasis to upper-tail quantiles. The coefficients β_m balance the different evidence sources, with displacement evidence included only when the corresponding quality-conditioned reference is available. Interpolation between the predicted quantiles yields a continuous piecewise-linear approximation of the conditional cumulative distribution function.

An observed evidence value is then evaluated according to its upper-tail probability under the corresponding class-quality condition:

$$J_m(\boldsymbol{x}_i) = -\log\left[1 - F_m\left(e_m(\boldsymbol{x}_i) \mid \hat{c}_i, \boldsymbol{q}_i\right) + \varepsilon\right]. \tag{12}$$

Thus, J_m reflects the conditional rarity of the observed evidence rather than its absolute magnitude. The semantic and contrastive terms are combined to form the primary anomaly response, $J_{\mathrm{D}}(\boldsymbol{x}_i) = \frac{1}{2}[J_{\mathrm{S}}(\boldsymbol{x}_i) + J_{\mathrm{C}}(\boldsymbol{x}_i)]$. The displacement evidence is used only for confirmation, with $G_{\mathrm{Q}}(\boldsymbol{x}_i) = 1 - \exp[-J_{\mathrm{Q}}(\boldsymbol{x}_i)]$. To prevent displacement evidence from independently determining the anomaly prediction, its contribution is activated according to the primary response through $g(\boldsymbol{x}_i) = \sigma((1 - \exp[-J_{\mathrm{D}}(\boldsymbol{x}_i)] - \tau)/T)$, where τ and T denote the activation threshold and temperature. The final point-wise anomaly score is

$$S_{\mathrm{QCAE}}(\boldsymbol{x}_i) = 1 - \exp\left\{-J_{\mathrm{D}}(\boldsymbol{x}_i)\left[1 + \lambda_{\mathrm{Q}} G_{\mathrm{Q}}(\boldsymbol{x}_i) g(\boldsymbol{x}_i)\right]\right\}, \tag{13}$$

where λ_{Q} controls the contribution of displacement-based confirmation.

5 Experiments

Datasets. We evaluate our method on four LiDAR anomaly segmentation datasets: nuScenes-OoD, SemanticKITTI-OoD, SemanticPOSS-OoD Mosco et al. (2026), and STU Nekrasov et al. (2025). The first three are mixed real-synthetic datasets based on 32-, 64-, and 40-beam LiDAR data, respectively, and provide both single-object and multi-object OOD settings. STU is a large-scale real-world dataset collected with a 128-beam LiDAR sensor, containing natural and deliberately placed anomalies for both point-level and object-level evaluation.

Implementation Details. We adopt MinkowskiNet Choy et al. (2019) as the backbone and follow the representation-learning protocol of LIDO Mosco et al. (2026). The representation network is trained for 64 epochs with a batch size of 4 on a single NVIDIA RTX 4090 GPU, using an initial learning rate of 0.24 and a voxel size of 0.05 m. The canonical anchors are updated with an EMA momentum of $\mu = 0.99$, while the quality-conditioned displacement uses $d_q = 4$ coefficients with a scale of $s = 0.25$ and $\lambda_{\mathrm{QCRL}} = 0.1$. The conditional quantile model is optimized separately on clean in-distribution evidence using AdamW to learn the class- and quality-conditioned evidence distributions. We use $K = 10$ quantiles and set the displacement-confirmation strength to $\lambda_{\mathrm{Q}} = 0.2$.

Evaluation Metrics. We evaluate point-level OOD detection using Average Precision (AP), Area Under the Receiver Operating Characteristic Curve (AUROC), and False Positive Rate at 95% True Positive Rate (FPR@95), following existing LiDAR anomaly segmentation studies Li et al. (2026b). For STU, we further report object-level metrics, including Recall Quality (RecallQ), Segmentation Quality (SQ), Recognition Quality (RQ), Unknown Quality (UQ), and Panoptic Quality (PQ), following the official evaluation protocol Nekrasov et al. (2025).

Table 1: Anomaly segmentation results on nuScenes-OoD, SemanticKITTI-OoD, and SemanticPOSS-OoD. The best results are highlighted in **bold**.

Dataset	Method	Single			Multi		
		AUROC ↑	FPR@95 ↓	AP ↑	AUROC ↑	FPR@95 ↓	AP ↑
nuScenes-OoD	RbA	32.65	100.00	3.56	71.60	98.68	7.95
	Max Logit	66.17	98.60	3.54	43.54	100.00	8.98
	Deep Ensemble	91.79	34.59	18.34	88.62	43.63	23.87
	LIDO	89.33	39.70	6.79	87.25	44.51	10.53
	QuaD (Ours)	**92.22**	**27.79**	**26.66**	**91.37**	**27.98**	**28.12**
SemanticKITTI-OoD	RbA	59.68	100.00	4.56	71.39	100.00	13.63
	Max Logit	76.49	99.63	4.84	84.05	99.32	13.84
	Deep Ensemble	92.87	27.69	6.20	92.19	28.64	12.04
	LIDO	93.36	31.19	10.60	89.89	39.04	9.42
	QuaD (Ours)	**95.80**	**21.87**	**22.08**	**93.12**	**28.27**	**20.59**
SemanticPOSS-OoD	RbA	49.09	100.00	0.13	50.42	100.00	0.26
	Max Logit	61.05	91.85	0.19	63.77	88.74	0.40
	Deep Ensemble	85.86	58.12	0.82	85.83	54.20	1.21
	LIDO	91.51	45.10	3.97	90.84	44.74	5.92
	QuaD (Ours)	**92.82**	**34.75**	**6.81**	**91.66**	**32.88**	**8.73**

Table 2: Anomaly segmentation performance on the STU validation set.

Method	Point-Level OOD			Object-Level OOD				
	AUROC ↑	FPR@95 ↓	AP ↑	RecallQ ↑	SQ ↑	RQ ↑	UQ ↑	PQ ↑
RbA	73.00	100.00	1.64	21.84	78.58	2.75	17.16	2.16
MaxLogit	87.27	68.76	2.02	26.64	79.26	2.06	21.12	1.63
Deep Ensemble	90.93	37.34	6.94	17.70	79.96	9.10	14.15	7.27
MC Dropout	65.76	79.82	0.17	3.54	74.36	3.48	2.63	2.59
Void Classifier	89.77	79.50	2.62	17.35	**81.27**	8.98	14.10	7.30
LIDO	95.05	34.86	27.53	31.92	74.82	28.38	23.88	21.23
QuaD (Ours)	**97.05**	**19.26**	**40.25**	**37.12**	75.74	**39.82**	**28.11**	**30.16**

5.1 Performance Comparisons

Quantitative Results. We compare QuaD with representative LiDAR anomaly segmentation baselines, including RbA Nayal et al. (2023), Max Logit Hendrycks & Gimpel (2017), MC Dropout Srivastava et al. (2014), Void Classifier Blum et al. (2021), Deep Ensemble Lakshminarayanan et al. (2017), and LIDO Mosco et al. (2026). These competing baselines are built upon Mask4Former3D Yilmaz et al. (2024), whereas LIDO and our method adopt MinkowskiNet Choy et al. (2019) as the backbone.

Table 1 reports the results on the three mixed real-synthetic OOD benchmarks. Our method consistently achieves the best performance across different LiDAR datasets under both single-object and multi-object OOD detection settings, with particularly clear improvements in AP and FPR@95. Compared with LIDO, our method substantially improves OOD discrimination by explicitly modeling quality-dependent representation variation. On SemanticKITTI-OoD, the AP for single-object OOD detection increases by 108.3%, while the consistent gains on nuScenes-OoD and SemanticPOSS-OoD further demonstrate the effectiveness of our method across different

Table 3: Closed-set semantic segmentation performance (mIoU, %). S and M denote the *Single* and *Multi* splits, respectively.

Method	nuScenes-OoD		KITTI-OoD		POSS-OoD	
	S	M	S	M	S	M
Standard	**72.75**	**72.62**	**64.99**	**64.84**	**57.07**	**57.05**
LIDO	60.61	60.44	61.34	61.19	55.63	55.58
QuaD (Ours)	60.72	60.53	61.50	61.44	55.21	55.15

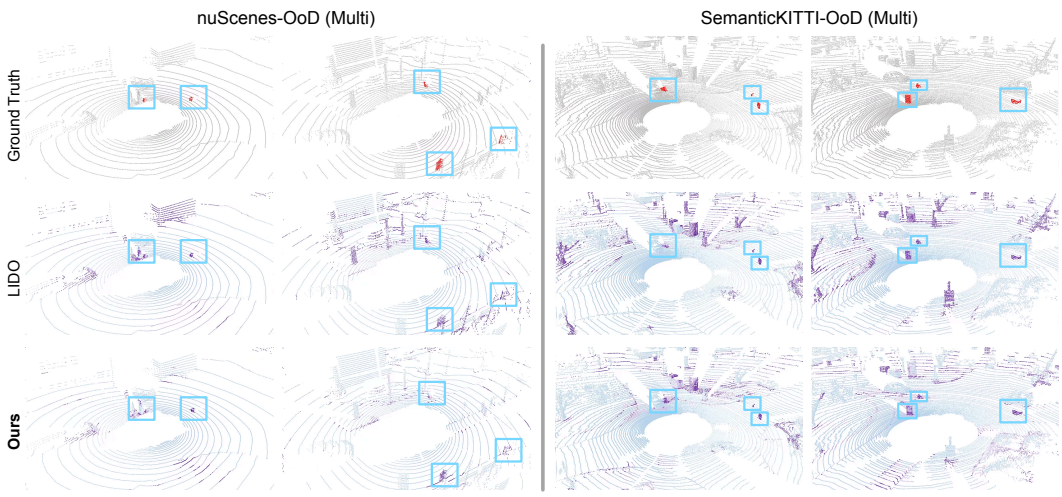

Figure 2: Qualitative multi-object anomaly segmentation results on nuScenes-OoD and SemanticKITTI-OoD. Ground-truth anomalous objects are highlighted by cyan bounding boxes, with anomalous points shown in red. Predicted anomaly scores below a visualization threshold are displayed in light blue, while higher scores are represented by a purple gradient with increasing intensity.

LiDAR sensing resolutions and point-density characteristics. A similar trend is observed on the real-world STU benchmark in Table 2, where our method further improves point-level OOD detection over LIDO while substantially reducing false positives. Since STU contains real anomalous objects rather than synthetically inserted ones, these results show that our method remains effective on real-world anomalies. Meanwhile, Table 3 shows that our method maintains closed-set semantic segmentation performance comparable to LIDO, with differences within 0.5 mIoU across all datasets and splits. This indicates that the improved OOD discrimination is achieved while largely preserving the original closed-set segmentation capability.

Qualitative Results. Figure 2 visualizes the anomaly prediction maps of LIDO and our method on nuScenes-OoD and SemanticKITTI-OoD under the multi-object setting. LIDO tends to produce false-positive responses on nearby above-ground objects and structures, such as vehicles and poles, as well as on distant ground surfaces and walls. In contrast, our method effectively suppresses these spurious responses while maintaining clear predictions on anomalous objects. These results demonstrate improved discrimination between admissible sensing-induced variation and genuine OOD deviations, leading to fewer false-positive responses in both nearby structural regions and distant scene areas.

5.2 Ablation Study

All ablation experiments are conducted on the single-object detection setting of the nuScenes-OoD test set.

Table 4: Contributions of the main components of QuaD.

QCRL	QCAE	AUROC ↑	FPR@95 ↓	AP ↑
		89.33	39.70	6.79
✓		90.79	34.37	20.94
✓	✓	**92.22**	**27.79**	**26.66**

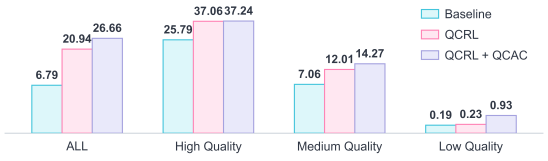

Figure 3: AP across observation-quality levels.

Component Ablation. We first evaluate the contributions of Quality-Constrained Representation Learning (QCRL) and Quality-Conditioned Anomaly Estimation (QCAE). As shown in Table 4, QCRL substantially improves detection performance while retaining the original anomaly

scoring strategy, indicating that explicitly accommodating quality-induced feature shifts provides a more reliable representation structure for OOD discrimination. Incorporating QCAE brings further improvements by interpreting anomaly evidence according to its class- and quality-dependent in-distribution distributions. Figure 3 provides further insight into these improvements across different observation qualities. QCRL primarily benefits high- and medium-quality observations, while QCAE provides additional gains for medium- and low-quality observations. These findings highlight the complementary roles of structured representation learning and conditional evidence interpretation, particularly when sensing degradation complicates anomaly discrimination.

Table 5: Comparison of representation strategies.

Method	AUROC ↑	FPR@95 ↓	AP ↑
Single	89.33	39.70	6.79
Multiple	89.67	35.88	7.83
Distance-based	89.82	33.27	8.01
QuaD	**92.22**	**27.79**	**26.66**

Table 6: Ablation of QCRL formulation.

Method	AUROC ↑	FPR@95 ↓	AP ↑
Canonical	89.84	35.88	7.82
+ Displacement	90.09	34.80	11.30
+ Recovery	90.38	34.60	16.87
Full	**90.79**	**34.37**	**20.94**

Representation Modeling Analysis. We next investigate whether prototype-based alternatives can effectively accommodate quality-dependent representation variation. As reported in Table 5, introducing multiple prototypes for each class yields only modest improvements, while partitioning prototypes according to sensing distance provides limited additional gains. Both approaches represent intra-class variation through a finite set of discrete references. In contrast, our method learns continuous quality-conditioned displacements from the canonical class structure, achieving substantially stronger OOD discrimination.

Ablation of Quality-Constrained Representation Learning. Table 6 examines the progressive formulation of QCRL, starting with canonical class anchors established from reliable high-quality observations. The *Canonical* variant directly uses these anchors as common references for degraded observations, implicitly treating the class structure as invariant to sensing quality. Allowing class-dependent, quality-conditioned displacements substantially improves AP, suggesting that degraded representations benefit from quality-dependent references rather than being constrained to fixed class centers. Requiring displacement-corrected features to recover their canonical structure further strengthens detection, while additional displacement regularization produces the best overall results by limiting unnecessary and excessive representation shifts.

Complementary Anomaly Evidence. We finally examine whether the learned displacement structure provides complementary information for anomaly estimation. As presented in Table 7, incorporating displacement evidence to confirm the primary anomaly response derived from semantic and contrastive evidence improves all three metrics, particularly AP and FPR@95. This improvement suggests that deviations from the expected quality-conditioned representation provide an effective additional cue for distinguishing anomalous observations from admissible sensing-induced variation.

Table 7: Contribution of displacement evidence to anomaly estimation.

Method	AUROC ↑	FPR@95 ↓	AP ↑
Semantic + Contrastive	91.75	29.83	24.62
+ Displacement	**92.22**	**27.79**	**26.66**

6 Conclusion

In this paper, we propose QuaD, a quality-conditioned framework for LiDAR OOD detection that mitigates the confusion between sensing-induced in-distribution variations and genuine OOD deviations. QuaD models normal feature variation through canonical class structures and class-dependent quality-conditioned displacements, enabling sensing-induced shifts to be distinguished from anomalous deviations. It further performs quality-conditioned anomaly estimation by interpreting semantic, contrastive, and displacement evidence according to semantic category and observation quality. Experimental results across multiple LiDAR OOD benchmarks validate the effectiveness of our method in more reliably distinguishing in-distribution variations from genuine OOD deviations.

AI USE STATEMENT

Generative AI tools were used solely to aid and polish the writing of this manuscript, including improving language clarity, grammar, and presentation. All scientific ideas, methodological designs, experimental analyses, and core contributions were developed by the authors. The framework diagrams were manually created by the authors, and all experimental visualizations were generated from actual experimental results using author-written Python code; no figures or experimental results were generated by AI. The authors reviewed and take full responsibility for all content presented in this paper.

REFERENCES

Angelika Ando, Spyros Gidaris, Andrei Bursuc, Gilles Puy, Alexandre Boulch, and Renaud Marlet. Rangevit: Towards vision transformers for 3d semantic segmentation in autonomous driving. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5240–5250. IEEE, 2023.

Hermann Blum, Paul-Edouard Sarlin, Juan Nieto, Roland Siegwart, and Cesar Cadena. Fishyscapes: A benchmark for safe semantic segmentation in autonomous driving. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 2403–2412. IEEE, 2019.

Hermann Blum, Paul-Edouard Sarlin, Juan Nieto, Roland Siegwart, and Cesar Cadena. The fishyscapes benchmark: Measuring blind spots in semantic segmentation. *International Journal of Computer Vision*, 129(11):3119–3135, 2021.

Senqi Cao and Zhongfei Zhang. Deep hybrid models for out-of-distribution detection. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4723–4733. IEEE, 2022.

Jun Cen, Peng Yun, Shiwei Zhang, Junhao Cai, Di Luan, Mingqian Tang, Ming Liu, and Michael Yu Wang. Open-world semantic segmentation for lidar point clouds. In *European Conference on Computer Vision*, pp. 318–334. Springer, 2022.

Robin Chan, Krzysztof Lis, Svenja Uhlemeyer, Hermann Blum, Sina Honari, Roland Siegwart, Pascal Fua, Mathieu Salzmann, and Matthias Rottmann. Segmentmeifyoucan: A benchmark for anomaly segmentation. *arXiv preprint arXiv:2104.14812*, 2021a.

Robin Chan, Matthias Rottmann, and Hanno Gottschalk. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5108–5117. IEEE, 2021b.

Ran Cheng, Ryan Razani, Ehsan Taghavi, Enxu Li, and Bingbing Liu. 2-s3net: Attentive feature fusion with adaptive feature selection for sparse semantic segmentation network. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12542–12551. IEEE, 2021.

Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 3070–3079. IEEE, 2019.

Giancarlo Di Biase, Hermann Blum, Roland Siegwart, and Cesar Cadena. Pixel-wise anomaly detection in complex driving scenes. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16913–16922. IEEE, 2021.

Kun Fang, Qinghua Tao, Kexin Lv, Mingzhen He, Xiaolin Huang, and Jie Yang. Kernel pca for out-of-distribution detection. *Advances in Neural Information Processing Systems*, 37:134317–134344, 2024.

Debargha Ganguly, Warren Morningstar, Andrew Yu, and Vipin Chaudhary. Forte: Finding outliers with representation typicality estimation. In *International Conference on Learning Representations*, volume 2025, pp. 23827–23864, 2025.

Ruiyuan Gao, Chenchen Zhao, Lanqing Hong, and Qiang Xu. Diffguard: Semantic mismatch-guided out-of-distribution detection using pre-trained diffusion models. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1579–1589. IEEE, 2023.

Matej Grcić, Petra Bevandić, and Siniša Šegvić. Densehybrid: Hybrid anomaly detection for dense open-set recognition. In *European Conference on Computer Vision*, pp. 500–517. Springer, 2022.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *International Conference on Learning Representations*, 2017.

Jordan SK Hu, Tianshu Kuai, and Steven L Waslander. Point density-aware voxels for lidar 3d object detection. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8459–8468. IEEE, 2022.

Rui Huang and Yixuan Li. Mos: Towards scaling out-of-distribution detection for large semantic space. In *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 8706–8715. IEEE, 2021.

Sanghun Jung, Jungsoo Lee, Daehoon Gwak, Sungha Choi, and Jaegul Choo. Standardized max logits: A simple yet effective approach for identifying unexpected road obstacles in urban-scene segmentation. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15405–15414. IEEE, 2021.

Hyeonseong Kim, Yoonsu Kang, Changgyoon Oh, and Kuk-Jin Yoon. Single domain generalization for lidar semantic segmentation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17587–17598. IEEE, 2023.

Lingdong Kong, Youquan Liu, Runnan Chen, Yuexin Ma, Xinge Zhu, Yikang Li, Yuenan Hou, Yu Qiao, and Ziwei Liu. Rethinking range view representation for lidar segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 228–240, 2023.

Michael Kösel, Marcel Schreiber, Michael Ulrich, Claudius Gläser, and Klaus Dietmayer. Revisiting out-of-distribution detection in lidar-based 3d object detection. In *2024 IEEE Intelligent Vehicles Symposium (IV)*, pp. 2806–2813. IEEE, 2024.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.

Wangkai Li, Zhaoyang Li, Yuwen Pan, Rui Sun, Yujia Chen, and Tianzhu Zhang. Adaptive augmentation-aware latent learning for robust lidar semantic segmentation. *arXiv preprint arXiv:2603.01074*, 2026a.

Zizhao Li, Zhengkang Xiang, Jiayang Ao, Feng Liu, Joseph West, and Kourosh Khoshelham. Neural distribution prior for lidar out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3035–3045, June 2026b.

Zizhao Li, Zhengkang Xiang, Jiayang Ao, Joseph West, and Kourosh Khoshelham. Relative energy learning for lidar out-of-distribution detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 239:334–346, 2026c.

Shiyu Liang, Yixuan Li, and Rayadurgam Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. *arXiv preprint arXiv:1706.02690*, 2017.

Krzysztof Lis, Krishna Nakka, Pascal Fua, and Mathieu Salzmann. Detecting the unexpected via image resynthesis. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 2152–2161, 2019.

Litian Liu and Yao Qin. Detecting out-of-distribution through the lens of neural collapse. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15424–15433. IEEE, 2025.

Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in neural information processing systems*, 33:21464–21475, 2020.

Yuyuan Liu, Choubo Ding, Yu Tian, Guansong Pang, Vasileios Belagiannis, Ian Reid, and Gustavo Carneiro. Residual pattern learning for pixel-wise out-of-distribution detection in semantic segmentation. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1151–1161. IEEE, 2023.

Haodong Lu, Dong Gong, Shuo Wang, Jason Xue, Lina Yao, and Kristen Moore. Learning with mixture of prototypes for out-of-distribution detection. In *International Conference on Learning Representations*, volume 2024, pp. 8613–8635, 2024.

Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. *Advances in neural information processing systems*, 31, 2018.

Yifei Ming, Yiyou Sun, Ousmane Dia, and Yixuan Li. How to exploit hyperspherical embeddings for out-of-distribution detection? *arXiv preprint arXiv:2203.04450*, 2022.

Simone Mosco, Daniel Fusaro, and Alberto Pretto. Learning to identify out-of-distribution objects for 3d lidar anomaly segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17204–17214, June 2026.

Nazir Nayal, Misra Yavuz, Joao F Henriques, and Fatma Güney. Rba: Segmenting unknown regions rejected by all. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 711–722, 2023.

Nazir Nayal, Youssef Shoeb, and Fatma Güney. A likelihood ratio-based approach to segmenting unknown objects. *International Journal of Computer Vision*, 133(10):6860–6872, 2025.

Alexey Nekrasov, Malcolm Burdorf, Stewart Worrall, Bastian Leibe, and Julie Stephany Berrio Perez. Spotting the unexpected (stu): A 3d lidar dataset for anomaly segmentation in autonomous driving. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11875–11885. IEEE, 2025.

Changgyoon Oh, Hyeonseong Kim, Daehyun We, Jongoh Jeong, Yujeong Chae, and Kuk-Jin Yoon. Doodle: Diffusion-based out-of-distribution learning for open-set lidar semantic segmentation. In *2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2273–2283. IEEE, 2026.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017a.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017b.

Shyam Nandan Rai, Fabio Cermelli, Dario Fontanel, Carlo Masone, and Barbara Caputo. Unmasking anomalies in road-scene segmentation. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4014–4023. IEEE, 2023.

Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(7): 1757–1772, 2012.

Matteo Sodano, Federico Magistri, Lucas Nunes, Jens Behley, and Cyrill Stachniss. Open-world semantic segmentation including class similarity. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3184–3194. IEEE, 2024.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6411–6420, 2019.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Chunwei Wang, Chao Ma, Ming Zhu, and Xiaokang Yang. Pointaugmenting: Cross-modal augmentation for 3d object detection. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11789–11798. IEEE, 2021.

Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, 38(5): 1–12, 2019.

Bichen Wu, Alvin Wan, Xiangyu Yue, and Kurt Keutzer. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud. In *2018 IEEE international conference on robotics and automation (ICRA)*, pp. 1887–1893. IEEE, 2018.

Bichen Wu, Xuanyu Zhou, Sicheng Zhao, Xiangyu Yue, and Kurt Keutzer. Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud. In *2019 international conference on robotics and automation (ICRA)*, pp. 4376–4382. IEEE, 2019a.

Wenxuan Wu, Zhongang Qi, and Li Fuxin. Pointconv: Deep convolutional networks on 3d point clouds. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9613–9622. Ieee, 2019b.

Chenfeng Xu, Bichen Wu, Zining Wang, Wei Zhan, Peter Vajda, Kurt Keutzer, and Masayoshi Tomizuka. Squeezesegv3: Spatially-adaptive convolution for efficient point-cloud segmentation. In *European Conference on Computer Vision*, pp. 1–19. Springer, 2020.

Shaocong Xu, Pengfei Li, Qianpu Sun, Xinyu Liu, Yang Li, Shihui Guo, Zhen Wang, Bo Jiang, Rui Wang, Kehua Sheng, et al. Lion: Learning point-wise abstaining penalty for lidar outlier detection using diverse synthetic data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 8951–8959, 2025a.

Xiang Xu, Lingdong Kong, Hui Shuai, and Qingshan Liu. Frnet: Frustum-range networks for scalable lidar segmentation. *IEEE Transactions on Image Processing*, 34:2173–2186, 2025b.

Jingkang Yang, Haoqi Wang, Litong Feng, Xiaopeng Yan, Huabin Zheng, Wayne Zhang, and Ziwei Liu. Semantically coherent out-of-distribution detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8281–8289. IEEE, 2021.

Jingkang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. Generalized out-of-distribution detection: A survey. *International Journal of Computer Vision*, 132(12):5635–5662, 2024.

Pengfei Yang, Feng Wu, Minyang Liu, Ting Zhong, and Fan Zhou. Beyond pillars: Advancing 3d object detection with salient voxel enhancement of lidar-4d radar fusion. *Pattern Recognition*, pp. 112841, 2025.

Kadir Yilmaz, Jonas Schult, Alexey Nekrasov, and Bastian Leibe. Mask4former: Mask transformer for 4d panoptic segmentation. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9418–9425. IEEE, 2024.

Yonggang Zhang, Jie Lu, Bo Peng, Zhen Fang, and Yiu-ming Cheung. Learning to shape in-distribution feature space for out-of-distribution detection. *Advances in Neural Information Processing Systems*, 37:49384–49402, 2024.

Jichun Zhao, Haiyong Jiang, Haoxuan Song, Jun Xiao, and Dong Gong. D 3 ctta: Domain-dependent decorrelation for continual test-time adaption of 3d lidar segmentation. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11864–11874. IEEE, 2025.

Lin Zhao, Siyuan Xu, Liman Liu, Delie Ming, and Wenbing Tao. Svaseg: Sparse voxel-based attention for 3d lidar point cloud semantic segmentation. *Remote Sensing*, 14(18):4471, 2022.

Xinge Zhu, Hui Zhou, Tai Wang, Fangzhou Hong, Yuexin Ma, Wei Li, Hongsheng Li, and Dahua Lin. Cylindrical and asymmetrical 3d convolution networks for lidar segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9939–9948, 2021.

A Appendix

A.1 Additional Implementation Details

Quality and Displacement Modeling. For quality estimation, the point range is clipped at $r_{\max} = 50$ m and normalized to obtain $\bar{r}_i$. Local voxel occupancy is computed within spatial cells of size $\epsilon_\rho = 1$ m, with the occupancy capped at $\rho_{\max} = 64$ before logarithmic normalization. We set the range–sparsity weighting coefficient to $\alpha = 0.5$ and use $\tau_{\mathrm{H}} = 0.15$ and $\tau_{\mathrm{L}} = 0.35$ to define the high-, medium-, and low-quality groups. The quality mapping $\boldsymbol{\phi}(\boldsymbol{q}_i)$ in Eq. (5) is implemented by a lightweight bias-free MLP. Specifically, the two-dimensional quality state $\boldsymbol{q}_i \in \mathbb{R}^2$ is projected to a hidden dimension $d_h = 16$ with a SiLU activation and then mapped to d_q quality coefficients, followed by $\tanh$ to bound their values. For each semantic class c, the displacement basis $\boldsymbol{B}_c \in \mathbb{R}^{d_q \times d}$ is independently learned in the contrastive feature space and initialized from $\mathcal{N}(0, \sigma_B^2)$ with $\sigma_B = 0.01$. For the QCRL objective in Eq. (8), we set $\lambda_{\mathrm{can}} = 0.5$, $\lambda_0 = 0.1$, and $\lambda_{\mathrm{mag}} = 0.01$.

Conditional Quantile Model. The conditional quantile function $Q_m(\xi \mid c, \boldsymbol{q})$ is implemented by a shared lightweight network conditioned on the semantic category, evidence type, and observation quality. The semantic category c and evidence type m are represented by embeddings of dimensions $d_c = 8$ and $d_m = 4$, respectively. The quality state $\boldsymbol{q} \in \mathbb{R}^2$ is encoded by an MLP with dimensions $2 \to 16 \to 16$ and SiLU activations. The three representations are concatenated and processed by two hidden layers of dimension 64, followed by an output layer producing the K conditional quantiles. To preserve their ordering, the predicted increments are constrained to be positive using Softplus, cumulatively accumulated along the quantile dimension, and normalized to $(0, 1)$. The quantile model is optimized separately on clean in-distribution evidence using AdamW for $N_{\mathrm{q}} = 100$ optimization steps with batch size $B_{\mathrm{q}} = 8192$, learning rate $\eta_{\mathrm{q}} = 10^{-3}$, and weight decay $\omega_{\mathrm{q}} = 10^{-5}$. The quantile levels are

$$\{\xi_k\}_{k=1}^{K} = \{0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 0.975, 0.99, 0.995\}.$$

For the objective in Eq. (11), we use $\kappa = 2.0$ and $(\beta_{\mathrm{S}}, \beta_{\mathrm{C}}, \beta_{\mathrm{Q}}) = (0.5, 0.5, 0.5)$. The objectosphere radius in the contrastive evidence is set to $r_o = 5.0$. For the confirmation mechanism in Eq. (13), we use an activation threshold $\tau = 0.4$ and temperature $T = 0.08$.

A.2 Detailed Evaluation Metrics

Point-Level Evaluation Metrics. Point-level OOD detection is evaluated using AUROC, FPR@95, and Average Precision (AP), with OOD points treated as positive samples and in-distribution points as negative samples. AUROC measures the overall separability between ID and OOD points across all possible anomaly-score thresholds. A higher AUROC indicates that OOD points are more consistently assigned larger anomaly scores than ID points. FPR@95 evaluates the false-positive rate at the operating point where 95% of OOD points are correctly detected, and therefore reflects the ability to suppress false alarms under high OOD recall. AP summarizes the precision–recall trade-off over different score thresholds. Since OOD points are typically much less frequent than ID points in LiDAR scenes, AP is particularly informative under the resulting class imbalance. Higher AUROC and AP and lower FPR@95 indicate better OOD detection performance.

Object-Level Evaluation Metrics. For STU Nekrasov et al. (2025), we additionally follow the official object-level evaluation protocol. A predicted anomaly instance is considered a true positive if its Intersection over Union (IoU) with a ground-truth anomaly exceeds 50%. Unmatched predictions and ground-truth instances are counted as false positives and false negatives, respectively.

Panoptic Quality (PQ) jointly evaluates the segmentation accuracy of matched instances and the correctness of instance recognition:

$$\mathrm{PQ} = \underbrace{\frac{\sum_{(p,g)\in \mathrm{TP}} \mathrm{IoU}(p,g)}{|\mathrm{TP}|}}_{\text{Segmentation Quality (SQ)}} \times \underbrace{\frac{|\mathrm{TP}|}{|\mathrm{TP}| + \frac{1}{2}|\mathrm{FP}| + \frac{1}{2}|\mathrm{FN}|}}_{\text{Recognition Quality (RQ)}}. \tag{14}$$

Here, p and g denote matched predicted and ground-truth anomaly instances, respectively. SQ reflects the segmentation quality of correctly matched anomalies, whereas RQ measures instance-level recognition performance.

STU further reports Unknown Quality (UQ), which focuses on the ability to retrieve and segment anomalous objects:

$$\mathrm{UQ} = \underbrace{\frac{\sum_{(p,g)\in \mathrm{TP}} \mathrm{IoU}(p,g)}{|\mathrm{TP}|}}_{\text{Segmentation Quality (SQ)}} \times \underbrace{\frac{|\mathrm{TP}|}{|\mathrm{TP}| + |\mathrm{FN}|}}_{\text{Recall Quality (RecallQ)}} . \tag{15}$$

Unlike PQ, UQ does not penalize false-positive instances and therefore emphasizes anomaly recall together with the segmentation quality of successfully detected objects. Higher SQ, RQ, RecallQ, UQ, and PQ indicate better object-level anomaly segmentation performance.

A.3 Computational Efficiency

We assess the computational overhead introduced by QuaD relative to LIDO in terms of additional model parameters and inference time. The learnable components of QCRL and QCAE introduce only approximately 8.4K parameters in total, accounting for less than 0.04% of the 21.7M parameters of LIDO. Thus, the overall model size remains effectively unchanged at 21.7M parameters. As shown in Figure 4(a), QCRL introduces only marginal inference overhead, increasing the per-scan runtime from 12.05 ms to 12.19 ms. This small increase is expected because QCRL preserves the original backbone and prediction heads and only introduces a low-dimensional quality mapping and lightweight class-specific displacement modeling. With the complete QCAE module, the inference time increases by only 3.04 ms compared with LIDO. The additional cost mainly arises from evaluating the conditional quantile model for point-wise anomaly evidence and performing the subsequent conditional CDF interpolation and displacement-confirmation operations. Although these operations introduce additional point-wise computation, the quantile network itself contains very few parameters and does not modify the main feature-extraction backbone. Overall, QuaD introduces negligible additional model parameters while maintaining a moderate inference overhead.

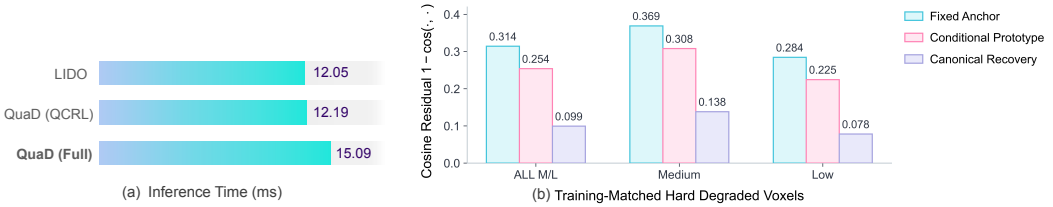

Figure 4: (a) Inference-time comparison of LIDO and QuaD variants. (b) Representation-space analysis of quality-conditioned feature displacement. Cosine residuals are measured against the fixed canonical anchor, the quality-conditioned prototype, and the canonical anchor after displacement correction for training-matched hard degraded observations. Lower values indicate better representation alignment.

A.4 Additional Ablation Studies

We conduct additional ablation studies on the single-object detection setting of the nuScenes-OoD test set to examine the representation behavior and key design choices of QuaD.

Quality-Conditioned Displacement Analysis. We analyze how the learned quality-conditioned displacement affects the representation geometry of degraded observations. Figure 4(b) compares their cosine residuals under different representation references. Across different degradation levels, the quality-conditioned reference consistently aligns more closely with the observed representations than the fixed canonical anchor, showing that the learned displacement captures systematic feature shifts associated with sensing quality. More importantly, removing the predicted displacement brings both medium- and low-quality representations closer to their canonical anchors. This two-sided behavior demonstrates that the displacement models a structured quality-induced shift rather than simply relaxing the class representation, allowing degraded features to retain their quality-dependent variation while remaining recoverable toward the canonical class structure.

Table 8: Ablation of the representation space constrained by QCRL.

Constraint Space	AUROC ↑	FPR@95 ↓	AP ↑
Shared	87.60	41.15	8.69
Semantic	85.74	39.54	5.31
Contrastive	**90.79**	**34.37**	**20.94**

Table 9: Ablation of the quantile-model architecture in QCAE.

Architecture	AUROC ↑	FPR@95 ↓	AP ↑
CNN	91.87	29.10	25.03
Transformer	91.98	28.45	26.12
MLP	**92.22**	**27.79**	**26.66**

Effect of Constraint Space. Table 8 compares the effect of imposing QCRL on different representation spaces. Applying the quality-dependent constraint to the shared feature space yields only limited improvement over the original LIDO baseline and adversely affects AUROC and FPR@95, while constraining the semantic feature space provides no consistent benefit. By contrast, operating in the contrastive space substantially improves all three metrics. This result is consistent with the role of the contrastive branch, which is explicitly optimized to organize in-distribution feature structure for anomaly discrimination and is therefore better suited to characterizing sensing-induced intra-class variation. Meanwhile, avoiding direct constraints on the semantic representation helps preserve the features used for category prediction.

Quantile-Model Architecture. We examine the sensitivity of QCAE to the architecture of the conditional quantile model. As shown in Table 9, the CNN-based model He et al. (2016) and the Transformer-based model Vaswani et al. (2017) both achieve competitive performance, while the lightweight MLP provides the best overall results. This relatively small architectural variation suggests that the benefit of QCAE mainly arises from modeling evidence under class and quality conditions rather than from increasing the complexity of the quantile predictor. Since its inputs are compact representations of semantic category, evidence type, and observation quality, an MLP is sufficient to capture their conditional dependencies without unnecessary architectural overhead.

Table 10: Ablation of conditioning factors in QCAE.

Conditioning	AUROC ↑	FPR@95 ↓	AP ↑
None	90.86	34.48	23.68
Class	89.58	35.28	13.30
Quality	91.03	63.64	23.47
Class + Quality	**92.22**	**27.79**	**26.66**

Table 11: Ablation of displacement-evidence fusion.

Fusion	AUROC ↑	FPR@95 ↓	AP ↑
Direct fusion	91.82	32.82	24.92
Confirmation	92.05	27.91	26.31
Gated confirmation	**92.22**	**27.79**	**26.66**

Effect of Conditioning Factors. A central design of QCAE is to characterize clean in-distribution evidence jointly according to semantic category and observation quality. Table 10 shows that removing either condition weakens this modeling considerably. Semantic-category conditioning alone cannot account for quality-dependent changes within the same category, whereas quality conditioning alone ignores the distinct evidence characteristics associated with different classes and produces poor false-positive control. Jointly incorporating both factors resolves these complementary limitations and achieves the strongest overall performance. The results therefore support the use of the joint conditional distribution $F_m(e \mid c, \boldsymbol{q})$ for interpreting anomaly evidence.

Displacement-Evidence Integration. We investigate how displacement evidence should be incorporated into anomaly estimation. Directly treating it as another fusion term improves anomaly estimation but produces more false positives than the confirmation-based alternatives. Using displacement evidence instead to corroborate the semantic–contrastive response yields a clear improvement, reflecting its intended role as an auxiliary indicator of inconsistency with the expected quality-conditioned representation. The response-dependent gate further restricts this confirmation to observations already supported by the primary anomaly evidence and achieves the best overall result. This comparison motivates the gated-confirmation formulation adopted in QCAE, rather than allowing displacement evidence to independently drive the anomaly decision.

A.5 Additional Visualization

Qualitative Comparison on STU. Figure 5 shows additional results on the STU dataset, covering anomalous objects with varying detection difficulty. The first two scenes contain chairs as anomalies.

LIDO assigns relatively high OOD scores to nearby ground regions and distant objects, whereas the responses of our method are more concentrated on the actual anomalies. The remaining four scenes involve small anomalous objects located on the ground, which are more difficult to identify. Our method successfully detects these objects in all four cases with clear high-score responses, while LIDO misses the anomalies in the fifth and sixth scenes. These examples demonstrate more localized anomaly responses and fewer irrelevant high-score regions, particularly for small and challenging objects.

Qualitative Comparison on SemanticPOSS-OoD. Additional results on SemanticPOSS-OoD are provided in Figure 6. This dataset contains relatively cluttered road environments with diverse surrounding objects and structures, including pedestrians, poles, fences, trunks, and vegetation. For anomalies located directly on the road surface, both methods generally produce clear high OOD responses. In the second and third scenes, where the anomalous objects appear away from the main road region, both methods exhibit missed detections. Nevertheless, our method generates fewer spurious responses on ground protrusions and distant objects, producing cleaner anomaly score maps in these complex environments.

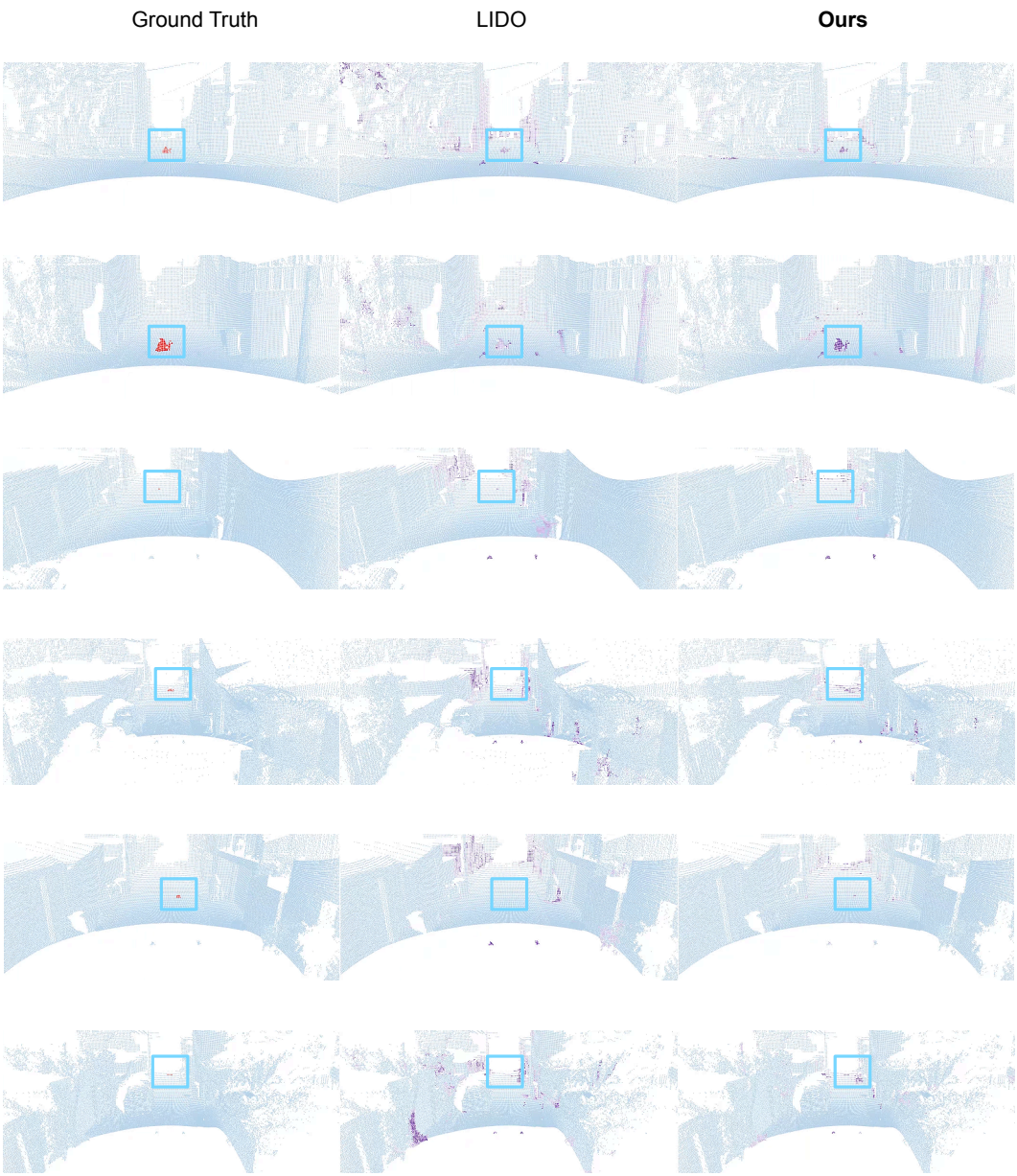

Figure 5: Qualitative comparison on the STU validation set. Ground-truth anomalous objects are highlighted by cyan bounding boxes, with anomalous points shown in red. Predicted anomaly scores below a visualization threshold are displayed in light blue, while higher scores are represented by a purple gradient with increasing intensity. Our method produces more concentrated anomaly responses with fewer false positives and more reliable detection of small anomalous objects.

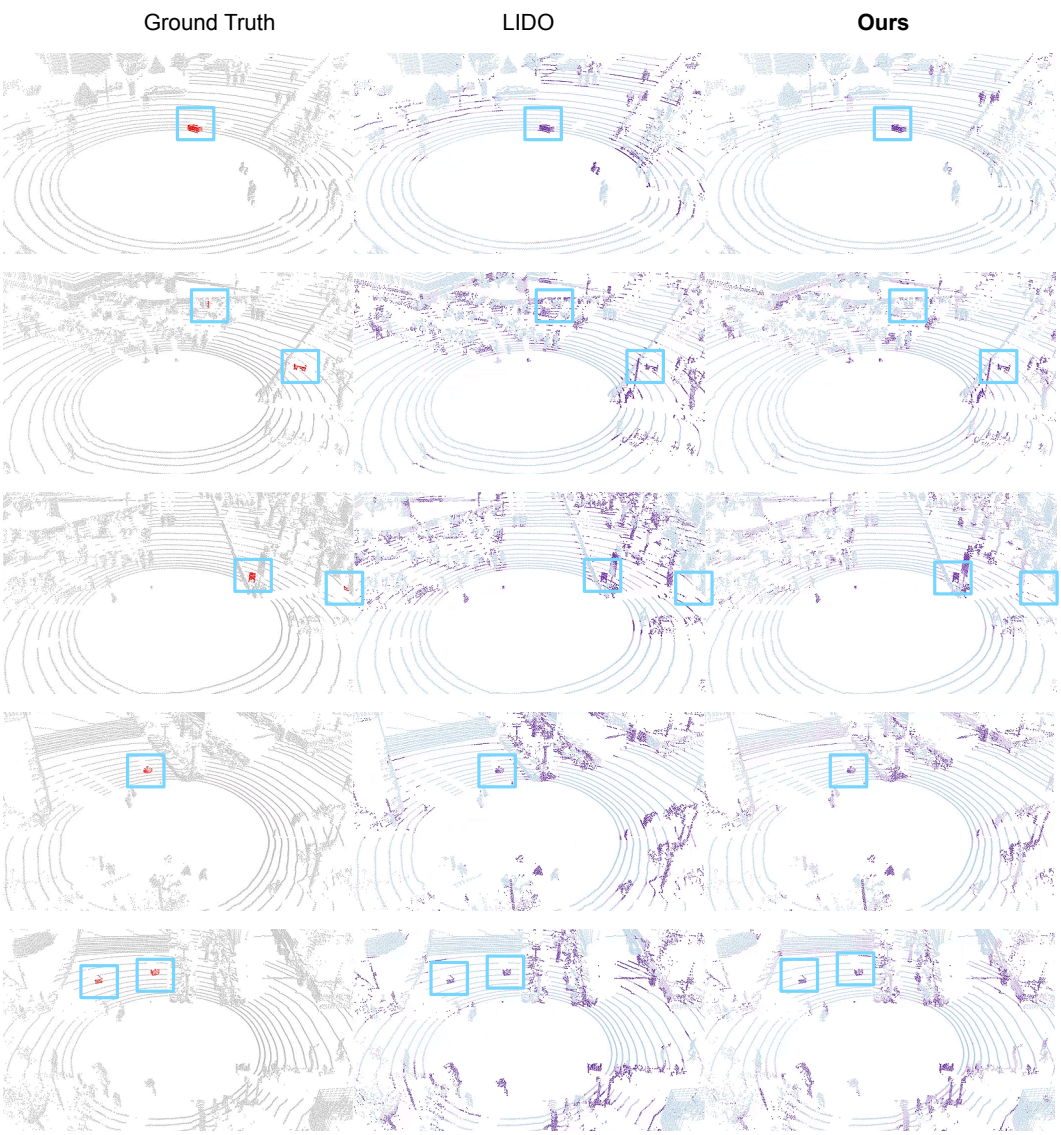

Figure 6: Qualitative multi-object anomaly segmentation results on the SemanticPOSS-OoD dataset. Ground-truth anomalous objects are highlighted by cyan bounding boxes, with anomalous points shown in red. Predicted anomaly scores below a visualization threshold are displayed in light blue, while higher scores are represented by a purple gradient with increasing intensity. Our method produces fewer false-positive responses on ground protrusions and distant objects, resulting in cleaner anomaly score maps in complex scenes.