SOLVING LIDAR INVERSE PROBLEMS USING LIDAR DIFFUSION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Inverse problems are widely studied across imaging domains, but formulations for LiDAR under real-world corruptions remain underexplored. LiDAR is frequently corrupted by adverse weather, which we cast as a LiDAR inverse problem with explicit measurement operators. However, when 2D inverse problem frameworks are naively applied to LiDAR range views, they struggle to handle the strong spatial correlations inherent in 2.5D data. This is because they typically rely on an unrealistic assumption of pixel independence. To address this limitation, we present a new diffusion model guidance mechanism, termed correlative guidance, explicitly designed to model these missing spatial dependencies. Our framework augments the standard L2 guidance, which assumes pixel independence, with our proposed correlative term. This new guidance is implemented as a tractable, robust patch-wise Earth Mover's Distance (EMD) that compares local depth histograms. By operating on local distributions, our correlative guidance enforces local geometric consistency, a property ignored by standard L2 guidance. Extensive experiments across eight distinct corruption types demonstrate the effectiveness of our method and establish a new state-of-the-art for solving LiDAR inverse problems.

1 INTRODUCTION

LiDAR sensors are crucial in 3D perception systems, particularly for safety-critical applications like autonomous driving and robotics. They provide precise geometric measurements of the surrounding environment, enabling robust scene understanding. However, these sensors' reliability is often compromised in real-world scenarios. LiDAR point clouds are frequently corrupted by adverse weather conditions such as fog and snow, which cause signal attenuation and range ambiguity. Furthermore, they are vulnerable to sensor malfunctions or occlusions, leading to missing beams and incomplete data. These corruptions severely degrade LiDAR data quality and can cause catastrophic failures in downstream perception tasks. Therefore, developing a robust method to obtain clean, complete LiDAR scans from corrupted or partial measurements is a critical, still-open challenge.

At its core, an *inverse problem* mathematically formalizes the task of recovering a true, high-fidelity signal from a corrupted measurement mapped by a known forward operator. This long-standing field of study, particularly in the image domain (Yu et al., 2011; Jin et al., 2017; Zhang et al., 2016; Lahiri et al., 2023; Ulyanov et al., 2018), encompasses critical tasks such as *image inpainting, deblurring*, and *super-resolution*. In their standard formulation, these tasks assume that the specific degradation operator, such as an inpainting mask or a blur kernel, is known. These problems are ill-posed, as a single measurement can correspond to a vast space of plausible signals.

Recently, score-based diffusion models (Song et al., 2020) have been actively explored as powerful learned priors to effectively mitigate this inherent ill-posedness, achieving state-of-the-art results across various 2D inverse problems (Song et al., 2020; Jalal et al., 2021; Chung & Ye, 2022; Chung et al., 2022b; Kadkhodaie & Simoncelli, 2021).

However, adapting these 2D frameworks to the LiDAR domain poses a fundamental challenge because of a modality mismatch between dense 2D images and sparse 3D point clouds. To bridge this gap, we adopt the prevailing strategy of projecting 3D data into a 2D range-view representation, reframing the 3D problem as a 2.5D one. We tackle the LiDAR inverse problem by adapting Diffusion Posterior Sampling (DPS) (Chung et al., 2022a) to this 2.5D data. DPS is a powerful 2D framework that guides a pre-trained diffusion model using a likelihood term derived from the measurement.

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.

While this framework provides the two necessary components, the prior and the likelihood, a naive translation of the likelihood term is fundamentally flawed. Standard DPS typically derives its guidance from an L2 loss. This implicitly assumes the measurement noise is white Gaussian, which dictates that all pixels are statistically independent (i.e., a diagonal precision matrix $\mathbf{K}_{diag}$). This assumption is fundamentally violated in LiDAR range views. A range view is a 2.5D projection of a 3D scene; adjacent pixels are not independent but are strongly correlated, as they often represent points on the same continuous 3D surface. The L2 loss completely ignores this inherent spatial structure. Therefore, establishing a well-defined likelihood tailored for this new LiDAR domain is a critical foundational step. Once properly formulated to account for these spatial dependencies, this likelihood not only enables robust inverse problem solving of known corruptions but also paves the way for advanced applications such as blind inverse problems (Chung et al., 2023a), where the exact degradation operator is unknown.

To address this limitation, we rethink the LiDAR likelihood. We argue that a faithful likelihood model must account for these spatial correlations (i.e., a non-diagonal precision matrix $\mathbf{K}_{off}$). We propose to augment the standard L2 guidance with a novel correlative guidance term, $\mathbf{g}_{corr}$, designed to approximate this neglected off-diagonal component. As in §4.4, we introduce a tractable surrogate for this guidance based on a patch-wise Earth Mover's Distance (EMD) on local depth histograms. This EMD term enforces local geometric consistency and is robust to non-overlapping distributions caused by severe corruptions, providing robust guidance.

Our main contributions are summarized as follows:

- To the best of our knowledge, this work is the first to formulate and solve a wide range of challenging LiDAR inverse problems using a score-based diffusion model.
- We propose a new likelihood formulation for LiDAR data, decomposing the guidance into a pixel-wise term and a novel "correlative guidance" component to model spatial dependencies.
- We introduce a tractable implementation for this correlative guidance using a patch-wise EMD on local depth histograms, which enforces local geometric consistency.
- We demonstrate the effectiveness of our framework on the SemanticKITTI benchmark across eight distinct corruption types, where it improves over the strongest baseline on nearly all metrics.

2 Related Work

2.1 Diffusion Models for Inverse Problems

Score-based diffusion models have emerged as powerful data priors for inverse problems, effectively mitigating the inherent ill-posedness across a wide range of degradations (Song et al., 2020; Kawar et al., 2021; 2022; Wang et al., 2022; Zhu et al., 2023; Feng et al., 2023; Feng & Bouman, 2023; Mardani et al., 2023; Wu et al., 2024b; Rozet et al., 2024). A prominent approach, Diffusion Posterior Sampling (DPS) (Chung et al., 2022a), mitigates the intractability of the likelihood gradient by using an approximation derived from Tweedie's formula (Efron, 2011). Our work extends this diffusion-based inverse problem framework to the LiDAR domain. This requires a generative prior for LiDAR, which has recently been made available via models trained on a 2D range-view representation (Zyrianov et al., 2022; Ran et al., 2024; Wu et al., 2024a; Nakashima & Kurazume, 2024). We adopt R2DM (Nakashima & Kurazume, 2024) as our pre-trained LiDAR prior and extend DPS with correlative guidance tailored to LiDAR range views.

2.2 LiDAR in Adverse Conditions

LiDAR sensors are crucial for 3D perception in autonomous driving (Zhou & Tuzel, 2018; Lang et al., 2019; Meyer et al., 2019; Li et al., 2021; Liu et al., 2022). Yet their measurements are often degraded by adverse weather conditions such as fog and snow, and sensor-level issues including missing beams and crosstalk. This has motivated extensive research on LiDAR scene denoising and adverse-weather perception (Heinzler et al., 2020; Yu et al., 2022; Seppänen et al., 2022; Xiao et al.,

2023; Zhao et al., 2024; Park et al., 2025), along with physically grounded simulation methods (Hahner et al., 2022; 2021; Kilic et al., 2025) that reproduce degradations to synthesize realistic corrupted LiDAR data. However, most approaches are highly specialized, often relying on supervised models trained for a single, known degradation type. Even recent diffusion-based methods, such as 3DR-DIFF (Yasas Mahima et al., 2024), still address this challenge as a supervised reconstruction task. In contrast, our approach address LiDAR degradations as a diffusion-based inverse problem, using the corruption methodology from Robo3D (Kong et al., 2023) as our forward measurement model.

3 BACKGROUND

Score-based diffusion models learn the time-dependent score function $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x})$ through denoising score matching. The learned score guides a reverse process that generates samples by progressively removing noise. For LiDAR data, recent works (Ran et al., 2024; Wu et al., 2024a; Nakashima & Kurazume, 2024) operate on 2D range-view representations obtained through spherical projection of 3D point clouds and their reflectances.

To solve inverse problems using diffusion posterior sampling (DPS), we consider the task of recovering an unknown signal $\mathbf{x_0} \in \mathbb{R}^n$ from a given measurement $\mathbf{y} \in \mathbb{R}^m$. The forward model is generally formulated as:

$$\mathbf{y} = \mathcal{A}(\mathbf{x_0}) + \mathbf{n}, \quad \mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}), \tag{1}$$

where $\mathcal{A}(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^m$ is the forward operator and $\mathbf{n}$ is the *white gaussian* measurement noise. To solve the ill-posed inverse problem, the goal is to sample from the posterior distribution $p(\mathbf{x}_0|\mathbf{y})$. To leverage a score-based diffusion model for inverse problem, the reverse process must be guided by the time-dependent posterior score. This score is formulated using Bayes' rule as follows:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{y}) = \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t) \tag{2}$$

Although the score-based diffusion model provides the prior score $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$, the likelihood score $\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)$ remains computationally intractable. To address this, Chung *et al.* (Chung et al., 2022a) proposed the following approximation:

$$p(\mathbf{y}|\mathbf{x}_t) \approx p(\mathbf{y}|\hat{\mathbf{x}}_0), \quad \hat{\mathbf{x}}_0 = \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t] \tag{3}$$

The clean data estimate $\hat{\mathbf{x}}_0$ is readily computable from $\mathbf{x}_t$ through Tweedie's theorem (Efron, 2011). Given the forward model, which assumes *white gaussian* noise $\mathbf{n}$, the likelihood score can be expanded into a tractable term as follows:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\hat{\mathbf{x}}_0) = -\frac{1}{2\sigma^2} \nabla_{\mathbf{x}_t} \|\mathbf{y} - \mathcal{A}(\hat{\mathbf{x}}_0)\|_2^2 \tag{4}$$

Combining the prior score $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$ with the tractable likelihood guidance in Eq. 4 enables posterior sampling. However, this pixel-wise guidance does not explicitly account for spatial relationships in LiDAR range views, motivating our proposed correlative guidance.

4 LIDAR INVERSE PROBLEMS

4.1 PROBLEM FORMULATION

Let a clean LiDAR scan $\mathbf{L}$ be an unordered set of 3D points with reflectance:

$$\mathbf{L} = \{(\mathbf{p}_i, r_i)\}_{i=1}^N \in \mathbb{R}^{N \times 4} \tag{5}$$

Here, $\mathbf{p}_i = (x_i, y_i, z_i)$ are the 3D coordinates and r_i is the reflectance value. We utilize a spherical projection function $\Pi : \mathbb{R}^{N \times 4} \mapsto \mathbb{R}^{2 \times H \times W}$ to map $\mathbf{L}$ into a 2-channel range view image, including depth map $\mathbf{d} \in \mathbb{R}^{1 \times H \times W}$ and reflectance $\mathbf{r} \in \mathbb{R}^{1 \times H \times W}$, which we denote as our clean data $\mathbf{x}_0$:

$$\mathbf{x}_0 = \Pi(\mathbf{L}) \in \mathbb{R}^{2 \times H \times W} \tag{6}$$

We model LiDAR corruptions (e.g., *Beam Missing, Fog, Snow*) as operators $\mathcal{C} : \mathbb{R}^{N \times 4} \mapsto \mathbb{R}^{N \times 4}$ acting in the 3D domain.Given a clean scan $\mathbf{L}$, the corrupted scan is $\mathbf{L}^c = \mathcal{C}(\mathbf{L})$, and its spherical projection forms our 2D measurement $\mathbf{y}$:

$$\mathbf{y} = \Pi(\mathbf{L}^c) \in \mathbb{R}^{2 \times H \times W} \tag{7}$$

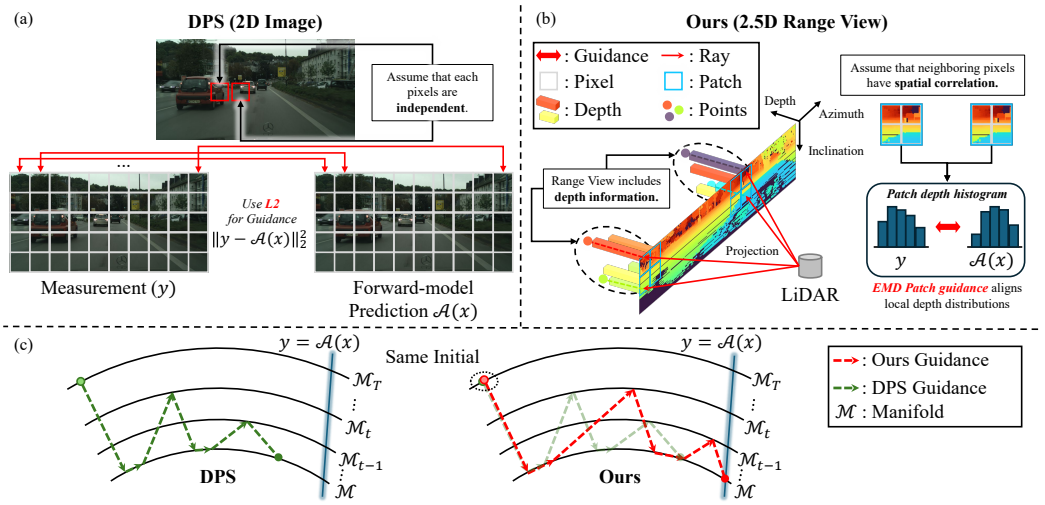

Figure 1: Overview of our framework. (a) Standard DPS uses pixel-independent L2 guidance, assuming diagonal precision and ignoring spatial correlations. (b) Our patch-wise EMD guidance models local dependencies in 2.5D LiDAR range views. (c) Without modeling off-diagonal correlations, DPS fails to reach the true measurement manifold, whereas ours converges more stably.

To connect the clean data $\mathbf{x}_0$ to the measurement $\mathbf{y}$ within the inverse problem framework, we define a composite forward operator $\mathcal{A}$. This operator first maps the 2D range view back to a 3D point cloud using the inverse spherical projection Π^{-1}, applies the 3D corruption $\mathcal{C}$, and projects the result back into the 2D range view:

$$\mathcal{A}(\cdot) = (\Pi \circ \mathcal{C} \circ \Pi^{-1})(\cdot) \tag{8}$$

Our goal is to recover the clean LiDAR data $\mathbf{x}_0$ from the corrupted measurement $\mathbf{y}$ using Diffusion Posterior Sampling (DPS) (§3). To this end, our framework consists of two components: a differentiable, corruption-aware forward operator that approximates the 3D corruption process in the 2D range-view domain (§4.2), and correlative guidance that complements standard L2 guidance by accounting for local depth relationships (§4.3-§4.4). Figure 1 provides an overview of our framework.

4.2 Corruption-aware Forward Operator

The DPS guidance (Eq. 4) requires backpropagating through the forward operator to compute the gradient with respect to the clean data estimate $\hat{\mathbf{x}}_0$. Hence, the forward operator must be differentiable. However, directly implementing the composite operator in (Eq. 8) is non-differentiable in practice. The 3D corruption $\mathcal{C}$ involves discrete operations, such as adding or dropping points, while the projection Π relies on a non-differentiable z-buffer operation. Both prevent gradient propagation.

To address this, we introduce a differentiable, *corruption-aware forward operator*, $\mathcal{A}$ that approximates the full 3D corruption process in the 2D range-view domain. This operator is derived from the known corruption model $\mathcal{C}$ and is assumed to be given at test time. The core idea is to pre-compute how the 3D corruption manifests in the 2D range view using two maps: (1) a *pixel mapping* $\mathbf{M}$, which links each measurement pixel $\mathbf{q}^c$ to its corresponding clean pixel $\mathbf{q}$, and (2) a *per-pixel corruption map* $\mathbf{C}$, which stores the corruption parameters $\mathbf{c}$ at $\mathbf{q}$.

To apply the parameters stored in $\mathbf{C}$, we define a differentiable, pixel-wise transform $\mathcal{T}_{\mathbf{c}}(\cdot) : \mathbb{R}^2 \mapsto \mathbb{R}^2$ that operates on a clean pixel's depth-reflectance vector $[d, r]^\top$:

$$\mathcal{T}_{\mathbf{c}}([d, r]^\top) = \begin{bmatrix} m(d + \beta) \\ m\alpha r \end{bmatrix}, \quad \mathbf{c} = (m, \alpha, \beta) \tag{9}$$

Here, the parameters model common corruptions: m represents a point drop (masking), α models signal attenuation, and β represents a range bias.

Using these components, our 2D-to-2D operator $\mathcal{A}$ computes a measurement from the clean estimate $\hat{\mathbf{x}}_0$. For each measurement pixel $\mathbf{q}^c$, it retrieves the corresponding clean pixel $\mathbf{q} = \mathbf{M}[\mathbf{q}^c]$ and applies

the corruption parameters $\mathbf{c} = \mathbf{C}[\mathbf{q}]$:

$$\mathcal{T}_{\mathbf{C}[\mathbf{q}]}(\hat{\mathbf{x}}_0[\mathbf{q}]) = \mathcal{A}(\hat{\mathbf{x}}_0)[\mathbf{q}^c] \tag{10}$$

Since $\mathcal{T}_\mathbf{c}$ is differentiable and both $\mathbf{M}$ and $\mathbf{C}$ remain fixed during sampling, the resulting operator $\mathcal{A}$ is differentiable with respect to $\hat{\mathbf{x}}_0$.

4.3 Rethinking the LiDAR Likelihood

With the differentiable forward operator $\mathcal{A}$ (Eq. 10), we can define the likelihood guidance for LiDAR inverse problems. Standard DPS uses the pixel-wise L2 distance between the measurement $\mathbf{y}$ and the forward-model prediction $\mathcal{A}(\hat{\mathbf{x}}_0)$, without explicitly accounting for neighboring pixel relationships. This is particularly limiting for LiDAR range views, where adjacent pixels often correspond to points on the same 3D surface and exhibit strong spatial correlations. To account for these local relationships, we introduce correlative guidance to complement the standard L2 guidance.

We motivate this formulation by revisiting the likelihood assumption in DPS. As described in §3, standard DPS assumes additive white Gaussian measurement noise, $\mathbf{n} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$. The "white" property implies a diagonal covariance matrix $\mathbf{\Sigma} = \sigma^2\mathbf{I}$, and consequently a diagonal precision matrix $\mathbf{K} = \mathbf{\Sigma}^{-1}$. Under this assumption, the general multivariate Gaussian log-likelihood, expressed using the residual $\mathbf{r} = \mathbf{y} - \mathcal{A}(\hat{\mathbf{x}}_0)$, simplifies to:

$$\log p(\mathbf{y}|\hat{\mathbf{x}}_0) \propto -\frac{1}{2}\mathbf{r}^T\mathbf{K}\mathbf{r} = -\frac{1}{2\sigma^2}(\mathbf{r}^T\mathbf{r}) \tag{11}$$

Taking the gradient of this expression yields the conventional L2 guidance in Eq. 4.

However, this diagonal formulation does not explicitly capture spatial dependencies in LiDAR range views. To motivate guidance that accounts for such dependencies, we consider a more general measurement model with a non-diagonal covariance matrix:

$$\mathbf{y} = \mathcal{A}(\mathbf{x}_0) + \mathbf{n}, \quad \text{where} \quad \mathbf{n} \sim \mathcal{N}(0, \mathbf{\Sigma}_L) \tag{12}$$

Here, $\mathbf{\Sigma}_L$ represents a covariance model that allows dependencies between measurement residuals. Its corresponding precision matrix is $\mathbf{K}_L = \mathbf{\Sigma}_L^{-1}$. Unlike the diagonal formulation used in standard DPS, this model allows correlations between different pixel locations to contribute to the likelihood.

Directly modeling $\mathbf{K}_L$ is intractable, as the exact correlation structure between all pixels is unknown. Instead, we decompose the precision matrix into its diagonal and off-diagonal components, $\mathbf{K}_L = \mathbf{K}_{diag} + \mathbf{K}_{off}$. Substituting this decomposition into the general Gaussian log-likelihood gives:

$$\log p(\mathbf{y}|\hat{\mathbf{x}}_0) \propto -\frac{1}{2}\mathbf{r}^T(\mathbf{K}_{diag} + \mathbf{K}_{off})\mathbf{r} = -\frac{1}{2}\mathbf{r}^T\mathbf{K}_{diag}\mathbf{r} - \frac{1}{2}\mathbf{r}^T\mathbf{K}_{off}\mathbf{r} \tag{13}$$

The first term represents the pixel-wise contribution, which we retain through the conventional L2 guidance in Eq. 4. The second term represents contributions from inter-pixel correlations, which are not explicitly modeled by standard DPS.

We refer to the additional guidance motivated by this off-diagonal component as **correlative guidance**. Since $\mathbf{K}_{off}$ is unknown, directly computing its gradient is intractable. We therefore retain the existing L2 guidance and introduce a tractable surrogate $\mathbf{g}_{corr}$ to capture local dependencies neglected by the pixel-wise formulation. Our resulting likelihood guidance is defined as:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\hat{\mathbf{x}}_0) = -\frac{1}{2\sigma^2}\nabla_{\mathbf{x}_t}\|\mathbf{y} - \mathcal{A}(\hat{\mathbf{x}}_0)\|_2^2 + \mathbf{g}_{corr} \tag{14}$$

The following section describes how we implement this correlative guidance using patch-wise Earth Mover's Distance (EMD) on local depth distributions.

4.4 Tractable Surrogate for Correlative Guidance

Following Eq. 14, we design a tractable surrogate for the correlative guidance $\mathbf{g}_{corr}$ using patch-wise Earth Mover's Distance (EMD). Specifically, we compare local depth distributions within corresponding patches of $\mathbf{y}$ and $\mathcal{A}(\hat{\mathbf{x}}_0)$, capturing local depth statistics without requiring exact pixel-wise alignment within each patch. Combined with standard L2 guidance, this encourages local depth consistency while maintaining measurement alignment.

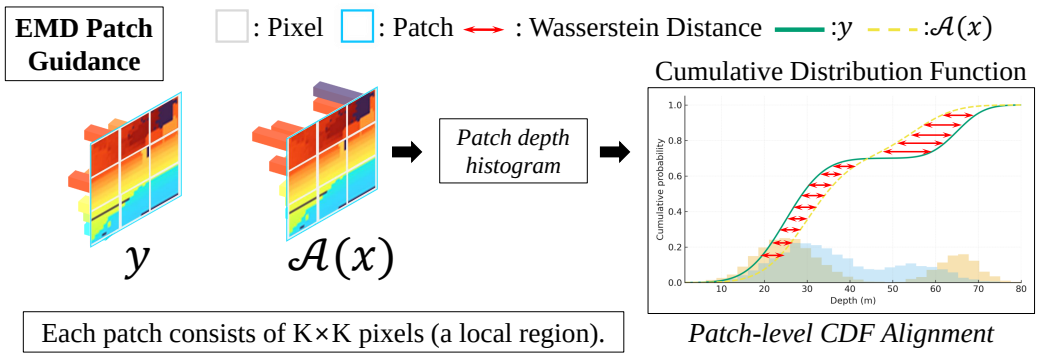

Figure 2: Patch-wise comparison of observed $\mathbf{y}$ and predicted $\mathcal{A}(x)$ depth distributions in the range-view. The EMD between local distributions is computed as the L1 distance between their CDFs.

We adopt EMD because it measures the distance between distributions while accounting for the distances between their values. In particular, EMD remains informative even when two depth distributions do not overlap, which can occur under severe LiDAR corruptions. Since our goal is to enforce consistency in local depth structure, we apply EMD only to the depth channel $\mathbf{d}$.

We define surrogate correlative log-likelihood $\log p_{EMD}$ as the negative sum of patch-wise EMD:

$$\log p_{EMD}(\mathbf{y}|\hat{\mathbf{x}}_0) = -\lambda_{EMD} \sum_{p} \text{EMD}\left(\mu(\mathbf{y}_p), \mu(\mathcal{A}(\hat{\mathbf{x}}_0)_p)\right) \tag{15}$$

Here, p indexes all patches, λ_{EMD} is a weighting coefficient, and $\mu(\cdot)$ denotes the 1D empirical distribution of the depth values within that patch.

We represent these distributions as depth histograms over N_{bins} predefined depth intervals. To enable gradient propagation, we use soft histograms instead of hard binning. Specifically, we place a Gaussian kernel at each transformed depth value in the u-domain and accumulate its responses over uniformly spaced bin centers. The kernel bandwidth is set proportional to the bin spacing Δu, smoothing the empirical distributions and providing stable gradients.

The EMD between the resulting histograms is efficiently computed as the L1 distance between their Cumulative Distribution Functions (CDFs), as illustrated in Fig. 2:

$$\text{EMD}(\mu_1, \mu_2) = \sum_{i=1}^{N_{\text{bins}}} |\text{CDF}(\mu_1)_i - \text{CDF}(\mu_2)_i| \tag{16}$$

The patch size and binning strategy are hyperparameters whose effects are analyzed in our experiments. Finally, the proposed correlative guidance $\mathbf{g}_{corr}$ is obtained by differentiating the surrogate objective through standard backpropagation:

$$\mathbf{g}_{corr} = \nabla_{\mathbf{x}_t} \log p_{EMD}(\mathbf{y}|\hat{\mathbf{x}}_0) \tag{17}$$

This gradient penalizes discrepancies between local depth distributions, complementing the pixel-wise L2 guidance with information about local depth relationships. Together, the two terms provide measurement guidance that accounts for both individual pixel values and local depth statistics. We further verify empirically that this descriptor encodes local geometry in *Appendix*.

5 Experiments

5.1 Experimental Settings

Datasets. We evaluate our method on two large-scale LiDAR datasets: SemanticKITTI (Behley et al., 2019) and KITTI-360 (Liao et al., 2022). We use the SemanticKITTI full validation split(4,071 frames) for evaluation. For KITTI-360, we uniformly sample 10% of the test split (2,637 frames) to ensure diverse scene coverage while maintaining efficiency.We simulate corruption conditions using the methodology adapted from Robo3D (Kong et al., 2023).

Table 1: Comparison across various corruption types. Best results in **bold**. ↓: lower is better.

		Point Cloud	Image	BEV		Depth	Reflectance
Corruption	Method	FPD ↓	FRD ↓	JSD $\times 10^2$ ↓	MMD $\times 10^4$ ↓	MAE ↓	MAE $\times 10^1$ ↓
Beam Missing	DPS (Chung et al., 2022a)	9.733	318.6	1.673	0.309	0.586	0.386
	DPS + TV	8.246	341.0	1.758	0.356	0.593	0.400
	Ours	**6.402** (+22.36%)	**302.8** (+4.96%)	**1.654** (+1.14%)	**0.286** (+7.44%)	**0.576** (+1.70%)	**0.385** (+0.26%)
Cross Sensor	DPS (Chung et al., 2022a)	20.84	390.9	2.194	0.421	0.694	0.678
	DPS + TV	45.91	424.1	2.339	0.428	0.687	0.694
	Ours	**12.64** (+39.35%)	**382.5** (+2.15%)	**2.059** (+6.15%)	**0.389** (+7.60%)	**0.676** (+1.60%)	**0.677** (+0.15%)
Crosstalk	DPS (Chung et al., 2022a)	14.79	111.2	0.947	0.037	0.207	**0.036**
	DPS + TV	23.79	102.3	1.081	0.038	0.211	0.042
	Ours	**9.103** (+38.45%)	**101.7** (+0.59%)	**0.819** (+13.52%)	**0.031** (+16.22%)	**0.199** (+3.86%)	**0.036**
Fog	DPS (Chung et al., 2022a)	**8.50**	123.4	1.441	0.036	0.291	**0.055**
	DPS + TV	97.37	184.4	1.514	0.045	0.296	0.072
	Ours	9.168	**119.3** (+3.32%)	**1.420** (+1.46%)	**0.035** (+2.78%)	**0.283** (+2.75%)	**0.055**
Incomplete Echo	DPS (Chung et al., 2022a)	15.38	140.1	1.446	0.260	0.268	**0.080**
	DPS + TV	36.64	134.9	1.663	0.287	0.297	0.084
	Ours	**8.938** (+41.90%)	**130.4** (+3.34%)	**1.341** (+7.26%)	**0.240** (+7.69%)	**0.264** (+1.49%)	**0.080**
Motion Blur	DPS (Chung et al., 2022a)	334.4	259.4	2.555	0.350	0.606	0.507
	DPS + TV	560.1	**228.2**	2.598	0.288	**0.533**	**0.486**
	Ours	**101.4** (+69.68%)	**228.2**	**1.592** (+37.69%)	**0.182** (+36.81%)	0.560	0.507
Snow	DPS (Chung et al., 2022a)	29.57	166.6	1.665	0.221	0.384	0.330
	DPS + TV	77.32	165.2	1.766	**0.186**	0.433	0.332
	Ours	**17.14** (+42.05%)	**157.8** (+4.48%)	**1.537** (+7.69%)	0.201	**0.376** (+2.08%)	**0.329** (+0.30%)
Wet Ground	DPS (Chung et al., 2022a)	17.51	523.1	2.037	0.881	0.514	0.307
	DPS + TV	25.65	659.0	2.843	1.557	0.642	0.387
	Ours	**12.28** (+29.87%)	**518.7** (+0.84%)	**1.982** (+2.70%)	**0.834** (+5.33%)	**0.500** (+2.72%)	**0.306** (+0.33%)

Evaluation Metrics. We evaluate our method using both distributional and reconstruction-based metrics, following standard protocols in recent LiDAR generation and restoration studies (Nakashima & Kurazume, 2024; Yasas Mahima et al., 2024; Zyrianov et al., 2022). For distributional evaluation, we report (1) Fréchet Range Distance (FRD) on range and reflectance images, (2) Fréchet Point Cloud Distance (FPD) (Shu et al., 2019) computed with a PointNet (Qi et al., 2017) pretrained on ShapeNet (Chang et al., 2015), and (3) bird's-eye-view (BEV) metrics including Jensen Shannon Divergence (JSD) and Maximum Mean Discrepancy (MMD) (Nakashima & Kurazume, 2024). For reconstruction fidelity, we measure the Mean Absolute Error (MAE) between generated and clean LiDAR range-view images.

Implementation Details. For the EMD guidance, we generally use a patch size of 3, a bin size of 32, a bandwidth of 0.5, and a *log* bin domain transformation. The λ_{EMD} value is adjusted for each corruption type. The guidance scale increases linearly over 128 sampling steps. LiDAR range-view image is set to a resolution of 64×1024. All experiments employ the pretrained R2DM (Nakashima & Kurazume, 2024) model. Per-corruption guidance, schedules and details of $\mathcal{A}$ are in *Appendix*.

Baselines. We select our baselines based on their compatibility with the non-linear forward measurement operators $\mathcal{A}$ defined in § 4.2. Therefore, standard inverse problem solvers that require a linear forward model, such as ADMM-TV, DDRM (Kawar et al., 2022), and DiffusionMBIR (Chung et al., 2023b), are excluded. We apply **DPS** (Chung et al., 2022a) with standard L2 guidance directly to the 2D LiDAR range view. This approach computes the likelihood approximation pixel by pixel, implicitly assuming that each pixel is spatially independent of its neighbors. Inspired by DiffusionMBIR (Chung et al., 2023b), we augment DPS with a Total Variation (TV) regularization term, **DPS+TV**. This variant enforces spatial smoothness along the 2D coordinates of the range-view representation. Comprehensive justifications for exclusions and details of baselines are in *Appendix*.

5.2 Comparison with Inverse Problem Solving Methods

Table 1 shows a quantitative comparison between prior methods and ours across eight corruption types with *Moderate* severity, following the LiDAR degradation taxonomy of Robo3D (Kong et al., 2023). Ours achieves overall performance improvement across nearly all metrics and corruption types compared to existing approaches, demonstrating enhanced reconstruction quality in both 2D range-view and 3D point-cloud domains. In particular, the gains are most pronounced in the point-cloud metric (FPD), where ours improves over the stronger baseline on seven of the eight corruptions. These improvements imply that our approach better exploits local geometric correlations, enabling more faithful 3D reconstruction and robust performance across diverse LiDAR corruptions.

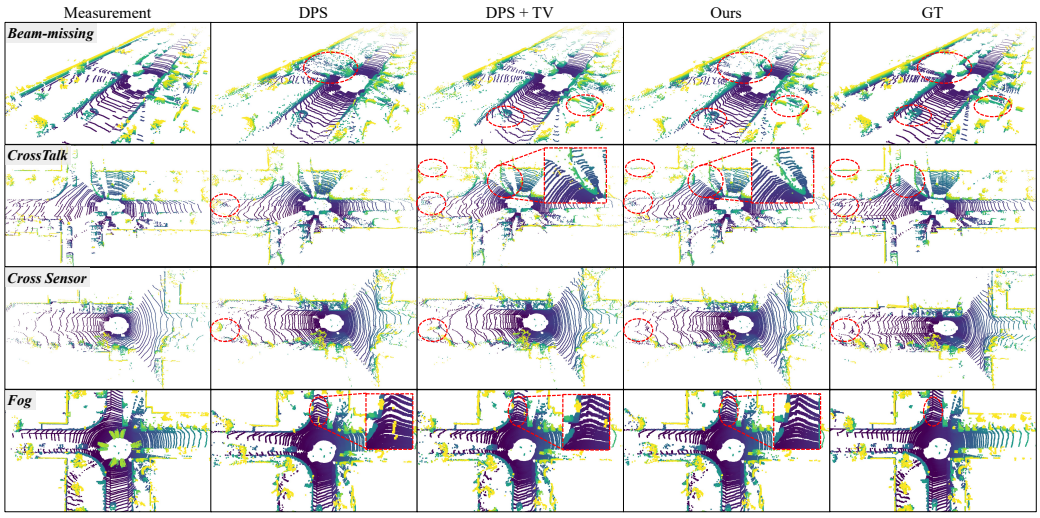

Figure 3: Qualitative comparison of LiDAR inverse problem solving under four corruptions.

Table 2: Comparison with supervised LiDAR denoisers on *Fog* and *Snow*.

			Point Cloud	Image	BEV		Depth	Reflectance
Corruption	Method	Setting	FPD ↓	FRD ↓	JSD $\times 10^2$ ↓	MMD $\times 10^4$ ↓	MAE ↓	MAE $\times 10^1$ ↓
Fog	WeatherNet (Heinzler et al., 2020)	Supervised	7.081	394.8	7.859	1.702	1.207	0.910
	4DenoiseNet (Seppänen et al., 2022)	Supervised	**4.737**	401.5	6.418	0.411	0.913	0.911
	TripleMixer (Zhao et al., 2025)	Supervised	5.859	421.4	12.71	11.22	2.017	0.889
	Ours	Zero-shot	9.168	**119.3**	**1.420**	**0.035**	**0.283**	**0.055**
Snow	WeatherNet (Heinzler et al., 2020)	Supervised	399.5	384.8	5.301	0.440	1.601	0.342
	4DenoiseNet (Seppänen et al., 2022)	Supervised	29.36	371.4	1.670	0.439	0.540	0.493
	TripleMixer (Zhao et al., 2025)	Supervised	521.2	669.1	6.562	1.010	2.450	0.425
	Ours	Zero-shot	**17.14**	**157.8**	**1.537**	**0.201**	**0.376**	**0.329**

Figure 3 illustrates these trends qualitatively. In *Beam Missing*, DPS produces elongated artifacts behind the vehicle, whereas our method successfully solves the inverse problem by leveraging correlations on the vehicle surface to reconstruct plausible structures. In *Crosstalk* and *Cross Sensor*, prior methods generate spurious road artifacts, while ours reconstruct points aligned with the ground. Furthermore, DPS+TV produces noisy points around walls under *Crosstalk* as shown in the red box, whereas ours yields clean boundaries. Both quantitative and qualitative results demonstrate that incorporating non-diagonal covariance into the LiDAR likelihood and introducing correlative guidance are crucial for achieving more accurate, geometrically consistent solutions. Additional qualitative comparisons are provided in *Appendix*.

5.3 Comparison with Supervised LiDAR Denoisers

As shown in Tab. 2, we compare our zero-shot framework with supervised LiDAR denoisers. These methods are developed for adverse weather, and *Fog* and *Snow* are the conditions they share with our benchmark, while our formulation applies to any degradation expressible as a differentiable forward operator. Unlike denoisers that primarily remove corrupted points while preserving observed ones, our framework reconstructs the entire scene, including missing regions. Since these denoisers only predict which measurements to discard, we inpaint the removed regions before evaluation. Under *Fog*, the denoisers alter only a small fraction of the measurement, so their output stays close to the input distribution and yields a low FPD. Leaving the corruption largely in place, however, costs them on every other metric. Our framework instead draws a complete scene from a learned LiDAR prior constrained by the known forward operator, which departs from the measurement but recovers the structure the denoisers leave missing. Nevertheless, ours achieves the best performance across the remaining five metrics under *Fog* and all metrics under *Snow* among the supervised denoisers. These results highlight the advantage of our framework in recovering complete LiDAR scans without corruption-specific training. A comparison against heuristic filters is provided in *Appendix*.

Table 3: Ablation Studies under the *Moderate, Beam Missing* condition on SemanticKITTI.

		Point Cloud	Image	BEV		Depth	Reflectance
Experiment	Method	FPD ↓	FRD ↓	JSD $\times 10^2$ ↓	MMD $\times 10^4$ ↓	MAE ↓	MAE $\times 10^1$ ↓
Corruption Severity (*Heavy*)	DPS (Chung et al., 2022a)	22.19	372.4	1.868	0.319	0.730	0.501
	Ours	**9.123**	**357.7**	**1.719**	**0.292**	**0.726**	**0.500**
Surrogate Design	SSIM	10.643	315.3	1.689	0.314	**0.572**	0.386
	Sorted-L1	6.422	315.9	1.917	0.361	0.729	**0.385**
	Hist-Intersection	9.920	313.3	**1.633**	0.294	0.577	0.386
	EMD	**6.402**	**302.8**	1.654	**0.286**	0.576	**0.385**
KITTI-360	DPS (Chung et al., 2022a)	4.390	444.7	1.811	**0.303**	0.729	0.397
	Ours	**2.840**	**426.0**	**1.763**	0.304	**0.717**	**0.396**

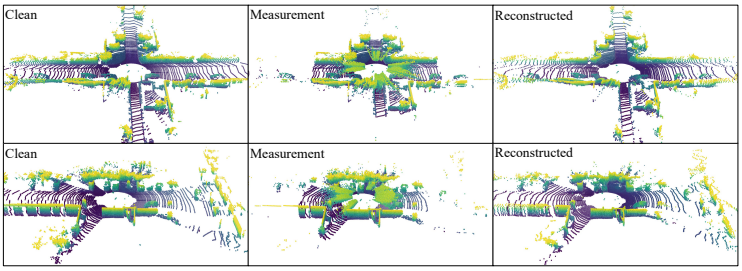

Figure 4: Reconstruction result (Approximate Forward Operator $\hat{\mathcal{A}}$, *Moderate, Fog*)

5.4 Ablation Study

Corruption Severity. We evaluate the robustness of our method under other corruption severity level, *Heavy*. As shown in Tab. 3, our model consistently outperforms DPS across different corruption severity levels. These results demonstrate that our approach exhibits superior robustness in solving the inverse problem compared to prior work, regardless of the degradation severity.

Surrogate Design. To validate the choice of EMD as a surrogate for correlative guidance, we compare it with SSIM, SortedL1, and histogram intersection (HistInter) under *Beam Missing*. EMD provides the most balanced performance across reconstruction and distributional metrics, supporting its effectiveness in capturing local depth relationships beyond pixel-wise guidance.

Experiments on KITTI-360 Dataset. To evaluate the generalizability of our method, we conducted a comparative experiment between DPS (Chung et al., 2022a) and Ours on the KITTI-360 dataset (Liao et al., 2022). The evaluation was performed on the KITTI-360 subset under the *Moderate, Beam Missing* corruption, using the R2DM (Nakashima & Kurazume, 2024) trained on the KITTI-360 training set. As shown in Tab. 3, our method consistently outperforms DPS across all metrics except BEV MMD, where the two are effectively tied. These results exhibit a trend consistent with the SemanticKITTI experiments, demonstrating the generalizability of our approach.

Additional ablations on the patch size and bin domain are provided in *Appendix*.

5.5 Posterior Sampling with the Approximate Forward Operator

We explore posterior sampling with an estimated surrogate operator $\hat{\mathcal{A}}$ when the exact forward operator $\mathcal{A}$ is unavailable. Using a fog-like corruption that suppresses points within 6 m, our framework reconstructs near-range geometry while preserving overall scene structure, as shown in Fig. 4. While not a comprehensive blind inverse problem benchmark, this proof-of-concept suggests the potential to extend our framework beyond the known-operator assumption.

6 Conclusion

In this work, we presented a diffusion-based framework for LiDAR inverse problems. We designed a differentiable, corruption-aware forward operator that approximates 3D degradations in the 2D range-view domain. To address spatial correlations neglected by standard pixel-wise guidance, we introduced correlative guidance, implemented using patch-wise EMD on local depth histograms. Experiments across eight corruption types demonstrate improved reconstruction quality over existing inverse solvers.

AI USE STATEMENT

In this work, we used generative AI tools for implementing methods and assisting with translation. We have not used generative AI tools for generating synthetic data sets, helping develop theoretical models or conceptual frameworks, formulating mathematical claims, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, cleaning and reformatting datasets, supporting qualitative and thematic data analysis, or interpreting results. Providing critical ingredients for proving mathematical claims and assisting in the writing of proofs are not applicable to this work, as it contains no formal proofs. Additionally, we used generative AI tools for drafting parts of a research paper, brainstorming, editing a research paper to improve readability, and identifying relevant literature. We have reviewed all AI-assisted work. All AI-assisted outputs were manually reviewed and verified by the authors for correctness, relevance, and consistency with the final manuscript. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9297–9307, 2019.

Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri. *Medical image analysis*, 80:102479, 2022.

Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022a.

Hyungjin Chung, Byeongsu Sim, and Jong Chul Ye. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12413–12422, 2022b.

Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. Parallel diffusion models of operator and image for blind inverse problems. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023a.

Hyungjin Chung, Dohoon Ryu, Michael T McCann, Marc L Klasky, and Jong Chul Ye. Solving 3d inverse problems using pre-trained 2d diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 22542–22551, 2023b.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Berthy T Feng and Katherine L Bouman. Variational bayesian imaging with an efficient surrogate score-based prior. *arXiv preprint arXiv:2309.01949*, 2023.

Berthy T Feng, Jamie Smith, Michael Rubinstein, Huiwen Chang, Katherine L Bouman, and William T Freeman. Score-based diffusion models as principled priors for inverse imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10520–10531, 2023.

Martin Hahner, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Fog simulation on real lidar point clouds for 3d object detection in adverse weather. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 15283–15292, 2021.

Martin Hahner, Christos Sakaridis, Mario Bijelic, Felix Heide, Fisher Yu, Dengxin Dai, and Luc Van Gool. Lidar snowfall simulation for robust 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16364–16374, 2022.

Robin Heinzler, Florian Piewak, Philipp Schindler, and Wilhelm Stork. Cnn-based lidar point cloud de-noising in adverse weather. *IEEE Robotics and Automation Letters*, 5(2):2514–2521, 2020.

Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G Dimakis, and Jon Tamir. Robust compressed sensing mri with deep generative priors. *Advances in neural information processing systems*, 34:14938–14954, 2021.

Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE transactions on image processing*, 26(9): 4509–4522, 2017.

Zahra Kadkhodaie and Eero Simoncelli. Stochastic solutions for linear inverse problems using the prior implicit in a denoiser. *Advances in Neural Information Processing Systems*, 34:13242–13254, 2021.

Bahjat Kawar, Gregory Vaksman, and Michael Elad. Snips: Solving noisy inverse problems stochastically. *Advances in Neural Information Processing Systems*, 34:21757–21769, 2021.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in neural information processing systems*, 35:23593–23606, 2022.

Velat Kilic, Deepti Hegde, A Brinton Cooper, Vishal M Patel, and Mark Foster. Lidar light scattering augmentation (lisa): Physics-based simulation of adverse weather conditions for 3d object detection. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2025.

Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in neural information processing systems*, 34:21696–21707, 2021.

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*, pp. 19994–20006, 2023.

Anish Lahiri, Gabriel Maliakal, Marc L Klasky, Jeffrey A Fessler, and Saiprasad Ravishankar. Sparse-view cone beam ct reconstruction using data-consistent supervised and adversarial learning from scarce training data. *IEEE Transactions on Computational Imaging*, 9:13–28, 2023.

Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12697–12705, 2019.

Zhichao Li, Feng Wang, and Naiyan Wang. Lidar r-cnn: An efficient and universal 3d object detector. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7546–7555, 2021.

Yiyi Liao, Jun Xie, and Andreas Geiger. Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45 (3):3292–3310, 2022.

Zhijian Liu, Haotian Tang, Alexander Amini, Xinyu Yang, Huizi Mao, Daniela Rus, and Song Han. Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation. *arXiv preprint arXiv:2205.13542*, 2022.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. *arXiv preprint arXiv:2305.04391*, 2023.

Gregory P Meyer, Ankit Laddha, Eric Kee, Carlos Vallespi-Gonzalez, and Carl K Wellington. Lasernet: An efficient probabilistic 3d object detector for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12677–12686, 2019.

Kazuto Nakashima and Ryo Kurazume. Lidar data synthesis with denoising diffusion probabilistic models. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 14724–14731. IEEE, 2024.

Junsung Park, Hwijeong Lee, Inha Kang, and Hyunjung Shim. No thing, nothing: Highlighting safety-critical classes for robust lidar semantic segmentation in adverse weather. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 6690–6699, 2025.

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, 2017.

Haoxi Ran, Vitor Guizilini, and Yue Wang. Towards realistic scene generation with lidar diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14738–14748, 2024.

François Rozet, Gérôme Andry, François Lanusse, and Gilles Louppe. Learning diffusion priors from observations by expectation maximization. *Advances in Neural Information Processing Systems*, 37:87647–87682, 2024.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems*, 35:36479–36494, 2022.

Alvari Seppänen, Risto Ojala, and Kari Tammi. 4denoisenet: Adverse weather denoising from adjacent point clouds. *IEEE Robotics and Automation Letters*, 8(1):456–463, 2022.

Dong Wook Shu, Sung Woo Park, and Junseok Kwon. 3d point cloud generative adversarial network based on tree structured graph convolutions. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3859–3868, 2019.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 9446–9454, 2018.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *arXiv preprint arXiv:2212.00490*, 2022.

Yang Wu, Kaihua Zhang, Jianjun Qian, Jin Xie, and Jian Yang. Text2lidar: Text-guided lidar point cloud generation via equirectangular transformer. In *European Conference on Computer Vision*, pp. 291–310. Springer, 2024a.

Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine Bouman. Principled probabilistic imaging using diffusion models as plug-and-play priors. *Advances in Neural Information Processing Systems*, 37:118389–118427, 2024b.

Aoran Xiao, Jiaxing Huang, Weihao Xuan, Ruijie Ren, Kangcheng Liu, Dayan Guan, Abdulmotaleb El Saddik, Shijian Lu, and Eric P Xing. 3d semantic segmentation in the wild: Learning generalized models for adverse-condition point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9382–9392, 2023.

K. T. Yasas Mahima, Asanka G. Perera, Sreenatha Anavatti, and Matt Garratt. 3dr-diff: Blind diffusion inpainting for 3d point cloud reconstruction and segmentation. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7414–7421, 2024. doi: 10.1109/IROS58592.2024.10802338.

Guoshen Yu, Guillermo Sapiro, and Stéphane Mallat. Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. *IEEE Transactions on Image Processing*, 21(5):2481–2499, 2011.

Ming-Yuan Yu, Ram Vasudevan, and Matthew Johnson-Roberson. Lisnownet: Real-time snow removal for lidar point clouds. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6820–6826. IEEE, 2022.

Hanming Zhang, Liang Li, Kai Qiao, Linyuan Wang, Bin Yan, Lei Li, and Guoen Hu. Image prediction for limited-angle tomography via deep learning with convolutional neural network. *arXiv preprint arXiv:1607.08707*, 2016.

Haimei Zhao, Jing Zhang, Zhuo Chen, Shanshan Zhao, and Dacheng Tao. Unimix: Towards domain adaptive and generalizable lidar semantic segmentation in adverse weather. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14781–14791, 2024.

Xiongwei Zhao, Congcong Wen, Xu Zhu, Yang Wang, Haojie Bai, and Wenhao Dou. Triplemixer: A triple-domain mixing model for point cloud denoising under adverse weather. *IEEE Transactions on Image Processing*, 34:7712–7727, 2025.

Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4490–4499, 2018.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1219–1229, 2023.

Vlas Zyrianov, Xiyue Zhu, and Shenlong Wang. Learning to generate realistic lidar point clouds. In *European Conference on Computer Vision*, pp. 17–35. Springer, 2022.

A Appendix

In this materials, we provide additional details and experimental results that could not be fully included in the main paper. We first elaborate on the proposed EMD Patch Guidance (B), including its formulation, intuition, and design choices. We then provide a deeper analysis of whether EMD Patch depth histograms capture meaningful spatial correlation through scene-wide retrieval experiments (C). In addition, we present further implementation details, including sampling configurations and baseline descriptions (D), as well as additional experiments on cost-performance trade-offs, mixed corruption, and additional ablation studies (E). Finally, we describe the details of each corruption map and forward operator (F), and provide additional qualitative results on SemanticKITTI and KITTI-360 (G).

B EMD Patch Guidance

In the main paper, we introduced the EMD Patch Guidance as a correlative term that enforces local consistency in the depth distributions and thereby approximates the effect of the off-diagonal precision matrix $\mathbf{K}_{off}$. Here, we provide a more detailed, concept-level description of the underlying formulation.

B.1 Depth-domain transform

Let $d \in [d_{\min}, d_{\max}]$ denote the metric depth. Rather than working directly in depth space, we first map d to a transformed "LiDAR-aware" domain:

$$u = g(d), \tag{B.1}$$

where g is chosen as a logarithmic transform. In our setting, we adopt a logarithmic transform of the form:

$$u = g(d) = \log(d + \varepsilon), \tag{B.2}$$

with a small constant $\varepsilon > 0$ to avoid numerical issues at very small depths. The logarithmic transform compresses large depths and relatively expands small depths, making the discrepancy measure

more sensitive to near-field errors while remaining well behaved for far-range points. We denote the transformed range by $[u_{\min}, u_{\max}] = [g(d_{\min}), g(d_{\max})]$.

Within this domain, we define N_{bins} uniformly spaced bin centers:

$$u_i = u_{\min} + (i-1)\,\Delta u, \quad i = 1, \ldots, N_{bins}, \tag{B.3}$$

where $\Delta u = |u_{\max} - u_{\min}|/(N_{bins} - 1)$ is the bin spacing and N_{bins} denotes the number of histogram bins used to discretize the transformed depth domain.

This uniform u-discretization enables our soft histogram construction and EMD computation. Following the depth bounds used in SemanticKITTI (Behley et al., 2019) and KITTI-360 (Liao et al., 2022), we set the depth range to $[1.45, 80]$ meters.

B.2 Patch-wise empirical distributions

For each pixel, we consider a local patch p of size $K \times K$ centered at that location in both the observed range-view $\mathbf{y}$ and the current prediction $\mathcal{A}(\hat{\mathbf{x}}_0)$. We treat the depth values in this patch as a small empirical sample from the local depth distribution, and we apply the depth-domain transform $u = g(d)$ to all valid depths in the patch. Significantly, we restrict our attention to depth values that lie within the predefined range $[d_{\min}, d_{\max}]$ and are marked as valid in the LiDAR mask; pixels outside this range or invalid returns (e.g., missing beams) are excluded before constructing the distribution. In this way, each patch p is represented by two comparable sets of transformed depth samples—one from the observed data and one from the prediction—defined over the same depth interval.

B.3 Soft histograms in the transformed domain

To obtain a smooth, differentiable approximation of the empirical distribution, we construct a soft histogram over the bin centers $\{u_i\}$ by placing a Gaussian kernel at each transformed depth value. For a given patch p, the unnormalized contribution to bin i is

$$h_p(i) = \sum_j w_{p,j} \exp\left(-\frac{1}{2}\left(\frac{u_{p,j} - u_i}{\sigma_u}\right)^2\right), \tag{B.4}$$

where $w_{p,j} \in \{0, 1\}$ is a validity weight indicating whether the j-th pixel in the patch contains a valid LiDAR return, and σ_u is the kernel bandwidth in the u-domain. We parameterize the bandwidth as

$$\sigma_u = \gamma\,\Delta u, \tag{B.5}$$

with a bandwidth multiplier $\gamma > 0$ (the *bandwidth* hyperparameter). This choice ties the amount of smoothing directly to the bin resolution, yielding stable, well-conditioned gradients. The resulting histogram $h_p \in \mathbb{R}^{N_{bins}}$ is normalized to obtain a discrete probability distribution:

$$P_p(i) = \frac{h_p(i)}{\sum_{k=1}^{N_{bins}} h_p(k)}, \quad i = 1, \ldots, N_{bins} \tag{B.6}$$

An analogous construction is applied to the observed patch to obtain $Q_p(i)$, corresponding to $\mu(\mathbf{y}_p)$.

B.4 CDF-based EMD in one dimension

Given the predicted and observed patch distributions P_p and Q_p on the same bin grid $\{u_i\}$, we compute their cumulative distribution functions (CDFs)

$$F_p(i) = \sum_{k=1}^{i} P_p(k), \qquad F_q(i) = \sum_{k=1}^{i} Q_p(k) \tag{B.7}$$

In one dimension, the Earth Mover's Distance (Wasserstein-1 distance) admits a simple expression in terms of these CDFs. Our patch-wise EMD is given by

$$\text{EMD}(\mu(\mathbf{y}_p), \mu(\mathcal{A}(\hat{\mathbf{x}}_0)_p)) := \sum_{i=1}^{N_{bins}} |F_p(i) - F_q(i)|\ \Delta u \tag{B.8}$$

This corresponds to the CDF–L1 formulation of the 1D EMD used in Eq. 16, up to the constant spacing factor Δu. Averaging this quantity over all valid patch centers yields the global EMD guidance:

$$\log p_{EMD}(\mathbf{y} \mid \hat{\mathbf{x}}_0) = -\lambda_{EMD} \sum_p \mathrm{EMD}\left(\mu(\mathbf{y}_p), \mu(\mathcal{A}(\hat{\mathbf{x}}_0)_p)\right), \tag{B.9}$$

where λ_{EMD} is the *guidance weight* controlling the overall strength of the correlative term.

B.5 Guidance schedule and hyperparameters

The final guidance term $\mathbf{g}_{corr}$ is obtained as the gradient of the surrogate log-likelihood (equivalently, the negative gradient of $\mathcal{L}_{EMD}$):

$$\mathbf{g}_{corr} = \nabla_{\mathbf{x}_t} \log p_{EMD}(\mathbf{y} \mid \hat{\mathbf{x}}_0) \tag{B.10}$$

This guidance is combined with the standard L2 term within the DPS framework (Chung et al., 2022a). In practice, we further modulate $\mathbf{g}_{corr}$ by a *guide scale* $s(t)$ that depends on the diffusion step t:

$$\tilde{\mathbf{g}}_{corr}(t) = s(t) \cdot \mathbf{g}_{corr} \tag{B.11}$$

In all experiments, we use the same simple linear guidance-scale schedule as described in the main paper: the scale starts from 1 and increases linearly to 10 over 128 sampling steps. We apply this schedule to the EMD Patch Guidance via the factor $s(t)$, so that the correlative term is weak at early diffusion steps and is gradually strengthened at later steps, primarily refining local depth structure once a coarse reconstruction has already been established.

In summary, the EMD Patch Guidance is governed by several interpretable hyperparameters:

- The guidance weight, which sets the overall strength of the distributional constraint.
- The guide scale schedule, which controls how strongly the guidance acts across diffusion steps.
- The patch size, which trades off local detail versus robustness to noise.
- The number of bins, which determines the resolution of the depth distributions.
- The bandwidth multiplier, which regulates the smoothness of the soft histograms.

These hyperparameters jointly shape how the EMD Patch Guidance enforces local depth-structure consistency and complements the pixel-wise L2 term in approximating the effect of the off-diagonal precision matrix $\mathbf{K}_{off}$.

C Can Depth Histograms Capture Spatial Correlation?

We further examine whether the Patch depth histogram encodes meaningful spatial correlation through a scene-wide spatial retrieval analysis. In the main framework, spatial consistency is enforced by minimizing the Patch EMD between predictions and measurements at the same pixel location. Here, instead, we evaluate the descriptor itself by retrieving geometrically similar pixels across different spatial locations.

To this end, we sample 81 scenes from the 4,071 validation frames of SemanticKITTI (Behley et al., 2019). Since consecutive frames in autonomous driving are highly redundant, we apply a temporal stride of 50 to obtain spatially diverse scenes. The central premise is simple: if the Patch depth histogram captures intrinsic local geometry, it should retrieve pixels with similar structural patterns even across diverse scenes. Therefore, successful retrieval of pixels from the same object instance or with similar surface topology provides direct evidence that the descriptor encodes spatial structure beyond point-wise depth. This analysis complements our main framework by showing that Patch EMD serves as a meaningful structural constraint rather than a simple depth-matching objective.

C.1 Baselines for Comparison

We compare EMD Patch with two baselines. *Random* reports the expected Precision@K ($P@K$) under uniform sampling from the valid retrieval pool, providing a lower bound that accounts for

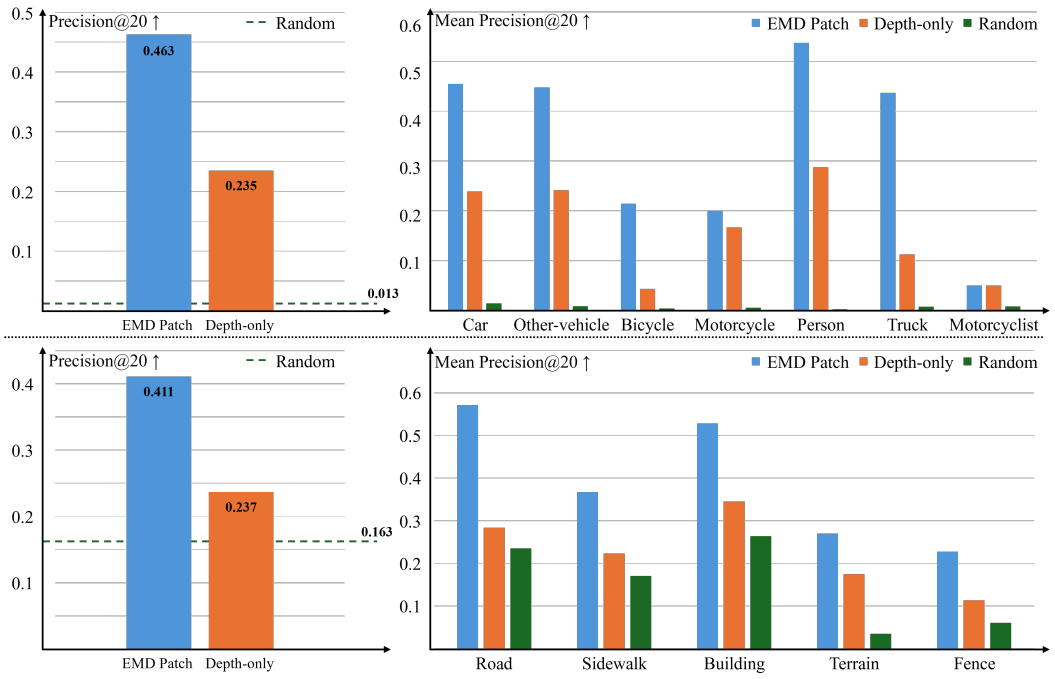

Figure C.1: Quantitative evaluation of spatial retrieval (Precision@20). The top row shows results for 'Things', and the bottom row shows results for 'Stuff' under a restricted ± 2.0m depth band. In both settings, EMD Patch consistently outperforms the *Depth-only* baseline, indicating that it captures local 3D structure beyond absolute depth similarity.

scene-wise class imbalance. *Depth-only* retrieves the Top-K pixels with the smallest absolute depth difference from the query. This is a particularly strong baseline in LiDAR range views, where similar-depth pixels often lie on the same iso-depth contour and frequently correspond to broad background regions. Hence, outperforming *Depth-only* indicates that EMD Patch captures local geometric context beyond absolute depth similarity.

C.2 Instance-level Spatial Correlation (Things)

We first evaluate whether EMD Patch can discriminate object-level structures such as cars, bicycles and pedestrians (person). For object instances, *Depth-only* retrieval is fundamentally limited: it tends to follow iso-depth contours across the scene and thus frequently mixes foreground objects with unrelated background regions at similar distance. In contrast, EMD Patch retrieves pixels that are substantially more concentrated within the queried instance.

As shown in the top row of Fig. C.1, EMD Patch achieves an aggregate $P@20$ of 0.463 across 81 scenes, compared with 0.235 for *Depth-only* and 0.013 for *Random*. The same trend holds across classes: for *Person*, EMD Patch reaches 0.5375 versus 0.2875 for *Depth-only*, and for *Car*, 0.4552 versus 0.2390. Qualitative examples in Fig. C.2 show the same behavior. For both *Car* and *Bicycle* queries, *Depth-only* spreads across unrelated regions with similar depth, whereas Patch EMD remains localized on the queried object. These results suggest that the Patch depth histogram captures instance-level structural cues that are not explained by absolute depth alone.

C.3 Surface Topology Discrimination (Stuff)

We next evaluate whether EMD Patch also captures surface topology for 'Stuff' classes, rather than acting as a depth proxy. To provide a favorable prior for *Depth-only* retrieval, we focus the search pool within a depth band (e.g., ± 2.0m), where absolute depth values are already closely aligned.

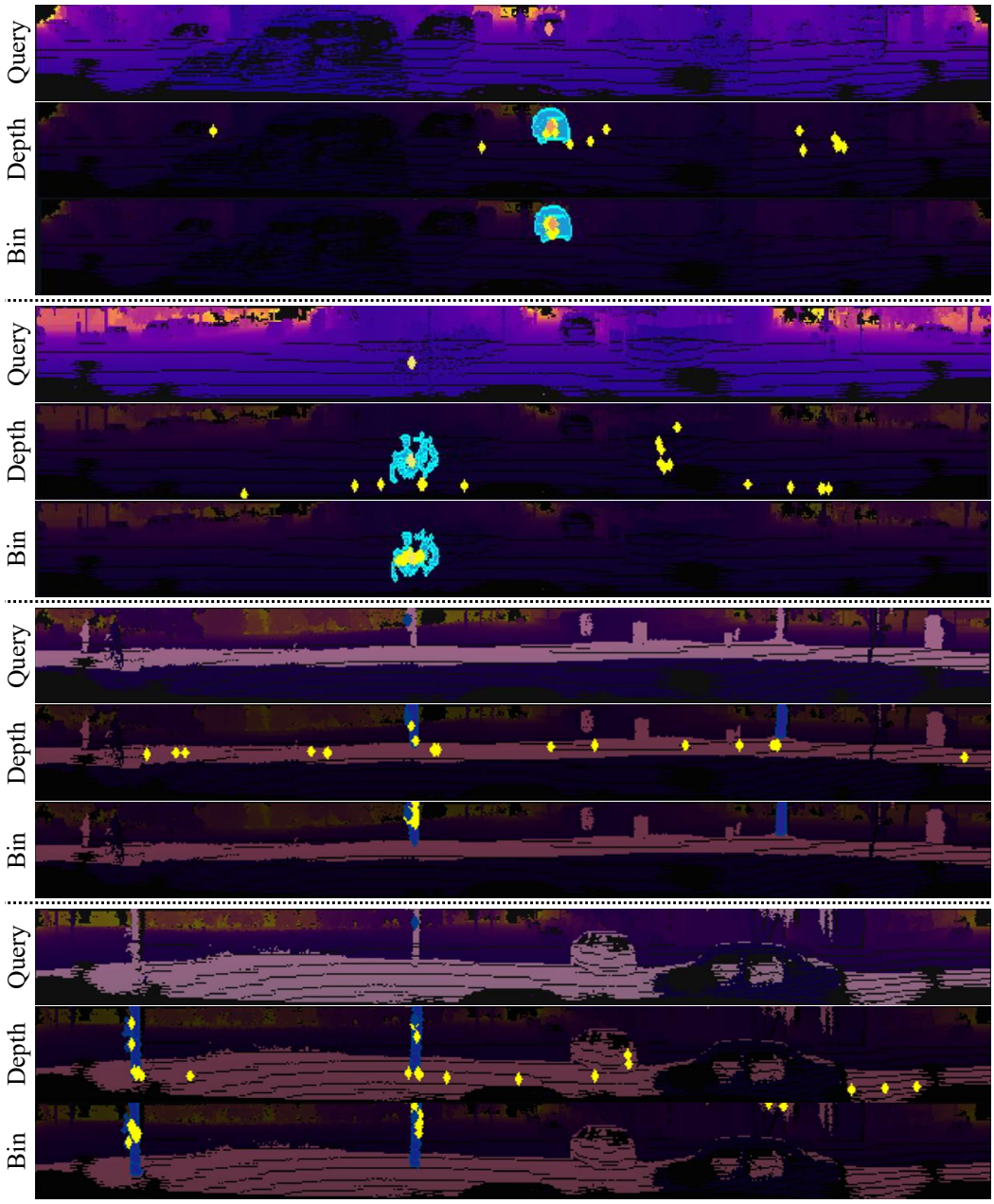

Figure C.2: Qualitative comparison of the Top-20 retrieval results for the query. From top to bottom, the rows illustrate queries for 'Things' (car, bicycle) and 'Stuff' (terrain, terrain) classes. *Depth-only* based retrieval searches points unrelated to the query, whereas EMD Patch histogram based retrieval searches points closely related to query.

Even under this constraint, EMD Patch maintains a clear advantage over *Depth-only*, as shown in the bottom row of Fig. C.1. It achieves an aggregate P@20 of 0.411, compared with 0.237 for *Depth-only*. The same tendency is observed across multiple classes, including road and building. Qualitative results in Fig. C.2 further illustrate this point: when querying *Terrain*, *Depth-only* is often distracted by nearby road regions with similar depth, while Patch EMD remains focused on terrain regions with similar local structure. This indicates that EMD Patch captures geometric patterns that allow it to distinguish structurally different regions even when their absolute depths are nearly identical.

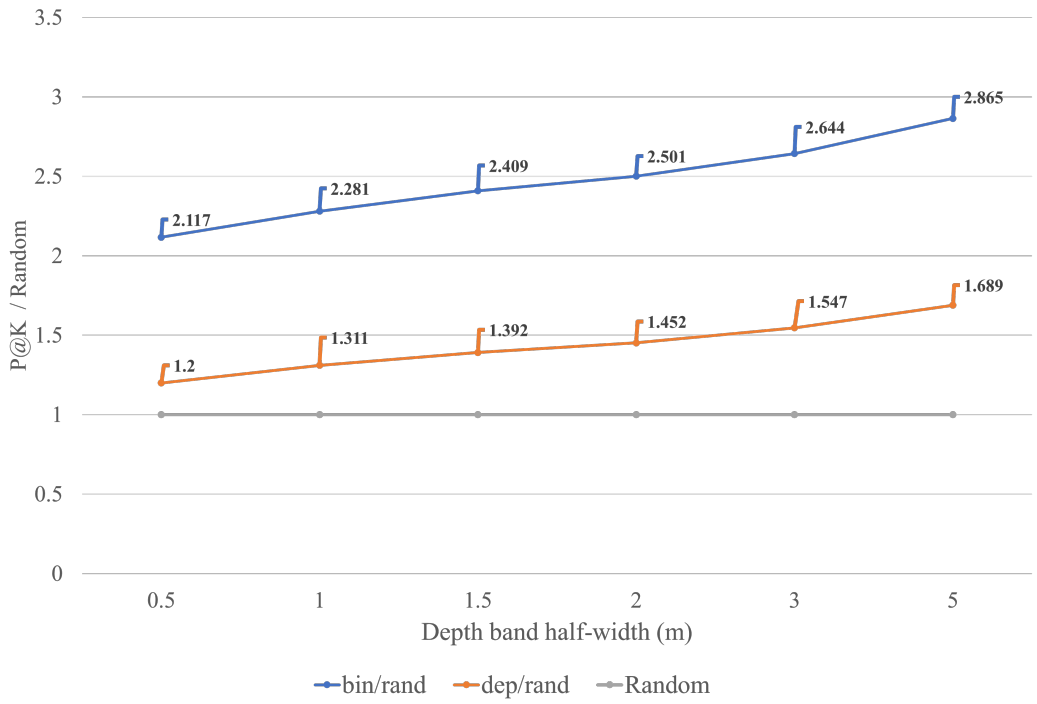

Figure C.3: Ablation study on varying depth-band constraints for 'Stuff' classes. The plot illustrates the retrieval performance ratio over the *Random* baseline across different depth-band half-widths (0.5m to 5.0m). As the band narrows, the *Depth-only* baseline degrades toward *Random*. Conversely, the EMD Patch depth histogram maintains a robust and consistent advantage, demonstrating that it effectively encodes surface topologies beyond absolute depth.

Ablation on Depth-Band Constraints (Stuff). To further validate this behavior, we vary the half-width of the depth band from 0.5m to 5.0m, as shown in Fig. C.3. As the band narrows, the performance of *Depth-only* drops toward the *Random* baseline, demonstrating that absolute depth becomes increasingly uninformative when retrieval candidates lie at nearly identical distances. In contrast, Patch EMD consistently outperforms *Depth-only* across all band widths, with an average margin of approximately 1.7×. This robustness across depth-band ranges indicates that the Patch depth histogram captures local geometric structure and surface topology, rather than relying on trivial iso-depth cues.

D Additional Implementation Details

D.1 Sampling Configurations

EMD Patch Guidance. The EMD Patch Guidance involves several hyperparameters. These include the guide scale, which controls how strongly the guidance term is applied across diffusion timesteps; the guidance weight; the patch size; the number of histogram bins; and the kernel bandwidth. As summarized in Table D.1, we specify these hyperparameters for each corruption type. By default, we use 32 histogram bins, a patch size of 3, and a kernel bandwidth of 0.5. Across different corruptions, we primarily vary the guidance weight, while the patch size and the guide-scale schedule are adjusted slightly for each corruption to better reflect each degradation's characteristics.

We note that the optimal guidance weight (λ_{EMD}) can vary across corruption types. This is expected and standard in gradient-based inverse solvers (Chung et al., 2022a), since different forward operators induce surrogate likelihood gradients with substantially different magnitudes and Lipschitz constants, making a single fixed scaling suboptimal or unstable. Accordingly, we tune the guidance strength per corruption via a coarse grid search on a small subset, which is sufficient in practice.

Table D.1: Corruption-specific hyperparameters for the proposed EMD Patch Guidance. The *Default* row shows the base configuration; "–" indicates that the default value is used.

Corruption	DPS Guide Scale	EMD Guide Scale	Weight	# Bins	Patch Size	Bandwidth
Default	1–10	1–10	300	32	3	0.5
Beam Missing	–	–	400	–	–	–
Cross Sensor	–	–	400	64	–	–
Crosstalk	–	–	–	–	–	–
Fog	–	–	120	–	–	–
Incomplete Echo	–	–	–	–	–	–
Motion Blur	1–3	–	2500	–	–	–
Snow	–	–	–	–	–	–
Wet Ground	–	–	150	–	–	–

Apart from the corruption-specific guidance weight, the only non-default setting is the guidance-scale range for Motion Blur. We found DPS guidance to be most stable when restricted to 1–3 for this corruption, and align the EMD Patch Guidance with the same schedule.

Diffusion Configuration. We configured our diffusion model to follow the training methodology outlined in R2DM (Nakashima & Kurazume, 2024). This approach employs a Variational Diffusion Model (Kingma et al., 2021), which is trained using an Efficient-UNet architecture (Saharia et al., 2022). Consistent with this approach, our pretrained model was trained to synthesize the range view depth map and range view reflectance data. For optimization, we used AdamW (Loshchilov & Hutter, 2017) with a learning rate $lr = 10^{-4}$, and momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The model was trained up to 300K steps, and we set the diffusion sampling steps to 128.

D.2 Baselines

Exclusion of ADMM-TV, DDRM, and DiffusionMBIR. In this section, we provide the rationale for excluding common inverse problem-solving baselines, such as DDRM (Kawar et al., 2022), DiffusionMBIR (Chung et al., 2023b), or the classical ADMM-TV. These methods fundamentally rely on the assumption that the forward measurement process is strictly linear.

Specifically, DDRM (Kawar et al., 2022) requires the Singular Value Decomposition (SVD) of the linear measurement matrix $\mathcal{A}$ to perform exact spectral filtering during the reverse diffusion process. Similarly, ADMM-TV and DiffusionMBIR (Chung et al., 2023b) formulate the data-consistency step as a TV-regularized least-squares optimization:

$$\min_x \frac{1}{2} \|y - \mathcal{A}x\|_2^2 + \lambda \|Dx\|_1 \tag{D.1}$$

Efficiently solving this step, typically via Conjugate Gradient (CG) or closed-form ADMM updates, strictly requires the forward operator $\mathcal{A}$ to be a linear transform.

However, LiDAR degradations are inherently non-linear. The mapping process involves complex geometric operations such as z-buffering, which cannot be formulated as a standard matrix multiplication. Because $\mathcal{A}$ cannot be expressed as a linear static matrix, computing the SVD for DDRM is mathematically impossible in this context. Furthermore, the linear ADMM updates and CG optimization required by ADMM-TV and DiffusionMBIR become entirely invalid. Moreover, while DiffusionMBIR regularizes 3D volumes by aggregating sliced 2D diffusion priors, adapting this stacking-based regularization is fundamentally incompatible with our framework, which natively operates on a single 2D range view. Therefore, we exclude these linear-model-based solvers from our baselines.

DPS+TV. Inspired by DiffusionMBIR (Chung et al., 2023b), we augment Diffusion Posterior Sampling (DPS) (Chung et al., 2022a) with a Total Variation (TV) regularization term. Since the CG-based data-consistency step of DiffusionMBIR is not applicable to all of our operators, we replace it with gradient descent and apply TV along the in-plane range-view coordinates, which yields a practical surrogate that explicitly optimizes both the data-fidelity and TV penalties within the diffusion

Table E.1: Quantitative comparison with heuristic filters and a supervised baseline under the *Moderate, Snow* corruption.

	Point Cloud	Image	BEV		Depth	Reflectance
Method	FPD ↓	FRD ↓	JSD $\times 10^2$ ↓	MMD $\times 10^4$ ↓	MAE ↓	MAE $\times 10^1$ ↓
ROR	1350	321.5	7.327	1.488	0.917	0.410
SOR	2480	330.1	7.366	1.470	0.946	0.412
CNN	338.7	974.7	3.543	0.581	0.620	0.928
Ours	**17.14**	**157.8**	**1.537**	**0.201**	**0.376**	**0.329**

sampling loop:

$$x_{t-1} \leftarrow \text{DiffusionReverseStep}(x_t) - \zeta_1 \nabla_{\mathbf{x}_t} \left\| y - \mathcal{A}(\hat{x}_0) \right\|_2^2 - \zeta_2 \nabla_{\mathbf{x}_t} \left\| D\hat{x}_0 \right\|_1 , \tag{D.2}$$

where $\hat{x}_0$ is the predicted clean range view at timestep t, and ζ_1, ζ_2 are the scaling factors for the measurement and TV guidance, respectively. Here, D denotes the discrete spatial gradient operator, computing finite differences along the 2D spatial coordinates (horizontal and vertical) of the range-view representation to enforce local smoothness (implemented via a smooth Charbonnier approximation for differentiability).

E Additional Experiments

E.1 Cost–Performance Trade-off Analysis

While diffusion models solve the inverse problem robustly, their iterative sampling inherently incurs higher inference costs compared to supervised regression models or heuristic filters. Sampling takes 17.7 s per frame on a single NVIDIA RTX 3090, against 17.6 s for DPS and 18.0 s for DPS+TV, so the correlative guidance itself accounts for under 1% of this cost. To justify this cost-performance trade-off, we evaluate our framework against a supervised regression baseline and standard point-cloud filters: Radius Outlier Removal (ROR) and Statistical Outlier Removal (SOR). Importantly, a strictly fair comparison between these approaches is fundamentally impossible. Our method solves the *inverse problem* in a *zero-shot* manner using the known forward operator $\mathcal{A}$, without any corruption-specific training. In contrast, to provide a comparative baseline, the supervised model is explicitly trained on the corrupted dataset for **two days** on a single NVIDIA RTX 3090, sharing the exact same EfficientUNet-34 backbone as our pretrained model.

We specifically highlight the *Snow* corruption, as its highly stochastic and spatially irregular noise patterns effectively expose the fundamental limitations of both heuristic and supervised approaches. As shown in Tab. E.1, heuristic filters (ROR, SOR) are computationally cheap but completely fail to handle such complex distributions, resulting in extremely high FPD scores. Furthermore, while the supervised baseline (CNN) benefits from extensive corruption-specific training and fast single-pass inference, it completely fails to preserve the underlying physical 3D point distribution under this stochastic corruption, leading to severe degradation in geometry-sensitive metrics (e.g., FPD, JSD, and MMD). In contrast, our zero-shot diffusion framework successfully maintains 3D structural fidelity. These results indicate that the inference overhead is a necessary trade-off: at roughly 18 s per frame and no corruption-specific training, our framework attains a distributional fidelity that neither the filters nor the two-day-trained regressor reaches.

E.2 Mixed Corruptions

In real-world scenarios, LiDAR sensors frequently encounter multiple overlapping corruptions simultaneously, such as concurrent adverse weather and sensor malfunctions. Unlike supervised baselines that require combinatorial retraining for every possible mixture, our zero-shot formulation elegantly handles these complex situations. Because we explicitly model the physical degradation through the forward operator $\mathcal{A}$, we can naturally represent mixed corruptions via mathematical composition. For example, a joint degradation of beam missing and fog can be strictly defined as a composite operator $\mathcal{A}_{mixed} = \mathcal{A}_{beam} \circ \mathcal{A}_{fog}$. As illustrated in Fig. E.1, our framework effectively solves the inverse problem using this composite operator, successfully recovering the 3D scene in its

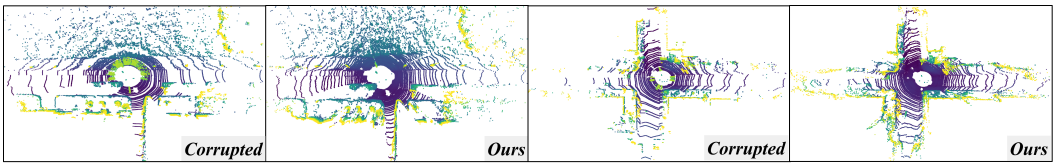

Figure E.1: Comparison of point cloud reconstruction under mixed corruption: *Beam Missing*, *Fog*.

Table E.2: Additional Ablation Studies of parameters for EMD Patch Guidance (Ours) under the *Moderate, Beam Missing* condition on SemanticKITTI.

		Point Cloud	Image	BEV		Depth	Reflectance
Experiment	Method	FPD ↓	FRD ↓	JSD $\times 10^2$ ↓	MMD $\times 10^4$ ↓	MAE ↓	MAE $\times 10^1$ ↓
EMD Loss Patch Size	Patch Size 3	6.402	**302.8**	**1.654**	**0.286**	0.576	**0.385**
	Patch Size 5	6.075	304.3	1.674	0.290	0.582	**0.385**
	Patch Size 7	**5.810**	304.6	1.694	0.293	0.585	**0.385**
Bin Domain Transformation	Log	6.402	**302.8**	**1.654**	**0.286**	**0.576**	0.385
	Linear	**5.384**	311.1	1.693	0.302	0.578	0.386
	Inverse	8.882	330.8	2.605	0.958	0.667	**0.384**

clean state without any model modification. This highlights the profound flexibility of our approach in addressing combinatorial real-world degradations.

E.3 Additional Ablation Studies

EMD Patch Size. We investigate the effect of varying the patch size in the EMD in Tab. E.2. Reducing the patch size consistently improves distributional metrics such as FRD, JSD, and MMD, indicating stronger local geometric consistency and better alignment in the BEV representation. In contrast, increasing the patch size yields slightly lower FPD, suggesting that larger patches emphasize global structural coherence at the cost of fine-grained fidelity. These results demonstrate that the patch size controls a trade-off between local detail preservation and global smoothness.

Bin Domain Transformation. We compare *linear*, *inverse*, and *log* bin domains in Tab. E.2, where the *log* domain performs best on most metrics. Although the *linear* domain achieves a lower FPD, the *log* domain consistently outperforms it across the remaining depth, image, and BEV metrics. Moreover, the *log* transformation reduces scale variations between near- and far-range depths, better accommodating the intrinsic distance-dependent sparsity of LiDAR measurements. We therefore adopt the *log* domain for its balanced reconstruction quality across multiple evaluation metrics.

F Details of Corruption Map

This section provides additional details on how to store specific 3D corruption $\mathbf{c}$ to corruption map $\mathbf{C}$ and the auxiliary information they record for DPS inference. All operators closely follow the corruption behavior defined in the official Robo3D (Kong et al., 2023) implementation, and we preserve the corresponding metadata required to reconstruct the observation. For all corruption types except *Motion Blur*, the corrupted point and its clean counterpart project to the same range-view pixel; thus, the recorded metadata is indexed by the pixel location of the corrupted point. In the case of *Beam Missing*, we only record the IDs of the removed scan rings because the corruption does not alter individual point attributes.

Crucially, while we utilize exact realizations to align with the DPS benchmarking protocol, we emphasize that the inversion remains fundamentally ill-posed. Since 3D corruptions involve non-linear projection operations (e.g., Z-buffering), **the inverse problem cannot be resolved by simple arithmetic subtraction.**

Beam Missing. We simulate missing scan lines by randomly dropping a set of LiDAR rings according to the severity level. For each point, we compute its ring index and remove all points whose indices belong to the sampled subset of "dropped" rings. The corrupted range view is then obtained by projecting only the remaining points. The forward operator records the IDs of the removed rings.

During the DPS update, the measurement operator is applied as a pure masking onto the observed support, with the keep-mask derived from the validity channel of the recorded corrupted range view so that $\mathcal{A}(\hat{x}_0)$ and the measurement share the same support.

Cross Sensor. We emulate cross-sensor mismatch by converting the original 64-beam LiDAR into an effectively lower-resolution sensor. For each severity level, we deterministically select a subset of scan rings to drop and remove all points belonging to those rings, mimicking a LiDAR with fewer vertical beams. From the remaining points, we further sub-sample along the scan order to reduce horizontal density before projecting them to the range view. During this process, the forward operator records the IDs of the dropped rings and a validity mask.

Crosstalk. We model LiDAR crosstalk as sparse, localized perturbations of individual returns. A small percentage of points is randomly selected based on severity level, and Gaussian noise is added to their 3D coordinates and intensities before projection. The corrupted range view is obtained from these perturbed points. During recording, we build a noise field by mapping the point-wise offsets to their range-view locations and store it together with a validity mask.

Fog. We reproduce the Robo3D fog corruption by directly using the integral lookup tables and parameter settings provided in the official implementation. For each severity level, we load the corresponding precomputed integral table and extinction parameters, then simulate fog effects following the original Robo3D logic. Each point undergoes distance-dependent attenuation of its return intensity ("hard" region), while points whose fog backscatter exceeds the attenuated return ("soft" region) additionally receive range scaling and stochastic range noise, also consistent with the Robo3D implementation. The augmented points are then projected to the range view. During recording, the operator stores hard-attenuation factors, fog-response ratios, range-scaling factors, a soft-region mask, the fog parameter α, and a validity mask, enabling the reapplication of the transformation.

Incomplete Echo. Using the provided SemanticKITTI learning map, we first convert fine-grained labels into coarse classes and identify a subset of classes for which echoes are likely to be missing. For each such class, we randomly drop a severity-dependent fraction of points, mimicking partial returns along the ray. The remaining points are projected to the range view to form the corrupted observation. The forward operator records the binary mask indicating where echoes are still present.

Motion Blur. Motion blur is modeled as a misalignment between clean and corrupted range views caused by sensor/ego motion during a scan. We first add a global 3D translation and small, axis-dependent jitter to all points, then project the perturbed points to obtain the corrupted range view and its z-buffer selection. For each selected point, we compute its original clean range-view pixel and record the corresponding 3D displacement (translation + jitter) on a clean-pixel grid, together with a correspondence mask and the clean xyz grid. This yields a displacement field defined in the clean coordinate system, which is later used to apply a differentiable re-projection–based blur operator to the clean prediction in DPS.

Snow. We adopt the official Robo3D snow simulator to model depth- and class-dependent snow attenuation. Using the provided SemanticKITTI learning map and sensor calibration, we estimate a distance-dependent noise threshold and simulate snow particles per scan channel with severity-dependent snowfall rate and terminal velocity. The simulator then decides, for each point, whether snow only attenuates its intensity or fully replaces its 3D coordinates and intensity with a snow-affected echo, followed by thresholding to remove low-SNR returns. The resulting augmented points are projected to the range view. During recording, the forward operator stores scaling factors (xyz), target intensities, and masks indicating original, removed, and changed pixels (intensity-only vs. full xyzi changes), allowing the application of the snow transformation.

Wet Ground. We re-implement the Robo3D wet-ground corruption in 3D using its helper functions and semantic mapping. Using the SemanticKITTI learning map, we first identify ground-like classes (road, parking, sidewalk, other ground) and restrict the water effect to those points. For a given severity, we set the water height and noise floor, then estimate a distance-dependent laser response and adaptive noise threshold from ground returns. With the ground transmittance function provided by Robo3D, we compute an effective wet-ground transmittance and derive new intensities for ground points, clamping them below the original intensity and dropping points whose wet intensity falls below the adaptive threshold. Non-ground points remain unchanged. The augmented point cloud is projected to the range view, and the forward operator records a reflectance scaling factor and a validity mask.

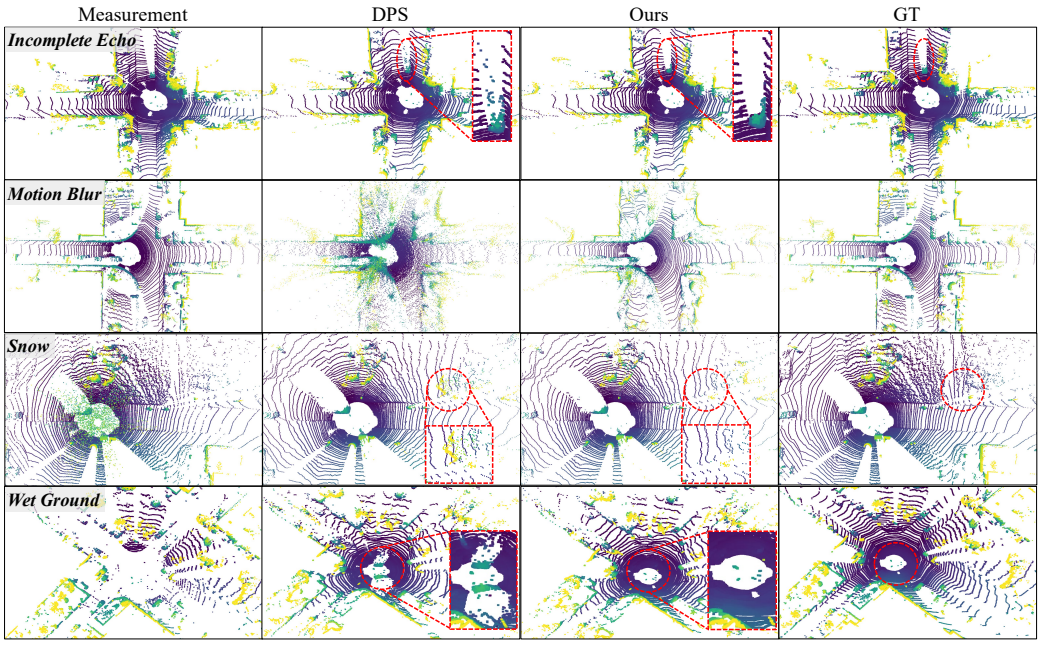

Figure G.1: Qualitative comparison of point cloud reconstruction under four corruptions: *Incomplete Echo*, *Motion Blur*, *Snow*, *Wet Ground*.

G Additional Qualitative Results

In this section, we present additional qualitative results and analyses on SemanticKITTI for corruption types that could not be included in the main paper due to space constraints, namely *Incomplete Echo*, *Motion Blur*, *Snow*, and *Wet Ground*. In addition, for the KITTI-360 dataset, we provide qualitative examples and analyses under the *Beam Missing* corruption.

G.1 SemanticKITTI

Figure G.1 presents the qualitative comparison results across four specific corruption types. For the *Incomplete Echo* corruption, DPS struggles to reconstruct the corrupted regions, particularly on vehicle surfaces, often generating residual noise or artifacts. In contrast, our method achieves clean and accurate reconstruction. Under *Motion Blur* corruption, DPS approach yields overall noisy and blurred reconstructions across the entire scene, whereas our proposed method produces natural and coherent results. Regarding the *Snow* corruption, DPS fails to fully denoise the individual snow particles and artifacts. In contrast, our framework effectively removes the snow while preserving the crispness of the underlying geometry. Finally, in the *Wet Ground* corruption, DPS exhibits poor ground plane reconstruction near the LiDAR. Our method successfully solves the inverse problem to recover a complete and natural ground surface. Collectively, these results indicate that our method provides more robust and geometrically consistent reconstruction across all corruption types.

G.2 KITTI-360

Figure G.2 presents the comparison results under the *Moderate, Beam Missing* corruption. The DPS tends to generate incoherent geometry or spurious noise particles during the reconstruction process. In contrast, our method achieves seamless beam completion. These results showcase the effectiveness and robust generalizability of our framework across various domains.

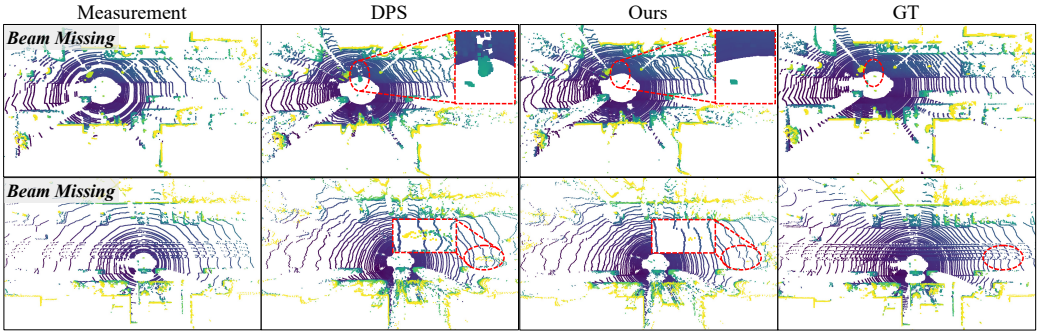

Figure G.2: Qualitative comparison of point cloud reconstruction on KITTI-360 under the *Beam Missing* corruption.