Structure-to-Intensity Diffusion for Adverse-Weather LiDAR Generation

Supplementary Material

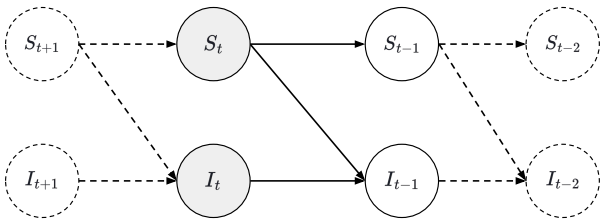


Figure 1. Probabilistic graphical model of the SiD reverse process. The graph visualizes the estimation of S_{t-1} depends only on S_t, not on I_t, reflecting physical causality.

A. Rationale

This section provides a detailed justification for the factorization of the reverse process transition used in our Structure-to-Intensity Diffusion (SiD). Recall that in the main paper, we decompose the reverse transition of the joint LiDAR generation process at timestep t as:

$$\begin{aligned} &p_\theta(S_{t-1},I_{t-1} \mid S_t, I_t, c) \\ &\quad = p_\theta(S_{t-1} \mid S_t, c) \cdot p_\theta(I_{t-1} \mid I_t, S_t, c). \end{aligned} \tag{1}$$

This decomposition is derived by applying two principled conditional independence assumptions to the exact chain rule of probability. Our derivation begins with the standard chain rule, which is always true and makes no assumptions:

$$\begin{aligned} &p_\theta(S_{t-1},I_{t-1} \mid S_t, I_t, c) \\ &\quad = p_\theta(S_{t-1} \mid S_t, I_t, c) \cdot p_\theta(I_{t-1} \mid S_{t-1}, S_t, I_t, c). \end{aligned} \tag{2}$$

We now simplify each of the two terms on the right-hand side.

Considering the asymmetric dependency in LiDAR sensing, where intensity I is strongly influenced by geometry S yet inferring geometry from intensity is fundamentally ill-posed due to ambiguity and noise, we treat intensity as explicitly conditioned on geometry. This implies that, given the precise current geometry S_t, the current intensity I_t provides no additional information for estimating the previous geometry S_{t-1}. This justifies the assumption $(S_{t-1} \perp I_t) \mid (S_t, c)$, which simplifies the first term:

$$p_\theta(S_{t-1} \mid S_t, I_t, c) = p_\theta(S_{t-1} \mid S_t, c). \tag{3}$$

We rely on the standard Markov assumption of diffusion models, which states that the state at $t-1$ depends only on the state at t. Therefore, when predicting I_{t-1} given the full current state (S_t, I_t), the other variable S_{t-1} from the previous step is redundant. This implies $(I_{t-1} \perp S_{t-1}) \mid (S_t, I_t, c)$, simplifying the second term:

$$p_\theta(I_{t-1} \mid S_{t-1}, S_t, I_t, c) = p_\theta(I_{t-1} \mid S_t, I_t, c). \tag{4}$$

Finally, by substituting Equation (3) and Equation (4) into the chain rule in Equation (2), we obtain the factorized expression in Equation (1), which makes the resulting generative process consistent with the physical and causal structure of LiDAR data.

Additionally, we further substantiate this dependency structure using a causal inference perspective [2, 3]. As illustrated in Figure 1, our model respects the physical asymmetry of LiDAR sensing: the estimation of S_{t-1} depends exclusively on S_t, reflecting that structural evolution is independent of instantaneous intensity.

B. Detailed Architecture

Our model couples a denoising backbone with a condition pathway to disentangle scene structure from weather factors and refine intensity reflectance predictions across stages. The backbone provides a hierarchical representation with skip fusion, while the condition pathway distills external cues and stage-wise priors into a compact modulation signal. Positional context and diffusion progress are treated as auxiliary information and fused into the same conditioning stream, steering feature transformations without disrupting locality. To complement local processing with long-range dependencies at modest cost, Vision Mamba [7, 8] layers are placed at semantically deep resolutions where global context is most beneficial.

Denoising Network. The detailed architecture of SiD backbone is illustrated in Figure 2. The backbone follows a symmetric encoder–decoder that progressively compresses and reconstructs multi-scale features with lateral connections preserving fine geometry. Conditioning signals, comprising weather semantics, stage context, and temporal progress, are fused into a latent vector that modulates normalization throughout the network, aligning representation updates with the target weather while maintaining consistent structure.

Condition Encoder. As depicted in Figure 3, we adopt a light-weight yet efficient architecture to embed weather context and geometry condition. The condition pathway mirrors the encoding motif in a compact form: it summarizes the previous stage's clean estimate (coupled with positional cues) into a refinement embedding that shares the

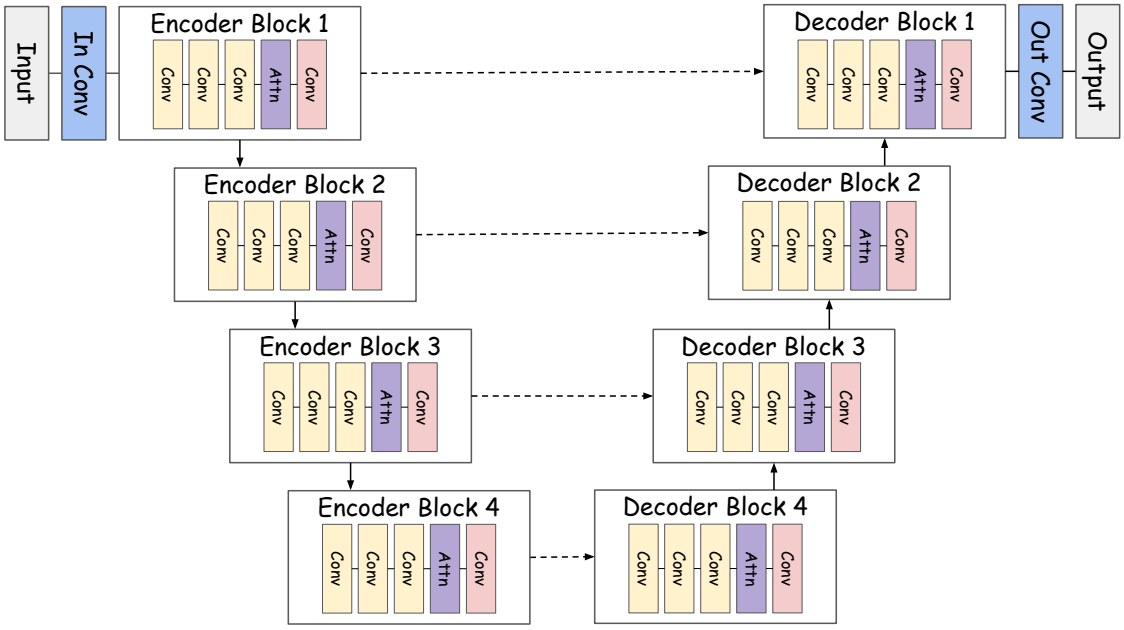


Figure 2. Network architecture of our denoiser backbone.

modulation interface with the backbone. Sequential convolutions enhance semantic abstraction, and intermittent Vision Mamba operations inject global coupling without inflating complexity. The pathway is initialized to exert negligible early influence, allowing the denoising backbone to stabilize; its contribution then grows to provide stage-aware guidance. The resulting refinement embedding is additively merged with the weather-derived conditioning, yielding a unified signal that drives consistent, progressively sharper reconstructions across stages.

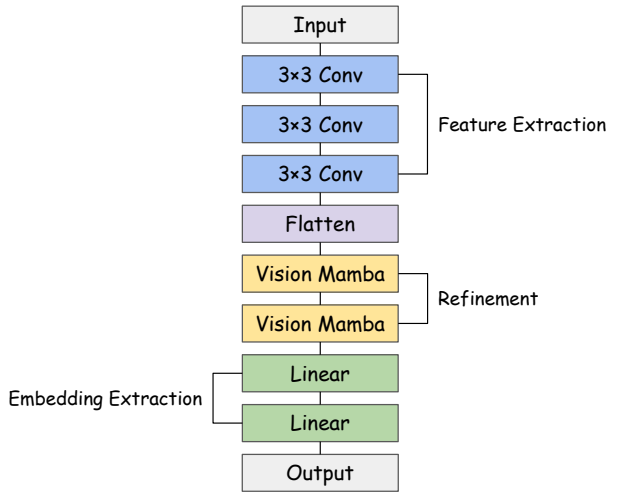


Figure 3. Network architecture of our condition encoder.

C. Design of Fréchet Minkowski Distance

The Fréchet Minkowski Distance (FMD) is designed to evaluate the similarity between real and generated point cloud features under adverse weather. Traditional metrics such as FPD and FRD depend on feature extractors (*e.g.*, PointNet++ [5] and RangeNet++ [4]) trained exclusively on clear-weather data and therefore exhibit large domain gaps when the input distribution shifts. In contrast, FMD leverages a weather-robust feature backbone to provide a more faithful and reliable comparison.

We adopt the Minkowski UNet trained on the STF [1] dataset via the SLidR framework [6] as the FMD feature extractor. By leveraging pixel-to-point distillation, SLidR transfers rich semantic priors from large 2D image encoders to 3D point clouds, yielding voxel features with exceptional stability and generalization. Crucially, grounding this cross-modal distillation in real-world adverse weather data renders the metric highly robust to environmental corruption, ensuring a consistent and reliable evaluation of both geometric integrity and reflectance patterns.

Given a LiDAR scan, SLidR outputs a D-dimensional feature vector for each point. These features jointly encode geometry and reflectance cues, providing a unified representation suitable for evaluating both structural fidelity and intensity consistency. To form a compact scene-level descriptor, we aggregate all point-level features within a scan using their mean and standard deviation. This pooling captures both the central tendency and intra-frame variability, which is important for weather-altered, spatially heteroge-

neous scans.

Let $(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$ and $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ denote the empirical mean and covariance of pooled features computed over real and generated scans, respectively. The Fréchet Minkowski Distance is then defined as:

$$\begin{aligned} \mathrm{FMD} = &\|\boldsymbol{\mu}_r - \boldsymbol{\mu}_g\|_2^2 \\ &+\mathrm{Tr}\Big(\boldsymbol{\Sigma}_r + \boldsymbol{\Sigma}_g - 2(\boldsymbol{\Sigma}_r\boldsymbol{\Sigma}_g)^{1/2}\Big)\,. \end{aligned} \tag{5}$$

By grounding the evaluation in weather-aware features and comparing full second-order statistics, FMD provides a stable and discriminative metric for assessing generative models in adverse-weather LiDAR reconstruction.

D. Study on RPWS

In this section, we conduct a further investigation into the effectiveness of our proposed Real-Prior Weather Simulation (RPWS). We comprehensively analyze the impact of adverse weather data simulation across distinct adverse weather conditions (*e.g.*, fog, rain, and snow), examining performance during both the data augmentation pre-training stage and the subsequent real-data fine-tuning stage. We benchmark our RPWS against the map-based data producer (MDP) augmentation strategy used in the baseline WeatherGen [8], while employing the exact same denoiser backbone for all experiments.

We first evaluate the generation quality in the pre-training stage, where models are trained solely on simulated data without access to real-world weather samples. As presented in Table 1, Table 2, and Table 3, our RPWS method consistently outperforms the MDP across the majority of metrics in all weather scenarios. Notably, in the fog test set shown in Table 1, our method achieves a substantial improvement in generation fidelity, with FPD and FRD scores significantly lower than the reproduced WeatherGen results. Similar advantages are observed in rain and snow conditions. This demonstrates that by incorporating physical priors, RPWS generates geometry and intensity degradations that are conceptually and statistically closer to the real-world distribution, offering a higher-quality starting point for the model.

Furthermore, it is crucial to evaluate the performance after fine-tuning, as the quality of the augmented data serves as a structural prior and provides the model with a critical starting point for the subsequent learning process. We validate this impact on real-world adverse-weather data, with results reported in Table 4, Table 5, and Table 6. Our method maintains a clear performance advantage over the baseline, proving that a better initialization leads to superior final results. For instance, in the challenging snow scenario (Table 6), our method achieves the lowest scores across Point Cloud and Image metrics, significantly surpassing both the originally reported and our reproduced metrics of WeatherGen. This consistent superiority indicates that RPWS not only serves as an effective data augmenter but also facilitates better model convergence, enabling the network to more accurately capture the complex characteristics of diverse weather phenomena.

Table 1. Quantitative comparison of weather simulation methods before fine-tuning. Metrics are evaluated on the STF [1] fog test set. Results with † indicate our reproduction.

Method	Point Cloud		Image	Bird's-Eye View	
	FMD ↓	FPD ↓	FRD ↓	MMD $\times 10^4$ ↓	JSD $\times 10^1$ ↓
WxGen† [8]	4.34	321.64	2722.71	15.97	2.26
Ours	**4.19**	**267.29**	**2324.00**	**15.29**	**2.14**

Table 2. Quantitative comparison of weather simulation methods before fine-tuning. Metrics are evaluated on the STF [1] rain test set. Results with † indicate our reproduction.

Method	Point Cloud		Image	Bird's-Eye View	
	FMD ↓	FPD ↓	FRD ↓	MMD $\times 10^4$ ↓	JSD $\times 10^1$ ↓
WxGen† [8]	8.56	292.46	2940.84	**10.23**	2.20
Ours	**6.40**	**139.33**	**2240.28**	11.85	**1.93**

Table 3. Quantitative comparison of weather simulation methods before fine-tuning. Metrics are evaluated on the STF [1] snow test set. Results with † indicate our reproduction.

Method	Point Cloud		Image	Bird's-Eye View	
	FMD ↓	FPD ↓	FRD ↓	MMD $\times 10^4$ ↓	JSD $\times 10^1$ ↓
WxGen [8]	-	221.73	1887.51	7.85	**1.59**
WxGen† [8]	6.11	271.12	2231.16	10.53	1.80
Ours	**3.80**	**73.64**	**1678.36**	**6.19**	1.84

Table 4. Quantitative comparison of weather simulation methods after fine-tuning on real-world adverse-weather data. Metrics are evaluated on the STF [1] fog test set. Results with † indicate our reproduction.

Method	Point Cloud		Image	Bird's-Eye View	
	FMD ↓	FPD ↓	FRD ↓	MMD $\times 10^4$ ↓	JSD $\times 10^1$ ↓
WxGen [8]	-	314.14	1968.90	8.08	2.66
WxGen† [8]	3.29	374.77	1955.05	5.96	2.25
Ours	**2.42**	**165.78**	**1137.46**	**3.41**	**1.88**

E. More Visualization Results

In this section, we present additional visualization results to further validate the effectiveness of our approach.

Table 5. Quantitative comparison of weather simulation methods after fine-tuning on real-world adverse-weather data. Metrics are evaluated on the STF [1] rain test set. Results with † indicate our reproduction.

Method	Point Cloud		Image	Bird's-Eye View	
	FMD ↓	FPD ↓	FRD ↓	MMD $\times 10^4$ ↓	JSD $\times 10^1$ ↓
WxGen [8]	-	86.40	1270.60	4.15	**0.93**
WxGen† [8]	4.65	132.81	1269.27	3.68	0.97
Ours	**4.24**	**64.97**	**1112.20**	**2.83**	1.11

Table 6. Quantitative comparison of weather simulation methods after fine-tuning on real-world adverse-weather data. Metrics are evaluated on the STF [1] snow test set. Results with † indicate our reproduction.

Method	Point Cloud		Image	Bird's-Eye View	
	FMD ↓	FPD ↓	FRD ↓	MMD $\times 10^4$ ↓	JSD $\times 10^1$ ↓
WxGen [8]	-	59.28	1241.66	1.71	0.77
WxGen† [8]	3.39	59.96	1295.31	2.28	0.87
Ours	**2.72**	**35.55**	**1201.89**	**1.69**	**0.63**

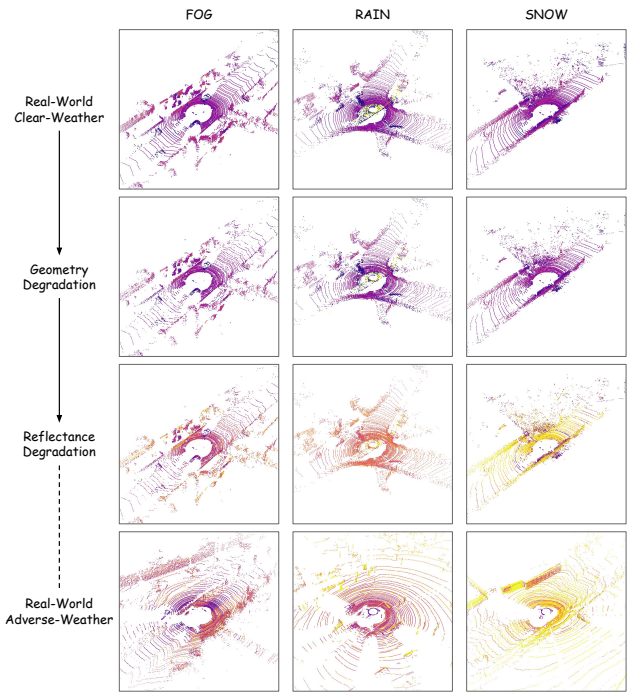


Figure 4. Visualization of RPWS process under different adverse weather conditions.

Visualization of RPWS Process. As illustrated in the main paper, our RPWS framework operates in two sequential stages: geometry degradation and reflectance degradation. We provide a step-by-step visualization of the point cloud evolution through this pipeline across different weather conditions in Figure 4.

Compared to the dense measurements of the original clear-weather input, the point cloud resulting from the geometry degradation stage becomes notably sparser and exhibits increased noise intensity. Subsequently, the reflectance degradation stage further modulates the intensity signal, ensuring that the values and distribution of the reflectance intensity align closely with the sensing characteristics observed in real-world adverse weather point clouds.

t-SNE Feature Distribution Alignment. We visualize the feature distributions of real and generated point clouds using t-SNE embeddings in Figure 5. The visualization covers three distinct weather conditions: fog, rain, and snow. As observed, the clear weather data (blue triangles) form a distinct cluster significantly separated from the real adverse weather data (red crosses), highlighting the substantial domain gap. The RPWS simulation (green pluses) occupies a separate region, serving as a physics-based initialization. Crucially, the features of our generated point clouds, both without fine-tuning (orange squares) and with fine-tuning (purple circles), exhibit a strong distributional overlap with the real adverse data. This alignment confirms that SiD effectively bridges the domain gap, synthesizing point cloud features that are semantically and structurally consistent with real-world sensor data under challenging conditions, rather than merely memorizing the clear-weather distribution.

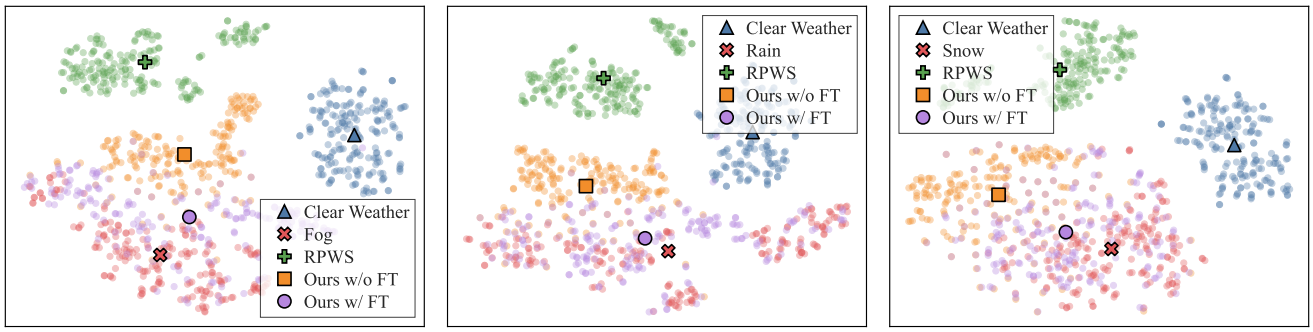
Qualitative Comparisons. We present additional qualitative results of generated point clouds under varying adverse weather conditions and provide a comparison with the STF [1] ground truth, as shown in Figure 6 (fog), Figure 7 (rain), and Figure 8 (snow).

Our model generates point clouds that exhibit high structural realism and coherence. Furthermore, the intensity values of our generated scans are closely aligned with the reflectance distribution observed in real-world data.

F. Limitations

While our proposed method demonstrates strong performance, we acknowledge certain limitations regarding fine-grained fidelity. Although the model captures global structure effectively, the generated scans may occasionally exhibit slight irregularities in the concentric ring patterns typical of mechanical LiDAR, and the boundaries of small or distant objects can appear slightly smoothed.

Additionally, our current framework is designed for static, single-frame generation. While this formulation successfully addresses data augmentation needs, it does not ex-



(a) t-SNE Distribution under Foggy Condition (b) t-SNE Distribution under Rainy Condition (c) t-SNE Distribution on Snowy Condition

Figure 5. t-SNE visualizations of feature distributions across different adverse weather conditions. The plots show the alignment between real clear-weather data (blue), real adverse-weather data (red), RPWS simulation results (green), and our generated results (orange and purple).

plicitly model the temporal dynamics required for coherent video synthesis. Extending our approach to the temporal domain, potentially serving as a component within autonomous driving world models, remains an exciting avenue for future exploration.

References

[1] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In *Proc. of the IEEE/CVF Conf. on Comput. Vis. and Pattern Recog. (CVPR)*, pages 11666–11676, 2020. 2, 3, 4

[2] Miguel A. Hernán and James M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC, 2020. 1

[3] Paul W. Holland. Statistics and causal inference. *J. Amer. Stat. Assoc.*, 81(396):945–960, 1986. 1

[4] Andres Milioto, Ignacio Vizzo, Jens Behley, and Cyrill Stachniss. RangeNet++: Fast and accurate LiDAR semantic segmentation. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 4213–4220, 2019. 2

[5] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 30: 5099–5108, 2017. 2

[6] Corentin Sautier, Gilles Puy, Spyros Gidaris, Alexandre Boulch, Andrei Bursuc, and Renaud Marlet. Image-to-LiDAR self-supervised distillation for autonomous driving data. In *Proc. of the IEEE/CVF Conf. on Comput. Vis. and Pattern Recog. (CVPR)*, pages 1234–1243, 2022. 2

[7] Renkai Wu, Yinghao Liu, Guochen Ning, Pengchen Liang, and Qing Chang. UltraLight VM-UNet: Parallel vision Mamba significantly reduces parameters for skin lesion segmentation. *Patterns*, 6(11):101298, 2025. 1

[8] Yang Wu, Yun Zhu, Kaihua Zhang, Jianjun Qian, Jin Xie, and Jian Yang. WeatherGen: A unified diverse weather generator for LiDAR point clouds via spider Mamba diffusion. In *Proc. of the IEEE/CVF Conf. on Comput. Vis. and Pattern Recog. (CVPR)*, pages 17019–17028, 2025. 1, 3, 4

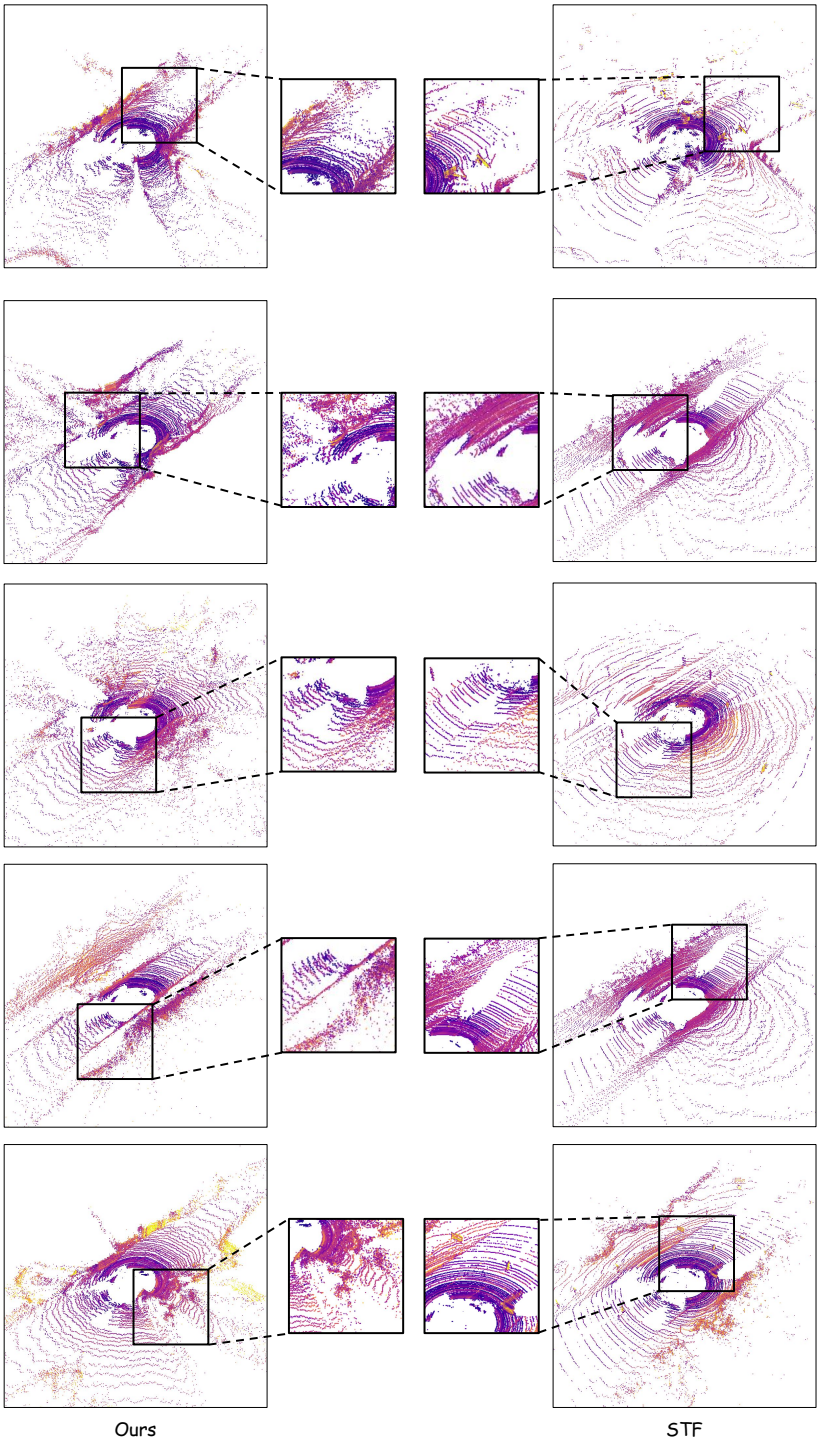


Figure 6. Visualization of generated LiDAR scan under foggy condition.

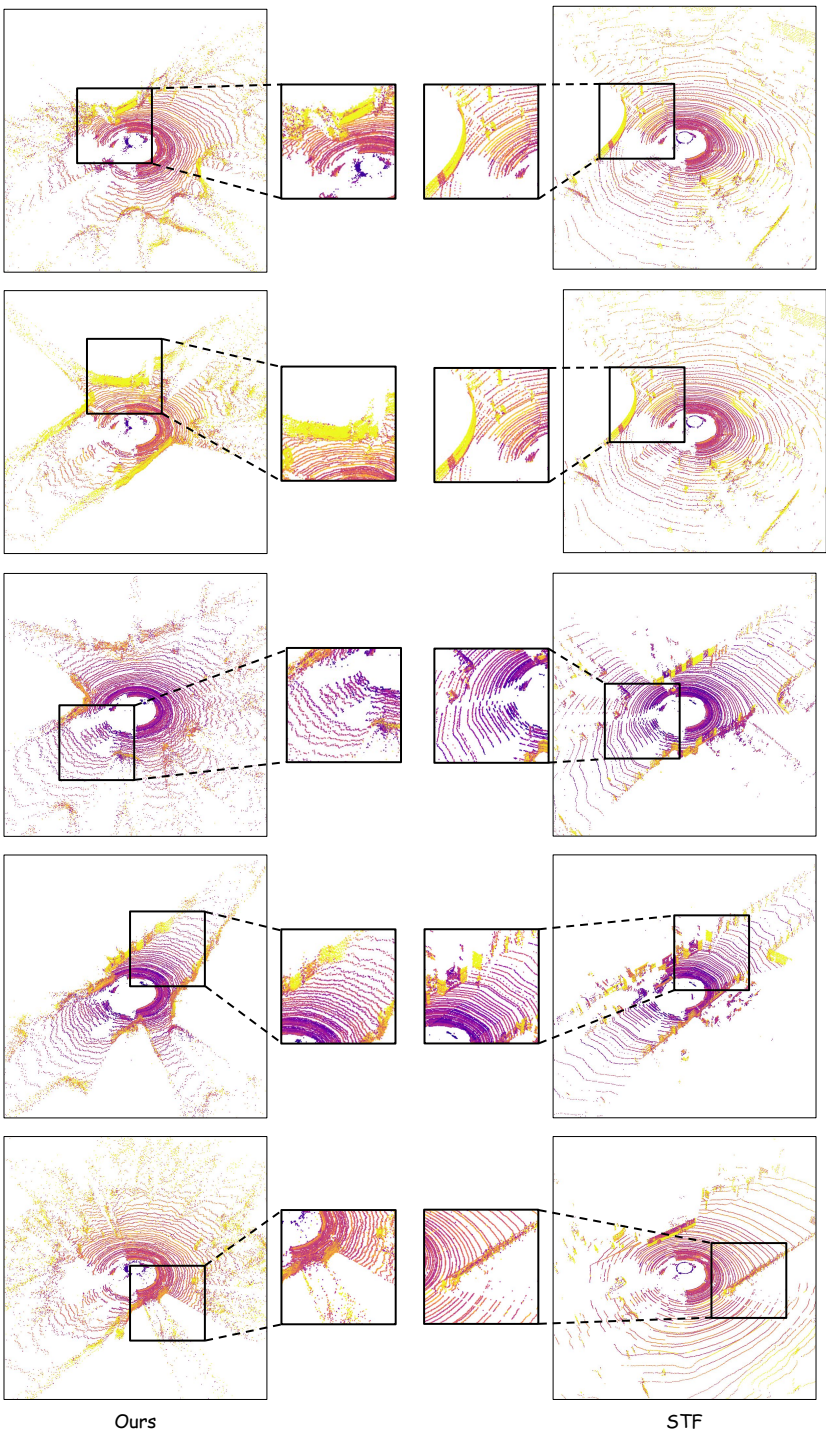


Figure 7. Visualization of generated LiDAR scan under rainy condition.

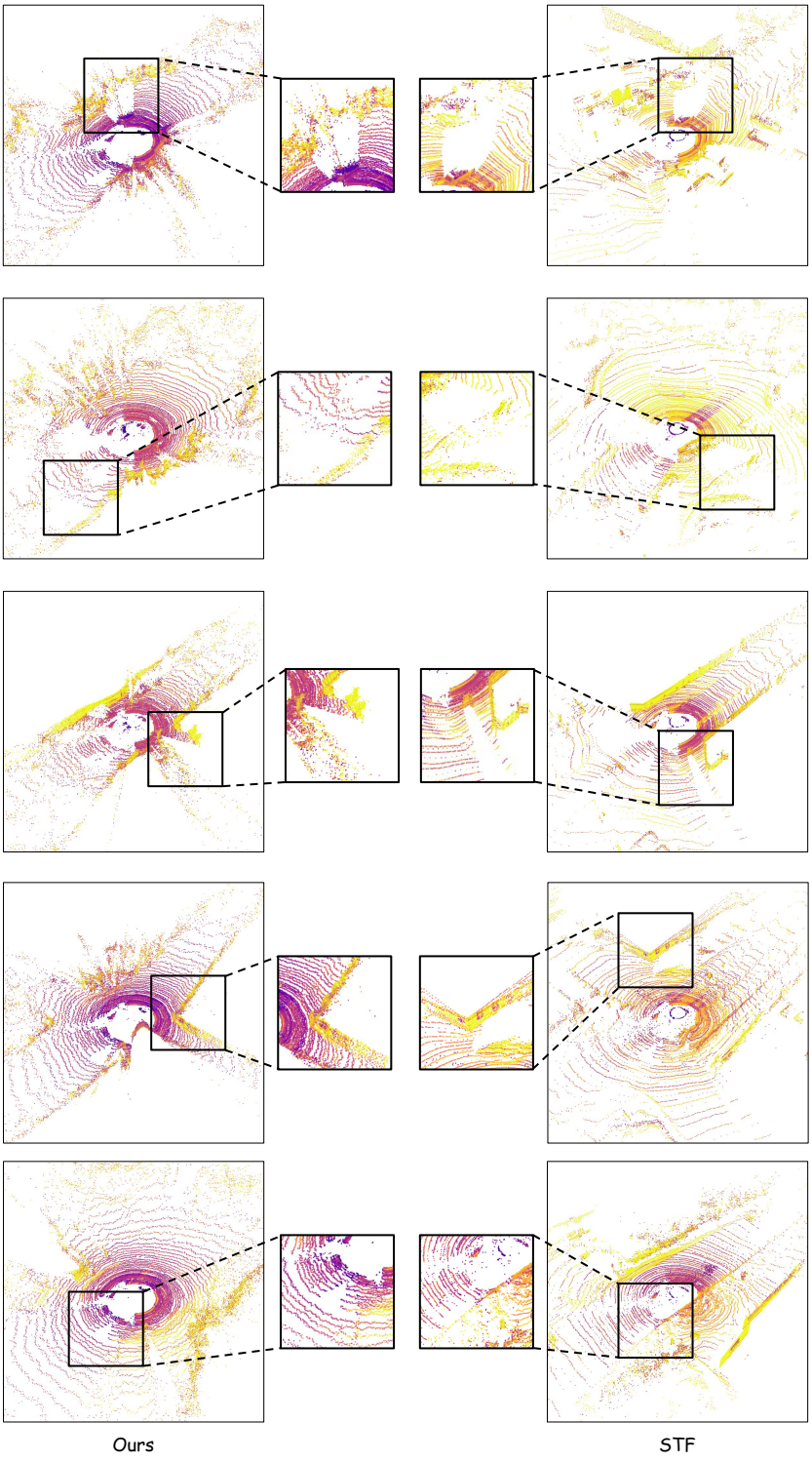


Figure 8. Visualization of generated LiDAR scan under snowy condition.