Diffusion-Based mmWave Radar Point Cloud Enhancement Driven by Range Images

Ruixin Wu, *Graduate Student Member, IEEE*, Zihan Li, Jin Wang, Xiangyu Xu, Zhi Zheng, *Graduate Student Member, IEEE*, Kaixiang Huang, and Guodong Lu

***Abstract*—Millimeter-wave (mmWave) radar has attracted significant attention in robotics and autonomous driving due to its robustness in harsh environments. However, the radar point clouds are typically sparse and noisy, which limits its futher development. Traditional mmWave radar enhancement approaches often struggle to leverage the effectiveness of diffusion models in super-resolution, largely due to the unnatural range-azimuth heatmap (RAH) or bird's eye view (BEV) representation. To address this issue, we pioneer the integration of range image representations into an image diffusion framework that leverages pre-trained image diffusion priors to generate dense and accurate 3D mmWave radar point clouds with LiDAR-like quality. Extensive evaluations on both public datasets and self-constructed datasets demonstrate that our approach provides substantial improvements, establishing a new state-of-the-art performance in generating truly three-dimensional LiDAR-like point clouds via mmWave radar.**

***Index Terms*—Range sensing, deep learning methods, representation learning.**

I. Introduction

In RECENT years, mmWave radar has found increasing applications in tasks such as Advanced Driver Assistance Systems (ADAS) and Navigation-Assisted Driving (NOA). And it has also been deployed in robotic platforms, including mobile robots [1], unmanned aerial vehicles (UAVs) [2], and unmanned surface vehicles (USVs) [3]. MmWave radar, with its compact structure and robust performance, remains reliable in extreme weather such as rain, snow, and fog, making it an ideal choice for environmental perception in harsh environments.

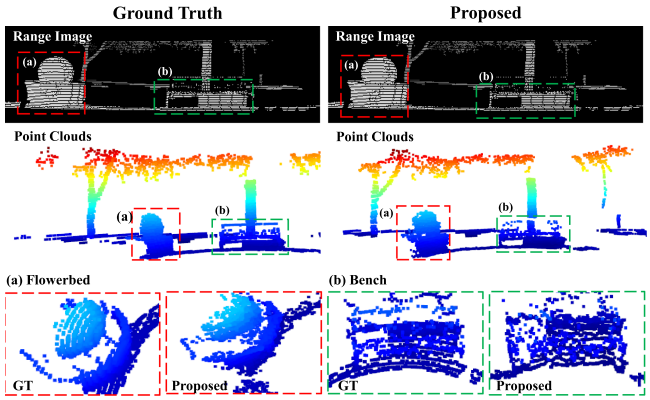

Fig. 1. Motivational comparison between Ground Truth and the Proposed method. The top row shows range images and the middle row displays the corresponding point clouds, with pseudo-color for height visualization only. Flowerbed (a) and bench (b), with zoomed-in views shown in the bottom row for both GT and Proposed. The results suggest that range images preserve rich semantic and geometric information and provide a diffusion-friendly proxy for point cloud enhancement.

However, mmWave radar is often used as an auxiliary sensor rather than a standalone solution. This is mainly due to two factors:

1) **Cluttering:** Complex electromagnetic propagation, including reflection, scattering, refraction, and diffraction, together with sidelobe effects, produces false returns and ghost points.
2) **Sparsity:** Compared with LiDAR, radar point clouds are sparse and lack fine geometric details, which limits performance in high-precision tasks such as detection, localization, and mapping.

To address these limitations, extensive research has focused on two stages of the signal processing pipeline: first, enhancing pre-processing to improve angular resolution and suppress false detections [4], [5], [6] and second, post-processing to refine geometric and detailed information while reducing the workload on the detector. The point clouds obtained by existing methods still have poor accuracy and density, limiting their suitability for backend tasks in various robotic systems. Therefore, the enhancement of mmWave radar remains an open problem.

Since it aims to denoise returns and densify missing regions, enhancing radar point clouds can be viewed as a super-resolution problem. This motivates diffusion-based solutions.

Received 9 September 2025; accepted 16 February 2026. Date of publication 13 March 2026; date of current version 3 April 2026. This article was recommended for publication by Associate Editor A. Vora and Editor J. Civera upon evaluation of the reviewers' comments. This work was supported in part by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang under Grant 2024C01020, Grant 2024C01170, and Grant 2025C01091, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LD24E050010, in part by the National Natural Science Foundation of China under Grant 52475033, and in part by Ningbo Science and Technology Project under Grant 2024Z295 and Grant 2025Z058. *(Ruixin Wu and Zihan Li contributed equally to this work.) (Corresponding author: Jin Wang.)*

Ruixin Wu, Jin Wang, Xiangyu Xu, Zhi Zheng, Kaixiang Huang, and Guodong Lu are with The State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou 310027, China, and also with the Zhejiang Key Laboratory of Industrial Big Data and Robot Intelligent Systems, Zhejiang University, Hangzhou 310058, China (e-mail: 22325053@zju.edu.cn; dwjcom@zju.edu.cn; 22325221@zju.edu.cn; z.z@zju.edu.cn; kaixianghuang@zju.edu.cn; lugd@zju.edu.cn).

Zihan Li is with the State Key Laboratory of Robotics and Systems, Department of Mechatronics Engineering, Harbin Institute of Technology, Harbin 150001, China (e-mail: 23S008053@stu.hit.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/LRA.2026.3673977, provided by the authors.

Digital Object Identifier 10.1109/LRA.2026.3673977

Recent studies [6], [7], [8] have explored diffusion models for radar perception by using structured radar representations as conditions. For example, Zhang et al. [6] apply RAH to generate LiDAR-like BEV maps, while other works convert radar point clouds into BEV to guide point cloud generation [7], [8]. These methods, including RAH and BEV, emphasize velocity cues or top-down layouts and often lack abundant pixel-level structure details and semantic context that diffusion backbones typically exploit, which limits representation learning.

To overcome this problem, we propose a novel approach that, for the first time, integrates range images with diffusion models to generate mmWave radar point clouds. Range images preserve clearer geometric structures and align better with natural-image priors, as illustrated in Fig. 1. This enables more effective conditioning and improves reconstruction quality.

We conducted extensive benchmark comparisons with existing methods on public datasets and seamlessly transferred the models trained on these datasets to our self-constructed datasets. Results show consistent improvements in point cloud quality and promising generalization.

In summary, the main contributions of this letter are as follows:

1) We propose a novel high-quality mmWave radar point cloud generation method based on a pre-trained diffusion model that can generate truly three-dimensional dense and accurate LiDAR-like point clouds.
2) Our work pioneers the integration of range image representations with diffusion models for mmWave radar super-resolution. By introducing human perception-aligned range images as a proxy representation of point clouds, we can fully leverage the prior knowledge embedded in pre-trained models.
3) We conducted extensive benchmark, generalization, and ablation experiments on both public and self-constructed datasets, which validated the superior performance of the proposed approach.

II. Related Work

Traditional radar processing, such as CFAR [9], [10], [11], [12] and MUSIC [13], generates point clouds that are sparse or noisy. Research on improving the quality of mmWave radar point clouds primarily focuses on improving pre-processing methods and post-processing methods.

A. Pre-Processing Methods

Cheng et al. [4] generated ground truth Range-Doppler Maps (RDM) from LiDAR point clouds and used them to train Generative Adversarial Networks (GANs) for building detectors. While this approach improves detection performance over traditional methods, it still relies on traditional DOA estimation, leading to noisy mmWave radar point clouds. Prabhakara et al. [5] supervised a U-Net using LiDAR point cloud labels and map low-resolution radar heatmaps to LiDAR-like point clouds, which preserves real objects while suppressing part of the noise. Han et al. [14] propose DenserRadar, which directly processes the raw 4D radar cube using a 3D U-Net. However, its computational overhead is enormous. Zhang et al. [6] supervised a diffusion model using LiDAR BEV images, which predicts LiDAR-like BEV images from paired radar RAH. Zhang et al. [15] propose RaLD, which uses spectrum-conditioned latent diffusion with a frustum-based LiDAR autoencoder to generate high-resolution 3D radar point clouds. Since generation is performed in the autoencoder latent space, it relies on task-specific autoencoder pre-training and a dedicated decoder for reconstruction.

Beyond LiDAR supervision, Fan et al. [16] generated labels using a dynamic 3D reconstruction algorithm, replacing LiDAR-based labels in RPDNet while achieving comparable performance.

B. Post-Processing Methods

Lu et al. [17] propose MilliMap, which uses a GAN and takes sparse maps obtained from Bayesian grid mapping. Geng et al. [18] first applied non-coherent integration and synthetic aperture accumulation methods to improve the density and angular resolution of radar point clouds, and then proposed the RDM network to suppress noise. However, both methods rely on accurate ego-motion estimation from other sensors, which is difficult to obtain in extreme environments. Cai et al. [19] combined traditional signal processing with an adaptive neural network to generate high-quality indoor point clouds from mmWave reflection signals, but the resulting point clouds still lack sufficient shape and detailed information. Wu et al. [8] project raw radar point clouds into a BEV representation and extracted features to condition a diffusion model for point cloud reconstruction. This representation is effective for top-down structure but provides limited range-view geometry, which can hinder learning fine-grained details.

In summary, existing methods either require additional sensors, accurate motion estimates, or task-specific representations to overcome the limitations of mmWave radar. Meanwhile, the reconstructed point clouds are still not dense or accurate enough, making it challenging to support autonomous navigation or other advanced tasks in complex environments.

III. Preliminaries

A. Diffusion Models

Diffusion models can be encapsulated within a unified generative modeling framework proposed by Karras et al. [20]. They conceptualize the diffusion process, which is represented as a stochastic differential equation (SDE):

$$dx = f(t)x\,dt + g(t)\,d\omega_t, \tag{1}$$

where ω_t denotes the standard Wiener process, $f(\cdot) : \mathbb{R} \to \mathbb{R}$ and $g(\cdot) : \mathbb{R} \to \mathbb{R}$ represent the drift and diffusion coefficients, respectively, with d indicating the dimensionality of the dataset.

The diffusion model learns to reverse a forward diffusion process, which corresponds to the forward SDE in (1) and is defined as follows:

$$q(x_t|x_0) := \mathcal{N}(x_t; s(t)x_0, s^2(t)\sigma^2(t)\mathcal{I}), \tag{2}$$

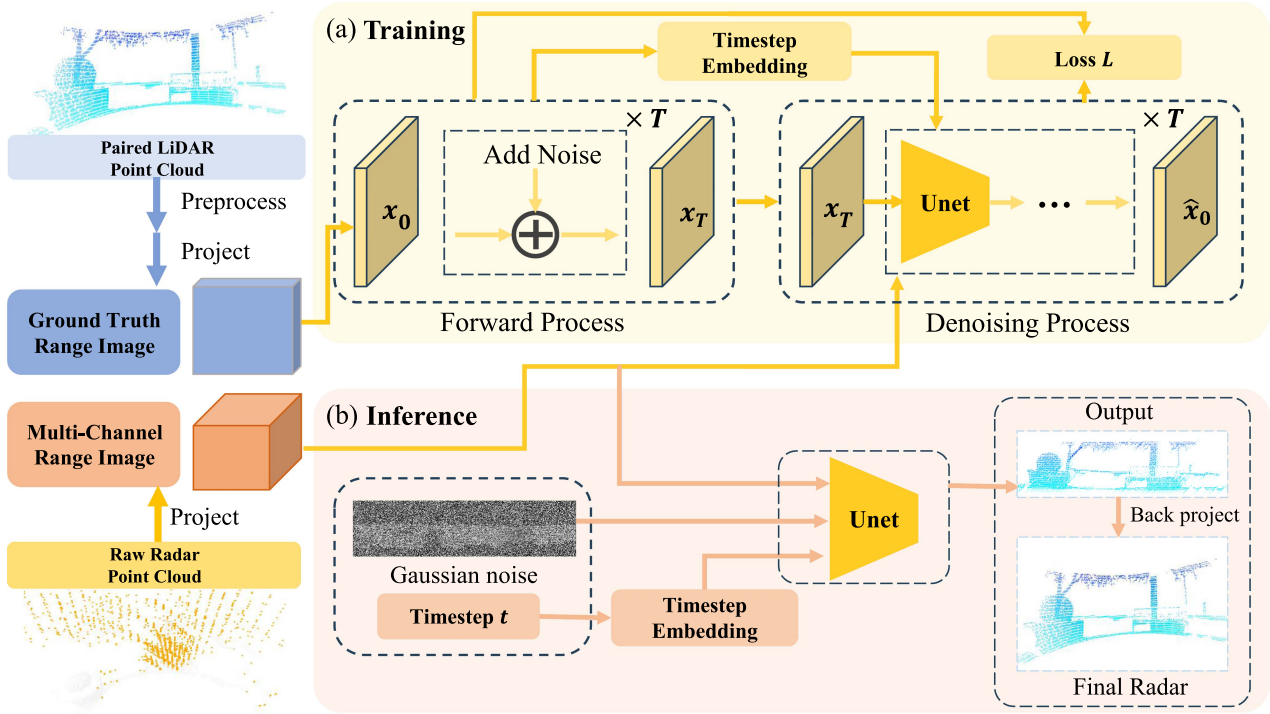

Fig. 2. Diagram of the proposed framework. (a) In training, the LiDAR and raw mmWave radar point clouds are projected into the range image x_0 and the multi-channel range image c. Then x_0 is corrupted to x_T by adding Gaussian noise for T diffusion steps; the step index $t \in \{1, \ldots, T\}$ is mapped to a timestep embedding and fed to the network together with x_t to recover x_0, conditioned on c. (b) In inference, the network directly predicts $\hat{x}_0$ from pure Gaussian noise with the same timestep embedding conditioned on c. Then $\hat{x}_0$ is back-projected to obtain the final high-quality mmWave radar point cloud.

where $q(\cdot|\cdot)$ represents the conditional probability, $\mathcal{N}(x; \mu, \Sigma)$ denotes the probability density function of $\mathcal{N}(\mu, \Sigma)$ evaluated at x, $t \in \{1, \ldots, T\}$, where T is the total number of steps in the diffusion chain, $s(t) = \exp(\int_0^t f(\xi)\, d\xi)$, and $\sigma(t) = \sqrt{\int_0^t \frac{g(\xi)^2}{s(\xi)^2}\, d\xi}$. Gaussian noise can be added to x_0 to yield any x_t in a single step. The forward process models a fixed Markov chain, and the noise depends on a variance schedule $s^2(t)\sigma^2(t)$.

The reverse process is defined by the following formula, with parameters θ:

$$p_\theta(x_0) := \int p_\theta(x_{0:T})\, dx_{1:T}, \tag{3}$$

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t), \tag{4}$$

$$p_\theta(x_{t-1}|x_t) := \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)). \tag{5}$$

By employing a fixed variance $\Sigma_\theta(x_t, t)$, our task reduces to learning the means of the reverse process, $\mu_\theta(x_t, t)$. Training is conventionally carried out through a reweighted variational bound on the maximum likelihood objective, with the loss defined as follows:

$$L := \mathbb{E}_{t,q}\left[\lambda_t \left\|\mu_t(x_t, x_0) - \mu_\theta(x_t, t)\right\|^2\right], \tag{6}$$

where $\mu_t(x_t, x_0)$ is the mean of the forward process posterior $q(x_{t-1}|x_t, x_0)$.

IV. Methods

This section specifically details the proposed method. Motivated by recent advances in image diffusion models for super-resolution, we integrate range images with diffusion models for point-cloud enhancement. Specifically, the range image projected from the LiDAR point cloud is used to supervise the training of the diffusion model, after which high-quality range images are restored from noise, conditioned by range data from mmWave radar. Ultimately, the range image is back-projected to generate high-quality mmWave radar point clouds. The system architecture is illustrated in Fig. 2.

A. Range Image Construction

Certain LiDAR systems, such as Velodyne, generate raw data in a format that resembles range images. Each column represents the distances measured by laser range-finders at a specific moment, while each row corresponds to varying rotational angles of the sensors. This implies that range images can serve as a proxy for point clouds. A range image, in essence, is a 2D array where each pixel contains the spherical coordinates and range of a point mapped onto its field of view. We transform each point $\Pi : \mathbb{R}^3 \to \mathbb{R}^2$ through the mapping $\mathbf{p_i} = (p_x, p_y, p_z)$ into spherical coordinates, and then into image coordinates, as follows:

$$\begin{pmatrix} i_\theta \\ i_\phi \end{pmatrix} = \begin{pmatrix} \lfloor (\arctan 2(p_y, p_x) - \theta_{\min})\, l_w r_\theta^{-1} \rfloor \\ \lfloor (\arccos(p_z, r) - \phi_{\min})\, l_h r_\phi^{-1} \rfloor \end{pmatrix}, \tag{7}$$

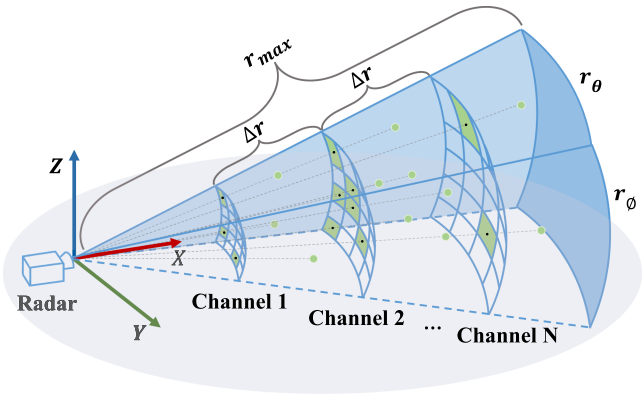

Fig. 3. Illustration of multi-channel range image point cloud proxy representation for mmWave Radar.

where $r = \|\mathbf{p_i}\|_2$ denotes the range of each point, $\theta_{\min}$ and $\phi_{\min}$ represent the minimum values of the azimuth and elevation angles, r_θ and r_ϕ are the ranges of the azimuth and elevation angles, l_w and l_h are the width and height of the range image, and $\lfloor \cdot \rfloor$ denotes the floor function.

B. Data Representation

Furthermore, projecting raw mmWave radar point clouds into a single-channel range image often causes pixel collisions, where multiple returns fall into the same pixel and overwrite each other, leading to occluded far returns and information loss. While similar multi-dimensional radar tensors have been used in recent works, our key design is range-aware channel semantics. We perform slicing operations along the range dimension of the mmWave radar point cloud and individually map the radar data from each slice, thus generating multi-channel range images as the conditions of the diffusion model. The necessity of multiple channels is also validated in Section V-B2. The process of constructing multi-channel range images is illustrated in Fig. 3.

Notably, the range image configuration, including resolution and channel capacity, critically affects projection quality. Low resolution or few channels increases quantization and overwriting, while overly large settings increase computation and reduce effective supervision density. We therefore re-project range images at different configurations back to point clouds and quantify information loss using metrics such as point retention and reconstruction error, as detailed in Section V-C. Based on these measurements, we select the range image configuration. In contrast, LiDAR point clouds with their angular resolution and relatively organized spatial distribution allow for the conversion to range images with minimal loss when a resolution approximating the LiDAR's angular resolution is chosen.

C. Diffusion-Based Range Image Prediction

As elaborated in Section III-A, the principal objective of the diffusion model is to learn the reverse process of the forward diffusion mechanism, which incrementally introduces noise to data samples according to the predefined noise schedule described in (2), where $s(t) = 1$ and $\sigma(t) = t$. Given a set of LiDAR range images x_0 and multi-channel radar range images c that are aligned both spatially and temporally, Gaussian noise is introduced and propagated to x_t. We follow the parameterization approach for the reverse process proposed by Karras et al. [20] to model the original data sample $D_\theta(x_t, t, c)$ conditioned on c. During the inference stage, we perform deterministic sampling using the probabilistic flow (PF) ODE defined by (1) to accelerate the inference process. The Heun method is employed for iterative solving. The PF ODE is defined as follows:

$$dx = -\dot{\sigma}(t)\sigma(t)\nabla_x \log p(x; \sigma(t))dt, \tag{8}$$

where the dot denotes a time derivative, $\nabla_x \log p(x; \sigma(t))$ represents the score function.

1) Design of the Network Architecture: The resolutions of range images from LiDAR and mmWave radar differ due to distinct point cloud densities. If the range images of the mmWave radar are directly concatenated with the noise at the input layer, size alignment through methods such as interpolation becomes necessary, leading to the inevitable loss of information. Consequently, we opt to embed conditions during the downsampling stage, thereby ensuring the maximum retention of the original data. Furthermore, as the pitch range of LiDAR is narrower than its horizontal range, the height and width of its range images are unequal. Thus, we introduce a horizontal sampling module, which effectively extracts the lateral features of the target, adjusts the feature dimensions, and accelerates the feature extraction process simultaneously.

2) Training Objective: We first optimize the following Mean Squared Error (MSE) loss to make $D_\theta(x_t, t, c)$ similar to x_0. The MSE loss is expressed as follows:

$$L_m = \|x_0 - D_\theta(x_t, t, c)\|_2^2. \tag{9}$$

However, since MSE loss is particularly sensitive to outliers, it may lead the model to overly focus on the few anomalous pixels within the image, thereby ignoring the broader enhancement in image quality. To mitigate this issue, we introduce the Learned Perceptual Image Patch Similarity (LPIPS) [21], which leverages a pre-trained deep network to extract features from both x_0 and $D_\theta(x_t, t, c)$. and then calculates the distance between these features to evaluate the perceptual similarity between the images. The LPIPS loss is articulated as follows:

$$L_p = \|g_p(x_0) - g_p(D_\theta(x_t, t, c))\|_2^2. \tag{10}$$

Furthermore, when transforming point clouds into range images, while the loss of geometric information remains negligible, it cannot be eliminated. To mitigate this geometric loss, drawing inspiration from the per-point coordinate supervision method in [6], we introduce per-pixel distance supervision. This involves calculating the distance loss for each pixel based on the predicted distance at each pixel location.

$$L_c = |x_0 - D_\theta(x_t, t, c)|. \tag{11}$$

Finally, our loss function consists of three components.

$$L = \lambda_m L_m + \lambda_p L_p + \lambda_c L_c. \tag{12}$$

TABLE I
QUANTITATIVE RESULTS ON THE COLORADAR DATASET. THE **BOLD** DENOTES THE BEST PERFORMANCE

Method	Aspen Lab			Hallways			Edgar Mine			Outdoor		
	CD(m) ↓	MHD(m) ↓	F-Score(%) ↑	CD(m) ↓	MHD(m) ↓	F-Score(%) ↑	CD(m) ↓	MHD(m) ↓	F-Score(%) ↑	CD(m) ↓	MHD(m) ↓	F-Score(%) ↑
OS-CFAR [12]	5.741	0.902	20.8	8.327	0.963	20.8	4.79	1.064	25.6	9.20	2.459	18.9
MUSIC [13]	11.782	0.722	4.6	18.277	0.694	4.0	17.605	1.054	1.6	13.653	2.038	4.1
RPDNet [4]	2.329	0.545	32.8	3.044	0.464	33.7	2.927	0.858	30.7	8.183	2.203	24.3
DenserRadar [14]	8.345	4.081	14.8	3.603	0.653	30.7	2.315	0.742	37.4	20.2	11.5	15.7
DreamPcd [18]	4.958	2.117	9.7	5.574	1.033	15.2	7.613	1.521	8.7	7.636	1.985	8.8
Ours	**2.259**	0.522	**48.6**	**2.754**	0.644	**36.3**	**2.042**	0.676	**48.8**	**6.615**	**1.948**	**26**
Ours-CD	2.290	**0.223**	42.8	2.834	**0.314**	39.8	2.547	**0.377**	47.8	10.963	2.110	19.9

where $\lambda_m, \lambda_p, \lambda_c$ denote the weight coefficients of each loss function, used to balance multi-objective optimization during model training

V. EXPERIMENTS AND RESULTS

In this section, we conduct benchmark comparisons, ablation experiments, and real-world experiments on both the public dataset (Coloradar) and the self-constructed dataset to validate the superior performance of our method.

A. Benchmark Comparisons

We performed benchmark comparisons with four baseline methods, including OS-CFAR [12], MUSIC [13], RPDNet [4], DenserRadar [14], Radar-Diffusion [6], and DreamPcd [18]. These baselines cover classical DOA-based pipelines, representative learning-based generation methods, and recent diffusion-related designs, all evaluated under a unified point-level protocol on ColoRadar. Additional methods are discussed in Section II but are excluded when public code is unavailable or insufficient for reproducible evaluation.

1) Dataset Configuration: ColoRadar provides LiDAR point clouds and raw mmWave ADC data for seven different environments with multiple trajectories. We use the four scenes in Table I. For each scene, the first three trajectories constitute the training set, and the rest are used for testing. For fair comparison, the dataset configurations of RPDNet [4], DenserRadar [14], Radar-Diffusion [6], and DreamPcd [18] methods are aligned with those adopted in this study.

2) Dataset Preprocessing: Due to the different operational ranges of mmWave radar and LiDAR, we preprocessed the raw sensor data. First, we align the radar and LiDAR using the extrinsic calibration and keep only the shared field of view. In addition, since mmWave radar is insensitive to low-reflectivity surfaces such as floors and ceilings, we applied Patchwork++ [22] to remove these point clouds from the LiDAR data, where the ceiling is removed by flipping the Z-axis and applying Patchwork++ again.

Additionally, considering that mmWave radar is not sensitive to low-reflectivity objects, we first employed DBSCAN to cluster the combined point cloud of mmWave radar and LiDAR. Then, we filter the LiDAR points using radar-derived labels to avoid supervising the model with structures that are consistently invisible to mmWave radar.

TABLE II
GENERALIZATION TESTING OF THE PROPOSED METHOD

Method	Corridor			Hall		
	CD(m) ↓	MHD(m) ↓	F-Score(%) ↑	CD(m) ↓	MHD(m) ↓	F-Score(%) ↑
OS-CFAR [12]	29.712	3.08	2.1	17.294	3.713	2.5
Ours	**7.615**	**1.058**	**11.4**	**8.835**	**2.584**	**7.6**

3) Implementation: In the process of mmWave point cloud processing, we extract the raw radar point cloud with CFAR, whose output is sensitive to the detection threshold. A higher threshold suppresses noise but removes valid returns and makes it difficult to match the LiDAR density. Therefore, to preserve as much raw mmWave radar information as possible, we configured the CFAR detector's threshold to approach zero.

To balance model capacity and information, we set the range image resolution of LiDAR and radar to 128×512 and 64×64, respectively. We adopted the same diffusion noise and timestamp settings as the EDM proposed by Karras et al. [20], and used the Heun deterministic sampler for sampling.

4) Qualitative Comparison: The mmWave radar point clouds generated by both the proposed and baseline methods, alongside the ground truth LiDAR point clouds, are presented in Figs. 1 and 4. The results demonstrate that our method outperforms all baseline methods in terms of density across all scenes. Our approach produces high-fidelity scene details, reconstructs the curved geometry of objects, and maintains sharp corner features under complex environmental conditions. In contrast to the OS-CFAR [12], and MUSIC [13], learning-based generation baselines [4], [6], [14], [18] recover more points, but their surface coverage remains incomplete and insufficiently dense. DenserRadar [14] achieves higher point density yet tends to exhibit structured stripe-like patterns in the wall regions, leading to incomplete geometric reconstruction. Radar-Diffusion [6] excels at reconstructing the BEV map but lacks elevation detail.

5) Quantitative Comparison: We begin by presenting the evaluation metrics employed in our quantitative comparison: Chamfer Distance (CD), Modified Hausdorff Distance (MHD), and F-Score. CD measures the mean nearest-neighbor distance between two point sets, MHD reflects the Hausdorff-style discrepancy, and F-Score is the harmonic mean of precision P and recall R under a fixed matching threshold. The results of the quantitative analysis are provided in Table I, while Fig. 7

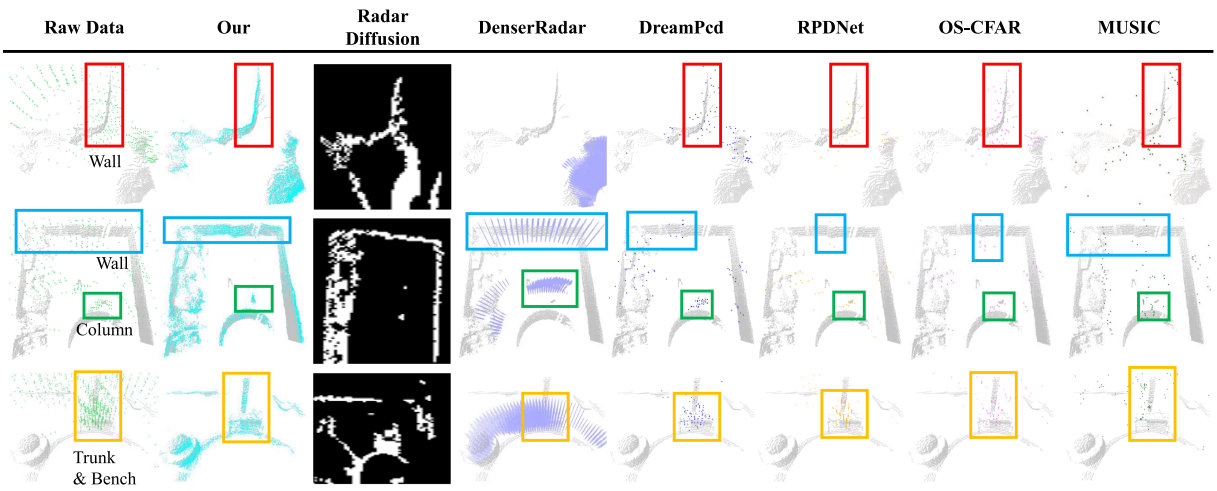

Fig. 4. Qualitative comparisons of single-frame 3D point clouds on the ColoRadar dataset. Rows show the Edgar Mine, Aspen Lab, and Outdoor scenes, and columns compare different methods, with the first column showing the raw input. For clearer visualization, the semi-transparent LiDAR ground truth is overlaid on each result, and the boxes highlight structures that are correctly reconstructed.

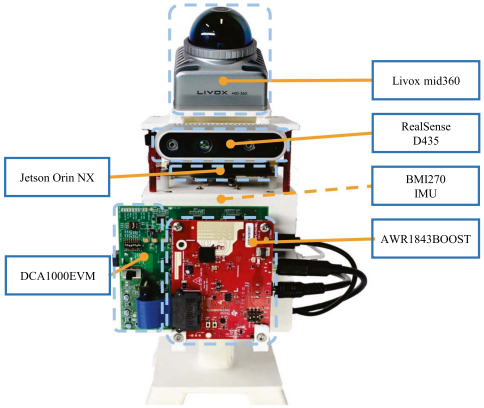

Fig. 5. Our customized handheld data collection platform. The LiDAR is mounted at an inclined angle to align with the field of view of the mmWave radar.

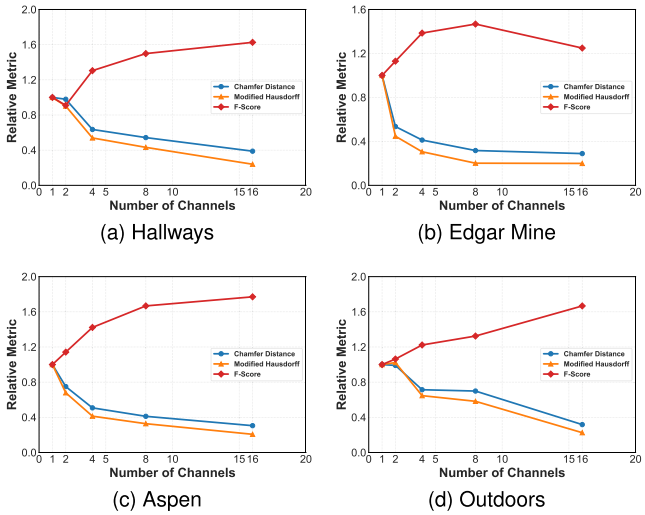

Fig. 6. Ablation study on the number of channels in mmWave radar range images. The y-axis denotes the relative proportion of each metric with respect to the case when the number of channels is one.

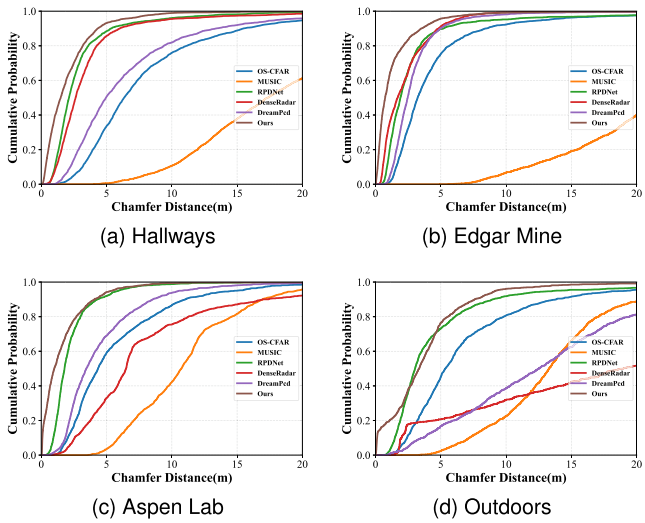

Fig. 7. The CDF curves of our method and the baseline methods on the ColoRadar dataset.

illustrates the cumulative distribution function (CDF) curves of the Chamfer Distance.

The results demonstrate that the proposed method significantly outperforms the baseline approaches across all scenes in the ColoRadar datasets. In the Hallways scene, RPDNet [4] secures a marginal advantage in the MHD, which is highly sensitive to outliers, owing to its robust filtering capabilities. However, our method performs better in CD and F-Score. This divergence underscores the inherent trade-off between outlier suppression and the preservation of surface details. While range-image-based approaches may introduce minor artifacts at depth discontinuities, thereby affecting the MHD, they achieve a more comprehensive retention of surface geometry and density. Consequently, this fidelity to the original structure grants them a distinct advantage in the CD and F-Score metrics, which are more indicative of the overall reconstruction quality.

TABLE III
ABLATION EXPERIMENT OF THE DATA REPRESENTATION FORMAT

Method	Hallways			Edgar Mine			Aspen Lab			Outdoor		
	CD(m) ↓	MHD(m) ↓	F-Score(%) ↑	CD(m) ↓	MHD(m) ↓	F-Score(%) ↑	CD(m) ↓	MHD(m) ↓	F-Score(%) ↑	CD(m) ↓	MHD(m) ↓	F-Score(%) ↑
Ours-BEV	2.452	0.548	**51.808**	5.127	1.154	30.374	2.462	0.723	47.234	20.155	8.358	18.113
Ours-RI	**2.259**	**0.522**	48.6	**2.754**	**0.644**	**36.3**	**2.042**	**0.676**	**48.8**	**6.615**	**1.948**	**26.0**

Note: Ours-RI denotes our method using Range Image as the data representation format, while Ours-BEV represents our approach with Bird's Eye View representation.

TABLE IV
PROJECTION-INDUCED INFORMATION LOSS ANALYSIS UNDER DIFFERENT RANGE IMAGE CONFIGURATIONS

Channels	Resolution	Hallways			Edgar Mine			Aspen Lab			Outdoor		
		CD(m) ↓	Ret.(%) ↑	Occ.(%)	CD(m) ↓	Ret.(%) ↑	Occ.(%)	CD(m) ↓	Ret.(%) ↑	Occ.(%)	CD(m) ↓	Ret.(%) ↑	Occ.(%)
8	64×64	1.842	22.82	3.27	1.790	41.47	1.93	2.143	30.63	2.10	2.290	33.46	2.03
16	64×64	**0.704**	**44.01**	3.15	**0.972**	**69.15**	1.61	**0.767**	**57.89**	1.98	**0.910**	**63.38**	1.92
32	64×64	**0.704**	**44.01**	1.58	**0.972**	**69.15**	0.80	**0.767**	**57.89**	0.99	**0.910**	**63.38**	0.96
16	32×32	0.706	34.67	9.94	0.963	56.84	5.28	0.775	47.62	6.51	0.911	52.63	6.36
16	64×64	0.704	**44.01**	3.15	0.972	**69.15**	1.61	0.767	**57.89**	1.98	0.910	**63.38**	1.92
16	128×128	**0.682**	**44.01**	0.79	**0.951**	**69.15**	0.40	**0.746**	**57.89**	0.49	**0.889**	**63.38**	0.48

B. Ablation Experiments

Beyond benchmark comparisons, we conducted ablation experiments to substantiate the necessity of the proposed method.

1) Data Representation Format: In this experiment, the data representation in our method is switched from range images to BEV, a format widely used in prior studies [6], [7], [8], while all other components remain unchanged. Table III reports the quantitative results: representing the data as range images yields substantially higher pointcloud quality than BEV. This suggests that range images, which align with human perceptual priors, more effectively exploit the capabilities of the diffusion model, thereby underscoring the efficacy of the proposed approach.

2) Range Image Channel Number: Furthermore, we performed an ablation study on the range image channels. Only the number of range image channels was altered in this experiment, and all other settings remained identical to the benchmark comparisons. The quantitative results in Fig. 6 show that multi-channel range images consistently outperform the single-channel setting. This is mainly due to the penetrative nature of mmWave radar, which often produces multiple returns along similar directions. With a limited channel capacity, these returns are forced to collide in the same angular bin and overwrite each other, leading to severe information loss. The experiment substantiates the imperative for employing multi-channel range images, with 16-channel configurations yielding better performance.

C. Projection-Induced Information Loss Analysis

To quantify the supervision loss introduced by range-image construction and to select the appropriate number of radar range image channels and resolution, we further analyze the information loss caused by the projection and back-projection process. Specifically, we project the raw radar point cloud into range images under different channel numbers and resolutions, and then back-project them to 3D for comparison. We report CD, Retention rate (Ret.), and Occupancy rate (Occ.) as quantitative metrics. The results are summarized in Table IV.

1) Effect of Channel Number: Increasing the channel number from 8 to 16 yields a clear improvement in both CD and Ret. across all scenes, confirming that additional channels effectively reduce overwrite-induced loss from many-to-one projection collisions. Increasing to 32 does not further improve CD or Ret., while Occ. drops substantially, indicating a much sparser tensor with higher computational cost and weaker effective supervision density. Therefore, we adopt 16 channels as the default configuration, which achieves the best balance between information preservation and efficiency.

2) Effect of Resolution: We also evaluate the impact of range image resolution while keeping all other settings identical. When the resolution is increased from 32×32 to 64×64, Ret. consistently improves across all four scenes, indicating that finer angular quantization reduces many-to-one projection collisions and alleviates information loss. Further increasing the resolution to 128×128 brings only marginal gains in CD and Ret., while Occ. decreases substantially, and the number of pixel computations increases fourfold, resulting in a sparser representation and higher computational overhead. Therefore, we ultimately adopt 64×64 as the range image resolution, which achieves the best trade-off between information preservation and computational efficiency.

D. Real-World Experiments

We further assess the generalization capability of the proposed method by employing our self-constructed datasets.

1) Data Collection Platform: We customized a handheld data collection platform, as shown in Fig. 5. It consists of an NVIDIA Jetson Orin NX for recording sensor data, a Livox Mid-360 LiDAR, and a BMI270 IMU to estimate real-time platform poses with the assistance of Fast-LIO [23]. Additionally, it is

equipped with a TI-AWR1843BOOST and TI-DCA1000EVM for collecting mmWave radar raw data, and an Intel RealSense D435 to capture image data for visualization. Using the handheld data collection platform, we respectively collected four trajectories in the hall and corridor, totaling 6,636 frames.

2) Generalization Capability Tests: We deployed the model trained on the ColoRadar dataset directly on our self-collected dataset, without any further fine-tuning, and benchmarked it against CFAR-based [12] baselines. This test has achieved robust generalization across heterogeneous scenes, LiDAR sensors, and radar parameterizations. Quantitative results are reported in Table II, with metrics computed using LiDAR point clouds as ground truth. Even under unseen scenarios and sensor configurations, the proposed method substantially outperforms conventional CFAR-based approaches, compellingly attesting to its efficacy and generalization capacity.

3) System Efficiency: We accelerate inference via diffusion consistency distillation, reducing sampling to 10 steps. On an NVIDIA RTX 4060 Ti, inference takes 212.1 ms on average, achieving about 5 Hz, which is sufficient for timely obstacle perception and trajectory replanning within a 16 m sensing range. We also observe that distillation improves the MHD, likely because it suppresses outlier points and spurious artifacts that dominate Hausdorff-type distances. In addition, runtime can be further reduced via FP16 or mixed-precision execution and TensorRT-based optimization.

VI. Conclusion

This letter proposes a novel approach for generating high-quality mmWave radar point clouds. By leveraging the strong generative capability of diffusion models, the proposed method enhances sparse and noisy single-frame radar point clouds to achieve LiDAR-like quality. Our approach employs range images aligned with human perception as proxy representations and combines them with image diffusion models for the first time for this task. Since the human-aligned projection makes range images resemble natural images, knowledge from pre-trained image diffusion models can be transferred effectively, substantially enhancing overall performance. The proposed method is validated on public and self-constructed datasets, significantly outperforming baseline methods in both point cloud quality and density.

In the future, we intend to explore the interdependencies between adjacent frames of mmWave radar data to further enhance the interframe continuity of mmWave radar point clouds and apply the proposed method to downstream robot tasks.

References

[1] Y. Lyu, L. Hua, J. Wu, X. Liang, and C. Zhao, "Robust radar inertial odometry in dynamic 3D environments," *Drones*, vol. 8, 2024, Art. no. 197.

[2] J. Zhang et al., "4DRadarSLAM: A 4D imaging radar SLAM system for large-scale environments based on pose graph optimization," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2023, pp. 8333–8340.

[3] Y. Cheng, H. Xu, and Y. Liu, "Robust small object detection on the water surface through fusion of camera and millimeter-wave radar," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 15243–15252.

[4] Y. Cheng, J. Su, M. Jiang, and Y. Liu, "A novel radar point cloud generation method for robot environment perception," *IEEE Trans. Robot.*, vol. 38, no. 6, pp. 3754–3773, Dec. 2022.

[5] A. Prabhakara et al., "High resolution point clouds from mmWave radar," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2023, pp. 4135–4142.

[6] R. Zhang, D. Xue, Y. Wang, R. Geng, and F. Gao, "Towards dense and accurate radar perception via efficient cross-modal diffusion model," *IEEE Robot. Automat. Lett.*, vol. 9, no. 9, pp. 7429–7436, Sep. 2024.

[7] K. Luan, C. Shi, N. Wang, Y. Cheng, H. Lu, and X. Chen, "Diffusion-based point cloud super-resolution for mmWave radar data," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2024, pp. 11171–11177.

[8] J. Wu et al., "DiffRadar: High-quality mmWave radar perception with diffusion probabilistic model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2024, pp. 8291–8295.

[9] M. Barkat, S. D. Himonas, and P. K. Varshney, "CFAR detection for multiple target situations," *IEE Proc. F Radar Signal Process.*, vol. 136, pp. 193–209, 1989.

[10] G. Minkler and J. Minkler, "CFAR: The principles of automatic radar detection in clutter," NASA, Washington, DC, USA, NASA STI/Recon Tech. Rep. A., vol. 90, 1990, Art. no. 23371.

[11] P. P. Gandhi and S. A. Kassam, "Analysis of CFAR processors in non-homogeneous background," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 427–445, Jul. 1988.

[12] H. Rohling, "Radar CFAR thresholding in clutter and multiple target situations," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-19, no. 4, pp. 608–621, Jul. 1983.

[13] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. TAP-34, no. 3, pp. 276–280, Mar. 1986.

[14] Z. Han et al., "DenserRadar: A 4D millimeter-wave radar point cloud detector based on dense LiDAR point clouds," in *Proc. IEEE 27th Int. Conf. Intell. Transp. Syst.*, 2024, pp. 930–936.

[15] R. Zhang, B. Zeng, S. Wang, F. Zhou, and W. Wang "Rald: Generating high-resolution 3d radar point clouds with latent diffusion," in *Proc. AAAI Conf. Artif. Intell.*, vol. 40, no. 22, 2026, pp. 18773–18781.

[16] C. Fan, S. Zhang, K. Liu, S. Wang, Z. Yang, and W. Wang, "Enhancing mmWave radar point cloud via visual-inertial supervision," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2024, pp. 9010–9017.

[17] C. X. Lu et al., "See through smoke: Robust indoor mapping with low-cost mmWave radar," in *Proc. 18th Int. Conf. Mobile Syst., Appl., Serv.*, 2020, pp. 14–27.

[18] R. Geng et al., "Dream-PCD: Deep reconstruction and enhancement of mmWave radar point cloud," *IEEE Trans. Image Process.*, vol. 33, pp. 6774–6789, 2024.

[19] P. Cai and S. Sur, "MilliPCD: Beyond traditional vision indoor point cloud generation via handheld millimeter-wave devices," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 6, pp. 1–24, 2023.

[20] T. Karras, M. Aittala, T. Aila, and S. Laine, "Elucidating the design space of diffusion-based generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, vol. 35, pp. 26565–26577.

[21] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 586–595.

[22] S. Lee, H. Lim, and H. Myung, "Patchwork++: Fast and robust ground segmentation solving partial under-segmentation using 3D point cloud," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2022, pp. 13276–13283.

[23] W. Xu and F. Zhang, "FAST-LIO: A fast, robust LiDAR-inertial odometry package by tightly-coupled iterated Kalman filter," *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 3317–3324, Apr. 2021.