RaLD: Generating High-Resolution 3D Radar Point Clouds with Latent Diffusion

Ruijie Zhang[1] Bixin Zeng[1] Shengpeng Wang[1] Fuhui Zhou[2] Wei Wang[3]

[1]Huazhong University of Science and Technology

[2]Nanjing University of Aeronautics and Astronautics [3]Wuhan University

AAAI-26 / IAAI-26 / EAAI-26

JANUARY 20–27, 2026 | SINGAPORE

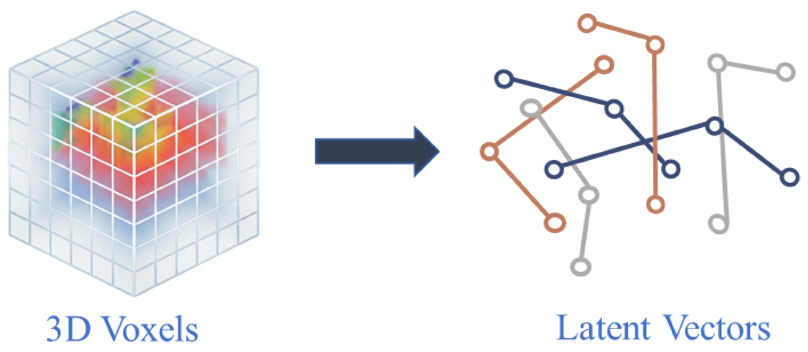
Overview

3D Radar Super-Resolution:

- *Goal:* Generate high-fidelity, LiDAR-like dense point clouds from sparse, noisy raw radar spectrums.

Existing Challenges: Prior generative methods rely on **dense voxel representations** (e.g., 3D voxel grids), which are **computationally expensive and struggle to preserve fine-grained structural details** due to high memory consumption and inherent sparsity.

Observation: Latent Diffusion Models (LDMs) can alleviate the burden of modeling unordered 3D data by operating in a **compact, lower-dimensional latent space**.



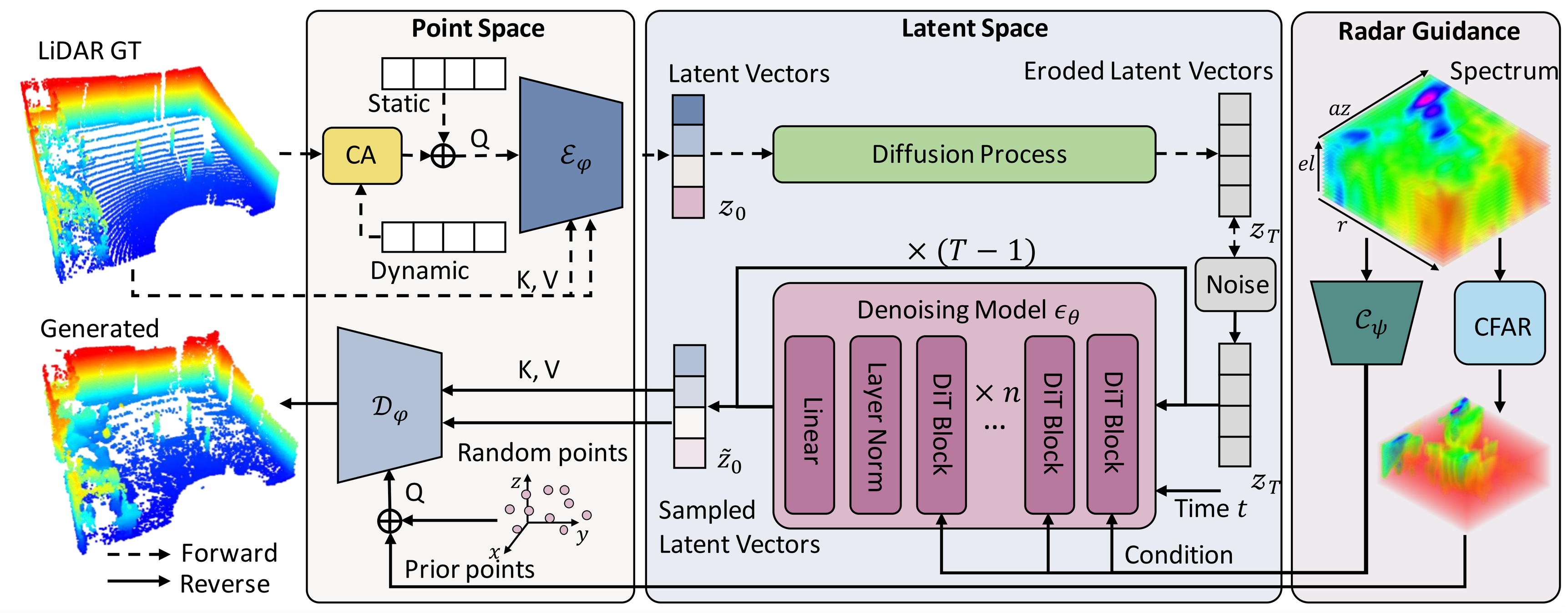
Innovations

Radar-based Latent Diffusion (RaLD) framework is the first to explore sparse, point-based latent diffusion for this task through **three key innovations**:

- **Frustum-Based Autoencoder**: A tailored architecture that aligns with the **polar sampling geometry** of radar and LiDAR, effectively preserving spatial regularity and capturing non-uniform point density across depth.
- **Order-Invariant Latent Encoding**: A hybrid strategy fusing static and dynamic queries to ensure **consistent latent representations** regardless of input point ordering, facilitating stable diffusion training.
- **Radar Spectrum Guidance**: Injects semantic and geometric cues directly from raw spectrums into the diffusion process, utilizing **CFAR-guided query initialization** to significantly improve decoding efficiency and structural accuracy.

3D Radar Latent Diffusion Model (RaLD)

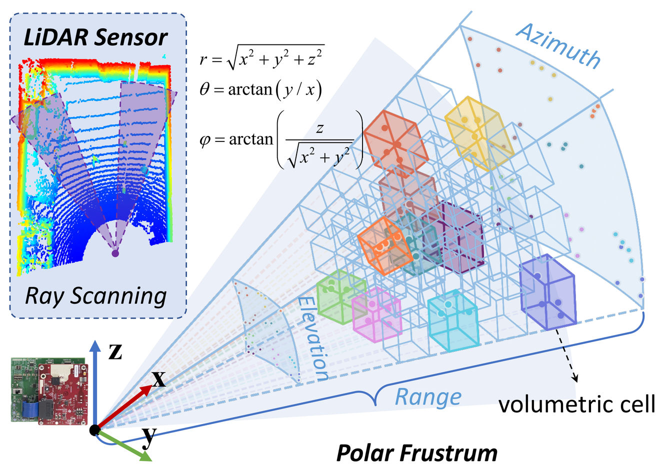
Pipeline of RaLD:



Frustum-Based LiDAR Autoencoder

Motivation: Cartesian voxels ignore the angular sampling pattern of sensors, causing non-uniform point distributions.

- **Polar Partitioning**: Space is divided into frustums $\mathcal{F}_{i,j,k}$ bounded by range (r), azimuth (θ), and elevation (ϕ).
- **Occupancy Query**: Compresses point cloud into latent vectors and reconstructs via continuous interpolation for occupancy $O(\mathbf{q})$.



Benefits:

- **Geometric Alignment**: Matches sensor ray-scanning, preserving spatial regularity.
- **Physically Grounded**: Facilitates learning occlusion relationships and provides consistent conditioning for radar spectrum.

Order-Invariant Latent Encoding

Problem: Point clouds are unordered sets. Traditional encoders may produce order-sensitive latents, causing inconsistent noise prediction targets and unstable diffusion training.

Hybrid Query Strategy: We design a token encoding scheme using both **static** and **dynamic** queries to ensure consistency:

- **Static Queries** ($\mathbf{Q}_s$): Fixed learned tokens acting as stable anchors to maintain consistent ordering.
- **Dynamic Queries** ($\mathbf{Q}_d$): Derived from input $\mathbf{P}$ to capture geometry-specific features via cross-attention.

Mechanism:

$$\mathbf{Q}_{enc} = \mathrm{Proj}(\mathbf{Q}_s + \mathrm{CrossAttn}(\mathbf{Q}_d, \mathbf{P}))$$

The final query $\mathbf{Q}_{enc}$ combines fixed structure with geometry-aware features

Result: This ensures order invariance, leading to a stable optimization trajectory and improved generalization for the diffusion model.

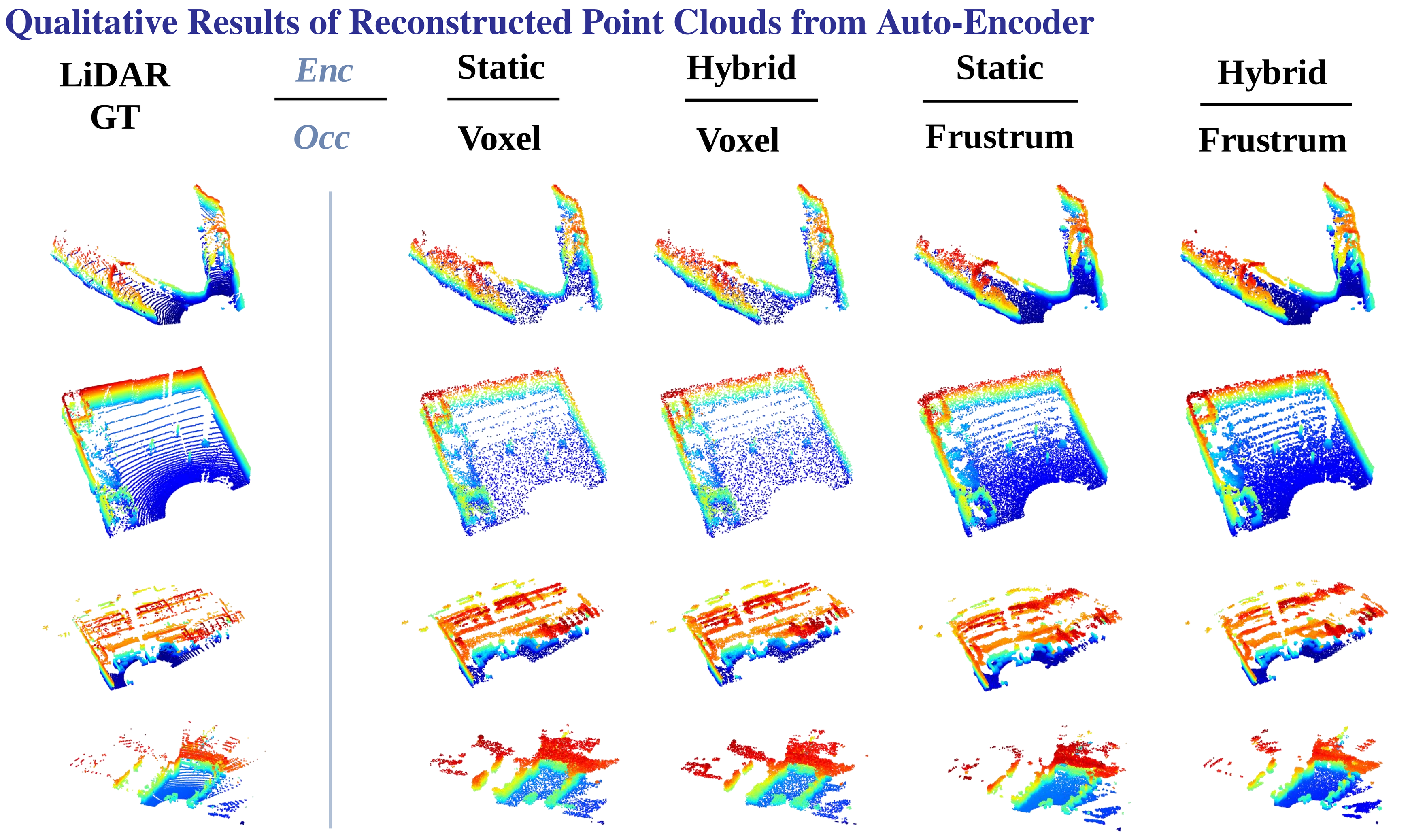
Main Experiments

We evaluate our framework on the ColoRadar and SDDiff datasets, both featuring synchronized radar spectrums and LiDAR point clouds. This presentation primarily highlights performance on the ColoRadar dataset.

LiDAR Auto-Encoder Performance Results

Scene	Encoder / Occupancy	Hybrid Voxel	Sample Frustum	Static Frustum	Hybrid Frustum
Aspen Lab	CD↓	0.133	**0.082**	0.090	0.088
	EMD↓	0.132	**0.083**	0.089	0.087
Hallways	CD↓	0.166	**0.094**	0.118	0.112
	EMD↓	0.162	**0.095**	0.118	0.112
ARPG Lab	CD↓	0.160	0.082	0.104	**0.081**
	EMD↓	0.155	0.083	0.104	**0.080**

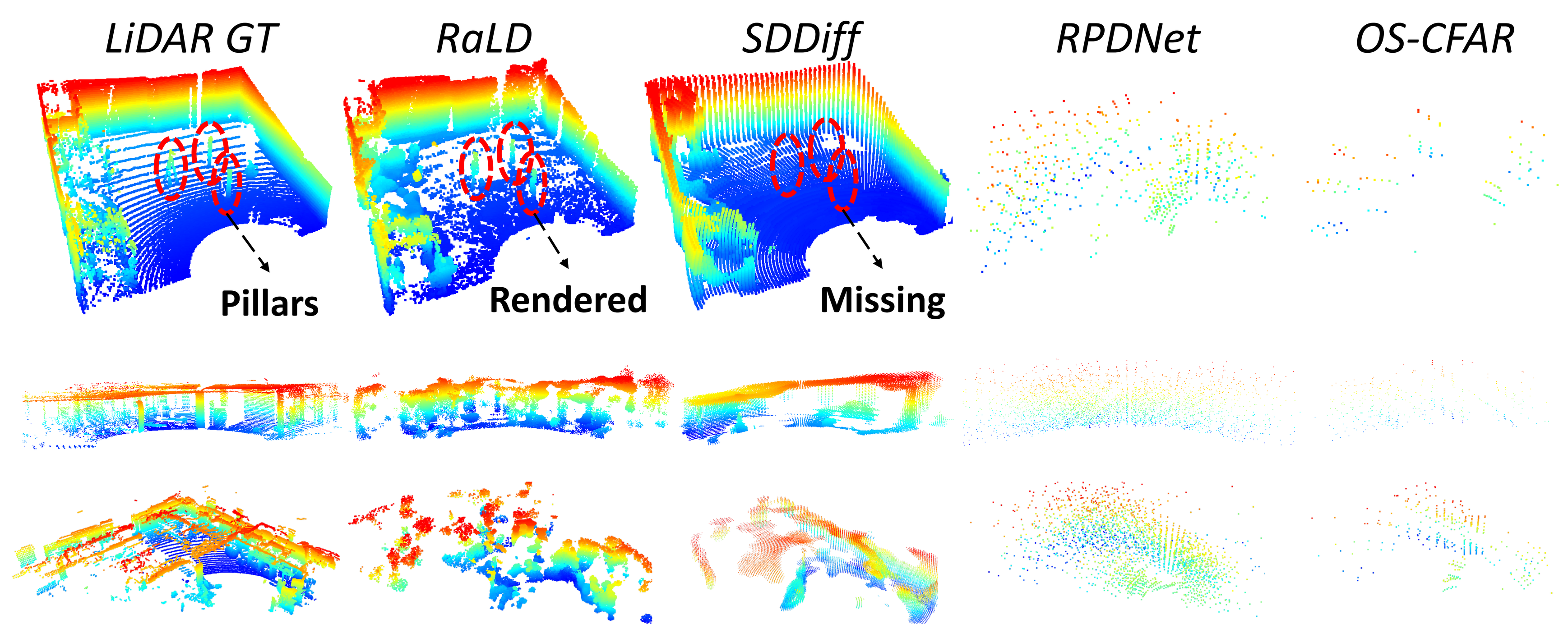
Qualitative Results of Reconstructed Point Clouds from Auto-Encoder



End-to-End Radar Point Cloud Generation Results

Model	Aspen Lab CD↓	Aspen Lab EMD↓	Hallways CD↓	Hallways EMD↓	ARPG Lab CD↓	ARPG Lab EMD↓
OS-CFAR[1]	1.175	1.342	1.098	1.387	1.076	1.163
RPDNet[2]	0.874	0.587	0.793	0.664	0.823	0.512
SDDiff[3]	0.385	0.386	0.581	0.603	0.497	0.505
RaLD	**0.339**	**0.356**	**0.576**	**0.515**	**0.488**	**0.450**

Qualitative Results of Generated Radar Point Clouds:



Ablation & Additional Results

Ablation Studies

Var.	Radar Enc.	CFAR Init	Aspen Lab CD↓	Aspen Lab EMD↓	Hallways CD↓	Hallways EMD↓	ARPG Lab CD↓	ARPG Lab EMD↓
(a)	w/o	w/	0.596	0.638	0.723	0.647	0.659	0.628
(b)	w/	w/o	0.348	0.381	0.586	0.545	0.535	0.547
(c)	w/	w/	**0.339**	**0.356**	**0.576**	**0.515**	**0.488**	**0.450**

"w/" and "w/o" indicate the presence and absence of each component.

Scene	Encoder Query / Occ. Type	Hybrid Voxel	Sample Frustum	Static Frustum	Hybrid Frustum
Aspen Lab	CD↓	0.397	0.366	0.390	**0.339**
	EMD↓	0.519	0.422	0.412	**0.356**
Hallways	CD↓	0.695	**0.562**	0.633	0.576
	EMD↓	0.770	0.566	0.540	**0.515**
ARPG Lab	CD↓	0.609	**0.475**	0.564	0.488
	EMD↓	0.766	0.511	0.513	**0.450**

Generation performance on different auto-encoder queries.

Model Scalability

Scale	Param (M)	Aspen Lab CD↓	Aspen Lab EMD↓	Hallways CD↓	Hallways EMD↓	ARPG Lab CD↓	ARPG Lab EMD↓
Depth = 12	101.87	0.367	0.388	0.591	0.518	0.493	**0.448**
Depth = 18	142.82	0.361	0.383	0.590	0.520	0.500	0.449
Depth = 24	183.77	**0.339**	**0.356**	**0.576**	**0.515**	**0.489**	0.450

Paper

Project

Code

References:

[1] Cheng. Y. *et al.* A novel radar point cloud generation method for robot environment perception. *IEEE TRO* 22.

[2] Richards, M. A.; et al. 2005. *Fundamentals of radar signal processing*, volume 1. McGraw-Hill, New York.

[3] Wang, S.*et al.*. 2025. SDDiff: Boost Radar Perception via Spatial-Doppler Diffusion *IJCAI* 25.