PHYSICS-GROUNDED NEURAL RADAR FIELDS FOR HIGH-FIDELITY NOVEL VIEW SYNTHESIS

Anonymous authors
Paper under double-blind review

ABSTRACT

Radar-based novel view synthesis (NVS) supports robust scene understanding and radar simulation for autonomous driving, particularly under poor illumination and adverse weather. However, most existing methods simplify radar measurement formation and primarily model scene responses, leaving radar-specific artifacts insufficiently represented. This limits the fidelity of synthesized data to real radar measurements. The resulting synthetic-to-real gap can hinder the use of synthesized data in downstream tasks. To narrow this gap, we propose *PhyRF*, a physics-grounded neural radar field framework for high-fidelity radar-based NVS. For points sampled along each radar beam, the neural radar fields predict scene occupancy, radar reflectivity, and artifact contribution. We propose a differentiable renderer grounded in radar measurement formation that combines these predictions according to their distinct physical roles to synthesize measurements. To help disentangle scene geometry from artifacts during training, we further introduce LiDAR measurements as additional supervision by optimizing a shared geometry field. Extensive experiments demonstrate that PhyRF synthesizes radar measurements with higher fidelity to real observations than existing methods.

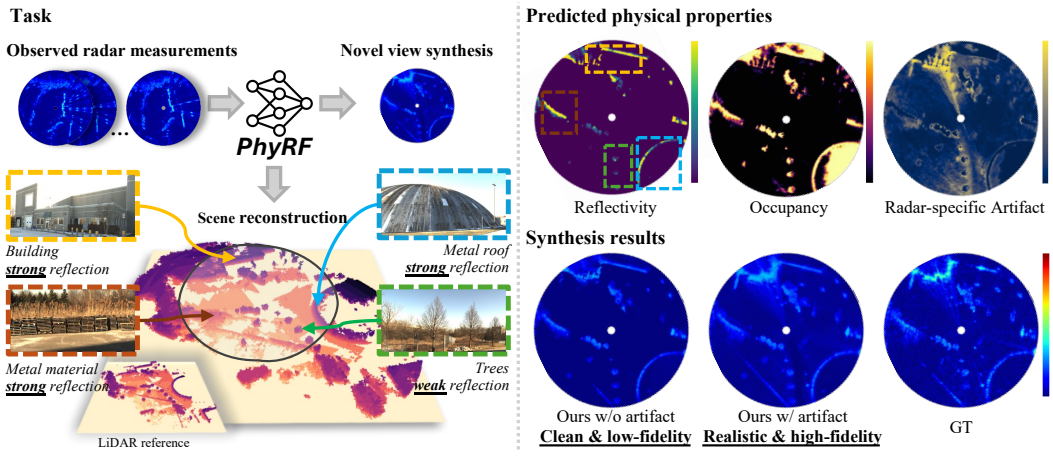

Figure 1: **Overview of PhyRF for radar novel view synthesis and scene reconstruction.** PhyRF predicts physically meaningful occupancy, reflectivity, and radar-specific artifacts. Rendering with artifacts produces realistic, high-fidelity radar measurements close to the ground truth, while disabling artifacts retains only radar responses from physical objects.

1 INTRODUCTION

Recent advances in computer vision have enabled high-fidelity 3D scene reconstruction and novel view synthesis (NVS), supporting autonomous driving tasks such as data simulation (Yang et al., 2023b; Zhao et al., 2025), motion planning (Hu et al., 2022; 2023), and scene understanding (Sadat et al., 2020; Muhammad et al., 2022). NVS has been widely studied using camera (Ost et al., 2021; Turki et al., 2023; Wang et al., 2023) and LiDAR data (Tao et al., 2024a; Zheng et al., 2024),

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.

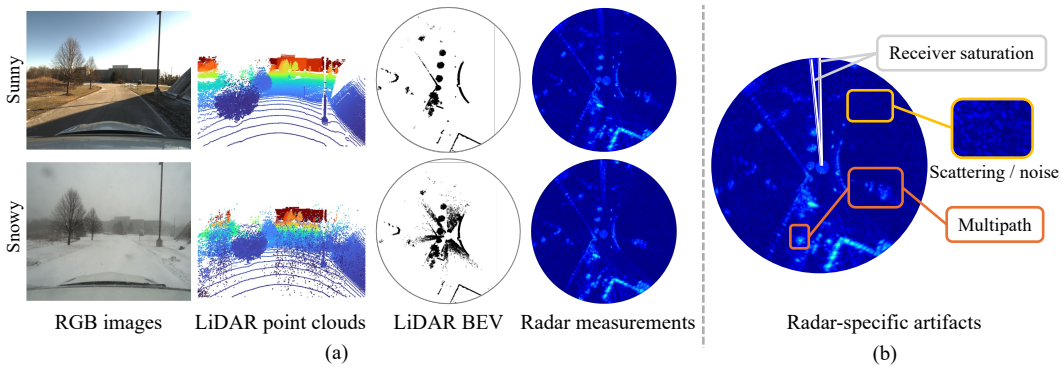

Figure 2: (a) Snowflake returns degrade LiDAR point clouds, while radar measurements contain radar-specific artifacts, making the two modalities complementary under challenging conditions. (b) Radar measurements contain artifacts such as scattering, multipath reflections, noise and receiver saturation, which are not directly associated with scene geometry.

but radar sensing remains relatively underexplored despite its extensive use in autonomous driving systems. Radar offers unique advantages, such as cost efficiency and robustness to adverse weather and low light (Venon et al., 2022; Fan et al., 2024; 2026). For example, Figure 2(a) shows that radar measurements are not affected in snowy conditions, but LiDAR point clouds are corrupted by snowflakes. Nevertheless, radar-based NVS is particularly challenging because radar measurements contain radar-specific artifacts as shown in Figure 2(b) During radar measurement formation, in addition to responses from physical objects, scattering, multipath reflections, receiver saturation, and hardware noise can introduce these undesired artifacts (Richards et al., 2005).

Most existing methods(Borts et al., 2024; Zhang et al., 2026) focus on modeling scene responses while leaving radar-specific artifacts unmodeled. They commonly apply a global threshold to suppress low-intensity regions and treat the retained high-intensity regions as responses from physical objects in the scene. As a result, their synthesized radar measurements lack the radar-specific artifacts present in real observations and tend to appear overly clean. This limits synthesis fidelity and can introduce a synthetic-to-real gap when the generated data are used for downstream tasks.

Moreover, existing methods (Borts et al., 2024; Zhang et al., 2026; Kung et al., 2025) simplify radar rendering by treating each measurement intensity as the received power from a corresponding scene point and predicting it independently. However, in real radar sensing, radar measurements are not formed point-wise. A finite-width radar beam can illuminate multiple targets, and their echoes are superimposed in the received signal. A discrete Fourier transform (DFT) converts this combined signal into a frequency spectrum. The resulting magnitude spectrum forms the radar measurement along the beam: each frequency bin maps to a range, and its magnitude gives the measured intensity at that range. Therefore, point-wise rendering cannot capture how multiple scene points jointly shape the frequency spectrum, limiting rendering fidelity.

To overcome these limitations, we propose *PhyRF*, a physics-grounded neural radar field framework for high-fidelity radar-based NVS. For each radar beam, we sample scene points and query the neural radar fields to predict three physical properties: scene occupancy, radar reflectivity, and artifact contribution. We then introduce a differentiable spectral renderer grounded in radar measurement formation to combine these predictions according to their distinct physical roles. The renderer computes the complex frequency response of each sampled point and sums the responses to form the radar measurement. To further separate scene responses from radar-specific artifacts, we introduce LiDAR as additional supervision for scene geometry. Directly supervising scene occupancy with LiDAR can mislead the learned geometry when LiDAR measurements are corrupted by snowflakes. To mitigate this issue, radar and LiDAR jointly optimize the shared geometry field through their respective reconstruction objectives, encouraging the learned geometry to capture scene structures consistently supported by both modalities. Radar responses that cannot be explained by the shared scene geometry are automatically modeled by the artifact field.

Our contributions are three-fold:

- We propose *PhyRF*, a physics-grounded neural radar field framework that jointly models scene responses and radar-specific artifacts for high-fidelity radar-based NVS. To disentangle scene responses from artifacts, radar and LiDAR jointly optimize a shared scene geometry field, while a separate artifact field captures the remaining radar responses.
- We introduce a differentiable spectral renderer that directly constructs the frequency spectrum forming the radar measurement. It computes and accumulates the complex frequency responses of sampled scene points to obtain this spectrum.
- We conduct extensive experiments demonstrating that PhyRF achieves state-of-the-art performance in high-fidelity radar measurement synthesis and produces physically meaningful estimates of scene occupancy, radar reflectivity, and radar-specific artifacts.

2 Related work

Neural Radiance Fields. Neural Radiance Fields (NeRFs) (Mildenhall et al., 2021) represent a 3D scene as a continuous neural field and synthesize novel views through differentiable volume rendering (Kajiya & Von Herzen, 1984). Subsequent works improve the efficiency and rendering fidelity of neural fields (Müller et al., 2022; Fridovich-Keil et al., 2023; Martin-Brualla et al., 2021; Barron et al., 2021). Early NeRF methods mainly focus on objects or small, bounded scenes, while later works extend neural fields to unbounded and large-scale urban environments (Barron et al., 2022; Rematas et al., 2022; Wang et al., 2023). Beyond RGB images, neural fields have also been extended to other sensing modalities. Several works jointly model camera and LiDAR observations (Tao et al., 2024b; Yang et al., 2023a), while others develop neural fields specifically for LiDAR measurements (Tao et al., 2024a; Huang et al., 2023; Xue et al., 2024). These studies demonstrate the applicability of neural fields to different sensing modalities.

Novel View Synthesis Using Radar. Radar novel view synthesis has been explored under different scene scales and sensing settings. Some works (Huang et al., 2024; Lu et al., 2026; Armouti et al., 2026) focus on small-scale or object-level radar reconstruction and synthesis settings. Other studies target large-scale outdoor environments. NeuRadar (Rafidashti et al., 2025) models urban scenes from sparse radar point clouds together with camera and LiDAR observations. In contrast, other large-scale works operate directly on higher-resolution range–azimuth maps acquired by spinning radars (Borts et al., 2024; Zhang et al., 2026; Kung et al., 2025; Borts et al., 2026). These measurements retain dense radar responses across range and azimuth. Our work follows this large-scale spinning-radar setting for scene reconstruction and novel view synthesis.

3 Preliminary

Neural Radiance Fields. Neural Radiance Fields (NeRFs) (Mildenhall et al., 2021) represent a 3D scene as a continuous function $(\sigma, \mathbf{c}) = F_\theta(\mathbf{x}, \mathbf{d})$, which maps a spatial location $\mathbf{x}$ and a viewing direction $\mathbf{d}$ to a volume density σ and an RGB color $\mathbf{c}$. For a pixel, we cast a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$, where $\mathbf{o}$ is the ray origin, $\mathbf{d}$ is a unit direction, and t is the distance from the origin. We sample N ordered distances t_i along the ray, yielding sample locations $\mathbf{x}_i = \mathbf{r}(t_i)$ and interval lengths $\delta_i = t_{i+1} - t_i$. The density σ_i at the i-th sample is converted into the corresponding opacity as:

$$\alpha_i = 1 - \exp(-\sigma_i \delta_i). \tag{1}$$

The opacity $\alpha_i \in [0, 1]$ represents the local termination probability within the i-th interval. The rendered pixel color is then obtained by accumulating the sample colors $\mathbf{c}_i$:

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i \mathbf{c}_i, \qquad T_i = \prod_{j=1}^{i-1}(1 - \alpha_j) = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right). \tag{2}$$

where T_i represents the fraction of radiance transmitted through all preceding intervals before reaching the i-th sample. Similarly, LiDAR neural fields (Tao et al., 2024a) represent a 3D scene as a continuous function $(\sigma, s, p) = F_\theta^{\mathrm{L}}(\mathbf{x}, \mathbf{d})$, which maps a spatial location $\mathbf{x}$ and a beam direction $\mathbf{d}$ to a volume density σ, return intensity s, and ray-drop probability p. The NeRF volume-rendering formulation described above is directly applied to LiDAR rays, where σ_i is converted into opacity

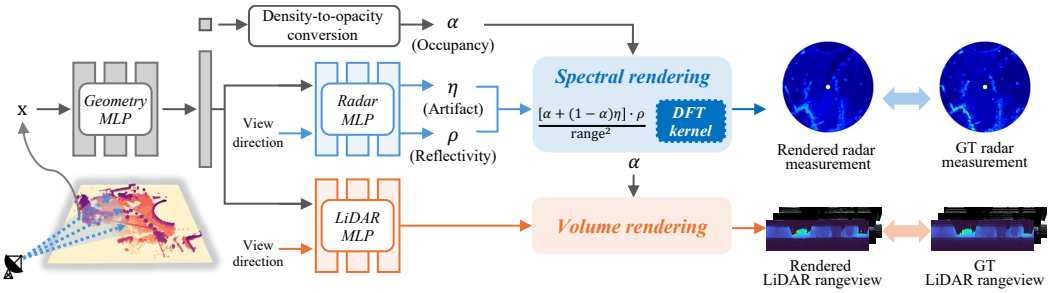

Figure 3: **Overview of PhyRF**. PhyRF represents the scene as a continuous neural field that predicts scene occupancy, radar reflectivity, and radar-specific artifacts. A spectral renderer combines these predictions to synthesize radar measurements, while a LiDAR branch provides geometric constraints to disentangle scene geometry from radar-specific artifacts.

α_i and transmittance T_i following the same definitions. The LiDAR depth, intensity, and ray-drop probability are then rendered as:

$$\hat{D}(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i t_i, \qquad \hat{I}(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i s_i, \qquad \hat{P}(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i p_i. \tag{3}$$

FMCW Radar Measurement Formation. An FMCW radar transmits a linear frequency-modulated chirp with slope S. For a static target at range R, the received echo is delayed by $\tau = 2R/c$, where c is the speed of light. After dechirping and low-pass filtering, the resulting intermediate-frequency (IF) signal can be expressed as

$$s_{\text{if}}(t) = A_{\text{if}} \exp\left(j(\phi_0 + 2\pi f_b t)\right), \tag{4}$$

where A_{if} is the IF signal amplitude, ϕ_0 is a time-independent phase term, and f_b is the beat frequency. For a fixed radar configuration, the beat frequency encodes the target range, and the IF signal amplitude is determined by the target reflectivity ρ and range:

$$f_b \propto R, \qquad A_{\text{if}} \propto \frac{\rho}{R^2}. \tag{5}$$

The IF signal is sampled at a sampling rate f_s, resulting in the discrete signal $x[n] = s_{\text{if}}(n/f_s)$, where $n = 0, \ldots, N_{adc} - 1$. A Hamming window $w[n]$ is applied to suppress sidelobes and reduce spectral leakage caused by finite-duration sampling. An N_{dft}-point DFT then transforms the windowed signal into a frequency spectrum, where different frequency bins correspond to different target ranges. For a target with beat frequency f_b, the resulting DFT response is

$$X[m] = A_{\text{if}} e^{j\phi_0} \sum_{n=0}^{N_{adc}-1} w[n] \exp\left[j2\pi \left(\frac{f_b}{f_s} - \frac{m}{N_{\text{dft}}}\right) n\right]. \tag{6}$$

A detailed derivation is provided in Appendix A.1.

4 Method

Given radar measurements and their corresponding sensor poses, PhyRF aims to reconstruct the scene as a continuous implicit representation based on neural fields. Furthermore, given a novel radar pose, PhyRF performs neural rendering to synthesize the novel radar measurement. Each radar measurement consists of multiple azimuth beams, with each beam recording a range profile. For each radar beam, we sample points and query the neural radar field to obtain scene opacity, radar reflectivity, and radar-specific artifacts. The proposed spectral renderer then combines these predictions across the beam to form the range profile according to the radar measurement formation process. During training, LiDAR observations provide additional geometric constraints to reduce ambiguity between scene geometry and radar-specific artifacts. Figure 3 illustrates the overall pipeline of the proposed method.

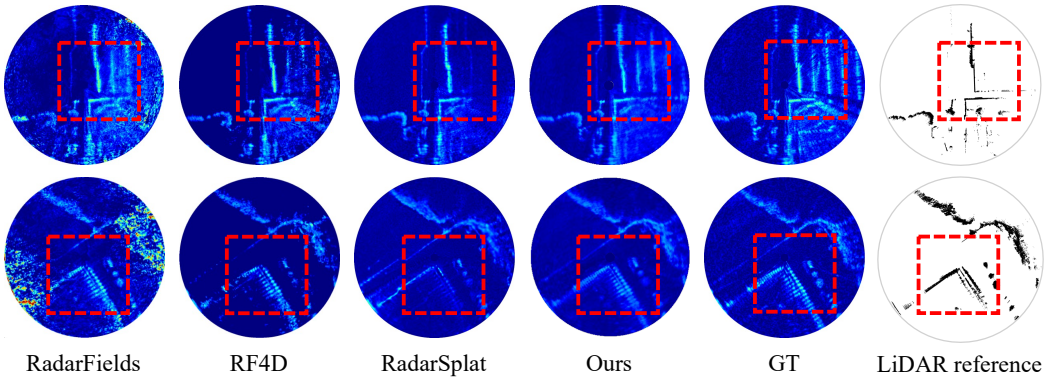

Figure 4: **Qualitative comparison of radar novel view synthesis.** Our method produces high-fidelity radar measurements that closely match the ground truth, preserving both scene responses and radar-specific artifacts. In contrast, RadarFields and RF4D do not synthesize radar-specific artifacts well, with RF4D producing overly clean results, while RadarSplat fails to faithfully recover the scene responses.

4.1 Neural Radar Fields

Each sampled 3D location $\mathbf{x}$ is encoded by a multi-resolution hash grid $\mathcal{H}$ (Müller et al., 2022) and processed by a geometry MLP f_{geo} to obtain the density σ and a spatial feature χ, i.e., $(\sigma, \chi) = f_{\text{geo}}(\mathcal{H}(\mathbf{x}))$. The density is converted into scene opacity α following Equation 1. The spatial feature χ is then concatenated with the encoded radar view direction and passed to f_{radar} to predict direction-dependent radar reflectivity ρ and artifact coefficient η, i.e., $(\rho, \eta) = f_{\text{radar}}(\chi, \mathcal{S}(\mathbf{d}^{\text{r}}))$, where $\mathcal{S}$ denotes the Spherical Harmonics (SH) encoding (Yu et al., 2021) of the view direction. We use ReLU for the non-negative reflectivity ρ and sigmoid for η.

4.2 Spectral Rendering

A spinning radar records a range profile for each azimuth beam. For each radar beam, we additionally sample an $N_s \times N_s$ grid of sub-beams within its horizontal and vertical beamwidths. Along each sub-beam, N_r points are uniformly sampled over the radar range, yielding sampled locations $\mathbf{x}_{q,i}$, where $q \in 1, \ldots, N_s^2$ and $i \in 1, \ldots, N_r$. The neural radar fields are queried at each sampled location to obtain scene opacity $\alpha_{q,i}$, radar reflectivity $\rho_{q,i}$, and artifact coefficient $\eta_{q,i}$. We then compute the probability of generating a radar response:

$$\epsilon_{q,i} = \alpha_{q,i} + (1 - \alpha_{q,i})\eta_{q,i}, \tag{7}$$

where $\alpha_{q,i}$ represents the scene-supported response probability and $(1 - \alpha_{q,i})\eta_{q,i}$ represents the additional artifact response probability. Hence, we can compute the reflected signal's amplitude at each range across all sub-beams.

$$A_i = \frac{1}{N_s^2} \sum_{q=1}^{N_s^2} \frac{\rho_{q,i}\epsilon_{q,i}}{r_i^2}, \tag{8}$$

where $\frac{1}{r_i^2}$ accounts for range-dependent amplitude attenuation. Following the FMCW measurement formation described in the preliminary, the range r_i determines the beat frequency and thus its location in the DFT spectrum. For an N_{dft}-point DFT, the corresponding range-bin spacing is $\Delta R_{\text{bin}} = cf_s/(2SN_{\text{dft}})$. Therefore, the i-th sampled range corresponds to the continuous DFT-bin coordinate $\kappa_i = r_i/\Delta R_{\text{bin}}$. Substituting this relation into the DFT response in Equation 6, and setting the phase term to zero in our renderer, gives

$$X_i[m] = A_i K(m, \kappa_i), \tag{9}$$

where the windowed DFT kernel is defined as

$$K(m, \kappa_i) = \sum_{n=0}^{N_{\text{adc}}-1} w[n] \exp\left(j2\pi \frac{\kappa_i - m}{N_{\text{dft}}} n\right). \tag{10}$$

Table 1: Quantitative comparison across Sunny, Snow, Rain, and Night scenes. Higher PSNR and SSIM indicate better performance.

	Sunny		Snow		Rain		Night	
Method	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
RadarFields	22.67	0.3269	22.29	0.3587	22.37	0.3132	24.52	0.3477
RF4D	23.69	0.2153	24.02	0.2637	23.70	0.1761	24.93	0.2056
RadarSplat	<u>25.68</u>	<u>0.5003</u>	<u>25.59</u>	<u>0.5163</u>	<u>26.54</u>	<u>0.4941</u>	<u>26.67</u>	<u>0.5121</u>
Ours	**26.65**	**0.5105**	**26.77**	**0.5223**	**27.07**	**0.5028**	**27.51**	**0.5273**

Here, m denotes the output range-bin index and N_{adc} is the number of sampled IF signal points. The kernel characterizes the contribution of a response located at the continuous range-bin position κ_i to the output bin m. In real radar sensing, echoes from multiple objects are first superimposed in the time-domain IF signal before the DFT is applied. Because the DFT is linear, this is equivalent to summing their individual DFT responses in the complex domain:

$$W[m] = \sum_{i=1}^{N_r} X_i[m] = \sum_{i=1}^{N_r} A_i K(m, \kappa_i). \tag{11}$$

The rendered range profile is finally obtained in the log-magnitude domain:

$$\widehat{Y}[m] = 20 \log_{10} \left(|W[m]| \right). \tag{12}$$

4.3 LiDAR-Guided Geometry Learning

The neural radar field and spectral renderer alone may lead to a trivial solution in which the model drives scene opacity α toward zero and explains most radar measurements using the artifact branch. This ambiguity arises because both α and the artifact coefficient η contribute to the rendered radar response, making scene responses and radar-specific artifacts difficult to disentangle.

LiDAR provides a complementary cue for resolving this ambiguity because it does not contain radar-specific artifacts. However, LiDAR measurements are not always a reliable direct supervision signal for scene occupancy, especially under snowy conditions where snowflake returns can corrupt the measurements. We therefore avoid treating LiDAR as ground-truth geometry and instead use LiDAR reconstruction to constrain the same geometry field used for radar reconstruction. Specifically, a LiDAR head f_{lidar} takes the spatial feature χ from f_{geo} and predicts LiDAR-specific attributes. These attributes are rendered together with the shared scene opacity α using the volume rendering formulation in Equation 3. As a result, the geometry field is optimized to explain observations from both sensing modalities, encouraging α to capture scene structures consistently supported by radar and LiDAR rather than modality-specific returns. This suppresses snow-induced LiDAR returns and radar-specific artifacts from being encoded as shared scene geometry, leaving radar responses unexplained by the shared geometry to be absorbed by the artifact field.

4.4 Training objective

During each iteration, radar beams and LiDAR rays are sampled from their corresponding measurements. The radar branch renders a range profile using the spectral renderer in Section 4.2, while the LiDAR branch renders depth, intensity, and raydrop following Equation 3. To supervise radar synthesis, we minimize the mean squared error (MSE) between the rendered range profile and the corresponding ground-truth range profile. For LiDAR reconstruction, we follow LiDAR-NeRF (Tao et al., 2024a). The final objective combines the radar and LiDAR reconstruction losses:

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{radar}} + \lambda_{\mathrm{lidar}} \mathcal{L}_{\mathrm{lidar}}, \tag{13}$$

where λ_{lidar} balances the contributions of two modalities.

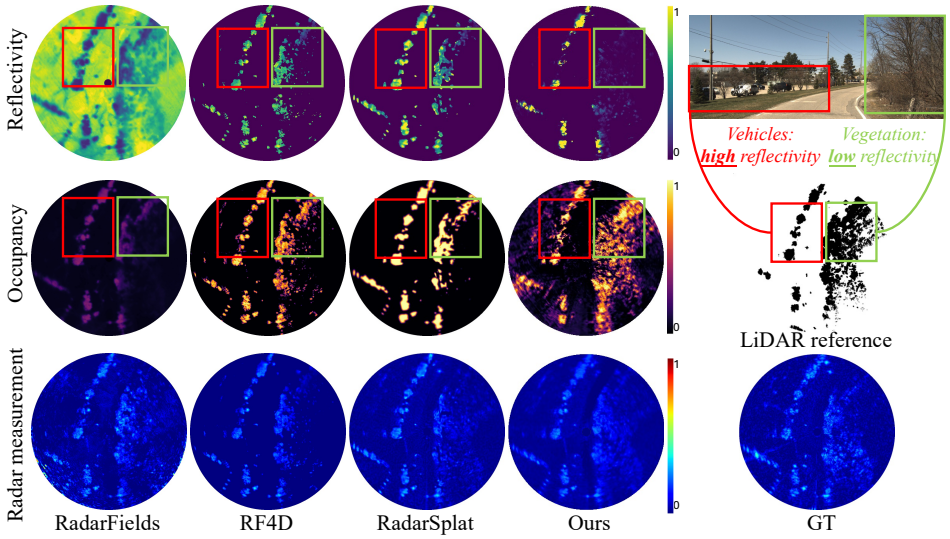

Figure 5: **Qualitative comparison of predicted occupancy and reflectivity.** Our method recovers clearer material-dependent reflectivity, with stronger reflectivity from vehicles and weaker, more diffuse reflectivity from vegetation.

5 Experiments

5.1 Experimental Setup

Datasets & Competing Methods. We compare PhyRF with state-of-the-art radar NVS methods, including RadarFields (Borts et al., 2024), RF4D (Zhang et al., 2026), and RadarSplat (Kung et al., 2025). All methods are evaluated on the same Boreas (Burnett et al., 2023) scenes used in RadarSplat, including 7 sunny, 2 nighttime, 2 rainy, and 2 snowy scenes.

Implementation Details. Each scene is trained independently on a single NVIDIA A6000 GPU. We optimize the model using Adam for 30,000 iterations, with an exponentially decaying learning rate from $5e{-}3$ to $5e{-}4$. At each iteration, we randomly select 4 azimuth beams from a radar scan. For sunny, rainy, and night conditions, λ_{lidar} is set to 1, and it is reduced to 0.2 for snowy scenes where LiDAR measurements are less reliable.

Evaluation metrics. For novel view synthesis, radar measurement quality is evaluated using peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). For occupancy estimation, extracted radar point clouds are evaluated against synchronized single-frame LiDAR point clouds. The evaluation metrics include root mean squared error (RMSE), relative Chamfer distance (RCD), accuracy (ACC), precision, recall, and F1 score. Details on radar point cloud extraction and LiDAR reference construction are provided in Appendix A.6.

5.2 Main Results

Rendering fidelity. As shown in Table 1, our method achieves the best overall performance across Sunny, Snow, Rain, and Night conditions. RadarFields and RF4D ignore radar-specific artifacts by directly removing weak responses before training. This preprocessing simplifies the radar measurements but discards information contained in the original radar spectrum, leading to poor SSIM. RadarSplat explicitly considers radar-specific artifacts and improves performance. Our method further improves the synthesis quality through spectral rendering and LiDAR-guided geometric regularization. Figure 4 compares the synthesized radar measurements with the ground truth. RadarFields and RF4D tend to produce relatively clean predictions, where many radar artifacts are suppressed or blurred. RadarSplat captures more radar-specific effects, but some scene responses are not recovered. In contrast, our method better preserves both the scene responses and radar-specific artifacts observed in the ground-truth measurements, resulting in higher radar measurement fidelity.

Table 2: Ablation of spectral radar rendering (SR), LiDAR-guided geometry learning (LGL), and the artifact field (AF) across four scene conditions. Higher PSNR and SSIM indicate better performance.

			Sunny		Snow		Rain		Night	
SR	LGL	AF	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
	✓	✓	24.28	0.4181	23.94	0.4085	24.56	0.4070	25.50	0.4431
✓			26.01	0.4769	24.76	0.4790	26.60	0.4701	26.95	0.4960
✓		✓	26.41	0.5003	25.80	0.5047	26.82	0.4871	27.40	0.5192
✓	✓		26.40	0.5002	25.83	0.5052	26.95	0.4958	27.35	0.5206
✓	✓	✓	**26.65**	**0.5105**	**26.77**	**0.5223**	**27.07**	**0.5028**	**27.51**	**0.5273**

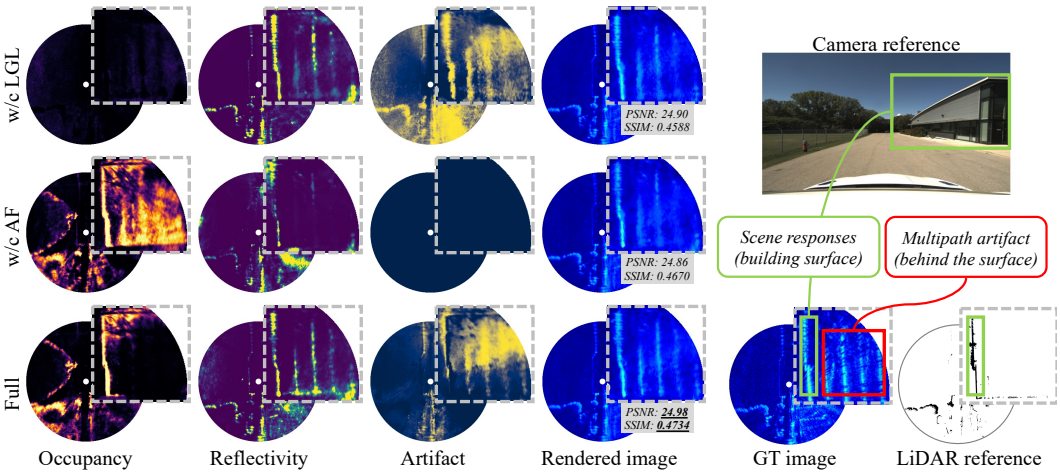

Figure 6: **Ablation of LiDAR-guided geometry learning (LGL) and the artifact field (AF).** Without LGL, the occupancy field is weakly constrained by physical geometry, and the artifact field explains most radar responses. Without AF, multipath responses behind the building surface are incorrectly absorbed into occupancy. The full model better separates responses from the physical scene from radar-specific artifacts.

Material-dependent reflectivity. Figure 5 compares the reconstructed occupancy and reflectivity fields, with LiDAR serving as a reference for scene geometry. All methods recover the major occupied regions, including vehicles and vegetation areas. However, existing methods show limited distinction in reflectivity between these regions. RadarFields further exhibits inconsistency between occupancy and reflectivity, where some occupied regions have very weak or vanishing reflectivity. In contrast, our method reveals clearer material-dependent reflectivity. Vehicles show stronger reflectivity due to their metallic bodies and regular surfaces. Vegetation produces weaker and more diffuse reflectivity because of its irregular structure and distributed scattering.

5.3 Ablation Study

Table 2 evaluates the effects of spectral rendering (SR), LiDAR-guided geometry learning (LGL), and the artifact field (AF). Replacing SR with point-wise rendering leads to a clear performance drop across all scene conditions. This shows that modeling the radar measurement in the spectral domain is important for reproducing realistic radar responses. Adding LGL or AF further improves the reconstruction quality. The full model achieves the best performance across all four conditions in both PSNR and SSIM. Figure 6 further illustrates how LGL and AF affect the learned scene representation. The building surface corresponds to the physical scene, while the responses behind the surface mainly arise from multipath. Without LGL, the occupancy field loses its correspondence with the physical scene, and much of the radar measurement is instead explained by the artifact field. Without AF, the multipath responses behind the building surface are absorbed into the occupancy field and incorrectly interpreted as scene geometry. With both LGL and AF, the building surface is preserved in the occupancy field, while the multipath responses are assigned to the artifact field.

Table 3: Comparison of occupancy estimation results averaged over all non-snowy scenes.

Method	RMSE (m) ↓	RCD ↓	ACC ↑	Precision ↑	Recall ↑	F1 ↑
RadarFields	3.2730	0.1302	68.49	44.51	95.27	59.48
RF4D	4.2305	0.1899	73.68	47.14	**97.16**	62.27
RadarSplat	2.6530	0.0795	69.71	52.75	94.63	67.36
Ours	**0.8430**	**0.0376**	**82.60**	**71.57**	85.47	**77.28**

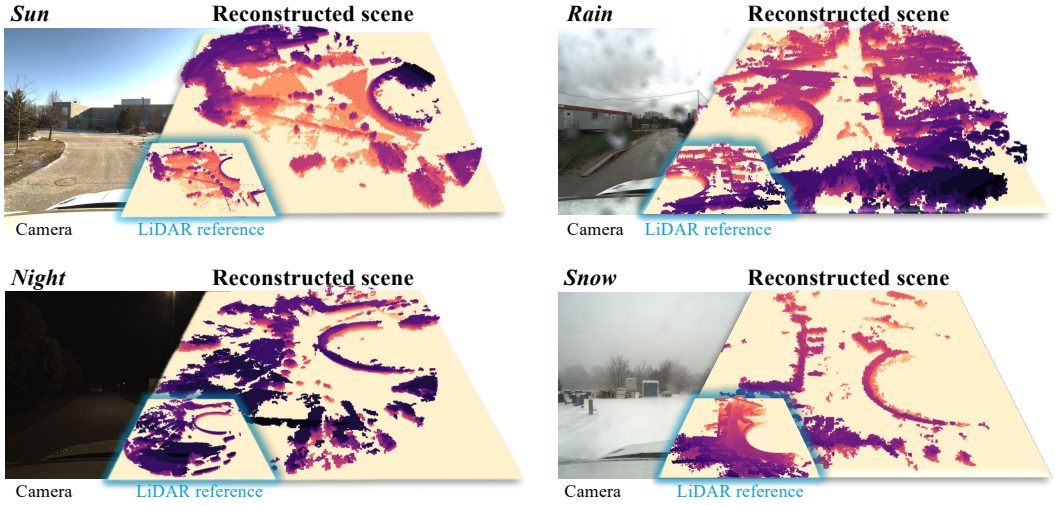

Figure 7: **3D scene reconstruction under diverse weather and lighting conditions.** PhyRF consistently recovers scene geometry across sunny, rainy, nighttime, and snowy environments. Even under heavy snow, it preserves major scene structures despite severe corruption of the LiDAR reference by snowflake returns. Camera images and LiDAR point clouds are provided for reference.

5.4 Scene Reconstruction

Occupancy estimation. As shown in Table 3, our method achieves the best RMSE, RCD, accuracy, precision, and F1 score, indicating substantially better geometric consistency with the LiDAR reference. Its recall is lower than that of the radar-only baselines. Radar beams have much wider angular support than LiDAR beams and can receive returns from targets behind the first LiDAR-visible surface, so radar-only reconstructions naturally cover a larger spatial region and obtain higher recall against the LiDAR reference. In our method, the additional LiDAR-based filtering removes part of this broader radar coverage. This improves geometric precision and suppresses unsupported structures, but also reduces recall.

3D Voxel grid reconstruction. As shown in Figure 7, our method consistently recovers scene geometry across sunny, rainy, nighttime, and snowy environments. Under heavy snow, the LiDAR reference is severely corrupted by snowflake returns and loses substantial scene structure. Our method can still recover scene geometry and synthesize faithful radar novel views.

6 Conclusion

This paper presents **PhyRF**, a physics-grounded neural radar field framework for high-fidelity radar novel view synthesis. A differentiable spectral renderer combines the frequency responses of multiple scene points along each radar beam according to the radar measurement formation process. In addition, radar and LiDAR jointly optimize a shared geometry field, while a separate artifact field captures radar-specific artifacts that do not correspond to the physical scene. Extensive experiments demonstrate PhyRF achieves state-of-the-art performance in high-fidelity radar measurement synthesis and produces physically meaningful estimates of scene occupancy, radar reflectivity, and radar-specific artifacts.

AI USE STATEMENT

In this work, we used generative AI tools to polish the writing and check experimental code for errors and consistency. All AI-assisted content and code-related suggestions were reviewed and verified by the authors. The experimental results, quantitative comparisons, and scientific claims were independently verified by the authors, and generative AI was not used to generate or modify experimental results. The authors take full responsibility for the final content of this work, including all text, results, claims, and artifacts produced with the aid of generative AI.

REFERENCES

Adnan Armouti, Yixuan Gao, and Rajalakshmi Nandakumar. mmir: Frequency-space inverse rendering for 3d millimeter-wave radar adc synthesis. *arXiv preprint arXiv:2608.28913*, 2026.

Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5835–5844. IEEE, 2021.

Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5460–5469. IEEE, 2022.

David Borts, Erich Liang, Tim Broedermann, Andrea Ramazzina, Stefanie Walz, Edoardo Palladin, Jipeng Sun, David Brueggemann, Christos Sakaridis, Luc Van Gool, et al. Radar fields: Frequency-space neural scene representations for fmcw radar. In *ACM SIGGRAPH 2024 Conference Papers*, pp. 1–10, 2024.

David Borts, Julian Ost, Shamik Basu, Tim Broedermann, Andrea Ramazzina, Christos Sakaridis, Mario Bijelic, and Felix Heide. CaRaFe: Camera-radar radiance fields for scene reconstruction. In *ACM SIGGRAPH 2026 Conference Papers*, 2026.

Keenan Burnett, David J Yoon, Yuchen Wu, Andrew Z Li, Haowei Zhang, Shichen Lu, Jingxing Qian, Wei-Kang Tseng, Andrew Lambert, Keith YK Leung, et al. Boreas: A multi-season autonomous driving dataset. *The International Journal of Robotics Research*, 42(1-2):33–42, 2023.

Junqiao Fan, Jianfei Yang, Yuecong Xu, and Lihua Xie. Diffusion model is a good pose estimator from 3d rf-vision. In *European Conference on Computer Vision*, pp. 1–18. Springer, 2024.

Junqiao Fan, Yunjiao Zhou, Yizhuo Yang, Xinyuan Cui, Jiarui Zhang, Lihua Xie, Jianfei Yang, Chris Xiaoxuan Lu, and Fangqiang Ding. M4human: A large-scale multimodal mmwave radar benchmark for human mesh reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 42836–42846, 2026.

Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-planes: Explicit radiance fields in space, time, and appearance. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12479–12488. IEEE, 2023.

Shengchao Hu, Li Chen, Penghao Wu, Hongyang Li, Junchi Yan, and Dacheng Tao. St-p3: End-to-end vision-based autonomous driving via spatial-temporal feature learning. In *European Conference on Computer Vision*, pp. 533–549. Springer, 2022.

Yihan Hu, Jiazhi Yang, Li Chen, Keyu Li, Chonghao Sima, Xizhou Zhu, Siqi Chai, Senyao Du, Tianwei Lin, Wenhai Wang, et al. Planning-oriented autonomous driving. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17853–17862. IEEE, 2023.

Shengyu Huang, Zan Gojcic, Zian Wang, Francis Williams, Yoni Kasten, Sanja Fidler, Konrad Schindler, and Or Litany. Neural lidar fields for novel view synthesis. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 18190–18200. IEEE, 2023.

Tianshu Huang, John Miller, Akarsh Prabhakara, Tao Jin, Tarana Laroia, Zico Kolter, and Anthony Rowe. Dart: Implicit doppler tomography for radar novel view synthesis. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24118–24129. IEEE, 2024.

James T Kajiya and Brian P Von Herzen. Ray tracing volume densities. *ACM SIGGRAPH computer graphics*, 18(3):165–174, 1984.

Pou-Chun Kung, Skanda Harisha, Ram Vasudevan, Aline Eid, and Katherine A Skinner. Radarsplat: Radar gaussian splatting for high-fidelity data synthesis and 3d reconstruction of autonomous driving scenes. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27596–27606. IEEE, 2025.

Jiachen Lu, Hailan Shanbhag, and Haitham Al Hassanieh. Seeing through boxes: Non-line-of-sight 3d reconstruction from radar signals. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1221–1230, June 2026.

Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In *CVPR*, 2021.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

Khan Muhammad, Tanveer Hussain, Hayat Ullah, Javier Del Ser, Mahdi Rezaei, Neeraj Kumar, Mohammad Hijji, Paolo Bellavista, and Victor Hugo C De Albuquerque. Vision-based semantic segmentation in scene understanding for autonomous driving: Recent achievements, challenges, and outlooks. *IEEE Transactions on Intelligent Transportation Systems*, 23(12):22694–22715, 2022.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102:1–102:15, July 2022. doi: 10.1145/3528223.3530127. URL https://doi.org/10.1145/3528223.3530127.

Julian Ost, Fahim Mannan, Nils Thuerey, Julian Knodt, and Felix Heide. Neural scene graphs for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2856–2865, 2021.

Mahan Rafidashti, Ji Lan, Maryam Fatemi, Junsheng Fu, Lars Hammarstrand, and Lennart Svensson. Neuradar: Neural radiance fields for automotive radar point clouds. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2479–2489. IEEE, 2025.

Konstantinos Rematas, Andrew Liu, Pratul P Srinivasan, Jonathan T Barron, Andrea Tagliasacchi, Thomas Funkhouser, and Vittorio Ferrari. Urban radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12932–12942, 2022.

Mark A Richards et al. *Fundamentals of radar signal processing*, volume 1. Mcgraw-hill New York, 2005.

Abbas Sadat, Sergio Casas, Mengye Ren, Xinyu Wu, Pranaab Dhawan, and Raquel Urtasun. Perceive, predict, and plan: Safe motion planning through interpretable semantic representations. In *European Conference on Computer Vision*, pp. 414–430. Springer, 2020.

Tang Tao, Longfei Gao, Guangrun Wang, Yixing Lao, Peng Chen, Hengshuang Zhao, Dayang Hao, Xiaodan Liang, Mathieu Salzmann, and Kaicheng Yu. Lidar-nerf: Novel lidar view synthesis via neural radiance fields. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 390–398, 2024a.

Tang Tao, Guangrun Wang, Yixing Lao, Peng Chen, Jie Liu, Liang Lin, Kaicheng Yu, and Xiaodan Liang. Alignmif: Geometry-aligned multimodal implicit field for lidar-camera joint synthesis. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21230–21240. IEEE, 2024b.

Haithem Turki, Jason Y Zhang, Francesco Ferroni, and Deva Ramanan. Suds: Scalable urban dynamic scenes. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12375–12385. IEEE, 2023.

Arthur Venon, Yohan Dupuis, Pascal Vasseur, and Pierre Merriaux. Millimeter wave fmcw radars for perception, recognition and localization in automotive applications: A survey. *IEEE Transactions on Intelligent Vehicles*, 7(3):533–555, 2022.

Zian Wang, Tianchang Shen, Jun Gao, Shengyu Huang, Jacob Munkberg, Jon Hasselgren, Zan Gojcic, Wenzheng Chen, and Sanja Fidler. Neural fields meet explicit geometric representations for inverse rendering of urban scenes. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8370–8380. IEEE, 2023.

Weiyi Xue, Zehan Zheng, Fan Lu, Haiyun Wei, Guang Chen, and changjun jiang. Geonlf: Geometry guided pose-free neural lidar fields. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 73672–73692. Curran Associates, Inc., 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/86ab6927ee4ae9bde4247793c46797c7-Paper-Conference.pdf.

Ze Yang, Yun Chen, Jingkang Wang, Sivabalan Manivasagam, Wei-Chiu Ma, Anqi Joyce Yang, and Raquel Urtasun. Unisim: A neural closed-loop sensor simulator. In *CVPR*, 2023a.

Ze Yang, Sivabalan Manivasagam, Yun Chen, Jingkang Wang, Rui Hu, and Raquel Urtasun. Reconstructing objects in-the-wild for realistic sensor simulation. In *IEEE International Conference on Robotics and Automation, ICRA 2023, London, UK, May 29 - June 2, 2023*, pp. 11661–11668. IEEE, 2023b. doi: 10.1109/ICRA48891.2023.10160535. URL https://doi.org/10.1109/ICRA48891.2023.10160535.

Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. Plenoctrees for real-time rendering of neural radiance fields. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5732–5741. IEEE, 2021.

Jiarui Zhang, Zhihao Li, Chong Wang, and Bihan Wen. Rf4d:neural radar fields for novel view synthesis in outdoor dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15387–15397, June 2026.

Guosheng Zhao, Chaojun Ni, Xiaofeng Wang, Zheng Zhu, Xueyang Zhang, Yida Wang, Guan Huang, Xinze Chen, Boyuan Wang, Youyi Zhang, et al. Drivedreamer4d: World models are effective data machines for 4d driving scene representation. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12015–12026. IEEE, 2025.

Zehan Zheng, Fan Lu, Weiyi Xue, Guang Chen, and Changjun Jiang. Lidar4d: Dynamic neural fields for novel space-time view lidar synthesis. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5145–5154. IEEE, 2024.

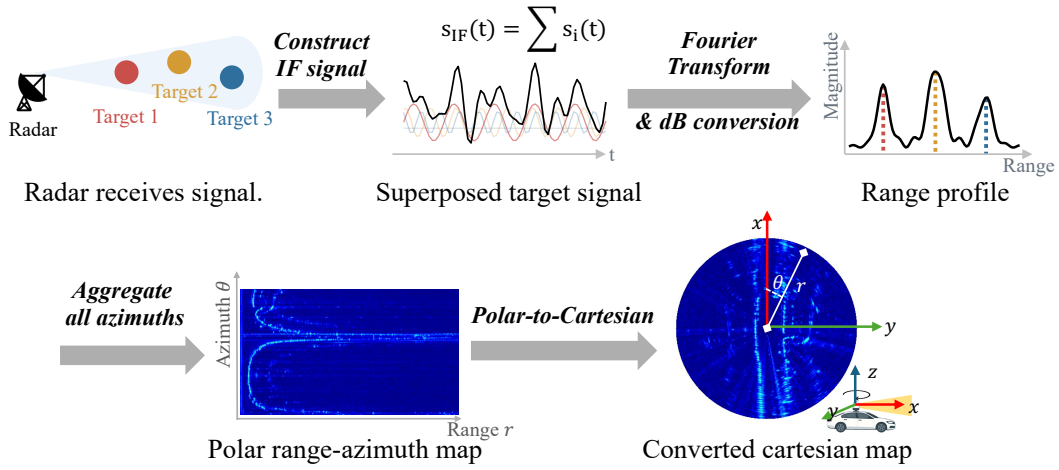

Figure 8: Spinning FMCW radar forward process: from scene to measurement

A Appendix

A.1 Radar Range Profile Derivation.

IF signal construction. An FMCW radar transmits frequency-modulated signals, commonly referred to as chirps. A transmitted chirp can be expressed as

$$s_{\text{tx}}(t) = A_{\text{tx}} \exp\left(j2\pi\left(f_c t + \frac{S}{2}t^2\right)\right), \tag{14}$$

where A_{tx} is the transmitted amplitude, f_c is the carrier frequency, and $S$is the chirp slope. A_{tx}, f_c and S are predefined by the radar configuration and remain fixed for a given sensing setup, independent of the scene.

For a static target at range R, the round-trip propagation delay is $\tau = \frac{2R}{c}$, where c is the speed of light. The received signal is a delayed and attenuated version of the transmitted chirp,

$$s_{\text{rx}}(t) = A_{\text{rx}} \exp\left(j2\pi\left(f_c(t-\tau) + \frac{S}{2}(t-\tau)^2\right)\right), \tag{15}$$

where A_{rx} denotes the received amplitude.

According to the classic radar power equation:

$$P_{rx} = \frac{P_{tx} G^2 \lambda^2 \sigma_{\text{RCS}}}{(4\pi)^3 R^4 L}, \tag{16}$$

where P_{tx} is the transmitted power, G is the antenna gain, λ is the wavelength and L represents system losses. These parameters are determined by the radar system and are independent of the scene. Since signal amplitude is proportional to the square root of power, the received amplitude satisfies:

$$A_{\text{rx}} \propto \sqrt{P_{\text{rx}}} \propto \frac{\rho}{R^2}, \tag{17}$$

where we define the radar reflectivity as $\rho = \sqrt{\sigma_{\text{RCS}}}$ for convenience,.

After dechirping and low-pass filtering, the intermediate-frequency (IF) signal can be expressed as

$$s_{\text{if}}(t) = A_{\text{if}} \exp\left(j(\phi_0 + 2\pi f_b t)\right), \tag{18}$$

where $A_{\text{if}} = A_{\text{tx}} A_{\text{rx}}$, ϕ_0 denotes the constant phase term and $f_b = S\tau$ is the beat frequency.

Therefore, for a fixed radar configuration, the resulting beat frequency for the target is determined by the target range, whereas the IF amplitude depends on both the target reflectivity and range:

$$f_b \propto R, \quad A_{\text{if}} \propto \frac{\rho}{R^2}. \tag{19}$$

Table 4: LiDAR reconstruction averaged over all scenes excluding snow.

Model	Depth RMSE ↓	CD ↓	F-score ↑	Depth SSIM ↑	Intensity MAE ↓
LiDAR-only	5.9205	0.9117	66.15	0.5029	0.07282
Ours	5.9456	0.9448	66.63	0.5024	0.07302

Discrete Fourier Transform. The IF signal is sampled at a sampling rate f_s, resulting the discrete signal $x[n] = s_{\text{if}}(n/f_s)$, where $n = 0, \ldots, N_{adc} - 1$. Then a Hamming window $w[n]$ is applied to the sampled signal to suppress sidelobes and reduce spectral leakage caused by finite-duration sampling. An N_{dft}-point DFT then transforms this signal into a frequency spectrum, where frequency bins correspond to different ranges.

We can then express this frequency spectrum for the target:

$$X[m] = A_{\text{if}} e^{j\phi_0} \sum_{n=0}^{N_{adc}-1} w[n] \exp\left[j2\pi \left(\frac{f_b}{f_s} - \frac{m}{N_{\text{dft}}} \right) n \right]. \tag{20}$$

DFT determines the range-bin (or frequency bin) spacing to be $\Delta R_{\text{bin}} = \frac{c f_s}{2 S N_{\text{dft}}}$.

Let $\kappa = \frac{N_{\text{dft}} f_b}{f_s} = \frac{N_{\text{dft}} \cdot 2SR}{f_s \cdot c} = \frac{R}{\Delta R_{\text{bin}}}$, Equation 20 gives

$$X[m] = A_{\text{if}} K(m, \kappa) = A_{\text{if}} e^{j\phi_0} \sum_{n=0}^{N_{adc}-1} w[n] \exp\left(j2\pi \frac{\kappa - m}{N_{\text{fft}}} n \right), \tag{21}$$

where $K(m, \kappa)$ denotes the windowed DFT kernel evaluated at the continuous range-bin location κ.

Therefore, for Q scatterers along the same radar beam, their range-DFT responses are coherently accumulated in the complex domain:

$$W[m] = \sum_{q=1}^{Q} X_q[m] = \sum_{q=1}^{Q} A_{if,q} K(m, \kappa_q) \tag{22}$$

where $X_k[m]$ denotes its DFT response. The radar magnitude is then expressed in the logarithmic domain as

$$Y[m] = 20 \log_{10} \left(|W[m]| \right). \tag{23}$$

A.2 Additional implementation details

PhyRF is implemented in PyTorch using the tiny-cuda-nn framework for acceleration. Each radar measurement is formed by weighted pooling over a 4×4 grid of subbeams to account for the finite radar beamwidth. Meanwhile, 2,048 LiDAR rays are randomly sampled from the synchronized LiDAR rangeviews for joint reconstruction. The hash-grid encoding contains 16 levels with two features per level. For spherical harmonics (SH) encoding, we use eight bands for radar view direction and four bands for LiDAR view direction.

A.3 LiDAR Novel view synthesis

Table 4 shows that LiDAR reconstruction quality is largely preserved after incorporating radar supervision. Compared with the LiDAR-only model, the full model achieves a slightly higher F-score, while the other metrics show only minor degradation. This behavior is expected, as joint optimization introduces additional radar constraints that may contain noisy or artifact-related responses and thus slightly perturb the learned geometry. Overall, the results indicate that incorporating radar does not substantially compromise the LiDAR NVS performance.

A.4 Artifact-Controllable Radar Rendering

As shown in Figure 9, our method supports artifact-controllable radar rendering. By omitting the artifact contribution during rendering, the model produces a clean radar image that better reflects

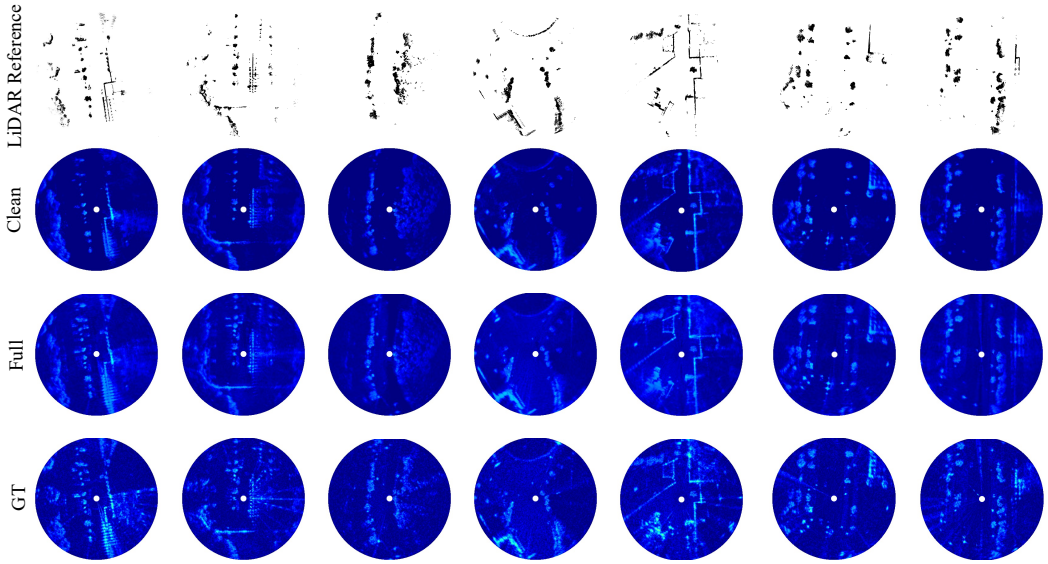

Figure 9: Artifact-controllable radar rendering. Our model can produce either clean radar measurements without radar-specific artifacts or full measurements with these artifacts. The clean rendering better reflects the underlying scene geometry, as indicated by the LiDAR reference, while the full rendering better matches the ground-truth radar measurements.

Table 5: Subbeam sampling ablation. Results are averaged over scenes within each category.

	Sunny		Snow		Rain		Night	
Grid	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
1×1	26.48	0.5008	26.62	0.5125	26.89	0.4927	27.28	0.5153
2×2	26.47	0.4999	26.41	0.5018	26.94	0.4934	27.33	0.5178
3×3	26.62	0.5090	26.48	0.5124	**27.08**	**0.5032**	27.49	**0.5275**
4×4	**26.65**	**0.5105**	**26.69**	**0.5184**	27.07	0.5028	**27.51**	0.5273

the underlying scene geometry, as indicated by the LiDAR reference. Including the artifact contribution recovers radar-specific effects and produces a full rendering that more closely matches the ground-truth radar measurement. The clean rendering can provide clearer geometric observations for downstream perception tasks, while the full rendering is better suited for high-fidelity radar synthesis and simulation.

A.5 Effect of supersampled beams

Table 5 evaluates different subbeam sampling densities. Overall, denser sampling improves rendering quality by better capturing the spatial extent of each radar beam. The 4×4 grid achieves the best PSNR in Sunny, Snow, and Night scenes, and the best SSIM in Sunny and Snow scenes. In Rain, 3×3 performs only marginally better, while its Night SSIM is also nearly identical to that of 4×4. These results show that 4×4 provides the most consistent performance across different conditions. Therefore, 4×4 subbeam sampling is adopted in all experiments.

A.6 Radar point cloud evaluation

We evaluate the synthesized radar point cloud against the corresponding single-frame LiDAR reference for each test frame. For all methods, radar points are extracted from locations with predicted occupancy $\alpha > 0.5$. The z coordinate is ignored, and the LiDAR reference is projected onto the radar plane and restricted to the radar sensing range. For our method, these points are further filtered using the rendered LiDAR point cloud, which is available because our model jointly synthesizes LiDAR and radar measurements.