CVPR #12

RISE: Single Static Radar-based Indoor Scene Understanding

Anonymous CVPR submission

Paper ID 12

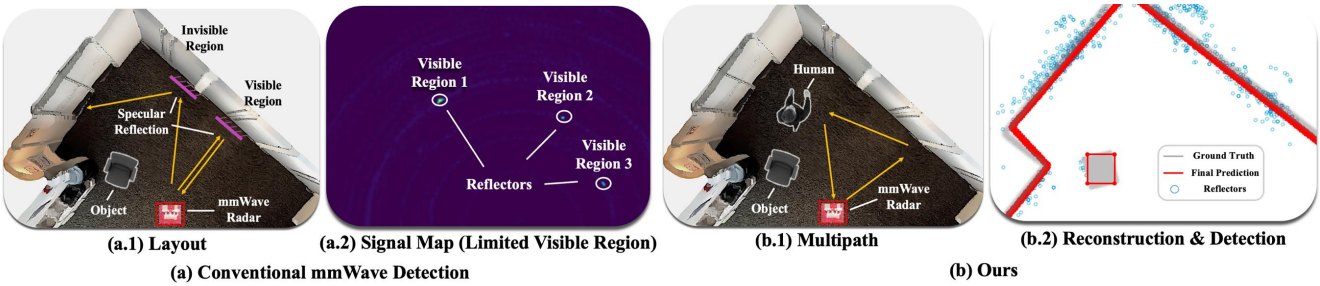

Figure 1. **RISE** is the first single-radar system for object-level indoor scene understanding. Under standard millimeter-wave (mmWave) sensing **(a.1)**, the radar directly observes only a small portion of the environment, producing sparse visible regions **(a.2)** due to strong specular reflections. To overcome this limitation, RISE performs **multipath inversion (b.1)** to exploit indirect reflections and recover surfaces that are not directly visible. Combined with our geometry-aware reconstruction pipeline, RISE enables accurate and complete indoor layout reconstruction and object detection **(b.2)**.

Abstract

Robust and privacy-preserving indoor scene understanding remains a fundamental open problem. While optical sensors such as RGB and LiDAR offer high spatial fidelity, they suffer from severe occlusions and introduce privacy risks in indoor environments. In contrast, millimeter-wave (mmWave) radar preserves privacy and penetrates obstacles, but its inherently low spatial resolution makes reliable geometric reasoning difficult. We introduce RISE, the first benchmark and system for single-static-radar indoor scene understanding, jointly targeting layout reconstruction and object detection. RISE is built upon the key insight that multipath reflections—traditionally treated as noise—encode rich geometric cues. To exploit this, we propose a Bi-Angular Multipath Enhancement that explicitly models Angle-of-Arrival and Angle-of-Departure to recover secondary (ghost) reflections and reveal invisible structures. On top of these enhanced observations, a simulation-to-reality Hierarchical Diffusion framework transforms fragmented radar responses into complete layouts reconstruction and object detection. Our benchmark contains 50,000 frames collected across 100 real indoor trajectories, forming the first large-scale dataset dedicated to single, static, radar-based indoor scene understanding. Extensive experiments show that RISE reduces the Chamfer Distance by 60% (down to 16 cm) compared to the state of the art in mmWave layout reconstruction, and delivers the first mmWave-based object detection, achieving 58% IoU. These results establish RISE as a new foundation for geometry-aware and privacy-preserving indoor scene understanding using a single static radar.

1. Introduction

Indoor scene understanding has long been a fundamental research problem in the vision community, with wide-ranging applications in smart homes [71], virtual reality (VR) [48], and augmented reality (AR) [72]. Classical approaches to this problem rely on optical sensors such as RGB cameras and LiDAR, which suffer from visual occlusions since they cannot image through common obstructions like walls or objects [60, 79] and also raise privacy concerns in smart home and corporate environments [39]. To overcome these challenges, researchers have recently started investigating the use of wireless signals – like WiFi or millimeter-waves (mmWaves) – for indoor scene reconstruction since such signals can penetrate everyday occlusions and are much less privacy intrusive than cameras [3, 4, 13, 15–17, 25, 37, 51, 78, 81]. Despite promising early results, existing wireless solutions either suffer from low resolution – providing only limited patches of the surrounding environment – or require mounting the wireless

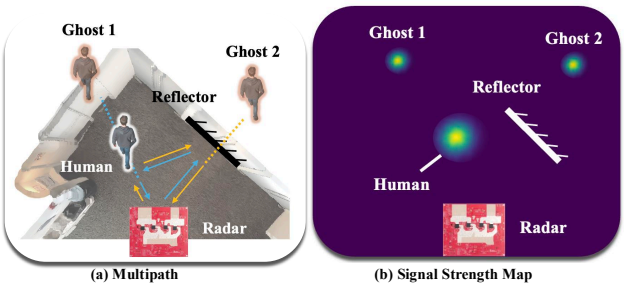

Figure 2. **Multipath-induced Ghosts:** Figure (a) illustrates the scenario in which ghost points appear. Figure (b) presents the corresponding XY heatmap generated by the radar.

sensor on a mobile robot to scan the environment – which increases deployment overhead, making them less desirable [7, 23, 30, 33, 34, 41, 44, 61, 64, 68, 80, 83].

In this paper, we ask the following question: *Can we use a single static mmWave wireless device to enable accurate indoor scene understanding?* A positive answer would enable leveraging a typical wireless setup – such as an existing access point or wireless router already deployed in smart homes – for this purpose. A main challenge to enable indoor layout reconstruction via static mmWave sensing is specularity [55]. Unlike optical sensors that capture diffuse reflections, mmWave signals primarily undergo mirror-like (specular) reflections (Fig.1(a.1)), resulting in only a limited visible region (Fig.1(a.2)). This means that when the sensor transmits a signal toward a surface, such as a wall, the signal reflects away at an angle rather than scattering in all directions. If the reflected signals do not return back to the sensor, the surface remains undetected Fig. 1 (a.2), leading to incomplete and ambiguous reconstructions [14, 21, 22, 27, 29, 35, 65, 67, 76].

To overcome this challenge, we introduce RISE, a novel benchmark and system that leverages human mobility to enable layout reconstruction and object detection using a single static mmWave sensor. When a person moves within the environment, their presence introduces *multipath effects*, where mmWave wireless signals undergo secondary reflections[1] before reaching the receiver, as in Fig.1 (b.1). To understand this better, let's consider a simple scenario shown in Fig.2. In addition to direct reflections from humans, the radar can also capture signals from higher-order reflection paths (e.g. path 1: radar → reflector → human → radar, and path 2: radar → human → reflector → radar.). These signals create multipath *ghost targets* that move in sync with the human (Ghost 1 and Ghost 2 in Fig.2) [26, 28, 36, 38, 42, 62]. While these multipath ghosts are traditionally treated as a noise source [32, 50], we exploit their geometric properties to infer environmental structures. By analyzing the temporal evolution of multipath reflections, RISE estimates a rich set of reflectors (blue dots in Fig.1(b.2)) that encode information about the surrounding layout. RISE then enhances this reconstruction through a hierarchical diffusion-based optimization model [8, 10, 12, 24, 58], filling in the missing details and generating the final reconstruction output (red line in Fig.1(b.2)). Contributions:

- **A new benchmark for radar-based indoor scene understanding.** We collect the first labeled dataset and system for indoor layout understanding using only a *single static mmWave radar*. This dataset provides annotations for structural layouts and object detection, enabling systematic study of radar perception for spatial reasoning.
- **A Bi-Angular Multipath Enhancement (BAME) module.** We introduce a novel signal enhancement technique that exploits the angular diversity of multipath reflections to reveal "ghost" signals corresponding to occluded walls and structures. The BAME module significantly amplifies these weak yet geometrically meaningful signals, offering a strong initial prior for layout reasoning.
- **A Simulation-to-Reality Hierarchical Diffusion (SRHD).** To achieve complete scene interpretation—including both wall layout reconstruction and object detection—we develop a radar simulation engine and train a diffusion-based generative model upon it. This approach bridges the gap between synthetic multipath radar signals and real-world measurements, substantially improving the completeness and accuracy of indoor layout recovery.

To rigorously evaluate our approach, we deploy the proposed system on a TI MMWCAS-RF-EVM Cascade mmWave radar and collect over **100 real-world trajectories** encompassing approximately **50,000 frames**. Quantitatively, **RISE** achieves a mean Chamfer distance of **16 cm** for wall-layout reconstruction and an **IoU of 58 %** for furniture detection—compared to the state-of-the-art baseline [9], which attains only **40 cm** Chamfer error and lacks furniture detection capability. These results establish **RISE** as the first comprehensive framework for mmWave-based indoor scene reconstruction, marking significant *qualitative and quantitative* advances in radar layout understanding.

2. Related Work

2.1. Optical-Based Scene Understanding

Optical methods, including RGB cameras, LiDAR, and their fusion, have been extensively used for scene understanding. Recent RGB-based approaches enable accurate object recognition [6, 19, 40, 43, 57] and semantic segmentation [31, 46, 59, 82] with the help of deep learning, while LiDAR-based methods provide precise depth information for 3D understanding [11, 52, 53, 66, 70, 84–86]. Combining RGB and LiDAR enhances spatial and semantic understanding, benefiting applications such as robotics and augmented reality [2, 18, 45, 69, 87]. Other optical-based meth-

[1] These are in addition to the direct reflections described earlier

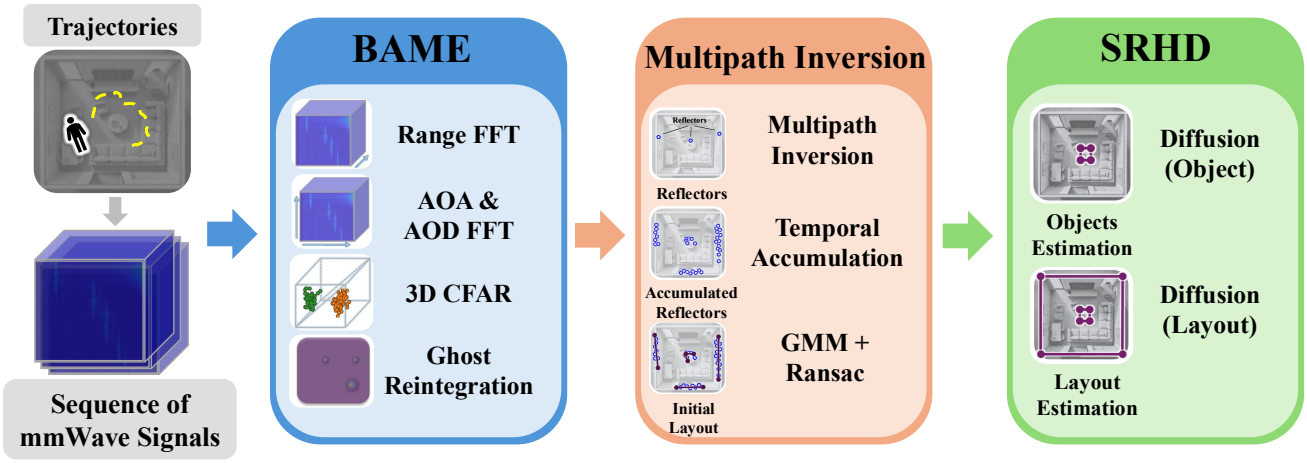

Figure 3. **Pipeline.** Given a sequence of mmWave signals, RISE first applies the Bi-Angular Multipath Enhancement (BAME) module to each frame to recover previously suppressed ghost paths. These enhanced observations are then processed by our multipath inversion module to estimate the underlying reflector geometry for every frame. By aggregating reflector estimates across the entire trajectory, RISE forms an initial, geometry-aware layout hypothesis. This coarse reconstruction is subsequently refined by our Sim2Real Hierarchical Diffusion (SRHD) model, which transforms fragmented radar observations into complete wall layouts and object masks.

ods rely on ambient lighting [63] or laser [5] to reconstruct the scene around the corner. However, all optical methods share limitations, including privacy concerns and occlusion challenges [39, 60, 79].

2.2. Wireless Signal-Based Scene Understanding

Past research utilizing wireless signals for scene reconstruction can be broadly categorized into two approaches. The first approach leverages Wi-Fi signals, where the system reconstructs the environment by analyzing reflections captured from multiple vantage points [75, 80]. However, these approaches either suffer from low resolution or have deployment overhead. The second category relies on mmWave signals, where a mmWave radar is mounted on a mobile agent (i.e. robot) to actively scan the environment and perform reconstruction [20, 47, 49, 54, 56, 73, 74, 77]. While this approach has proven effective in providing accurate scene reconstruction, the requirement for having a mobile radar restricts its applicability. Recent work has demonstrated the possibility of using a static mmWave radar and multipath for layout reconstruction [9]. However, this method has limitations in fully reconstructing the scene due to blind spots in its field of view and inconsistencies in ghost visibility. Moreover, it is primarily effective for reconstructing large surfaces, such as walls, while lacking the ability to identify or differentiate smaller objects like furniture. In contrast, RISE addresses these challenges by leveraging a Bi-Angular Multipath Enhancement module and a Sim2Real Hierarchical Diffusion framework, effectively mitigating blind spot and ghost visibility issues. Additionally, RISE extends reconstruction capabilities beyond walls, enabling the detection of smaller objects within the scene.

3. Benchmark

Dataset Collection. To enable standardized evaluation and comparison, we establish the **RISE-Indoor Benchmark**, the first dataset dedicated to **indoor scene understanding using a single static mmWave radar**. All data was collected using a TI MMWCAS-RF-EVM Cascade mmWave sensor deployed at a fixed height of 1.2 m, synchronized with an Intel Realsense depth camera for ground-truth capture. The benchmark contains over **100 human motion trajectories**, each approximately 30 s long, resulting in **50,000 radar frames** at 20 Hz. Data were gathered across **11 diverse indoor environments**—including offices, corridors, laboratories, and furnished living rooms—under varying furniture arrangements and wall geometries. To ensure coverage diversity, five volunteers walked along random trajectories, producing rich multipath reflections from multiple incidence and reflection angles.

Annotation Protocol. Each radar frame is temporally aligned with its corresponding depth frame via extrinsic calibration and manual verification. We provide two levels of annotation: **(i) Structural Layout Annotations.** Wall boundaries and room contours are manually labeled on top-down projections of the depth data and cross-checked with physical floor-plan measurements. Each frame therefore includes a ground-truth *layout polygon* representing the wall geometry. **(ii) Object Annotations.** Items such as tables, cabinets, and sofas are annotated as axis-aligned bounding boxes in the same 2D coordinate system. These annotations serve as reference for the object detection.

Evaluation Metrics. The benchmark defines two primary tasks: *layout reconstruction* and *object detection*. *Layout Reconstruction.* We report: Chamfer Distance between predicted and ground-truth wall points, reflecting geomet-

ric accuracy. F1-score at a 15 cm tolerance threshold to measure boundary completeness. *Object Detection.* We use: Intersection-over-Union(IoU) between predicted and ground-truth bounding boxes. Dice Coefficient to quantify segmentation consistency.

4. Methodology

In this paper, we introduce RISE, a novel system that leverages mmWave multipath reflections for accurate indoor layout reconstruction and furniture detection. Section 4.1 introduces the preliminary concepts underlying our approach. Section 4.2 provides a detailed geometric analysis for multipath-based layout understanding. Section 4.3 investigates the visibility inconsistency of multipath ghosts and proposes the **Bi-Angular Multipath Enhancement (BAME)** module to enhance ghost signal observability. Finally, Section 4.4 presents a **Sim2Real Hierarchical Diffusion (SRHD)**, which generates synthetic training data and employs a hierarchical diffusion model to reconstruct accurate indoor layouts and detect furniture from radar observations. The whole pipeline is shown in Fig. 3.

4.1. Preliminaries

RF Signal. An RF signal's phase depends on travel distance; sampling captures its amplitude and phase [1]. The sampled signal is

$$s = Ae^{-j2\pi r/\lambda} \tag{1}$$

where r is the traveled distance, λ is the wavelength, and A is the amplitude.

Angle Estimation. For an N-element antenna array, the received signal power from direction θ can be estimated as:

$$P(\theta) = \sum_{n=1}^{N} s_n e^{j2\pi(nd\cos\theta)/\lambda} \tag{2}$$

where s_n is the received signal at the n-th antenna, d is the spacing between antennas, and λ is the wavelength [1]. In practice we use the standard sum-of-phasors/beamforming formulation (Eq. 2) to build range–angle heatmaps which serve as the primary observation for ghost identification and later geometric inference.

4.2. Multipath Inversion

Ghost Target Formation and Identification. As shown in Fig. 4, when a static radar sensor S observes a moving human target H, additional reflections—*ghost targets*—appear due to multipath propagation. We focus on the first- and second-order ghosts, denoted as G_1(following $s \to c_1 \to h \to s$), G_1'(following $s \to h \to c_1 \to s$), G_2(following $s \to h \to c_1 \to h \to s$) and G_2'(following $s \to c_1 \to h \to c_1 \to s$), corresponding to two- and three-bounce reflections [68]. Given the rapid attenuation of higher-order paths, only these four ghosts are considered. Their visibility in the range–angle map depends on scene geometry and signal processing.

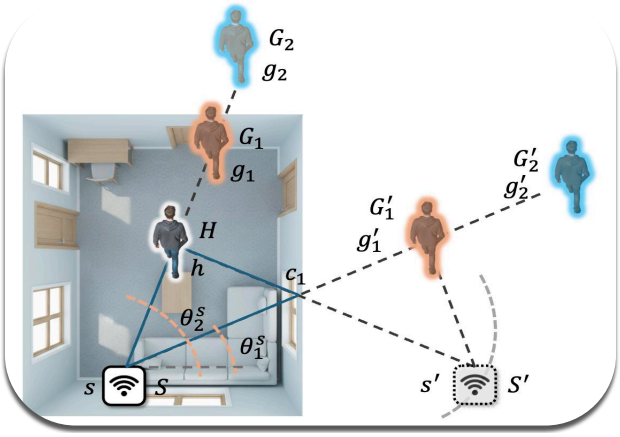

Figure 4. **Multipth Inversion.** Geometric relationships between ghost targets and reflectors.

We begin by using the Constant False Alarm Rate (CFAR) detection method, followed by RANSAC clustering to identify several high power regions ("clusters") within the mmWave image. Then, we can identify which of these clusters correspond to different ghost targets in four steps: H: select clusters with $m_i > \tau \max(m_i)$ where m_i is the received magnitude of the i$^{\text{th}}$ cluster and choose the one with the smallest range. G_1**:** same direction as H but slightly larger range. G_1'**:** same range as G_1 but different direction. G_2'**:** larger range than G_1' and aligned with $\vec{sg_1'}$. Two points are considered identical if their range difference is below 15 cm or angular difference below 15°, consistent with real-world spatial resolution. Detailed thresholds and pseudocode are provided in the Appendix.

Reflector Point Estimation. The reflector point corresponding to G_1' can be estimated as:

$$sc_1 = \frac{2|\vec{sg_1'}|^2 - 2|\vec{sg_1'}||\vec{sh}|}{2|\vec{sg_1'}| - |\vec{sh}|\cos(\theta_2^s - \theta_1^s) - |\vec{sh}|} \tag{3}$$

The proof can be found in the appendix. We first apply a Gaussian Mixture Model (GMM) to group the estimated reflector points into spatially coherent clusters, and then use RANSAC line fitting to infer the dominant line structure.

4.3. Bi-Angular Multipath Enhancement (BAME)

Accurately identifying ghost targets is essential for reconstructing reliable reflector geometry. However, during experiments, we found that ghost visibility changes drastically across frames, as shown in Fig. 5(a). To increase the ghost detection capability, we propose algorithm named Bi-Angular Multipath Enhancement (BAME).

Motivation. The main reason lies in the difference

CVPR #12

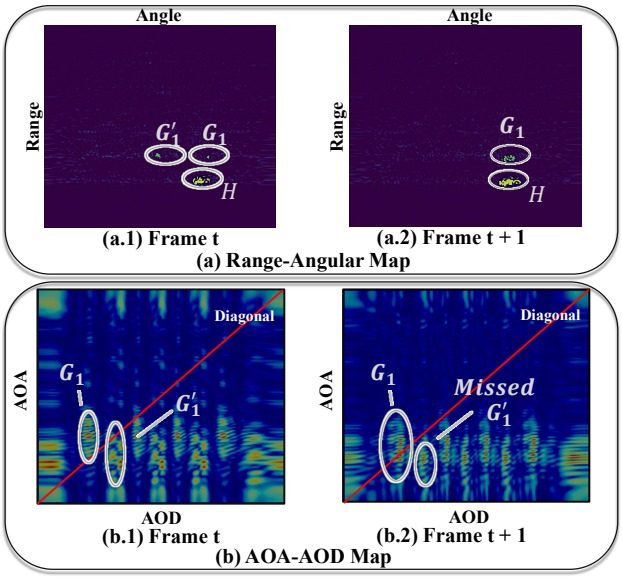

Figure 5. **Bi-Angular Multipath Enhancement (BAME).** Figure (a.1) shows a case where the ghost targets can be recovered using conventional beamforming, and Fig. (a.2) shows a case where the ghost target G_1' cannot be recovered using conventional beamforming. Figure (b) illustrates the corresponding AOA–AOD response at the range of the ghost targets G_1' and G_1. Notably, G_1' disappears in the standard range–angle map (Fig. (a.2)) because its AOA and AOD are not aligned with the diagonal, causing it to be suppressed by the AOA = AOD assumption.

between the **Angle of Arrival (AOA)** and **Angle of Departure (AOD)** of the reflected paths. For example, as shown in Fig. 4, ghost G_1' has AOA $= \theta_1^s$ and AOD $= \theta_2^s$. This asymmetry is crucial because most conventional radar beamforming formulations (Eq. 2) implicitly assume AOA = AOD, causing reflections with mismatched AOA and AOD—such as G_1'—to be suppressed or disappear.

Mathematical insight. If we separate the processing of AOA and AOD information, we obtain a more general formulation to find the received power at a given AOA, θ_{AOA}, AOD, θ_{AOD}, and range, r:

$$S(\theta_{\text{AOA}}, \theta_{\text{AOD}}, r) = \sum_{i=1}^{N_r} \left(\sum_{k=1}^{N_t} s_{ikr}\, e^{j2\pi(kd_t \cos\theta_{\text{AOD}})/\lambda} \right) e^{j2\pi(id_r \cos\theta_{\text{AOA}})/\lambda}. \tag{4}$$

where N_r and N_t are the number of receivers and transmitters, and d_r and d_t are the spacing between successive receivers and transmitters, respectively. s_{ikr} is the signal from the k$^{\text{th}}$ transmitter to the i$^{\text{th}}$ receiver, at a fixed range r (after an FFT along the range dimension). Here, the inner summation encodes the AOD contribution and the outer summation represents the AOA component. The corresponding 2D response map (at a fixed range r) is illustrated in Fig. 5(b). However, in standard MIMO radar processing, the receiver and transmitter indices $\{r_i\}_{i=1}^{N_r}$ and $\{t_j\}_{j=1}^{N_t}$ are combined into a single "virtual array" index, such that Eq. 2 becomes:

$$S(\theta, r) = \sum_{i=1}^{N_r} \sum_{k=1}^{N_t} s_{ikr}\, e^{j2\pi(d_r i + d_t k)\cos\theta/\lambda}. \tag{5}$$

This formulation implicitly merges AOA and AOD into a single equivalent angular dimension, effectively assuming AOA $\approx$ AOD. While this simplification is adequate for direct reflections, it suppresses multipath components where the two angles differ. Thus, first-order ghosts such as G_1'—which possess distinct AOA and AOD values—are partially or completely invisible in the resulting range–angle map.

Empirical observation. As illustrated in Fig. 5(a), ghost G_1' is visible in frame t but disappears in frame $t{+}1$. In frame t, both G_1 and G_1' lie along the diagonal of the Range–AOA–AOD map (left of Fig. 5(b)), making them detectable. However, in frame $t{+}1$, only G_1 remains on the diagonal, while G_1' shifts off-diagonal and consequently vanishes from the range–angle response (right of Fig. 5(b)). This temporal inconsistency demonstrates that conventional diagonal-only beamforming fails to capture multipath components with distinct AOA–AOD pairs.

Proposed Method. To overcome the above limitation, we introduce **Bi-Angular Multipath Enhancement (BAME)**, a two-angle radar processing pipeline that models both AOA and AOD dimensions. BAME reconstructs the full Range–AOA–AOD cube through the follows:

1. **Range FFT:** Perform a fast Fourier transform along the range dimension to obtain the distance spectrum.
2. **Bi-Angular Beamforming:** Apply Eq. 4 separately to receiver and transmitter dimensions to compute independent AOA and AOD responses.
3. **3D CFAR Detection:** Detect high-energy clusters within the Range–AOA–AOD cube using a three-dimensional CFAR filter.
4. **Ghost Reintegration:** For clusters exhibiting distinct AOA and AOD values, reinsert them into the range–angle map as valid ghost reflections.

Effect. This bi-angular formulation recovers the missing off-diagonal reflections that conventional processing suppresses, enabling consistent detection of both G_1 and G_1'. Consequently, the reconstructed range–angle map becomes more complete and geometrically faithful, forming a robust foundation for subsequent layout reconstruction.

4.4. Sim2Real Hierarchical Diffusion (SRHD)

Using the outputs of the Bi-Angular Multipath Enhancement (BAME) and Multipath Inversion modules, we obtain an initial reconstruction of the indoor layout. While this estimate captures the coarse spatial structure, it remains fragmented—comprising disjoint line segments that fail to form continuous walls or provide reliable object detection. Inspired by the strong reasoning capability of modern diffusion-based generative models across vision and

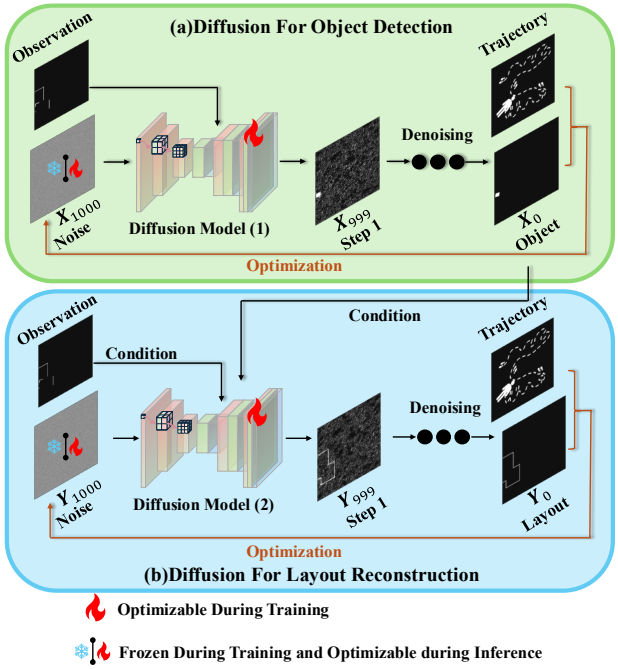

Figure 6. **Sim2Real Hierarchical Diffusion (SRHD).** In the first stage, RISE predicts the furniture-based segmentation map. In the second stage, RISE estimates the layout-based segmentation map.

3D perception, RISE introduces a **geometry-aware diffusion framework** that bridges the gap between fragmented radar observations and complete indoor layout reconstruction. However, training such a model requires a large corpus of radar-to-layout correspondences, which is unavailable due to the novelty of this sensing modality. To overcome this data scarcity, we design a high-fidelity **layout simulation engine** and a **hierarchical diffusion architecture** that jointly enable robust, geometry-consistent reconstruction from partial and noisy radar observations.

Layout Simulation Engine. We construct a simulator based on a corpus of 35,000 real-world indoor floor plans. Each layout is first converted into a 2D structural skeleton, and a virtual radar sensor is randomly placed at arbitrary positions and orientations. Multiple rays are then emitted from the radar, and only the first intersection point per ray is retained, simulating the physical propagation property of mmWave signals—where only the nearest surface contributes to the reflection. To emulate real-world multipath interactions, we insert N randomly positioned bounding boxes as objects and select one reflective line segment per box to model stochastic material-dependent scattering.

Data Augmentation. To enhance generalization from simulation to reality, we apply three augmentation strategies (Fig. 11 Appendix.) that inject structural noise:

- **Random Missing.** Randomly remove line segments within selected angular intervals Θ_{del} to simulate occluded or incomplete radar coverage.
- **Random Rotation.** Rotate all foreground pixels (x_p, y_p) within a chosen angular range Θ_{rot} by angle α_i around the radar's center (x_c, y_c):

$$\begin{bmatrix} x'_p \\ y'_p \end{bmatrix} = \begin{bmatrix} \cos\alpha_i & -\sin\alpha_i \\ \sin\alpha_i & \cos\alpha_i \end{bmatrix} \begin{bmatrix} x_p - x_c \\ y_p - y_c \end{bmatrix} + \begin{bmatrix} x_c \\ y_c \end{bmatrix}. \quad (6)$$

This augmentation mimics sensor pose errors and small orientation drifts.

- **Random Scaling.** Scale each foreground pixel (x_j, y_j) radially relative to the radar position (x_r, y_r) by a random factor a_i:

$$\begin{bmatrix} x'_p \\ y'_p \end{bmatrix} = \begin{bmatrix} x_r \\ y_r \end{bmatrix} + a_i \left(\begin{bmatrix} x_p \\ y_p \end{bmatrix} - \begin{bmatrix} x_r \\ y_r \end{bmatrix} \right), \quad (7)$$

which simulates depth uncertainty and propagation delay variations.

These augmentations improve robustness by exposing the model to spatial noise and non-ideal radar conditions.

Hierarchical Diffusion Model. With the augmented dataset, we train a two-stage diffusion model (Fig. 6) that reconstructs both semantic and geometric completeness:

- **Stage 1: Object Detection.** Given a partial observation $\mathbf{O}$, the first diffusion model f_1 predicts a binary object detection map $\mathbf{X}_0 \in \{0,1\}^{H\times W}$:

$$\mathbf{X}_0 = f_1(\mathbf{O}, \mathbf{X}_{1000}; \theta_1), \quad (8)$$

where $\mathbf{X}_{1000}$ denotes Gaussian noise. This step provides semantic priors that guide reasoning in the next stage.

- **Stage 2: Wall Structure Prediction.** The second model f_2 takes both $\mathbf{O}$ and $\mathbf{X}_{1000}$ as inputs to reconstruct the full wall configuration $\mathbf{Y}_{1000}$:

$$\mathbf{Y}_0 = f_2(\mathbf{O}, \mathbf{Y}_{1000}, \mathbf{X}_0; \theta_2). \quad (9)$$

A channel-attention fusion module integrates geometric and semantic cues, enforcing spatial coherence.

Spatial Consistency Optimization. During inference, we further refine the latent noise variables $\mathbf{X}_{1000}$ and $\mathbf{Y}_{1000}$ using an overlap-based optimization that enforces plausibility between human motion and generated structures:

$$\mathcal{L}_{\text{overlap}} = \sum_{(x_t,y_t)\in\mathcal{T}} \Big[\mathbb{1}[(x_t, y_t) \in \mathcal{W}] + \mathbb{1}[(x_t, y_t) \in \mathcal{B}] \Big], \quad (10)$$

where $\mathcal{T}$ denotes the trajectory of human motion, and $\mathcal{W}$ and $\mathcal{B}$ represent wall and object masks, respectively. Here, $\mathbb{1}[\cdot]$ is the indicator function that equals 1 if a trajectory point collides with an obstacle and 0 otherwise. Minimizing $\mathcal{L}_{\text{overlap}}$ iteratively adjusts the latent noise inputs, eliminating wall–trajectory and object–trajectory intersections. This reverse optimization ensures that the final reconstruction is physically valid—walls remain static and non-penetrable, while human trajectories traverse only feasible free-space regions. A discussion is presented in our Appendix.

method	Metric	Type	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	AVG
BRL [22]	IoU (%) ↑	S	00.00	00.00	00.00	00.00	00.00	12.93	00.00	00.00	00.00	00.00	00.00	01.17
	Dice (%) ↑	S	00.00	00.00	00.00	00.00	00.00	20.03	00.00	00.00	00.00	00.00	00.00	01.82
EMT [9]+SRHD	IoU (%) ↑	M	00.00	00.00	15.61	00.00	00.00	25.77	23.44	19.93	00.00	00.00	00.00	07.70
	Dice (%) ↑	M	00.00	00.00	26.13	00.00	00.00	43.98	38.65	30.54	00.00	00.00	00.00	12.66
RISE(Ours)	IoU (%) ↑	M	58.70	96.02	29.48	47.90	55.24	47.12	64.70	37.91	64.43	74.19	59.87	57.78
	Dice (%) ↑	M	73.93	96.68	43.64	64.26	70.24	61.55	71.20	48.46	77.41	84.86	70.52	69.34

Table 1. **Object Detection Results:** This table presents the object detection results of our method and baselines, evaluated using Intersection over Union (IoU) and Dice coefficient metrics. "S" denotes single-frame input, and "M" denotes multi-frame input.

5. Experimental

5.1. Experimental Setting

Tasks. We evaluate RISE on two core tasks: *layout reconstruction* and *object detection*. For layout reconstruction, we report Chamfer Distance to quantify geometric accuracy and F1-score (15 cm tolerance) to measure boundary completeness. For object detection, we adopt standard 2D metrics: Intersection-over-Union (IoU) and the Dice coefficient. **Baselines.** Since RISE is the first system to perform *both* layout reconstruction and object detection using a *single static mmWave radar*, no prior method directly matches our setting. For object detection, we compare against: (1) the state-of-the-art learning-based single-radar detector BRL [22]; (2) the multi-frame radar layout reconstruction method EMT [9]. Because EMT does not support object detection, we augment its outputs using the same post-processing used in RISE to provide a fair comparison. We rename it as EMT [9] + SRHD. For layout reconstruction, we directly compare against EMT [9], a strong existing multi-frame radar layout reconstruction baseline.

5.2. Object Detection

Tab. 1 and Fig. 8 present the object detection results. As shown in Fig. 8, the baseline method (EMT [9]) fails to recover reflectors from small objects, making it unable to produce meaningful object detections. To enable a fair comparison, we report two additional baselines in Tab. 1: (1) **EMT [9] + SRHD**, where we apply our Sim2Real Hierarchical Diffusion module on top of the EMT output, and (2) **BRL** [22], a deep-learning–based single-radar object detector. Across all settings, RISE delivers substantially stronger performance. By combining Ghost Signal Enhancement with Sim2Real Hierarchical Diffusion, RISE reliably localizes objects and generalizes to complex indoor scenes. RISE achieves an **IoU of 57.78** and a **Dice score of 69.34**, far surpassing the best baseline (IoU = 7.70, Dice = 12.66).

5.3. Layout Reconstruction

Fig. 8 also presents qualitative comparisons for layout reconstruction between our method and the baseline **EMT** [9]. Since EMT [9] is inherently capable of wall-layout reconstruction, we evaluate it in its original form without combining it with SRHD. We further summarize

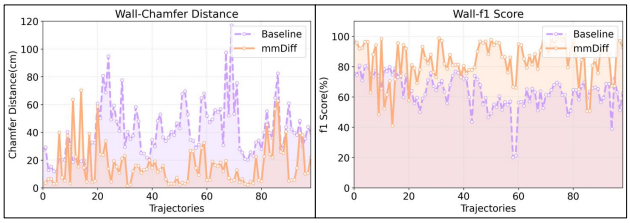

Figure 7. **Wall Reconstruction Across 100 Trajectories.** Comparison of our method and the baseline (**EMT** [9]) over 100 real-world trajectories, evaluated using Chamfer Distance and F1-score for wall layout reconstruction.

quantitative results across 100 trajectories in Fig. 7. Compared with EMT [9], RISE produces substantially more complete and structurally coherent layouts, particularly in regions where multipath coverage is sparse. Quantitatively, RISE achieves a markedly lower average Chamfer distance of **16.03 cm** (vs. **39.06 cm** for EMT) and a significantly higher F1-score of **83.63** (vs. **63.43**). These results demonstrate the effectiveness of our Bi-Angular Multipath Ghost Enhancement and Sim2Real Hierarchical Diffusion framework in recovering both global room geometry and fine-grained structural details.

5.4. Effect of Trajectory Length

Our method employs 30-second trajectories for layout reconstruction. To assess the impact of trajectory length on performance, we analyze the results as the trajectory length decreases shown in Fig. 9. Both our method and the baselines (**EMT** [9]) exhibit performance degradation as the trajectory length shortens. However, the decline in F1-score is more pronounced in our method. This is attributed to the nature of the F1-score computation, which takes into account predictions within a specific error threshold. As the trajectory length decreases, the diffusion model generates more diverse outputs due to the reduced input information, which consequently has a greater impact on the F1-score. Notably, even with 40% of the full-length trajectories, RISE still achieves a lower Chamfer distance compared to the baseline performance using the full-length trajectories.

5.5. Ablation Study

To assess the contribution of key components in our framework, we conduct an ablation study, the results of which are summarized in Tab. 2. We evaluate the impact of three crit-

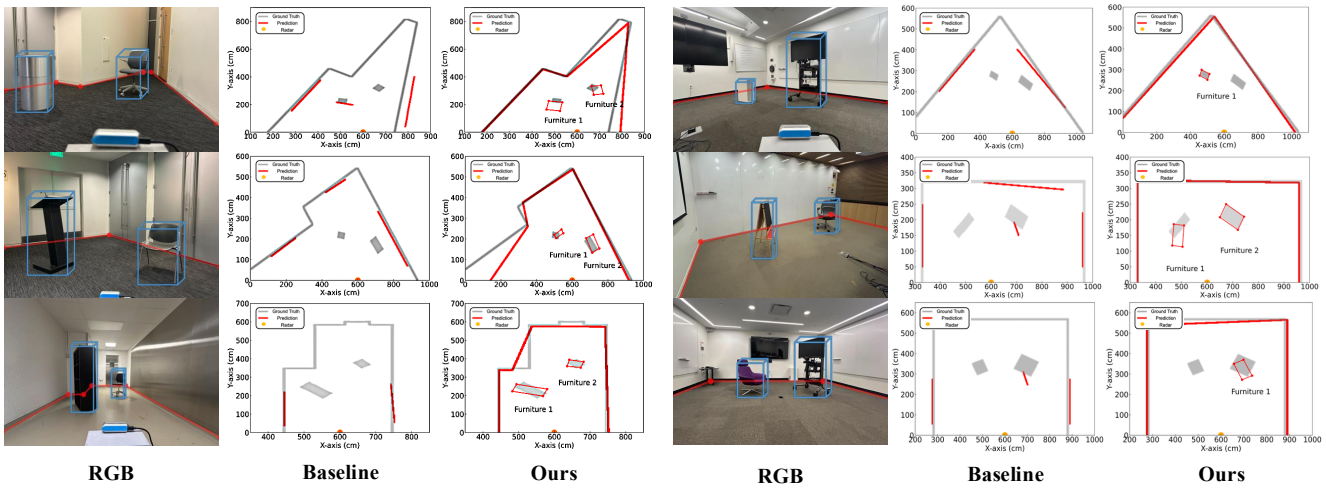

Figure 8. **Comparison Between Our Method and the Baseline.** The first column shows the RGB reference images of the corresponding scenes. The second column presents the reconstruction results produced by EMT [9]. The third column shows the results generated by our method, including both layout reconstruction and object detection.

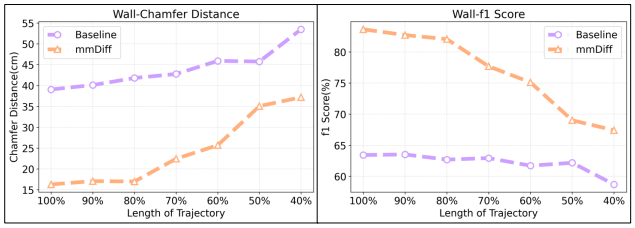

Figure 9. **Results Across Varying Trajectory Lengths.** Comparison between our method and the baseline **EMT** [9]. The figure reports layout reconstruction performance—Chamfer distance and F1-score—under shorter human trajectories, illustrating the robustness of RISE when input coverage becomes limited.

Method/Metric	F1-score ↑	Chamfer ↓
Baseline	63.43 %	39.06 cm
RISE w/ G	73.37 %	32.31 cm
RISE w/ G, D	78.84 %	19.82 cm
RISE w/ G, D, R	83.63 %	16.32 cm

Table 2. Ablation study results showing the impact of different components on F1-score and Chamfer Distance.

ical components: Ghost Signal Enhancement, the diffusion model, and reverse optimization. The configurations tested include: **Model w/ G**, which includes only Ghost Signal Enhancement; **Model w/ G, D**, which includes both Ghost Signal Enhancement and the diffusion model; and **Model w/ G, D, R**, the full model incorporating Ghost Signal Enhancement, the diffusion model, and reverse optimization.

Our findings reveal a progressive improvement in Chamfer distance, which underscores the contribution of each module in enhancing reconstruction accuracy: 1) The inclusion of Ghost Signal Enhancement reduces the Chamfer distance from **39.06 cm** to **32.31 cm**, highlighting its ability to improve initial detection by enhancing reflector visibility. 2) The addition of the diffusion model further refines the reconstruction, reducing the Chamfer distance to **19.82 cm**, demonstrating its ability to generate more accurate layout details by leveraging generative modeling. 3) The incorporation of reverse optimization leads to the final refinement, achieving a Chamfer distance of **16.32 cm**. This stage fine-tunes the results by optimizing the generative outputs, further improving the quality and precision of the reconstructed layout. The progressive reduction in Chamfer distance across these configurations demonstrates the complementary nature of each component. Ghost Signal Enhancement aids in initial detection, the diffusion model enhances overall reconstruction, and reverse optimization refines the final output, collectively resulting in a substantial improvement in reconstruction accuracy.

6. Conclusion

In conclusion, this paper introduces a novel method for indoor layout reconstruction using a single static mmWave sensor. By leveraging human mobility-induced multipath effects and integrating a generative model, our system, RISE, effectively tackles challenges such as specularity and incomplete layout reconstruction. Experimental results demonstrate that RISE achieves a Chamfer distance of 16 cm for layout reconstruction and an IoU of 58% for furniture detection, outperforming conventional approaches. This work highlights the potential of privacy-preserving mmWave sensing for applications in smart homes, security, and virtual/augmented reality, and lays the groundwork for future improvements in more complex environments, such as multi-person scenarios and shorter trajectories.

7. Acknowledgments.

We thank the anonymous reviewers and the Signal Kinetics group for their help and feedback. This research is sponsored by NSF (Award #2313234), Amazon MIT Science Hub, and the MIT Media Lab. Laura Dodds is supported by the Amazon Robotics Thriving Stars Fellowship.

References

[1] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. Capturing the human figure through a wall. *ACM Transactions on Graphics (TOG)*, 34(6):1–13, 2015. 4

[2] Alireza Asvadi, Luis Garrote, Cristiano Premebida, Paulo Peixoto, and Urbano J Nunes. Multimodal vehicle detection: fusing 3d-lidar and color camera data. *Pattern Recognition Letters*, 115:20–29, 2018. 2

[3] Geonho Bang, Kwangjin Choi, Jisong Kim, Dongsuk Kum, and Jun Won Choi. Radardistill: Boosting radar-based object detection performance via knowledge distillation from lidar features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15491–15500, 2024. 1

[4] Oded Bialer and Yuval Haitman. Radsimreal: Bridging the gap between synthetic and real data in radar object detection with simulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15407–15416, 2024. 1

[5] Katherine L Bouman, Vickie Ye, Adam B Yedidia, Frédo Durand, Gregory W Wornell, Antonio Torralba, and William T Freeman. Turning corners into cameras: Principles and methods. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2270–2278, 2017. 3

[6] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020. 2

[7] Yujeong Chae, Hyeonseong Kim, and Kuk-Jin Yoon. Towards robust 3d object detection with lidar and 4d radar fusion in various weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15162–15172, 2024. 2

[8] Shoufa Chen, Peize Sun, Yibing Song, and Ping Luo. Diffusiondet: Diffusion model for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 19830–19843, 2023. 2

[9] Weiyan Chen, Hongliu Yang, Xiaoyang Bi, Rong Zheng, Fusang Zhang, Peng Bao, Zhaoxin Chang, Xujun Ma, and Daqing Zhang. Environment-aware multi-person tracking in indoor environments with mmwave radars. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(3):1–29, 2023. 2, 3, 7, 8

[10] Xingyu Chen and Xinyu Zhang. Rf genesis: Zero-shot generalization of mmwave sensing through simulation-based data synthesis and generative diffusion models. In *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems*, pages 28–42, 2023. 2

[11] Liang Cheng, Jianya Gong, Manchun Li, and Yongxue Liu. 3d building model reconstruction from multi-view aerial imagery and lidar data. *Photogrammetric Engineering & Remote Sensing*, 77(2):125–139, 2011. 2

[12] Guoxuan Chi, Zheng Yang, Chenshu Wu, Jingao Xu, Yuchong Gao, Yunhao Liu, and Tony Xiao Han. Rf-diffusion: Radio signal generation via time-frequency diffusion. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 77–92, 2024. 2

[13] Han Cui, Shu Zhong, Jiacheng Wu, Zichao Shen, Naim Dahnoun, and Yiren Zhao. Milipoint: A point cloud dataset for mmwave radar. *Advances in Neural Information Processing Systems*, 36:62713–62726, 2023. 1

[14] Fangqiang Ding, Andras Palffy, Dariu M Gavrila, and Chris Xiaoxuan Lu. Hidden gems: 4d radar scene flow learning using cross-modal supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9340–9349, 2023. 2

[15] Fangqiang Ding, Zhen Luo, Peijun Zhao, and Chris Xiaoxuan Lu. milliflow: Scene flow estimation on mmwave radar point cloud for human motion sensing. In *European Conference on Computer Vision*, pages 202–221. Springer, 2024. 1

[16] Fangqiang Ding, Xiangyu Wen, Yunzhou Zhu, Yiming Li, and Chris Xiaoxuan Lu. Radarocc: Robust 3d occupancy prediction with 4d imaging radar. *Advances in Neural Information Processing Systems*, 37:101589–101617, 2024.

[17] Anthony Fuller, Koreen Millard, and James Green. Croma: Remote sensing representations with contrastive radar-optical masked autoencoders. *Advances in Neural Information Processing Systems*, 36:5506–5538, 2023. 1

[18] Michael Fürst, Shriya TP Gupta, René Schuster, Oliver Wasenmüller, and Didier Stricker. Hperl: 3d human pose estimation from rgb and lidar. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 7321–7327. IEEE, 2021. 2

[19] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015. 2

[20] Francesco Guidi, Anna Guerra, and Davide Dardari. Personal mobile radars with millimeter-wave massive arrays for indoor mapping. *IEEE Transactions on Mobile Computing*, 15(6):1471–1484, 2015. 3

[21] Yongchang Hao, Mengyao Zhai, Hossein Hajimirsadeghi, Sepidehsadat Hosseini, and Frederick Tung. Radar: Fast long-context decoding for any transformer. In *The Thirteenth International Conference on Learning Representations*. 2

[22] Yiduo Hao, Sohrab Madani, Junfeng Guan, Mohammed Alloulah, Saurabh Gupta, and Haitham Hassanieh. Bootstrapping autonomous driving radars with self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15012–15023, 2024. 2, 7

[23] Steven M Hernandez and Eyuphan Bulut. Wifi sensing on the edge: Signal processing techniques and challenges for real-world systems. *IEEE Communications Surveys & Tutorials*, 25(1):46–76, 2022. 2

[24] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 2

[25] Yuan-Hao Ho, Jen-Hao Cheng, Sheng Yao Kuan, Zhongyu Jiang, Wenhao Chai, Hsiang-Wei Huang, Chih-Lung Lin, and Jenq-Neng Hwang. Rt-pose: A 4d radar tensor-based 3d human pose estimation and localization benchmark. In *European Conference on Computer Vision*, pages 107–125. Springer, 2024. 1

[26] Boyang Hu, Hui Tian, Wei Ni, Shaoshuai Fan, Wanli Ni, and Ekram Hossain. Multipath identification, user localization, and environment mapping in radio slam. *IEEE Transactions on Communications*, 2024. 2

[27] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24118–24129, 2024. 2

[28] Jihye Kim, Jaehoon Jung, Sohee Lim, and Seong-Cheol Kim. Multipath suppression using 2d angle analysis based on the mimo fmcw radar. *IEEE Transactions on Instrumentation and Measurement*, 2024. 2

[29] Jisong Kim, Minjae Seong, and Jun Won Choi. Crt-fusion: Camera, radar, temporal fusion using motion information for 3d object detection. *Advances in Neural Information Processing Systems*, 37:108625–108648, 2024. 2

[30] Youngseok Kim, Juyeb Shin, Sanmin Kim, In-Jae Lee, Jun Won Choi, and Dongsuk Kum. Crn: Camera radar net for accurate, robust, efficient 3d perception. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17615–17626, 2023. 2

[31] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023. 2

[32] Maria Koutsoupidou, Helena Cano-Garcia, Roberto L Pricci, Shimul C Saha, George Palikaras, Efthymios Kallos, and Panagiotis Kosmas. Study and suppression of multipath signals in a non-invasive millimeter wave transmission glucose-sensing system. *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, 4(3):187–193, 2019. 2

[33] Wei-Yu Lee, Martin Dimitrievski, David Van Hamme, Jan Aelterman, Ljubomir Jovanov, and Wilfried Philips. Carbnet: Camera-assisted radar-based network for vulnerable road user detection. In *European Conference on Computer Vision*, pages 294–310. Springer, 2024. 2

[34] Huadong Li, Minhao Jing, Wang Jin, Shichao Dong, Jiajun Liang, Haoqiang Fan, and Renhe Ji. Sparse beats dense: Rethinking supervision in radar-camera depth completion. In *European Conference on Computer Vision*, pages 127–143. Springer, 2024. 2

[35] Teng Li, Liwen Zhang, Youcheng Zhang, Pengcheng Pi, Zongqing Lu, Qingmin Liao, and Zhe Ma. Adapkc: Peakconv with adaptive peak receptive field for radar semantic segmentation. *Advances in Neural Information Processing Systems*, 37:136545–136575, 2024. 2

[36] Xuhong Li, Xuesong Cai, Erik Leitinger, and Fredrik Tufvesson. A belief propagation algorithm for multipath-based slam with multiple map features: A mmwave mimo application. In *2024 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 269–275. IEEE, 2024. 2

[37] Yu-Jhe Li, Shawn Hunt, Jinhyung Park, Matthew O'Toole, and Kris Kitani. Azimuth super-resolution for fmcw radar in autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17504–17513, 2023. 1

[38] Mingchao Liang, Erik Leitinger, and Florian Meyer. Direct multipath-based slam. *arXiv preprint arXiv:2409.20552*, 2024. 2

[39] Huichen Lin and Neil W Bergmann. Iot privacy and security challenges for smart home environments. *Information*, 7(3): 44, 2016. 1, 3

[40] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 2

[41] Zhiwei Lin, Zhe Liu, Zhongyu Xia, Xinhao Wang, Yongtao Wang, Shengxiang Qi, Yang Dong, Nan Dong, Le Zhang, and Ce Zhu. Rcbevdet: radar-camera fusion in bird's eye view for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14928–14937, 2024. 2

[42] Ruizhi Liu, Xinghui Song, Jiawei Qian, Shuai Hao, Yue Lin, and Hongtao Xu. A data-driven method for indoor radar ghost recognition with environmental mapping. *IEEE Transactions on Radar Systems*, 2024. 2

[43] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer, 2016. 2

[44] Yang Liu, Feng Wang, Naiyan Wang, and ZHAO-XIANG ZHANG. Echoes beyond points: Unleashing the power of raw radar data in multi-modality fusion. *Advances in Neural Information Processing Systems*, 36:53964–53982, 2023. 2

[45] Zhijian Liu, Haotian Tang, Alexander Amini, Xinyu Yang, Huizi Mao, Daniela L Rus, and Song Han. Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation. In *2023 IEEE international conference on robotics and automation (ICRA)*, pages 2774–2781. IEEE, 2023. 2

[46] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 2

[47] Chris Xiaoxuan Lu, Stefano Rosa, Peijun Zhao, Bing Wang, Changhao Chen, John A Stankovic, Niki Trigoni, and Andrew Markham. See through smoke: robust indoor mapping with low-cost mmwave radar. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, pages 14–27, 2020. 3

[48] Alessandro Manni, Damiano Oriti, Andrea Sanna, Francesco De Pace, and Federico Manuri. Snap2cad: 3d indoor envi-

ronment reconstruction for ar/vr applications using a smartphone device. *Computers & Graphics*, 100:116–124, 2021. 1

[49] Arvind Narayanan, Eman Ramadan, Rishabh Mehta, Xinyue Hu, Qingxu Liu, Rostand AK Fezeu, Udhaya Kumar Dayalan, Saurabh Verma, Peiqi Ji, Tao Li, et al. Lumos5g: Mapping and predicting commercial mmwave 5g throughput. In *Proceedings of the ACM internet measurement conference*, pages 176–193, 2020. 3

[50] Jeong-Ki Park, Jae-Hyun Park, and Kyung-Tae Kim. Multipath signal mitigation for indoor localization based on mimo fmcw radar system. *IEEE internet of things journal*, 11(2): 2618–2629, 2023. 2

[51] Guohao Peng, Heshan Li, Yangyang Zhao, Jun Zhang, Zhenyu Wu, Pengyu Zheng, and Danwei Wang. Transloc4d: Transformer-based 4d radar place recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17595–17605, 2024. 1

[52] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 2

[53] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017. 2

[54] Kun Qian, Zhaoyuan He, and Xinyu Zhang. 3d point cloud generation with millimeter-wave radar. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4):1–23, 2020. 3

[55] Kun Qian, Lulu Yao, Xinyu Zhang, and Tse Nga Ng. Millimirror: 3d printed reflecting surface for millimeter-wave coverage expansion. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, pages 15–28, 2022. 2

[56] M Mahbubur Rahman, Ryoma Yataka, Sorachi Kato, Pu Wang, Peizhao Li, Adriano Cardace, and Petros Boufounos. Mmvr: Millimeter-wave multi-view radar dataset and benchmark for indoor perception. In *European Conference on Computer Vision*, pages 306–322. Springer, 2024. 3

[57] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016. 2

[58] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2

[59] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 2

[60] Sebastian Schneider, Michael Himmelsbach, Thorsten Luettel, and Hans-Joachim Wuensche. Fusing vision and lidar-synchronization, correction and occlusion reasoning. In *2010 IEEE Intelligent Vehicles Symposium*, pages 388–393. IEEE, 2010. 1, 3

[61] Akash Deep Singh, Yunhao Ba, Ankur Sarker, Howard Zhang, Achuta Kadambi, Stefano Soatto, Mani Srivastava, and Alex Wong. Depth estimation from camera image and mmwave radar point cloud. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9275–9285, 2023. 2

[62] Neharika Valecha, Jesus Omar Lacruz, Michael Lentmaier, Joerg Widmer, and Fredrik Tufvesson. Angle estimation using mmwave rss measurements with enhanced multipath information. In *2024 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2024. 2

[63] Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Moungi Bawendi, Diego Gutierrez, and Ramesh Raskar. Femto-photography: capturing and visualizing the propagation of light. *ACM Transactions on Graphics (ToG)*, 32(4):1–8, 2013. 3

[64] Yingjie Wang, Jiajun Deng, Yao Li, Jinshui Hu, Cong Liu, Yu Zhang, Jianmin Ji, Wanli Ouyang, and Yanyong Zhang. Bi-lrfusion: Bi-directional lidar-radar fusion for 3d dynamic object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13394–13403, 2023. 2

[65] Ziye Wang, Yiran Qin, Lin Zeng, and Ruimao Zhang. High-dynamic radar sequence prediction for weather nowcasting using spatiotemporal coherent gaussian representation. *arXiv preprint arXiv:2502.14895*, 2025. 2

[66] Bichen Wu, Alvin Wan, Xiangyu Yue, and Kurt Keutzer. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud. In *2018 IEEE international conference on robotics and automation (ICRA)*, pages 1887–1893. IEEE, 2018. 2

[67] Jialong Wu, Mirko Meuter, Markus Schoeler, and Matthias Rottmann. Sparseradnet: Sparse perception neural network on subsampled radar data. In *European Conference on Computer Vision*, pages 52–69. Springer, 2024. 2

[68] Peilun Wu, Jiahui Chen, Shisheng Guo, Guolong Cui, Lingjiang Kong, and Xiaobo Yang. Nlos positioning for building layout and target based on association and hypothesis method. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–13, 2023. 2, 4

[69] Danfei Xu, Dragomir Anguelov, and Ashesh Jain. Pointfusion: Deep sensor fusion for 3d bounding box estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 244–253, 2018. 2

[70] Yan Xu, Xinge Zhu, Jianping Shi, Guofeng Zhang, Hujun Bao, and Hongsheng Li. Depth completion from sparse lidar data with depth-normal constraints. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2811–2820, 2019. 2

[71] Aimin Yang, Chunying Zhang, Yongjie Chen, Yunxi Zhuansun, and Huixiang Liu. Security and privacy of smart home systems based on the internet of things and stereo matching algorithms. *IEEE Internet of Things Journal*, 7(4):2521–2530, 2019. 1

[72] Ming-Der Yang, Chih-Fan Chao, Kai-Siang Huang, Liang-You Lu, and Yi-Ping Chen. Image-based 3d scene reconstruction and exploration in augmented reality. *Automation in Construction*, 33:48–60, 2013. 1

[73] Ali Yassin, Youssef Nasser, Ahmed Y Al-Dubai, and Mariette Awad. Mosaic: Simultaneous localization and environment mapping using mmwave without a-priori knowledge. *IEEE Access*, 6:68932–68947, 2018. 3

[74] Ryoma Yataka, Adriano Cardace, Perry Wang, Petros Boufounos, and Ryuhei Takahashi. Retr: Multi-view radar detection transformer for indoor perception. *Advances in Neural Information Processing Systems*, 37:19839–19869, 2024. 3

[75] Fangrui Yu, Shisheng Guo, Xiaojian Hao, Jiahui Chen, Nian Li, Guolong Cui, Lingjiang Kong, and Xiaobo Yang. At-blr: Aoa and td based multimaterial building layout reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 3

[76] Jingtong Yue, Zhiwei Lin, Xin Lin, Xiaoyu Zhou, Xiangtai Li, Lu Qi, Yongtao Wang, and Ming-Hsuan Yang. Roburcdet: Enhancing robustness of radar-camera fusion in bird's eye view for 3d object detection. *arXiv preprint arXiv:2502.13071*, 2025. 2

[77] Zhiyuan Zeng, Jie Wen, Jianan Luo, Gege Ding, and Xiongfei Geng. Dense 3d point cloud environmental mapping using millimeter-wave radar. *Sensors*, 24(20):6569, 2024. 3

[78] Liwen Zhang, Xinyan Zhang, Youcheng Zhang, Yufei Guo, Yuanpei Chen, Xuhui Huang, and Zhe Ma. Peakconv: Learning peak receptive field for radar semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17577–17586, 2023. 1

[79] Tianshu Zhang, Buzhen Huang, and Yangang Wang. Object-occluded human shape and pose estimation from a single color image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7376–7385, 2020. 1, 3

[80] Yang Zhang, Jiahui Chen, Shisheng Guo, Xiaobo Yang, and Guolong Cui. Building layout tomographic reconstruction via commercial wifi signals. *IEEE Internet of Things Journal*, 8(20):15500–15511, 2021. 2, 3

[81] Youcheng Zhang, Liwen Zhang, Pengcheng Pi, Teng Li, Yuanpei Chen, Shi Peng, and Zhe Ma. Tarss-net: Temporal-aware radar semantic segmentation network. *Advances in Neural Information Processing Systems*, 37:4906–4933, 2024. 1

[82] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017. 2

[83] Lingjun Zhao, Jingyu Song, and Katherine A Skinner. Crkd: Enhanced camera-radar object detection with cross-modality knowledge distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15470–15480, 2024. 2

[84] Kaichen Zhou, Jia-Xing Zhong, Sangyun Shin, Kai Lu, Yiyuan Yang, Andrew Markham, and Niki Trigoni. Dynpoint: Dynamic neural point for view synthesis. *Advances in Neural Information Processing Systems*, 36:69532–69545, 2023. 2

[85] Kaichen Zhou, Jia-Wang Bian, Jian-Qing Zheng, Jiaxing Zhong, Qian Xie, Niki Trigoni, and Andrew Markham. Manydepth2: Motion-aware self-supervised monocular depth estimation in dynamic scenes. *IEEE Robotics and Automation Letters*, 2025.

[86] Kaichen Zhou, Yuhan Wang, Grace Chen, Xinhai Chang, Gaspard Beaudouin, Fangneng Zhan, Paul Pu Liang, and Mengyu Wang. Page-4d: Disentangled pose and geometry estimation for 4d perception. *arXiv preprint arXiv:2510.17568*, 2025. 2

[87] Saman Zia, Buket Yuksel, Deniz Yuret, and Yucel Yemez. Rgb-d object recognition using deep convolutional neural networks. In *Proceedings of the IEEE International conference on computer vision workshops*, pages 896–903, 2017. 2

8. Experiments

Data Process: The ground truth is generated using a RealSense depth camera. Specifically, the depth camera captures the 3D point cloud of the surrounding environment. This point cloud is then transformed into the radar coordinate system using pre-calibrated intrinsic and extrinsic matrices. To create ground truth for wall estimation, we project all 3D points onto the radar's horizontal (XZ) plane by collapsed along Z. This projection provides a reliable reference for evaluating our wall detection algorithm. For object detection, we first use the depth camera to identify the positions of objects in the scene. Then, using geometric cues from the depth data and simple measurement rules, we estimate the actual dimensions of each object. Based on the estimated position and size, we construct 2D bounding boxes in the radar coordinate system, which serve as ground truth for evaluating object detection performance.

9. Detail for Section - 4.2. Multipath Inversion

9.1. Ghost Target Formation and Identification

To make the ghost and human target identification process transparent, we provide a clear step-by-step pseudocode description in Algorithm 1.

9.2. Reflector Point Estimation

9.2.1. First-Bounce Ghost

The first-bounce ghost point, denoted as G_1', is geometrically related to the source S' as follows:

$$2|sg_1'| = |sh| + |hc_1| + |sc_1|, \tag{11}$$

where $|sg_1'|$ represents the distance from the radar to the first-bounce ghost point, $|sh|$ is the distance from the radar to the human being, $|hc_1|$ is the distance from the human being to the reflector point, and $|sc_1|$ is the distance from the radar to the reflector point.

Applying the cosine law to the triangle formed by S, C_1, and H, we obtain:

$$|sh|^2 + |sc_1|^2 - 2|sh||sc_1|\cos(\theta_2^s - \theta_1^s) = |hc_1|^2. \tag{12}$$

where $\theta_2^s - \theta_1^s$ is the angle between $\vec{sh}$ and $\vec{sg_1}$ From the above relationships, we derive $|sc_1|$ as:

$$|sc_1| = \frac{2|sg_1'|^2 - 2|sg_1'||sh|}{2|sg_1'| - |sh|\cos(\theta_2^s - \theta_1^s) - |sh|} \tag{13}$$

The coordinates of the reflector point C_1 can then be determined as:

$$c_1 = s + \frac{\vec{sg'}}{|sg'|} \cdot |sc_1| \tag{14}$$

where $\frac{\vec{sg'}}{|sg'|}$ is the unit vector in the direction of sg'.

Algorithm 1: Ghost Target Formation and Identification

Input: Clusters $\mathcal{C} = \{(r_i, \theta_i, m_i)\}$ obtained from CFAR detection, where r_i denotes the range, θ_i the angle, and m_i the reflection magnitude. Threshold ratio $\tau = 0.4$; Range tolerance $\delta_r = 0.15$m; Angle tolerance $\delta_\theta = 15°$.

Output: Human H, first-order ghosts G_1, G_1', second-order ghosts G_2, G_2'.

Step 1: Identify Human H;

$M_{\max} \leftarrow \max_i m_i$

$\mathcal{C}_{\text{valid}} \leftarrow \{c \in \mathcal{C} \mid m_c > \tau M_{\max}\}$

$H \leftarrow \arg\min_{c \in \mathcal{C}_{\text{valid}}} r_c$

Step 2: Identify G_1 (same direction, slightly larger range);

$\mathcal{G}_1 \leftarrow \{c \in \mathcal{C} \mid |\theta_c - \theta_H| < \delta_\theta,\ 0 < r_c - r_H < \delta_r\}$

$G_1 \leftarrow \arg\min_{c \in \mathcal{G}_1} r_c$ if $\mathcal{G}_1 \neq \emptyset$, else $\emptyset$

Step 3: Identify G_1' (same range, different direction);

$\mathcal{G}_{1'} \leftarrow \{c \in \mathcal{C} \mid |r_c - r_{G_1}| < \delta_r,\ |\theta_c - \theta_{G_1}| \geq \delta_\theta\}$

$G_1' \leftarrow \arg\max_{c \in \mathcal{G}_{1'}} m_c$ if $\mathcal{G}_{1'} \neq \emptyset$, else $\varnothing$

Step 4 (Optional): Identify G_2 (same direction as H, farther than G_1);

$\mathcal{G}_2 \leftarrow \{c \in \mathcal{C} \mid |\theta_c - \theta_H| < \delta_\theta,\ r_c > r_{G_1}\}$

$G_2 \leftarrow \arg\min_{c \in \mathcal{G}_2} r_c$ if $\mathcal{G}_2 \neq \emptyset$, else $\varnothing$

Step 5: Identify G_2' (aligned with $\vec{G_1'}$, farther range);

$\mathcal{G}_{2'} \leftarrow \{c \in \mathcal{C} \mid r_c > r_{G_1'},\ |\theta_c - \theta_{G_1'}| < \delta_\theta\}$

$G_2' \leftarrow \arg\min_{c \in \mathcal{G}_{2'}} r_c$ if $\mathcal{G}_{2'} \neq \emptyset$, else $\varnothing$

return H, G_1, G_1', G_2, G_2'

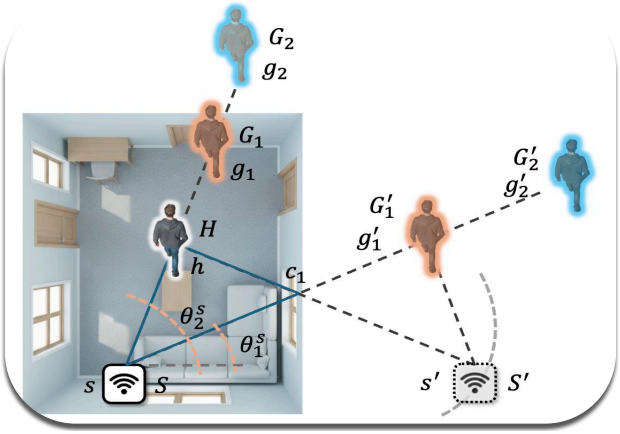

Figure 10. **Multipath Inversion.** Geometric relationships between ghost targets and reflectors.

CVPR #12

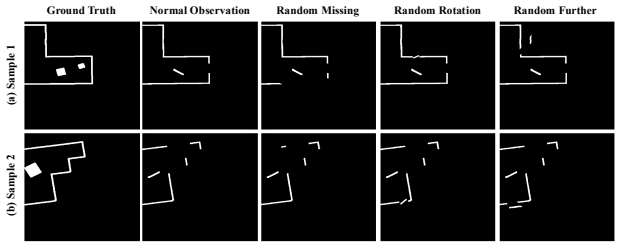

Figure 11. **Illustration of Simulator and Data Augmentation.** The first column represents the ground truth in the simulator, while the second column shows partial observations resulting from occlusion. The third column depicts randomly missing regions, the fourth column illustrates random rotations, and the fifth column demonstrates random scaling.

9.2.2. Second-Bounce Ghost

The second-bounce ghost point, denoted as G_2', follows a geometric relationship similar to the first-bounce case. The total path length from the radar to G_2' can be expressed as:

$$|sg_2'| = |hc_1| + |sc_1|, \tag{15}$$

where $|sg_2'|$ represents the total distance from the radar to the second-bounce ghost point, $|hc_1|$ is the distance from the human to the reflector, and $|sc_1|$ is the distance from the radar to the reflector.

Using the cosine law and substituting the known relationships, we derive:

$$|sc_1| = \frac{|sg_2'|^2 - |sh|^2}{2|sg_1'| - 2|sh|\cos(\theta_2^s - \theta_1^s)}, \tag{16}$$

where θ_1^s and θ_2^s denote the respective angles between the reflection points and the radar.

Subsequently, the remaining steps outlined in the first-bounce ghost section can be applied to determine the location of the mirror radar.

9.3. Post Processing Reflector Points

After estimating reflector points from ghost targets, we obtain a sparse and noisy set of 2D samples that approximate the underlying wall and object boundaries. To extract coherent geometric structures, we refine these points in two stages.

First, we apply a Gaussian Mixture Model (GMM) to group the reflector points into spatially coherent clusters. The GMM models the distribution of reflector points as a mixture of Gaussian components, which enables soft probabilistic assignments and naturally captures elongated point distributions associated with physical surfaces. This clustering step separates points belonging to different walls or objects while remaining robust to uneven sampling and local noise.

Next, for each cluster, we perform RANSAC line fitting to recover the dominant structural direction and reject outliers introduced by multipath noise or incomplete reflections. RANSAC iteratively samples minimal point sets, fits line hypotheses, and selects the model with the largest inlier set under a point-to-line distance threshold. This procedure produces clean and geometrically consistent line segments for each cluster.

The resulting set of RANSAC-refined line structures forms a robust initial layout estimate, which is subsequently fed into our diffusion-based refinement module to obtain complete wall boundaries and object footprints.

10. Discussion

RISE demonstrates that a single static mmWave radar can reliably recover indoor layout and object information by leveraging multipath reflections and generative modeling. While our system has several limitations, we emphasize that it remains highly practical and broadly useful in many real-world settings.

A key limitation is that RISE relies on human motion to stimulate diverse multipath paths. This motion provides the geometric variation necessary for uncovering occluded regions, and completely static environments remain challenging. However, this requirement aligns well with many deployment scenarios: homes, clinics, elder-care facilities, offices, and rehabilitation environments naturally contain frequent human movement. In such settings, RISE can operate fully passively and unobtrusively, without requiring additional devices, active user participation, or intrusive sensors. Even short or routine movements, walking through a hallway or moving to sit in a chair, can generate sufficient multipath diversity for reliable reconstruction.

Second, blind regions inherent to fixed sensor placement cannot be eliminated entirely, although our bi-angular enhancement significantly reduces them. These blind spots typically correspond to uncommon reflection geometries and often do not affect the global layout.

Third, the current system outputs 2D top-down geometry rather than full 3D meshes or semantic object models. While this limits fine-grained reconstruction, top-down geometry is already sufficient for many high-impact applications such as indoor navigation, elder monitoring, health assessment, fall detection, and safety analysis—domains where privacy constraints prohibit cameras.

Despite these limitations, RISE remains an effective, privacy-preserving perception system capable of recovering meaningful indoor structure from a single static radar. It offers a lightweight, low-cost alternative to camera- or LiDAR-based approaches, especially in privacy-sensitive environments. These strengths, combined with its ability to operate passively using natural human motion, underscore its practical value and establish a foundation for future advances in radar-based indoor understanding.