X-band Radar Non-Line-of-Sight Imaging (Supplementary Document)

Dongyu Du[1,2*] Mingkun Zhao[1*] Yutong Yang[3] Dominik Scheuble[3] Xiaolong Huang[1]
Zijian Shao[1] Mario Bijelic[1,4] Kaushik Sengupta[1] Felix Heide[1,4]
[1] Princeton University [2] University of Toronto [3] Mercedes-Benz AG [4] Torc Robotics

This supplementary document provides additional technical details and results for our X-band radar NLOS imaging system. Sec. 1 describes the hardware configuration of our prototype, the data acquisition pipeline, and statistics of the collected real-world dataset. Sec. 2 details the cross-modality simulation framework, including the X-band and 77 GHz radar and the SPAD–LiDAR simulator, together with statistics of the synthetic dataset. Sec. 3 expands on the neural reconstruction method, specifying the architectures of the dense prediction and geometry-aware recovery modules, as well as training procedures and hyperparameters. Finally, Sec. 4 presents baseline methods, additional simulated and experimental results that complement the main paper.

Contents

*indicates joint first authorship.

1. Experimental Setup

1.1. Experimental Prototype

X-band Radar Prototype. As illustrated in Fig. 1, the experimental X-band radar platform is an FMCW radar system built on FPGA (Xilinx ZCU102, AMD), integrating high-speed waveform generation, data acquisition, and digital signal processing. The host computer communicates with the FPGA Processing System over Ethernet to control each capture. The FPGA generates FMCW chirps with a bandwidth of $B = 700\,\text{MHz}$ and $N_c = 2^7$ chirps within a frame period of $T_f = 0.70\,\text{ms}$. It interfaces with the mixed-signal front end (MxFE, Analog Devices) via HPC connectors for high-speed conversion. The MxFE samples the waveform at a rate of $f_s = 1.5\,\text{GHz}$ and outputs a $4.5\,\text{GHz}$ FMCW signal, which is then up-converted to $10\,\text{GHz}$ by the XUD1a module and amplified to $-17\,\text{dBm}$ by a power amplifier.

Beam steering is performed using the beamforming front-end (STINGRAY, Analog), which includes an embedded amplifier and phase shifter. The board drives a 2×8 patch antenna array to transmit the radar signal and uses a separate 2×8 patch antenna array to receive the echo. The received signal is mixed with the reference signal to generate the beat frequency,

$$f_{\text{beat}} = \frac{2RS}{c},$$

where S is the chirp slope, c is the speed of light, and R is the distance of interest. Additional radar parameters are summarized in Table 1.

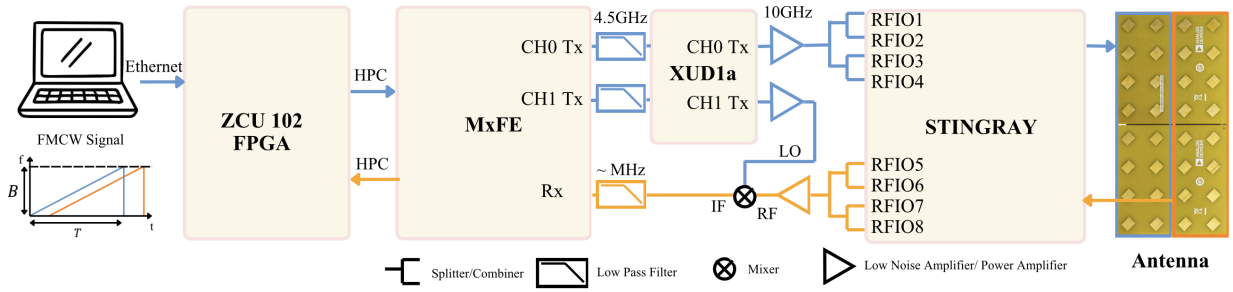

Figure 1. **X-band Radar Platform.** The host computer interfaces with the FPGA to control each capture, and the processing system of the FPGA controls the Analog-to-digital converter MxFE as well as the up-converter XUD1a to output 10 GHz FMCW signal to the antenna control board STINGRAY. The received signal is mixed with the reference signal and the distance information can be decoded from the beating frequency.

Table 1. Parameters of X-band Radar Platform.

Quantity	Symbol	Value
Bandwidth	B	$700\,\text{MHz}$
Number of Samples	N	2^{20}
Sampling Rate	f_s	$1.5\,\text{GHz}$
Number of Chirps	N_c	2^7
Central Frequency	f_c	$10\,\text{GHz}$
Wavelength	$\lambda = c/f_c$	$0.03\,\text{m}$
Single Chirp Period	$T_c = \frac{N}{f_s N_c}$	$5.46\,\mu\text{s}$
Chirp Slope	$S = B/T_c$	$1.28 \times 10^{14}\,\text{Hz/s}$
Total Frame Period	$T_f = N_c\, T_c$	$0.70\,\text{ms}$

System Interface. We develop an integrated interface that enables the X-band radar, LiDAR, and camera to capture simultaneously. The X-band radar is a custom 10 GHz FMCW system built on Analog Devices' hardware platform (details provided in the next section). The LiDAR (Ouster OS1) provides dense ground-truth point clouds with 128 channels and a range of up to 200 m. The camera (Aoni A30) supplies the corresponding ego view.

Doppler Limitation. In principle, our X-band radar system supports Doppler estimation through FMCW processing. However, under the current acquisition setting, the coherent observation time is short, yielding a coarse velocity resolution (approximately 21 m/s in our setup). This is insufficient for the motion scales considered in our target scenarios, so Doppler information is not used in the current reconstruction pipeline. In future implementations, increasing the number of chirps per frame or redesigning the acquisition and processing pipeline could better support velocity estimation. Incorporating Doppler cues could further improve robustness in dynamic scenes and help distinguish between different propagation paths.

1.2. Experimental Data Acquisition and Post-Processing

For each acquisition, the radar scans from –60° to 60° in 5° increments, collecting 2^{20} sampling points per angle. The scan is controlled by the phase shifter, which computes the required phase shift from the detection angle in real time. Its phase accuracy is 2.8° with 6-bit resolution. The acquisition time per scan is 0.3 s, limited by the readout speed. The raw data and post-processing code will be released. The frequency spectrum is estimated using Welch's method [6], and each static scene is scanned multiple times and averaged to reduce noise. Background is measured by scanning an open field and is used for background subtraction to mitigate flicker noise. The noise power spectral density (–74.24 dBm) is estimated by averaging background measurements across angles and is used in the simulation. The LiDAR point cloud is calibrated to the radar frame such that the origin and orientation are aligned. For each scene with occluded NLOS objects, multiple captures are taken from different viewpoints, with some recording the NLOS object's location. The point clouds captured from different positions are mapped to the radar coordinate frame so that the ground-truth positions of all objects in the scene are aligned. The positions of the NLOS objects, LOS objects, and the relay wall are annotated from a z-slice of the aligned point cloud and serve as ground truth for supervising the NLOS reconstruction algorithm.

1.3. Experimental Dataset Statistics

As reported in Fig. 2, we collected 122 captures across representative scenes to evaluate the proposed method under typical driving conditions. The dataset includes outdoor and indoor scenes, with outdoor data acquired during both daytime and nighttime, where nighttime conditions are challenging even for direct line-of-sight visibility. Across all captures, we include a wide range of hidden objects (cars, bicycles, etc.) and relay structures (building facades, walls, containers, etc.), leading to substantial variability in relay-wall type, incidence angle, and target layout. These captures are randomly partitioned into disjoint training and test splits at the sequence level, so that evaluation scenes are never seen during training. As shown in Sec. 4.3 of the main manuscript, our method generalizes well across this diverse set of real-world conditions and reliably reconstructs objects in hidden regions.

Figure 2. **Experimental Scene Examples.** We collected indoor and outdoor scenes under both daytime and nighttime conditions, encompassing diverse scenarios such as parking lots, blocks, and streets with varying relay wall materials and hidden objects.

2. Cross-Modal Simulation

To assess the benefit of long wavelengths for NLOS recovery in simulation, we compare the imaging performance of three representative sensing modalities: a SPAD-based LiDAR [3, 4, 7, 15] operating at an 850 nm wavelength, a 77 GHz automotive radar ($\lambda = 3.9$ mm) [13], and our proposed X-band radar system ($\lambda = 3$ cm). These systems span more than two orders of magnitude in wavelength and thus probe very different propagation and scattering regimes, from highly diffusive, fine-resolution optical sensing to increasingly specular, long-range microwave sensing.

To enable a fair cross-modality comparison, we construct a unified NLOS simulation framework in which all three modalities observe the same hidden scenes and relay geometries under matched viewpoint and occluder configurations.

In the following, we first describe the radar-side simulation, including the full X-band and 77 GHz signal models, antenna configurations, and implementation details, followed by statistics of the synthetic radar datasets used in our experiments. We then present the SPAD-based LiDAR simulation, specifying the pulsed source, temporal sampling, photon-counting and noise models, and how the same NLOS scenes are instantiated in the LiDAR setting. This provides details to reproduce our cross-modality comparison.

2.1. Simulation

2.1.1. Implementation Details

Existing automotive simulators such as CARLA [2] and AirSim [14] typically operate at the rendered-sensor level: radar returns are synthesized from geometric visibility, range, and simple RCS or noise models, without explicitly modeling the RF propagation chain or multi-bounce NLOS transport. In contrast, our radar simulator is an end-to-end forward model that follows the NLOS transport formulation in the main paper and explicitly comprises radar antenna simulation, scenario rendering, FDTD-based material reflectance estimation, multi-path ray tracing, and coherent RA-map formation. For clarity, we first restate the forward model and then describe how each simulation block instantiates its corresponding terms.

Forward Model. We denote the transmitting antenna position by $\mathbf{l}$, the receiving antenna position by $\mathbf{s}$, the relay wall by a plane

$$\Pi = \{\mathbf{x} = (x, y, z) \in \mathbb{R}^3 \mid \mathbf{n}^\top(\mathbf{x} - \mathbf{p}) = 0\},$$

with normal $\mathbf{n}$, root point $\mathbf{p}$, and the hidden object surface by O. A radar waveform $g(t)$ is emitted at angle $\boldsymbol{\theta}_t$, reflects at a first wall point $\mathbf{w}_1 \in \Pi$ with reflectance $\rho_\Pi(\mathbf{w}_1)$, scatters at an object point $\mathbf{o} \in O$ with reflectance $\rho_O(\mathbf{o})$, and returns via a second wall point $\mathbf{w}_2 \in \Pi$ with reflectance $\rho_\Pi(\mathbf{w}_2)$ to the receiver steered to angle $\boldsymbol{\theta}_r$. The received field is

$$\begin{aligned}\phi(\boldsymbol{\theta}_t, \boldsymbol{\theta}_r, t) = \iiint_{\Pi, O} & B_t(\boldsymbol{\theta}_t)\, B_r(\boldsymbol{\theta}_r)\, \rho_\Pi(\mathbf{w}_1)\, \rho_O(\mathbf{o})\, \rho_\Pi(\mathbf{w}_2) \\ & \times A(\mathbf{w}_1, \mathbf{o}, \mathbf{w}_2)\, g\left(t - \frac{L(\mathbf{w}_1, \mathbf{o}, \mathbf{w}_2)}{c}\right) \mathrm{d}\mathbf{w}_2 \mathrm{d}\mathbf{o} \mathrm{d}\mathbf{w}_1,\end{aligned} \tag{1}$$

where B_t and B_r are the transmit and receive beam patterns, A is the path attenuation, c is the propagation speed, and the total path length is

$$L(\mathbf{w}_1, \mathbf{o}, \mathbf{w}_2) = r_{\mathbf{l}\mathbf{w}_1} + r_{\mathbf{w}_1\mathbf{o}} + r_{\mathbf{o}\mathbf{w}_2} + r_{\mathbf{w}_2\mathbf{s}}, \quad r_{\mathbf{x}\mathbf{y}} = \|\mathbf{y} - \mathbf{x}\|.$$

For long-wavelength radar, the wall response contains a pronounced specular component. In our simulator we explicitly model this specular lobe and approximate the corresponding BRDF term by a Dirac kernel that enforces the law of reflection at $\mathbf{w}_1$:

$$\rho_\Pi(\mathbf{w}_1) \approx \alpha(\mathbf{w}_1)\, \delta\big(\mathbf{n}^\top \mathbf{r}_{\mathbf{w}_1\mathbf{o}} - \mathbf{n}^\top \mathbf{r}_{\mathbf{l}\mathbf{w}_1}\big), \tag{2}$$

where $\alpha(\mathbf{w}_1)$ is a material- and angle-dependent specular gain factor obtained from the FDTD wave-propagation stage of our pipeline. This collapses the surface integrals in (1) to the specular wall points $\mathbf{w}_1^\star, \mathbf{w}_2^\star$ and yields

$$\begin{aligned}\phi(\boldsymbol{\theta}_t, \boldsymbol{\theta}_r, t) \approx \int_O & \alpha(\mathbf{w}_1^\star)\alpha(\mathbf{w}_2^\star)\, B_t(\boldsymbol{\theta}_t)\, B_r(\boldsymbol{\theta}_r) \\ & \times A(\mathbf{w}_1^\star, \mathbf{o}, \mathbf{w}_2^\star)\, \rho_O(\mathbf{o})\, g\Big(t - \tfrac{L(\mathbf{w}_1^\star, \mathbf{o}, \mathbf{w}_2^\star)}{c}\Big)\, \mathrm{d}\mathbf{o}.\end{aligned} \tag{3}$$

Using mirror symmetry, the wall Π is treated as a virtual transparent interface: each object point $\mathbf{o}$ is mapped to a mirrored point $\mathbf{o}' = R_\Pi(\mathbf{o})$ behind the wall, and the two specular bounces become a single virtual line-of-sight path of length $L'(\mathbf{o}')$. The forward model reduces to

$$\phi(\boldsymbol{\theta}_t, \boldsymbol{\theta}_r, t) \approx \int_{O'} B_t(\boldsymbol{\theta}_t)\, B_r(\boldsymbol{\theta}_r)\, \tilde{\alpha}(\mathbf{o}')\, \tilde{A}(\mathbf{o}')\, \tilde{\rho}_O(\mathbf{o}')\, g\Big(t - \tfrac{L'(\mathbf{o}')}{c}\Big)\, \mathrm{d}\mathbf{o}'. \tag{4}$$

For specular or corner reflections, the effective path attenuation is modeled as

$$\tilde{A}(\mathbf{o}') = \frac{1}{L'(\mathbf{o}')^2}, \tag{5}$$

corresponding to spherical spreading over the total virtual range.

Below we describe how each block in Fig.5 of the main paper numerically instantiates these terms.

Radar Antenna Simulation. We model each radar front-end starting from the element pattern and array layout, and explicitly tie these to the beam-pattern terms $B_t(\boldsymbol{\theta}_t)$ and $B_r(\boldsymbol{\theta}_r)$ in Eqs. (1)–(4). Each radiating element is parameterized by an idealized directive pattern with azimuth and elevation half-power beamwidths of azHPBW $= 80^\circ$ and elHPBW $= 60^\circ$, a sidelobe floor of -25 dB, peak gain of 4.1 dBi, and a back-baffled response (no radiation into the backward hemisphere). The transmit power is fixed to $P_{\text{tx}} = 1.6$ mW and the receiver is assigned an effective noise figure of 8 dB; these parameters set the overall SNR level in the simulated RA-maps. The element patterns and powers are then combined with the array geometry and digital beamforming weights to synthesize the array factors $B_t(\boldsymbol{\theta}_t)$ and $B_r(\boldsymbol{\theta}_r)$ on a discrete steering grid that matches the real hardware configuration.

For the 77 GHz radar, we adopt a standard automotive configuration [13] with a 1×57 uniform linear array and approximately half-wavelength spacing in azimuth, achieving an angular resolution of about 1.8°. The X-band (10 GHz) radar uses a 4×8 uniform planar phased array matching our prototype, with inter-element spacings chosen close to $\lambda/2$ in both azimuth and elevation. This aperture provides an angular resolution of roughly 12.6°. In the simulator, both systems share the same transmit power and noise settings so that differences in NLOS performance primarily arise from wavelength, aperture size, and the resulting differences in the specular gain factor $\alpha(\mathbf{w}_1)$ and beam patterns B_t, B_r in the forward model.

Scenario Rendering. The urban scenes, relay walls, and hidden objects are constructed manually as 3D triangle meshes with material labels. The relay wall defines the plane Π and its normal $\mathbf{n}$, while hidden objects provide the surface O and candidate points $\mathbf{o}$ that contribute to the integral. For each object surface point, we precompute its mirrored position $\mathbf{o}' = R_\Pi(\mathbf{o})$ and store the geometry necessary for evaluating the virtual path length $L'(\mathbf{o}')$ in Eq. (4).

FDTD Propagation (Material-dependent Reflectance). To characterize how different wall materials reflect long-wavelength radar, we run 2D FDTD simulations for representative surfaces under varying incidence angles and wavelengths (X-band and 77 GHz). For each material and incidence angle, the FDTD solver produces the angular distribution of the scattered field, from which we extract the specular lobe and its energy. This energy ratio directly defines the specular gain factor $\alpha(\mathbf{w}_1)$ in the BRDF approximation of Eq. (2), while the residual non-specular energy is treated as diffuse loss and absorbed into the effective attenuation and noise models. In the forward model of Eqs. (3)–(4), $\alpha(\mathbf{w}_1)$ therefore encodes how strongly a given wall point contributes to the mirror-like NLOS path. FDTD results show that X-band exhibits a much stronger, narrower specular lobe (larger $\alpha(\mathbf{w}_1)$) than 77 GHz for typical urban materials, which is precisely the behavior we exploit in our NLOS simulations.

To further examine whether wall transmission needs to be modeled at X-band, we perform an FDTD simulation of 10 GHz wave propagation through a 40 cm concrete wall and place detectors on the reflection and transmission sides to measure the corresponding field intensities. The results show the transmission attenuation is much larger than the reflection attenuation, indicating that the transmitted energy is negligible compared with the reflected path in our setting. This validates our assumption that the dominant NLOS contribution arises from mirror-like reflection, while transmission through typical outdoor concrete relay walls can be safely ignored.

Multi-path Ray Tracing. We implement a geometric ray tracer that combines the rendered scene geometry, the mirror mapping, and the FDTD-based material tables. For each steering pair (θ_t, θ_r) and each candidate object point $\mathbf{o}$, the tracer first finds the specular wall point $\mathbf{w}_1^\star$ that satisfies the law of reflection implied by (2). In the general non-cofocal case, the transmit and receive paths intersect the relay wall at two specular points, $\mathbf{w}_1^\star$ and $\mathbf{w}_2^\star$. Here, however, we simulate a cofocal configuration, in which the transmit and receive paths share the same relay-wall point, so that $\mathbf{w}_2^\star = \mathbf{w}_1^\star$. We then check visibility and occlusion along the path $(\mathbf{l} \to \mathbf{w}_1^\star \to \mathbf{o} \to \mathbf{w}_1^\star \to \mathbf{s})$. The total path length is computed as $L(\mathbf{w}_1^\star, \mathbf{o}, \mathbf{w}_1^\star)$, and the corresponding virtual range $L'(\mathbf{o}')$ is evaluated, with the path attenuation $\tilde{A}(\mathbf{o}')$ computed via (5), optionally including additional atmospheric losses.

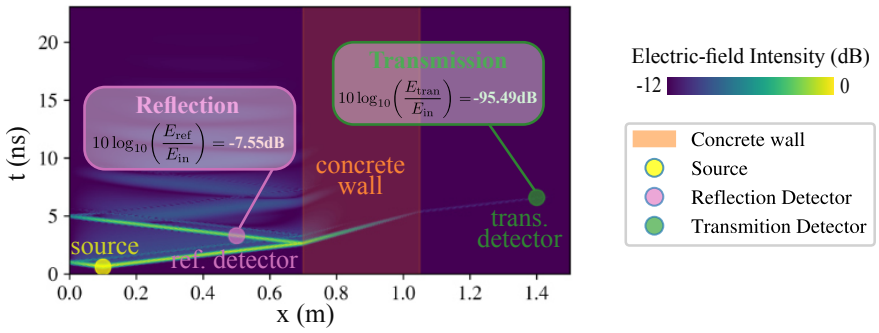

Figure 3. **FDTD simulation of 10 GHz wave propagation through a concrete wall.** The simulated field shows a dominant specular reflection from the wall, while the transmitted component is heavily attenuated and can be neglected in our setting.

The resulting complex weight

$$W(\mathbf{o}') = B_t(\mathbf{o}';\boldsymbol{\theta}_t)\, B_r(\mathbf{o}';\boldsymbol{\theta}_r)\, \tilde{\alpha}(\mathbf{o}')\, \tilde{A}(\mathbf{o}')\, \tilde{\rho}_O(\mathbf{o}')$$

is stored for each $(\boldsymbol{\theta}_t, \boldsymbol{\theta}_r, \mathbf{o}')$ triple and will be used to synthesize time-domain FMCW returns.

FMCW Waveform, Coherent detection, and RA-map Formation. Both the X-band and 77 GHz radars are driven by the same linear FMCW waveform, so that differences in NLOS performance are not confounded by range resolution. The transmitted chirp is

$$g(t) = \exp\left[j\left(2\pi f_c t + \pi(B/T_c)\, t^2\right)\right], \tag{6}$$

where the sweep bandwidth is fixed to $B = 700$ MHz, the ADC sampling rate to $F_s = 1.5$ GHz, the number of chirps per frame to $N_c = 2^7$, and the total number of fast-time samples per frame to $N_{\text{sample}} = 2^{20}$. The chirp duration is therefore

$$T_c = \frac{N_{\text{sample}}/F_s}{N_c},$$

which yields a range resolution of $\Delta R = c/(2B) \approx 0.21$ m for both modalities. The carrier frequency f_c is set to $f_c \approx 77$ GHz and $f_c \approx 10$ GHz for the millimeter-wave and X-band systems, respectively.

Given the path weights $W(\mathbf{o}')$ and virtual ranges $L'(\mathbf{o}')$ from Eq. (4), we synthesize the received time-domain waveform by sampling $t_n = n/F_s$ over the frame and summing the delayed chirps

$$\phi(\boldsymbol{\theta}_t, \boldsymbol{\theta}_r, t_n) \approx \sum_{\mathbf{o}'} W(\mathbf{o}')\, g\left(t_n - \tfrac{L'(\mathbf{o}')}{c}\right).$$

Coherent detection is implemented by mixing ϕ with the conjugate of the transmitted chirp and applying a low-pass filter, yielding the baseband beat signal

$$\begin{aligned} \tilde{m}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_r, t) &= \mathrm{LPF}\{\phi(\boldsymbol{\theta}_t, \boldsymbol{\theta}_r, t)\, g^*(t)\} \\ &\approx \int_{O'} W(\mathbf{o}')\, \exp\left[j\left(2\pi f_b(\mathbf{o}')\, t + \varphi(\mathbf{o}')\right)\right] \mathrm{d}\mathbf{o}', \end{aligned} \tag{7}$$

whose dominant frequency components encode the round-trip delays $L'(\mathbf{o}')$. For each steering pair $(\boldsymbol{\theta}_t, \boldsymbol{\theta}_r)$, we apply a 1D FFT along the fast-time dimension of $\tilde{m}$ to obtain the beat-frequency spectrum, map it to range bins using the known chirp parameters (B, T_c), and assemble the resulting range profiles over all steering angles into a simulated RA map. Complex Gaussian noise consistent with the assumed noise figure is added at the baseband prior to the FFT, so that the final RA statistics match those of the physical radar front-ends.

2.1.2. Synthetic Dataset Statistics

We synthesize a large-scale dataset using the proposed radar simulation pipeline, comprising 2,160 RA maps with corresponding ground-truth annotations. The dataset covers diverse outdoor environments, including urban streets, parking lots, residential blocks, and mixed traffic intersections, as illustrated in Fig. 4. Within these scenes, the

Figure 4. **Synthetic NLOS Dataset Scene Types.** The scenes cover diverse urban layouts (streets, parking lots, residential blocks) with varied relay interfaces (building facades, walls, vegetation) and hidden objects including cars, trucks, buses, and bicycles.

occluding relay surfaces are also varied: the dominant intermediate reflectors can be building facades, concrete walls, glass fronts, sound barriers, trees, and other vertical structures, leading to a wide range of wall normals, roughness levels, and specular gains. This diversity in relay geometry and material properties is crucial for testing the robustness of our X-band NLOS model.

The hidden targets include a broad set of everyday objects such as passenger cars, trucks, buses, vans, and bicycles, with different poses, aspect angles, and distances to both the relay wall and the radar platform. Multiple objects may be present in the same hidden region, producing overlapping LOS and mirror-NLOS responses in the RA domain. For each simulated frame, we randomize scene layouts, object categories, and positions to ensure that the dataset spans a wide variety of occlusion configurations and multi-bounce path topologies.

For every RA map, we generate geometric ground truth by evaluating the forward model on the underlying 3D scene. Specifically, we record the 3D coordinates of all visible LOS points on objects, their corresponding mirror-NLOS (mNLOS) virtual points behind the relay plane, and the true hidden NLOS object points that give rise to these echoes. Each non-empty RA cell is thus annotated with its associated 3D point(s) and a discrete label indicating whether the return is LOS, mNLOS, or NLOS. These annotations are used to supervise our reconstruction pipeline and to quantitatively evaluate localization accuracy in all experiments.

2.2. Transient SPAD-LiDAR Simulation

Our SPAD-based LiDAR simulator instantiates a physically grounded single-photon time-of-flight model on the same hidden scenes and relay geometries as the radar simulation. Instead of directly rendering range–azimuth (RA) maps, we first generate transient photon histograms for each visible path using a pulsed laser and SPAD detection model, and then project these histograms back onto the RA grid to obtain a LiDAR RA representation that is directly comparable to the radar measurements.

Forward Model. We simulate a confocal SPAD-based LiDAR configuration [3, 8] to generate NLOS signal, where a pulsed laser and a single-pixel SPAD share the same point $\mathbf{w} \in \Pi$ on the relay wall and are scanned jointly across Π. For a fixed wall point $\mathbf{w}$, light travels from $\mathbf{w}$ to a hidden point $\mathbf{o}$ and back to $\mathbf{w}$, with one-way path length $L(\mathbf{w}, \mathbf{o}) = \|\mathbf{o} - \mathbf{w}\|_2$, so that the round-trip delay is $2L(\mathbf{w}, \mathbf{o})/c$. The ideal transient flux measured at $\mathbf{w}$ can be

written as

$$\tau(\mathbf{w},t) = \int_O \frac{\rho_O(\mathbf{o})}{L(\mathbf{w},\mathbf{o})^4}\, s\!\left(t - \frac{2L(\mathbf{w},\mathbf{o})}{c}\right) \mathrm{d}\mathbf{o}, \tag{8}$$

where $\rho_O(\mathbf{o})$ denotes the object reflectance, $s(t)$ is the system impulse response (emitted pulse convolved with receiver response), and the factor L^{-4} accounts for two-way free-space spreading and Lambertian cosine terms along the wall–object–wall path. Thus, for each scanned wall point $\mathbf{w}$ we obtain a 1D transient $\tau(\mathbf{w},t)$ given by an integral over the hidden volume O. We discretize O into surface elements indexed by n, with centers $\mathbf{o}_n$, ranges $L_n = L(\mathbf{w},\mathbf{o}_n)$, and reflectances ρ_n. Equation (8) is then approximated as

$$\tau(\mathbf{w},t) \approx \sum_n \frac{\rho_n}{L_n^4}\, s\!\left(t - \frac{2L_n}{c}\right). \tag{9}$$

Photon-counting and Poisson Model. We model a pulsed laser at wavelength $\lambda = 850$ nm with average power P_{avg}, repetition rate f_{rep}, and total acquisition time T_{meas}. The pulse energy is

$$E_{\text{p}} = \frac{P_{\text{avg}}}{f_{\text{rep}}}, \quad E_{\text{ph}} = \frac{hc}{\lambda}$$

denotes the photon energy, with h Planck's constant and c the speed of light. For a path corresponding to a discrete element $\mathbf{o}_n$ at range L_n and incidence angle ψ_n, we propagate E_{p} through the optical chain (beam divergence, footprint, BRDF, and receiver étendue) and obtain a returned energy $E_{\text{echo}}(L_n,\psi_n)$. The expected number of detected signal photons per pulse is

$$\mu_{\text{sig}}(L_n,\psi_n) = \frac{E_{\text{echo}}(L_n,\psi_n)}{E_{\text{ph}}}\,\text{QE}, \tag{10}$$

where QE is the SPAD quantum efficiency. For each laser repetition, the number of detected signal photons from this path is drawn as

$$N_{\text{sig}} \sim \text{Poisson}\left(\mu_{\text{sig}}(L_n,\psi_n)\right), \tag{11}$$

and each photon is assigned a timestamp centered at $2L_n/c$, blurred by the system temporal point-spread function and SPAD timing jitter.

Ambient photons are modeled as a homogeneous Poisson process with effective rate $\Phi_{\text{amb}}^{\text{eff}}$. Over the total acquisition time T_{meas}, the number of ambient detections is

$$N_{\text{amb}} \sim \text{Poisson}\left(\Phi_{\text{amb}}^{\text{eff}} T_{\text{meas}}\right), \tag{12}$$

with timestamps sampled uniformly in $[0, T_{\text{meas}}]$. Signal and ambient timestamps are merged and sorted, after which SPAD non-idealities are applied: a dead time τ_{dead} enforces a minimum separation between accepted detections, and only the first detection within each laser repetition period is retained. In the implementation, the retained timestamps are first quantized to a temporal resolution of Δt_{quant}, and are then histogrammed over one repetition period $T_{\text{rep}} = 1/f_{\text{rep}}$ using N_{bin} bins. This yields a 1D transient histogram for each wall position w, with histogram bin width

$$\Delta t_{\text{hist}} = \frac{T_{\text{rep}}}{N_{\text{bin}}} = \frac{1/f_{\text{rep}}}{N_{\text{bin}}}.$$

At high photon flux this model naturally reproduces pile-up: early bins saturate while later bins are depleted due to the combination of dead time and first-photon selection.

Optical BRDF and Fresnel Reflectance. To compute $E_{\text{echo}}(L_n,\psi_n)$ in Eq. (10), we use an analytic BRDF model that combines Fresnel reflectance with a Rayleigh roughness term. Let n and k denote the real and imaginary parts of the complex refractive index $n_2 = n + jk$ of the surface, and let θ_i be the incidence angle between the incoming direction and the surface normal. For an interface between air ($n_1 = 1$) and the material (n_2), the Fresnel reflection coefficients for TE and TM polarization are

$$\begin{aligned} \Gamma_{\text{TE}}(\theta_i) &= \frac{n_1\cos\theta_i - n_2\cos\theta_t}{n_1\cos\theta_i + n_2\cos\theta_t}, \\ \Gamma_{\text{TM}}(\theta_i) &= \frac{n_2\cos\theta_i - n_1\cos\theta_t}{n_2\cos\theta_i + n_1\cos\theta_t}, \end{aligned} \tag{13}$$

where θ_t is the refracted angle given by Snell's law $n_1 \sin\theta_i = n_2 \sin\theta_t$. In the implementation, we define an effective Fresnel power factor by taking an equal-weight coherent average of the TE/TM reflection amplitudes and then squaring the magnitude,

$$|\Gamma(\theta_i)|^2 = \left|\tfrac{1}{2}\left(\Gamma_{\mathrm{TE}}(\theta_i) + \Gamma_{\mathrm{TM}}(\theta_i)\right)\right|^2. \tag{14}$$

Surface roughness with rms height σ_{rough} attenuates the coherent component according to a Rayleigh factor

$$\rho_{\text{rough}}(\theta_i) = \exp\left[-\left(\frac{4\pi\sigma_{\text{rough}}\cos\theta_i}{\lambda}\right)^2\right]. \tag{15}$$

The effective optical reflectance used in the simulation is then

$$\rho_{\text{opt}}(\theta_i) = |\Gamma(\theta_i)|^2\,\rho_{\text{rough}}(\theta_i), \tag{16}$$

which scales the BRDF term in $E_{\text{echo}}(L_n, \psi_n)$ and thus directly influences the expected signal photon count in Eq. (10).

Histogram Generation. To ensure that the LiDAR simulation sees the same geometric content as the radar, we start from the ground-truth radar RA maps produced by the radar simulator in Sec. 2.1.1. Each RA cell is labeled as either single-bounce LOS or double-bounce mNLOS. For a given scene and azimuth angle θ (corresponding to a wall point $\mathbf{w}(\theta) \in \Pi$), we read out all nonempty cells along that azimuth and obtain the corresponding path lengths $L_{\text{LOS},k}(\theta)$ and $L_{\text{mNLOS},k}(\theta)$. Each such cell is treated as a distinct optical path between $\mathbf{w}(\theta)$ and a hidden point $\mathbf{o}_k$, and we instantiate a confocal SPAD–LiDAR measurement for that path using the forward model in the previous paragraph: we compute the expected signal photon rate from the radiometric chain, draw Poisson-distributed signal and ambient photons over T_{meas}, apply dead time and first-photon selection, and bin timestamps into N_{bin} time bins. This yields one transient histogram $h_{\text{LOS},k}(t)$ for each LOS path and $h_{\text{mNLOS},k}(t)$ for each mNLOS path, where the latter includes an additional two-bounce attenuation factor derived from the optical BRDF. For each azimuth, the final transient at $\mathbf{w}(\theta)$ is then obtained by superposing all contributions,

$$h(\mathbf{w}(\theta), t) = \sum_k h_{\text{LOS},k}(t) + \sum_k h_{\text{mNLOS},k}(t). \tag{17}$$

RA Projection. To map the transient histograms back into a LiDAR RA representation, we convert the time-bin centers t_j at each wall position $\mathbf{w}(\theta)$ into range via $R_j = ct_j/2$, interpolate the histogram values $h(\mathbf{w}(\theta), t_j)$ onto the global radar range axis, and accumulate them into the corresponding RA column at azimuth θ. This procedure yields a synthetic LiDAR RA map $\mathrm{RA}_{\text{syn}}^{\text{LiDAR}}$ with the same range limits, number of range bins, and azimuth grid as the radar RA maps, ensuring a fair cross-modality comparison.

Simulation Parameters. Table 2 summarizes the main parameters of the SPAD–LiDAR simulation, including wavelength, laser power, repetition rate, system impulse-response width, SPAD timing jitter and dead time, quantum efficiency, ambient count rate, and the temporal sampling configuration. We adopt a moderately powered pulsed source and a single-pixel SPAD receiver with realistic timing characteristics, representative of state-of-the-art 850 nm LiDAR systems [4–6, 10].

The unambiguous range of the pulsed system is $R_{\text{amb}} = c/(2f_{\text{rep}}) \approx 10$ m; in our simulations, we discard targets whose radar-derived range exceeds R_{amb} to ensure that the SPAD LiDAR operates in its non-aliased regime. The RA grid (range limits, number of bins, and azimuth field of view) is chosen to match the radar configuration, enabling a direct cross-modality comparison.

3. Neural Reconstruction

Our neural reconstruction pipeline is designed to compensate for the inherently low angular resolution of X-band radar measurements, which leads to strong range–angle ambiguities and severely blurred RA responses. To address this, we formulate NLOS recovery as a learned mapping from raw RA data to hidden-scene geometry using two tightly coupled modules. A dense prediction module operates directly on complex-valued RA maps and produces high-resolution confidence maps for both LOS and mNLOS returns, effectively upsampling the angular resolution while disentangling different path types. A geometry-aware recovery module then leverages the predicted LOS/mNLOS

Table 2. Parameters of the simulated SPAD-based LiDAR system.

Quantity	Symbol	Value
Wavelength	λ	850 nm
Laser average power	P_{avg}	1.6 mW
Pulse repetition rate	f_{rep}	15 MHz
Measurement time	T_{meas}	0.03 s
System IRF FWHM	FWHM_{sys}	200 ps
SPAD timing jitter (rms)	σ_{jitter}	16 ps
SPAD dead time	τ_{dead}	4 ns
Quantum efficiency	QE	0.28
Receiver aperture radius		5 mm
Transmit / receive optics efficiency	$\eta_{\text{tx}}, \eta_{\text{rx}}$	$0.6, 0.6$
Effective ambient count rate	$\Phi_{\text{amb}}^{\text{eff}}$	3×10^7 s^{-1}
Time bins per period	N_{bin}	512
Timestamp quantization resolution	Δt	55 ps (≈ 8.3 mm one-way)
Field of view (azimuth)		$[-60^\circ, 60^\circ]$
Range window	$[R_{\min}, R_{\max}]$	$[0, 109.4]$ m

responses together with the mirror-reflection model of the wall to invert the NLOS transport and recover the true positions of hidden targets.

In the following, we first describe the architectural details of these two modules. We then present our training and implementation details, such as loss functions and optimization strategy, and finally summarize all hyperparameters to facilitate reproducibility of our neural reconstruction results.

3.1. Architectural Details

3.1.1. Dense Prediction Module

To extract precise peaks and path types from low–angular-resolution RA measurements, we formulate the first module as a dense, heatmap-based LOS/mNLOS prediction problem directly on the complex RA domain. Instead of applying CFAR or other local thresholding independently to each beam, the network jointly reasons over the full range–azimuth field to (i) sharpen blurred lobes into well-localized peaks and (ii) assign each peak to either a direct LOS return or a mNLOS reflection. This formulation explicitly turns the ill-posed deblurring and classification problem into a structured dense prediction task with strong spatial context.

We build this module on top of a Swin-UNet backbone [1], which we adapt to complex-valued radar data. Given a complex RA measurement $\kappa \in \mathbb{C}^{H\times W}$, we stack the real and imaginary parts of the complex RA map and pad them to three channels and partition it into non-overlapping patches that are embedded as tokens for the encoder. These tokens are processed by a hierarchy of Swin Transformer blocks [5] and patch-merging layers, whose shifted-window attention naturally captures long-range dependencies along both the range and azimuth axes. A symmetric decoder with patch-expanding layers then gradually restores the original spatial resolution while fusing encoder and decoder features through skip connections, which preserves fine spatial structure that is critical for resolving closely spaced targets. The stage-wise layer configuration of this backbone is summarized in Table 3. Given a complex RA measurement with physical size $H\times W = 512\times256$, we stack its real and imaginary parts and zero-pad along the azimuth axis to obtain a square 512×512 input for the network.

The final decoder features are projected by a 1×1 convolution into a two-channel tensor and passed through a sigmoid activation to obtain a dense confidence map $\mathbf{c} \in [0, 1]^{H\times W\times2}$, where the two channels encode the per-pixel probabilities of LOS points $\mathbf{w}$ and mNLOS reflections $\mathbf{o}'$, respectively. A local-maximum filter on $\mathbf{c}$ then extracts sparse peak sets for each class, suppressing low-confidence responses while retaining only well-supported detections. These peaks serve simultaneously as high-precision range–angle estimates and as class-labeled geometric anchors for the subsequent geometry-aware recovery module.

3.1.2. Geometry-aware Recovery Module

The geometry-aware recovery module operates on the sparse set of detected LOS and mNLOS points and refines the analytic mirror reconstruction using a lightweight attention-based point network. We first recall the geometry-aware

Table 3. Network configuration of the dense prediction module.

Stage	Operation	#Blocks	Heads	Output size
Input	RA map	–	–	$512 \times 512 \times 3$
Patch embed	4×4 patch, stride 4	–	–	$128 \times 128 \times 96$
Enc-1	Swin blocks + PatchMerging	2	3	$64 \times 64 \times 192$
Enc-2	Swin blocks + PatchMerging	2	6	$32 \times 32 \times 384$
Enc-3	Swin blocks + PatchMerging	6	12	$16 \times 16 \times 768$
Bottleneck	Swin blocks	2	24	$16 \times 16 \times 768$
Dec-3	PatchExpand (upsample $\times2$)	–	–	$32 \times 32 \times 384$
Dec-2	concat skip (Enc-2), Swin_up & PatchExpand	6	12	$64 \times 64 \times 192$
Dec-1	concat skip (Enc-1), Swin_up & PatchExpand	2	6	$128 \times 128 \times 96$
Dec-0	concat skip (Patch embed), Swin_up	2	3	$128 \times 128 \times 96$
Output	FinalPatchExpand_X4 & 1×1 Conv	–	–	$512 \times 512 \times 2$

reflection model used in the main paper. For each detected mNLOS point $\mathbf{o}'$, the hidden point $\mathbf{o}$ is reconstructed as

$$\mathbf{o} = R_\Pi(\mathbf{o}') + \Delta\mathbf{o}_\theta = \big(\mathbf{o}' - 2(\mathbf{n}^\top\mathbf{o}' + b)\mathbf{n}\big) + \Delta\mathbf{o}_\theta, \tag{18}$$

where $\mathbf{n}$ and b denote the normal and offset of the local relay wall, $R_\Pi(\cdot)$ is the analytic mirror mapping with respect to this wall, and $\Delta\mathbf{o}_\theta$ is a learned residual correction. The architecture described below is precisely designed to provide this residual term in a geometry-aware manner.

Geometric Preprocessing. We apply DBSCAN clustering to the detected LOS points to identify wall segments, and within each cluster we fit a line via RANSAC to obtain a stable local wall normal $\mathbf{n}$ and offset b. Each mNLOS point $\mathbf{o'}_i$ is then associated with its most plausible LOS cluster by minimizing the angular difference in polar coordinates between $\mathbf{o'}_i$ and all LOS points. Given the assigned cluster and fitted wall parameters $(\mathbf{n}_i, b_i)$, we form a purely geometric estimate

$$\mathbf{o}_{\text{geom},i} = \mathbf{o}'_i - 2\big(\mathbf{n}_i^\top\mathbf{o}'_i + b_i\big)\mathbf{n}_i. \tag{19}$$

which corresponds to Eq. (18) with $\Delta\mathbf{o}_\theta = \mathbf{0}$. This estimate is accurate for ideal planar, specular reflections but becomes biased in the presence of wall curvature, sparse sampling, or clustering errors.

Local Patch Encoding and Attention. To correct these biases, we parameterize the residual term $\Delta\mathbf{o}_{\theta,i}$ in Eq. (18) with a small neural network, *ResidualReflectNet*, conditioned on the local LOS structure around each mNLOS point. For every $\mathbf{o'}_i$, we collect a local "patch" of up to K nearest LOS points $\{\mathbf{w}_{i,j}\}_{j=1}^K$ from the same DBSCAN cluster and express them in a local frame relative to $\mathbf{o'}_i$. A shared point encoder $\phi(\cdot)$ and a query encoder $\psi(\cdot)$ produce point features $\mathbf{e}_{i,j} = \phi(\mathbf{w}_{i,j} - \mathbf{o'}_i)$ and a query embedding $\mathbf{q}_i = \psi(\mathbf{o'}_i)$. We then use dot-product attention to aggregate information over the patch

$$\alpha_{i,j} = \frac{\exp(\mathbf{q}_i^\top\mathbf{e}_{i,j})}{\sum_{j'=1}^K \exp(\mathbf{q}_i^\top\mathbf{e}_{i,j'})}, \qquad \mathbf{v}_{\text{patch},i} = \sum_{j=1}^K \alpha_{i,j}\,\mathbf{e}_{i,j}, \tag{20}$$

where $\alpha_{i,j}$ are attention weights and $\mathbf{v}_{\text{patch},i}$ is a compact descriptor summarizing local wall geometry (orientation, curvature, sampling density) around $\mathbf{o'}_i$.

State Encoding and Residual Prediction. Besides local geometry, the residual also depends on the global reflection state. We construct a low-dimensional state vector by concatenating the mNLOS point, the fitted line parameters, and a reliability flag,

$$\mathbf{s}_i = [\,\mathbf{o'}_i,\ \mathbf{n}_i,\ b_i,\ \eta_i\,],$$

where $\eta_i \in \{0, 1\}$ indicates whether $\mathbf{o'}_i$ was attached to a valid wall segment or treated as noise. A state encoder MLP $\chi(\cdot)$ maps this vector to a state embedding $\mathbf{v}_{\text{state},i} = \chi(\mathbf{s}_i)$. The patch and state embeddings are then concatenated and passed through a decoder MLP $f_\theta(\cdot)$ to predict the residual and final hidden point

$$\mathbf{z}_i = [\,\mathbf{v}_{\text{patch},i},\ \mathbf{v}_{\text{state},i}\,], \quad \Delta\mathbf{o}_{\theta,i} = f_\theta(\mathbf{z}_i), \quad \mathbf{o}_i = \mathbf{o}_{\text{geom},i} + \Delta\mathbf{o}_{\theta,i}. \tag{21}$$

Table 4. Network configuration of the geometry-aware recovery module.

Block	Operation	Output shape
Input	mNLOS point, 16 LOS neighbours, line params, flag	$(128, 2),\ (128, 16, 2),\ (128, 3),\ (128, 1)$
Patch encoder	Neighbor MLP $2 \to 64 \to 128 \to 128$	$(128, 16, 128)$
Query encoder	Query MLP $2 \to 64 \to 128$	$(128, 128)$
Attention aggregation	Dot-product attention over 16 neighbours	$(128, 128)$
State encoder	State MLP $6 \to 64 \to 64$	$(128, 64)$
Decoder	Decoder MLP $192 \to 256 \to 128 \to 2$	$(128, 2)$

This design cleanly separates interpretable geometric reasoning from data-driven refinement. DBSCAN and RANSAC provide explicit estimates of the relay wall and analytic mirror mapping (Eq. (18)), while the attention-based ResidualReflectNet only learns local, context-dependent corrections (Eqs. (20)–(21)) informed by neighboring LOS structure and the global line state. As a result, the geometry-aware recovery module remains lightweight and stable, yet can compensate for wall curvature, roughness, and sparse or noisy LOS samples, yielding accurate hidden-object locations even under challenging NLOS configurations.

We set the number of LOS neighbours per mNLOS point to 16, batch size to 128 and the embedding dimension to 64. The network configuration is summarized in Table 4.

3.2. Training Details

We train the neural reconstruction method by supervising the dense prediction and geometry-aware recovery modules separately on the synthetic radar dataset described in Sec. 2.1.2. All experiments are implemented in PyTorch and run on a single GPU; we use the same train/validation/test splits for all methods to enable fair comparison.

Data Splits and Preprocessing. From the 2,160 simulated X-band RA maps, we construct non-overlapping training, validation, and test sets using fixed index lists stored on disk. Each RA sample is a complex-valued tensor in range–azimuth coordinates; we represent it as a two-channel real-valued input (real and imaginary parts) and apply per-sample normalization before feeding it to the network. The corresponding ground-truth labels consist of (i) dense LOS/mNLOS annotations on the RA grid and (ii) 3D coordinates for all LOS, mirror-NLOS, and true NLOS points, projected into the relay-wall coordinate system for the geometry-aware module.

Dense Prediction Module. We supervise the dense prediction module using Gaussian heatmaps and a focal loss adapted from CenterNet [17]. For ground truth generation, we create a two-channel Gaussian heatmap $\mathbf{Y} \in [0, 1]^{H \times W \times 2}$. Each channel $k \in \{1 = \text{LOS}, 2 = \text{mNLOS}\}$ corresponds to one reflection class. Given the ground-truth peak locations $\mathcal{P}_k = \{(x_i, y_i)\}$ for class k on the RA grid, the target heatmap value at position (x, y) for that channel is generated using a Gaussian kernel centered at the nearest peak

$$Y_{xyk} = \max_{(x_i, y_i) \in \mathcal{P}_k} \exp\left[-\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma^2}\right], \tag{22}$$

ensuring that $Y_{xyk} = 1$ only at the exact peak locations (x_i, y_i).

Training minimizes a heatmap focal loss between the predicted confidence map $\mathbf{c} \in [0, 1]^{H \times W \times 2}$ and the target heatmap $\mathbf{Y}$

$$\mathcal{L}_{\text{heat}} = -\frac{1}{N} \sum_{x,y,k} \begin{cases} (1 - c_{xyk})^{\alpha} \log(c_{xyk}), & Y_{xyk} = 1, \\ (1 - Y_{xyk})^{\beta} (c_{xyk})^{\alpha} \log(1 - c_{xyk}), & \text{otherwise}, \end{cases} \tag{23}$$

where α and β control the focusing strength ($\alpha = 2,\ \beta = 4$ in all experiments). The loss is normalized by $N = \sum_{x,y,k} \mathbf{1}_{\{Y_{xyk}=1\}}$, representing the total number of ground truth peaks across all classes. This objective encourages accurate localization of reflection peaks while jointly distinguishing LOS and mNLOS peaks.

In addition to the Gaussian heatmap-based focal loss, we evaluated a standard binary segmentation counterpart, where pixel-wise classification is supervised using the binary LOS/mNLOS ground truth without Gaussian filtering. However, this approach yielded uncompetitive performance and led to training instability. We attribute this failure to

the extreme sparsity of the LOS/mNLOS peaks: the imbalance between background and foreground pixels means the sparse binary targets fail to provide sufficient gradient signals to guide optimization. This confirms that the Gaussian heatmap formulation is superior to naive binary segmentation for this task.

Geometry-aware Recovery Module. We train *ResidualReflectNet* to predict the residual term $\Delta\mathbf{o}_\theta$ in Eq. (18). For each detected mNLOS point $\mathbf{o'}_i$ in the training set, we first construct the analytic mirror reconstruction $\mathbf{o}_{\text{geom},i}$ by applying the fitted wall parameters $(\mathbf{n}_i, b_i)$ to Eq. (18) with $\Delta\mathbf{o}_\theta = \mathbf{0}$. The ground-truth hidden point $\mathbf{o}_i^{\text{gt}}$ is known from the synthetic scene geometry, allowing us to define a residual supervision signal

$$\Delta\mathbf{o}_{\theta,i}^{\text{gt}} = \mathbf{o}_i^{\text{gt}} - \mathbf{o}_{\text{geom},i}. \tag{24}$$

Given the query point $\mathbf{o'}_i$, its local LOS patch and state vector, the network outputs a residual prediction $\Delta\mathbf{o}_{\theta,i}^{\text{pred}}$ (cf. Eq. (21)). Then, the training objective is designed as an ℓ_1 regression loss over these residuals,

$$\mathcal{L}_{\text{res}} = \frac{1}{N}\sum_{i=1}^{N} \left\|\Delta\mathbf{o}_{\theta,i}^{\text{pred}} - \Delta\mathbf{o}_{\theta,i}^{\text{gt}}\right\|_1, \tag{25}$$

where N is the number of mNLOS samples in a mini-batch. This loss directly measures the absolute error of the predicted correction in meters, and empirically provides better robustness to outliers and heavy-tailed geometric errors than an ℓ_2 objective.

In practice, we backpropagate the residual loss primarily on geometrically reliable mNLOS samples, while still providing the reliability flag η_i as an input feature so that the network can distinguish clean from ambiguous configurations. Although the analytic mirror mapping captures a first-order planar approximation, the ground-truth residuals $\Delta\mathbf{o}_{\theta,i}^{\text{gt}}$ encode higher-order effects induced by wall curvature, surface roughness, clutter, and sparse LOS sampling. Consequently, the residual field is highly structured and heterogeneous across scenes, and the geometry-aware module must exploit both the local wall context and the global line state to produce consistent corrections, rather than simply fitting small random offsets.

3.3. Hyperparameters

Both the dense prediction and geometry-aware recovery modules are trained with the same global optimization settings. We use the Adam optimizer with default PyTorch parameters (betas $0.9/0.999$, no weight decay) and an initial learning rate of 1×10^{-4}. Training is performed on a single GPU with mixed-precision enabled, and we monitor performance on a held-out validation split to select the final checkpoint.

For the dense prediction module, we train the model for 100 epochs using a mini-batch size of 4. To balance accuracy and computational speed, we adopt the Swin-T (Tiny) variant as the encoder, initialized with ImageNet [12] pretrained weights following [1]. Consistent with the original Swin-UNet hyperparameters, we set the input patch size to 4×4, the window size to 8×8, the embedding dimension to 96, and the MLP expansion ratio to 4.

For the geometry-aware recovery module (*ResidualReflectNet*), we use a mini-batch size of 128 mNLOS points and train for 100 epochs. The local LOS patch size is fixed to $K = 16$ neighbors per mNLOS point. The internal embedding dimensions are set to 64 for the query and point encoders, 128 for the aggregated patch feature, and 64 for the state embedding; the decoder MLP uses hidden widths of 256 and 128 and outputs a 2D residual. The residual regression is optimized with an ℓ_1 loss in meters.

3.4. Ablation Experiments

We perform two ablation experiments on the simulated dataset to isolate the contributions of the dense prediction backbone and the geometry-aware reconstruction head. In all cases, we keep the training schedule, loss functions, and supervision signals identical to the full model and only vary the component under study.

Dense Prediction Backbone. We first replace the Swin-UNet dense prediction module with a conventional UNet [11], keeping the input RA representation, output heatmap definition, and post-processing identical. Both networks are trained to produce two-channel LOS/mNLOS confidence maps on the same synthetic training split. We evaluate dense prediction quality using a macro-averaged F1 score over LOS and mNLOS peaks (Macro-F1) and the corresponding macro Chamfer distance (Macro-CD), and then feed the detected peaks into the same geometry-aware recovery module to measure the final NLOS reconstruction F1 and CD. As shown in Table 5, Swin-UNet substantially improves

Table 5. Ablation experiments on simulated data.

Ablation		Dense Prediction		Reconstruction	
		Macro-F1↑	Macro-CD↓	F1↑	CD↓
Dense Prediction	UNet [11]	0.67	44.0	0.27	11.93
	Swin-UNet [1]	**0.85**	**17.3**	**0.45**	**8.16**
Reconstruction	w/o residual	0.85	17.3	0.40	9.20
	w/ residual	0.85	17.3	**0.45**	**8.16**

peak localization and classification (Macro-F1 from 0.67 to 0.85, Macro-CD from 44.0 to 17.3), which in turn yields a large gain in NLOS reconstruction performance (F1 from 0.27 to 0.45, CD from 11.93 to 8.16). This confirms that transformer-based global context is particularly beneficial for resolving low–angular-resolution radar lobes.

Geometry-aware Residual Head. We compare the full geometry-aware recovery module against a purely analytic variant without residual learning. In the "w/o residual" variant, hidden points are obtained by applying only the mirror reflection in Eq. (19), using the same line estimation pipeline but setting the learned residual $\Delta\mathbf{o}_\theta$ to zero. Since both variants share the same dense predictions, their Macro-F1 and Macro-CD are identical; the ablation therefore directly measures how much the residual head improves the NLOS reconstruction given the same inputs. Table 5 shows that the residual correction increases reconstruction F1 from 0.40 to 0.45 and reduces CD from 9.20 to 8.16, indicating that the network learns to compensate for wall curvature, surface roughness, and other non-idealities that violate the ideal planar mirror assumption.

4. Additional Results

4.1. Baseline Methods

To validate the effectiveness of our neural reconstruction approach, we compare it against three radar-based reconstruction methods that we adapt to the LOS/mNLOS setting considered in this work. All baselines operate on the same X-band RA inputs as our method and are evaluated at two levels: (i) dense prediction of LOS and mNLOS responses on the RA grid, and (ii) final NLOS object reconstruction when the outputs of (i) are fed into our geometry-aware recovery module. In the following, we provide additional detail on all baseline methods.

CFAR-NLOS [13]. This baseline is a classical Constant False Alarm Rate (CFAR) detector. CFAR estimates a noise floor for each RA cell by pooling statistics over a local neighborhood (excluding guard cells around the test cell) and adapts the detection threshold so that the probability of false alarm remains approximately constant across the map. This yields a set of high-SNR "blobs" in RA space, but CFAR has no ability to distinguish LOS from mNLOS returns. Scheiner et al. [13] extended CFAR to NLOS recovery by assuming a known relay wall and using an auxiliary LiDAR to estimate LOS wall normals; CFAR peaks that project consistently onto the wall are treated as LOS, and farther detections behind the wall are interpreted as mNLOS. To apply this approach in our setting, we provide CFAR with privileged LOS information: we run CFAR on the RA maps, and then associate each detection with the nearest ground-truth LOS point in RA space. Detections within a small range–angle tolerance of a LOS ground-truth cell are labeled as LOS, while all remaining detections are labeled as mNLOS. This gives CFAR access to the same wall information that Scheiner et al. obtain from LiDAR, and allows us to directly compare its LOS/mNLOS separation with our dense prediction module.

Further Than CFAR [10]. This baseline is the learned "Further than CFAR" model [10], originally proposed as a heatmap-based radar detector. In its original form, the network predicts an occupancy map indicating the presence of objects beyond CFAR detections. We adapt this architecture to our problem by replacing the final layer with a two-channel output that produces per-pixel logits for LOS and mNLOS classes on the RA grid. The backbone, input encoding, and training protocol follow the original design, but the supervision is now given by our LOS/mNLOS ground-truth labels derived from the simulation. At test time, we threshold the predicted confidence maps and extract

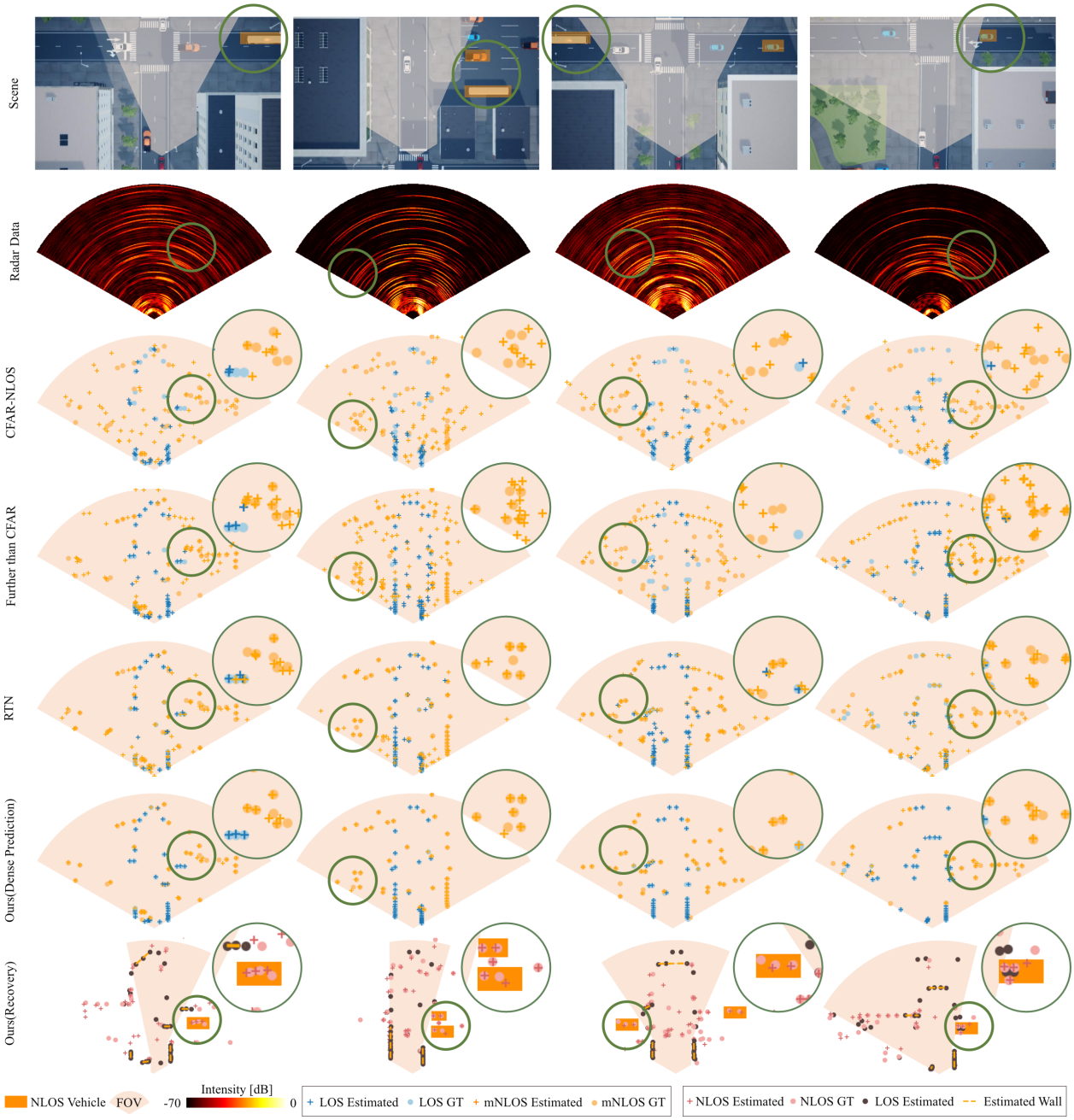

Figure 5. **Additional Qualitative NLOS Reconstructions in Simulation.** Each column corresponds to a different scenario (top row), with the second row showing the simulated X-band radar RA data. Rows three to five visualize LOS/mNLOS detections from CFAR-NLOS [13], Further than CFAR [10], and RTN [9], respectively, while the last two rows show our dense prediction output and the final geometry-aware recovery. Green circles highlight regions where our method more accurately localizes and classifies hidden objects compared to the baselines, and the solid rectangular blocks in the bottom row illustrate the precise reconstruction of NLOS vehicles in the hidden scene.

LOS/mNLOS point sets, which are then used both for dense prediction evaluation and as inputs to our geometry-aware recovery module.

RTN [9]. This baseline is the Radar Tensor Network (RTN) [9], a network architecture originally developed for object detection in range–angle(-Doppler) space. RTN predicts object-level bounding boxes and classes; we adapt it

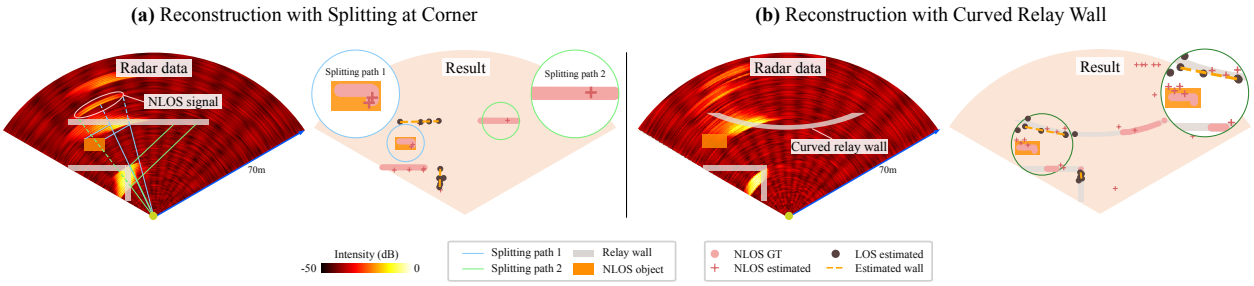

Figure 6. **Reconstruction with Beam Splitting at Corner and Curved Relay Wall.** (a) A wide beam incident on a corner leads to path splitting; our method can correctly trace multiple propagation paths and recover the scene. (b) The method remains robust under non-planar (curved) relay walls and achieves accurate reconstruction.

to LOS/mNLOS recovery by replacing the detection head with a head that predicts LOS/mNLOS/noise classification scores as well as another head that predicts a two-dimensional offset vector that is added to the respective anchor points. We use their variant with the 2D convolutional backbone (2D-DCB), as this is a good fit for our RA input grid. For fair comparison, we adopt the initial anchor points to our resolution. The network is trained end-to-end using the same complex RA inputs and LOS/mNLOS labels as in the "Further Than CFAR" baseline. At inference time, anchors classified as noise are suppressed, and for LOS/mNLOS points, the two-dimensional offsets are added to each anchor position to output the final point cloud.

Evaluation Protocol. For all three baseline methods, we first compare their dense LOS/mNLOS predictions against ground truth using per-class F1 scores and Chamfer Distance (CD) on the RA grid. This directly measures how accurately each method localizes and classifies LOS and mNLOS responses. To isolate the effect of dense prediction quality from geometric inversion, we then feed each method's predicted LOS/mNLOS point sets into the same geometry-aware recovery module and evaluate the resulting NLOS reconstructions using F1 and CD. Under this unified geometric back-end, improvements in final NLOS reconstruction quality can be attributed to more accurate and better calibrated LOS/mNLOS predictions, highlighting the advantage of our proposed dense prediction module over both classical CFAR and existing learned baselines.

4.2. Additional Simulation Results

Figure 5 reports additional qualitative results on synthetic scenes. Across all test scenarios, our method first produces accurate dense predictions in the RA domain: the dense prediction module yields well-localized and cleanly separated LOS (blue) and mNLOS (orange) point sets, with significantly fewer missed detections and spurious responses than CFAR-NLOS, Further than CFAR, and RTN. These reliable LOS/mNLOS estimates then provide informative geometric anchors for the geometry-aware recovery module, which leverages the predicted wall structure and mirror relations to reconstruct compact, correctly recovered NLOS vehicle clusters behind the relay wall (bottom row). In contrast, errors in the baseline dense predictions carry over to the recovered layouts, resulting in fragmented, displaced, or entirely missing hidden objects, which underlines the advantage of our coupled dense prediction and geometry-aware reconstruction modules.

Reconstruction with Beam Splitting at Corner. At X-band frequencies, the longer wavelength leads to a wider effective beamwidth, which can cause beam splitting when interacting with wall edges or corners [16]. In such cases, different portions of the beam follow distinct multipath trajectories, resulting in mixed LOS and NLOS returns that are challenging to disentangle.

Our method addresses this issue by leveraging both local geometric cues and global contextual information to jointly reason about multiple propagation paths. This enables accurate tracing of both LOS and NLOS components and allows the model to distinguish between different paths. As shown in Fig. 6(a), our approach successfully reconstructs the scene under beam-splitting conditions.

Reconstruction with Curved Relay Wall. In practical scenarios, relay walls might be non-planar, and surface curvature introduces deviations from ideal specular reflection. Unlike flat surfaces, where a single reflection normal

defines the propagation path, curved walls induce spatially varying normals, leading to angular distortions and reconstruction errors if not properly modeled. This effect becomes more pronounced under limited angular resolution, making accurate NLOS reconstruction more challenging.

Our method addresses this issue through the geometry-aware reconstruction module, which explicitly models local reflection geometry and learns residual corrections to account for deviations from ideal mirror reflections. This allows the model to remain robust to moderate curvature in the relay surface.

We evaluate our method on curved relay walls and observe accurate reconstruction up to a curvature of $0.025\,\mathrm{m}^{-1}$, as shown in Fig. 6(b). The results demonstrate that our approach effectively compensates for curvature-induced distortions and maintains reliable scene recovery.

4.3. Additional Experimental Results

Real-world NLOS reconstruction is more challenging than the synthetic setting due to calibration errors, unknown wall geometry, diffuse clutter and non-ideal antenna patterns. These effects break some assumptions used in the forward model and lead to stronger speckle, side-lobe artefacts, and spurious multi-path responses in the measured RA maps. As a result, classical detectors and learned baselines tend to produce many more false positives on the relay wall as well as missed detections in the far range, which directly degrades the quality of the recovered hidden scene.

Figure 7 provides additional qualitative real-world results. We evaluate our method on daytime and nighttime scenarios with different NLOS objects (cars, vans, bicycles) and relay materials (concrete barriers, metal containers, building facades). Across all scenes, the dense prediction module produces compact, well-localized LOS and mNLOS clusters despite the strong artefacts in the raw radar data, whereas CFAR-NLOS, Further Than CFAR, and RTN frequently hallucinate extended point clouds along the relay wall or miss distant NLOS vehicles. The geometry-aware recovery module then uses these reliable LOS/mNLOS estimates to reconstruct consistent BEV layouts of the hidden vehicles behind the wall (bottom row), closely matching the manually annotated ground truth and maintaining accurate relative positions at ranges up to $40\,\mathrm{m}$ (corresponding to an $80\,\mathrm{m}$ round-trip path). In contrast, errors in the baselines' dense predictions carry over to the recovered layouts, producing fragmented, displaced, or missing hidden objects and underscoring the benefit of our coupled dense prediction and geometry-aware reconstruction modules. Together with the quantitative comparisons in the main paper, these results confirm that the proposed approach remains accurate and robust under realistic sensing conditions.

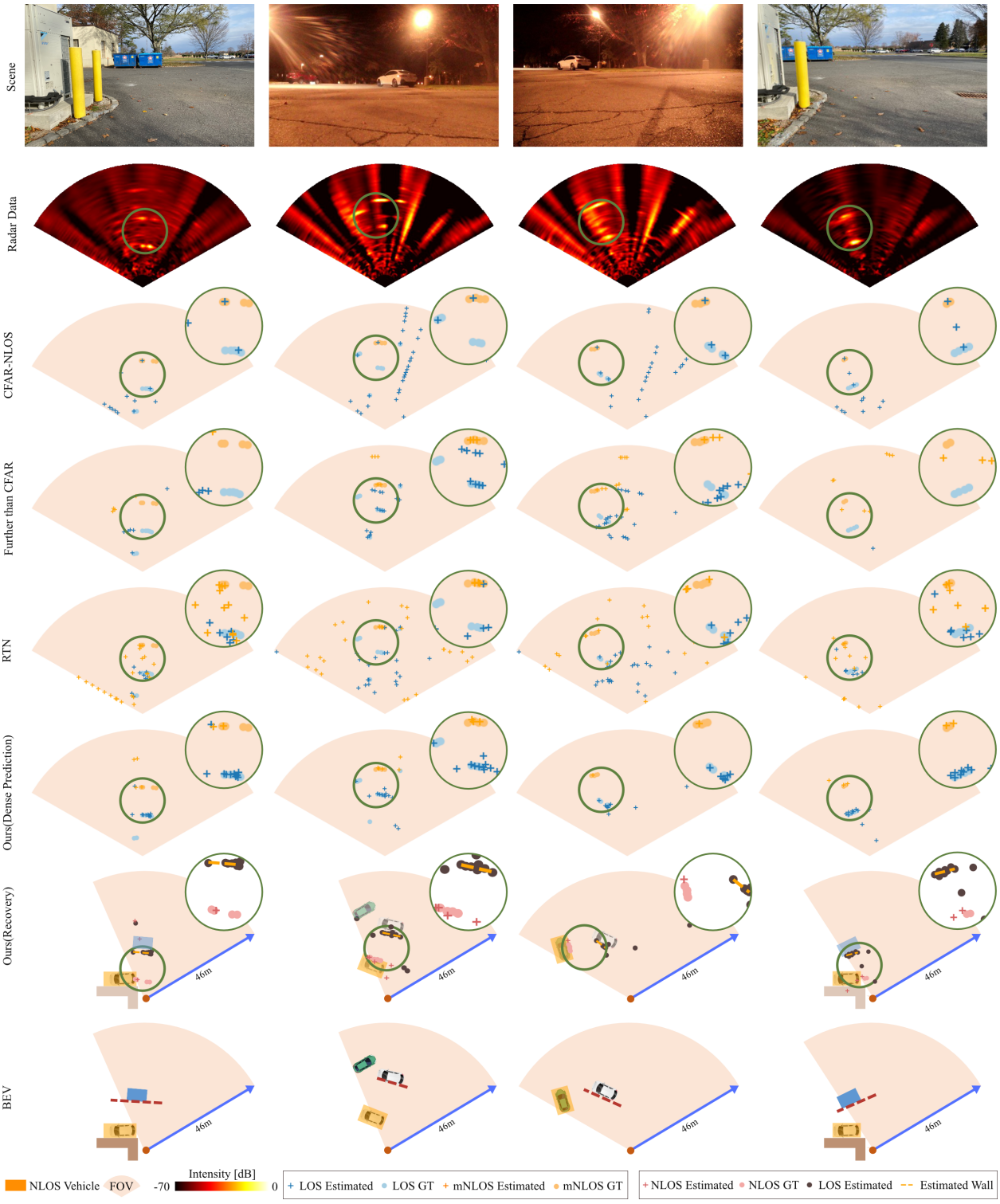

Figure 7. **Additional Experimental Results.** Each column corresponds to a different scenario (top row), with the second row showing the experimental X-band radar RA data. Rows three to five visualize LOS/mNLOS detections from CFAR-NLOS [13], Further than CFAR [10], and RTN [9], respectively, with the last two rows showing the geometry-aware recovery compared to the bird's-eye view of the ground truth scene. Green circles highlight regions where our method more accurately localizes and classifies hidden objects compared to the baselines, and the solid rectangular blocks in the bottom row illustrate the precise reconstruction of NLOS vehicles in the hidden scene.

References

[1] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-UNet: UNet-like pure transformer for medical image segmentation. In *European Conference on Computer Vision (ECCV)*, pages 205–218, 2022. 9, 12, 13

[2] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *Conference on robot learning*, pages 1–16. PMLR, 2017. 3

[3] David B. Lindell, Matthew O'Toole, and Gordon Wetzstein. Wave-based non-line-of-sight imaging using fast f-k migration. *ACM Transactions on Graphics*, 38(4):116:1–116:13, 2019. 2, 6

[4] Xintong Liu, Jianyu Wang, Zhupeng Li, Zuoqiang Shi, Xing Fu, and Lingyun Qiu. Non-line-of-sight reconstruction with signal–object collaborative regularization. *Light: Science & Applications*, 10(1):198, 2021. 2

[5] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10012–10022, 2021. 9

[6] Alan V. Oppenheim and Ronald W. Schafer. *Digital Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1975. 2

[7] Matthew O'Toole, David B. Lindell, and Gordon Wetzstein. Real-time non-line-of-sight imaging. In *ACM SIGGRAPH 2018 Emerging Technologies*, pages 1–2. 2018. 2

[8] Matthew O'Toole, David B Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555(7696):338–341, 2018. 6

[9] Dong-Hee Paek, Seung-Hyun Kong, and Kevin Tirta Wijaya. K-radar: 4d radar object detection for autonomous driving in various weather conditions. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022. 14, 17

[10] Ignacio Roldan, Andras Palffy, Julian F. P. Kooij, Dariu M. Gavrila, Francesco Fioranelli, and Alexander Yarovoy. See further than CFAR: A data-driven radar detector trained by lidar. In *Proceedings of the 2024 IEEE Radar Conference (RadarConf24)*, 2024. 13, 14, 17

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 12, 13

[12] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015. 12

[13] Nicolas Scheiner, Florian Kraus, Fangyin Wei, Buu Phan, Fahim Mannan, Nils Appenrodt, Werner Ritter, Jurgen Dickmann, Klaus Dietmayer, Bernhard Sick, and Felix Heide. Seeing around street corners: Non-line-of-sight detection and tracking in the wild using doppler radar. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2065–2074. IEEE, 2020. 2, 4, 13, 14, 17

[14] Shital Shah, Debadeepta Dey, Chris Lovett, and Ashish Kapoor. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and service robotics: Results of the 11th international conference*, pages 621–635. Springer, 2017. 3

[15] Shumian Xin, Sotiris Nousias, Kiriakos N. Kutulakos, Aswin C. Sankaranarayanan, Srinivasa G. Narasimhan, and Ioannis Gkioulekas. A theory of fermat paths for non-line-of-sight shape reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6793–6802, 2019. 2

[16] Shichao Yue, Hao He, Peng Cao, Kaiwen Zha, Masayuki Koizumi, and Dina Katabi. Cornerradar: Rf-based indoor localization around corners. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(1):1–24, 2022. 15

[17] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. Objects as points. *arXiv preprint arXiv:1904.07850*, 2019. 11