

Reflection-Aware Reasoning for Non-Line-of-Sight Pedestrian Localization

Byeonggyu Park¹, Mingu Jeon^{2, (✉)}, and Seong-Woo Kim^{1, (✉)}

¹ Seoul National University, Seoul, Republic of Korea
`{bg.park, snwoo}@snu.ac.kr`

² Pusan National University, Busan, Republic of Korea
`mingujeon@pusan.ac.kr`

Abstract. Reliable localization of non-line-of-sight (NLOS) pedestrians is critical for safe urban autonomous driving, yet it remains highly challenging in ego-dynamic outdoor environments, where ego-vehicle motion makes radar multipath propagation complex and noisy. In this paper, we present a reflection-aware framework for NLOS pedestrian localization with a moving ego-vehicle in outdoor testbed scenarios. Our framework fuses front-view camera images and 2D radar point clouds to infer reflection orders and reflective surface distributions in bird’s-eye-view space. It then uses physics-guided ray tracing to reconstruct distorted reflection paths and localize the hidden pedestrian. We validate the framework in outdoor testbed scenarios under ego-dynamic conditions. The results demonstrate the effectiveness of the proposed framework for NLOS pedestrian localization with a moving ego-vehicle.

Keywords: NLOS perception, mmWave radar, multipath reflections, multimodal perception, pedestrian localization, collision avoidance

1 Introduction

Urban driving environments contain dense structural occlusions caused by buildings, walls, and noise barriers. Such structures create non-line-of-sight (NLOS) regions where pedestrians are fully hidden from direct observation. In urban environments, sudden pedestrian emergence from occluded areas poses a persistent safety challenge. In 2023, approximately 84% of pedestrian fatalities in the United States occurred in urban areas, and pedestrian deaths accounted for about 18% of all traffic fatalities [22, 26]. The prevalence of fatalities in structurally complex urban settings underscores the need for perception beyond direct line-of-sight (LOS), as illustrated in Fig. 1.

To prevent accidents caused by pedestrians emerging from occluded regions, recent studies have explored methods for detecting objects in NLOS areas. Representative approaches include vehicle-to-everything (V2X) systems [12, 13, 18], reflective wave-based sensing modalities such as acoustic sensing [9, 28] and mmWave radar [5, 10, 11]. V2X-based methods can provide information about

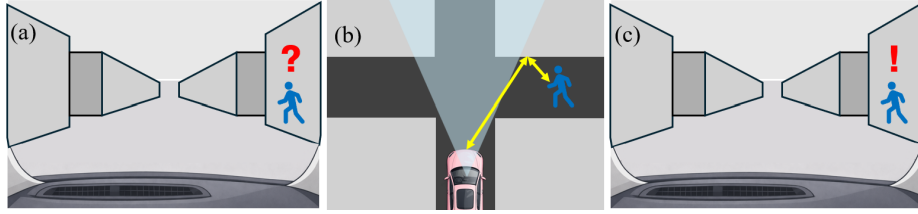


Fig. 1: Illustration of an NLOS pedestrian scenario from the ego-vehicle perspective. (a) While driving through a narrow alley, regions between buildings fall outside the LOS, making direct detection of pedestrians difficult. (b) By leveraging reflective wave propagation, indirect spatial evidence can be obtained even when the target is occluded. (c) Based on reflection-aware reasoning, the position of the NLOS object can be inferred despite the absence of direct visibility.

hidden objects through communication, but require prior infrastructure deployment [8]. Sound-based NLOS perception leverages diffraction and reflection properties of acoustic waves [1, 21, 28], yet often assumes predefined spatial layouts and remains sensitive to environmental noise, making robust localization in urban driving environments challenging. mmWave radar-based methods exploit multipath reflections to estimate occluded objects [29, 37]. These approaches typically rely on predefined geometric assumptions or rule-based modeling to interpret reflection paths [10, 37] and are commonly validated in controlled indoor environments [17, 36]. Consequently, most existing NLOS perception methods [1, 5, 9–11, 24, 29, 37] assume static sensing setups and stable propagation patterns, which limits their applicability in outdoor ego-dynamic scenarios.

Achieving reliable NLOS perception with radar point cloud data (PCD) in driving environments presents several challenges. Under ego-dynamic conditions, radar observations vary across frames due to vehicle motion and are often corrupted by measurement noise and clutter. Therefore, identifying radar points that originate from NLOS pedestrians becomes challenging. In addition, observations of NLOS pedestrians appear spatially distorted because radar signals propagate through multi-bounce reflection paths. As a result, accurate estimation of reflective surfaces is required to reconstruct reflection paths and localize occluded objects.

To address these challenges, we propose a practical reflection-aware learning framework. The proposed framework consists of two stages: Reflection-aware representation learning and Physics-guided ray tracing. In the reflection-aware representation learning stage, feature-level fusion is performed between front-view camera images providing visual cues and semantic information in the LOS region and 2D radar PCD containing quantitative range measurements and reflection characteristics. Based on the fused representations, the model performs reflection-type point segmentation to classify radar points according to their reflection mechanisms. In addition, reflective surface distributions are estimated as heatmaps in bird’s-eye-view (BEV) space. In the physics-guided ray tracing

Table 1: Comparison of representative NLOS perception and localization studies. Existing approaches often rely on predefined geometry or rule-based reasoning and are mostly evaluated in static or controlled settings. In contrast, the proposed method performs learning-based NLOS point extraction and reflective surface estimation, and is validated under ego-dynamic conditions.

Method	Sensors	Methodology		Validation Setting			
		Learned Points	Learned Surface	Outdoor	Vehicle Val.	Multi-Target	Ego Motion
[9]	Mic Array			✓	✓		
[37]	Radar			✓		✓	
[10]	Radar			✓	✓	✓	
[16]	Radar	✓	✓				✓
[31, 32]	Radar				✓		
[27]	Radar, LiDAR	✓		✓	✓		
[11, 24]	Radar, Camera		✓	✓	✓	✓	
Proposed	Radar, Camera	✓	✓	✓	✓	✓	✓

stage, the estimated reflective structures and the identified radar points are used to interpret reflection paths and physically reconstruct propagation trajectories.

To evaluate the effectiveness of the proposed method, experiments are conducted in a constructed outdoor T-junction intersection testbed. The evaluation is performed under both ego-static and ego-dynamic conditions using the average Euclidean localization error. The proposed method achieves localization errors of 0.44 m and 1.01 m for LOS and NLOS pedestrians under ego-static conditions, and 0.54 m and 1.23 m under ego-dynamic conditions, outperforming existing approaches [24, 27]. The results demonstrate reliable localization of NLOS pedestrians even under ego-dynamic driving conditions.

The main contributions of this paper are as follows:

- We propose a practical reflection-aware framework for NLOS pedestrian localization with a moving ego-vehicle in outdoor testbed scenarios.
- We develop a reflection-aware method that combines multimodal representation learning and physics-guided ray tracing to recover distorted radar propagation and localize hidden pedestrians.
- We present experimental validation of the proposed method with a moving ego-vehicle in an outdoor testbed intersection.

Code, data, and models are available at <https://bingurrr.github.io/reflection-aware-nlos>.

2 Related Work

Acoustic-based NLOS perception. Several approaches have leveraged acoustic waves for non-line-of-sight perception by exploiting their reflection and diffraction properties [1, 2, 21, 28]. Prior work has estimated object directions using microphone arrays [2] or inferred positions by analyzing acoustic reflection paths [1–3]. However, these approaches are primarily evaluated in indoor or static environments with limited noise, restricting applicability to urban driving scenarios.

In outdoor settings, sound-based methods have been explored for vehicle state classification [9] and NLOS vehicle localization using particle filters [7]. Nevertheless, these approaches either focus on coarse state inference or rely on predefined geometric maps, limiting their ability to infer spatial structure directly from acoustic observations in dynamic environments.

Radar-based NLOS perception. Unlike acoustic waves, mmWave radar provides accurate range measurements, enabling indirect observation of occluded objects through multipath reflections [29, 32, 37]. Early studies explored around-the-corner radar sensing using geometric reflection modeling and multipath exploitation [31, 32]. Later works further investigated localization methods based on multipath radar signals and joint estimation of propagation parameters [37]. While effective in controlled settings, such approaches often rely on predefined wall or corner geometry and prior scene knowledge, limiting their applicability to unknown or dynamic urban environments. Learning-based methods have also been introduced to analyze radar point clouds for NLOS perception. These approaches classify radar points into reflection categories or detect NLOS-associated points using data-driven models [15, 27]. More recent studies have explored high-resolution mmWave imaging and RF sensing techniques to reconstruct occluded objects behind reflectors [5, 17]. However, such systems are typically limited to indoor or short-range scenarios with static targets. Recent radar localization studies continue to exploit multipath propagation for NLOS target positioning [4, 33]. Nevertheless, many of these methods rely on auxiliary structural information or controlled environments, limiting their applicability to outdoor driving scenarios.

Camera–radar fusion for NLOS perception. To compensate for sparse and noisy radar measurements, several studies have explored camera–radar fusion for object detection and scene understanding, where visual semantic cues complement radar observations [6, 20, 35]. However, these approaches are primarily designed for LOS perception and do not explicitly address multipath effects in NLOS scenarios. For NLOS settings, prior work has applied particle filtering to detect occluded pedestrians [23] or leveraged camera-derived structural cues, such as road layout and reflective surfaces, to guide radar interpretation [11, 24]. Nevertheless, radar point interpretation remains largely based on predefined heuristics rather than data-driven modeling of multipath propagation.

Table 1 summarizes representative NLOS perception and localization studies with respect to their methodological components and validation settings. Prior studies have made substantial progress in around-corner sensing, target localization, and radar-based reconstruction; however, they are predominantly evaluated in static, controlled, or ego-static scenarios. In contrast, our paper addresses outdoor ego-dynamic NLOS pedestrian localization using a moving vehicle-mounted radar-camera platform, including multi-target localization scenarios.

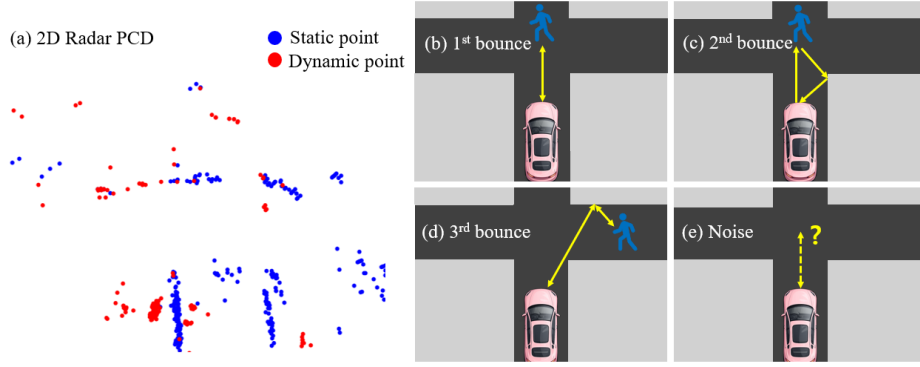


Fig. 2: Reflection path diversity in mmWave radar observations. (a) mmWave 2D Radar PCD is observed in BEV and appears sparse and noisy, making informative observations difficult to identify. (b)–(d) illustrate representative multipath propagation cases (1st, 2nd, and 3rd bounce), where the geometric strategy for object localization varies depending on the reflection order. Although higher-order reflections may exist in practice, we model up to third-order reflections considering mmWave attenuation characteristics. (e) shows structurally inconsistent noise that further complicates reflection path interpretation.

3 Problem Definition

To formally describe the reflection reasoning problem addressed in this work, we first define the radar reflection path representation used for NLOS interpretation.

Reflection Path Definition. Each radar point corresponds to a propagation path characterized by the number of reflections the signal undergoes before returning to the sensor. In this work, we consider up to third-order reflections, taking into account the attenuation characteristics of mmWave signals. Accordingly, each radar point is categorized into first, second, or third-order reflection paths. Furthermore, depending on whether a dynamic object is involved along the propagation path, each point is additionally classified as either object-originated or surface-originated reflection, as shown in Fig. 2. This labeling scheme jointly encodes the reflection order and the physical origin of the signal, resulting in a structured reflection-path representation.

Input Representation. We consider a time-discrete sequence of synchronized sensor measurements indexed by $t \in \{1, \dots, T\}$ where T denotes the total number of time steps in the sequence. At each time step t , the ego-vehicle observes a front-view RGB image $\mathbf{I}_t$ and a 2D Radar PCD $\mathbf{R}_t = \{\mathbf{r}_{t,i}\}_{i=1}^{N_t}$.

Each radar point is represented as

$$\mathbf{r}_{t,i} = (x_{t,i}, y_{t,i}), v_{t,i}, \rho_{t,i},$$

where $(x_{t,i}, y_{t,i})$ denotes planar coordinates in the ego-centric BEV frame, $v_{t,i}$ is the Doppler velocity, and $\rho_{t,i}$ represents the radar cross section (RCS). The

index i identifies an individual radar point at time t , and N_t denotes the number of radar points in $\mathbf{R}_t$, which varies across frames.

Problem Formulation. Given synchronized inputs $(\mathbf{I}_t, \mathbf{R}_t)$ at time t , the proposed model predicts two complementary intermediate representations: (i) point-wise reflection-path labels $\hat{\mathbf{Y}}_t = \{\hat{y}_{t,i}\}_{i=1}^{N_t}$ and (ii) a reflective surface probability map in BEV space $\hat{\mathbf{S}}_t \in [0, 1]^{H_{\text{bev}} \times W_{\text{bev}}}$.

Let f_θ denote the proposed reflection-aware model parameterized by θ . The model is formally defined as

$$f_\theta : (\mathbf{I}_t, \mathbf{R}_t) \rightarrow (\hat{\mathbf{Y}}_t, \hat{\mathbf{S}}_t). \quad (1)$$

$\hat{\mathbf{Y}}_t$ and $\hat{\mathbf{S}}_t$ are combined through a physics-guided NLOS inference operator $\mathcal{T}(\cdot)$:

$$\hat{\mathbf{X}}_t = \mathcal{T}(\hat{\mathbf{Y}}_t, \hat{\mathbf{S}}_t). \quad (2)$$

The final objective is to minimize the localization error e_t between the predicted object position $\hat{\mathbf{X}}_t$ and the ground-truth position $\mathbf{X}_{\text{gt},t}$:

$$e_t = \left\| \hat{\mathbf{X}}_t - \mathbf{X}_{\text{gt},t} \right\|_2. \quad (3)$$

4 Method

4.1 Overview

To achieve this, we develop a learning-based framework that performs feature-level fusion between front-view camera images, which provide structural and semantic cues in LOS regions, and radar point clouds containing quantitative range measurements and reflection characteristics, as shown in Fig. 3. The model predicts reflection-type labels for radar points and estimates reflective surface distributions in BEV space, which are then integrated into a physics-guided ray tracing module to interpret reflection paths and infer the final NLOS pedestrian location.

4.2 Reflection-Aware Representation Learning

While cameras provide geometric and visual layout of the scene, radar offers physical observations of NLOS objects through multipath propagation. To preserve the complementary properties of both modalities, we perform feature-level fusion to learn reflection-aware representations for NLOS pedestrian localization.

The input front-view image $\mathbf{I}_t$ is transformed into BEV space using a Lift-Splat-Shoot (LSS)-based image encoder f_{cam} [25]. The LSS encoder predicts per-pixel depth distributions and projects image features onto the BEV plane, generating a spatially structured visual representation:

$$\mathbf{F}_{\text{cam}} = f_{\text{cam}}(\mathbf{I}_t), \quad \mathbf{F}_{\text{cam}} \in \mathbb{R}^{C \times H_{\text{bev}} \times W_{\text{bev}}}. \quad (4)$$

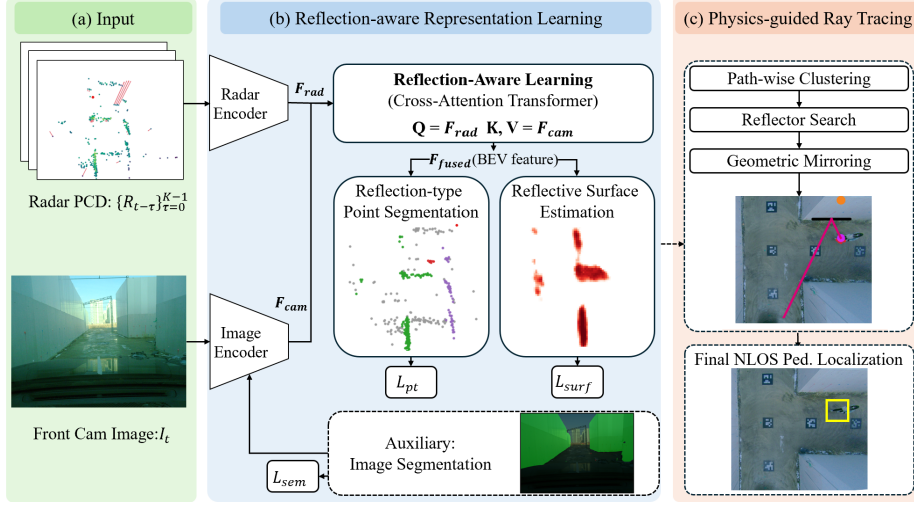


Fig. 3: Overall framework of the proposed reflection-aware NLOS localization method. (a) **Input.** A temporal stack of radar point clouds $\{R_{t-\tau}\}_{\tau=0}^{K-1}$ and a synchronized front-view image I . (b) **Reflection-aware representation learning.** Radar point features F_{rad} and BEV-aligned visual features F_{cam} are extracted and fused via cross-attention to obtain F_{fused} , from which reflection-type point labels and a reflective surface heatmap are predicted. (c) **Physics-guided ray tracing.** Reflection-consistent clusters are formed, candidate reflectors are identified, and the NLOS pedestrian position is recovered via geometric mirroring.

Here, C denotes the feature channel dimension, and H_{bev} , W_{bev} denote the spatial resolution of the BEV feature map.

To stabilize semantic feature learning, an auxiliary semantic segmentation head is attached to the image encoder during training. The auxiliary loss is defined as a cross-entropy term $\mathcal{L}_{sem}$. This head is used only during training and discarded at inference time.

For radar, the PCD observed at a single time step t is extremely sparse and noisy. To mitigate this limitation, we aggregate radar observations from the past K frames. Specifically, $\{\mathbf{R}_{t-\tau}\}_{\tau=0}^{K-1}$ are ego-motion compensated and aligned to the current coordinate frame before accumulation. The aggregated radar points are encoded in BEV space using a radar encoder f_{rad} :

$$\mathbf{F}_{rad} = f_{rad}(\{\mathbf{R}_{t-\tau}\}_{\tau=0}^{K-1}), \quad \mathbf{F}_{rad} \in \mathbb{R}^{C \times H_{bev} \times W_{bev}}. \quad (5)$$

The radar BEV feature $\mathbf{F}_{rad}$ and camera BEV feature $\mathbf{F}_{cam}$ are fused via a transformer-based cross-attention module. To allow sparse radar features to reference dense spatial structures, we set radar features as queries and camera features as keys and values:

$$\mathbf{Q} = \mathbf{W}_Q \mathbf{F}_{rad}, \quad \mathbf{K} = \mathbf{W}_K \mathbf{F}_{cam}, \quad \mathbf{V} = \mathbf{W}_V \mathbf{F}_{cam}. \quad (6)$$

The fused representation is then obtained as

$$\mathbf{F}_{\text{fused}} = \text{Softmax} \left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}} \right) \mathbf{V}. \quad (7)$$

The resulting fused representation $\mathbf{F}_{\text{fused}}$ jointly encodes the physical reflection characteristics captured by radar and the structured spatial cues provided by the camera. This reflection-aware feature representation serves as the shared input for both the reflection-type point segmentation and reflective surface estimation modules.

4.3 Reflection-Type Point Segmentation

To interpret multipath propagation, we classify the reflection order of each radar point in $\mathbf{R}_t$ at the point level. Specifically, the BEV feature vector corresponding to the spatial location of $\mathbf{r}_{t,i}$ is sampled from $\mathbf{F}_{\text{fused}}$ and passed through a point segmentation head g_θ to predict the reflection-type logits $\mathbf{p}_{t,i}$:

$$\mathbf{p}_{t,i} = g_\theta(\mathbf{F}_{\text{fused}}(\mathbf{r}_{t,i})), \quad \mathbf{p}_{t,i} \in \mathbb{R}^{|\mathcal{C}|}, \quad (8)$$

where $\mathcal{C}$ denotes the set of reflection-type classes corresponding to different reflection orders. The predicted reflection label is obtained as

$$\hat{y}_{t,i} = \arg \max_{c \in \mathcal{C}} \mathbf{p}_{t,i}^{(c)}. \quad (9)$$

The point segmentation head is applied independently to each radar point and trained as a multi-class classification problem using a point-wise cross-entropy loss ℓ_{CE} :

$$\mathcal{L}_{\text{pt}} = \frac{1}{N_t} \sum_{i=1}^{N_t} \ell_{\text{CE}}(\mathbf{p}_{t,i}, y_{t,i}), \quad (10)$$

where $y_{t,i}$ is the ground-truth reflection-type label for point $\mathbf{r}_{t,i}$, and N_t denotes the number of radar points at time t .

Through this process, informative observations associated with NLOS pedestrians are identified from sparse and noisy $\mathbf{R}_t$ at the reflection-order level. This decomposition separates mixed multipath signals into physically interpretable reflection paths, which subsequently serve as inputs to reflective surface estimation and physics-based NLOS reconstruction.

4.4 Reflective Surface Estimation

Although point segmentation identifies higher-order reflection points, recovering the true location of an NLOS pedestrian requires additional geometric constraints that describe the structures responsible for the observed reflections. Since NLOS observations arise from reflections on surrounding environmental

structures rather than direct object returns, the location and shape of reflective surfaces must be estimated to interpret the underlying propagation paths.

To estimate reflective structures responsible for multipath propagation, the fused BEV representation $\mathbf{F}_{\text{fused},t}$ is used to predict a reflective surface probability map at time step t :

$$\hat{\mathbf{S}}_t = h_\phi(\mathbf{F}_{\text{fused},t}), \quad \hat{\mathbf{S}}_t \in [0, 1]^{H_{\text{bev}} \times W_{\text{bev}}}, \quad (11)$$

where h_ϕ denotes the reflective surface prediction head parameterized by ϕ . Let (u, v) denote a grid coordinate in BEV space. Each element $\hat{\mathbf{S}}_t(u, v)$ represents the probability that a reflective surface exists at location (u, v) .

To supervise reflective surface estimation, we construct a pseudo ground-truth surface map $\mathbf{S}_t$. Because radar PCD from a single frame are highly sparse, radar observations from temporally adjacent frames are aggregated after ego-motion compensation. Specifically, radar PCD from K_h past frames and K_f future frames relative to time step t are aligned to the coordinate frame at time t . The aggregated radar point set $\tilde{\mathbf{R}}_t$ is defined as

$$\tilde{\mathbf{R}}_t = \bigcup_{\tau=-K_h}^{K_f} \mathbf{R}_{t+\tau}. \quad (12)$$

Among the aggregated radar points, only those classified as first-order surface reflections are retained to form a candidate reflective surface set $\mathbf{R}_t^{\text{surf}}$. Each selected point is then projected onto the BEV grid using anisotropic Gaussian splatting to generate the pseudo surface map:

$$\mathbf{S}_t(u, v) = \max_j \exp \left(-\frac{1}{2} \left[\frac{d_{\parallel,j}^2}{\sigma_{\parallel}^2} + \frac{d_{\perp,j}^2}{\sigma_{\perp}^2} \right] \right), \quad (13)$$

where (u, v) denotes a BEV grid location. The terms $d_{\parallel,j}$ and $d_{\perp,j}$ represent the distances from grid cell (u, v) to point $\mathbf{r}_j$ along the estimated tangent and normal directions of the reflective surface, respectively. The parameters $\sigma_{\parallel}$ and $\sigma_{\perp}$ control anisotropic smoothing, with $\sigma_{\parallel} > \sigma_{\perp}$ to preserve elongated reflective structures while suppressing diffusion across the normal direction.

Reflective surface estimation is formulated as a binary classification problem over BEV grid cells and optimized using the binary cross-entropy loss ℓ_{BCE} :

$$\mathcal{L}_{\text{surf}} = \ell_{\text{BCE}}(\hat{\mathbf{S}}_t, \mathbf{S}_t). \quad (14)$$

4.5 Physics-Guided Ray Tracing

In this stage, $\hat{\mathbf{y}}_{t,i}$ and $\hat{\mathbf{S}}_t$ are used to recover the location of NLOS pedestrians through physics-based ray tracing. The module operates as a deterministic inference procedure grounded in geometric optics and reflection physics. It consists of three steps: (i) path-wise clustering, (ii) ray-guided reflector search, and (iii) geometric mirroring.

Path-wise Clustering. Given the $\mathbf{R}_t$, radar points are first separated according to $\hat{y}_{t,i}$. The first-order reflection point set $\mathcal{P}_t^{(1)}$ and third-order reflection point set $\mathcal{P}_t^{(3)}$ are defined as

$$\mathcal{P}_t^{(1)} = \{\mathbf{r}_{t,i} \mid \hat{y}_{t,i} = 1\}, \quad \mathcal{P}_t^{(3)} = \{\mathbf{r}_{t,i} \mid \hat{y}_{t,i} = 3\}. \quad (15)$$

Second-order reflections are excluded from geometric inference because they typically occur under LOS conditions and are accompanied by more reliable first-order reflections.

DBSCAN clustering is then applied to $\mathcal{P}_t^{(1)}$ and $\mathcal{P}_t^{(3)}$ to obtain representative cluster centers, denoted as $\{\bar{\mathbf{r}}_k^{(1)}\}$ and $\{\bar{\mathbf{r}}_m^{(3)}\}$, respectively. The centers $\bar{\mathbf{r}}_k^{(1)}$ approximate the true object location, while $\bar{\mathbf{r}}_m^{(3)}$ correspond to higher-order reflections that do not coincide with the actual object position.

Ray-Guided Reflector Search. Each third-order cluster center $\bar{\mathbf{r}}_m^{(3)}$ is assumed to originate from reflections on surrounding structures. Therefore, the corresponding reflective surface must lie along the ray connecting the radar origin $\mathbf{o}$ and $\bar{\mathbf{r}}_m^{(3)}$. The unit ego-ray direction is defined as

$$\mathbf{d}_m = \frac{\bar{\mathbf{r}}_m^{(3)} - \mathbf{o}}{\|\bar{\mathbf{r}}_m^{(3)} - \mathbf{o}\|_2}. \quad (16)$$

Candidate reflector positions are sampled along this ray. For each candidate position, the reflection likelihood is evaluated using the reflective surface probability map $\hat{\mathbf{S}}_t$. The reflector geometry $(\mathbf{c}_m, \mathbf{n}_m)$ is selected as the candidate that maximizes the accumulated reflection consistency.

Geometric Mirroring. Given the estimated $(\mathbf{c}_m, \mathbf{n}_m)$ and $\bar{\mathbf{r}}_m^{(3)}$, the NLOS pedestrian position is recovered via geometric mirroring with respect to the reflective surface. The recovered position $\hat{\mathbf{x}}^{(3)}$ is computed as

$$\hat{\mathbf{x}}_m^{(3)} = \bar{\mathbf{r}}_m^{(3)} - 2((\bar{\mathbf{r}}_m^{(3)} - \mathbf{c}_m) \cdot \mathbf{n}_m)\mathbf{n}_m. \quad (17)$$

This operation corresponds to the geometric reflection of $\bar{\mathbf{r}}_m^{(3)}$ across the estimated planar reflector, consistent with third-order reflection propagation.

Final Localization. The final set of object location candidates at time t is constructed by combining the first-order reflection cluster centers with the mirrored third-order estimates:

$$\hat{X}_t = \{\bar{\mathbf{r}}_k^{(1)}\}_k \cup \{\hat{\mathbf{x}}_m^{(3)}\}_m. \quad (18)$$

Spatially adjacent candidates are merged to obtain a refined object location estimate. This physics-guided inference procedure leverages the predicted reflection types and the estimated reflective surface geometry to enable reliable NLOS pedestrian localization without requiring prior scene structure information.

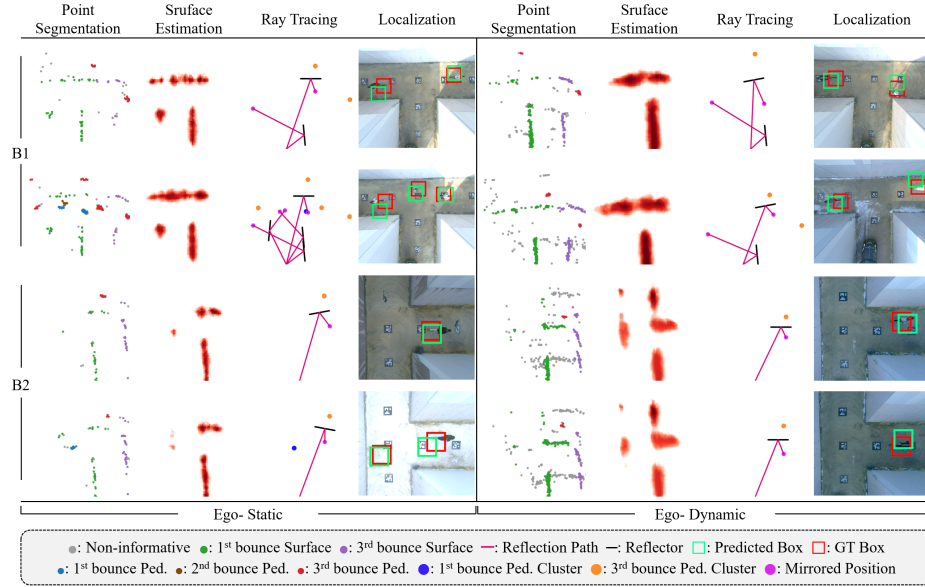


Fig. 4: Qualitative results of the proposed reflection-aware NLOS localization framework under ego-static and ego-dynamic settings. Columns show intermediate stages of the pipeline: point segmentation, reflective surface estimation, ray tracing, and final localization. The final column visualizes ground-truth and predicted pedestrian locations.

5 Dataset

The dataset was collected in a constructed urban intersection testbed designed to capture NLOS pedestrian scenarios. The ego vehicle is equipped with a front-view RGB camera, a 77 GHz mmWave radar, a LiDAR, and a wheel encoder. To obtain accurate ground-truth for occluded pedestrians, an overhead camera was installed above the center of the intersection to capture a top-down view of the entire scene. All sensor data were synchronized using ROS. Radar data were acquired using a TI AWR2944EVM sensor [30]. Raw radar signals were processed into radar PCD following the methodology described in [19]. The radar PCD generation itself is not part of the proposed contribution.

The dataset consists of scenarios where one to three pedestrians move in an intersection environment. Pedestrians appear from occluded side roads and move across the intersection while the ego vehicle either remains stationary or moves through the intersection. Two spatial configurations are defined based on the intersection geometry. B1 corresponds to the case where a frontal wall blocks the opposite side of the intersection, creating strong NLOS conditions. B2 represents the configuration where the opposite road is open, allowing the ego vehicle to drive straight or turn right. Both ego-static and ego-dynamic driving conditions are included in the dataset. In the ego-dynamic setting, the ego ve-

Table 2: Point-wise reflection path classification performance. Overall results are reported using Accuracy and Macro-F1, while precision, recall, and F1-score are additionally reported for the NLOS pedestrian reflection class corresponding to third-bounce reflections, which are critical for NLOS position reconstruction.

Method	Scenario	Accuracy	Macro-F1	3rd-bounce pedestrian		
				Precision	Recall	F1-score
Kraus et al. [14]	Ego-static	0.686	0.661	0.757	0.639	0.693
	Ego-dynamic	0.677	0.635	0.553	0.863	0.674
Wu et al. [34]	Ego-static	0.930	0.890	0.884	0.931	0.907
	Ego-dynamic	0.610	0.481	0.250	0.501	0.334
Proposed	Ego-static	0.946	0.915	0.897	0.935	0.916
	Ego-dynamic	0.899	0.814	0.744	0.718	0.731

hicle drives through the testbed at speeds up to 22.00 km/h, and the evaluation covers NLOS pedestrians with a radar-based NLOS observation range of up to 17.92 m. In total, the dataset contains 44 training scenarios and 13 validation scenarios under ego-static conditions, and 57 training scenarios with 10 evaluation scenarios under ego-dynamic conditions. Detailed dataset information is provided in the Supplementary material.

6 Experimental Results

Experiments are conducted using the dataset described in Section 5. Evaluation is performed under both ego-static and ego-dynamic conditions across the two spatial configurations B1 and B2. The proposed method is evaluated from three perspectives: reflection-type point segmentation, reflective surface estimation, and final NLOS pedestrian localization. Qualitative results are illustrated in Fig. 4. Detailed experimental results are provided in the Supplementary material.

For a fair comparison, all radar-based baselines and our method are evaluated using the same ego-motion-compensated radar PCD, which is generated by FMCW processing, CFAR detection, clustering, coordinate transformation, and wheel-encoder-based motion compensation. Thus, motion compensation is applied consistently to all compared methods rather than only to the proposed method. All methods are evaluated under the same ego-static or ego-dynamic protocol, association rule, and average localization-error metric.

6.1 Reflection-Type Point Segmentation Performance

We evaluate the proposed reflection-type point segmentation on the dataset defined in Section 5 to assess its ability to identify informative radar points for NLOS localization. The task focuses on distinguishing pedestrian-originated reflections from environmental reflections and noise.

Performance is evaluated using class-wise precision, recall, and F1-score, and compared with baseline methods [14, 34]. Experiments are conducted under both ego-static and ego-dynamic conditions.

The proposed method achieves 94.6% accuracy and a macro-F1 of 0.915 in ego-static settings, and 89.9% accuracy with a macro-F1 of 0.814 in ego-dynamic settings, as summarized in Table 2. Since NLOS pedestrians are observed through third-order reflections, accurate classification of third-order points is particularly important. The proposed approach maintains strong performance for third-bounce classification in both ego-static and ego-dynamic conditions.

Compared to baseline methods [14, 34], the proposed method consistently improves overall segmentation performance as well as third-order classification accuracy, demonstrating its effectiveness in isolating radar observations relevant to NLOS pedestrians.

6.2 Reflective Surface Estimation Performance

We evaluate the accuracy of the inferred reflective surfaces, which are used for NLOS object localization. Quantitative evaluation is performed using pseudo ground-truth wall segments extracted from LiDAR.

Given that the dataset environment consists of linear walls in a T-junction, reflective surfaces are approximated as line segments using DBSCAN clustering and RANSAC-based line fitting on LiDAR point clouds. The inferred reflector geometry is then compared against these LiDAR-derived pseudo labels.

To assess geometric consistency, we measure both reflection point error (RPE) and reflection angle error (RAE). RPE is defined as the Euclidean distance between the predicted reflection point $\mathbf{p}_{\text{pred}}$ and the LiDAR-derived ground-truth reflection point $\mathbf{p}_{\text{gt}}$:

$$\text{RPE} = \|\mathbf{p}_{\text{pred}} - \mathbf{p}_{\text{gt}}\|_2. \quad (19)$$

Here, $\mathbf{p}_{\text{gt}}$ is defined as the intersection between the ray connecting the radar origin and the third-order cluster center and the LiDAR-extracted reflector. $\mathbf{p}_{\text{pred}}$ is defined analogously using the predicted reflector. RAE measures the angular difference between the predicted reflector direction $\mathbf{d}_{\text{pred}}$ and the LiDAR-derived reflector direction $\mathbf{d}_{\text{gt}}$:

$$\text{RAE} = \arccos \left(\left| \frac{\mathbf{d}_{\text{pred}} \cdot \mathbf{d}_{\text{gt}}}{\|\mathbf{d}_{\text{pred}}\| \|\mathbf{d}_{\text{gt}}\|} \right| \right) \times \frac{180}{\pi}. \quad (20)$$

Quantitative results are summarized in Table 4. Under ego-static conditions, the proposed method achieves an RPE of 0.280 m in intersection layout B1 and 0.415 m in layout B2, with corresponding RAEs of 7.175° and 6.698°. Under ego-dynamic conditions, the RPE remains 0.285 m in layout B1 and 0.357 m in layout B2, while the RAE is 7.207° and 5.997°, respectively. These results demonstrate consistent sub-meter reflection point accuracy and reflection angle errors below 8° across different spatial configurations under both ego-static and ego-dynamic conditions.

Table 3: Quantitative comparison of pedestrian localization performance.
Results are reported in terms of average localization error (AE, m).

Ego Vehicle Status	Static		Dynamic	
	LOS (m)	NLOS (m)	LOS (m)	NLOS (m)
Scheiner et al. [27]	1.02	2.36	1.36	4.74
Park et al. [24]	0.89	1.69	1.41	3.23
Proposed	0.44	1.01	0.54	1.23

Table 4: Quantitative evaluation of reflective surface estimation.

Ego Status	Scenario	RPE (m)	RAE (deg)
Static	B1	0.280	7.175
	B2	0.415	6.698
Dynamic	B1	0.285	7.207
	B2	0.357	5.997

Table 5: System performance.

Metric	Value
Reflection-aware learning (ms)	55.22
Physics-guided ray tracing (ms)	16.80
FPS	13.85
Early detection (s)	0.77

6.3 Object Localization Performance

Localization accuracy is evaluated using the average localization error, defined as the mean localization distance described in Eq. 3. Evaluation is conducted under both ego-static and ego-dynamic conditions, with results reported separately for LOS and NLOS pedestrian localization. As summarized in Table 3, the proposed method achieves an average localization error of 0.44 m for LOS pedestrians and 1.01 m for non-line-of-sight pedestrians under ego-static conditions. Under ego-dynamic conditions, the error is 0.54 m for LOS pedestrians and 1.23 m for NLOS pedestrians.

Compared with geometry-based baselines, the proposed approach substantially improves NLOS pedestrian localization accuracy. Under ego-static conditions, the NLOS localization error decreases from 1.69 m to 1.01 m, corresponding to an improvement of approximately 40%. Under ego-dynamic conditions, the error decreases from 3.23 m to 1.23 m, yielding a reduction of more than 60%. The increase in NLOS localization error from ego-static to ego-dynamic conditions remains limited, indicating robustness against motion-induced disturbances.

Finally, Table 5 reports the runtime performance of the proposed framework. The reflection-aware learning stage takes 55.22 ms per frame, and the physics-guided ray-tracing module takes 16.80 ms, resulting in an overall throughput of 13.85 FPS on an RTX 3090 GPU. The model is lightweight, with 1.17M parameters and 55.04 GMACs per frame, corresponding to 110.08 GFLOPs when one multiply-accumulate operation is counted as two floating-point operations. The semantic auxiliary head is used only during training and therefore introduces no additional inference cost.

7 Conclusion & Discussion

This paper presented a reflection-aware learning framework for NLOS pedestrian localization in intersection scenarios. The proposed method jointly learns reflection-type point segmentation and reflective surface estimation, and integrates them with a physics-guided ray tracing module for structured multipath interpretation. This formulation enables data-driven decomposition of multipath signals while preserving physical consistency.

Experimental results demonstrate improved NLOS localization accuracy under both ego-static and ego-dynamic conditions. The proposed framework detects NLOS pedestrians on average 0.77 s earlier than first-order radar observations by leveraging third-order multipath reflections. The system operates near real time, achieving 72 ms end-to-end latency corresponding to 13.85 FPS.

Further analysis shows that accurate reflection-point detection alone is insufficient for reliable NLOS pedestrian inference. Localization error increases significantly when reflective surface estimation becomes inaccurate, indicating a strong interdependence between reflection-type segmentation and reflective surface reasoning.

Overall, the results highlight that combining reflection-aware representation learning with physics-guided reasoning enables reliable and early NLOS perception beyond direct LOS.

Acknowledgements

This work was supported by Seoul National University, including the provision of experimental equipment and the construction of the testbed environment for data collection. This work was also supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Ministry of Trade, Industry and Energy (MOTIE) (P0020536, HRD Program for Industrial Innovation), a National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (MSIT) (No. RS-2026-25518893), and the Advanced GPU Utilization Support Program funded by the Ministry of Science and ICT (MSIT), Republic of Korea. This research was also supported by the Institute of Engineering Research at Seoul National University.

References

1. An, I., Lee, D., Choi, J.w., Manocha, D., Yoon, S.e.: Diffraction-aware sound localization for a non-line-of-sight source. In: 2019 International Conference on Robotics and Automation (ICRA). pp. 4061–4067. IEEE (2019)
2. An, I., Son, M., Manocha, D., Yoon, S.E.: Reflection-aware sound source localization. In: 2018 IEEE international conference on robotics and automation (ICRA). pp. 66–73. IEEE (2018)
3. Boger-Lombard, J., Slobodkin, Y., Katz, O.: Towards passive non-line-of-sight acoustic localization around corners using uncontrolled random noise sources. *Scientific Reports* **13**(1), 4952 (2023)

4. Chen, J., Yang, X., Qiu, C., Zhu, Z., Wu, P., Xu, Z., Guo, S., Cui, G.: Joint localization of los and nlos targets with clutter mitigation via multipath exploitation radar. *IEEE Transactions on Radar Systems* **3**, 549–561 (2025). <https://doi.org/10.1109/TRS.2025.3550023>
5. Dodds, L., Shanbhag, H., Guan, J., Gupta, S., Hassanieh, H.: Around the corner mmwave imaging in practical environments. In: *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. pp. 953–967 (2024)
6. Dong, X., Zhuang, B., Mao, Y., Liu, L.: Radar camera fusion via representation learning in autonomous driving. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 1672–1681 (2021)
7. Hao, M., Ning, F., Wang, K., Duan, S., Wang, Z., Meng, D., Xie, P.: Acoustic non-line-of-sight vehicle approaching and leaving detection. *IEEE Transactions on Intelligent Transportation Systems* **25**(8), 9979–9991 (2024). <https://doi.org/10.1109/TITS.2024.3353749>
8. Huang, T., Liu, J., Zhou, X., Nguyen, D.C., Azghadi, M.R., Xia, Y., Han, Q.L., Sun, S.: V2x cooperative perception for autonomous driving: Recent advances and challenges. *arXiv preprint arXiv:2310.03525* (2023)
9. Jeon, M., Cho, J.K., Kim, H.Y., Park, B., Seo, S.W., Kim, S.W.: Non-line-of-sight vehicle localization based on sound. *IEEE Transactions on Intelligent Transportation Systems* **26**(2), 2321–2338 (2025). <https://doi.org/10.1109/TITS.2024.3510582>
10. Jeon, M., Park, B., Kim, H.Y., Kang, Y., Choi, B., Cho, H., Kim, B., Lee, S., Seo, S.W., Kim, S.W.: Non-line-of-sight multi-target localization in t-junctions using ray tracing of mmwave radar. In: *2025 IEEE Intelligent Vehicles Symposium (IV)*. pp. 1779–1786 (2025). <https://doi.org/10.1109/IV64158.2025.11097630>
11. Kim, H.Y., Park, B., Choi, B., Cho, H., Kim, B., Lee, S., Jeon, M., Seo, S.W., Kim, S.W.: Radar-based nlos pedestrian localization for darting-out scenarios near parked vehicles with camera-assisted point cloud interpretation. In: *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 21352–21359 (2025). <https://doi.org/10.1109/IROS60139.2025.11246930>
12. Kim, S.W., Liu, W., Ang, M.H., Frazzoli, E., Rus, D.: The impact of cooperative perception on decision making and planning of autonomous vehicles. *IEEE Intelligent Transportation Systems Magazine* **7**(3), 39–50 (2015)
13. Kim, S.W., Qin, B., Chong, Z.J., Shen, X., Liu, W., Ang, M.H., Frazzoli, E., Rus, D.: Multivehicle cooperative driving using cooperative perception: Design and experimental validation. *IEEE Transactions on Intelligent Transportation Systems* **16**(2), 663–680 (2014)
14. Kraus, F., Scheiner, N., Ritter, W., Dietmayer, K.: Using machine learning to detect ghost images in automotive radar. In: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. pp. 1–7. IEEE (2020)
15. Kraus, F., Scheiner, N., Ritter, W., Dietmayer, K.: The radar ghost dataset – an evaluation of ghost objects in automotive radar data. In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 8570–8577 (2021). <https://doi.org/10.1109/IROS51168.2021.9636338>
16. Lai, H., Lan, Z., Zhao, M.: Non-line-of-sight 3d reconstruction with radar. *Advances in Neural Information Processing Systems* **38**, 72844–72869 (2026)
17. Lai, H., Zheng, Z., Zhao, M.: Rf-based 3d slam rivaling vision approaches. In: *Proceedings of the 31st Annual International Conference on Mobile Computing and Networking*. pp. 170–185 (2025)

18. Lee, S., Jung, Y., Park, Y.H., Kim, S.W.: Design of v2x-based vehicular contents centric networks for autonomous driving. *IEEE Transactions on Intelligent Transportation Systems* **23**(8), 13526–13537 (2021)
19. Lee, S., Yoon, J., Kim, H., Kim, J., Kim, B.: Ego vehicle speed estimation with automotive corner radar under velocity ambiguity using optimized mobilenetv3. *IEEE Access* (2025)
20. Lim, T.Y., Ansari, A., Major, B., Fontijne, D., Hamilton, M., Gowaikar, R., Subramanian, S.: Radar and camera early fusion for vehicle detection in advanced driver assistance systems. In: Machine learning for autonomous driving workshop at the 33rd conference on neural information processing systems. vol. 2 (2019)
21. Lindell, D.B., Wetzstein, G., Koltun, V.: Acoustic non-line-of-sight imaging. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6780–6789 (2019)
22. National Highway Traffic Safety Administration: Pedestrians: 2023 data (traffic safety facts). <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813727> (2025), dOT HS 813 727, Accessed: 2026-02-20
23. Palffy, A., Kooij, J.F., Gavrilu, D.M.: Detecting darting out pedestrians with occlusion aware sensor fusion of radar and stereo camera. *IEEE Transactions on Intelligent Vehicles* **8**(2), 1459–1472 (2022)
24. Park, B., Kim, H.Y., Choi, B., Cho, H., Kim, B., Lee, S., Jeon, M., Kim, S.W.: mmwave radar-based non-line-of-sight pedestrian localization at t-junctions utilizing road layout extraction via camera. In: 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 19661–19668 (2025). <https://doi.org/10.1109/IROS60139.2025.11246461>
25. Philion, J., Fidler, S.: Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3d. In: European conference on computer vision. pp. 194–210. Springer (2020)
26. Safe Transportation Research and Education Center (SafeTREC): Traffic safety facts: Pedestrian safety. <https://safetrec.berkeley.edu/2025-safetrec-traffic-safety-facts-pedestrian-safety> (2025), accessed: 2026-03-02
27. Scheiner, N., Kraus, F., Wei, F., Phan, B., Mannan, F., Appenrodt, N., Ritter, W., Dickmann, J., Dietmayer, K., Sick, B., et al.: Seeing around street corners: Non-line-of-sight detection and tracking in-the-wild using doppler radar. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2068–2077 (2020)
28. Singh, V., Knisely, K.E., Yönek, S.H., Grosh, K., Dowling, D.R.: Non-line-of-sight sound source localization using matched-field processing. *The Journal of the Acoustical Society of America* **131**(1), 292–302 (2012)
29. Sume, A., Gustafsson, M., Herberthson, M., Janis, A., Nilsson, S., Rahm, J., Orbom, A.: Radar detection of moving targets behind corners. *IEEE Transactions on Geoscience and Remote Sensing* **49**(6), 2259–2267 (2011). <https://doi.org/10.1109/TGRS.2010.2096471>
30. Texas Instruments: AWR2944/AWR2944P Evaluation Module User’s Guide (2024), <https://www.ti.com/lit/spruj22>, sPRUJ22, revised September 2024, Accessed: 2026-02-22
31. Thai, K.P.H., Rabaste, O., Bosse, J., Chonavel, T.: Grlt particle filter for tracking nlos target in around-the-corner radar. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 3216–3220 (2018). <https://doi.org/10.1109/ICASSP.2018.8462391>

32. Thai, K.P.H., Rabaste, O., Bosse, J., Poullin, D., Hinostroza, I., Letertre, T., Chonavel, T.: Around-the-corner radar: Detection and localization of a target in non-line of sight. In: 2017 IEEE Radar Conference (RadarConf). pp. 0842–0847 (2017). <https://doi.org/10.1109/RADAR.2017.7944320>
33. Wu, P., Guo, S., Chen, J., Chen, H., Cui, G., Kong, L., Yang, X.: A two-stage nlos target positioning method based on 0-1 non-coherent binary accumulation. *IEEE Transactions on Vehicular Technology* **74**(6), 8866–8878 (2025). <https://doi.org/10.1109/TVT.2025.3542117>
34. Wu, X., Jiang, L., Wang, P.S., Liu, Z., Liu, X., Qiao, Y., Ouyang, W., He, T., Zhao, H.: Point transformer v3: Simpler faster stronger. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4840–4851 (2024)
35. Yao, S., Guan, R., Huang, X., Li, Z., Sha, X., Yue, Y., Lim, E.G., Seo, H., Man, K.L., Zhu, X., et al.: Radar-camera fusion for object detection and semantic segmentation in autonomous driving: A comprehensive review. *IEEE Transactions on Intelligent Vehicles* **9**(1), 2094–2128 (2023)
36. Zhou, K., Dodds, L., Afzal, S.S., Adib, F.: Rise: Single static radar-based indoor scene understanding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 32194–32205 (2026)
37. Zhu, Z., Guo, S., Chen, J., Xue, S., Xu, Z., Wu, P., Cui, G., Kong, L.: Non-line-of-sight targets localization algorithm via joint estimation of dod and doa. *IEEE Transactions on Instrumentation and Measurement* **72**, 1–11 (2023). <https://doi.org/10.1109/TIM.2023.3323003>