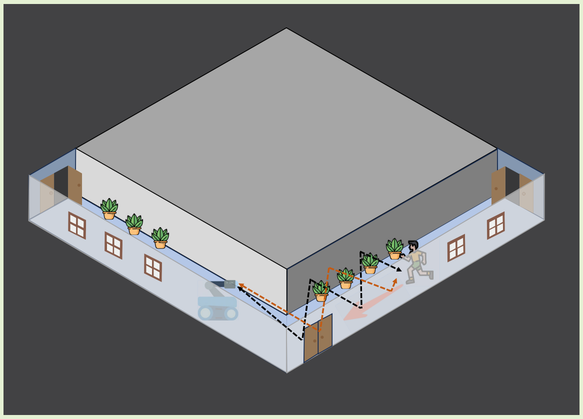
Tracking the Occluded Indoor Target With Scattered Millimeter Wave Signal

Yinda Xu, *Graduate Student Member, IEEE*, Xinjue Wang, *Student Member, IEEE*,
Juhani J. Kupiainen, Joonas Säe, Jani Boutellier, *Senior Member, IEEE*,
Jari Nurmi, *Senior Member, IEEE*, and Bo Tan, *Member, IEEE*

***Abstract*—The popularity of mobile robots in factories, warehouses, and hospitals has raised safety concerns about human–machine collisions, particularly in nonline-of-sight (NLoS) scenarios such as corners. Developing a robot capable of locating and tracking humans behind the corners will greatly mitigate risk. However, most of them cannot work in complex environments or require a costly infrastructure. This article introduces a solution that uses the reflected and diffracted millimeter wave (mmWave) radio signals to detect and locate targets behind the corner. Central to this solution is a localization convolutional neural network (L-CNN), which takes the angle-delay heatmap of the mmWave sensor as input and infers the potential target position. Furthermore, a Kalman filter is applied after L-CNN to improve the accuracy and robustness of estimated locations. A red-green-blue-depth (RGB-D) camera is attached to the mmWave sensor as the annotation system to provide accurate position labels. The results of the experimental evaluation demonstrate that our data-driven approach can achieve remarkable positioning accuracy at the 10-cm level without extensive infrastructure. In particular, the approach effectively mitigates the adverse effects of diffraction and multibounce phenomena, making the system more resilient.**



***Index Terms*— Angle-delay estimation, convolutional neural network (CNN), cross-modal training, frequency-modulated continuous-wave (FMCW) radar, indoor positioning, nonline-of-sight (NLoS) tracking, robotics.**

Received 28 April 2024; revised 6 August 2024; accepted 9 August 2024. Date of publication 23 August 2024; date of current version 14 November 2024. This work was supported in part by the Academy of Finland within Secure and Privacy Preserving Healthcare in the Residential Environment with Multimodal Distributed Data and Decentralized AI (SPHERE-DNA) Project under Grant 345681; in part by the Algorithmic Design of 5G Positioning, Sensing and Security Functions (5G-PSS) Project through Business Finland under Grant 6868/31/2021; and in part by the DIOR Project through the European Union's MSCA RISE Program under Agreement 10100828. The associate editor coordinating the review of this article and approving it for publication was Dr. Avik Santra. *(Corresponding author: Bo Tan.)*

Yinda Xu is with the Department of Electrical Engineering and Automation, Aalto University, 00076 Espoo, Finland (e-mail: yinda.xu@aalto.fi).

Xinjue Wang is with the Department of Information and Communications Engineering, Aalto University, 00076 Espoo, Finland (e-mail: xinjue.wang@aalto.fi).

Juhani J. Kupiainen was with the Faculty of Information Technology and Communication Sciences, Tampere University, 33720 Tampere, Finland. He is now with Insta Advance Oy, 33210 Tampere, Finland (e-mail: juhani@jkupiainen.fi).

Joonas Säe, Jari Nurmi, and Bo Tan are with the Tampere Wireless Research Centre, Faculty of Information Technology and Communication Sciences, Tampere University, 33720 Tampere, Finland (e-mail: joonas.sae@tuni.fi; jari.nurmi@tuni.fi; bo.tan@tuni.fi).

Jani Boutellier is with the School of Technology and Innovations, University of Vaasa, 65200 Vaasa, Finland (e-mail: jani.boutellier@uwasa.fi).

Digital Object Identifier 10.1109/JSEN.2024.3447271

I. INTRODUCTION

THE recent development in artificial intelligence (AI) and robotics applications has prominently positioned these technologies in logistics, manufacturing, and healthcare [1], [2], [3]. In these domains, AI-assisted robots execute essential tasks, from managing bulky items in warehouses to delivering medicines in mixed human environments. Central to these tasks is the capability of robots to navigate and interact safely within its environment. This navigation primarily relies on advanced sensor and signal processing systems that help robots detect obstacles and prevent collisions in various indoor scenarios. Leveraging cameras and light detection and ranging (LiDAR) [4] systems, robots can have accurate and high-resolution perceptions of their surrounding spaces and line-of-sight (LoS) obstacles. However, when it comes to nonline-of-sight (NLoS) situations, such as detecting a person behind a corner, traditional sensing methods are less effective. Sensors relying on visible light or lasers are constrained in NLoS conditions because light or laser signals cannot penetrate or sufficiently reflect around typical obstacles. This limitation increases the risk of collisions, particularly in scenarios with limited response times. Given these challenges, it is important to develop innovative solutions for robots to detect and track personnel behind corners before entering the range of visual sensors.

Optical and radio sensors are mainstream approaches to enhancing the response duration of robots for collision prevention. Optical sensors can detect NLoS targets by capturing shadows [5], [6], and disentangling scattered light information to reconstruct spatial images [7], [8], [9]. However, in intricate indoor settings, these optical techniques encounter sensitivity challenges. For example, the performance of optical sensors can degrade significantly due to factors like varying lighting conditions, occlusions, and reflective surfaces. In contrast, sensors using radio signals exhibit advantages like multipath propagation [10], [11], [12], [13], [14], [15] and penetration capabilities [16], [17], making them more suited for NLoS target detection and tracking. Nevertheless, the requirements for their application are higher: the need for prior corner geometry knowledge, complexities in managing multibounce and diffraction propagation, and the trade-off between penetration capability and position estimation resolution. For instance, radars operating at high carrier frequency and wider bandwidth usually offer finer range resolution, thus, more position estimation accuracy, however, they have less capability to penetrate obstacles due to the short wavelength. A detailed literature review of state-of-the-art methods is given in Section II.

In this article, we propose an innovative system to accurately infer and track target locations using millimeter wave (mmWave) beat frequency signal. Our approach integrates a classic direction-finding method with advanced deep learning techniques, specifically a convolutional neural network (CNN) [18], to enhance detection capabilities. Additionally, a Kalman filter is added to further rectify the CNN inference results. This combination is inspired by studies [19], [20]. Our system consists of two main components: the mmWave sensor and the annotator. The mmWave sensor captures and processes raw radio frequency (RF) data, while the annotator utilizes red-green-blue-depth (RGB-D) cameras to provide precise location labels. The major contributions in our design are listed as follows.

1) *Effectiveness in Complex Environments:* Our system demonstrates robust performance in complex indoor environments characterized by elevators, doors, and walls with varied materials. It not only outperforms prior models that commonly fail under such conditions but also ensures consistent operational reliability.
2) *Privacy Enhancement:* We limit the use of the RGB-D camera strictly to the training phase. After training, the absence of the camera in the operational phase guarantees the privacy of system.
3) *Easy Deployment:* Our method is a data-driven approach. The training stage of our system does not require the measurement of floor plan and prior knowledge of electromagnetic information of the environment that are required in many other RF-based indoor localization. The lightweight inference stage of our method indicates a low deployment threshold for scenarios where low-power processors are used.

The article is organized as follows. Section II presents related works and explores existing research on the topic. Section III focuses on radar signal modeling and target positioning, providing detailed explanations. Section IV explains experimental devices and the methodology adopted. Section V discusses the positioning results obtained from the experiments. Finally, Section VI concludes the article and discusses future work.

II. RELATED WORKS

The studies on perceiving targets behind corners can be broadly categorized into two main genres: optical methods and RF-based methods.

A. Optical Methods

In recent years, there has been a surge in the development of novel optical methods for sensing objects around corners. The ShadowCam algorithm, introduced in [5], distinguishes between "dynamic" and "static" shadows created by obstacles through registered image sequences. This classification helps identify dynamic obstacles or subtle changes in illumination. However, the dependence of the algorithm on additional infrastructure, such as AprilTags, is a notable limitation. An improvement is proposed in [6], where the system transitions to utilizing visual odometry methods for image sequence registration. Despite this enhancement, the approach still has limitations, including the need for close camera proximity to the corner and the requirement for the light source to be positioned well behind individuals to create discernible shadows. It is worth noting that both of these shadow-based techniques primarily serve the purpose of detecting the presence of individuals around the corner and do not offer precise object localization.

Certain methods focus on the detection of scattered light from these objects to locate objects behind the corner. While reflected light carries valuable scene information, it often undergoes mixing due to diffuse reflections before reaching the image sensor. To overcome this problem, the study detailed in [7] uses a combination of ultra-short pulsed signals and computational reconstruction algorithms to disentangle image information for NLoS localization. However, this approach comes with a high cost, primarily due to the need of the ultra-short pulsed laser and a streak camera as illuminator and receiver, respectively. To reduce infrastructure costs, researchers proposed several methods based on conventional laser systems [8], [9] to obtain NLoS location information. The approach described in [8] takes advantage of the spatial variation of indirect reflections under varying illumination conditions to image occluded objects. The work in [9] employs spectral estimation theory to establish a noise model for NLoS correlography and subsequently develops a speckle correlation-based technique to recover occluded objects. Both methods are computationally expensive due to the requirement of a large training set.

Digital refocusing algorithms [21] have been leveraged to allow cameras form images through foreground occlusions. Although this technology has gained impressive performance [22] through deep learning, it is only suitable for perceiving through sparse obstacles, such as fences or vegetation.

Although optical methods have shown promise for indoor localization, accurately locating people around corners in complex indoor environments remains a challenging task. Many optical-based methods require controlled illumination conditions, which can be difficult to meet in indoor environments. Additionally, the use of cameras in optical solutions raises privacy concerns, since they capture image data, which could be vulnerable to compromise if accessed by unauthorized entities through hacking or other means.

B. RF-Based Methods

Compared to optical methods, RF-based methods offer advantages such as improved robustness in indoor environments and fewer concerns about privacy leakage. Studies such as [10], [11], [12], [13], [14], [15], use indirect path signals, commonly known as bounced signals, to locate individuals behind corners. The authors [10], [11], [12], [13], [14], [15] universally adopt ray tracing to address the limitations posed by obstructed or heavily attenuated direct path signals. By tracing indirect RF paths and using the known geometric information of the environment, these approaches successfully determine the positions of individuals behind corners. However, these multipath exploitation methods require accurate geometry information for ray tracing models, and their performance can be affected by variations in wall materials, reflectivity, and floor layout, especially in complex indoor environments. Furthermore, these methods usually do not consider the existing factors. For example, the work in [12] overlooks multiple-bounce signals, while the research in [10] and [11] neglects to account for diffraction effects.

On the other hand, some work [16], [17] demonstrated the possibility of using direct path signals from the targets. These approaches typically start by disentangling the direct path from the indirect paths and then using the direct path information for target localization. For example, Chuo et al. [16] introduces a neural network based timing interpolation technique, which enables an efficient distinction between indirect and direct paths. In [17], they employ the discrete inverse Fourier transform method to effectively separate the direct path signals from the indirect path signals. However, these methods are not well-suited for corner situations due to issues like blockage or severe signal attenuation when signals traverse multiple walls, as evidenced by prior research [10], [12], [15], [23], [24], [25]. Studies such as [26], [27], [28] have demonstrated the feasibility of applying neural networks in mmWave radar data for target tracking in indoor LoS scenarios. It is worth noting that the mmWave signal is capable of penetrating the obstacle for the detection and tracking of NloS objects in [27]. However, the only limitation of a 3 mm occlusion is that it may not realistically represent typical indoor NLoS scenarios.

Our data-driven mmWave corner sensor does not require any prior knowledge about propagation effects. The model is capable of learning RF signal propagation around corners, eliminating the need for detailed layout information, unlike the ray tracing approach. Furthermore, its superiority lies in its adaptability to various factors that affect signal propagation, such as elevators, walls of different materials, multiple bounce scenarios, and diffraction.

III. SYSTEM COMPOSITION, TARGET POSITIONING, AND TRACKING

Our system comprises two primary components: an mmWave sensor and an annotation subsystem. The mmWave sensor collects raw data, while the annotation subsystem captures real-time subject location, both sent to the PC. After preprocessing the I/Q samples and location data, the PC trains a CNN model. The subsequent deep learning inference and filtering on the PC enable precise behind-corner people localization.

A. Signal Modeling of the Phased Array mmWave Sensor

In this work, a phased array frequency-modulated continuous-wave (FMCW) radar [29] is used as the mmWave sensor. The FMCW chirp signal is transmitted over M antennas and received by N antennas after propagation in the complex environment. The transmitting and receiving antennas are co-located, phase coherent, and linearly arranged with half-wavelength spacing. This configuration adheres to the principles of a uniform linear array (ULA) [30]. A single FMCW chirp signal can be written as

$$z(t) = e^{j2\pi f_c t + \pi \frac{B}{T_d} t^2} \tag{1}$$

where f_c denotes the carrier wave frequency, T_d represents the chirp signal ramp duration, and B is the FMCW sweep bandwidth. Due to limitations of the hardware, only analog beamforming (ABF) can be applied to transmitting antennas. The transmitted signal after ABF is

$$s_t(t) = \sum_{m=1}^{M} a_m(\theta) z(t). \tag{2}$$

Here, M is the number of the transmitting elements, a_m is the phase shift applied on the mth transmitting element

$$a_m(\theta) = e^{-j\frac{2\pi}{\lambda}(m-1)d\sin(\theta)}$$

where λ is the wavelength, $d = \lambda/2$ denotes the array element spacing, and θ is the steering direction (set to zero in this article).

The hardware used in this work provides full digital beamforming (DBF) capability on the receiving side. By the principle of ULA, the received mmWave sensing signal $\mathbf{s}_r(t)$ is

$$\mathbf{s}_r(t) = \sum_{k=1}^{K} \alpha_k \mathbf{a}_r(\theta_k) s_t(t - \tau_k) + \mathbf{w}(t). \tag{3}$$

Here, K indicates the number of signal paths. Parameters α_k, θ_k, τ_k represent the amplitude, angular position, and delay of the kth arrival signal, respectively. The receiving steering vector associated with the kth arrival signal, $\mathbf{a}_r(\theta_k)$, is expressed as

$$\mathbf{a}_r(\theta_k) = \left[1, e^{-j\frac{2\pi}{\lambda}d\sin(\theta_k)}, \ldots, e^{-j\frac{2\pi}{\lambda}(N_r-1)d\sin(\theta_k)}\right]^{\mathsf{T}}$$

where N_r denotes the number of the receiving array elements, and $(\cdot)^{\mathsf{T}}$ is the transpose operator. The last term $\mathbf{w}(t) \in \mathbb{C}^{N_r}$

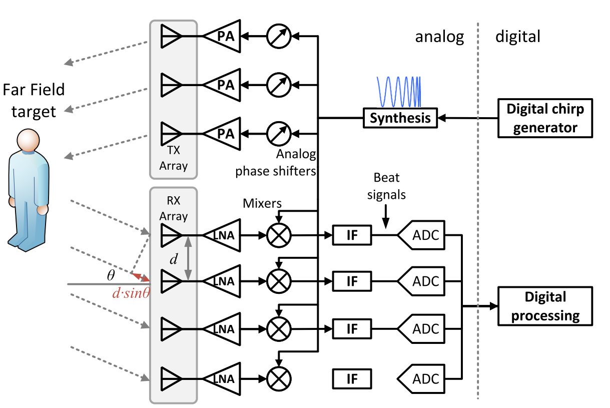


Fig. 1. FMCW radar system. Upon mixing the received signal **Rx** with the transmitted signal **Tx** in a radar system, a new signal referred to as the beat signal is generated. This beat signal represents the difference in frequency between signals **Tx** and **Rx**.

represents the receiver noise vector, which follows a Gaussian distribution with a zero mean.

For the coherent FMCW radar used in this work, the receiving signal $\mathbf{s}_r(t)$ is deramped by the reference signal, which is a tap-off from the transmitting signal $z(t)$, to obtain a beat signal (as illustrated in Fig. 1). The time domain beat signal in the n_rth receiving element is defined as

$$\mathbf{s}_r^{n_r}(t)=\sum_{k=1}^{K}\alpha_k a_r^{n_t}(\theta_k)e^{j2\pi\left(f_c\tau_k+\frac{B}{T_d}\tau_k t+\frac{B}{2T_d}\tau_k^2\right)}+w(t) \qquad (4)$$

where $a_r^{n_t}(\theta_k)=e^{-j(2\pi/\lambda)(n_r-1)d\sin(\theta_k)}$.

After the intermediate frequency (IF) filtering and analog-to-digital conversion (ADC), the sampled time-domain beat signal can be written in the discrete form

$$\mathbf{s}_r^{n_r}(i)=\sum_{k=1}^{K}\alpha_k a_r^{n_t}(\theta_k)e^{j2\pi\left(f_c\tau_k+\frac{B}{T_d}\tau_k\frac{n}{f_s}+\frac{B}{2T_d}\tau_k^2\right)}+w(i) \qquad (5)$$

for $i=0,\ldots,I$, where f_s denotes the sampling frequency and the number of samples is $I=T_d f_s$. $\mathbf{s}_r^{n_r}\in\mathbb{C}^I$. After deramping and ADC over multiple coherent receiving chains, the resulting deramped FMCW radar signal can be organized into a radar information matrix $\mathbf{S}_r=[(\mathbf{s}_r^1)^\mathsf{T},(\mathbf{s}_r^2)^\mathsf{T},\ldots,(\mathbf{s}_r^{N_r})^\mathsf{T}]^\mathsf{T}\in\mathbb{C}^{N_r\times I}$ as shown in Fig. 2.

In Section III-B1, we describe the process of generating a delay-angle heatmap from the deramped samples of a phased array mmWave radar, subsequently employing this heatmap for the precise positioning of human targets.

B. Target Positioning

In LoS scenarios, mostly outdoor, the target position can be derived from the angle and distance estimates of a single or multiple radar node. This simple approach often fails in indoor applications due to complicated multipath propagation.

While studies like [31] have demonstrated that multipath components (MPCs) can enhance the accuracy of indoor localization, challenges arise in intricate propagation environments. Specifically, when occurrences such as multiple bounces, diffractions, or penetrations are involved, localizing and tracking the target become significantly more challenging.

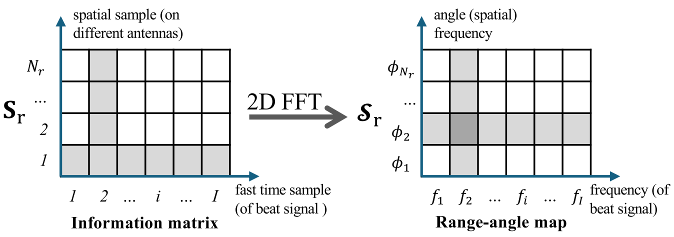


Fig. 2. Left matrix is the FMCW radar information matrix. The right matrix is the pseudospectrum generated by applying 2D-FFT algorithm to the information matrix.

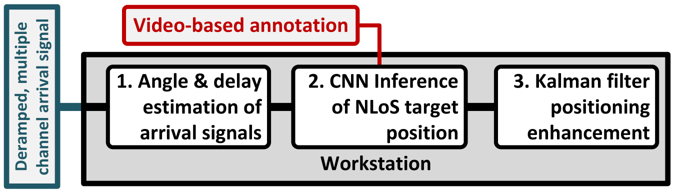


Fig. 3. mmWave phased array radar target positioning process includes three steps: first, obtaining the range-angle spectrum heatmap from the received radio signal after complex propagation; second, utilizing CNN to infer the NLoS target location using the range-angle profile; and finally, enhancing positioning accuracy through Kalman filtering.

To address these complexities, this article proposes a novel approach: leveraging the CNN combined with the Kalman filter, rooted in the preliminary delay-angle heatmap, to infer the target position. The processing flow is shown in Fig. 3. We recognize that integrating neural networks with Kalman filters has been explored in existing literature [19], [20], [32]. However, our primary focus is on the first two steps in Fig. 3. The integration is secondary to our main objective. Advanced Kalman filter design may offer further enhancements, however, is beyond the scope of our current work. Thus, we leave it as a future work and use the basic Kalman filters in this article.

1) Preliminary Angle and Range Estimation: For the information matrix $\mathbf{S}_r$ phased array FMCW radar, the samples on each column and each row are called spatial and fast-time samples, respectively. For the ULA subject to the far-field condition, the changing rate between the adjacent spatial samples is linearly proportional to the azimuth angle θ. The phase rotating rate can be represented by $e^{-j(2\pi/\lambda)d\sin(\theta)}$. Thus, peaks on the Fourier spectrum of spatial samples indicate azimuth angles of potential targets, or in other words, directions of targets. For FMCW radar, peaks on the Fourier spectrum of fast-time samples (beat signal) indicate the distances of potential targets [33]. Therefore, the most straightforward approach to estimate angle and range is searching the peak(s) on the pseudorange-angle spectrum. This spectrum is obtained via taking two-dimension fast Fourier transformation (2D-FFT) of the information matrix, $\mathbf{S}_r$, as depicted in Fig. 2

$$\mathcal{S}_r(\phi_{n_r},f_i)=\sum_{n_r=0}^{N_r-1}\sum_{f_i=0}^{I-1}\mathbf{S}_r(n_r,i)e^{-2\pi j\left(\frac{\phi_{n_r}}{N_r}n_r+\frac{f_i}{I}i\right)} \qquad (6)$$

where i and n_r are the fast-time and spatial sample indexes, and f_i and ϕ_{n_r} are the frequency and spatial-frequency sample indexes. The frequency f_i has a linear relationship with the signal propagation distance. The spatial frequency ϕ_{n_r} indicates the direction of arrival. In the simple LoS scenario,

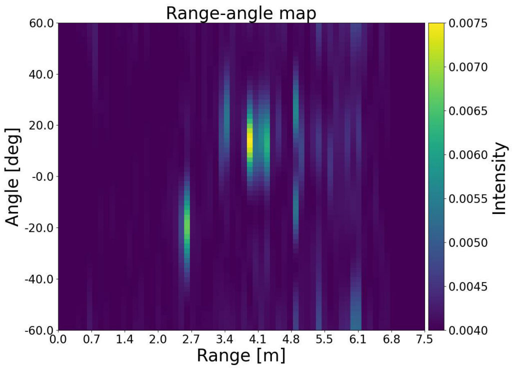


Fig. 4. 2D-FFT spectrum (NLoS) for range and angle estimation. Note that due to the complex environment and multipath propagation in the NLoS scenario, there are more than one peak in the 2D-FFT spectrum.

the peaks on the 2-D pseudospectrum $\mathcal{S}_r$, which can also be called range-angle heatmap, directly represent the direction and distance of the target. However, in NLoS scenarios, the presence of diffraction, penetration, and multiple reflections makes it difficult to derive the target position directly from the range-angle heatmap. An example of the angle-range estimation is shown in Fig. 4.

2) Deep Learning Inference of Preliminary Estimation: As discussed above, the strongest peak on the range-angle heatmap $\mathcal{X}_r$ may not directly indicate the position of a person in NLoS scenarios due to complex signal propagation. To address this issue, a localization CNN (L-CNN) is used to predict the location of the person. Compared to common CNN applications, our L-CNN solves a regression problem rather than a classification problem. The input data of L-CNN consist of radar range-angle heatmaps, whereas the annotation (or ground truth) labels are obtained from RGB-D-based data. The L-CNN (shown in Fig. 5), inspired from [34], comprises five convolutional layers. To avoid overfitting and to improve convergence, each convolutional layer is followed by activation function and batch normalization. Rectified linear units (ReLU) are used as activation layers. Additionally, to enhance the learning of abstract features and reduce the sensitivity to small input variations and noise, a maximum-pooling layer is included after the last convolutional layer. Furthermore, max-pooling downsampling reduces the number of learning parameters in the following layers. Two fully connected layers are then applied to capitalize on the information extracted from the features. Using range-angle heat maps as input to deep learning inference, the predicted position of the target can be obtained. Mean square error (MSE) serves as the loss function in the model training process, and a comprehensive explanation of this training procedure will be provided in Section IV-D. Using range-angle heatmaps as input for deep learning inference, the predicted position of the target (x, y) can be obtained.

3) Kalman Filter: During the deep learning inference phase, the tentative position of the target is identified as (x, y). This position, however, tends to be approximate since it solely relies on individual time frames without considering sequential ones. To enhance the accuracy of this position estimation, we apply the Kalman filter [35] technique. This algorithm estimates a joint probability distribution over the variables for each time frame, considering dynamic models. Consequently, the estimations achieved with the Kalman filter, which takes into account the temporal dynamics, are more accurate than those omit such considerations.

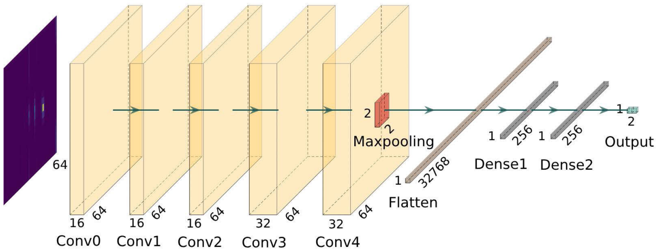


Fig. 5. L-CNN architecture comprises five convolutional layers, each followed by activation function and batch normalization. The network also includes a max-pooling layer and two fully connected layers. The input data are the range-angle heatmap derived from 2D-FFT processing, and the output of the network provides the predicted position of the target in the form of (*x*, *y*).

We consider a discrete-time linear dynamical system and model its behavior through a set of equations and matrices, which we describe sequentially below. The state vector $\boldsymbol{\omega}_n$ and observation vector $\boldsymbol{\psi}_n$ are represented as

$$\boldsymbol{\omega}_n = \mathbf{F}\boldsymbol{\omega}_{n-1} + \mathbf{q}_n \tag{7}$$

$$\boldsymbol{\psi}_n = \mathbf{G}\boldsymbol{\omega}_n + \mathbf{r}_n \tag{8}$$

where $\mathbf{F} \in \mathbb{R}^{4\times4}$ is state transition matrix, $\mathbf{G} \in \mathbb{R}^{2\times4}$ is observation matrix, $\mathbf{q}_n \in \mathbb{R}^{4\times1}$ and $\mathbf{r}_n \in \mathbb{R}^{2\times1}$ are system and observation noise vectors, respectively.

1) The **state vector** $\boldsymbol{\omega}_n \in \mathbb{R}^{4\times1}$ represents the current state of the system under estimation. This vector provides an internal snapshot of the state of the system at a given time

$$\boldsymbol{\omega}_n = [\, x_n \quad y_n \quad \dot{x}_n \quad \dot{y}_n \,]^{\mathsf{T}}.$$

2) The **observation vector** $\boldsymbol{\psi}_n \in \mathbb{R}^{2\times1}$ is derived from sensor readings or other data sources. In our case, this vector encompasses the spatial position of the target as determined by deep learning inference

$$\boldsymbol{\psi}_n = \begin{bmatrix} x_n & y_n \end{bmatrix}^{\mathsf{T}}.$$

3) The **state transition matrix F** describes the evolution of the system over time

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}$$

where Δt is the time interval, aligning with the radar frame interval $1/30$ s in our case.

4) The **observation matrix G** describes the relationship between the internal state of the system and the external observations or measurements

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \tag{10}$$

5) The **system noise vector** $\mathbf{q}_n$ and **observation noise vector** $\mathbf{r}_n$ represent the inherent noise in the system and observations, respectively. Their expectation and covariances are

$$\mathbb{E}\{\mathbf{q}_n\} = \mathbf{0}, \quad \mathrm{Cov}\{\mathbf{q}_n\} = \mathbf{Q} \tag{11}$$

$$\mathbb{E}\{\mathbf{r}_n\} = \mathbf{0}, \quad \mathrm{Cov}\{\mathbf{r}_n\} = \mathbf{R} \tag{12}$$

where $\mathbf{0}$ is a zero vector. The values of the process noise covariance matrix $\mathbf{Q} \in \mathbb{R}^{4\times4}$ are chosen based on prior knowledge of the behavior of the system (in our case, the noise is the unknown target acceleration); the values of the observation noise covariance matrix $\mathbf{R} \in \mathbb{R}^{2\times2}$ are determined by characteristics of the measurement sensors and their known or estimated noise properties.

Here, we choose the piecewise white noise model as the process noise model, and the process noise covariance matrix $\mathbf{Q}$ can be represented as

$$\mathbf{Q} = \sigma_a^2 \begin{bmatrix} \frac{\Delta t^4}{4} & 0 & \frac{\Delta t^3}{2} & 0 \\ 0 & \frac{\Delta t^4}{4} & 0 & \frac{\Delta t^3}{2} \\ \frac{\Delta t^3}{2} & 0 & \Delta t^2 & 0 \\ 0 & \frac{\Delta t^3}{2} & 0 & \Delta t^2 \end{bmatrix} \tag{13}$$

where σ_a^2 is the expected variation in target acceleration. The observability matrix of our system is

$$\mathbf{O} = \begin{bmatrix} \mathbf{G} \\ \mathbf{GF} \\ \mathbf{GF}^2 \\ \mathbf{GF}^3 \end{bmatrix}. \tag{14}$$

After substitutingthe values of matrixs $\mathbf{G}$ and $\mathbf{F}$ obtained from (9) and (10), we find the rank of the observability matrix is four, which indicates the system is fully observable. Although the estimated velocities may not be very accurate, the estimated positions have shown improvement when employing a Kalman filter, considering our system as a discrete-time linear dynamical system. The Kalman filter operates in an iterative manner, refining the status of the system, $\boldsymbol{\omega}$, based on the observation, $\boldsymbol{\psi}$. This iterative process consists of two core steps.

1) *Prediction Step:* We predict the state of the system based on the previous state and the dynamic model of the system. Formally, it is represented by

$$\hat{\boldsymbol{\omega}}_{n|n-1} = \mathbf{F}\hat{\boldsymbol{\omega}}_{n-1|n-1} \tag{15}$$

$$\mathbf{U}_{n|n-1} = \mathbf{F}\mathbf{U}_{n-1|n-1}\mathbf{F}^{\mathsf{T}} + \mathbf{Q} \tag{16}$$

where $\hat{\boldsymbol{\omega}}_{n-1|n-1} \in \mathbb{R}^{4\times1}$ denotes the conditional expectation of $\boldsymbol{\omega}_{n-1}$ given the data $\boldsymbol{\psi}_{1:n-1}$ and similar for the state covariance matrix $\mathbf{U}_{n-1|n-1} \in \mathbb{R}^{4\times4}$. They are outcomes from the last measurement update.

2) *Measurement Step:* In this step, the estimated state is revised based on new observations

$$\mathbf{D}_n = \mathbf{U}_{n|n-1}\mathbf{G}^{\mathsf{T}}\left(\mathbf{G}\mathbf{U}_{n|n-1}\mathbf{G}^{\mathsf{T}} + \mathbf{R}\right)^{-1} \tag{17}$$

$$\hat{\boldsymbol{\omega}}_{n|n} = \hat{\boldsymbol{\omega}}_{n|n-1} + \mathbf{D}_n\left(\boldsymbol{\psi}_n - \mathbf{G}\hat{\boldsymbol{\omega}}_{n|n-1}\right) \tag{18}$$

$$\mathbf{U}_{n,n} = \mathbf{U}_{n|n-1} - \mathbf{D}_n\left(\mathbf{G}\mathbf{U}_{n|n-1}\mathbf{G}^{\mathsf{T}} + \mathbf{R}\right)\mathbf{D}_n^{\mathsf{T}} \tag{19}$$

where $\mathbf{D}_n \in \mathbb{R}^{4\times2}$ is the Kalman gain.

To start the iteration, we set initial states as

$$\hat{\boldsymbol{\omega}}_{0|0} = \begin{bmatrix} x_0 & y_0 & \dot{x_0} & \dot{y_0} \end{bmatrix}^{\mathsf{T}} \tag{20}$$

$$\mathbf{U}_{0|0} = \begin{bmatrix} \sigma_x^2 & 0 & 0 & 0 \\ 0 & \sigma_y^2 & 0 & 0 \\ 0 & 0 & \sigma_{\dot{x}}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\dot{y}}^2 \end{bmatrix}. \tag{21}$$

Here, $\sigma_{(\cdot)}$ represents the standard deviation of specific variables, and initial values such as x_0 and y_0 originate from the first localization output of the CNN. By assuming the initial velocity of the target is 0, we set $\dot{x_0}$ and $\dot{y_0}$ to $\mathbf{0}$. The result of this process, $\hat{\boldsymbol{\omega}}_{n|n} = [\,\hat{x}_n \;\; \hat{y}_n \;\; \hat{\dot{x}}_n \;\; \hat{\dot{y}}_n\,]^{\mathsf{T}}$, offers the estimated location and velocity for every iteration. Utilizing the processed data through the Kalman filter provides enhanced accuracy in location determination compared to relying solely on the unfiltered observations.

IV. EXPERIMENTAL SETUP

This section gives detailed information on the hardware system design, experiment environment, procedure, data collection, preprocessing and annotation, and neural network model training and tuning. The dataset, which includes the raw radar data, ground truth, processed data, the trained L-CNN model, and other related metadata, is open via IEEE Dataport [36].

A. Experimental Hardware and Setting

1) mmWave Sensor Subsystem: Our mmWave sensor subsystem consists of a 3Tx-4Rx, 60 ∼ 64 GHz FMCW mmWave radar (TI[1], IWR6843ISK with TI[1], MMWAVEICBOOST) [37], and a raw data capture board (TI[1], DCA1000EVM). The firmware code is flashed into mmWAVEICBOOST. After starting the radar module, the beat frequency signal (shown in Fig. 1) will be recorded by the IWR6843 and transmitted to DCA1000EVM via the high-speed low-voltage differential signaling (LVDS) interface. The captured beat frequency data are then imported to the host PC over 1 Gb/s Ethernet for preprocessing. Fig. 6 shows the architecture of the mmWave sensor subsystem. The radar module provides 30 frames (2D-FFT spectrum obtained from Section III-B1) per second. A radar configuration file is used to configure the radar system. Some important radar setting parameters (shown in Table I) are based on this.

2) Annotation Subsystem: The annotation system is based on an Intel[1] RealSense L515 RGB-D camera. The example RGB and depth images are shown in Fig. 7. Object detection is performed using the You Only Look Once (YOLO) v4 model [38] that is pretrained with the COCO dataset [39],

---

[1] Registered trademark.

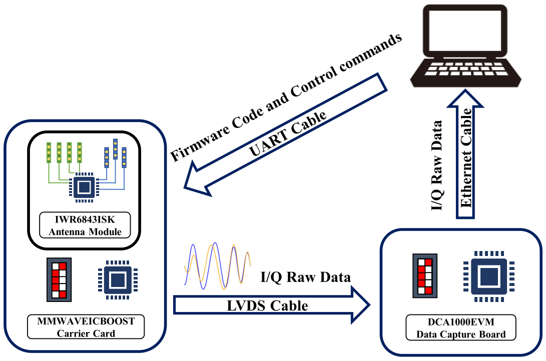


Fig. 6. Architecture of the mmWave sensor subsystem involves several key components. The PC flashes firmware code to the mmWAVEICBOOST carrier card and controls the recording process by sending commands through the UART port. The antenna module records beat signal data, which is subsequently transmitted to the data capture board via an LVDS cable. Finally, the data are transported to the PC over an Ethernet cable.

TABLE I
RADAR CONFIGURATION PARAMETERS

Parameters	Value
Number of Tx channels	3
Number of Rx channels	4
Start frequency	60 GHz
Stop frequency	64 GHz
Bandwidth	4 GHz
Frame rates	33.3 /s
Rx gain	30 dB
Maximum unambiguous range	7.5 m
Maximum radial velocity	2.41 m/s
Range detection threshold	15 dB
ADC sample duration	11.64 µs
Frequency slope	100 MHz/µs
Ramp duration	40 µs
Number of ADC samples per chirp	64
Number of chirps per frame	64
Number of frames per second	64

including various human targets. Thus, the YOLO v4 model can be directly used on the RGB image in our project for human target bounding box extraction without further training. For depth perception of the target, an initial alignment of the depth image with the RGB image is conducted. Once the target is identified within the RGB image using YOLO, the pixel data within the bounding box becomes accessible. Given that the depth and RGB images are aligned, the 3-D coordinates of the center of the bounding box [as shown in Fig. 7(a)] can be determined using the intrinsic and extrinsic camera properties combined with depth information. This location effectively represents the position of the target in the real world. This location information serves as ground truth labels for training radio sensing data (range-angle heatmap). The camera captures images at a consistent rate of 30 frames per second (fps). However, the overall output fps of the annotation system vary since the YOLO model may not always detect the target in every frame. We will explain the strategy to handle this fact in Section IV-C.

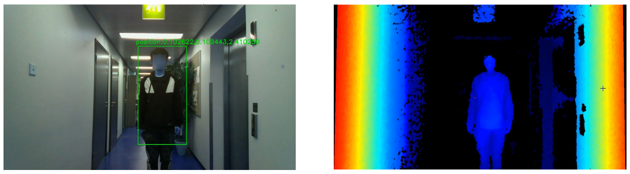

(a) (b)

Fig. 7. (a) RGB and (b) aligned depth images. The depth image in (b) has already been aligned to the RGB images in (a). YOLO operates on RGB images, marking detected targets with bounding boxes. Then, the 3-D coordinates of the targets are calculated using the pixel and depth information from the central point of the bounding box, along intrinsic and extrinsic camera information.

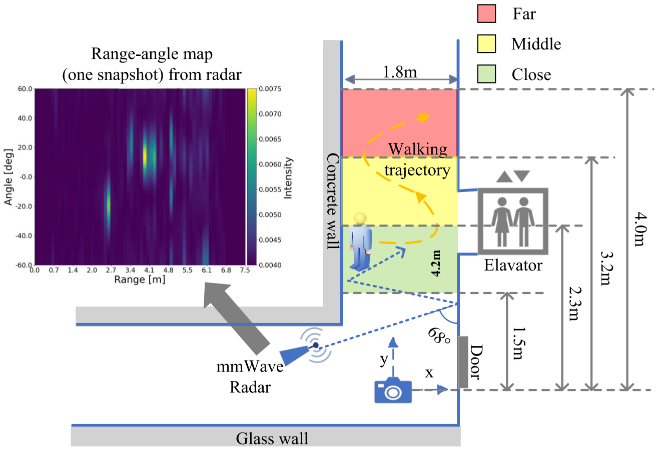


Fig. 8. NLoS experimental layout. Radar emits signals at a 68° angle toward the wall, collecting bounced radio signals. The subject moves in undefined (random) trajectories within the area of interest. Elevator and door introduce complicating factors.

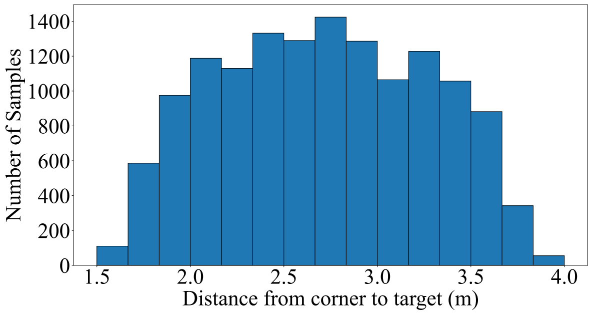


Fig. 9. Data Diversity: this figure presents a histogram depicting the distances from the corner to the target. It is important to note that the exact location of the corner (origin) is where the camera is.

B. Experiment Scene

The experiment scene is shown in Fig. 8. The mmWave sensor is placed along the wall to avoid the LoS between the sensor and the target. A person walks randomly in the area of interest, which is filled with green, yellow, and pink colors in Fig. 8. In the experiment, 13 948 frames are recorded for training. The variation of distances from the corner to the target is illustrated as a histogram in Fig. 9. Furthermore, a distinct test dataset was generated by having a different individual, not present in either the training or validation datasets, navigate the area in a zigzag pattern (the yellow curve shown in Fig. 8).

C. Data Preparation

After applying the 2D-FFT to the beat signal data, we obtain a series of range-angle heatmaps. These heatmaps are then integrated with ground-truth labels to construct the training dataset. This integration poses challenges due to two primary issues.

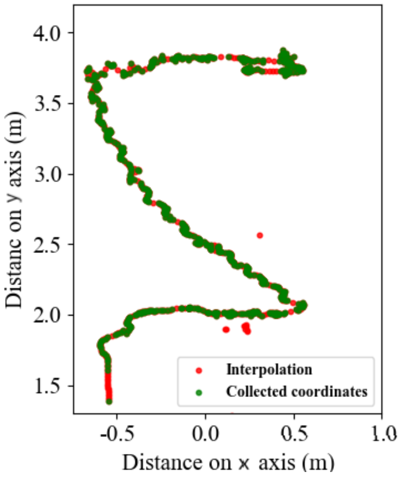


Fig. 10. Linear Interpolation: in the plot, green dots represent the coordinates collected by the ground truth system, while red dots indicate coordinates generated through linear interpolation. The origin corresponds to the fixed camera location (exact position of the corner).

1) *Temporal Mismatch:* The sensor and annotation subsystems have different start and end recording times, causing a misalignment in data duration.
2) *Frame Discrepancy:* Despite both subsystems operating at the same frame rate, the annotation subsystem occasionally misses frames, mainly when the YOLO model fails to detect the target.

To address the first issue, we resolve this by clipping the data to a common recording duration and discarding redundant frames from the sensor system outside this period. With regard to the second issue, under the assumption that velocity changes of the target are negligible, we have implemented a solution that involves the application of linear interpolation to the annotation data. This approach is depicted in Fig. 10. Subsequently, we align frames to ensure each heatmap correctly matches its ground-truth label.

D. Model Tuning and Training

After processing, we obtain the training dataset consisting of heatmaps paired with location labels and then proceed to train the L-CNN model. The objective of training is to minimize the difference (i.e., the loss) between the output of the network and the provided location label. After establishing the neural network structure (discussed in Section III-B2), the subsequent stage involves determining the hyperparameters. Some important hyperparameters are listed in Table II. To avoid slow convergence and local minima when using the small learning rate, as well as divergence with large learning rates, a constant learning rate is set to 2×10^{-4} after fine tuning. All the other hyperparameters, such as number of epochs and size of convolutional filters, are fine-tuned through exhaustive tests. We split the dataset into a training set and a validation set in a 7:3 ratio. The stochastic gradient descent (SGD) is selected as the optimizer. MSE is chosen as the loss function. Compared to mean absolute error (MAE), MSE penalizes larger prediction errors more severely.

To tune the process noise model in the Kalman filter, we have conducted a grid search to find the optimal value of σ_a^2. Our results have indicated that $\sigma_a^2 = 1.2$ minimizes the average Euclidean error of the filtered data. The relationship between σ_a^2 and the average Euclidean distance error is illustrated in Fig. 11.

TABLE II
HYPERPARAMETERS FOR TRAINING

Name of Hyperparameter	Contents
Training and validation data ratio	7 : 3
Optimizer	SGD
Loss function	MSE
Learning rate	2×10^{-4}
Batch size	256
Number of epochs	200
Size of filters	3×3
Fully connected layers	2
Number of neurons in fully connected layer	256

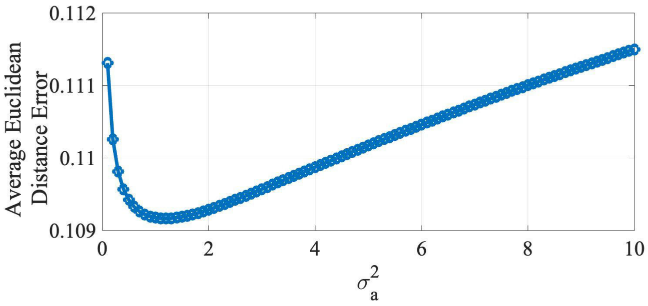


Fig. 11. Relationship between σ_a^2 and average Euclidean distance error.

V. EXPERIMENTAL RESULTS AND DISCUSSION

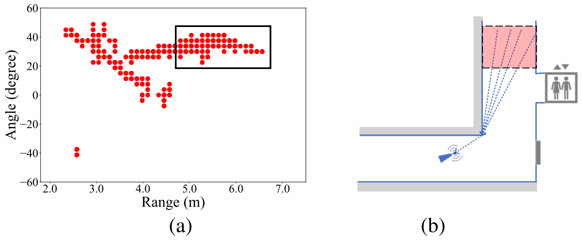
A. Angle and Range Estimation Result and Diffraction Phenomenon

Using the 2D-FFT method discussed in Section III-B1, we can obtain preliminary results for angle and range estimation. These results, based on 2D-FFT, are depicted in Fig. 4. However, it is important to note that these results should be considered preliminary, as they only present information regarding the angle and distance of the signal upon arrival without considering any reflections that may occur when the signal reflects off walls. If we extract all the peaks from the range-angle heatmaps in the test dataset, the result can be observed in Fig. 12.

Incorporating a ray tracing model or ray backpropagation with this approach is not likely to yield satisfactory localization accuracy, as the presence of peak points does not consistently indicate the correct angle and distance. In particular, when the target is above 4 m, the arrival angle of the signal remains consistently below 20°. This indicates the diffraction effects at that particular moment, and the bounced signal weakened as a result of multiple bounces. The diffracted signal is stronger than bounced signal, which leads to erroneous predictions.

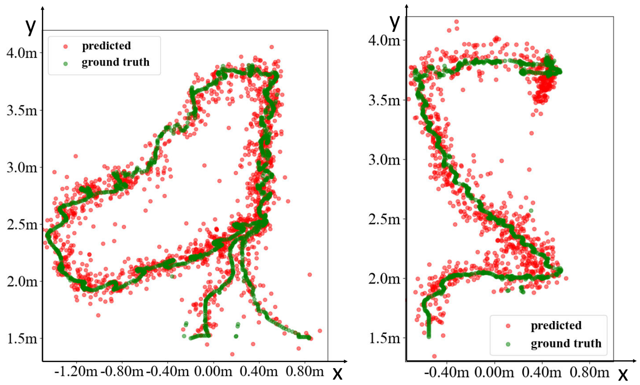
B. Results Based on Data Driven Methods

On the other hand, by using our data-driven solution, the problems discussed above can be overcome because of its powerful ability to handle complex data. Fig. 13 displays the localization results obtained using mmWave radio measurements and the trained L-CNN for both LoS and NLoS



(a) (b)

Fig. 12. (a) is obtained by extracting peaks of 2D-FFT range-angle heatmaps from test dataset. The peak is the strongest signal on the range-angle map. Upon observing the figure, it becomes evident that as the distance increases the predominant signal sources primarily fall within the 20°–40° range. It suggests there (bounded range) might be diffraction phenomenon. Note the origin here is the radar instead of the camera. (b) illustrates the diffraction phenomenon and where it might happen (red area).



(a) (b)

Fig. 13. Tracking results of (a) LoS scenario and (b) NLoS scenario (shown in Fig. 8). Green dots refer to ground truth points from RGB-D pixels, and red dots refer to the predicted results from the mmWave radio data. The result of the Los scenario is better than the one of the NLoS scenarios. It is reasonable as RF signals in LoS travel with less interference than ones in NLoS. Note here the origin is the corner (camera).

TABLE III
MAE of LoS and NLoS Scenarios

Scenario	Mean Absolute Error		
	e_x	e_y	e_{xy}
LoS	0.0561 m	0.0701 m	0.0898 m
NLoS	0.0845 m	0.0974 m	0.1465 m
NLoS with Kalman	0.0647 m	0.0721 m	0.1092 m

scenarios. Table III shows the MAE of the tests. From the results, we can easily find that the result of the LoS scenario is clearly superior to the NLoS scenario. Table III also shows that the Kalman filter effectively improves the localization accuracy in the NLoS scenario. The effect can also be observed in Fig. 14.

In Table IV, we compare the accuracy of the our method with related works. It is important to note that this comparison may not be entirely fair, as the research studies used varied error metrics and were conducted under different experimental conditions. For better comparison, we have selectively presented only those works that utilize MAE and root mean square error (RMSE) as their evaluation metrics. In addition, the speed of the computation of each step is shown in Table V, which indicates the potential for real-time operation and embedded deployments.

TABLE IV
Accuracy Comparison of Related Works

Research	Methods	Localization Error		
		e_x	e_y	e_{xy}
[13]*	Ray Tracing	0.247 m	0.247 m	×
[14]*	Ray Tracing	×	×	0.283 m
[16]*	Direct Path	×	×	0.460 m
[17]*	Direct Path	×	×	0.980 m
[25]$^+$	Doppler	×	×	0.120 m
Our Work$^+$	L-CNN+Kalman	0.065 m	0.072 m	0.109 m

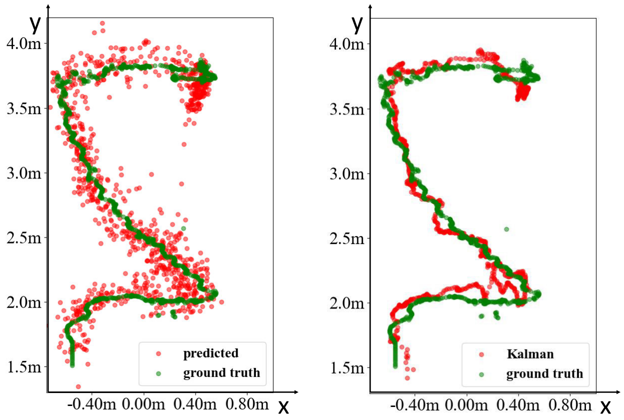
Note: * indicates the work uses RMSE, $^+$ indicates the work uses MAE

TABLE V
Computation Speed for Each Step in the System

Step / Device	AMD Ryzen7 4800H with NVIDIA GeForce GTX 1060 TI	Intel Core I5-8350U without GPU
Model Training time	1190 s	26791 s
One heatmap Generation	214 ms	493 ms
One Network Inference	2.2 ms	2.9 ms

TABLE VI
MAE for Test Data From Different Distances

Distance Level	MAE without filters		
	e_x	e_y	e_{xy}
Close (1.5-2.3 m)	0.0633 m	0.0679 m	0.1033 m
Middle (2.4-3.2 m)	0.0679 m	0.0807 m	0.1185 m
Far (3.3-4.0 m)	0.1088 m	0.1275 m	0.1927 m



(a) (b)

Fig. 14. Comparison between with/without the Kalman filter. It is easy to find that after the Kalman filter, the result is closer to the ground truth (annotation). (a) NLoS without filters. (b) NLoS with Kalman filter.

Our training data have a fairly proper distribution. However, after evaluating the accuracy of the model based on the distance from the Table VI, we notice that the prediction performance degrades when the target is far from the sensor. The reason is that the signal experiences more bounces between the walls to reach a far target, and the pattern is more difficult to determine. Additionally, as we mentioned earlier, the diffraction phenomenon may also account for this. However, the performance of the model is compromised when the target is close to the wall on the right side. This situation warrants a deeper investigation to uncover the underlying reasons behind this observation.

As Doppler is an important radar output parameter besides the angle and range, it is worth understanding the effect of the Doppler in the localizing the NLoS target with the mmWave radar. In this work, we also test our method by feeding the L-CNN with combined range-Doppler and range-angle maps during the training and inference. However, the tracking performance using the combined input shows significant degradation compared with the range-angle only input. The potential reasons for the degradation are: 1) the complicated geometry and propagation make the Doppler measurements in this work barely represent the actual direction and velocity of the target and 2) high-resolution Doppler measurement relies on longer slow time processing duration. However, the longer slow time duration introduces more ambiguity for the time varying velocity, which is the case in our experiment. Therefore, without proper processing, the Doppler measurements in this work could only be used as an indication of movement, rather than a complementary direction and velocity measurement for interpretation target position. To take advantage of Doppler to improve target tracking, a new neural network architecture needs to be developed. In particular, a neural network architecture that is able to represent the sequential relation between the current and previous locations will potentially make use of the Doppler measurement. The design of the new neural network architecture will be within the scope of our future work.

VI. CONCLUSION

In this article, we propose a novel system capable of detecting and localizing targets behind corners, substantiated by real-world experiments. Our approach employs mmWave sensing for data capture and camera-based annotation, leading to precise target localization. This framework demonstrates potential for enhancing NLoS detection capabilities.

Our research represents progress toward a comprehensive corner radar system, but reaching full-scale functionality remains an objective for future work. At present, our system is optimized for scenarios with a stationary radio source and a single target. The next phase involves expanding the capability of the system for collision prevention by accurately detecting and tracking multiple targets simultaneously. We aim to develop an advanced neural network that distinguishes between different entities, integrated with a robust tracking mechanism. Currently, this model has to be trained for every different corner scenario, despite its lightweight training requirements. Future enhancements will focus on generalizing this solution to diverse environments. The ultimate goal is to enable this system to identify targets regardless of radio source location, radio orientation, corner geometry, and wall composition.

REFERENCES

[1] D. Thangam and A. Sathish, "Transforming Indian industries through artificial intelligence and robotics in industry 4.0," *Int. J. Mech. Eng. Technol.*, vol. 9, no. 10, pp. 835–845, Oct. 2018.

[2] S. W. Kim, J. H. Kong, S. W. Lee, and S. Lee, "Recent advances of artificial intelligence in manufacturing industrial sectors: A review," *Int. J. Precis. Eng. Manuf.*, vol. 23, no. 1, pp. 111–129, Jan. 2022.

[3] I. B. Oran and H. R. Cezayirlioglu, "AI–robotic applications in logistics industry and savings calculation," *J. Organizational Behav. Res.*, vol. 6, no. 1, pp. 148–165, 2021.

[4] B. Zhou, D. Xie, S. Chen, H. Mo, C. Li, and Q. Li, "Comparative analysis of SLAM algorithms for mechanical LiDAR and solid-state LiDAR," *IEEE Sensors J.*, vol. 23, no. 5, pp. 5325–5338, Mar. 2023.

[5] F. Naser et al., "ShadowCam: Real-time detection of moving obstacles behind a corner for autonomous vehicles," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Maui, HI, USA, Nov. 2018, pp. 560–567.

[6] F. Naser et al., "Infrastructure-free NLoS obstacle detection for autonomous cars," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Macau, China, Nov. 2019, pp. 250–257.

[7] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," *Nat. Commun.*, vol. 3, no. 3, pp. 745, Mar. 2012.

[8] W. Chen, S. Daneau, C. Brosseau, and F. Heide, "Steady-state non-line-of-sight imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 6783–6792.

[9] C. A. Metzler et al., "Deep-inverse correlography: Towards real-time high-resolution non-line-of-sight imaging," *Optica*, vol. 7, no. 1, p. 63, Jan. 2020.

[10] K.-P.-H. Thai et al., "Around-the-corner radar: Detection and localization of a target in non-line of sight," in *Proc. IEEE Radar Conf.*, Seattle, WA, USA, May 2017, pp. 842–847.

[11] P. Setlur, T. Negishi, N. Devroye, and D. Erricolo, "Multipath exploitation in non-LOS urban synthetic aperture radar," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 1, pp. 137–152, Feb. 2014.

[12] Q. Zhao, G. Cui, S. Guo, W. Yi, L. Kong, and X. Yang, "Millimeter wave radar detection of moving targets behind a corner," in *Proc. 21st Int. Conf. Inf. Fusion (FUSION)*, Cambridge, U.K., Jul. 2018, pp. 2042–2046.

[13] J. Hyun, T. Oh, H. Lim, and H. Myung, "UWB-based indoor localization using ray-tracing algorithm," in *Proc. 16th Int. Conf. Ubiquitous Robots (UR)*, 2019, pp. 98–101.

[14] A. Tayebi, J. Gomez, F. Saez de Adana, and O. Gutierrez, "The application of ray-tracing to mobile localization using the direction of arrival and received signal strength in multipath indoor environments," *Prog. Electromagn. Res.*, vol. 91, pp. 1–15, 2009.

[15] T. Johansson, Å. Andersson, M. Gustafsson, and S. Nilsson, "Positioning of moving non-line-of-sight targets behind a corner," in *Proc. Eur. Radar Conf. (EuRAD)*, London, U.K., Oct. 2016, pp. 181–184.

[16] L.-X. Chuo, Z. Luo, D. Sylvester, D. Blaauw, and H.-S. Kim, "RF-echo: A non-line-of-sight indoor localization system using a low-power active RF reflector ASIC tag," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw.*, vol. 5, Snowbird, UT, USA, Oct. 2017, pp. 222–234.

[17] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single WiFi access point," in *Proc. 13th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, Santa Clara, CA, USA, 2016, pp. 165–178.

[18] N. Phukan, S. Mohine, A. Mondal, M. S. Manikandan, and R. B. Pachori, "Convolutional neural network-based human activity recognition for edge fitness and context-aware health monitoring devices," *IEEE Sensors J.*, vol. 22, no. 22, pp. 21816–21826, Nov. 2022.

[19] T. Kim and T.-H. Park, "Extended Kalman filter (EKF) design for vehicle position tracking using reliability function of radar and LiDAR," *Sensors*, vol. 20, no. 15, p. 4126, Jul. 2020.

[20] B. Zhao, D. Zhu, T. Xi, C. Jia, S. Jiang, and S. Wang, "Convolutional neural network and dual-factor enhanced variational Bayes adaptive Kalman filter based indoor localization with Wi-Fi," *Comput. Netw.*, vol. 162, Oct. 2019, Art. no. 106864.

[21] V. Vaish, B. Wilburn, N. Joshi, and M. Levoy, "Using plane+ parallax for calibrating dense camera arrays," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jul. 2004, pp. 2–9.

[22] J. Hur, J. Y. Lee, J. Choi, and J. Kim, "I see-through you: A framework for removing foreground occlusion in both sparse and dense light field images," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, Jan. 2023, pp. 229–238.

[23] M. Gustafsson, A. Andersson, T. Johansson, S. Nilsson, A. Sume, and A. Orbom, "Extraction of human micro-doppler signature in an urban environment using a 'sensing-behind-the-corner' radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 2, pp. 187–191, Jan. 2016.

[24] R. Linnehan and J. Schindler, "Multistatic scattering from moving targets in multipath environments," in *Proc. IEEE Radar Conf.*, Pasadena, CA, USA, May 2009, pp. 1–6.

[25] N. Scheiner et al., "Seeing around street corners: Non-line-of-sight detection and tracking in-the-wild using Doppler radar," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1–10.

[26] C. Wu, F. Zhang, B. Wang, and K. R. Liu, "mmTrack: Passive multi-person localization using commodity millimeter wave radio," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, Oct. 2020, pp. 2400–2409.

[27] P. Zhao et al., "Human tracking and identification through a millimeter wave radar," *Ad Hoc Netw.*, vol. 116, May 2021, Art. no. 102475.

[28] H. Liu, X. Liu, X. Xie, X. Tong, and K. Li, "PmTrack: Enabling personalized mmWave-based human tracking," in *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Dec. 2023, pp. 1–30.

[29] J. Jung, S. Lim, J. Kim, and S.-C. Kim, "Digit recognition using FMCW and UWB radar sensors: A transfer learning approach," *IEEE Sensors J.*, vol. 23, no. 16, pp. 18776–18784, Aug. 2023.

[30] V. Dakulagi and J. He, "Improved direction-of-arrival estimation and its implementation for modified symmetric sensor array," *IEEE Sensors J.*, vol. 21, no. 4, pp. 5213–5220, Feb. 2021.

[31] K. Witrisal et al., "High-accuracy localization for assisted living: 5G systems will turn multipath channels from foe to friend," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 59–70, Mar. 2016.

[32] P. Vaishnav and A. Santra, "Continuous human activity classification with unscented Kalman filter tracking using FMCW radar," *IEEE Sensors Lett.*, vol. 4, no. 5, pp. 1–4, May 2020.

[33] A. G. Stove, "Linear FMCW radar techniques," *IEE Proc. F Radar Signal Process.*, vol. 139, no. 5, pp. 343–350, Oct. 1992.

[34] S. Yue, H. He, P. Cao, K. Zha, M. Koizumi, and D. Katabi, "CornerRadar: RF-based indoor localization around corners," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 6, no. 1, pp. 1–24, Mar. 2022.

[35] S. R. Jondhale and R. S. Deshpande, "Kalman filtering framework-based real time target tracking in wireless sensor networks using generalized regression neural networks," *IEEE Sensors J.*, vol. 19, no. 1, pp. 224–233, Jan. 2019.

[36] Y. Xu and B. Tan. (2024). *Tracking the Occluded Indoor Target With Scattered Millimeter Wave Signal*. [Online]. Available: https://dx.doi.org/10.21227/bca8-3r93

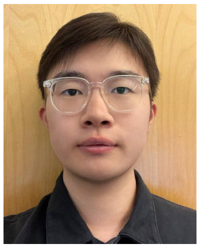
[37] C. Kumar, *User's Guide 60GHz mmWave Sensor EVMs*, TI, TI Blvd., Dallas, TX, USA, May 2022.

[38] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[39] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," 2014, *arXiv:1405.0312*.

Yinda Xu (Graduate Student Member, IEEE) was born in Ningbo, Zhejiang, China, in 1999. He received the M.Sc. (Tech) degree in automation and electrical engineering from Aalto University, Espoo, Finland, in 2023.

From 2022 to 2023 he was a Research Assistant at the Electrical Engineering Unit, Tampere University, Tampere, Finland. He is currently a Postdoctoral Researcher with the Electrical Engineering School, Aalto University. His research interests include the Internet of Things, smart buildings, and artificial intelligence.



Xinjue Wang (Student Member, IEEE) is currently pursuing the Ph.D. degree with the Department of Information and Communications Engineering, Aalto University, Espoo, Finland.

Juhani J. Kupiainen was born in Rauma, Finland, in 1997. He received the M.Sc. (Tech) degree in computer science from Tampere University, Tampere, Finland, in 2023.

From 2021 to 2022 he was a Research Assistant at the Electrical Engineering Unit, Tampere University. Since 2022, he has been a Software Engineer with Insta Advance Oy, Tampere.

Joonas Säe received the D.Sc. (Tech) degree from Tampere University of Technology, Tampere, Finland, in 2017.

He is a Staff Scientist with Tampere University. His current research interests include 5G mobile networks, the IoT technologies, and UAV communications.

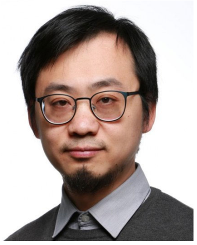
Jani Boutellier (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees from the University of Oulu, Oulu, Finland, in 2005 and 2009, respectively.

He is an Associate Professor with the School of Technology and Innovation, University of Vaasa, Vaasa, Finland. His research interests include computer vision, machine learning, and edge computing.

Jari Nurmi (Senior Member, IEEE) received the M.Sc., Lic.Tech., and D.Sc. (Tech) degrees from Tampere University of Technology (TUT), Tampere, Finland, in 1988, 1990, and 1994, respectively.

He has been working as a Professor with the Electrical Engineering Unit, Tampere University, TAU (formerly Tampere University of Technology, TUT), since 1999. He held various research, education, and management positions at TUT, since 1987, and was the Vice President of the SME VLSI Solution Oy from 1995 to 1998. He has supervised 32 Ph.D. and over 150 M.Sc. theses and has been the opponent or reviewer of about 50 Ph.D. theses for other universities worldwide. He has co-edited five Springer books and has published over 400 international conference and journal articles and book chapters. He is the Director of the National DELTA Doctoral Training Network of 200 Ph.D. students, a coordinator of the European Doctoral Training Network APROPOS, and the Head of A-WEAR European Joint Ph.D. Program at TAU.

Prof. Nurmi is a member of the technical committee on VLSI Systems and Applications at IEEE CASS.



Bo Tan (Member, IEEE) received the Ph.D. degree from the Institute for Digital Communications (IDCOM), The University of Edinburgh, Edinburgh, U.K., in 2013.

From 2012 to 2016, he conducted multiple postdoctoral research projects at the University College London, London, U.K., and the University of Bristol, Bristol, U.K., contributing to passive radar design and applications in security and healthcare. From 2017 to 2018, he was a Lecturer at Coventry University, Coventry, U.K. Since 2019, he has been a Tenure Track Assistant Professor at Tampere University, Tampere, Finland. His research interests include radio sensing and connectivity for intelligent machines.

Dr. Tan is the PI and coordinator of projects funded by the Academy of Finland, Business Finland, and Horizon European. He is the Vice Chair of the IEEE Finland AP/ED/MTT Chapter, an Associate Editor of IEEE WIRELESS COMMUNICATIONS LETTERS, and a reviewer of multiple IEEE/ACM/IET journals and conferences in wireless communications, radar, pervasive computing, and sensing.