

RDGait: A mmWave Based Gait Recognition System for Complex Indoor Environments Using Single-chip Radar

DEQUAN WANG, University of Science and Technology of China, China

XINRAN ZHANG, University of Science and Technology of China, China

KAI WANG, University of Science and Technology of China, China

LINGYU WANG, University of Science and Technology of China, China

XIAORAN FAN, Google, USA

YANYONG ZHANG*, University of Science and Technology of China, China and Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, China

In this paper, we aim to study millimeter-wave-based gait recognition in complex indoor environments, focusing on dealing with multipath ghosts and supporting rapid deployment to new environments. We design a ghost detection algorithm based on velocity change patterns. This algorithm relies solely on velocity estimation, requiring no environmental priors or multipath modeling. Hence, it is suitable for single-chip millimeter-wave radar with low angular resolution and can be conveniently deployed in new indoor settings. In addition, we build a gait recognition network based on an attention-based Recurrent Neural Network (RNN) to extract spatiotemporal-velocity features from RD heatmaps.

We have evaluated RDGait in two scenarios: a corridor scenario and a crowded office scenario, with 125 volunteers of different genders and ages ranging from 6 to 63. RDGait achieves a user recognition accuracy exceeding 95% among 125 candidates in both scenarios. We have further deployed RDGait in two additional scenarios using the pretrain-finetune approach. With minimal user registration data, RDGait could achieve satisfactory (> 90%) recognition accuracy in these new environments considering different radar placements, heights, and number of co-existing users.

CCS Concepts: • Human-centered computing → Ubiquitous and mobile computing.

Additional Key Words and Phrases: mmWave radar, human identification, ghost identification

ACM Reference Format:

Dequan Wang, Xinran Zhang, Kai Wang, Lingyu Wang, Xiaoran Fan, and Yanyong Zhang. 2024. RDGait: A mmWave Based Gait Recognition System for Complex Indoor Environments Using Single-chip Radar. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 3, Article 134 (September 2024), 31 pages. <https://doi.org/10.1145/3678552>

*Corresponding author.

Authors' Contact Information: Dequan Wang, University of Science and Technology of China, Hefei, Anhui, China, wdq15588@mail.ustc.edu.cn; Xinran Zhang, University of Science and Technology of China, Hefei, Anhui, China, zxrr@mail.ustc.edu.cn; Kai Wang, University of Science and Technology of China, Hefei, Anhui, China, wkzcm@163.com; Lingyu Wang, University of Science and Technology of China, Hefei, Anhui, China, wly18654382511@gmail.com; Xiaoran Fan, Google, Irvine, California, USA, vanxf@google.com; Yanyong Zhang, University of Science and Technology of China, Hefei, Anhui, China and Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, Anhui, China, yanyongz@ustc.edu.cn.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 2474-9567/2024/9-ART134

<https://doi.org/10.1145/3678552>

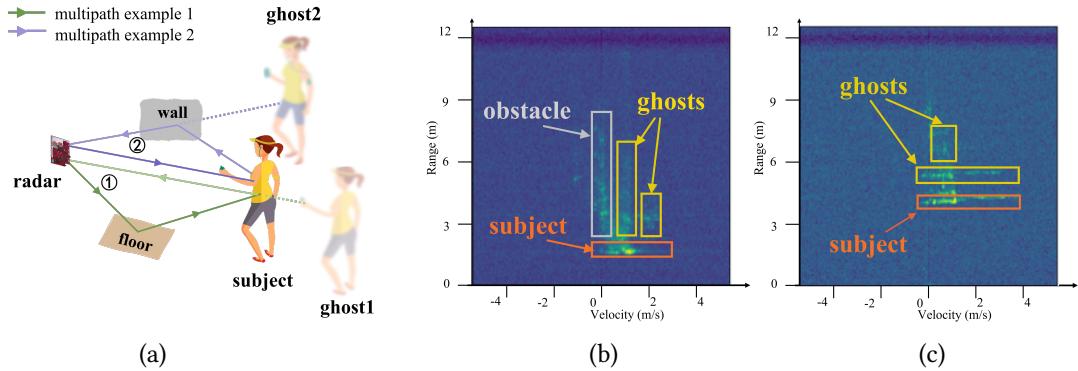


Fig. 1. (a) The green line and the purple line illustrate two examples of multipath. Two Ghost targets are generated when radar signals undergo additional reflections off surfaces. The darker green or purple represents stronger signal intensity. (b) and (c) illustrate the RD Heatmaps that contain points corresponding to real subjects, ghost targets, and obstacles.

1 Introduction

Smart space, including smart offices, homes, etc., has emerged as one of the most impactful applications in recent years that drives various technological advancements, ranging from smart care [16, 58, 63], health monitoring [12, 15, 46], to intelligent security [25, 55]. The increasing demands for these applications underscore the need for enhanced system capabilities, particularly in providing personalized services tailored to each user. To meet this need, the key is the ability to offer remote identity recognition, which has sparked widespread attention towards passive biometrics-based user identification, especially **gait recognition** [41, 45, 54, 62].

A person's gait describes her distinctive walking pattern, including factors such as stride length, pace, and rhythm [19]. The gait feature is commonly acknowledged for its uniqueness and can be utilized for user recognition purposes, using data collected from various sensors, such as cameras [4, 10, 26, 47], pressure sensors [24, 32, 44], WiFi [22, 42, 56], and ultrasonic radars [28, 59]. Compared to the aforementioned sensors, millimeter-wave (mmWave) technology presents unique advantages, such as high resolution, non-intrusiveness, immunity to lighting conditions, and privacy-friendliness. These attributes make mmWave a compelling alternative for indoor gait recognition [14]. For example, mmGaitNet [31] collects point clouds by placing two mmWave radar models to extract features for each attribute of the point cloud. MU-ID [61] utilizes range-time spectrograms to extract users' spatial-temporal lower limb motion features.

Despite the promises, mmWave-based gait recognition brings new challenges. The primary challenge lies in the presence of multipath ghosts in indoor environments. When the radar signals encounter subjects, obstacles, or reflective surfaces on the propagation path, the multipath effect takes place [17]. This phenomenon results in reflections arriving at different times. The presence of these delayed reflections leads to ghosts, as illustrated in Fig. 1. These ghosts could mislead the gait recognition system in various ways. They can result in false detections; they can also overlap with a subject, together forming a different target. Approaches that identify and remove ghost targets often rely on multipath environment-dependent features [13, 17, 21], rendering it cumbersome to deploy the systems in different settings. This in turn can undermine the wide adoption of a mm-wave-based gait recognition system.

In this work, we present RDGait, an approach that aims to identify subjects robustly through gait recognition in complex indoor environments. Our approach takes the radar's Range-Doppler (*RD* for short) heatmaps, hence the name. When the subjects enter the monitored area, we first calculate the Fast Fourier Transform (FFT) of the signals and generate RD Heatmaps [53]. In the preprocessing, we filter out the background and extract the dense

regions as clusters[43], with each cluster corresponding to one of the following three situations: (1) an actual subject, (2) a ghost target, and (3) a mix of the two. We apply a set of normal walking constraints to iteratively split mixed clusters into "minimum" ones that correspond to either a single subject or a single ghost. In corner cases such as a dad holding his baby, we believe it is acceptable to treat the dad and baby pair as a single subject. We refer to these minimum clusters as the targets in our system. Next, we engage a target tracking algorithm to associate clusters across multiple heatmap frames, obtaining each target's trajectory.

The next task is to discriminate the actual subjects and ghosts from the list of targets. A class of ghost recognition algorithms is geometric-model-based. For example, the work in [17, 27, 64] utilizes ray tracing, which models the paths of multipath reflections, deducing the positions of ghost targets through the subject's location and the positions of reflective surfaces based on the laws of reflection. These methods require modeling large surfaces such as walls, floors, and ceilings in the room, increasing the system's deployment cost. On the other hand, large furniture such as cabinets, dining tables, and doors may change position, rendering the previous model ineffective. The approach introduced in [29] presents a ghost detection method relying on angle consistency. The method proposed in [5] avoids manual mapping by simultaneously performing ghost detection and constructing reflection maps. These two articles have made pioneering contributions to the problem of model-free ghost detection. However, when utilizing lower-cost, lower angular resolution millimeter-wave radars, both angle variations and mapping quality are affected by the angular resolution, resulting in performance degradation.

To tackle this problem, we take a close look at the obtained target's trajectories and find that some target pairs (i.e., a subject and her corresponding ghost) exhibit correlation in their velocities. However, the correlation between the velocities is hard to capture as their relationship varies – e.g., sometimes one's velocity is twice as fast as the other one, sometimes in the opposite direction, and sometimes slower than the other one. To solve the myth, we utilize the principles of the Doppler effect to analyze the relationship between the velocities of ghosts and subjects. We find that their velocity change patterns are often the same. As such, when one picks up speed, the other also shows faster speed; when one slows down, the other also shows a decrease in velocity. This trend is more pronounced when we focus on the torso velocity. Accordingly, we propose a torso velocity change pattern-based ghost detection algorithm. This algorithm solely leverages the velocity correlation between the subject and ghost, requiring no prior information, being independent of angles, and avoiding specific modeling of multipath reflections. Because multiple-channel information is not required, our algorithm is equally applicable to single-transmit single-receive millimeter-wave radar.

In addition, to fully exploit the spatio-temporal-velocity information in the RD Heatmaps, we devise a gait recognition network based on attention to extract gait features and generate identification results. Given the significant differences in gait phases, we design an attention-LSTM module to assign varying weights to different frames. Furthermore, we also carefully consider the need for smooth deployments to different environments. We attempt to achieve this through a pretrain-finetune approach. For the pretraining, we collect a large dataset as the supporting set. After pretraining the model, we can deploy the trained system in different environments. For example, we can deploy a RDGait system in a smaller office. To facilitate the deployment of RDGait, we need to go through the following steps: (1) *radar placement*, adjusting the positioning and elevation of the radar based on specific situations in the room, (2) *user registration*, collecting a small amount of user gait data, (3) *model fine-tuning*, using the registration data to fine-tune the previously trained model. Following the aforementioned steps, we achieved rapid adaptation to new scenarios using a small amount of registration data.

We have implemented a RDGait prototype system using a TI IWR6843BOOST single-chip radar [49]. RD represents Range-Doppler. We have collected a supporting data set from 125 subjects, ages ranging from 6 to 63, from a crowded office and a spacious corridor. In total, we have over 50 hours of supporting data. Additionally, we have fine-tuned and tested the RDGait model in 2 different scenarios, each with a small amount of registration data, to show that the proposed system can be conveniently deployed across environments. We summarize our contributions below:

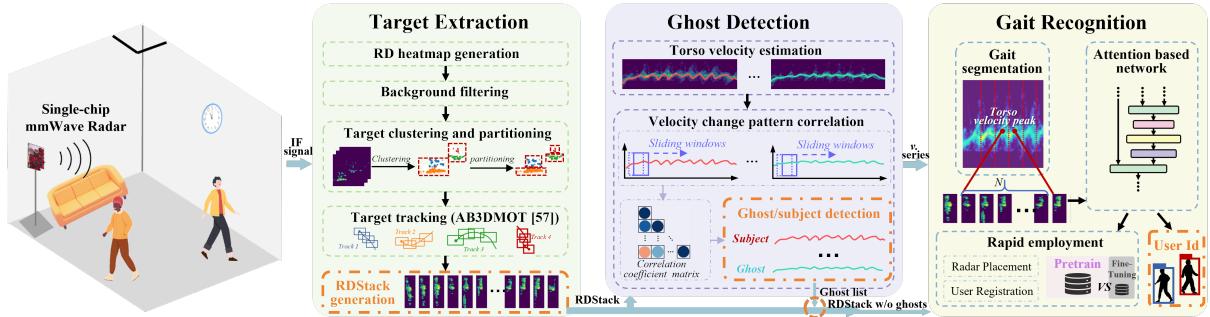


Fig. 2. The overview of our RDGait system, consists of three main modules: target extraction, ghost detection, and gait recognition. We annotate the output module of each part with **orange** boxes and use v . to stand for the abbreviation of velocity.

- (1) We develop a gait recognition framework named RDGait based on single-chip mmWave radar, which can accurately recognize user gaits in complex indoor environments, even when multiple users are present and can be conveniently deployed in different indoor environments.
- (2) We devise a series of techniques towards robust gait recognition, including walking constraints guided clustering and partitioning, torso velocity change pattern correlation-based ghost detection, and an attention-based gait recognition deep learning network. .
- (3) We conduct a large-scale and thorough evaluation of RDGait in the real world, with a supporting set featuring 2 scenarios and 125 volunteers spanning a wide age range (60 years). Our experimental results adequately show the effectiveness of RDGait, demonstrating a remarkable gait recognition top-1 accuracy of 96.5% among the 125 users. We have carefully studied the system performance with different system parameters such as radar placement and elevation and our system demonstrates great robustness in all settings. When deploying the system in new environments, we could achieve a recognition accuracy exceeding 90% using only 1 minute of registration data per person.

2 PRELIMINARY

In this section, we introduce the principles of range and velocity measurement using FMCW millimeter-wave radar, along with the physiological background of human walking. We then investigated the generation and characteristics of far-off multipath ghosting, exploring methods for removing ghost targets.

2.1 Doppler effect

The Doppler effect describes the frequency change of radiofrequency waves due to the relative motion of the radar or the subject. If a subject moves away from the radar, the observed RF signal will be shifted toward a lower frequency. Conversely, if a subject moves toward the radar, the observed RF signal will be shifted toward a higher frequency. The formula for the Doppler shift (Δf) in terms of frequency is given by:

$$\frac{\Delta f}{f} = \frac{v}{c} \cos\theta, \quad (1)$$

where Δf is the change in frequency, f is the source frequency, v is the relative velocity between the radar and the subject, c is the speed of light, θ is the angle between the relative velocity and the direction of the wave. mmWave radar employs the Doppler effect to calculate the frequency shift between the received and transmitted waves, facilitating the measurement of a subject's velocity.

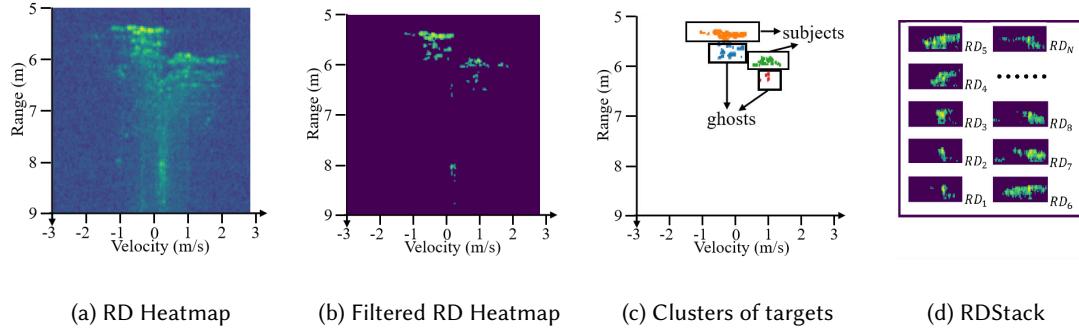


Fig. 3. Illustration of an RD Heatmap (a), an RD Heatmap with background removed (b), target clusters (c), and the RDStack over N frames (d). They are the intermediate results in our Target Extraction module.

2.2 Principles of radar sensing

RDGait employs a single-chip radar operating with a sequence of FMCW signals to detect human motion. Initially, the radar's transmitting antenna continuously sends out FMCW signals. These signals are reflected to the radar's receiving antenna when they encounter subjects/objects. Specifically, when a subject is moving, the reflected signals result in a Doppler frequency shift.

Upon the signals' return to the receiving antenna, a frequency mixer and a low-pass filter are employed to merge these signals and generate the intermediate frequency (IF) signals. The IF signal includes the frequency and phase differences between the received and transmitted signals. The frequency difference is caused by two factors: the time delay due to the propagation distance and the Doppler shift induced by velocity. Next, two consecutive fast Fourier transforms (FFT) are used to transform the signal into the frequency domain, to obtain **range** and **Doppler velocity** information of the reflecting objects. The range represents the straight-line distance from the radar to the object. The IF signals are processed into a **range-Doppler heatmap** (RD Heatmap). RD Heatmap is a two-dimensional matrix, where the values in the matrix represent the intensity of reflected signals at a range r and velocity v , which is shown in Fig 1(b),(c).

2.3 Background of Human Walking

Gait refers to specific movement patterns and biological characteristics generated when individuals walk. These features encompass step length, arm swing, torso sway, foot clearance, and so on [19]. Gait features are determined by the unique biomechanics and body structure of an individual during walking.

mmWave radar can be used for detecting and recognizing gait. The body reflects the mmWave signals. Signals reflected from different parts of the body, such as the head, arms, and legs, exhibit specific characteristics. By analyzing the signals, the system can extract gait features such as gait cycle, step length, walking speed [51], and even the position of each joint [60]. Compared to these extracted features, the RD Heatmaps contain more information – the distribution of distances between different body parts, velocity distribution, and the joint distribution of distance and velocity. Therefore, in RDGait, we employ a deep learning approach to extract gait features from the RD Heatmaps.

2.4 Principle of multipath ghosting

This section focuses on the causes of multipath effects and the characteristics of ghost targets. In radar systems, multipath effects occur when radar signals encounter multiple reflective surfaces or subjects in the environment

before reaching the receiver. These reflections can create multiple paths for the radar signal to travel, resulting in the reception of multiple delayed versions of the original signal. FMCW radar calculates range based on the time delay, where a greater time delay corresponds to a longer range. Therefore, in the mmWave radar heatmap or point cloud, false detections, known as **ghost targets**, occur in areas farther away than the actual target range, as shown in Fig. 1. Let's examine a specific process of multipath effects:

Doppler frequency shift under multipath effects. We assume that a millimeter-wave signal undergoes N reflections before returning to the radar receiver. After each reflection, the reflecting surface becomes the new wave source. The relative velocity between the reflecting surface and the wave source is v_i , and its angle to the signal is θ_i . Therefore, the Doppler frequency shift generated by the i -th reflector is: $\frac{\Delta f_i}{f} = \frac{v_i}{c} \cos\theta_i$. We sum up the Doppler frequency shifts from all N reflections to obtain the overall Doppler frequency shift:

$$\frac{\Delta f}{f} = \sum_{i=1}^N \frac{v_i}{c} \cos\theta_i. \quad (2)$$

Therefore, the Doppler velocity of the ghost target is $v' = \frac{1}{2} \sum_i v_i \cos\theta_i$, θ_i is the angle between the velocity of the i -th reflecting object/subject and the direction of the wave. Furthermore, we assume that the distance traveled by the signal between each reflection is L_i . The range of the ghost target is:

$$r' = \sum_{i=0}^N L_i / 2. \quad (3)$$

This indicates that the range of the ghost target is greater than that of the object.

3 Overview of RDGait

3.1 Design Goals

RDGait is designed to support robust gait recognition in various indoor environments and achieve rapid deployment in new scenarios, focusing on the following goals.

- **High recognition accuracy.** RDGait should extract more comprehensive gait information to enhance its ability to distinguish among different individuals.
- **Environment-independent ghost detection.** RDGait should achieve environment-independent ghost detection and ensure high precision and recall in ghost detection.
- **Rapid deployment.** RDGait should enable rapid deployment of new environments while using a small amount of user registration data.

3.2 System Overview

RDGait aims to remove indoor multipath ghosts, achieve accurate gait recognition, and quickly deploy in new environments. The framework of RDGait is shown in Fig. 2. Below we provide an overview of RDGait.

Target Extraction. RDGait first normalizes and filters the input heatmap, and then clusters it into multiple targets. RDGait associates targets across several frames, achieving continuous tracking, and generates a range-Doppler stack (RDStack) for each target.

Velocity-Correlation Based Ghost Detection. We determine which targets in the set are ghost targets. RDGait initially estimates the torso velocity, creating a velocity series for each target. Subsequently, RDGait computes velocity correlations between pairs to determine which ones are ghost targets.

Attention Based Gait Recognition. Leveraging the periodicity in the walking process, We first segment the RDStack into samples of fixed size. Next, we employ a carefully designed attention-based network to extract gait

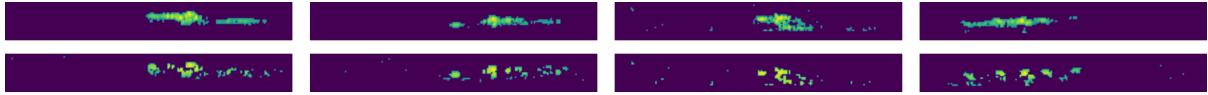


Fig. 4. The performance of background filtering for RDGait (up) and CA-CFAR (bottom)

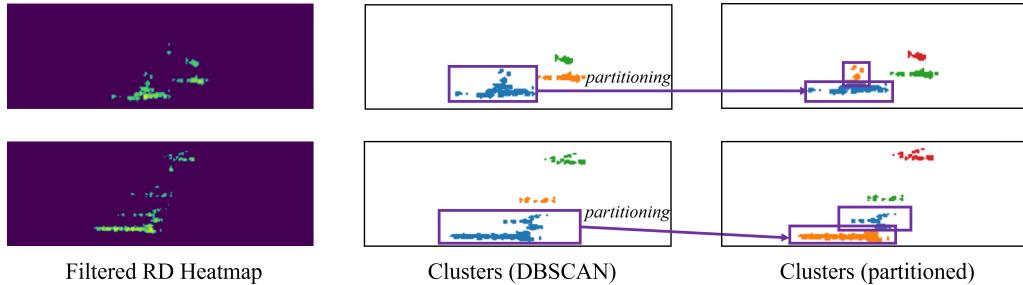


Fig. 5. Illustration of cluster partitioning. The DBSCAN algorithm generates mixed clusters, with each cluster containing a subject and a nearby ghost. Our cluster partitioning algorithm outputs "minimum" clusters, each corresponding to a single subject or a single ghost, by applying walking constraints onto each mixed cluster.

features. We collected a gait dataset of 125 individuals as a supporting set for pre-training and then fine-tuned the model in a new scenario to achieve rapid deployment.

4 Design of RDGait

When the subjects enter the monitored area, we calculate FFT of the echo signals and generate RD Heatmaps, as described in 2.2. Subsequently, our data preprocessing step (§4.1) first filters out background clutters, and then performs clustering and tracking to obtain targets. Furthermore, ghost detection is used to identify and remove ghosts. Finally, through the attention-based gait recognition network (§4.2), RDGait can achieve accurate user recognition with less user registration data.

4.1 Data Preparation

In the beginning, we apply FFT to the radar IF signal, resulting in an RD Heatmap. We denote this RD Heatmap as:

$$RDH = \{p_s^i\} + \{p_c^i\} + \{p_g^i\} + \{p_b^i\}, \quad (4)$$

where $\{p_s^i\}$, $\{p_c^i\}$, $\{p_g^i\}$, $\{p_b^i\}$ is the set of subjects, clutter, ghost targets, and background, respectively. Here, i is the index of each point in the set. In RDGait, we aim to recognize and track the *subjects*, who move around in the environment. *Ghost targets* are false targets due to multipath reflections (The details can be found in Sec. 2.4), which exhibit similar intensity in the heatmap, prone to being mistaken for subjects. The *clutter* points refer to noises or mixed signals caused by reflections from static objects. The *background* points refer to the relatively weak signals generated in places where no objects are present.

In the data preprocessing phase, we start with target extraction (Sec. 4.1.1) to remove background and clutter. Then, a ghost detection (Sec. 4.1.2) module is engaged to identify and remove ghost targets.

4.1.1 Target extraction. In this section, we filter out the background and clutter points, cluster the remaining points as targets, and continuously track the targets across frames.

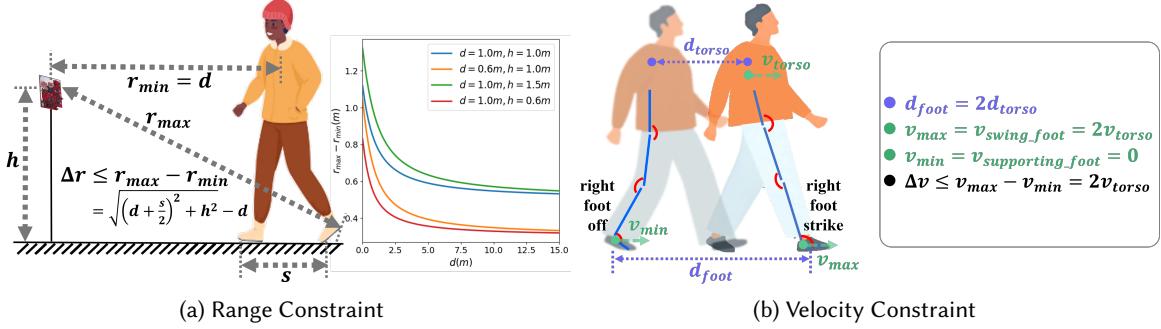


Fig. 6. Two walking constraints. (a) illustrates a schematic representation of a person’s nearest and farthest distances from the radar while walking. (b) illustrates a “step”, representing the process right foot from off to strike, where d_{torso} represents the displacement of the torso, and d_{foot} signifies the displacement of the foot.

Background filtering. We first employ the MinMax algorithm [50] to normalize each frame’s heatmap, rendering the values of each point in the RD Heatmap between 0 and 1. If the resulting point value is smaller than the predetermined threshold, we will set it to 0 to remove it:

$$I(r, v) = \begin{cases} I(r, v), & \text{if } I(r, v) \geq I_{th} \\ 0, & \text{else} \end{cases}, \quad (5)$$

where $I(r, v)$ is the intensity of a point located at a range of r and a velocity of v , and I_{th} is the threshold of intensity. In our implementation, we empirically set I_{th} as 0.5. A higher I_{th} will lead to ignoring weaker motions. Instead, too many background points will be retained with a lower I_{th} . The RD Heatmaps before and after filtering are shown in Fig. 3(a) and (b) respectively. During human walking, the distribution of body velocity is continuous. This is represented on an RD heatmap, where many adjacent Doppler bins display high intensities. In our experiments, since CFAR uses local reference windows for noise estimation, we found that if many bins in the Doppler domain have high intensities, CFAR will select the bins with the higher intensities as valid points, rather than retaining all bins containing velocity information. Our algorithm and CA-CFAR’s performance in background filtering is shown in Fig. 4. The image’s top portion shows our algorithm’s performance, while the bottom portion shows the performance of CA-CFAR. CA-CFAR filters out some information along the velocity dimension.

Target clustering and partitioning. Although the clutter has been removed, the subject, static objects, and ghost targets still need to be further classified. To address this, we employ Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [9]. DBSCAN is a clustering algorithm that groups data points based on their density in a spatial space, identifying dense regions as clusters while marking sparser areas as noise. We show the clustering process in Fig. 3(c). However, nearby targets that overlap with each other may be mistakenly identified as a single target, as illustrated in Fig. 5. To address this issue, we propose to consider the walking constraints in the clustering module.

The first constraint is *range constraint*, which specifies the constraint on the radar-subject distances. If we use r_{min} to denote the minimum distance between the radar and a subject, and r_{max} to denote the maximum distance, then we assume the distance between a valid subject and the radar falls within the range (r_{min}, r_{max}) . As shown in Fig. 6(a), if we consider the geometric details involved in the radar deployment and human walking, we further

have:

$$\Delta r = r_{max} - r_{min} = \sqrt{(d + \frac{s}{2})^2 + h^2} - d, \quad (6)$$

where h is the height of the radar, d is the horizontal distance between the person and the radar, and s is the person's step length. The average step length for an adult is between 0.6m and 0.8m [19]. According to Eq. 6, when s takes the value in the range of [0m, 15m], the maximum value of Δr is 1.0m in our setup. Considering arm movements and the energy leakage in the range dimension, we have appropriately extended this maximum range to 1.2m.

The second constraint is *velocity constraint*, which specifies the constraint on the maximum and minimal velocity when a person is walking. During walking, legs undergo two distinct phases, swinging and supporting, as shown in Fig. 6(b). If we assume the torso speed of a person walking is v_s , then the maximum speed during walking is $v_{max} = 2v_s$ (at the swinging foot), and the minimum speed is $v_{min} = 0$ (at the supporting foot). Therefore, the valid velocity of a person walking falls within the range (0, $2v_s$). The walking velocity of adults is approximately between 1.1m/s to 1.7m/s [19]. Therefore, we set the maximum velocity for valid clusters to 3.4m/s.

If the clusters extracted by the DBSCAN satisfy these two constraints, we regard them as valid. For the invalid clusters that do not satisfy the constraints, we recursively divide them into two or more clusters (see Alg. 1). Specifically, we iterate over all the clusters. For each cluster k , we perform M attempts of the k-means [30] clustering. If at an attempt all the resulting clusters satisfy the two constraints, the algorithm stops and outputs the clusters. Fig. 5 shows the clusters after partitioning. The cluster with constrained accurately separates multiple partially overlapping targets.

Target tracking. Tracking associates clustering results across different frames for obtaining a continuous motion trajectory. To monitor the continuous changes in gait, we employed AB3DMOT [57] to track the trajectory of targets. AB3DMOT first uses the Kalman filter [20] to track the clusters and employs the Hungarian algorithm [23] to associate clusters between consecutive frames. Then, we define a region of interest for each target, referred to as the *range-Doppler region* (RDR region):

$$RDR = \{(r, v) | |r - r_c| < \frac{s}{2} \text{ and } |v| < v_m\}, \quad (7)$$

where (r, v) is the point coordinates in heatmap, r_c is the cluster center after tracking, s is the length of one step and v_m represents the radar's maximum detectable velocity. When two target clusters are nearby, especially during a face-to-face encounter, additional target clusters emerge within the RD region. By removing the heatmap points associated with other target clusters, we obtain the RDR specific to the target cluster. After completing the tracking of a target, we combine all the RDRs associated with that target to form its range-Doppler Stack (RDS), as shown in Fig 3.

4.1.2 Velocity-correlated based ghost detection. After target extraction, we obtain range-Doppler stacks (RDS) for the extracted targets, which contain both real subjects and ghost targets. We next discuss how to discriminate between them. Let us begin our discussion with an observation of the velocity distributions of the extracted targets.

Observation. Intuitively, we take the viewpoint that there is a strong correlation in velocity between a subject and her/his corresponding ghost targets. Inspired by this thought, we carefully examine the velocity distributions and observe that *The velocities of ghosts are sometimes slower than that of the real subjects, sometimes twice as fast, and sometimes in the opposite direction, while the velocities of different subjects are not correlated at all.* We illustrate it and the corresponding RD Heatmap in Fig. 7.

Analysis of the ghost velocity. Next, we explain the above observations through a more formal analysis. Suppose a ghost target is caused by the radar wave being reflected N times. According to Eq. 2, we get the velocity

Algorithm 1: Walking Constraints Guided Clustering and Partitioning

Input: Data points $X = \{x_i = (r_i, v_i)\}$. Maximum range span R_{\max} , Maximum velocity span V_{\max} .

- 1 Use DBSCAN to cluster the point set X and obtain the initial clusters $C = \{c_1, c_2, \dots, c_n\}$;
- 2 **for** $i = 1$ to n :
- 3 **if** $\max_range(c_i) - \min_range(c_i) > R_{\max}$ **or** $\max_vel(c_i) - \min_vel(c_i) > V_{\max}$:
- 4 **for** $k = 2$ to N :
- 5 **repeat**
- 6 $c_i^1, c_i^2, \dots, c_i^k \leftarrow \text{K-Means}(k, c_i)$;
- 7 **for** $j = 1, 2, \dots, k$:
- 8 **if** $\max_range(c_i^j) - \min_range(c_i^j) > R_{\max}$ **or** $\max_vel(c_i^j) - \min_vel(c_i^j) > V_{\max}$:
- 9 **continue**;
- 10 $C.\text{pop}(c_i)$;
- 11 $C.\text{push}(c_i^1, c_i^2, \dots, c_i^k)$;
- 12 **break** and jump to line 3;
- 13 **until** M times;
- 14 **return** Clusters C .

of the ghost target v' as:

$$v' = \frac{1}{2} \sum_i^N v_i \cos \theta_i, \quad (8)$$

where v_i is the relative velocity between the reflecting surface (the i -th object/subject) and the wave source (the $(i-1)$ -th object/subject), and θ_i is the angle between the relative velocity and the direction of the wave.

We illustrate three typical multipath scenarios that generate ghosts in Fig. 7(b)-(d). If the radar waves undergo two reflections in opposite directions with a person during the propagation process, as shown in Fig. 7(b). Assuming the person's velocity is v , four Doppler frequency shifts occur, according to Eq. 8, the velocity of the ghost is $2v$. Similarly, the velocities of ghosts in Fig. 7(c) and Fig. 7(d) are $\frac{1+\cos\theta}{2}v$ and $-\frac{1-\cos\theta}{2}v$, respectively. Here, θ represents the angle between the incident wave and the velocity of the human body. These scenarios further confirm our observations above.

In particular, it is most common for only one subject to appear in the reflection path, given the person's velocity of v . Assume that there are K reflections occurring on the subject. For each reflection process on the subject, two Doppler frequency shifts occur: the first occurs when the wave is emitted from the object toward the subject, and the second occurs when the wave is emitted from the subject toward the object. There is no Doppler frequency shift when a wave propagates from a stationary object to another. Therefore, there are a total of $2K$ Doppler frequency shifts. the velocity of the ghost target is calculated as:

$$v' = \frac{1}{2} \sum_i^N v_i \cos \theta_i = \frac{1}{2} \sum_i^{2K} v \cos \theta_i = \frac{1}{2} v \sum_i^{2K} \cos \theta_i = v f(\theta_i), \quad f(\theta_i) = \frac{1}{2} \sum_i^{2K} \cos \theta_i, \quad (9)$$

We find that the velocity of the ghost target is directly proportional to the subject's velocity, with a proportionality factor of $f(\theta_i)$. The factor $f(\theta_i)$ varies with θ_i , which changes when the subject moves. Suppose we take a small time window during which the subject barely moves and $f(\theta_i)$ can be considered constant, marked as f_0 . Therefore, we have $v' = f_0 \cdot v \propto v$. As the subject starts moving around, the velocity of the ghost target will change accordingly; the target's speed increases when the subject picks up speed, while the target's velocity decreases when the subject slows down.

That is, the velocity change pattern of the ghost target is highly correlated with that of the subject. Meanwhile, the velocity change patterns of different users are drastically different, hence not correlated at all. Therefore, we

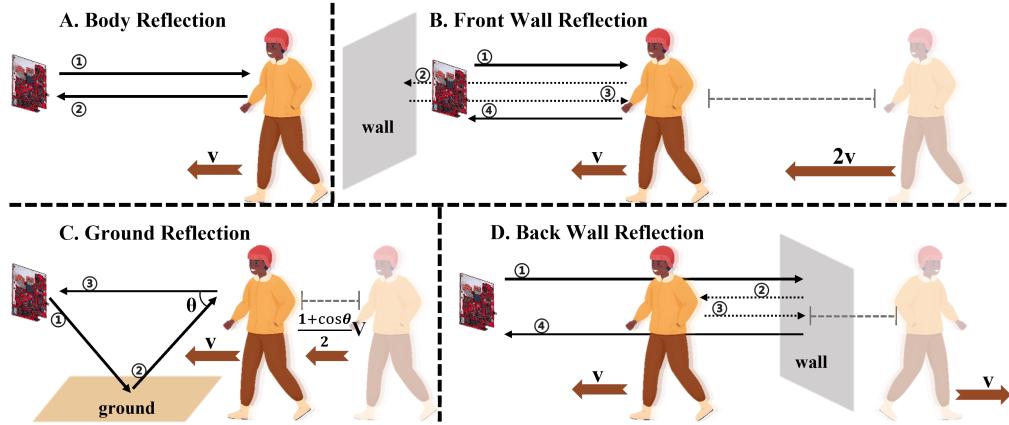


Fig. 7. Illustration of detecting a real subject and several ghost targets. (a) The reflection path generates real subjects. (b),(c), and (d) represent three typical reflection paths generating ghost targets, formed by reflections off the front wall, ground, and back wall, respectively. The process of emitting and reflecting signals is represented by a numerical sequence. The sequence numbers ①②③④ represent the propagation path of radar signals.

propose to mine the correlation between each target's velocity change patterns to identify the ghost target. That is, if there is a strong correlation in the velocity change patterns between two targets over the same period (e.g., 1 second), then one of them is most probably the ghost.

When there is a significant difference in velocity between two subjects, their velocity change patterns will be markedly different. Next, we illustrate that even when two subjects have similar velocities (including magnitude and direction), their velocity change patterns are still distinct. We assume two subjects are walking in the same area facing the same direction, divided into the following three scenarios: In the first scenario, depicted in Fig. 8(a), the step frequencies of the two subjects differ. Subject 1 takes 5 steps within 2.5 seconds, while Subject 2 takes only 4 steps. The velocity change patterns of the two subjects are unrelated. In the second scenario, depicted in Fig. 8(b), the step frequencies of the two subjects are the same, but the peak velocities of each step occur at different times. They exhibit a high degree of overlap after a temporal translation. However, due to the lack of time alignment, there is no clear proportional relationship between their velocities. In the third scenario, depicted in Fig. 8(c), one subject precisely mimics the movements of the other, resulting in velocity-time curves that closely align. Despite this high degree of synchronization, the differences in individual gait patterns often lead to variations most of the time. For example, within the sections marked by black circles in Fig. 8(c), there is a difference in the minimum velocities; Subject 2's minimum velocity is markedly lower than that of Subject 1. Therefore, even when one person follows and mimics the gait of another, our algorithm can still differentiate between the subject and the ghost in most cases. We also discuss it in the experimental section (§7).

Ghost detection. We utilize the above insight to design a ghost detection algorithm (see Alg. 2). First, we extract the torso velocity time series of all targets. Then, we calculate the matrix representing the velocity change pattern correlation between all targets and discriminate the ghost targets. We present the details below.

• **Torso velocity estimation.** Although the velocity changes in body parts such as the arms and legs are more evident, they are susceptible to obstruction. Therefore, we choose to extract the velocity of the torso, which is stable and less prone to obstruction. We first attempt to extract torso velocity using the percentile method discussed in [51]. The percentile method calculates the weighted median of target velocities utilizing the micro-Doppler spectrum of the target and can be used to extract torso velocities. Specifically, this method

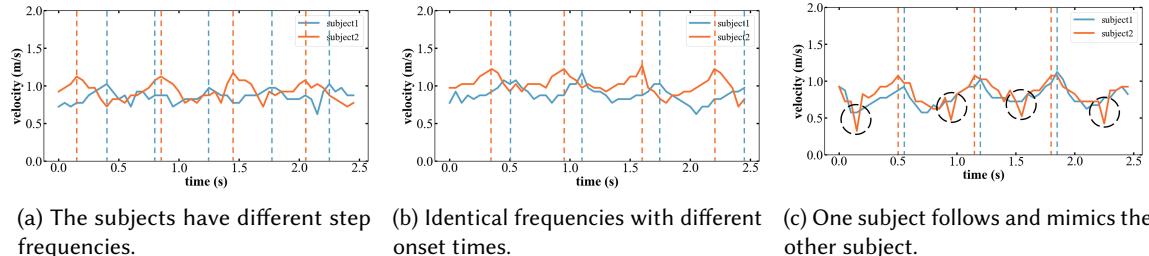


Fig. 8. The torso velocity over time for two subjects during the same time interval. The dashed line represents the onset of each step, indicated by the peak velocity.

calculates the cumulative intensity distribution for each time slice:

$$P(v, t) = \frac{\sum_{v=v_{min}}^v y(v, t)}{\sum_{v=v_{min}}^{v_{max}} y(v, t)}, \quad (10)$$

where $P(v, t)$ is the percentile in velocity at v and time at t , and $y(v, t)$ is the intensity in velocity at v and time at t of the micro-Doppler spectrum. The algorithm considers the velocity $v(t)$ to be torso velocity when $P(v, t) = 50\%$. The percentile algorithm performs well in calculating the torso velocities of subjects. However, as to the ghost targets, which are always occluded by subjects/objects, their torso velocity is no longer the median of the velocities. This results in the torso velocity extracted using the percentile algorithm leaning more towards the unobstructed parts, leading to incorrect estimation of torso velocity.

We, however, observe that *the torso of the target occupies the highest radar cross-section (RCS)*. No matter whether it's a real subject or a ghost, the intensity of the torso in the heatmap is always the highest. Inspired by this observation, We devise a simple yet effective *torso velocity estimation* method. We first extract the top N points with the highest intensity for each frame in the RDStack, which are identified as the torso points:

$$RDR'_{r,v} = \begin{cases} RDR_{r,v}, & \text{if point at } (r, v) \text{ in } topN(RDR) \\ 0, & \text{otherwise} \end{cases}, \quad (11)$$

where RDR is a frame in the RDStack. Then, we sum the RDS along the range dimension to obtain the micro-Doppler spectrum corresponding to torso motions. We calculate *the percentile of torso velocity* using the micro-Doppler spectrum as shown in Eq. 10. When the velocity $v(t)$ satisfies $P(v, t) = 50\%$, we designate $v(t)$ as the *torso velocity time series*. Fig. 9(a) illustrates the torso velocity time series for a subject and the associated ghost target by RDGait. Fig. 9(b) illustrates the torso velocity time series for the subject and ghost by Percentile method.

• Velocity change pattern corelation. After obtaining the torso velocity series for each target, we calculate the correlation coefficients between them in pairs. There are various ways of computing correlation coefficients between two series; we have considered three popular ways in our system – Pearson correlation coefficient [7], Spearman correlation coefficient [33], and Kendall's Tau correlation coefficient [1].

Suppose we have two velocity series $V_a = (v_{a,1}, v_{a,2}, \dots, v_{a,N})$ and $V_b = (v_{b,1}, v_{b,2}, \dots, v_{b,N})$. The Pearson correlation coefficient is a statistical measure used to assess the strength of a linear relationship between two continuous series calculated as:

$$r_{ab}(0, N) = \frac{\sum_{i=0}^N (V_{a,i} - \bar{V}_a)(V_{b,i} - \bar{V}_b)}{\sqrt{\sum_{i=0}^N (V_{a,i} - \bar{V}_a)^2 \sum_{i=0}^N (V_{b,i} - \bar{V}_b)^2}}. \quad (12)$$

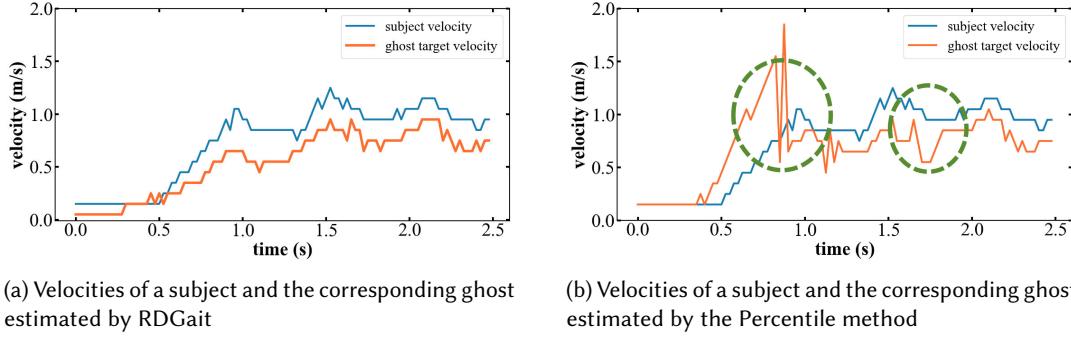


Fig. 9. Results of velocity estimation by RDGait (a) and Percentile (b). In (a), we observe that the two velocities exhibit a high correlation in their change patterns. The same level of correlation is not observed in (b), in which we use the green circle to mark the uncorrelated areas due to unstable ghost velocity estimation.

Algorithm 2: Ghost Detection

```

Input: Targets list, RDStack for each target
1 for each  $target_i \in \text{targets list}$ :
    // Step 1: Torso Velocity Estimation
2   for each  $RDR_t \in \text{RDStack}$ :
3     forall  $(r, v) \in RDR_t$  do
4       if  $(r, v) \notin \text{TopN}(RDR_t)$ :
5          $RDR_{r,v}^t = 0$ ; /* Eq. 11 */
6        $MD_v^t = \sum_r RDR_{r,v}^t; E_{all}^t = \sum_v MD_v^t;$ 
7       for  $v' = 0, 1, \dots, v_m$ :
8         if  $\sum_{v=0}^v MD_{v'}^t > E_{all}^t / 2$ :
9            $v^t = v'$ ; break; /* Eq. 10 */
10  Combine any two targets to form target pairs, denoted as set  $\{(target_i, target_j)\}$ 
11 for each pair  $(i, j)$  in pairs:
    // Step 2: Correlation coefficient Calculation
12   for  $k = 0, 1, \dots, M - 1$ :
13      $r_k = \text{Pearson}(v_i[k : N + k], v_j[k : N + k])$ ; /* Eq. 12 */
14      $r = \max\{r_0, r_1, \dots, r_{M-1}\}$ ; /* Eq. 13 */
15     if  $r > r_{th}$ :
16       if  $range_i > range_j$ :
17         ghost_list.append(range_i);
18       else:
19         ghost_list.append(range_j);
20 return ghost_list

```

The Spearman correlation coefficient and Kendall's tau correlation coefficient are both statistical measures used to assess the rank correlation (ordinal relationship) between two sequences.

For two sequences, to reduce the impact of velocity measurement errors, we calculate M correlation coefficients using a sliding window approach with a sliding step of 1, and choosing the maximum value as the correlation coefficient:

$$r_{ab}^* = \max\{r_{ab}(0, N), r_{ab}(1, N+1), \dots, r_{ab}(M-1, N+M-1)\}. \quad (13)$$

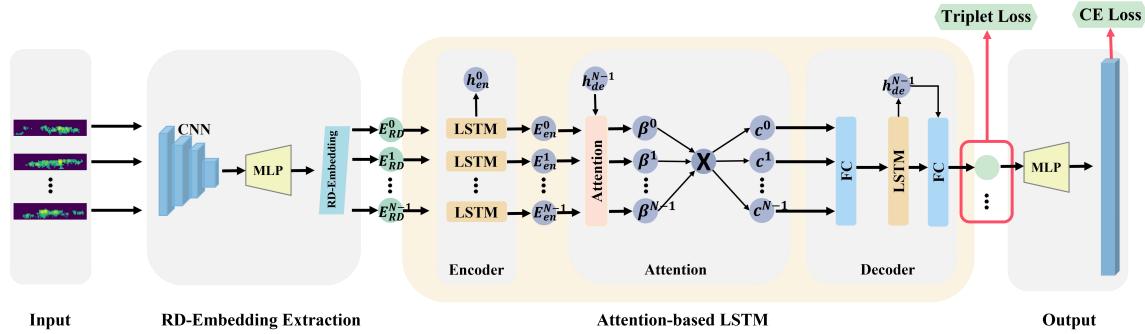


Fig. 10. The overview of our attention-based Gait Recognition Network.

A coefficient of 1 indicates a positive correlation, a coefficient of -1 indicates a negative correlation, and the larger the absolute value of the correlation, the more similar the series is. We form a correlation matrix by combining pairwise correlation coefficients. When the absolute value of the correlation between two targets is greater than a preset threshold value r_{th} , one of them must be the ghost target. According to Eq. 3, the range of the ghost will be larger than that of the subject. Therefore, we consider the one with the larger range as a ghost target. The setting of the threshold requires a balance between precision and recall, which will be discussed in Sec. 7. After obtaining a ghost list, we clear the RDStack of ghost targets.

4.2 Attention Based Gait Recognition

In our data preparation phase, we filter out background and dynamic ghost targets from the original RD Heatmap and obtain a range-Doppler stack (RDStack) for each subject. In this section, we propose an attention-based gait recognition network that takes the RDS data as input. Moreover, we carefully take into account the need for rapid deployments across diverse environments. We attempt to accomplish this through a pretrain-finetune approach. Our network architecture is illustrated in Fig. 10.

4.2.1 Samples preparation. For each subject, we have obtained the RDStack, which will be fed into the network as input. Here, we need to discretize the data into fixed-size samples. Also, we aim for each sample to start from the same gait phase (e.g., the moment when the torso velocity reaches its maximum or minimum). Specifically, we use the previously estimated torso velocity series and extract peaks and valleys. We select the subsequent 20 frames following the peaks and valleys to form a gait sample of the size $20 \times 11 \times 255$, where 11 and 255 are the dimensions for range and velocity, respectively.

4.2.2 Network design. In this section, we design an attention-based network. The network consists of three modules: (1) the RD-feature extraction module that is used to extract range-velocity features for each frame, (2) the attention-based LSTM that is used to extract temporal features between frames, and the output module that employs an MLP for classification.

RD Feature extraction. This module extracts the range-Doppler features of each frame in the gait sample. In detail, the feature extraction module consists of four convolutional layers with a kernel size of 2×2 , and adopts Leaky ReLU as the activation function. Each convolutional layer is followed by a 2×2 max pooling layer. Each gait sample consists of a continuous sequence of N frames. We input the N frames in chronological order into the RD-feature module, obtaining N embeddings of 128 dimensions each. We refer to these embeddings as *range-Doppler embeddings* (RD embeddings), marked as $E_{RD}^t \in \mathcal{R}^{128 \times 1}$. Each RD embedding only contains the

features of the current frame, which represents the joint distribution of range and velocity at a specific moment during the motions.

Attention-based LSTM. To extract the temporal features of gait, we design an RNN module. The output of a traditional RNN depends on the current input and the hidden state from the previous time step. For the gaits, each time step is different and should have different weights. Therefore, we use an attention-based RNN model, where the weights are dynamically computed through an attention mechanism. Inspired by [40], we devise an attention-based LSTM network for extracting temporal gait features. Our model consists of three modules: an encoder, an attention module, and a decoder:

- Encoder. We input the N frames of RD embeddings into the encoder, re-encoding them as $E_{en}^i, i \in \{0, 1, \dots, N - 1\}$. We use LSTM as the encoder, which can be represented as:

$$E_{en}^i = f_{LSTM}(E_{en}^{i-1}, E_{RD}^i), \quad (14)$$

where f_{LSTM} denotes the LSTM layer.

- Attention. We first use fully connected layers and the hidden states from the encoder to generate intermediate attention weights l :

$$l^t = W_2 \tanh(W_1 [E_{en}; h_{de}^{N-1}]), \quad (15)$$

where h_{de}^{N-1} is the decoder hidden state, the square bracket represents concatenation, and W_1 and W_2 are the weights to be optimized. Then, the intermediate attention weights undergo a softmax calculation to obtain the normalized attention weights β^t . The attention weight β^t represents the importance of the i -th encoder hidden state for the recognition. Thereafter, the context vector can be computed as:

$$c^t = \sum_{j=0}^{N-1} \beta^t E_{en}. \quad (16)$$

- Decoder. The decoder module processes the context vector through an LSTM and a fully connected layer to generate a 128-dimensional embedding, which we refer to as gait embedding. Finally, we employ an MLP to classify the gait embeddings.

Loss Function. During the training stage, the loss function consists of two parts: the first part is the cross-entropy for user identification, and the second part is the triplet loss representing embedding separability. The purpose of triplet loss is to make the embeddings of samples from the same class closer on the feature map and those from different classes farther apart. Triplet loss requires three samples: an anchor sample, a positive sample (similar to the anchor), and a negative sample (dissimilar to the anchor). Its computation is typically expressed as:

$$L = L_{CE} + \lambda L_{Triplet} = L_{CE} + \lambda \max(d(A, P) - d(A, N) + \alpha, 0), \quad (17)$$

where $d(A, P)$ represents the distance between the anchor and positive samples, $d(A, N)$ represents the distance between the anchor and negative samples, α is a predefined margin parameter, λ is used to balance the magnitudes of the two losses.

4.2.3 Rapid employment. We attempt to achieve rapid employment through a pretrain-finetune approach. The initial pretraining phase involves the collection of a substantial dataset, serving as the supporting set for our model. This dataset diversity ensures a robust pertaining model. Subsequently, the trained model undergoes deployment across diverse environments. For instance, deploying RDGait in a more confined space like a smaller office setting. To streamline and optimize the deployment of RDGait, we meticulously follow a structured sequence of steps:

- *Radar Placement:* We focus on the precise positioning and elevation adjustments of the radar. This step is tailored to the specific conditions within the room, ensuring optimal performance.

- *User Registration*: To fine-tune our model, we conduct user registration, collecting a small amount of gait data from users. This step is pivotal for the system to adapt effectively to the unique characteristics of individuals in the new environment.
- *Model Fine-Tuning*: Leveraging the gathered registration data, we fine-tune the pre-trained model. This iterative process hones the model's ability to recognize and adapt to the distinct gait features encountered in the new environment.

5 implementation

We have implemented RDGait in Python. RDGait contains a data preparation and recognition network module. The data preparation module was modified based on Openradar [34], and the tracking module employed AB3DMOT [57] for human tracking. The recognition network module is implemented using PyTorch [35].

As shown in Fig. 11(a), RDGait is developed on a single-chip millimeter-wave radar, IWR6843boost [49]. We use the DCA1000 evaluation module [48] for real-time data capture and streaming for the radar sensor. The IWR6843Boost sensor was configured to use 2 transmitter antennas and 4 receive antennas. We utilize two horizontally aligned transmitting antennas (TX1 and TX2). During processing, we accumulate the RD heatmap data from eight virtual antennas into a single heatmap. In the following section, we employ MVDR beamforming with eight antennas for angle estimation, generating a point cloud as a baseline. The start and end frequencies were set to 60GHz and 64GHz, respectively. And the Frequency Slope was set to 119.975MHz/ μ s. The numAdcSamples was set to 256, and the numChirpsPerLoop was set to 255. With the above configuration, the radar system has a range resolution of 5.9cm and a maximum unambiguous range of 15m. The radial velocity that RDGait can measure is 6.8m/s, with a velocity resolution of 0.05m/s. The frame rate of the radar was set at 20 frames per second (fps). While running, the radar is connected to a laptop with an AMD Ryzen 9 4900H CPU and an NVIDIA GeForce RTX 2060 GPU. The deep learning model was trained on an NVIDIA GeForce RTX 3090 GPU.

6 Evaluation of Gait Identification Performance

We report the detailed evaluation results below.

6.1 Dataset

We conducted a comprehensive evaluation of RDGait through laboratory studies with volunteers. Our studies were approved by the Institutional Review Board (IRB) of our institution. A total of 125 volunteers were invited to generate the gait dataset¹, refer as the supporting dataset, including 66 males and 59 females, with ages ranging from 6 to 63 years. The average age of volunteers is 28.6 years, with a standard deviation of 13.06 years. We established two sessions for the supporting dataset, namely the Crowded-Office session and the Spacious-Corridor session.

Session I: Crowded-Office. The first test field is an office scenario ($3m \times 10m$), as shown in Fig. 11(b), wherein the radar is placed in the middle of the hallway at an elevation of 1.0m. There are various office furniture items, including desks, chairs, benches, and partitions, which easily generate many ghost targets.

Session II: Spacious-Corridor. The second test field is a corridor scenario ($6m \times 13m$), as shown in Fig. 11(c), wherein the radar is placed in the middle of the corridor at an elevation of 1.0m. In this scenario, volunteers have a broader walking area.

Volunteers were instructed to walk back and forth within the designated area, first walking towards the radar and then walking away, as illustrated in Fig. 11. In the experiment, we only specified the walking direction for volunteers and did not require volunteers to follow a predefined trajectory. In each session, volunteers went

¹<https://github.com/DQ-WDQ/RDGait>

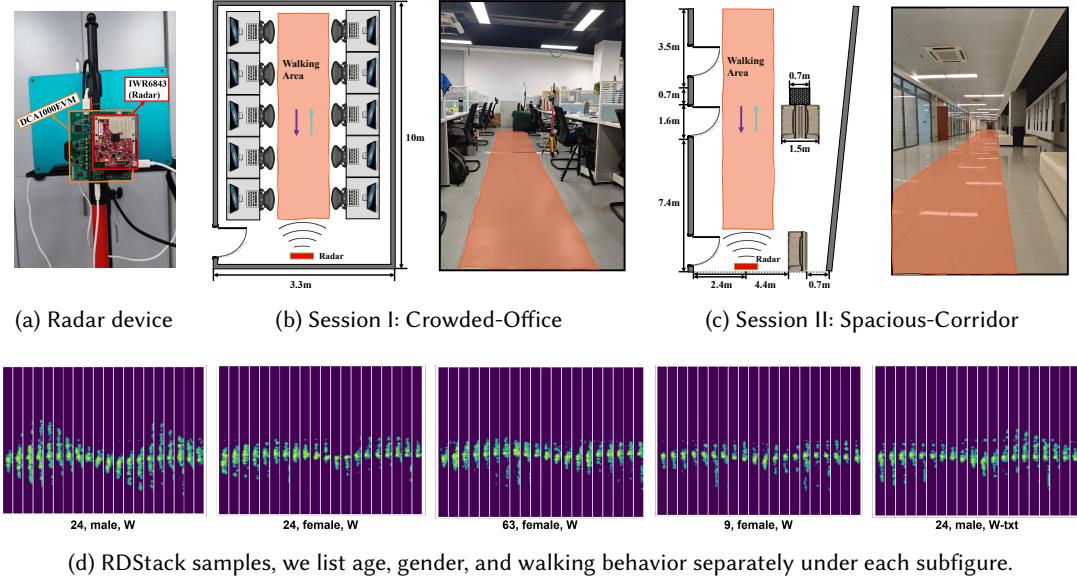


Fig. 11. Experimental equipment and scenarios for supporting dataset. The arrows indicate the walking direction and the orange area in the real-life image represents the walking area. With this guidance, the volunteers were free to choose their walking trajectories.

through five distinct walking behaviors: normal walking (W), walking with a backpack (W-BP), walking with a handbag (W-HB), walking while calling (W-Call), and walking while texting (W-txt). In particular, the fifth behavior, W-txt was excluded for seniors and children due to behavioral habits. Each volunteer performed these behaviors in both scenarios for 2 minutes, contributing to a total data collection period of 20 minutes per volunteer. The collective data from all volunteers amounted to 2500 minutes (approximately 42 hours).

We preprocess each frame's range-Doppler heatmap using the proposed method in Sec. 4, segmenting them into samples. Each sample consists of 1 second (20 frames) of radar data. To compare with other gait recognition methods, besides RDStack, we also preprocess the data into three other formats. Some RDStack samples of different volunteers with different ages, genders, and walking behaviors are shown in Fig. 11(d). In the figure, we can observe differences among individuals in terms of velocity, cadence, arm swing, and other aspects. When the subject walks while typing, the motion caused by arm movement disappears from the graph. The first format is the micro-Doppler spectrogram (MD for short), which is derived by summing RDStacks along the range dimension, resulting in samples sized at 20×255 . The second format, the range-time spectrogram (RT for short, utilized in MU-ID [61]), comprises samples sized at 20×70 . Finally, the last format is the point cloud (PC for short, employed in mmGaitNet [31]), where each frame is sized at $N \times 4 \times 20$, with N representing the number of extracted points. Each point has four attributes: (x, y, v, snr) , where $x = r \cdot \cos\theta$, $y = r \cdot \sin\theta$, r is a range, snr is signal-to-noise ratio, and θ is the angle of the point, estimated by Minimum Variance Distortionless Response (MVDR) beamforming.

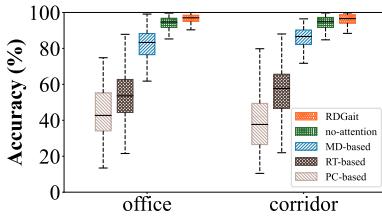


Fig. 12. Overall gait recognition accuracy results over a total of 125 users.

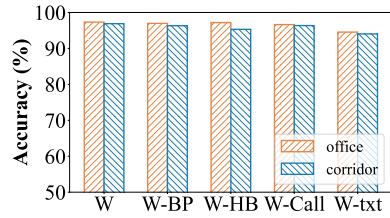


Fig. 13. Gait recognition accuracy across different walking behaviors.

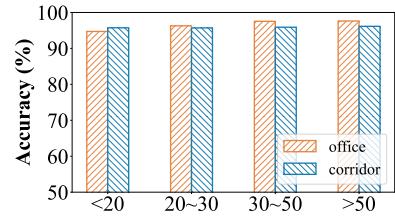


Fig. 14. Gait recognition accuracy across different age groups.

6.2 Evaluation metrics

In this set of experiments, we are trying to train and test a relatively generic model using a large supporting dataset with volunteers and walking behavior variations. Then in the next phase, such a model can be easily fine-tuned to fit different deployment settings. Toward this goal, we adopt the following training plan and evaluation metrics. We evaluate the overall performance of RDGait system by k -fold cross-validation. In the pretraining phase, the supporting set is split into k smaller sections. Each section is used for validation, and the remaining $k - 1$ sections are used for training, where $k = 5$ in our experiment.

Evaluation Metrics. In our evaluation, we mainly report the following metrics:

- **Accuracy.** The ratio between the number of correct gait predictions and the total number of predictions.
- **Confusion matrix.** The element C_{ij} in the confusion matrix C , located in the i -th row and j -th column, represents the proportion of samples that the model predicted as belonging to class j among the samples belonging to class i .

6.3 Gait Identification Performance Evaluation

Baselines. To evaluate RDGait, We selected four baselines to process four different data formats:

- **PC-based.** We use a network structure of mmGaitNet [31] for the point cloud-based method. This architecture utilizes a layer of ResNet18 [18] to extract features for each attribute of the point cloud individually. Subsequently, the features of each attribute are concatenated and fused using spatiotemporal convolution. In cases where the number of points in the real point cloud is less than 128, mmGaitNet padded the points by multi-copying them as padding points. The input dimension is $4 \times 20 \times 128$.
- **RT-based.** We adopted a network structure of MU-ID [61] for the range-time spectrogram-based method. This architecture employs four convolutional layers and one fully connected layer to extract gait features and perform classification from the spectrogram. To accommodate the differences in input dimensions, we adjusted the padding of each convolutional layer to 1 and the last max-pooling operation to 2×2 .
- **MD-based.** For the method based on the micro-Doppler spectrogram, we employed the network structure of IDRAd [52]. IDRAd utilizes a CNN to extract gait features and perform classification from the micro-Doppler spectrogram. In our adaptation, we remove a max-pooling layer and add a convolutional layer to improve the performance.
- **No-attention.** To evaluate the effectiveness of the attention-based LSTM in RDGait, we replaced it with a classic LSTM.

Overall identification accuracy. Fig. 12 shows the overall user identification accuracy for two scenarios across different methods. The average accuracy of RDGait in the two scenarios is 96.5% and 95.8%, respectively, delivering the best performance among all baselines. As a comparison, the identification accuracy of other methods is as

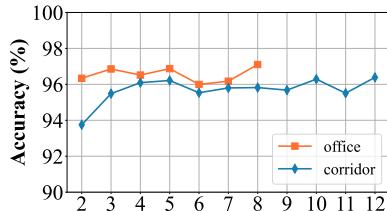


Fig. 15. Gait identification accuracy across different radar distances.

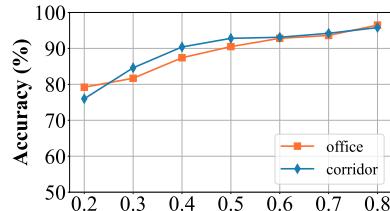


Fig. 16. Gait identification accuracy across different training size.

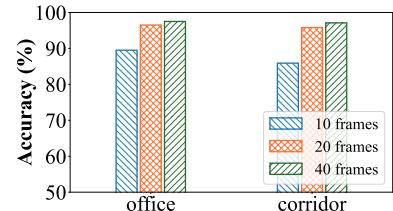


Fig. 17. Gait identification accuracy across different sample duration.

follows: PC-based (44.8%, 39.3%), RT-based (54.5%, 56.3%), and MD-based (82.0%, 85.8%). This suggests that RDGait is more proficient at extracting discriminative gait features in scenarios involving a larger number of people. The range information and velocity information in a person’s gait are coupled with each other. Compared to RDStack, other data formats lack certain gait information. Point cloud data is relatively sparse, range-time spectrograms lack velocity information, and micro-Doppler spectrograms lack range information.

Furthermore, the no-attention version of RDGait performs much better than other algorithms, but still lower than our full-fledged RDGait, with an average accuracy of 94.0% and 93.8% in the two scenarios. This demonstrates that the attention LSTM module can effectively extract the temporal features of gait compared to the plain LSTM. Additionally, we find that the identification accuracy of our algorithm is slightly lower in corridor scenes compared to office scenes. This could be attributed to two factors: corridors are longer (12m) compared to office spaces (8m), and they offer more spaciousness, providing volunteers with a greater variety of walking routes.

Gait identification accuracy across different walking behaviors. Fig. 13 reports the user identification accuracy based on pre-designed walking behaviors. Among all the walking behaviors, normal walking (W) achieved the highest recognition accuracy in both scenarios, reaching 97.3% and 96.8%, respectively. In contrast, walking while texting (W-txt) demonstrates the lowest identification accuracy at 94.5% and 94.0%, respectively. While texting, individuals tend to experience dispersed attention, reduced arm swing, altered walking rhythm, and other factors that can disrupt the normal walking pattern, thereby resulting in slightly deformed gaits. However, we note that across various walking behaviors, our algorithm can deliver satisfactory identification accuracy, which shows its robustness across walking behaviors.

Gait identification accuracy across different age groups. To assess the identification performance of RDGait across individuals of varying ages, we categorized the volunteers into four age groups: under 20, 20-30, 30-50, and over 50. The number of individuals in each group is 10, 88, 11, and 16, respectively. The results of volunteers in these four age groups in two scenarios are shown in Fig. 14. Among them, the identification accuracy is highest for volunteers aged over 50 in two scenarios, reaching 97.6% and 96.2%, whereas the accuracy is lowest in the age group under 20, at 94.7% and 95.7%. As shown in Fig. 19, we found that three children under the age of 12 had accuracy rates of 87.3%, 90.9%, and 90.3%, respectively, which were notably below the average level. However, the recognition accuracy for the remaining six children under 12 was around 95%, aligned with the average level. This is attributed to the fact that lively and active children often exhibit more playful behaviors while walking, leading to an increased instability in their gait. Overall, the recognition accuracy of RDGait is high across different age groups, indicating the robustness of RDGait in extracting gait features across various age ranges.

Gait identification accuracy across different radar distances. To evaluate the robustness of RDGait for the distance between radar and subjects, we partitioned radar samples into various distance intervals, ranging from [0 – 2m] to [11 – 12m]. Subsequently, we evaluated the recognition accuracy for each distance interval. Fig. 15 shows the user identification accuracy in two scenarios at varying distance intervals. Different distances

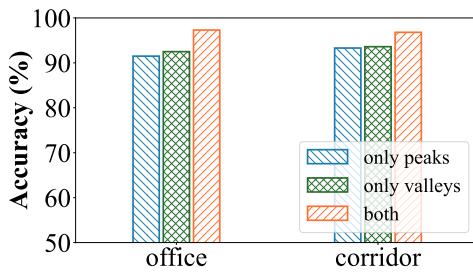


Fig. 18. Gait identification accuracy across different sample start times.

-	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.98	0.02	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.01	0.91	0.03	0.00	0.00	0.01	0.00	0.00
4	0.00	0.00	0.01	0.94	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.83	0.00	0.01	0.00	0.07
6	0.00	0.00	0.01	0.00	0.00	0.99	0.02	0.00	0.02
7	0.00	0.00	0.01	0.00	0.02	0.00	0.96	0.01	0.01
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.97	0.00
9	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.99

Fig. 19. Confusion matrix for children under 12 years old.

influence the intensity of the reflected signals. As the distance increases, the signal intensity decreases. The experimental results demonstrate that RDGait still achieves an accuracy of over 95% at a relatively far distance, indicating the robustness of RDGait across different distances.

However, we observe that the accuracy in the [0 – 2m] interval in the Spacious-Corridor scenario is relatively low, standing at 93.8%. Upon examining this data, we discovered that when users were in close proximity to the radar, such as at distances less than 1.0m, the radar, due to limitations in its field of view, couldn't capture the complete walking pattern of individuals.

Gait identification accuracy across different training size. To investigate the effect of training size in the pretraining phase, we selected 20% of samples as the test set, varying the training data size from 20% to 80% of samples, in increments of 10%. The recognition accuracy at different training set proportions is shown in Fig. 16. The results show that as the training set proportion increases from 0.2 to 0.5, the accuracy grows rapidly, surpassing 90%. Beyond a training set proportion of 0.5, the growth tends to stabilize. With a training set proportion of 0.5, RDGait has already learned a substantial portion of the gait features.

Gait identification accuracy across different sample duration. To investigate the impact of sample duration on the performance of RDGait, we have conducted experiments using sample sizes of 10, 20, and 40 frames, corresponding to sample duration of 0.5s, 1.0s, and 2.0s, respectively. The results are shown in Fig. 17. When the frame count drops to 10 frames (0.5s), the accuracy in the two scenarios decreases to 89.5% and 85.9%, respectively. When the frame count is increased to 40 frames (2s), the accuracy rises to 97.5% and 97.1%, respectively. We observed that extending the length of each sample from 1s to 2s led to a slight improvement in accuracy (around 1%). However, this enhancement comes at the cost of increased training time and a longer system response time. In RDGait, we have opted for a duration of 1.0s for each sample, striking a balance between accuracy and response time.

Gait identification accuracy across different sample start times. We choose different time points as the starting point for a gait sample, including the points of maximum torso velocity (referred to as *peaks*) and the points of minimum torso velocity (referred to as *valleys*). We assessed the accuracy of using either peaks or valleys as the starting point for gait samples, as well as the accuracy when using both, as depicted in Fig. 18. The results indicate that the accuracy achieved by using peaks alone or valleys alone is comparable. However, utilizing both peaks and valleys leads to an enhancement of 3-5 percentage points compared to relying on either one individually. This suggests that generating samples with different starting points increases the diversity of the dataset, thereby enhancing the robustness of feature extraction.

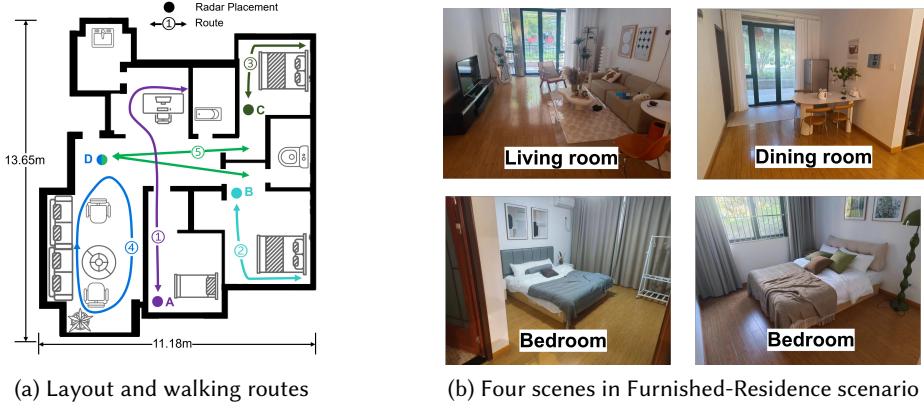


Fig. 20. Experimental equipment and scenarios for Furnished-Residence scenario.

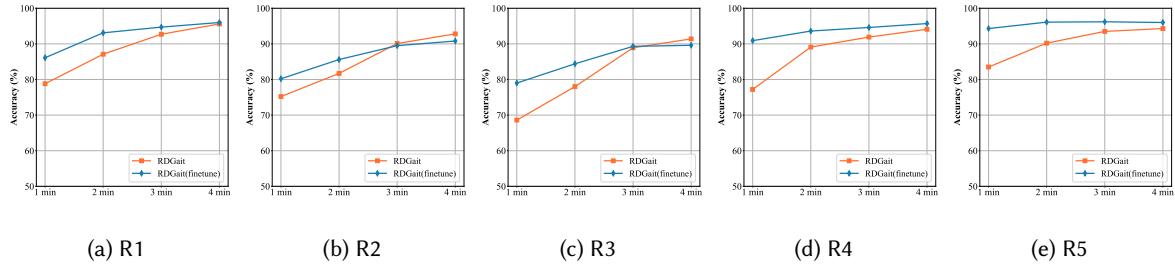


Fig. 21. Gait recognition accuracy of five routes across different training sizes.

6.4 Deployment of RDGait in Different Environments

Now that we have trained the RDGait model and demonstrated its effectiveness in gait recognition, we next discuss how we can deploy the trained system in different environments. For example, we can deploy a RDGait system in a much smaller office or home. To facilitate the deployment of RDGait, we need to go through the following steps: (1) *radar placement*, adjusting the positioning and elevation of the radar based on specific situations in the room, (2) *user registration*, collecting a small amount of user gait data, (3) *model fine-tuning*, using the registration data to fine-tune the previously trained model. We take the viewpoint that a ready and smooth deployment process is critically dependent on a quick and easy registration step. That is, the less registration data we require, the smoother the deployment is.

Experiment setup. Here, we test how easily RDGait can adapt to new deployments in different scenarios. To thoroughly evaluate this aspect, we chose different deployment environments:

6.4.1 Deployment in Furnished-Residence scenario. To assess the real-life performance of RDGait in household scenarios, we employed 11 volunteers (6 male, 5 female) to collect gait data in a house measuring $13.65m \times 11.18m$, furnished with various pieces of furniture. The layout of the rooms and the walking routes of the volunteers are illustrated in Fig. 20. The four dots in the figure represent the positions of radars, and the five lines with

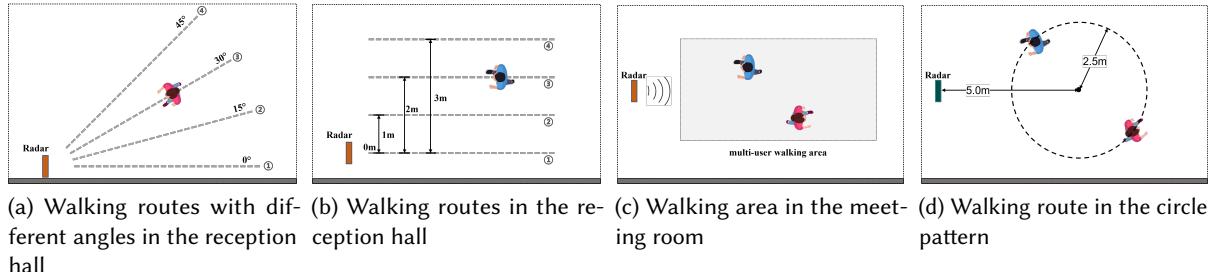


Fig. 22. Walking routes and areas in deployment scenarios.

arrows represent the walking routes (marked as R1, R2, R3, R4, and R5). The radar is placed at a height of 1.2 meters. Routes 1, 2, 3, and 4 experience walking direction changes and obstruction by furniture. R1 and R4 are obstructed by tables and chairs, while R2 and R3 encounter obstructions from beds and include a 90-degree turn. Each volunteer walked back and forth along each route for 5 minutes, resulting in 275 minutes of radar data.

We use data from 1, 2, 3, and 4 minutes for registration respectively, reserving the remaining data for testing. The identification accuracy of each route is shown in Fig. 22. The result shows that when using 4 minutes of the data for registration, the identification accuracies for five routes are 96.0%, 92.8%, 91.4%, 95.7%, and 96.0%, all exceeding 90%. Conversely, with only 1 minute of data for registration, the identification accuracies for the five routes are 86.1%, 80.2%, 79.0%, 90.9%, and 94.3%. We observed that the network pre-trained on the supporting dataset exhibited an average increase of 8.4% in identification accuracy over the retrained network when utilizing 1-minute registration data. The identification accuracy for R2 and R3 was lower compared to the other routes, suggesting that variations in velocity direction and occlusion can impact the extraction of gait features in RDGait. Furthermore, when using 4-minute registration data, the gait recognition accuracy of RDGait (fine-tuned) was slightly lower than that of RDGait. This is because the data with significant obstruction and turns did not appear in the supporting dataset, indicating the importance of the richness of the supporting dataset.

6.4.2 Deployment for different environmental parameters. To test the impact of different environmental parameters on RDGait, we selected two scenarios:

- **Session I: Reception-hall deployment scenario.** The first deployed field is a reception hall ($20m \times 20m$). To test different radar placements, we positioned the radar along the walls. We designed different walking routes in this scenario.
- **Session II: Meeting-room deployment scenario.** The second deployed field is a meeting room ($5m \times 10m$). There are two rows of long tables, accompanied by several chairs positioned on both sides. We positioned the radar at the end of the aisle.

In addition, we also varied the placement of the radar as well as the walking patterns of the users, illustrated in Fig. 22. More details will be presented in each specific experiment below.

To evaluate the recognition performance of RDGait in diverse deployment environments, we recruited 10 additional volunteers who were not included in the supporting dataset. In this set of experiments, volunteers walked for 2 minutes in each setting. We used data from one minute for training and the remaining minute for testing. In the following experiments, we employed RDGait without retrain as the baseline, marked as RDGait, and used RDGait with a pre-trained model on the supporting set as the test method, marked as RDGait (**fine-tune**).

Gait identification accuracy across different radar angles. To evaluate the impact of angles on RDGait, we had volunteers walk directly toward the radar at angles of 0° , 15° , 30° , and 45° in the open-hall deployment

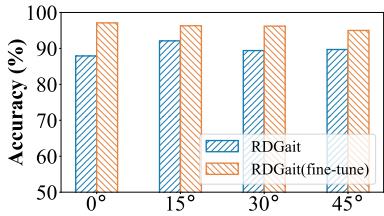


Fig. 23. Gait identification accuracy across different radar angles.

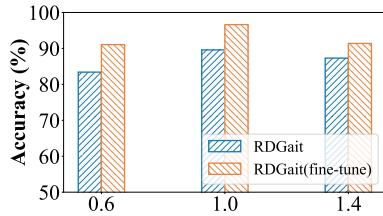


Fig. 24. Gait identification accuracy across different radar heights.

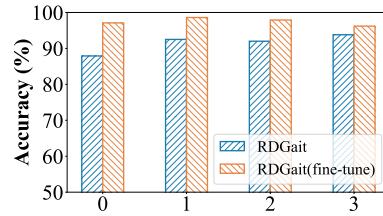


Fig. 25. Gait identification accuracy across different routes.

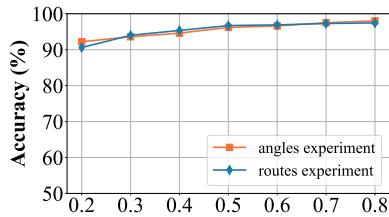


Fig. 26. Gait identification accuracy for different registration data sizes for hall deployment.

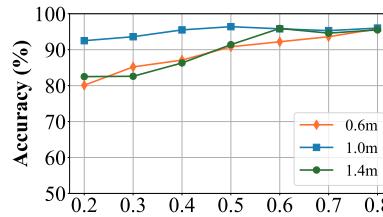


Fig. 27. Gait identification accuracy for different registration data sizes for meeting-room deployment.

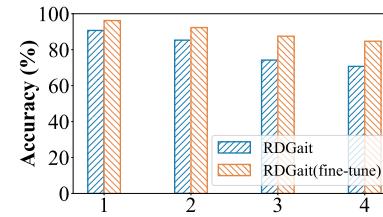


Fig. 28. Gait identification accuracy when multiple users are present. The accuracy degrades with the number of coexisting users.

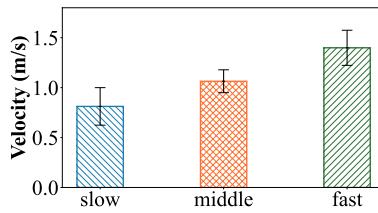


Fig. 29. Velocity distribution.

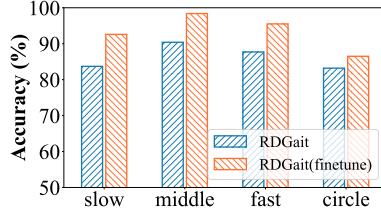


Fig. 30. Gait identification accuracy across different velocities.

scenario, as shown in Fig. 22(a). Influenced by the directionality of the radar beam and the presence of sidelobes, the energy of radar signals varies at different angles, attenuating significantly as the angles increase. Fig. 23 shows the recognition accuracy of RDGait under different walking angles. The results indicate that as the angle increases, the RDGait recognition accuracy slightly decreases, dropping from 97.1% to 95.0%. With just 1 minute of training data to fine-tune, RDGait achieved recognition accuracy of over 95% at a 45-degree angle, showcasing its robustness across angles.

Gait identification accuracy across radar elevation. To test the impact of radar elevations on RDGait, as shown in Fig. 22(c), we had volunteers walking back and forth within the monitoring area in the meeting room deployment scenario. We set the radar on elevations of 0.6m, 1.0m, and 1.4m, respectively. Fig. 24 demonstrates the recognition accuracy of RDGait under different radar elevations. The results of RDGait at three different radar elevations are 91.1%, 96.6%, and 91.4%, respectively. Differences in radar installation elevations result in variations in both distance and velocity measurements for human walking, resulting in notable distinctions in the gait features. As the supporting set was collected with the radar elevation set to 1.0m, the pre-trained model learned more gait patterns at this elevation. Therefore, the accuracy is highest at this specific elevation. However, when using the pre-trained model, there is still an improvement of 4.1% at the elevation of 0.6m and 3.0% at the

elevation of 1.4m. This indicates that we have also learned features in the pre-trained model independent of radar elevation.

Gait identification accuracy across different speeds. To evaluate the impact of speed on RDGait, We collected volunteer data at different speeds, including slow, medium, and fast speeds. We did not prescribe a specific range of speed; instead, volunteers determined their walking speed based on their own experience. According to measurements, the average speeds of volunteers during slow, medium, and fast walking were 0.81 m/s, 1.06 m/s, and 1.40 m/s, respectively. The distribution of walking speeds among volunteers is shown in Fig. 29. To assess the impact of changes in speed direction on RDGait, we designed an experiment where volunteers walked clockwise and counterclockwise in a circular route, as depicted in Fig. 25(d), for 2 minutes each. Fig. 30 demonstrates the identification accuracy of RDGait under different speeds. The results of RDGait at four different walking patterns are 92.6%, 98.4%, 95.5%, and 86.5%. The result shows that the gait identification accuracy for walking in circles is significantly lower than the other three types. This is due to the constant change in speed direction during circular motion, while millimeter waves can only capture velocity information in one direction, making it challenging to extract globally consistent gait features.

Gait identification accuracy across different routes. To evaluate the influence of walking routes on RDGait, as depicted in Fig. 22(b), volunteers were instructed to walk along four routes parallel to the radar boresight line in the open-hall deployment scenario. The distances from the radar boresight line were set at 0m, 1m, 2m, and 3m, respectively. The velocity measured by radar is the Doppler velocity, referring to the velocity toward the radar, which is influenced by the angle between the direction of the velocity and the line of sight. When users walk in a direction not aligned with the line of sight, the radar-measured velocity changes with the movement. This leads to inconsistency in the extracted gait features. Fig. 25 demonstrates the recognition accuracy of RDGait under different walking routes. The results on the four routes are 97.1%, 98.6%, 97.9%, and 96.2%, showing RDGait's robustness across routes.

Gait identification accuracy across different training size. To investigate the effect of training size in the register phase, we selected 20% of samples as the test set, varying the training data size from 20% to 80% of samples, in increments of 10%. Fig. 26 shows the recognition accuracy under different proportions of training sets in both angle and routes experiments. The results indicate that RDGait achieved an accuracy exceeding 90% with only 20% of the data, and the accuracy continues to improve as the proportion of the training set increases. Fig. 27 shows the recognition accuracy under different proportions of training sets in elevation experiments. The results show that, for different radar elevations from the supporting set, RDGait can achieve a recognition accuracy of 80% even with a low training set proportion. When the training set proportion exceeds 50%, the accuracy reaches 90%.

Gait identification accuracy with multiple users. To evaluate the performance of RDGait in a multi-user scenario. We formed volunteer groups randomly, with 2, 3, and 4 individuals in each group, enabling members within each group to walk together simultaneously. For each group of volunteers, we instructed them to walk simultaneously within the radar detection area, as shown in Figure (c). Volunteers were asked to maintain an appropriate distance (0.5m) between themselves during the walking process. For each group size, we gathered data in 5 sets, each spanning 1 minute. Before the experiment, we had each volunteer walk individually for three minutes, as part of the user registration process. As shown in Fig. 28, the accuracy for single users, and groups of two, three, and four are 96.2%, 92.3%, 87.5%, and 84.7%, respectively. When the number of co-existing users increases, the recognition accuracy degrades. our current system works best when the number of users is fewer than 4.

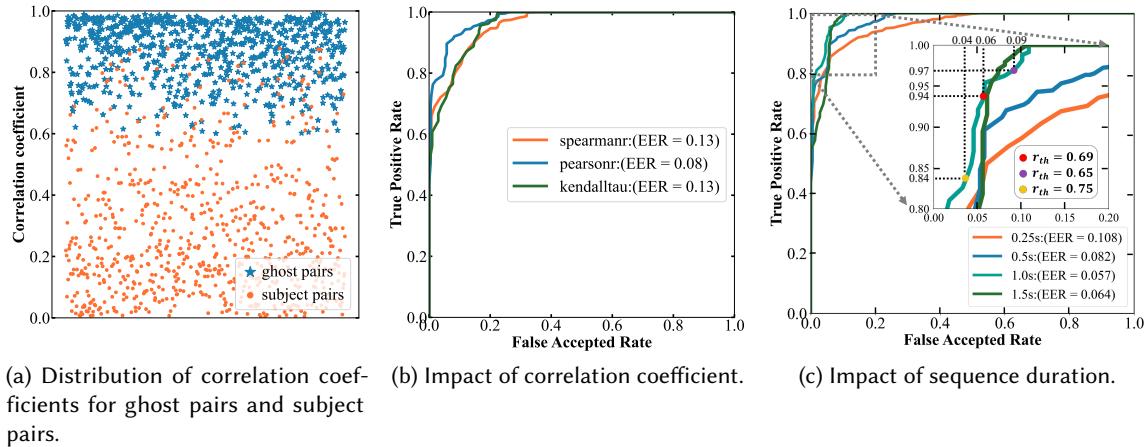


Fig. 31. Evaluation on ghost detection. (b) and (c) depict ROC curves drawn with different calculation methods and parameters.

7 Evaluation of ghost detection performance

Accurate ghost detection is a prerequisite for user identification. After completing the preprocessing, we use the algorithm designed in Sec. 4.1.2 to implement ghosting detection.

7.1 Dataset

To evaluate the ghost detection algorithm, we selected a portion of the data from the acquired dataset. Due to the relatively challenging annotation, we did not utilize the entire dataset. After preprocessing the radar echo signals, we obtain multiple target samples appearing in the same period. Each sample contains the torso velocity sequence of the target over N frames of time. Each target sample undergoes manual labeling. If the target sample belongs to a subject, we label it as s_{id} , and if the target sample is a ghost target, we label it as g_{id} . We pair the samples, combining s_{id} and g_{id} belonging to the same period and ID into a *ghost pair*. Similarly, we form a *subject pair* by pairing s_i and s_j from different IDs belonging to the same period. We consider ghost pairs as positive instances and subject pairs as negative instances. We annotated a total of 756 positive instances and 812 negative instances.

7.2 Experiment Setup

Classification method. We employed three different correlation algorithms to calculate the correlation for ghost pairs and subject pairs: the Pearson coefficient, Spearman coefficient, and Kendall Tau coefficient. For a pair, we employ a correlation method to calculate the correlation in a sliding-window manner. We use the absolute maximum value among them as the correlation coefficient for the pair. Fig. 31(a) shows the distribution of correlation coefficients for 600 positive samples (ghost pair) and 600 negative samples (subject pair), using the Pearson coefficient as the correlation method, with N set to 20.

Evaluation metrics: In our evaluation, we mainly report the following metrics:

- **False Positive Rate (FPR).** FPR represents the proportion of instances that are incorrectly classified as positive among all actual negative instances.

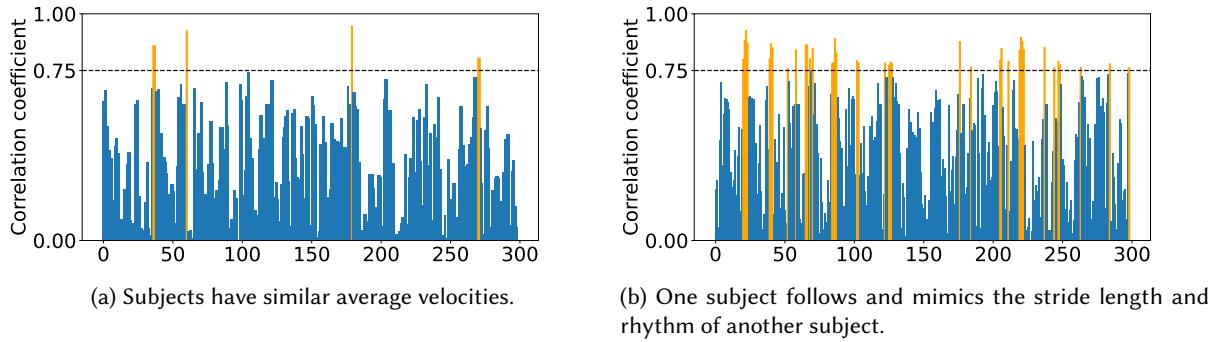


Fig. 32. The correlation in velocity change patterns between two subjects. the dashed line represents the threshold(0.75).

- **True Positive Rate (TPR).** TPR represents the proportion of instances that are correctly classified as positive among all actual positive instances.
- **Receiver Operating Characteristics (ROC).** This curve is plotted with the FPR on the x-axis and the TPR on the y-axis. For each parameter study, we varied the value of r_{th} from 0.0 to 1.0 with an increment of 0.01 and had a total of 100 threshold values to calculate FPRs and TPRs.
- **Equal Error Rate (EER).** To report the threshold-independent performance, we also consider EER which is the percentage of errors when we have $FAR = TNR$.

7.3 Ghost Detection Performance Evaluation

Ghost detection with different correlation calculation methods. Fig. 31(b) shows the ROC curves under different correlation calculation methods, which are the Pearson coefficient, Spearman coefficient, and Kendall Tau coefficient. In this experiment, we uniformly use a frame number of $N = 10$, corresponding to a duration of 0.5s for all three methods. The results indicate that the equal error rate of the Pearson coefficient (0.08) is significantly lower than that of the Spearman coefficient (0.13) and Kendall Tau correlation (0.13). In this work, we use the Pearson correlation coefficient to calculate correlations unless otherwise stated.

Ghost detection with different sequence lengths. To choose the most appropriate sequence length, we tested durations of 0.25s (5 frames), 0.5s (10 frames), 1.0s (20 frames), and 1.5s (30 frames) as candidate durations. Fig. 31(c) shows the ROC curves under different durations. The experiment indicates that when the duration $T = 1.0s$, the equal error rate is the lowest, with a value of 0.057. Following that, for $T = 1.5s$ and $T = 0.5s$, the EER values are 0.064 and 0.082, respectively. For the relatively short duration (0.25s to 0.5s), the randomness in velocity estimation takes precedence, impacting the calculation of correlation. For longer durations, the subject's movement distance cannot be treated as a "small quantity", rendering equation 9 no longer applicable. Therefore, RDGait employs $N = 20$ (1.0s).

Ghost detection with different threshold values. As shown in Fig. 31(c), when $r_{th} = 0.69$, the precision and recall values are equal, 94% for both. A decrease in r_{th} leads to lower precision and higher recall. When $r_{th} = 0.65$, we have a precision of 91% and recall of 97%; when $r_{th} = 0.75$, we have a precision of 96%, and recall is 84%.

Performance on subjects with similar velocities. To evaluate our algorithm in situations where subjects are moving at similar speeds, We collected data of two volunteers of similar build walking in the same direction at the same speed. The test is divided into two groups. In the first group, both individuals walked at the same

speed and in the same direction without any other specifications. In the second group, one individual walked behind the other, mimicking the stride and rhythm of the front person. We collected 300 pairs of subject pairs from each group and adopted the parameters from the previous section, setting the threshold (r_{th}) at 0.75. The result is shown in Fig. 32. In the first group, only 6 subject pairs are mistakenly identified as ghost pairs, resulting in only an error rate of 2% with an average correlation coefficient of 0.32. In the second group, 39 subject pairs are mistakenly identified as ghost pairs, resulting in an error rate of 13%, with an average correlation coefficient of 0.45. This indicates that when one subject follows and mimics another subject's gait, their velocity change patterns may sometimes overlap more, leading to false detections. This suggests that under normal circumstances when two people have similar speeds, it is generally unlikely for one person to be mistakenly identified as a ghost. However, when one person follows and mimics another, there is a certain probability of being mistaken as a ghost. Considering that such following-mimicking mode is not common in daily life, our algorithm rarely misidentifies the subject as a ghost.

8 Related Work

Sensor-based Gait Recognition. Various sensors can be used to extract unique features of gait, including cameras, wearable devices, WIFI, and ultrasound. For instance, vision-based methods [4, 10, 26, 47] extract silhouette and skeleton data of the body from each image frame and process them in a network to extract features for gait recognition. Vision-based methods have achieved notable advancements; however, they come with drawbacks such as sensitivity to changes in lighting conditions and privacy concerns. Li et al. [24] propose a gait sensor system based on a floor-mounted Triboelectric Nanogenerator (TENG), in which gaits are recognized through electrical signals. However, accelerometer requires user wearing, and floor-mounted TENG needs placement beneath the floor, limiting their practical applications. Wang et al. [56] utilize Channel State Information (CSI) measurements to achieve gait recognition and through-wall gait recognition. Pokkunuru et al. [39] employ CNN to extract gait features from WiFi CSI. Fan et al. [11] propose a multi-path gait recognition system based on WiFi-video cross-modal fusion by simulating ideal CSI based on video. Xu et al. [59] conduct user gait recognition by analyzing the Doppler effects of acoustics signals generated by various body parts during walking. Altaf et al. [2] employ footstep sounds to analyze user gait. However, the former WiFi-based method can only recognize one person, and the latter ultrasonic-based approach may cause harm to pets, imposing restrictions on their applications.

mmWave-based Gait Recognition. Millimeter-wave technology is attracting growing attention for its precise distance and velocity resolution, as well as its decreasing cost. Yang et al. [61] construct user-specific gait models using features in users' spatiotemporal lower limb motion. Vandersmissen et al. [52] and Pegoraro et al. [36] employ deep neural networks to extract gait features from the generated micro-Doppler spectra, achieving gait recognition for both single and multiple individuals. Meng et al. [31] and Cheng and Liu [6] deploy deep learning networks to extract spatiotemporal information from 4-D radar point clouds to perform user identification and re-identification. Cao et al. [3] achieve cross-modal vision-RF human re-identification (ReID) using rgbd camera and mmWave Radar. Different from existing solutions, we propose RDGait, which extracts spatiotemporal-velocity features from continuous range-Doppler heatmaps. To reduce false positives, we have meticulously designed a ghost detection algorithm based on velocity correlation. The above algorithms have not taken into account the impact of multipath ghosts.

Ghost detection on mmWave radar. Data-driven approaches use neural networks to directly obtain the speed and position of ghosting from range-angle heatmaps [37], range-Doppler heatmaps [37], or range-angle-Doppler heatmaps [38]. Esposito et al. [8] employ a U-net architecture to directly generate ghost-free images from tomographic images. Data-driven approaches require collecting a large amount of training data, making them costly and challenging to generalize across different environments. Model-based methods utilize prior knowledge

of the environment and the positions of radar and objects to model the propagation path of electromagnetic waves. Pegeraro et al. [36] simply treat targets outside the effective area as a ghost and can remove only a small amount. Zheng et al. [64] assume a known distance between the radar and the reflecting surface, aiming to eliminate ghosting by leveraging the distance relationship between targets and ghosting. However, this method becomes ineffective in the presence of multiple reflecting surfaces. Many methods based on automotive radar [21, 27] utilize ray tracing techniques with prior road information (or additional reflectors) to iteratively remove ghosting based on the geometric relationships between targets. Model-based methods require prior environmental knowledge, which increases the deployment cost of the system. Moreover, once the environment changes, the model becomes obsolete. Recently, there have been advancements in model-based algorithms[5, 13, 29] that do not require environmental prior. Chen et al. [5] utilize the relationships between targets and several ghosts to construct a reflection map (i.e., reflecting surfaces within the environment) while simultaneously removing ghosts. Liu et al. [29] use angle consistency to determine whether a subject is a ghost. The aforementioned approaches address the challenge of prior knowledge dependency in model-based algorithms and have demonstrated promising outcomes. However, they demand angular estimation, rendering them unsuitable for cost-effective single-channel radar systems. Feng et al. [13] discover a linear relationship between the target, 2nd-order ghost, and 3rd-order ghost on the range-Doppler heatmap, and use the three-point linear relationship to remove ghosts. The method proposed by Feng et al. [13] only utilizes range-Doppler heatmaps and does not incorporate angle information, which is also useful for single-channel radar. However, 3rd-order ghosts, having undergone multiple reflections, are more susceptible to being obscured by noise. Without 3rd-order ghosts, the algorithm would be unable to remove 2nd-order ghosts. The algorithm we propose is not based on modeling; it solely leverages the correlation between velocity change patterns resulting from the superposition of velocities induced by multiple Doppler effects. To the best of our knowledge, our work represents the first algorithm solely reliant on velocity information for ghost removal for mmWave radar.

9 Conclusion

We have presented the design and implementation of RDGait, a gait recognition system for complex indoor environments. In RDGait, we introduce a novel ghost detection method based on velocity change patterns. This method is environment-independent and requires no prior modeling of ghost targets. Additionally, we design a gait recognition network based on attention-LSTM, aiming to extract richer spatiotemporal-velocity information accurately. Furthermore, we can rapidly deploy the model through a pre-train fine-tune approach. Moving forward, we will continue to investigate the various issues that lead to wide deployment, such as environmental obstruction, and further enhance the robustness of the system.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62332016).

References

- [1] Hervé Abdi. 2007. The Kendall rank correlation coefficient. *Encyclopedia of Measurement and Statistics*. Sage, Thousand Oaks, CA (2007), 508–510.
- [2] M Umair Bin Altaf, Taras Butko, and Bieng-Hwang Juang. 2015. Acoustic gaits: Gait analysis with footstep sounds. *IEEE Transactions on Biomedical Engineering(TBME)* 62, 8 (2015), 2001–2011.
- [3] Dongjiang Cao, Ruofeng Liu, Hao Li, Shuai Wang, Wenchao Jiang, and Chris Xiaoxuan Lu. 2022. Cross vision-rf gait re-identification with low-cost rgb-d cameras and mmwave radars. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies(IMWUT)* 6, 3 (2022), 1–25.
- [4] Hanqing Chao, Kun Wang, Yiwei He, Junping Zhang, and Jianfeng Feng. 2022. GaitSet: Cross-View Gait Recognition Through Utilizing Gait As a Deep Set. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 44, 7 (2022), 3467–3478.

- [5] Weiyan Chen, Hongliu Yang, Xiaoyang Bi, Rong Zheng, Fusang Zhang, Peng Bao, Zhaoxin Chang, Xujun Ma, and Daqing Zhang. 2023. Environment-aware Multi-person Tracking in Indoor Environments with MmWave Radars. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies(IMWUT)* 7, 3 (2023), 1–29.
- [6] Yuwei Cheng and Yimin Liu. 2021. Person reidentification based on automotive radar point clouds. *IEEE Transactions on Geoscience and Remote Sensing (TGRS)* 60 (2021), 1–13.
- [7] Israel Cohen, Yiteng Huang, Jingdong Chen, Jacob Benesty, Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. *Noise reduction in speech processing* (2009), 1–4.
- [8] Giuseppe Esposito, Ilaria Catapano, Giovanni Ludeno, Francesco Soldovieri, and Gianluca Gennarelli. 2023. A Deep Learning Strategy for Multipath Ghosts Filtering via Microwave Tomography. *IEEE Transactions on Geoscience and Remote Sensing(TGRS)* 62 (2023), 1–14.
- [9] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, Vol. 96. 226–231.
- [10] Chao Fan, Yunjie Peng, Chunshui Cao, Xu Liu, Saihui Hou, Jiannan Chi, Yongzhen Huang, Qing Li, and Zhiqiang He. 2020. GaitPart: Temporal Part-Based Model for Gait Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [11] Jinmeng Fan, Hao Zhou, Fengyu Zhou, Xiaoyan Wang, Zhi Liu, and Xiang-Yang Li. 2022. WiVi: WiFi-video cross-modal fusion based multi-path gait recognition system. In *2022 IEEE/ACM 30th International Symposium on Quality of Service (IWQoS)*. IEEE, 1–10.
- [12] Xiaoran Fan, David Pearl, Richard Howard, Longfei Shangguan, and Trausti Thormundsson. 2023. APG: Audioplethysmography for Cardiac Monitoring in Hearables. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking (MobiCom)*. 1–15.
- [13] Ruoyu Feng, Eddy De Greef, Maxim Rykunov, Hichem Sahli, Sofie Pollin, and Andre Bourdoux. 2021. Multipath ghost recognition for indoor mimo radar. *IEEE Transactions on Geoscience and Remote Sensing (TGRS)* 60 (2021), 1–10.
- [14] Claudio Filipi Gonçalves dos Santos, Diego de Souza Oliveira, Leandro A. Passos, Rafael Gonçalves Pires, Daniel Felipe Silva Santos, Lucas Pascotti Valem, Thierry P. Moreira, Marcos Cleison S. Santana, Mateus Roder, Jo Paulo Papa, et al. 2022. Gait recognition based on deep learning: A survey. *ACM Computing Surveys (CSUR)* 55, 2 (2022), 1–34.
- [15] Hannah Friederike Fischer, Daniela Wittmann, Alejandro Baucells Costa, Bo Zhou, Gesche Joost, and Paul Lukowicz. 2021. Masquare: A functional smart mask design for health monitoring. In *Proceedings of the 2021 ACM International Symposium on Wearable Computers (ISWC)*. 175–178.
- [16] Xiaonan Guo, Jian Liu, and Yingying Chen. 2016. Automatic personal fitness assistance through wearable mobile devices: Poster. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking (MobiCom)*. 437–438.
- [17] Carl Gustafson, Katsuyuki Haneda, Shurjeel Wyne, and Fredrik Tufvesson. 2013. On mm-wave multipath clustering and channel modeling. *IEEE transactions on antennas and propagation* 62, 3 (2013), 1445–1455.
- [18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition(CVPR)*. 770–778.
- [19] James Oat JudgeRoy, B Davis III, and Sylvia Ounpuu. 1996. Step length reductions in advanced age: the role of ankle and hip kinetics. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 51, 6 (1996), M303–M312.
- [20] Rudolph Emil Kalman. 1960. A new approach to linear filtering and prediction problems. *Journal of basic Engineering* 82, 1 (1960), 35–45.
- [21] Alexander Kamann, Patrick Held, Florian Perras, Patrick Zaumseil, Thomas Brandmeier, and Ulrich T Schwarz. 2018. Automotive radar multipath propagation in uncertain environments. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 859–864.
- [22] Belal Korany, Chitra R Karanam, Hong Cai, and Yasamin Mostofi. 2019. XModal-ID: Using WiFi for through-wall person identification from candidate video footage. In *The 25th Annual International Conference on Mobile Computing and Networking (MobiCom)*. 1–15.
- [23] Harold W Kuhn. 1955. The Hungarian method for the assignment problem. *Naval research logistics quarterly* 2, 1-2 (1955), 83–97.
- [24] Jiarong Li, Zihan Wang, Zihao Zhao, Yuchao Jin, Jihong Yin, Shao-Lun Huang, and Jiyu Wang. 2021. TribоГait: A deep learning enabled triboelectric gait sensor system for human activity recognition and individual identification. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp/ISWC)*. 643–648.
- [25] Tao Li, Dianqi Han, Jiawei Li, Ang Li, Yan Zhang, Rui Zhang, and Yanchao Zhang. 2021. Your home is insecure: Practical attacks on wireless home alarm systems. In *IEEE Conference on Computer Communications (INFOCOM)*. 1–10.
- [26] Xiang Li, Yasushi Makihara, Chi Xu, Yasushi Yagi, Shiqi Yu, and Mingwu Ren. 2020. End-to-end model-based gait recognition. In *Proceedings of the Asian conference on computer vision (ACCV)*.
- [27] Yunda Li and Xiaolei Shang. 2022. Multipath ghost target identification for automotive MIMO radar. In *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*. IEEE, 1–5.
- [28] Jie Lian, Changlai Du, Jiadong Lou, Li Chen, and Xu Yuan. 2023. EchoSensor: Fine-Grained Ultrasonic Sensing for Smart Home Intrusion Detection. *ACM Transactions on Sensor Networks (TOSN)* 20, 1 (2023), 1–24.

- [29] Hankai Liu, Xiulong Liu, Xin Xie, Xinyu Tong, and Keqiu Li. 2024. PmTrack: Enabling Personalized mmWave-based Human Tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 7, 4 (2024), 1–30.
- [30] James MacQueen et al. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1. 281–297.
- [31] Zhen Meng, Song Fu, Jie Yan, Hongyuan Liang, Anfu Zhou, Shilin Zhu, Huadong Ma, Jianhua Liu, and Ning Yang. 2020. Gait recognition for co-existing multiple people using millimeter wave sensing. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, Vol. 34. 849–856.
- [32] Mostafa Mirshekari, Jonathon Fagert, Amelie Bonde, Pei Zhang, and Hae Young Noh. 2018. Human gait monitoring using footstep-induced floor vibrations across different structures. In *Proceedings of the 2018 ACM international joint conference and 2018 international symposium on pervasive and ubiquitous computing and wearable computers (UbiComp/ISWC)*. 1382–1391.
- [33] Leann Myers and Maria J Sirois. 2004. Spearman correlation coefficients, differences between. *Encyclopedia of statistical sciences* 12 (2004).
- [34] Edwin Pan, Jingning Tang, Dash Kosaka, Ruihao Yao, and Arjun Gupta. 2019. OpenRadar. <https://github.com/presenseradar/openradar>.
- [35] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems (NeurIPS)* 32 (2019).
- [36] Jacopo Pegoraro, Francesca Meneghelli, and Michele Rossi. 2020. Multiperson continuous tracking and identification from mm-wave micro-Doppler signatures. *IEEE Transactions on Geoscience and Remote Sensing (TGRS)* 59, 4 (2020), 2994–3009.
- [37] Jacopo Pegoraro and Michele Rossi. 2022. Human tracking with mmWave radars: A deep learning approach with uncertainty estimation. In *2022 IEEE 23rd International Workshop on Signal Processing Advances in Wireless Communication (SPAWC)*. IEEE, 1–5.
- [38] Jacopo Pegoraro, Domenico Solimini, Federico Matteo, Enver Bashirov, Francesca Meneghelli, and Michele Rossi. 2020. Deep learning for accurate indoor human tracking with a mm-wave radar. In *2020 IEEE Radar Conference (RadarConf20)*. IEEE, 1–6.
- [39] Akash Pokkunuru, Kalvik Jakkala, Arupjyoti Bhuyan, Pu Wang, and Zhi Sun. 2018. NeuralWave: Gait-based user identification through commodity WiFi and deep learning. In *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 758–765.
- [40] Yao Qin, Dongjin Song, Haifeng Cheng, Wei Cheng, Guofei Jiang, and Garrison W Cottrell. 2017. A dual-stage attention-based recurrent neural network for time series prediction. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*. 2627–2633.
- [41] Veenu Rani and Munish Kumar. 2023. Human gait recognition: A systematic review. *Multimedia Tools and Applications (MTA)* (2023), 1–35.
- [42] Yili Ren, Yichao Wang, Sheng Tan, Yingying Chen, and Jie Yang. 2022. Person re-identification using wifi signals. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking (MobiCom)*. 829–831.
- [43] Lior Rokach and Oded Maimon. 2005. Clustering methods. *Data mining and knowledge discovery handbook* (2005), 321–352.
- [44] Ali Hussein Sabry, Ahmed Sabry, Wan Zuha Wan Hasan, Mohd Nazim Mohtar, Raja Mohd Kamil Raja Ahmad, Hafiz Rashidi Ramli, Ang Swee Peng, and Zainidi Haji Abdul Hamid. 2018. Foot plantar pressure distribution modeling based on image processing. In *IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*. 1–4.
- [45] Alireza Sepas-Moghaddam and Ali Etemad. 2022. Deep gait recognition: A survey. *IEEE transactions on pattern analysis and machine intelligence (TPAMI)* 45, 1 (2022), 264–284.
- [46] Xingzhe Song, Boyuan Yang, Ge Yang, Ruirong Chen, Erick Forno, Wei Chen, and Wei Gao. 2020. SpiroSonic: monitoring human lung function via acoustic sensing on commodity smartphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking (MobiCom)*. 1–14.
- [47] Torben Teepe, Ali Khan, Johannes Gilg, Fabian Herzog, Stefan Hörmann, and Gerhard Rigoll. 2021. Gaitgraph: Graph convolutional network for skeleton-based gait recognition. In *IEEE International Conference on Image Processing (ICIP)*. 2314–2318.
- [48] TI. 2019. DCA1000 evaluation module for real-time data capture and streaming. <https://www.ti.com/tool/DCA1000EVM>.
- [49] TI. 2021. IWR6843 intelligent mmWave sensor standard antenna plug-in module. <https://www.ti.com/tool/IWR6843ISK>.
- [50] J v. Neumann. 1928. Zur theorie der gesellschaftsspiele. *Mathematische annalen* 100, 1 (1928), 295–320.
- [51] Philip van Dorp and FCA Groen. 2008. Feature-based human motion parameter estimation with radar. *IET Radar, Sonar & Navigation* 2, 2 (2008), 135–145.
- [52] Baptist Vandersmissen, Nicolas Knudde, Azarakhsh Jalalvand, Ivo Couckuyt, Andre Bourdoux, Wesley De Neve, and Tom Dhaene. 2018. Indoor person identification using a low-power FMCW radar. *IEEE Transactions on Geoscience and Remote Sensing (TGRS)* 56, 7 (2018), 3941–3952.
- [53] Jack L Walker. 1980. Range-Doppler imaging of rotating objects. *IEEE Transactions on Aerospace and Electronic systems (TAES)* 1 (1980), 23–52.
- [54] Changsheng Wan, Li Wang, and Vir V Phoha. 2018. A survey on gait recognition. *ACM Computing Surveys (CSUR)* 51, 5 (2018), 1–35.
- [55] Yinxin Wan, Xuanli Lin, Kuai Xu, Feng Wang, and Guoliang Xue. 2023. Extracting Spatial Information of IoT Device Events for Smart Home Safety Monitoring. In *IEEE Conference on Computer Communications (INFOCOM)*. 1–10.

- [56] Wei Wang, Alex X Liu, and Muhammad Shahzad. 2016. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. 363–373.
- [57] Xinshuo Weng, Jianren Wang, David Held, and Kris Kitani. 2020. 3d multi-object tracking: A baseline and new evaluation metrics. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 10359–10366.
- [58] Yadong Xie, Fan Li, Yue Wu, and Yu Wang. 2021. HearFit: Fitness monitoring on smart speakers via active acoustic sensing. In *IEEE Conference on Computer Communications (INFOCOM)*. 1–10.
- [59] Wei Xu, ZhiWen Yu, Zhu Wang, Bin Guo, and Qi Han. 2019. Acousticid: gait-based human identification using acoustic signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 3, 3 (2019), 1–25.
- [60] Hongfei Xue, Yan Ju, Chenglin Miao, Yijiang Wang, Shiyang Wang, Aidong Zhang, and Lu Su. 2021. mmMesh: Towards 3D real-time dynamic human mesh construction using millimeter-wave. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*. 269–282.
- [61] Xin Yang, Jian Liu, Yingying Chen, Xiaonan Guo, and Yucheng Xie. 2020. MU-ID: Multi-user identification through gaits using millimeter wave radios. In *IEEE Conference on Computer Communications (INFOCOM)*. 2589–2598.
- [62] Jia Zhang, Rui Xi, Yuan He, Yimiao Sun, Xiuzhen Guo, Weiguo Wang, Xin Na, Yunhao Liu, Zhenguo Shi, and Tao Gu. 2023. A survey of mmWave-based human sensing: Technology, platforms and applications. *IEEE Communications Surveys & Tutorials* (2023).
- [63] Shaohu Zhang and Anupam Das. 2021. A 2-FA for home voice assistants using inaudible acoustic signal. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking (MobiCom)*. 834–836.
- [64] Jingyue Zheng, Xingyu Gu, Shisheng Guo, and Guolong Cui. 2021. Multipath False Target Removal for Indoor Localization. In *International Conference on Autonomous Unmanned Systems*. Springer, 1367–1378.