FlowGait: Enabling Robust Long-Term Gait Recognition Across Real-World Covariates with mmWave Radar

Dequan Wang
University of Science and Technology of China
HeFei, China
wdq15588@mail.ustc.edu.cn

Chenming He
University of Science and Technology of China
Hefei, China
hechenming@mail.ustc.edu.cn

Lingyu Wang
University of Science and Technology of China
Hefei, China
lywang19@mail.ustc.edu.cn

Chengzhen Meng
University of Science and Technology of China
Hefei, China
czmeng@mail.ustc.edu.cn

Xiaoran Fan
Independent Researcher
Sunnyvale, California, USA
vanxf@google.com

Yanyong Zhang
University of Science and Technology of China
Hefei, China
yanyongz@ustc.edu.cn

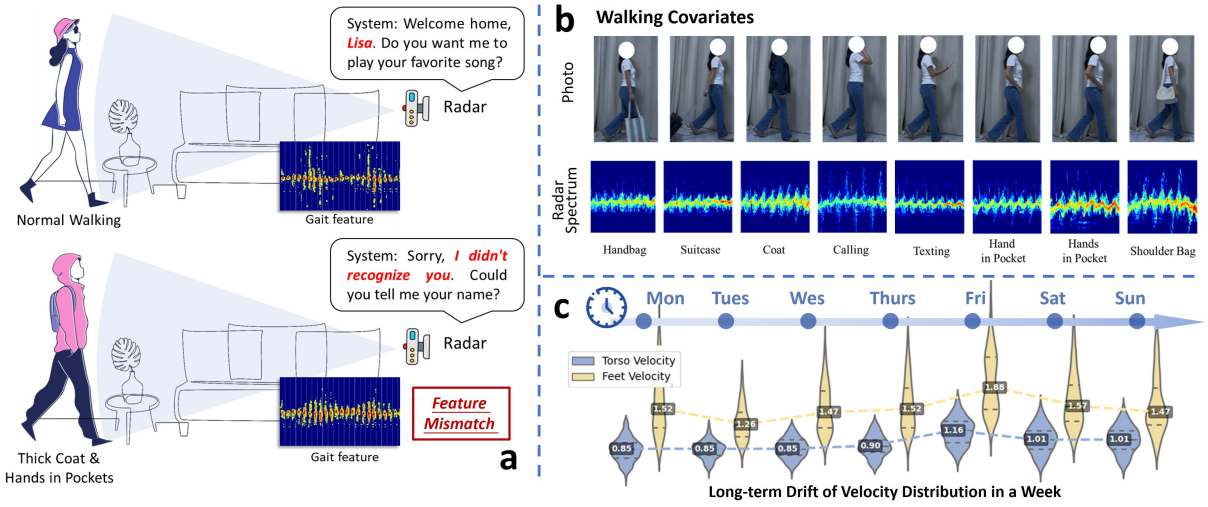


Figure 1: Challenges in robust gait recognition. (a) A Practical Failure Case: An illustrative example where a registered user is rejected by the system because her current walking pattern deviates significantly from the enrolled 'normal' data. (b) Diversity of Covariates: Various daily factors (including clothing, carried items, and actions) introduce complex distortions to the radar spectrograms, causing the recognition failure shown in (a). (c) Temporal Variability: Gait is not static; the violin plots show significant natural fluctuations in walking velocity over a week, posing challenges for fixed-model systems.

Abstract

Gait recognition enables proactive and personalized smart home interactions, but its long-term reliability is challenged by the non-static nature of gait. Covariates like carrying items and clothing induce a persistent domain shift that degrades traditional, static models. To solve this, we introduce `FlowGait`, a mmWave-based framework designed for robust, long-term adaptation. It combines self-training with continual learning, allowing the model to daily align with a user's evolving gait by learning from readily available unlabeled data. It features a specialized transformer network for radar spectrogram analysis and a novel two-stage labeling algorithm that leverages the gait's hierarchical nature to assign pseudo-labels to the unlabeled data accurately. Evaluated on three challenging datasets from 47 volunteers (covering 12 gait-covariates, 11 routes, and two weeks), `FlowGait` achieves high accuracies of 94.8 (cross-covariate), 98.6% (cross-route), and 95.5% (cross-day). Notably, for the long-term dataset, it reduced performance decay from 13.6% to just 1.4%, demonstrating its real-world robustness.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

CHI '26, Barcelona, Spain

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790623

CCS Concepts

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; **Human computer interaction (HCI)**.

Keywords

Gait Recognition, Smart Home, mmWave

ACM Reference Format:

Dequan Wang, Chenming He, Lingyu Wang, Chengzhen Meng, Xiaoran Fan, and Yanyong Zhang. 2026. FlowGait: Enabling Robust Long-Term Gait Recognition Across Real-World Covariates with mmWave Radar. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3772318.3790623

1 Introduction

Gait, a unique walking pattern of each individual, provides a reliable biometric identifier. Smart home systems can leverage this trait to distinguish between residents and deliver a range of customized automations. For instance, if a child's gait is identified nearing a designated dangerous zone, such as a staircase or balcony, the system can initiate a series of automated responses: locking the door, alerting the parents, and minimizing the possibility of potential household accidents. Similarly, once the system recognizes a user's gait as they enter the bedroom, it can activate a unique welcome routine, such as setting the preferred temperature, adjusting the lighting, or playing their dedicated music. Within the smart home system, the purpose of gait recognition is to provide an interactive environment that is truly **unobtrusive**, **proactive**, and **personalized**.

As an emerging biometric modality, gait recognition has been explored using various sensors, including cameras [11, 19, 32, 49], Wi-Fi [29, 47, 59], and ultrasound [33, 63]. However, these sensors suffer from some limitations, including privacy issues, limited resolution, and being unfriendly to pets. In contrast, millimeter-wave (mmWave) radar offers distinct advantages: it is high-resolution, non-contact, robust to lighting variations, and privacy-preserving. These qualities make mmWave radar an ideal candidate for indoor recognition tasks.

However, a major challenge is that gait is not as stable as biometrics like the face or fingerprints. For example, wearing a heavy coat restricts a person's natural motion, resulting in a distinct gait pattern characterized by a stiffer posture and suppressed arm swing. In the feature space, this physical adaptation translates to a domain shift, frequently causing the recognition system to fail, as illustrated in Fig. 1(a). This variability is compounded by numerous other factors, including carrying items, clothing, and phone-related activities (like calling). These factors are collectively known as **gait covariates**. Furthermore, variations in gait can stem from **walking routes** and **temporal factors** arising from a person's mood [15, 40], fatigue [22, 25], or health status [6, 61]. Fig. 1(b) shows the mmWave spectrograms under different gait covariates, which exhibit significant variations. Fig. 1 (c) illustrates significant day-to-day fluctuations in the velocity distribution of the same user over one week. Therefore, the major challenge for practical gait recognition lies in overcoming the performance degradation caused by this inherent instability during long-term, real-world deployment.

Existing methodologies typically involve feature extraction from radar-generated point clouds [13, 26, 38, 54], micro-Doppler spectrograms [44, 53], or heatmaps [56]. While these methods have demonstrated high performance, their validation typically relies on gait data collected over short durations in controlled laboratory environments. To handle gait variations, some methods [45, 64] employ domain adaptation techniques to learn domain-invariant features. This strategy is impractical for real-world deployment because it requires collecting labeled data multiple times. Moreover, its 'train-once' paradigm cannot account for the longitudinal nature of gait, given that a finite training set can never fully capture a user's future variability. Consequently, long-term gait adaptation remains a largely unexplored gap between research and practical application.

To overcome these challenges, we propose `FlowGait`, a novel mmWave-based gait recognition framework that integrates self-training and continual learning for robust long-term adaptation. Our approach is founded on the hypothesis that gait evolves as a continuous **flow**. This principle enables a practical enrollment process: We collect an enrollment-labeled dataset from each user once to fine-tune a general model into a personalized one. During daily operation, the system passively observes new, unlabeled walks as gaits naturally change. This incoming data is used to continually refine the model, ensuring it stays synchronized with each user's evolving patterns. `FlowGait` is effective not only for long-term adaptation but also for recognition across diverse covariates and routes, as the system progressively learns a wide range of gait patterns from simple to complex. However, there are three challenges.

A prerequisite for self-training is a robust feature recognition network whose extracted features satisfy the cluster assumption [74] across different gait covariates. This means that features from the same person remain tightly clustered despite variations in their gait. A key challenge is learning robust features from the uniquely structured mmWave radar heatmaps. These heatmaps are highly elongated and strip-like, featuring a large velocity dimension (>100 pixels) in contrast to small range and time dimensions (10-20 pixels). This elongated structure fundamentally limits the efficacy of standard CNNs, whose local kernels struggle to model the long-range dependencies essential for capturing a holistic gait pattern. To address this, we introduce a Transformer-based network specifically designed for the strip-shaped heatmaps. Its self-attention mechanism is designed to establish a global receptive field, making it suited to modeling the holistic patterns within these elongated heatmaps and leading to superior cross-domain performance.

The second challenge is ensuring data availability for continuous learning, which requires accurately labeling incoming data. A common approach is pseudo-labeling, where the model assigns labels to high-confidence predictions in the unlabeled set. However, this method faces a fundamental trade-off. A high confidence threshold ensures label accuracy but misses diverse or novel samples, while a low threshold captures more diversity but risks introducing labeling errors. To resolve this issue, we introduce a **two-stage step-traversal labeling** algorithm. We define **traversal** as a user's complete journey from an origin to a destination, composed of a series of discrete **steps**. Traversal and step are two hierarchical levels of walking; a step is the atomic unit of walking, while a traversal is a complete walking session. Our algorithm first assigns a preliminary label to each step. Then, instead of relying on individual step confidence, it evaluates the collective confidence of the entire traversal using a novel traversal labeling criterion. This approach enables the accurate labeling of low-confidence steps within a high-confidence

traversal, thereby expanding the training data with diverse samples without sacrificing accuracy.

The last challenge comes from learning continuously from an ever-growing stream of unlabeled data, which introduces two problems. First, constant data accumulation increases storage requirements and training durations. Second, training exclusively on new data leads to catastrophic forgetting—the loss of previously acquired knowledge. To address this, we employ a Core-set mechanism. This process operates on a daily cycle: new unlabeled data is used for self-training, and high-confidence pseudo-labeled samples are integrated into the existing training pool. Crucially, this combined pool is then distilled into a fixed-size core set. By capping the data volume, this approach ensures stable, predictable training times. Simultaneously, it mitigates catastrophic forgetting by replaying the core set in subsequent training iterations.

To validate our proposed system, `FlowGait`, we collected three comprehensive datasets with 47 volunteers. (1) *Cross-covariate Dataset:* involving 12 distinct gait covariates, with only "Normal Walking" data used for enrollment. (2) *Cross-routes Dataset:* spanning 11 different routes, with only "Straight Walking" data used for enrollment. (3) *Cross-day Dataset:* collected over two weeks for long-term evaluation, with only "Day-1" data used for enrollment. (4) *Demographic Dataset:* involving 10 elderly participants and 6 children to assess inclusivity. On these respective tasks, `FlowGait` achieved recognition accuracies of 94.8%, 98.6%, and 95.5%, respectively. Notably, for the long-term task, `FlowGait` reduced performance decay from 13.6% to just 1.4%, demonstrating its real-world robustness. In the demographic evaluation, the system exhibited exceptional stability, achieving 97.7% and 97.5% accuracy for the elderly and child groups, respectively. Furthermore, in simulated family unit scenarios (e.g., Multi-Generation households), the accuracy exceeded 99%, attributed to distinct inter-group gait differences. To promote further research, our datasets and code are publicly available[1]. Validation on a consumer-grade laptop demonstrated the system's ability to perform real-time inference, with model updates completing in approximately 6 minutes. In summary, the main contributions of `FlowGait` can be summarized as follows:

- We propose `FlowGait`, a long-term, self-learning mmWave-based gait recognition framework designed to adapt to real-world covariates. To the best of our knowledge, `FlowGait` is the first millimeter-wave gait recognition system designed for long-term deployment.
- We introduce a Transformer-based feature extraction network specifically designed for the strip-shaped mmWave heatmaps, achieving SOTA performance on cross-covariances recognition using a public dataset.
- We propose a two-stage step-traversal labeling algorithm. By jointly evaluating the labels and confidences at both the step and traversal levels, it achieves more accurate annotation of unlabeled data.
- We implement `FlowGait` using commercial off-the-shelf (COTS) mmWave radar and evaluate its performance on three datasets: a cross-covariate dataset, a cross-route dataset, and a cross-day dataset. Using only the Normal Walking, Straight Walk, or Day-1 data as the labeled set, our method achieved recognition accuracies of 94.8%, 97.9%, and 96.6%, respectively. Critically, it demonstrated superior robustness for vulnerable groups (elderly and children) and family scenarios (>99%), proving its viability for diverse smart home applications.

[1]https://github.com/DQ-WDQ/FlowGait

2 Related Work

2.1 Sensor-based Gait Recognition

A variety of sensors—including cameras, wearable devices, Wi-Fi, and ultrasound—are employed to capture distinctive gait signatures. Vision-based methods, for instance, typically extract bodily silhouettes and skeletal data from image frames, which are then processed by a neural network to derive gait-specific features [11, 19, 32, 49, 73]. Consequently, much of the research in this domain has focused on the challenges of cross-covariate [28, 31, 75], cross-view[9, 62], and in-the-wild[18, 48, 67] gait recognition. However, despite their notable success, the inherent privacy concerns associated with these methods render them unsuitable for deployment in private environments such as homes. Wi-Fi-based gait recognition has garnered widespread attention due to the ubiquity of Wi-Fi signals [46, 59, 68, 69]. The majority of these approaches extract features from Channel State Information (CSI), employing either traditional feature engineering or deep learning-based techniques. Recently, several studies have also begun to investigate the impact of walking direction [70, 71] and non-gait behaviors [65] on Wi-Fi-based gait recognition. However, Wi-Fi-based approaches are highly sensitive to the environment and struggle to handle multi-person scenarios. Other gait recognition approaches rely on specialized hardware, such as wearable accelerometers or floor-mounted sensors. For instance, Li et al. [30] propose a system using a floor-mounted Triboelectric Nanogenerator (TENG) that identifies gaits via electrical signals. However, the reliance on wearable devices or the need for intrusive under-floor installation for such sensors can limit their practical applicability and user acceptance. Acoustic signals represent another modality for gait recognition. For example, Xu et al. [63] analyze the Doppler shifts in acoustic signals reflected from a user's body, while Altaf et al. [1] focus on the distinct sounds of footsteps for identification. However, these acoustic-based approaches also face struggles to distinguish between multiple individuals. Moreover, the performance of acoustic methods in general is highly susceptible to interference from ambient noise, which limits their real-world robustness.

2.2 mmWave-based Gait Recognition

Millimeter-wave (mmWave) technology is gaining significant traction for many applications [12, 23, 24, 34, 36, 37, 55, 57, 58, 72], owing to its high-precision range and velocity measurements and increasing affordability. Early research on gait recognition explored various data representations. Several studies employed deep neural networks to extract features from micro-Doppler spectrograms to recognize single or multiple individuals [44, 53]. And some works leveraged the spatiotemporal information from 4D radar point clouds for user identification and re-identification [13, 26, 38, 54]. Other notable approaches include modeling the movement of lower limbs [66], performing cross-modal vision-RF ReID [8], synthesizing mmWave data from video [20], and using Range-Doppler

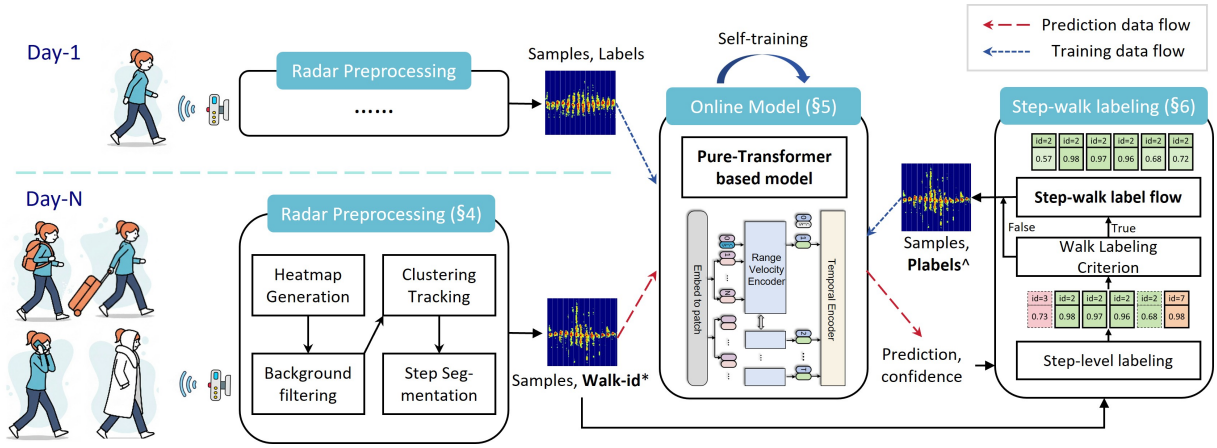


Figure 2: Architecture of FlowGait. The system utilizes a self-training loop to adapt to unlabeled daily data. The Step-Traversal module (right) aggregates predictions from all steps within a single walking event (Traversal-id*) to filter out unreliable samples. This ensures that only high-confidence pseudo-labels (Plabels^) are generated and fed back to update the Transformer model.

heatmaps with a "step" as the sample unit [56]. However, a common limitation in much of the prior works is the lack of consideration for real-world deployment, such as recognition across different gait covariates (e.g., carrying a bag), walking routes, or over extended time periods (cross-day). Recently, efforts have begun to address these domain shifts. For instance, Meng et al. [39] utilized contrastive learning on point clouds for deployment in unseen domains, while Pinyoanuntapong et al. [45] used a semi-supervised method on micro-Doppler data to mitigate these shifts. Nevertheless, these recent solutions have their own drawbacks. The method by [45], for example, defines an entire walking pass as a single sample, which is impractical in scenarios where users may turn or alter their path midway. Furthermore, this approach requires several days of labeled data for effective training, increasing the data collection burden. In summary, the current literature lacks a gait recognition system capable of robust, long-term deployment following a single enrollment. Addressing this limitation, our work proposes a millimeter-wave gait recognition system oriented towards achieving both long-term operational stability and cross-covariate robustness.

3 PRELIMINARY

In this section, we first introduce the principles of mmWave radar perception and how gait recognition works. Following that, we describe the human factors that influence gait.

3.1 Gait Recognition in HCI

While traditional biometrics like face and voice recognition prioritize accuracy, deploying sensing technologies in domestic environments requires a deeper consideration of user experience and social norms. mmWave gait recognition emerges as a compelling alternative, offering a paradigm shift from active authentication to seamless, privacy-preserving interaction. Its core advantages include:

Unobtrusive, Passive Sensing. Unlike methods requiring active cooperation (e.g., facing a camera), gait recognition is completely passive. Users simply walk naturally within the radar's range, eliminating the need to pause or pose, thus ensuring a seamless user experience.

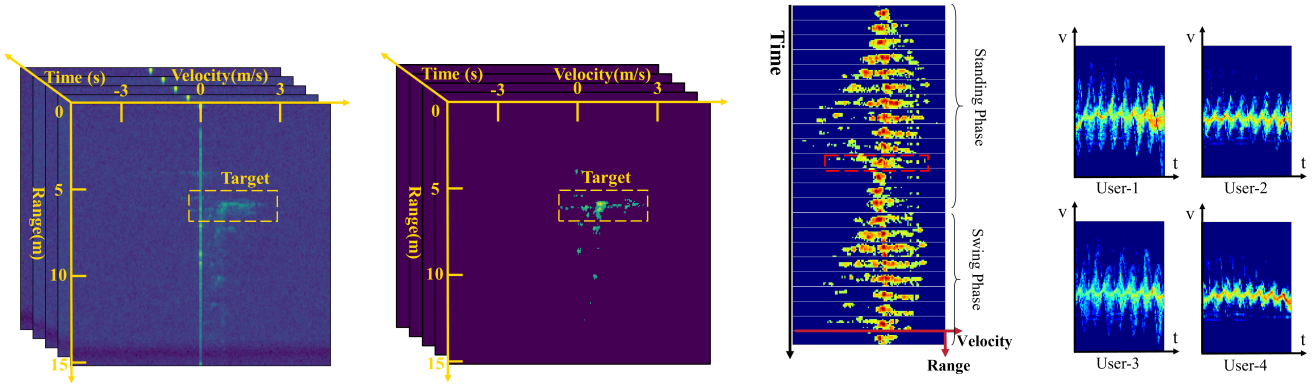
Human-Centered Privacy. HCI literature emphasizes that accepting home sensing technology relies on the trade-off between perceived utility and privacy intrusion [2]. Our choice of mmWave radar aligns with "privacy-enhancing sensing," as studies show users express significantly lower anxiety towards abstract RF signals compared to cameras [14]. Furthermore, addressing the "Contextual Integrity" [41] of the home, our system ensures that gait data—a highly personal biometric—remains local and is used strictly for user-authorized interactions, preventing context-collapse where data might be misused by third parties.

Robust Anti-Spoofing. While visual and fingerprint systems can be deceived by masks or films, mmWave captures the dynamic, holistic posture of an individual's walk. This complex biometric signature is extremely difficult to forge, offering superior security.

From Passive Tools to Proactive Environments. These advantages enable a transition to "Ambient Intelligence." Rather than awaiting explicit commands (e.g., "Siri, play music"), the environment proactively senses and anticipates needs. As a user moves between rooms, the system can provide "follow-me" services—seamlessly transferring media or adjusting lighting—while maintaining continuous, non-intrusive authentication. Beyond identification, the multi-purpose sensor simultaneously monitors gait health (e.g., for elderly care), creating a holistic, secure, and responsive living space.

3.2 Principles of mmWave Radar Sensing

mmWave radar can effectively estimate the range and velocity of a target. A typical radar system includes transmitting (Tx) and receiving (Rx) antennas. The Tx antenna emits a linearly frequency-modulated signal, known as a chirp. This signal reflects off a target, and the resulting echo is captured by the Rx antenna. Inside the radar, a mixer combines the received and transmitted signals to generate an Intermediate Frequency (IF) signal. The frequency of



(a) Raw range-Doppler heatmaps (b) Clean range-Doppler heatmaps (c) Range-Doppler stack (d) micro-Doppler

Figure 3: Visualization of Different Radar Spectrograms.

the IF signal (f_{IF}) is directly proportional to the target's range (d) due to the signal's time-of-flight. The range is calculated as: $d = f_{IF}c/2S$, where S is the chirp's frequency slope and c is the speed of light. For a moving target, the slight change in its range between consecutive chirps induces a phase shift (ω) in the IF signal. This phase difference, often called the Doppler phase shift, is proportional to the target's radial velocity (v). The velocity is determined by: $v = \lambda\omega/4\pi T_c$, where λ is the wavelength, and T_c is the time interval between chirps.

A target's motion features are primarily extracted from two key representations: the **Range-Doppler (RD) heatmap** and the **micro-Doppler (mD) spectrum**. An RD heatmap is generated by applying a two-dimensional Fast Fourier Transform (FFT) to the raw Intermediate Frequency (IF) signal (I). The resulting heatmap (H) maps the target's range and velocity distribution across range bins and Doppler bins at a single frame. To capture dynamic motion, such as a full gait cycle, these individual heatmaps are stacked over time (T) to form an **RD Stack**. From RD Stack, the **mD spectrogram** is derived by summing along the range dimension. This spectrogram visualizes the temporal evolution of the target's velocity distribution. While the RD sequence provides richer data for superior gait recognition, the mD spectrogram is more intuitive and visually interpretable. Fig. 3(c) visualizes a sample RD Stack, and Fig. 3(d) contrasts the distinct mD spectrograms produced by different targets, underscoring their utility in identifying unique gait patterns. In this paper, we use the RD Stack for gait recognition but employ micro-Doppler spectrograms for visualization.

3.3 Factors that Affect Human Gait

Human gait is an inherently complex biomechanical process that adapts to a variety of external factors. It is continuously influenced by conditions such as **clothing**, **carrying behaviors**, and **phone-related actions** (e.g., calling). Collectively, these influential factors are termed **gait covariates**. For instance, restrictive apparel like a thick coat can impede the natural articulation of knee and elbow joints, resulting in attenuated stride length and limb movement. Similarly, carrying an object such as a suitcase disrupts the body's kinematic symmetry, often suppressing the arm swing on the carrying side and thereby degrading overall gait stability. These variations manifest as distinct patterns in the micro-Doppler spectrograms captured by mmWave radar, as illustrated in Fig. 1(b). The spectrograms reveal how different covariates alter the signal's distribution. For example, the signature corresponding to rolling a suitcase exhibits a compressed and more chaotic structure, which is indicative of diminished limb swing and compromised gait stability.

The radar echo is also critically affected by other factors. The user's **walking route** is a dominant factor, with the **angle of trajectory** relative to the radar being especially critical. Since radar sensors measure the radial component of a target's velocity, kinematic details from limb motion are best resolved when the subject moves directly along the radar's boresight. For oblique routes, the sensor captures only a projection of the true velocity vector, leading to an incomplete representation of the motion. Furthermore, complex maneuvers like turning introduce significant non-stationarity into the signal, as the velocity and angle of various body parts change dynamically.

Finally, the **temporal dimension** introduces inherent, long-term variability. An individual's physiological state (e.g., health[6, 61], fatigue[22, 25]) and affective state (i.e., mood[15, 40]) are not static, leading to natural fluctuations in gait over time. This intrinsic, longitudinal variability means that achieving robust gait recognition over extended periods remains a significant challenge, even when external covariates are meticulously controlled.

4 Overview

4.1 Problem Definition

This paper presents a mmWave-based gait recognition system designed for robust, long-term deployment. We collected three comprehensive datasets with 31 volunteers. (1) *Cross-covariate Dataset:* involving 12 distinct gait covariates, with only "Normal Walking" data used for enrollment. *Cross-route Dataset:* spanning 11 different routes, with only "Straight Walking" data used for enrollment. (3) *Cross-day Dataset:* collected over two weeks for long-term evaluation, with only "Day-1" data used for enrollment. In all experimental settings, the enrollment data constitutes the sole source of labeled

information; all other data remains unlabeled. Our system is designed to leverage this small labeled set, along with a portion of the unlabeled data, to perform recognition on the remaining data.

4.2 FlowGait Architecture

As illustrated in Fig. 2, FlowGait is primarily composed of three components:

Data Preprocessing: The data preprocessing module transforms the raw radar signals into sets of samples, where each set constitutes a complete traversal consisting of multiple steps.

Step-level Gait Recognition: The Online Model is a Transformer-based feature extractor, specifically designed for the long, strip-shaped mmWave heatmaps. It generates a feature embedding and a predicted ID for each step sample.

Step-Traversal Self-training: The module features a two-stage step-traversal labeling algorithm generating pseudo-labels for unlabeled data. This module also includes a continual learning pipeline for long-term deployment.

5 Data Preprocessing

This section details the process for generating the network's standard input from raw radar IF signals. The process involves two stages: first, converting the raw signals into a sequence of preprocessed heatmaps; and second, segmenting these heatmaps into discrete, step-level samples.

5.1 Heatmap Generation and Target Extraction

Raw Heatmap Generation. As described in Section 3.2, we first generate a Range-Doppler heatmap by applying Range-FFT and Doppler-FFT to the IF signal. During the Doppler-FFT stage, the Moving Target Indication (MTI) algorithm [4] is employed to suppress static clutter. The resulting RD heatmap is a two-dimensional matrix where higher values indicate a stronger signal from a potential target at a specific range and velocity. A typical heatmap, as shown in Fig. 3(a), is composed of the dynamic target signal, residual static clutter, and background noise.

Background Filtering. To remove background noise, we first apply Min-Max normalization to each Range-Doppler heatmap, scaling all values to $[0, 1]$. Next, a fixed threshold is used to filter the noise: pixel values below this threshold are zeroed out, while those above are retained. This approach is chosen over CFAR-based methods because it ensures that all bins containing velocity information are preserved [56]. The filtered heatmap is shown in Fig. 3(b).

Target Clustering and Tracking. To isolate the target, the filtered heatmap is further processed using clustering and tracking. First, we employ the DBSCAN algorithm [17] on each frame to group points based on density in the range-Doppler space, identifying dense regions as targets and classifying sparse points as noise. The centroid of the densest cluster is designated as the target center, (r_0, v_0). Subsequently, a Kalman filter[60] is used to smooth the trajectory of these centers across frames, yielding a refined coordinate (r_c^i, v_c^i) for each frame i. Finally, we crop the heatmap around this smoothed coordinate to retain only the relevant information near the target:

$$RD_i = \left\{ (r, v) \middle| |r - r_c^i| < \frac{s}{2} \text{ and } |v| < v_m \right\}, \tag{1}$$

where s and v_m are the range and velocity thresholds, respectively. We adopt the values as calculated in [56]. Finally, we stack the cropped RD heatmaps from a complete walking process to generate a range-Doppler Stack (RDStack):

$$\text{RDStk} = \{ RD_i \mid 1 \le i \le N_T \}, \tag{2}$$

where N_T is the number of frames, is determined by the walking duration. Fig. 3(c) shows a sample of an RDStack.

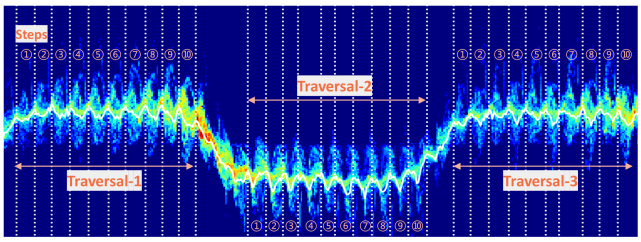


Figure 4: Step Segmentation. The white solid line represents the estimated torso velocity. The dashed lines indicate the local maxima of the velocity, which are used for segmentation. The labels "traversal-1", "traversal-2", ... are the traversal numbers, and ①, ②, ... are the step numbers.

5.2 Step Segmentation

Once a complete gait sequence (RDStack) is acquired, it must be segmented into standardized, fixed-length samples for network input. While some methods use longer spectrograms spanning several steps—an approach suitable for CNNs—this can be problematic. For example, abrupt stops or turns can create anomalous patterns in the spectrogram that degrade recognition performance. To mitigate this, we adopt a step-level segmentation strategy. Using a single gait step as the fundamental input unit allows our system to generate a prediction for each step, yielding greater robustness to irregular movements. We called a complete RDStack a **traversal** as a continuous walk from a start point to an endpoint, composed of multiple **steps**. A traversal begins when a user starts walking or enters a monitored area and ends when they stop or exit. Sharp turns split a single walk into two separate traversals. Once a traversal is fully identified, we then perform step segmentation to isolate the individual steps within it.

To further enhance model performance, we standardize each sample by aligning it to a consistent starting phase within the gait cycle, thereby reducing intra-class variation. The key to this alignment is identifying step boundaries from the subject's velocity profile. We estimate the torso velocity using the percentile method proposed in [52], which defines it as the weighted median of velocities in the micro-Doppler spectrum. We first generate the mD heatmap by summing the RDStack along the range dimension:

$$MD = \sum_r RDStk[:, r, :]. \tag{3}$$

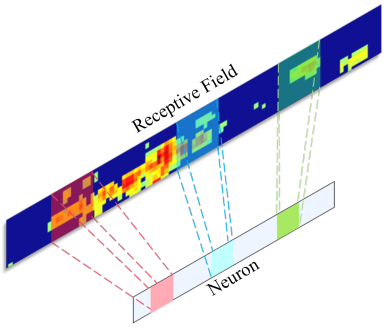


Figure 5: Receptive Fields for CNN, analogous to viewing a long scroll painting through a square window.

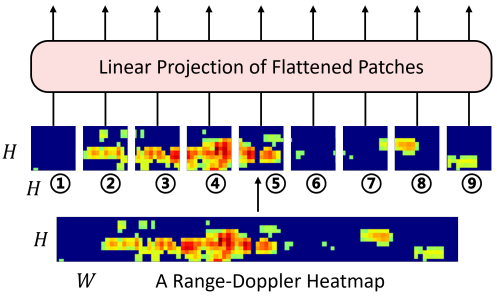


Figure 6: Heatmap Patching. We segment the heatmap into $N = W/H$ square patches, using the range dimension(H) as the side length for each patch.

Then, we estimate the torso velocity (v_{torso}) using the percentile method:

$$v_{\text{torso}} = \left\{ v_i \mid \frac{\sum_{k=v_{\min}}^{v_i} \text{MD}(i,k)}{\sum_{k=v_{\min}}^{v_{\max}} \text{MD}(i,k)} = 0.5 \right\}. \tag{4}$$

We then identify the start of each gait step by locating the time of the minimum value in the estimated torso velocity sequence. This starting time, denoted as t_{min}, is used to extract an individual sample (RDSample) as follows:

$$RDSample = RDStack[t_{min}, t_{min} + T, :, :], \tag{5}$$

where T is the fixed temporal length of the sample. In our configuration, we set $T = 20$ frames, which corresponds to a duration of 1 second.

We similarly utilize torso velocity for traversal segmentation. Significant fluctuations in torso velocity occur during walking pauses or sharp turns. This happens because mmWave radar measures radial velocity; therefore, when the target's walking direction changes relative to the radar, the *measured* torso velocity changes accordingly even if the walking speed remains constant. Specifically, we split a sequence into two separate traversals whenever the torso velocity fluctuates by more than 30%.

6 Step-level Gait Recognition

In this section, our goal is to design a feature extraction network for the RDStack that generalizes well. This requires the network to meet two criteria: it must accurately recognize gait across different covariates for unseen users, and features from the same user must form tight clusters in the feature space. The core challenge stems from a structural mismatch between conventional models and the input data. As shown in Fig. 5, range-Doppler heatmaps are elongated, with a dominant velocity dimension and compact range and time dimensions. The architectural design of traditional CNNs, defined by limited and isotropic receptive fields, is fundamentally mismatched with this data structure. This inherent bias towards local feature extraction prevents them from modeling the crucial long-range dependencies in Doppler velocity patterns, leading to suboptimal performance. To overcome this challenge, we introduce a feature extraction network based on a two-layer Transformer architecture, which is inherently adept at modeling such long-range contextual relationships. This chapter details our proposed methodology, including the dataset selection for pre-training, the architecture of our `FlowGait` Model, and a comprehensive performance evaluation.

6.1 `FlowGait` Model

To accommodate the long strip shape of the heatmap, we propose a Transformer-based feature extraction network for millimeter-wave heatmaps, inspired by the Vision Transformer (ViT)[16] and its adaptations for video recognition[3]. As depicted in Fig. 7, our model processes an input tensor of size $[T, H, W]$, which represents T consecutive range-Doppler heatmaps of size $[H, W]$. The structure of `FlowGait` is shown in Fig. 7.

Intra-Frame Feature Extractor. Our feature extraction process begins by partitioning each range-Doppler heatmap to model its internal spatial relationships. To account for the heatmap's elongated structure (where velocity dimension $W \gg H$), we divide it into $M = W/H$ square patches of size $H \times H$. Following the ViT framework, these patches ($x \in \mathbb{R}^{H \times H}$) are linearly projected via a matrix $\mathbf{E}$ and flattened into a 1D token sequence. A learnable classification token (z_cls) is prepended, and a position embedding (p) is addeed to preserve spatial information:

$$z = [z_{cls}; \mathbf{E}x_1; \mathbf{E}x_2; ...; \mathbf{E}x_M] + p. \tag{6}$$

The resulting tokens ($z \in \mathbb{R}^D$) are then processed through an L-layer Transformer encoder. As shown in Fig. 7(b), each layer sequentially applies a Multi-head Self-Attention (MSA) mechanism and a Multilayer Perceptron (MLP) module, with Layer Normalization (LN) preceding each block and residual connections applied after:

$$y^l = \text{MSA}(\text{LN}(z^l)) + z^l, \tag{7}$$

$$z^{l+1} = \text{MLP}(\text{LN}(y^l)) + y^l. \tag{8}$$

The final output of the range-velocity encoder is the state of the classification token from the last layer, denoted as $h = z^L_{cls} \in \mathbb{R}^d$.

Inter-Frame Relation Encoder. To model the temporal relationships, the frame-level feature vectors $\{h_i\}$ are first concatenated. A new, learnable classification token f_{cls} is then prepended to the sequence. Similar to the spatial encoder, we apply a linear projection $\mathbf{E}$ to each feature vector and add a position embedding p to retain

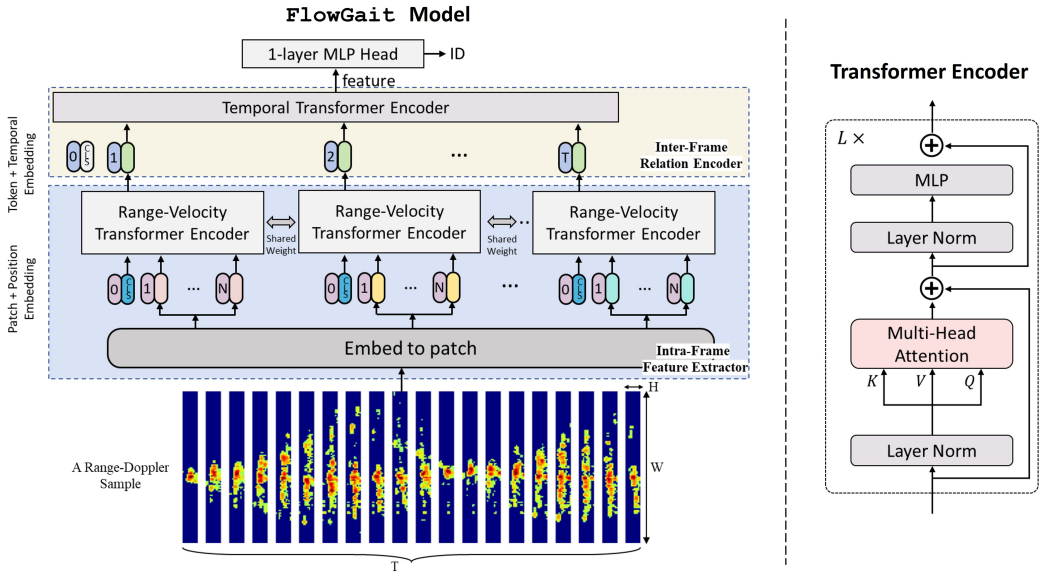


Figure 7: FlowGait Model. FlowGait model consists of two transformer encoders in series. The first encoder interacts between tokens extracted from the same temporal index to produce a latent representation per time index. The second transformer models interactions between time steps.

Table 1: Combined Evaluation for Top-1 (%) and mAP. Best results are in bold

Top-1(%)/mAP	NM	BP	SB	CA	TX	avg
RDGait	96.3/0.72	90.1/0.70	86.4/0.68	87.5/0.70	82.3/0.66	88.5/0.69
GaitSet	91.5/0.35	71.3/0.31	71.5/0.31	74.7/0.31	61.0/0.29	74.0/0.31
mD-ViT	87.4/0.64	79.1/0.59	80.3/0.59	76.6/0.59	75.6/0.56	79.8/0.59
FlowGait Model (ours)	**97.2/0.83**	**91.3/0.76**	**91.1/0.74**	**92.6/0.75**	**86.2/0.67**	**91.7/0.75**

temporal order:

$$f = [f_{cls}; \mathbf{E}h_1; \mathbf{E}h_2; ...; \mathbf{E}h_T] + p. \tag{9}$$

This aggregated sequence f is then processed by a temporal Transformer encoder consisting of L_T layers. This encoder models the interactions between tokens across different temporal steps. Finally, the state of the classification token from the temporal encoder's output is passed to a single-layer MLP for final classification.

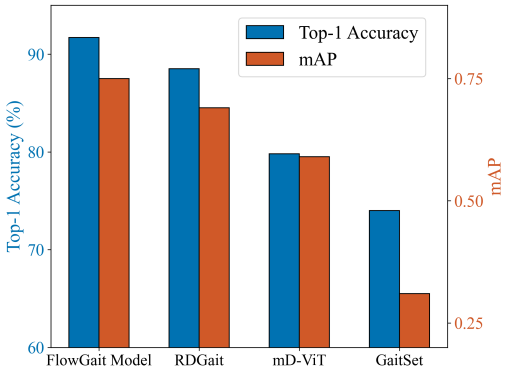
Our model FlowGait can be categorized as a "late fusion" strategy for temporal information. It operates by first extracting per-frame spatial-velocity features and subsequently aggregating them to form a final representation for classification. In contrast to CNN-based methods, FlowGait leverages a self-attention mechanism that computes the interrelationships and relative importance between every patch in the heatmap. This allows it to capture long-range dependencies across distant regions, thereby achieving a truly global receptive field.

6.2 Preliminary Evaluation of FlowGait Model

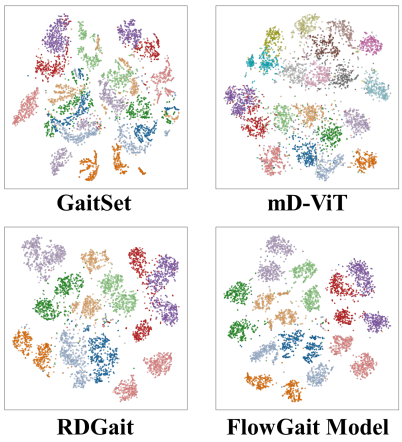
6.2.1 Dataset. To validate the model's generalization performance, we perform cross-covariate gait recognition on unseen users using the RDGait dataset[56]. RDGait dataset comprises recordings from 125 subjects across two scenarios and five gait covariates: normal walking (NM), backpack (BP), shoulder bag (SB), calling (CA), and texting (TX). The raw data is processed into tensors with dimensions of [20,11,220] corresponding to time, range, and velocity. For our experiments, we exclusively use data from Scenario 2, which includes 110 subjects. This set is partitioned into a training set of 100 subjects and a test set of the remaining 10, ensuring strict subject independence. The model is trained on all gait covariates from the training set and is not fine-tuned on the test set. We structure the test set into gallery and query sets. The gallery set for each test subject is composed of 20% of their normal walking (NM) data. The query set contains the remaining 80% of their NM data plus all of their data from the other four covariates (BP, SB, CA, TX).

6.2.2 Evaluation Metrics. During testing, feature vectors are extracted for all samples in the query and gallery sets. For each query, we rank all gallery samples based on their Euclidean distance to the query vector, where smaller distances indicate higher similarity. The performance is quantified using Top-1 Accuracy and mean Average Precision (mAP):

- **Top-1 Accuracy**. The proportion of queries where the highest-ranked sample is a correct match.



(a) Gait Recognition Accuracy for Unseen Users.



(b) t-SNE Visualization of Features from Various Models. Different colors represent different IDs.

Figure 8: Performance and Feature Visualization of the `FlowGait` Model.

- **Average Precision (AP)**. For a given query, the average of precision values is calculated at the rank of each correctly matched gallery sample.
- **Mean Average Precision (mAP).** The mean of AP scores computed over all queries.

6.2.3 Baselines. We selected several networks for visual gait recognition and mmWave gait recognition as our baselines, as follows:

- **RDGait[56]**. RDGait is a state-of-the-art millimeter-wave gait recognition network that uses a CNN and a carefully designed attention-based LSTM.
- **GaitSet[10]**. A classic visual gait algorithm, using a CNN for feature extraction and set-pooling for aggregation.
- **mD-ViT.** A standard ViT model to extract features from micro-Doppler spectrograms.

6.2.4 Overrall Performance. Fig. 8(a) shows the top-1 accuracy and mAP of the cross-covariate recognition for unseen users. The results show that our proposed model, `FlowGait` Model, achieves a Top-1 accuracy of 91.7% and an mAP of 0.75, performing the best among all baseline methods. In contrast, the SOTA mmWave gait recognition method, RDGait, achieves a Top-1 accuracy of 88.5% and an mAP of 0.69. Our method improves upon RDGait by 3.2% and 0.06, respectively. An approach directly adapted from GaitSet performs poorly (74.0% and 0.31), which indicates that directly applying methods from the visual domain is not suitable. Additionally, the method using micro-Doppler signatures yields a lower accuracy (79.8%), which suggests that range-Doppler stacks contain more information and offer higher resolution than micro-Doppler spectrograms.

Table 1 details the specific accuracy and mAP for each gait covariate. `FlowGait` achieves the highest recognition accuracy across all covariates. Notably, for the four simpler covariates—normal walking(NM), backpack(BP), shoulder bag(SB), and calling(CL)—the accuracy of each exceeds 90%. This satisfies our requirement for feature robustness when gait variations are not very large. As the accuracy indicates, the texting (TX) scenario is the most challenging. This is because texting while walking disrupts a person's pace and balance, causing a significant deviation in their gait pattern.

Fig. 8(b) visualizes the feature spaces of different models using t-SNE[35], with colors denoting subject IDs. Among them, `FlowGait` generates the most compact and discriminative feature clusters, demonstrating superior feature separation. Interestingly, we discovered that while gaits towards and away from the radar form a unified cluster during training, they naturally split into two sub-clusters in the test set. We attribute this to the differing radar cross-sections (RCS) of the body in opposing directions. Based on this, we modified our network to treat each direction as a separate identity during training and testing. This ensures each identity forms a single, tight cluster, and we merge the results only at the final output.

7 Step-Traversal Self-training

The reliability of gait recognition is often compromised by its high variability compared to static biometrics like faces or fingerprints. Real-world factors—from the clothes someone wears to their level of fatigue—can significantly alter their walking pattern, leading to critical identification failures. Our work confronts this problem by introducing an adaptive self-training approach. Our approach leverages abundant, unlabeled data to enable the network to learn and adapt to a user's diverse gait patterns. As shown in Fig. 2, the trained model processes the unlabeled dataset to generate predictions, which are subsequently assigned as pseudo-labels to these samples. This newly augmented dataset is then fed back into the model for further training. To generate a larger and more accurate set of pseudo-labels, we introduce a **two-stage step-traversal labeling algorithm**. A **traversal** is a user's complete walking process from an origin to a destination, composed of a series of discrete **steps**. By jointly evaluating the labels and confidence scores at both the step and traversal levels, this algorithm enables label propagation to low-confidence samples and achieves more accurate annotation of unlabeled data.

7.1 Pretained Model

First, a robust pre-trained model is essential to ensure that features extracted under different gait covariates meet the cluster assumption in the feature space. To this end, we pre-train our `FlowGait`

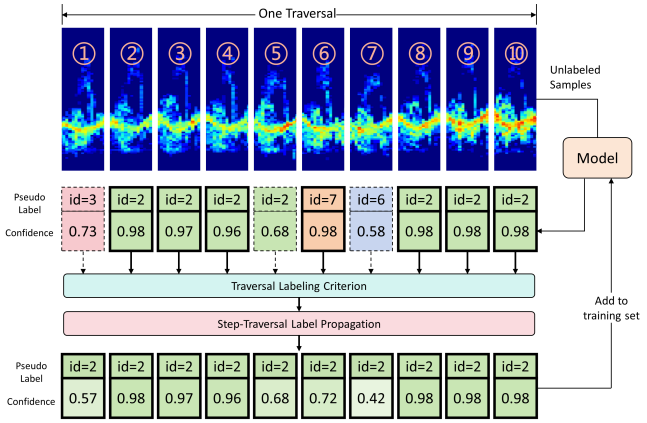


Figure 9: The Effect of the Step-Traversal Labeling Algorithm. To address incorrect labels and omissions in the step-level pseudo-labeling, we perform a label propagation once the labels in a complete traversal meet our traversal labeling criterion. This process corrects erroneous pseudo-labels and adds previously excluded, low-confidence steps to the training set. As a result, the model learns to recognize a more diverse range of gaits.

model on the RDGait dataset[56]. During pre-training, we treat samples of a subject walking towards and away from the radar as two distinct identities. This separation is critical because the body's radar reflection characteristics are highly orientation-dependent. Without this strategy, features for these two directions might cluster during training but would remain separate in the test set (see Sec. 6.2.4).

7.2 Self-training

Self-training is a semi-supervised learning approach where a model enhances its own performance by generating pseudo-labels from abundant unlabeled data and using them as new training samples. This methodology is particularly effective for gait recognition due to the temporal nature of gait: it exhibits short-term stability but undergoes long-term evolution. Self-training allows the model to continuously adapt to these gradual changes, leveraging the short-term consistency to confidently label new data and learn from the long-term shifts. The self-training in our system is divided into two stages: a supervised warm-up and pseudo-label-based semi-supervised fine-tuning. We assume that an initial batch of labeled data, referred to as "enrollment data", is readily available, marked as: $D_L = \{(x_i^l, y_i^l)\}_{i=1}^{N_L}$, where N_L is the number of samples, x_i^l is a sample, and y_i^l is its corresponding label. During the system's operational phase, new, unlabeled data is acquired, denoted by $D_U = \{x_j^u\}_{j=1}^{N_U}$. The process unfolds as follows: first, the model is warmed up using only the labeled enrollment data (D_L). Subsequently, we leverage the unlabeled data (D_U) to generate pseudo-labels for the second fine-tuning stage, allowing the model to adapt and improve.

During the self-training process, we freeze the model's feature extraction module. This approach serves multiple purposes: it preserves the robust knowledge acquired from the large-scale pre-training dataset, prevents overfitting to new users, and enhances training efficiency by eliminating gradient computations for the frozen layers. Building on this frozen feature extractor, we then attach and train a new classification head, which is a 3-layer MLP with dimensions (D_M, D_M, C), where D_M is the intermediate feature dimension and C is the number of classes.

7.2.1 Supervised Warm-up. We begin with a supervised warm-up phase before semi-supervised fine-tuning. This initial stage uses only the enrollment data to bring the pre-trained network to a stable and robust state before introducing unlabeled data. Concretely, during the first T_W epochs, the model is trained exclusively on the enrollment data by optimizing only for the supervised loss L_S:

$$L_S = \frac{1}{N_L} \sum_{i=1}^{N_L} H(y_i^l, f_\theta(x_i^l)), \tag{10}$$

where $H(y, p)$ is the cross-entropy (CE) loss between the true label y and the predicted probability distribution p.

7.2.2 Semi-supervised Fine-tuning. We address semi-supervised learning for unlabeled data through a pseudo-labeling approach. The central principle of this method is to utilize the model's own predictions on the unlabeled set to assign pseudo-labels, which are subsequently treated as true labels for further training. Herein, we introduce the two-stage step-traversal labeling algorithm. This algorithm is designed to leverage high-confidence **step** samples by propagating their labels to associated samples along the same **traversal**.

Step-level Labeling: Initially, for each unlabeled sample $x_j^u \in D_U$, the model f_θ generates a C-dimensional probability vector p_j^u. The initial pseudo-label $\hat{y}_j^u$ and its confidence score $conf_j^u$ are determined as follows:

$$p_j^u = f_\theta(x_j^u); \quad \hat{y}_j^u = \arg\max(p_j^u); \quad conf_j^u = \max(p_j^u). \tag{11}$$

To mitigate confirmation bias from noisy labels, we introduce a confidence threshold $\tau_c \in [0, 1]$. A pseudo-label is assigned to a sample only if its maximum predicted probability exceeds this threshold.

Traversal-level Label Propagation: A **traversal** is a user's walking sequence from start to finish, composed of multiple **steps** (unlabeled samples), denoted by $P = \{x_1^p, x_2^p, ..., x_{N_p}^p\}$.. Our goal is to assign a single, highly reliable label to the entire traversal.

A class $\hat{c}$ is designated as the **definitive label** for a traversal P if it represents a dominant proportion of high-confidence steps within the traversal. For each class c, we calculate proportion $\pi(c)$ of high-confidence steps:

$$\pi(c) = \frac{1}{N_p} \sum_{k=1}^{N_p} \mathbb{I}(\hat{y}_k^p = c \wedge conf_k^p > \tau_c), \tag{12}$$

where $\hat{y}_k^p$ and $conf_k^p$ are the initial pseudo-label and confidence for the k-th step, and $\mathbb{I}(\cdot)$ is the indicator function.

Let $\hat{c} = \arg\max_c \pi(c)$ be the class with the highest proportion. We define the **traversal labeling criterion**: if the highest proportion exceeds a traversal threshold $\tau_p \in [0, 1]$, we assign $\hat{c}$ as the definitive traversal label L_P:

$$L_P = \hat{c} \quad \text{if} \quad \pi(\hat{c}) > \tau_p. \tag{13}$$

Once determined, this label L_P is propagated to all steps within the traversal, updating their pseudo-labels to ensure consistency. This process corrects initial misclassifications and incorporates even low-confidence samples into the training set under a high-quality label, as illustrated in Fig. 9. If no class meets the threshold τ_p, the steps retain their original, confidence-filtered pseudo-labels. Different values of τ_p represent a more conservative or aggressive mechanism for assigning the determinant label. In our experiments, we set τ_c and τ_p to 0.95 and 0.5, respectively, and we analyze the impact of varying these values in our evaluation.

Constructing the Pseudo-Labeled Dataset: We now construct a new pseudo-labeled dataset D'_U by collecting all unlabeled samples that have received a high-quality label through our process.

A sample x_j^u with its final pseudo-label $\tilde{y}_j^u$ is added to D'_U if it meets one of two conditions:

1. **Traversal-level Label**: Its traversal $P(j)$ was assigned a definitive label $L_{P(j)}$. In this case, its final label is $\tilde{y}_j^u = L_{P(j)}$. All steps from such a traversal are included.
2. **Step-level Label**: Its traversal $P(j)$ did not get a definitive label, but the step's own confidence is high ($conf_j^u > \tau_c$). In this case, its final label is its own initial one, $\tilde{y}_j^u = \hat{y}_j^u$.

The resulting set $D'_u = \{(x_j^u, \tilde{y}_j^u\}$ contains all unlabeled samples that we are confident about, with their pseudo labels.

Loss Function: With the high-quality set D'_u defined, the unsupervised loss L_U is simply the standard cross-entropy loss computed over this set:

$$L_U = \frac{1}{|D'_U|} \sum_{(x_u, \tilde{y}_u) \in D'_U} H(\tilde{y}_u, f_\theta(x_u)), \tag{14}$$

where $|D'_U|$ is the number of samples in the new pseudo-labeled dataset. The total loss L for training is the weighted sum of the supervised and unsupervised losses:

$$L = L_S + \alpha L_U, \tag{15}$$

where α is a balance coefficient. During backpropagation, the pseudo-labels $\tilde{y}_u$ are treated as fixed constants (i.e., we stop their gradients). This same label propagation logic is also applied during inference.

7.3 Continuous Learning Pipeline

In the process of real-world deployment, the self-training method still faces several challenges. First, as new data continuously arrives and accumulates over time, it leads to increasing storage requirements and progressively longer training durations. Second, an excessive amount of new data can cause the model to become biased towards recent samples, leading to catastrophic forgetting. To address these issues, we implement a continuous learning pipeline based on core-set replay. Our model update pipeline is shown in Fig. 10. Our pipeline begins at the enrollment stage, where the collected data forms a labeled dataset. Starting from the second day, the system enters a daily update cycle:

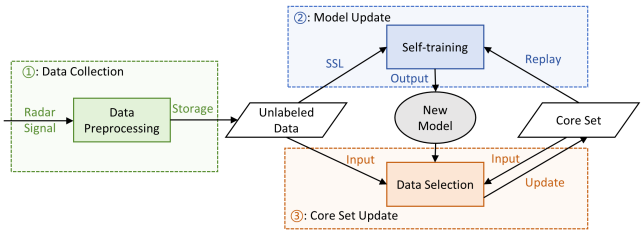


Figure 10: `FlowGait` Continuous Learning Pipeline. Executing in the order of 1,2,3.

Data Collection: During the user's daily activity, the system automatically collects their gait data. This data is then aggregated at the end of each day to form an unlabeled dataset.

Model Update: The system performs self-training using the "core set" (i.e., the labeled data) and the newly generated unlabeled data from the day. The core set is a fixed-size collection comprising the initial labeled enrollment data and pseudo-labeled daily data. During self-training, this core set is replayed to maintain the model's accuracy on historical patterns, thereby preventing catastrophic forgetting. To control the training duration, our system trains for a fixed number of epochs in each update cycle.

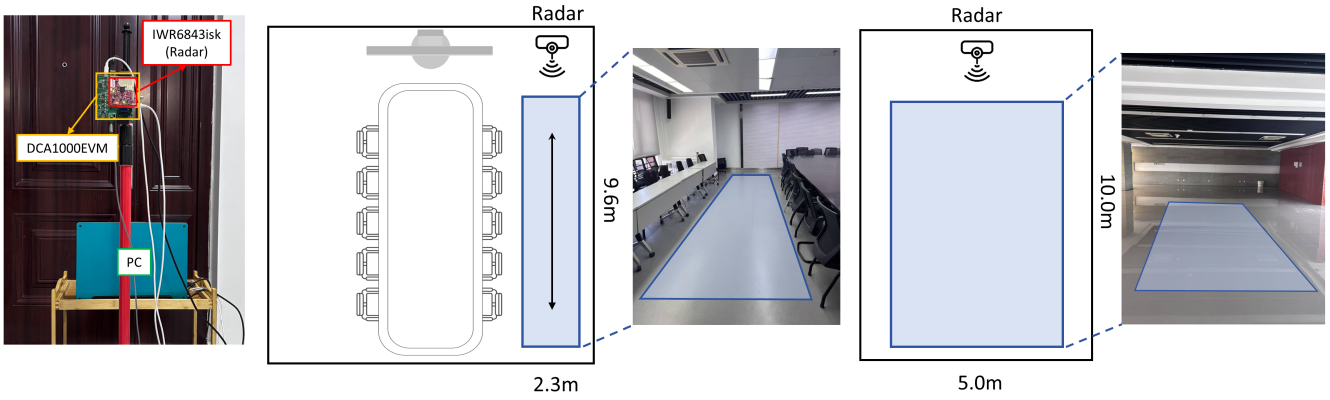
Core Set Update: After the model training is complete, we generate a new core set by screening both the unlabeled data and the existing core set. First, the unlabeled data is fed into the new model, and low-quality samples are filtered out using the pseudo-labeling algorithm described in the previous section. Next, we perform proportional downsampling on the combination of this high-confidence pseudo-labeled data and the old core set to yield the new, updated core set. This approach ensures that the new core set not only retains a portion of historical data but also increases the proportion of recent data, effectively keeping the model synchronized with the user's evolving gait. In this stage, we incorporate both high- and low-confidence samples without applying further filtering based on their individual scores. This strategy is justified by two key reasons. First, samples with lower step-level confidence have already been validated by the traversal labeling criterion; since the aggregated traversal-level confidence is sufficiently high, the reliability of these samples is ensured. Second, a core objective of our approach is to explicitly retain these 'outlier' data points, as they often capture critical temporal variations in the user's gait patterns.

In summary, through the pipeline described above, we achieve dynamic synchronization with the user's gait while constraining the training time for each update to a low and stable value.

8 Evaluation on Cross-Covariate/Route Recognition

8.1 Implementation

We have implemented `FlowGait` in Python. The data preparation module was modified based on OpenRadar [42]. The network and self-training module are implemented using PyTorch [43]. For the training stage, we use a batch size of 256 and the Adam optimizer with an initial learning rate of 1×10^{-4}, which decreases according to a cosine annealing schedule. For all models, the training stage is



(a) Radar device (b) Scenario I: Conference room (c) Session II: Hall

Figure 11: Experimental Equipment and Scenarios for Our Dataset. The arrows indicate the walking direction, and the blue area in the real-life image represents the walking area.

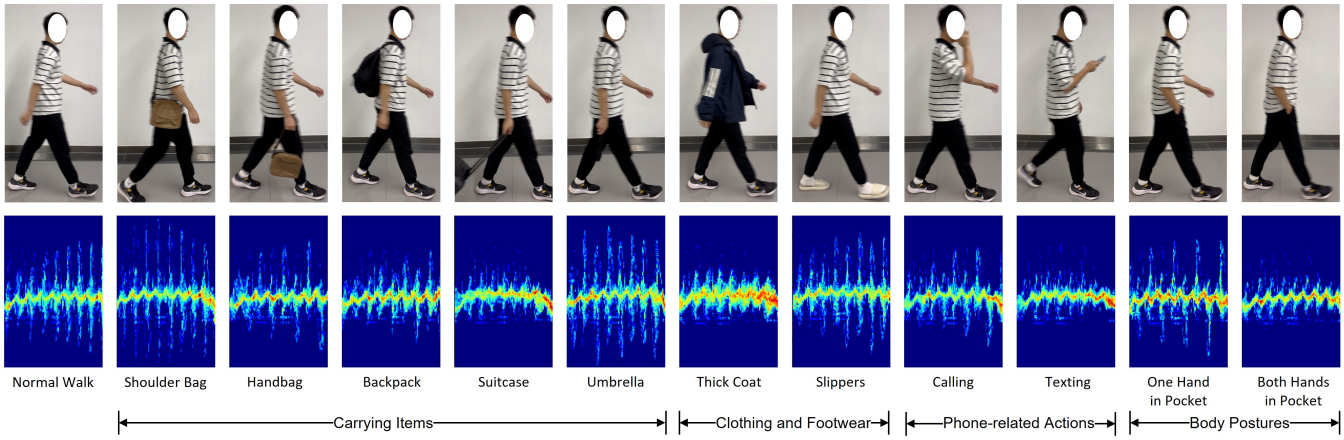


Figure 12: Visualization of Different Walking-covariates and the Corresponding Micro-Doppler Spectrum in Cross-Covariates Dataset.

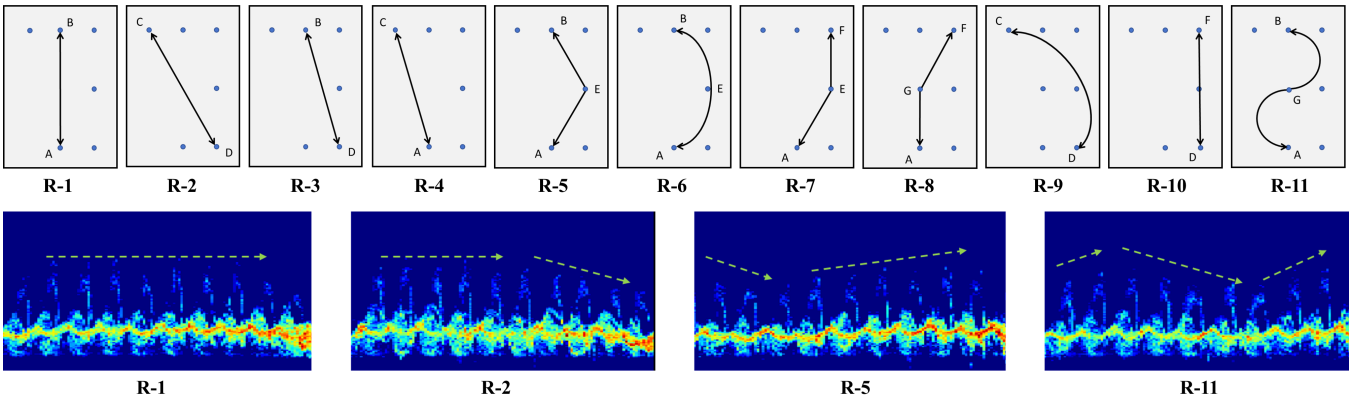


Figure 13: Visualization of the Experimental Routes and their Corresponding Micro-Doppler Spectrum in Cross-Routes Dataset. With the radar at the origin, the y-axis extends forward, and the x-axis extends to the right. The coordinates for the points in the figure are: $A(0.0, 2.4)$, $B(0.0, 8.8)$, $C(-1.6, 8.8)$, $D(1.6, 2.4)$, $E(1.6, 5.6)$, $F(1.6, 8.8)$, $G(0.0, 5.6)$. The green arrows in the micro-Doppler spectrums represent radial velocity changes.

conducted over a maximum of 100 epochs. Parameters T_W is set to 20, τ_c is set to 0.95, τ_p is set to 0.5, and α is 1.0.

As shown in Fig. 11(a), FlowGait is developed on a single-chip millimeter-wave radar, IWR6843boost [51]. We use the DCA1000 evaluation module [50] for real-time data capture. The start and end frequencies were set to $60GHz$ and $64GHz$, respectively. And the Frequency Slope was set to $119.975MHz/\mu s$. 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 range of $15m$. The radial velocity that FlowGait can measure is $6.8m/s$, with a velocity resolution of $5cm/s$. 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.

8.2 Data Collection and Baselines

We conducted a comprehensive laboratory study with 24 volunteers to evaluate FlowGait's cross-covariates/routes performance. Our studies were approved by the Institutional Review Board (IRB) of our institution. The dataset included 24 participants (14 male, 10 female), aged 20–35 years, with heights ranging from 1.52–1.91 m and weights from 41–110 kg. Our dataset is composed of two distinct subsets: a **Cross-Covariate dataset** (12 participants), a **Cross-Route dataset** (12 participants). Each volunteer took part in only one experiment. The two datasets are introduced below:

- **Cross-Covariate dataset.** The Cross-Covariate dataset includes 12 participants and **12 distinct gait covariates**, which are: Carrying Items (Shoulder Bag, Handbag, Backpack, Suitcase, Umbrella), Clothing and Footwear (Thick Coat, Slippers), Phone-related Actions (Calling, Texting), Body Posture (One Hand in Pocket, Both Hands in Pockets), and Normal Walking. Fig. 12 shows the walking photos and corresponding micro-Doppler spectrograms for the different gait covariates, demonstrating that the spectrograms vary significantly across different covariates. Data collection took place in a conference room, as shown in Fig. 11(b). Participants were instructed to walk naturally for two minutes for each covariate condition and for four minutes for Normal Walking. Detailed descriptions of each covariate and its abbreviation are provided in Table 6 in the Appendix.
- **Cross-Route dataset.** The Cross-Route dataset contains data from 12 participants navigating **11 distinct routes**: five straight-line routes, three polyline routes, and three curved routes. The data was collected in a large hall, as shown in Fig. 11(c). For each designated route, volunteers were instructed to walk back and forth along the route for two minutes. The different walking routes are shown in Fig. 13. Each route is denoted as R-1, R-2, ..., R-11. Fig. 13 also shows the micro-Doppler spectrograms of the same volunteer walking along different routes. In the figure, the green dashed line indicates the velocity trend, showing that the spectrograms for different routes exhibit varying degrees of scaling in the velocity dimension.

After preprocessing the data as described in Sec. 5, each data point consists of a sample, a user ID, covariate-ID/route-ID, and a **traversal-ID**. Each **traversal-ID** represents a complete, single traversal—either forward or backward—meaning that different steps from the same traversal. We then partitioned each dataset into labeled, unlabeled, and test sets. For the Cross-Covariate dataset, only data from the Normal Walking (NM) was used as the labeled set. For the Cross-Route dataset, data from the first two routes (R-1, R-2) formed the labeled set. In both scenarios, all remaining datas were split evenly: 50% for the unlabeled training set and 50% for the test set, guaranteeing no traversal-ID overlap between partitions. During the unlabeled training phase, the model had no access to user IDs or covariate-ID/route-ID labels, but it retained the traversal-ID information to group steps belonging to the same traversal.

Baselines. We compared FlowGait against four baseline methods to evaluate its performance:

- **Origin.** This is the origin FlowGait model pre-trained on the RDGait dataset [56]. Features are extracted directly without any fine-tuning. During testing, feature similarity is measured using the L2 distance.
- **Supervised.** Based on Origin, use labeled enrollment data for supervised training.
- **SSL (Own Cov.).** This semi-supervised learning (SSL) approach employs pseudo-labeling. For each evaluation, it is trained exclusively on unlabeled data from the specific target covariate (e.g., only "handbag" data is used when testing the "handbag" condition).
- **SSL (All Cov.)**. In this semi-supervised approach, the model is trained using pseudo-labeling on all available unlabeled data from the entire training set.

8.3 Overall Performance

The overall performance of our proposed method and the baseline methods is summarized in Table 2. Our proposed method, FlowGait (ours), consistently achieves the best performance across both challenging scenarios, reaching a Top-1 accuracy of 94.8% on the Cross-Covariates dataset and 98.6% on the Cross-Routes dataset.

The results reveal a clear and logical performance progression. The Origin FlowGait Model, which relies solely on pre-trained features for L2 distance comparison, establishes the baseline performance at 80.4% and 82.9% for the two scenarios, respectively. Supervised fine-tuning on the labeled data provides a notable improvement (83.6% and 87.7%), confirming the benefit of task-specific adaptation.

Table 2: A Comparison of Method Performance across Different Scenarios.

Method	Cross-Covariates	Cross-Routes
Origin	80.4%	82.9%
Supervised	83.6%	87.7%
SSL (Own Cov.)	86.4%	93.1%
SSL (All Cov.)	88.8%	94.4%
FlowGait (ours)	**94.8%**	**98.6%**

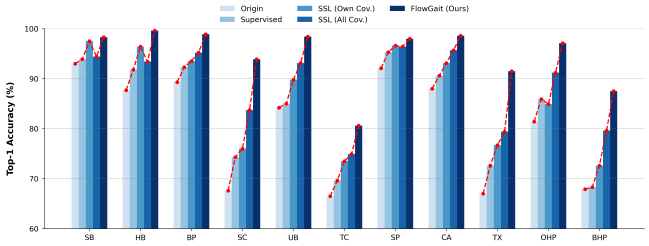


Figure 14: Performance for Cross-Covariates Gait Recognition.

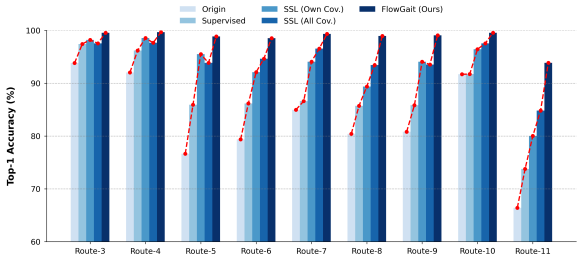


Figure 15: Performance for Cross-Routes Gait Recognition.

A significant leap in performance is achieved with semi-supervised learning (SSL). The SSL (Own Cov.) approach, which uses pseudo-labeling on unlabeled data from the same condition, substantially boosts accuracy to 86.4% and 93.1%. Furthermore, leveraging the full unlabeled data in SSL (All Cov.) provides an even greater advantage, pushing the accuracy to 88.8% and 94.4%. Finally, `FlowGait`, with its proposed two-stage step-traversal labeling algorithm, delivers the largest performance gain, validating the superior efficacy of our method in leveraging unlabeled data.

8.4 Performance Across Different Covariates

The gait recognition results for the Cross-Covariate dataset are detailed in Fig. 14. Our proposed method, `FlowGait`, demonstrates superior performance across all conditions, delivering the most significant gains in challenging scenarios.

Overall Performance. Our system consistently outperforms all baselines. On simpler covariates such as Shoulder Bag (SB), Handbag (HB), and Calling (CA), it achieved near-perfect accuracy rates, all exceeding 98%. More importantly, on difficult covariates like Suitcase (SC) and Texting (TX), our method boosted accuracy from baseline levels of 67% to over 91%, an improvement of more than 24 percentage points. Even on the most challenging conditions, Thick Coat (TC) and Both Hands in Pockets (BHP), it achieved robust accuracies of 80.6% and 87.5%, respectively.

The Value of Diverse Unlabeled Data. A key finding emerges from comparing the two semi-supervised learning (SSL) strategies. The SSL (All Cov.) method, which uses all available unlabeled data, consistently outperforms SSL (Own Cov.). This suggests the model benefits from an implicit "easy-to-hard" curriculum, where data from simpler covariates aids in learning features for more complex ones. This effect is most pronounced on difficult tasks like SC and BHP, where using all unlabeled data improved accuracy by over 7 percentage points compared to using only target-specific data. Conversely, for some simple covariates (e.g., SB, HB), this diverse training set caused a slight accuracy decrease. This highlights an expected trade-off: leveraging a wide range of data boosts overall generalization at the minor cost of specialization on a single, simple task, which in turn helps mitigate overfitting.

Analysis of Challenging Covariates. The difficulty of certain covariates stems from their significant biomechanical impact. Thick Coat (TC) severely restricts the natural movement of joints, while Both Hands in Pockets (BHP) forces unnatural, compensatory leg movements to maintain balance. It's worth noting that the BHP gait was described by most volunteers as awkward and unrepresentative of their normal walk. The TC scenario, however, points to a key strength of our approach. In reality, people's clothing choices are continuous, for example, from short sleeves to long sleeves, to sweaters, and then to heavy coats. In such continuous gait variations, our method can better learn new gaits progressively, which we will introduce in the Sec. 9.

Impact of the Step-Traversal Labeling Algorithm. Finally, the two-stage step-traversal labeling algorithm provides a consistent and significant performance boost. Its introduction improved accuracy by 2.4% to 9.4% across different covariates during the training phase. This enhancement carried over to the testing phase, providing a further improvement of 0.9% to 2.8%. This result validates the effectiveness of our hierarchical approach, which combines step-level evidence with traversal-level context for more robust label propagation.

8.5 Performance Across Different Route

As shown in Fig. 15, `FlowGait` (Ours) demonstrates outstanding performance and significant improvements across all 9 walking routes. It achieved average identification accuracies of 99.6% on straight-line routes (R-3, R-4, R-10), 99.0% on polyline routes (e.g., R-5, R-7, R-8), and 97.2% on curved routes (R-6, R-9, R-11), outperforming the baseline by 7.1%, 18.3%, and 21.7%, respectively. Notably, our system's accuracy exceeded 93% on all routes and surpassed 98% on 8 out of the 9 routes, confirming its superior generalization capabilities for cross-route recognition.

The primary challenge in cross-route recognition is the change in viewing angle and postural adjustments at turns, while a person's fundamental walking rhythm remains relatively consistent. Our system excels here because the labeled training routes (R-1 and R-2) expose the model to the velocity signatures of both straight and diagonal walking. Our two-stage step-traversal labeling algorithm then effectively leverages this knowledge. By identifying high-confidence steps within a new, unlabeled traversal, it propagates their labels to other, less certain steps from that same traversal, efficiently teaching the model to recognize a user's gait from a variety of new perspectives. The most challenging route was R-11 (93.8% accuracy), which involves frequent, sharp turns that disrupt body balance beyond simple angular changes. Nevertheless, our method still achieved a massive 25.9 percentage point improvement over the baseline on this difficult route.

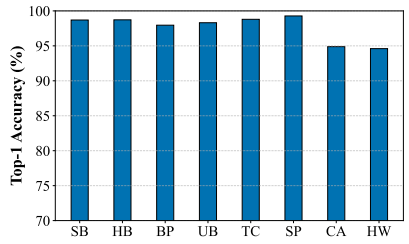


Figure 16: Evaluation on Elders Group.

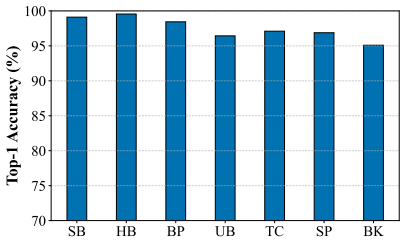


Figure 17: Evaluation on Children Group.

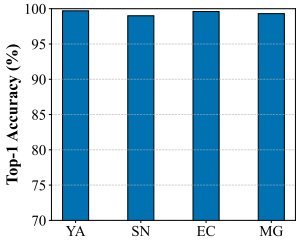


Figure 18: Evaluation on Families.

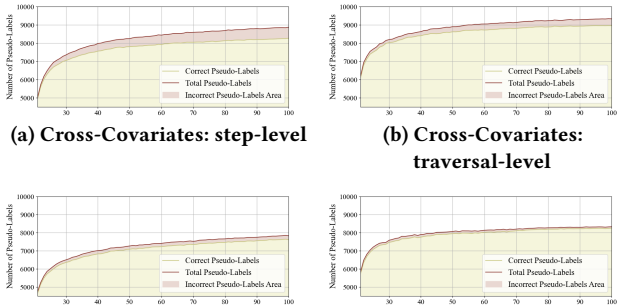
Importance of Labeled Routes. To validate the importance of the enrollment labeled data, we conducted an ablation study, training the model with only R-1 (straight) or only R-2 (diagonal) as the labeled set. As shown in Fig. 22, using both routes yielded a final accuracy of 98.6%. Using only the straight route (R-1) resulted in a significantly lower accuracy of 84.2%, while the diagonal route (R-2) alone achieved 90.2%. This confirms that both routes are crucial. The straight route (R-1) provides a stable anchor for a user's baseline gait features, while the diagonal route (R-2) is essential for learning features from varying perspectives. Therefore, we recommend that for any deployment in empty spaces, the enrollment data for a gait recognition system should include, at a minimum, samples from both straight and diagonal walking routes.

8.6 Generalization to Unseen Conditions

To evaluate the model's ability to generalize to entirely new scenarios, we conducted a leave-one-out cross-validation. In this setup, we iteratively held out one covariate or route as a test set, trained the model on all remaining conditions (both labeled and unlabeled), and then evaluated performance on the completely unseen held-out data. The model demonstrated strong generalization capabilities. In the leave-one-covariate-out tests, the average accuracy was 93.8%, and in the leave-one-route-out tests, it was 97.3%. These results represent only a minor performance drop of 1.0% and 1.3% compared to the main experiment, confirming that our system can effectively generalize to novel gait patterns not encountered during training.

8.7 Performance on Elderly and Child Subjects

To evaluate the generalization capability and robustness of the proposed system across different demographics, particularly for groups with distinct gait patterns, we conducted extensive cross-covariate experiments on elderly and child subjects. We recruited 10 elderly participants and 6 children to assess cross-covariate recognition performance. Our studies were approved by the Institutional Review Board (IRB) of our institution. The mean age and standard deviation were 59.5 ± 4.28 for the elderly group and 9.67 ± 1.75 for the children group. Adhering to human-centric principles, we surveyed the volunteers regarding their daily walking habits and adjusted the covariates accordingly. We excluded covariates that are uncommon for these groups, including Suitcase (SC), Texting (TX), One Hand in Pocket (OHP), and Both Hands in Pockets (BHP). Specifically, we added Walking while Holding Waist (HW) for the elderly; for the children, we additionally removed Calling (CA) and



(a) Cross-Covariates: step-level **(b) Cross-Covariates: traversal-level**

(c) Cross-Routes: step-level **(d) Cross-Routes: traversal-level**

Figure 19: Pseudo-Label Statistics over Epochs. The red region represents the incorrect pseudo-labels.

introduced Walking while Holding a Book(BK). The data collection scenarios and procedures remained consistent with those of the primary cross-covariate dataset. However, considering the limited physical stamina of the elderly and children, ample rest intervals were provided between sessions. The detailed Top-1 accuracy distributions are presented in Fig. 16 and Fig. 17.

As illustrated in Fig. 16, the algorithm demonstrates exceptional stability in the elderly group, achieving a mean accuracy of 97.7%. Under the six scenarios, including Satchel Bag (SB), Hand Bag (HB), Backpack (BP), Using Umbrella (UB), Thick Coat (TC), and Slippers (SP), the algorithm demonstrates exceptional recognition precision, with accuracies consistently maintained above 98%. Due to the limb deformation caused by "Walking while Holding Waist" and "Calling," the recognition accuracies are slightly lower, reaching 94.6% and 94.9%, respectively. These results attest to the system's high reliability in the daily activity scenarios of the elderly. As illustrated in Fig. 17, the algorithm also exhibits excellent performance in the child group, achieving a high mean recognition accuracy of 97.5%. Notably, compared to other age groups, carrying heavy objects (e.g., umbrellas and books) exerts the most significant impact on children. This is attributed to the limited physical strength of children, as carrying objects more easily disrupts their walking balance.

8.8 Performance Across Different Groups

To validate the recognition accuracy of `FlowGait` across different family compositions, we constructed simulated family units by randomly sampling subjects from the Cross-Covariate, Elderly, and

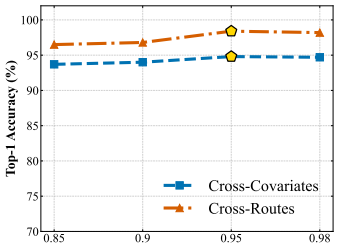


Figure 20: Evaluation on Different Confidence Thresholds.

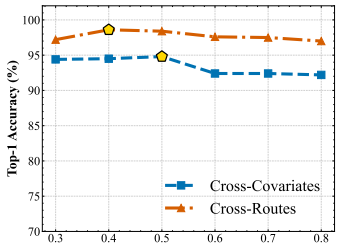


Figure 21: Evaluation on Different Traversal Thresholds.

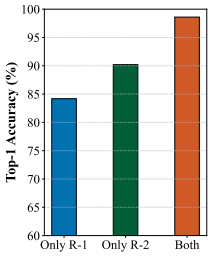


Figure 22: Accuracy across Different Labeled Datasets.

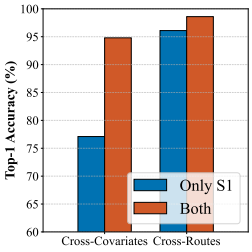


Figure 23: Accuracy across Different Pre-training Models.

Child datasets. Specifically, four distinct family types were defined: 1) Young-Adults (YA), consisting of 2 adults; 2) Standard-Nuclear (SN), comprising 2 adults and 2 children; 3) Elderly-Care (EC), including 2 adults and 2 elderly subjects; and 4) Multi-Gen (MG), containing 2 adults, 2 children, and 2 elderly subjects. The evaluation protocol followed the same metrics as the Cross-Covariate experiments. To ensure robustness, we performed five rounds of random sampling for each family type using different volunteers and reported the average accuracy. The results, as illustrated in Fig. 19, indicate that `FlowGait` achieves consistently high performance across all four family scenarios, with average accuracies exceeding 99%. Notably, these results surpass those obtained from individual group evaluations. This improvement is attributed to the significant inter-group gait differences—such as the slower walking speed of the elderly and the shorter stride length of children—which make distinguishing between members in a mixed group easier. These findings demonstrate the promising potential of our system for practical applications in smart home environments.

8.9 Ablation Study

8.9.1 Effecacy of Two-stage Labeling Algorithm. To validate our proposed two-stage labeling algorithm, we analyzed the quantity and quality of pseudo-labels it generated compared to a standard step-level-only approach. Fig. 19 visualizes the number of correct (yellow) and incorrect (red) pseudo-labels generated during each training epoch. The results clearly show that while both methods generate more pseudo-labels as training progresses, our two-stage labeling algorithm consistently produces more correct labels while simultaneously suppressing incorrect ones. For instance, at the 100th epoch of the cross-covariate task, our method increased the number of correct labels by 8.4% and reduced incorrect labels by 38.6%. The benefit was even more pronounced in the cross-route task, where our method yielded 8.3% more correct labels and 63.1% fewer incorrect ones. These statistics confirm that our two-stage labeling algorithm significantly improves the quality of the pseudo-labeled set, providing a more reliable signal for the self-training process.

8.9.2 Impact of Pre-training Data. To investigate the influence of the pre-trained model's quality, we compared two pre-training strategies: using the full RDGait dataset versus using only the Scenary 1 subset. As shown in Fig. 23, the choice of pre-training data significantly impacts performance, especially in the cross-covariate task. When pre-trained on Scenary 1, the model's accuracy dropped to 77.1% (a 17.7% decrease) on the cross-covariate task and 96.1% (a 2.5% decrease) on the cross-route task compared to pre-training on the full dataset. This result underscores that robust generalization, particularly across diverse covariates, is highly dependent on a large and varied pre-training dataset.

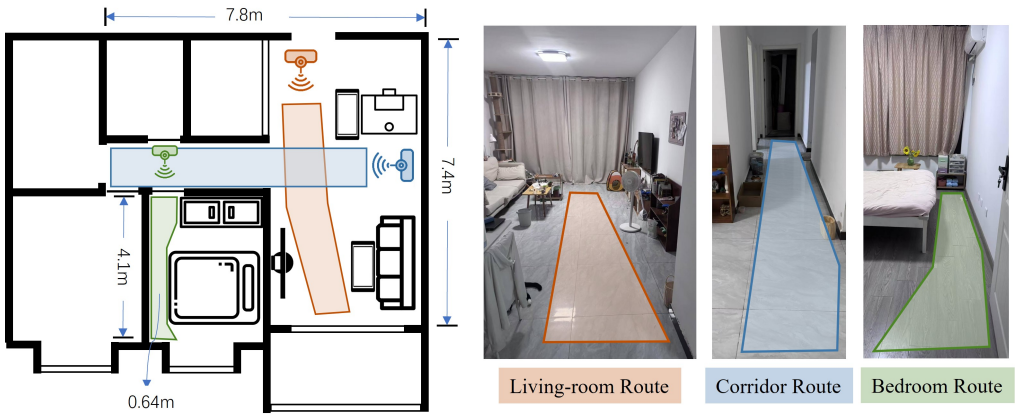
8.9.3 Impact of Key Parameters. We conducted an ablation study to evaluate the impact of the confidence threshold τ_c and the traversal threshold τ_p and determine their optimal values. The results are presented in Fig. 20 and Fig. 21. First, to find the optimal confidence threshold, we fixed the traversal threshold at 0.5 and tested τ_c values of 0.85, 0.9, 0.95, and 0.98. As shown in Fig. 20, the model achieved the best performance when τ_c was set to 0.95. Using this optimal confidence threshold, we then evaluated the impact of the traversal threshold τ_p , testing values from 0.3 to 0.8. The results (Fig. 21) show that the optimal τ_p is task-dependent. The highest accuracy was achieved with $\tau_p = 0.5$ for the cross-covariate task and $\tau_p = 0.4$ for the cross-route task. This suggests that a higher traversal threshold is better for more challenging tasks to prevent error propagation, while a lower threshold is advantageous for simpler tasks to include more unlabeled data.

9 Evaluation on Long-term Deployment

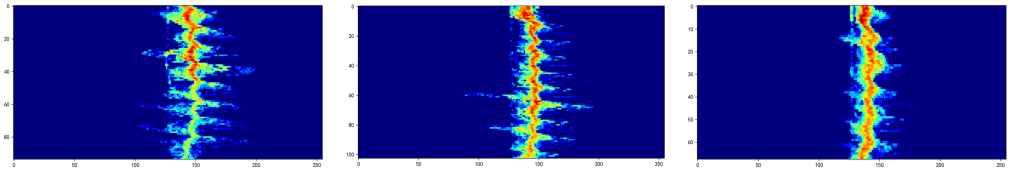
To validate `FlowGait`'s viability for real-world deployment, we performed a comprehensive evaluation targeting its long-term performance, security, and user experience. This evaluation comprised three parts. First, we collected a two-week longitudinal gait dataset to assess the stability of the system's recognition accuracy over time and to quantify the computational overhead (i.e., training time and latency). Second, we evaluated its efficacy in detecting abnormal users. Finally, we gathered feedback from participants through a user study.

9.1 Dataset

We collected a Cross-Day dataset by collecting data from 7 participants (3 male, 4 female) over two weeks in a real-house setting. Our studies were approved by the Institutional Review Board (IRB) of our institution. As depicted in Fig. 24(a), the experimental area featured three distinct routes: a spacious, 6-meter curved **living room route**; a straight 7-meter **corridor route**; and a narrow 4-meter **bedroom route**, with a width of just 0.64 meters beside the bed.



(a) Experimental Equipment and Scenarios for the Cross-Day Dataset.



(b) Micro-Doppler Spectrum for Different Rooms. The x-axis is Doppler-velocity bins, and the y-axis is the number of frames.

Figure 24: Illustration of the experimental environment and corresponding data of the Cross-day dataset.

Each participant was asked to collect data on 10 separate days over two weeks. The protocol required a four-minute walk per route on the initial day, with the duration shortened to two minutes for all subsequent days. To capture naturalistic walking patterns, participants received no specific instructions on their walking style. Some volunteers participated in data collection after a workout or Pilates session, others after their baby had fallen asleep, and some right after returning from work. As shown in the micro-Doppler spectrograms in Fig. 24(b), a single traversal along the living room and corridor routes typically took about 100 frames (approx. 8 steps), whereas the shorter bedroom route took about 70 frames (approx. 5 steps). Notably, the spectrogram for the narrow bedroom route appears more cluttered.

9.2 Baselines

To simulate a realistic deployment scenario, we designated the data from the first day as the sole source of the labeled dataset. Our model was pre-trained on the RDGait dataset[56]. We proceeded to evaluate the following methods:

- **Supervised**. A baseline model was fine-tuned on the labeled data from Day 1.
- **`FlowGait`**. Our method. For each day's test (on Day N), it performs semi-supervised training using the labeled data from Day 1 and the unlabeled data from Days 2 to N-1.
- **`FlowGait-Continuous`**. `FlowGait` using the continuous learning pipeline.

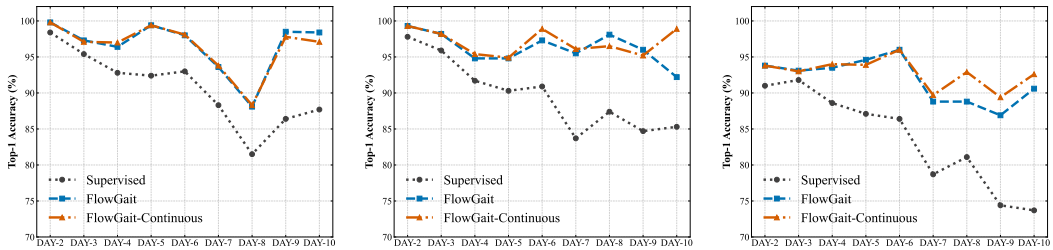
9.3 Performence

9.3.1 Overall Performance. The overall performance of our proposed model compared to the baseline is detailed in Table 3, with day-by-day results shown in Fig. 25. Our model achieved average recognition accuracies of 96.5%, 97.1%, and 92.8% for the living room, corridor, and bedroom routes, respectively, outperforming the baseline by 6.0%, 7.4%, and 9.2%. These results confirm our model's superior accuracy and stability in long-term operation. A significant performance degradation was observed for the baseline method over the two weeks. Specifically, its accuracy declined from 98.4% to 86.4% (living room), 97.8% to 84.7% (corridor), and 91.0% to 73.7% (bedroom). In stark contrast, our method achieves a high level of performance, posting Day-10 accuracies of 97.1%, 98.6%, and 92.6%. `FlowGait` reduced performance decay from 13.6% to just 1.4% for average.

To statistically validate these trends, we performed a Mann-Kendall Trend Test (Table 4). The baseline method exhibited a statistically significant downward trend, with p-values of 0.0091,

Table 3: Comparison of Gait Recognition Accuracy in Different Routes.

Method	Living room	Corrider	Bedroom
Supervised	90.6%	89.7%	83.6%
`FlowGait`	**96.6%**	**96.2%**	92.3%
`FlowGait-Continuous`	96.5%	**97.1%**	**92.8%**



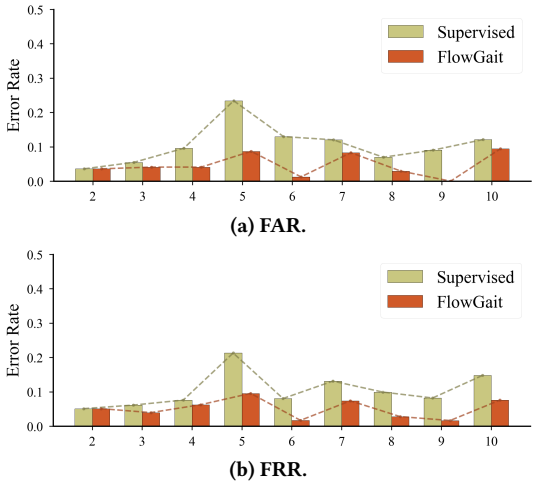
(a) Top-1 Accuracy of Living-room. **(b) Top-1 Accuracy of Corrider.** **(c) Top-1 Accuracy of Restroom.**

Figure 25: Overall Gait Recognition Accuracy in Different Routes.

Table 4: Mann-Kendall Trend Test Results for Different Routes and Methods.

	Living-room Route		Corridor Route		Bedroom Route	
Method	*p*-value	Slope	*p*-value	Slope	*p*-value	Slope
Supervised	**0.0091**	−1.5000	**0.0091**	−1.6833	**0.0012**	−2.3607
`FlowGait`	0.4655	−0.1929	0.2084	−0.4524	0.2084	−0.8167
`FlowGait-Continuous`	0.2084	−0.3688	0.6750	−0.0750	0.1753	−0.1917

Note: Slope refers to the Theil-Sen slope estimator. Bold values indicate a statistically significant trend (p < 0.05).



(a) FAR. **(b) FRR.**

Figure 26: FAR and FRR for Abnormal User Detection across Various Days.

0.0091, and 0.0012 for the three scenarios (all $p < 0.05$). Conversely, our model showed no significant trend, with p-values between 0.1 and 0.4. Furthermore, by analyzing the slope of the performance curves, we found that our method reduced the rate of degradation by 4.06-fold, 22.4-fold, and 12.3-fold across the three routes, yielding an average 12.9-fold mitigation in performance decay.

9.3.2 Training Time and Latency. To assess the system's viability for edge deployment, we benchmarked its computational efficiency on a representative consumer-grade laptop equipped with an AMD Ryzen 7 7435H CPU, 32GB of RAM, and an NVIDIA GeForce RTX 4060 Laptop GPU (8GB VRAM). For the training phase, we configured a batch size of 128, utilized NVIDIA's Automatic Mixed Precision (AMP) for acceleration, and set the coreset to 600 samples per ID. The results for each stage are presented in Fig. 27. We observed that the coreset construction time was negligible, averaging just 8.0 seconds per session. The main training time initially increased but quickly stabilized between 290 and 390 seconds. This plateau is attributed to the coreset reaching saturation around the fourth day, after which the daily training set size remains relatively constant. Additionally, during the training process, the average memory usage was 300M. Our model size is 10.9M. The system's real-time processing performance was also benchmarked on the same hardware, with latency measured using a batch size of one. The results, illustrated in Fig. 28, show a mean data preprocessing latency of 33.53 ms per frame and a model inference latency of 7.16 ms per sample. This low overall latency validates the system's capability for deployment in real-time applications.

9.3.3 Abnormal User Detection. The system's capability for abnormal user detection was evaluated using data collected exclusively from the corridor route. For each experimental setup, four participants were randomly assigned as authorized users and one as an unauthorized user. The evaluation protocol simulated a day-by-day operational scenario. To assess performance on a given Day N, the model was first trained on the cumulative unlabeled data from the four authorized users, spanning from Day 1 to Day N-1. Subsequently, the trained model was tested on the Day N data from all five participants. To ensure the statistical robustness of our results, this entire validation process was repeated five times for each day, each with a different random selection of users. Performance was quantified using two standard metrics: the False Acceptance Rate (FAR)—the proportion of unauthorized users accepted—and the False Rejection Rate (FRR)—the proportion of authorized users rejected. As illustrated in Fig. 26, our method maintained average

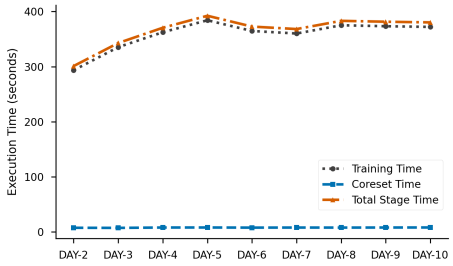


Figure 27: Daily Training Time.

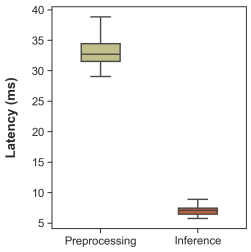


Figure 28: System Latency.

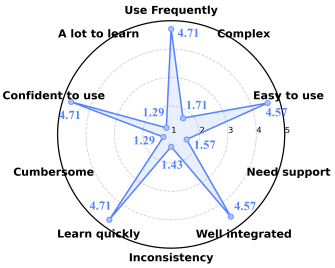


Figure 29: SUS Results.

FAR and FRR values below 5%. This performance represents a 2.24-fold reduction in error rates compared to the supervised baseline, demonstrating a substantial enhancement in system security.

9.3.4 User Study. We evaluated the usability of FlowGait using the System Usability Scale (SUS) [7], a standardized 10-item questionnaire. The system achieved an overall SUS score of 89.95 out of 100, which is considered "excellent" (a score $\geq$ 85 [5]). This high score is supported by volunteers' ratings, as shown in Fig. 29. Participants found the system easy to use (4.57 ± 0.73), well-integrated (4.57 ± 0.49), and suitable for frequent use (4.71 ± 0.70). Furthermore, the straightforward single-enrollment process contributed to a high learnability score (4.71 ± 0.45), and users expressed strong confidence in the system (4.71 ± 0.45).

User Acceptance. To evaluate the acceptance of FlowGait, we conducted interviews with a diverse group of potential users. We provided a detailed explanation of the system's working principles and its potential applications in smart home interactions. Their feedback is presented below:

> **Mother**: *"This system would be helpful for monitoring my child's safety and activities when I am away."*
>
> **Senior**: *"A tool that can monitor the daily gait health of my partner and myself is exactly what we need."*
>
> **Engineer**: *"The 'follow-me' concept is impressive. Being able to trigger commands from any location makes the environment feel truly personalized."*
>
> **Student**: *"Living in a shared apartment makes me wary of strangers. I strongly prefer a system that ensures identification while preserving privacy."*

The feedback affirms the practical utility of FlowGait for security and personalization. In particular, the responses from the mother and senior highlight its value in assisted living contexts. In summary, FlowGait shows strong potential for adoption in real-world environments.

Table 5: User feedback and responses.

User Concerns	Response
"Is the radiation from this device harmful to my health?"	**Safety Compliance.** mmWave is non-ionizing radiation. Our devices operate at milliwatt-level transmission power, remaining far below the strict exposure limits imposed by international safety guidelines (e.g., ICNIRP [27], FCC [21]). Thus, they pose no known health risks.
"What if a hacker breaks in? Can they steal my gait data?"	**Local Processing.** Our system adopts an edge-computing architecture. It processes data entirely on-device without requiring cloud uploads. Consequently, raw, high-dimensional mmWave data never leaves the user's private space.
"How can I be sure this technology won't be used against me (e.g., for surveillance)?"	**Purpose Limitation & Control.** We strictly adhere to purpose limitation principles. Data usage is restricted solely to the specific purpose consented to by the user (e.g., authentication). Secondary usage—such as employee surveillance or insurance risk assessment—is strictly prohibited. Furthermore, users retain full control via a physical "off-switch" and the right to permanently delete their data.

Feedback and Concerns. During the interviews, participants raised valid concerns regarding health, data privacy, and ethical implications. In response, we explained that we ensure physical safety by utilizing low-power, non-ionizing mmWave radiation that strictly adheres to international standards. To protect privacy and prevent misuse, the system employs on-device edge processing to keep data local and enforces rigorous purpose limitations, granting users full control via physical switches and data deletion rights. We summarize these inquiries and our detailed responses in Table 5.

10 Discussion and Future Work

Dynamic Environments: Environmental variables can also impact gait recognition performance. For example, different furniture layouts may alter walking trajectories and occlude signals, just as the sensor's placement (height and angle) affects the captured reflections. Our initial findings indicate that FlowGait is robust to these environmental and positional changes. We will explore in detail in future work.

Multi-User identification: FlowGait currently only supports single-person gait recognition, but it can be naturally extended to multi-person scenarios. By using multi-antenna data for angle estimation, the gaits of different users can be separated in the three dimensions of range, velocity, and angle. However, due to resolution limitations, it still cannot handle close-proximity situations, such as multiple people walking side-by-side. Furthermore, multi-person tracking introduces additional complexities, as interactions can alter gait patterns through phenomena like cadence synchronization and phase locking.

Large-Scale Pre-training Dataset: Motivated by our finding that pre-training on large, diverse datasets significantly enhances model

robustness and performance, our future work will focus on developing an extensive, multi-modal gait dataset. To this end, we plan to deploy automated collection systems in public areas, integrating both cameras and mmWave radar. A state-of-the-art vision model will facilitate automatic annotation, enabling the creation of a rich repository of naturalistic, intervention-free gait data to further advance research in this domain.

Longitudinal Gait Health Monitoring: A promising application for `FlowGait` is long-term gait monitoring for health. As a key health indicator, gait changes in elders can signal an increased risk of falls and other adverse events. We envision `FlowGait` as a non-intrusive tool that tracks gait parameters over time to provide early warnings of significant deviations. This could flag underlying health issues and allow for earlier intervention.

11 Conclusion

In this work, we present `FlowGait`, a self-learning millimeter-wave gait recognition framework designed for robust, long-term deployment in real-world environments. Our system pioneers a transformer-based network for feature extraction that outperforms state-of-the-art methods, coupled with a novel two-stage step-walk labeling algorithm for accurate semi-supervised adaptation. Extensive evaluations on a COTS mmWave radar across three challenging datasets demonstrate its superior performance. Using minimal initial labeled data, `FlowGait` achieved recognition accuracies of 94.8%, 97.9%, and 96.6%. Critically, it mitigated long-term performance decay by a factor of 12.9, proving its viability for continuous operation in dynamic, real-world settings. We hope this work will contribute to the development of adaptive biometric systems and inspire further research into long-term deployment in real-world settings.

Acknowledgments

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

References

[1] M Umair Bin Altaf, Taras Butko, and Biing-Hwang Juang. 2015. Acoustic gaits: Gait analysis with footstep sounds. *IEEE Transactions on Biomedical Engineering(TBME)* 62, 8 (2015), 2001–2011.

[2] Noah Apthorpe, Yan Shvartzshnaider, Arunesh Mathur, Dillon Reisman, and Nick Feamster. 2018. Discovering smart home internet of things privacy norms using contextual integrity. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 2, 2 (2018), 1–23.

[3] Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lučić, and Cordelia Schmid. 2021. Vivit: A video vision transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*. 6836–6846.

[4] Augusto Aubry, Antonio De Maio, Vincenzo Carotenuto, and Alfonso Farina. 2016. Radar phase noise modeling and effects-part I: MTI filters. *IEEE Trans. Aerospace Electron. Systems* 52, 2 (2016), 698–711.

[5] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[6] Kristen M Beavers, Daniel P Beavers, Denise K Houston, Tamara B Harris, Trisha F Hue, Annemarie Koster, Anne B Newman, Eleanor M Simonsick, Stephanie A Studenski, Barbara J Nicklas, et al. 2013. Associations between body composition and gait-speed decline: results from the Health, Aging, and Body Composition study. *The American journal of clinical nutrition* 97, 3 (2013), 552–560.

[7] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[8] 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.

[9] Tianrui Chai, Annan Li, Shaoxiong Zhang, Zilong Li, and Yunhong Wang. 2022. Lagrange motion analysis and view embeddings for improved gait recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20249–20258.

[10] Hanqing Chao, Kun Wang, Yiwei He, Junping Zhang, and Jianfeng Feng. 2021. GaitSet: Cross-view gait recognition through utilizing gait as a deep set. *IEEE transactions on pattern analysis and machine intelligence* 44, 7 (2021), 3467–3478.

[11] 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.

[12] 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* 7, 3 (2023), 1–29.

[13] 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.

[14] Eun Kyoung Choe, Sunny Consolvo, Jaeyeon Jung, Beverly Harrison, and Julie A Kientz. 2011. Living in a glass house: a survey of private moments in the home. In *Proceedings of the 13th international conference on Ubiquitous computing*. 41–44.

[15] Fani Deligianni, Yao Guo, and Guang-Zhong Yang. 2019. From emotions to mood disorders: A survey on gait analysis methodology. *IEEE journal of biomedical and health informatics* 23, 6 (2019), 2302–2316.

[16] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

[17] 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.

[18] Chao Fan, Saihui Hou, Yongzhen Huang, and Shiqi Yu. 2023. Exploring deep models for practical gait recognition. *arXiv preprint arXiv:2303.03301* (2023).

[19] 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)*.

[20] Yuxin Fan, Yong Wang, Hang Zheng, and Zhiguo Shi. 2024. Video2mmPoint: Synthesizing mmWave Point Cloud Data From Videos for Gait Recognition. *IEEE Sensors Journal* (2024).

[21] Federal Communications Commission (FCC). 2021. RF Safety Guidelines. [Online]. https://www.fcc.gov/general/radio-frequency-safety-0.

[22] Stephanie Grobe, Rumit Singh Kakar, Matthew Lee Smith, Ranjana Mehta, Timothy Baghurst, and Ali Boolani. 2017. Impact of cognitive fatigue on gait and sway among older adults: A literature review. *Preventive medicine reports* 6 (2017), 88–93.

[23] Chenming He, Chengzhen Meng, Chunwang He, Xiaoran Fan, Beibei Wang, Yubo Yan, and Yanyong Zhang. 2024. See Through Vehicles: Fully Occluded Vehicle Detection with Millimeter Wave Radar. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 740–754.

[24] Chenming He, Rui Xia, Chengzhen Meng, Xiaoran Fan, Dequan Wang, Haojie Ren, Jianmin Ji, and Yanyong Zhang. 2025. Ghost Points Matter: Far-Range Vehicle Detection with a Single mmWave Radar in Tunnel. arXiv:2509.06639 [cs.NI] https://arxiv.org/abs/2509.06639

[25] Jorunn L Helbostad, Sara Leirfall, Rolf Moe-Nilssen, and Olav Sletvold. 2007. Physical fatigue affects gait characteristics in older persons. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 62, 9 (2007), 1010–1015.

[26] Yanyan Huang, Yong Wang, Kun Shi, Chaojie Gu, Yu Fu, Cheng Zhuo, and Zhiguo Shi. 2023. HDNet: Hierarchical dynamic network for gait recognition using millimeter-wave radar. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1–5.

[27] International Commission on Non-Ionizing Radiation Protection (ICNIRP). 2020. Guidelines for Limiting Exposure to Electromagnetic Fields (100 kHz to 300 GHz). [Online]. https://www.icnirp.org/cms/upload/publications.

[28] Xin Jin, Tianyu He, Kecheng Zheng, Zhiheng Yin, Xu Shen, Zhen Huang, Ruoyu Feng, Jianqiang Huang, Zhibo Chen, and Xian-Sheng Hua. 2022. Cloth-changing person re-identification from a single image with gait prediction and regularization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 14278–14287.

[29] 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.

[30] Jiarong Li, Zihan Wang, Zihao Zhao, Yuchao Jin, Jihong Yin, Shao-Lun Huang, and Jiyu Wang. 2021. TriboGait: 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.

[31] Xiang Li, Yasushi Makihara, Chi Xu, Yasushi Yagi, and Mingwu Ren. 2020. Gait recognition via semi-supervised disentangled representation learning to identity and covariate features. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 13309–13319.

[32] 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)*.

[33] 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.

[34] Alan Liu, Yu-Tai Lin, and Karthikeyan Sundaresan. 2024. View-agnostic Human Exercise Cataloging with Single MmWave Radar. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 3 (2024), 1–23.

[35] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, Nov (2008), 2579–2605.

[36] Chengzhen Meng, Yifan Duan, Chenming He, Dequan Wang, Xiaoran Fan, and Yanyong Zhang. 2024. mmplace: Robust place recognition with intermediate frequency signal of low-cost single-chip millimeter wave radar. *IEEE Robotics and Automation Letters* 9, 6 (2024), 4878–4885.

[37] Chengzhen Meng, Chenming He, Dequan Wang, Yuxuan Xiao, Lingyu Wang, Xiaoran Fan, Lu Zhang, and Yanyong Zhang. 2025. Gr-fall: A fall detection system with gait recognition for indoor environments using siso mmwave radar. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 3 (2025), 1–26.

[38] 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.

[39] Zhen Meng, Anfu Zhou, Huadong Ma, Qian Zhang, et al. 2024. CRNet: Robust Millimeter Wave Gait Recognition Method Based on Contrastive Learning. *APSIPA Transactions on Signal and Information Processing* 13, 4 (2024).

[40] Johannes Michalak, Nikolaus F Troje, Julia Fischer, Patrick Vollmar, Thomas Heidenreich, and Dietmar Schulte. 2009. Embodiment of sadness and depression—gait patterns associated with dysphoric mood. *Psychosomatic medicine* 71, 5 (2009), 580–587.

[41] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.

[42] Edwin Pan, Jingning Tang, Dash Kosaka, Ruihao Yao, and Arjun Gupta. 2019. OpenRadar. https://github.com/presenseradar/openradar.

[43] 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).

[44] Jacopo Pegoraro, Francesca Meneghello, 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.

[45] Ekkasit Pinyoanuntapong, Ayman Ali, Kalvik Jakkala, Pu Wang, Minwoo Lee, Qucheng Peng, Chen Chen, and Zhi Sun. 2023. Gaitsada: Self-aligned domain adaptation for mmwave gait recognition. In *2023 IEEE 20th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*. IEEE, 218–226.

[46] Akarsh 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.

[47] 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.

[48] Yiyang Su, Minchul Kim, Feng Liu, Anil Jain, and Xiaoming Liu. 2024. Open-set biometrics: Beyond good closed-set models. In *European Conference on Computer Vision*. Springer, 243–261.

[49] 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.

[50] TI. 2019. DCA1000 evaluation module for real-time data capture and streaming. https://www.ti.com.com/tool/DCA1000EVM.

[51] TI. 2021. IWR6843 intelligent mmWave sensor standard antenna plug-in module. https://www.ti.com/tool/IWR6843ISK.

[52] Philip van Dorp and FCA Groen. 2008. Feature-based human motion parameter estimation with radar. *IET Radar, Sonar & Navigation* 2, 2 (2008), 135–145.

[53] 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.

[54] Chunyu Wang, Peixian Gong, and Lihua Zhang. 2022. Stpointgcn: Spatial temporal graph convolutional network for multiple people recognition using millimeter-wave radar. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 3433–3437.

[55] Dequan Wang, Yifan Duan, Xiaoran Fan, Chengzhen Meng, Jianmin Ji, and Yanyong Zhang. 2022. Maroam: Map-based radar slam through two-step feature selection. *arXiv preprint arXiv:2210.13797* (2022).

[56] Dequan Wang, Xinran Zhang, Kai Wang, Lingyu Wang, Xiaoran Fan, and Yanyong Zhang. 2024. Rdgait: A mmwave based gait user recognition system for complex indoor environments using single-chip radar. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies(IMWUT)* 8, 3 (2024), 1–31.

[57] Lingyu Wang, Kai Wang, Dequan Wang, You Zuo, Chenming He, Chengzhen Meng, Xiaoran Fan, Haojie Ren, and Yanyong Zhang. 2025. mmMUSE: An mmWave-based Motion-resilient Universal Speech Enhancement System. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 4 (2025), 1–33.

[58] Lingyu Wang, You Zuo, Dequan Wang, Chenming He, Chengzhen Meng, Xinran Zhang, Xiaoran Fan, and Yanyong Zhang. 2026. FeelWave: Enabling Emotion-Aware Voice Interaction through Noise-Robust mmWave Emotion Sensing. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. ACM New York, NY, USA, 24 pages.

[59] 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.

[60] Greg Welch, Gary Bishop, et al. 1995. An introduction to the Kalman filter. (1995).

[61] Daniel K White, Tuhina Neogi, Michael C Nevitt, Christine E Peloquin, Yanyan Zhu, Robert M Boudreau, Jane A Cauley, Luigi Ferrucci, Tamara B Harris, Susan M Satterfield, et al. 2013. Trajectories of gait speed predict mortality in well-functioning older adults: the Health, Aging and Body Composition study. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* 68, 4 (2013), 456–464.

[62] Chi Xu, Yasushi Makihara, Xiang Li, Yasushi Yagi, and Jianfeng Lu. 2020. Cross-view gait recognition using pairwise spatial transformer networks. *IEEE Transactions on Circuits and Systems for Video Technology* 31, 1 (2020), 260–274.

[63] 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.

[64] Dawei Yan, Panlong Yang, Fei Shang, Feiyu Han, Yubo Yan, and Xiang-Yang Li. 2025. Pushing the Limits of WiFi-Based Gait Recognition Towards Non-Gait Human Behaviors. *IEEE Transactions on Mobile Computing* 24, 7 (2025), 6137–6153. doi:10.1109/TMC.2025.3540863

[65] Dawei Yan, Panlong Yang, Fei Shang, Feiyu Han, Yubo Yan, and Xiang-Yang Li. 2025. Pushing the Limits of WiFi-Based Gait Recognition Towards Non-Gait Human Behaviors. *IEEE Transactions on Mobile Computing* (2025).

[66] 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.

[67] Dingqiang Ye, Chao Fan, Jingzhe Ma, Xiaoming Liu, and Shiqi Yu. 2024. Biggait: Learning gait representation you want by large vision models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 200–210.

[68] Jin Zhang, Zhuangzhuang Chen, Chengwen Luo, Bo Wei, Salil S Kanhere, and Jianqiang Li. 2022. MetaGanFi: Cross-domain unseen individual identification using WiFi signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–21.

[69] Jie Zhang, Zhanyong Tang, Meng Li, Dingyi Fang, Petteri Nurmi, and Zheng Wang. 2018. CrossSense: Towards cross-site and large-scale WiFi sensing. In *Proceedings of the 24th annual international conference on mobile computing and networking*. 305–320.

[70] Lei Zhang, Cong Wang, and Daqing Zhang. 2021. Wi-PIGR: Path independent gait recognition with commodity Wi-Fi. *IEEE Transactions on Mobile Computing* 21, 9 (2021), 3414–3427.

[71] Yi Zhang, Yue Zheng, Guidong Zhang, Kun Qian, Chen Qian, and Zheng Yang. 2021. GaitSense: Towards ubiquitous gait-based human identification with Wi-Fi. *ACM Transactions on Sensor Networks (TOSN)* 18, 1 (2021), 1–24.

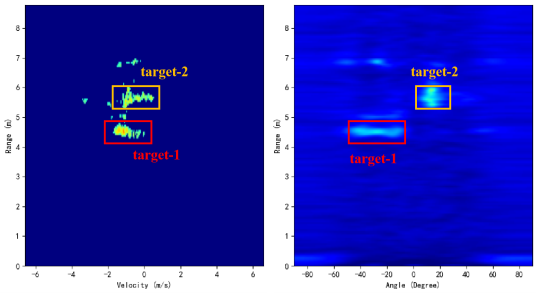
[72] Langcheng Zhao, Rui Lyu, Hang Lei, Qi Lin, Anfu Zhou, Huadong Ma, Jingjia Wang, Xiangbin Meng, Chunli Shao, Yida Tang, et al. 2024. AirECG: Contactless electrocardiogram for cardiac disease monitoring via mmWave sensing and cross-domain diffusion model. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 3 (2024), 1–27.

[73] Jinkai Zheng, Xinchen Liu, Boyue Zhang, Chenggang Yan, Jiyong Zhang, Wu Liu, and Yongdong Zhang. 2024. It takes two: Accurate gait recognition in the wild via cross-granularity alignment. In *Proceedings of the 32nd ACM International Conference on Multimedia*. 8786–8794.

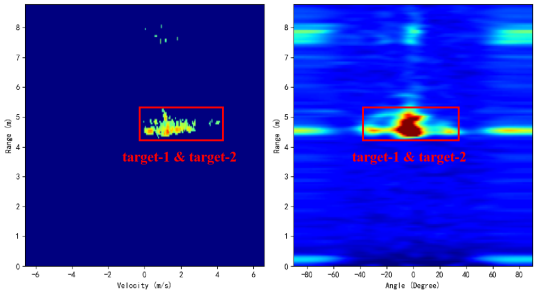
[74] Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty. 2003. Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the 20th*

International conference on Machine learning (ICML-03). 912–919.

[75] Shinan Zou, Chao Fan, Jianbo Xiong, Chuanfu Shen, Shiqi Yu, and Jin Tang. 2024. Cross-covariate gait recognition: A benchmark. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 7855–7863.



(a) General Scenario.



(b) Walking Side-by-Side.

Figure 30: Range-Doppler Heatmaps and Range-Azimuth Heatmaps for Multi-user Scenarios.

A Appendix: Feasibility Analysis for Multi-user Scenarios

In this section, we evaluate scenarios where multiple users coexist in the same space to demonstrate the applicability of the mmWave radar gait recognition system in multi-user environments. Fig. 30a illustrates the Range-Doppler (RD) and Range-Azimuth (RA) heatmaps in a scenario where two distinct users coexist. The angles in the RA heatmap are computed using Capon beamforming. When the two targets are non-overlapping across the range, velocity, and azimuth dimensions, we can employ an algorithm to separate them. Specifically, when the targets do not overlap in the range-velocity dimension, we directly perform clustering on the Range-Doppler heatmap to separate the multiple targets. Conversely, when there is overlap in the range-velocity dimension, but the targets are distinct in the azimuth dimension, we first perform clustering in the angular dimension to separate their spectra before reverting to the Range-Doppler domain. The subsequent steps remain consistent with those described in the paper.

Prior studies have sought to address the challenge of separating targets in close proximity, employing techniques such as constrained cluster re-partitioning[56]. However, challenges persist in multi-user scenarios. The first arises in the 'side-by-side' walking case shown in Fig. 30b, where signals from two targets merge due to spectral leakage and the limited angular resolution of mmWave radars (approx. 15° for a 2Tx-4Rx horizontal configuration). The second involves severe occlusion, where the signal from a rear user is obstructed by a front user, resulting in signal loss and identification failure. While these scenarios are common in real-world applications, they fall outside the scope of this paper and will be discussed in future work.

Type	Covariate	Abbr.	Key Biomechanical/Behavioral Characteristics
Carrying Items	Shoulder Bag	SB	Asymmetrical load; compensated by contralateral trunk lean and altered arm swing.
	Handbag	HB	Unilateral load restricts arm swing; induces subtle compensatory torso shifts.
	Backpack	BP	Symmetrical posterior load; causes a compensatory forward lean of the trunk.
	Suitcase	SC	Torsional load on the trunk from a fixed pulling arm; alters rotational dynamics.
	Umbrella	UB	Fixed, elevated arm eliminates swing; gait is subject to aerodynamic forces.
Clothing and Footwear	Thick Coat	TC	Added mass and bulk constrain motion; results in shorter stride and reduced arm swing.
	Slippers	SP	Poor foot fixation; leads to a shuffling motion with shorter strides to keep them on.
Phone-related Actions	Calling	CA	Cognitive load (dual-tasking); reduces speed, increases variability, and lowers awareness.
	Texting	TX	Downward head posture and lack of visual feedback cause a very slow, cautious gait.
Body Posture	One Hand in Pocket	OHP	Unilateral arm swing restriction; compensated by increased trunk and pelvis rotation.
	Both Hands in Pockets	BHP	Bilateral arm swing restriction; reduces dynamic stability, requiring more core/hip motion.
	Normal Walking	NM	Symmetrical, unconstrained gait with coordinated, reciprocal limb movement.

Table 6: Biomechanical and Behavioral Characteristics under Different Conditions