XGait: Cross-Modal Translation via Deep Generative Sensing for RF-based Gait Recognition

Huanqi Yang[1,2], Mingda Han[3], Mingda Jia[4], Zehua Sun[1,2], Pengfei Hu[3], Yu Zhang[5], Tao Gu[5], Weitao Xu[1,2]*
[1]City University of Hong Kong Shenzhen Research Institute,
[2]City University of Hong Kong, [3]Shandong University,
[4]Xi'an Jiaotong University, [5]Macquarie University

ABSTRACT

Radio Frequency (RF)-based gait recognition has emerged as a promising technology to authenticate individuals in a pervasive and unobtrusive way. However, a fundamental challenge remains in collecting extensive data of the same user in the same environment. To address this challenge, this paper introduces XGait, a cross-modal gait recognition framework that does not require the prior deployment of RF devices or explicit data collection. The key idea is to leverage the signals of the Inertial Measurement Unit (IMU), which is widely available in modern mobile devices, to simulate the RF signals that would be generated if the same person walked near RF devices. Despite the straightforward idea, several technical challenges need to be addressed due to the diversity of RF devices, the intrinsic difference between IMU signals and RF signals, and the complexity of gait. First, we propose an RF spectrogram generation method to consistently extract essential RF gait data features across different RF signals. Secondly, we propose a generative network-enabled IMU-to-RF translation approach that accurately converts IMU data to RF data. Finally, we design an RF gait spectrogram-specific transformer model to further improve the recognition performance. We conduct a comprehensive evaluation of XGait, involving thirty subjects in three different environments, utilizing three RF devices and seven mobile devices. Experimental results show that XGait consistently achieves over 99% Top-3 accuracy in various scenarios.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing**.

KEYWORDS

Gait Recognition, RF Sensing, Generative Model

ACM Reference Format:
Huanqi Yang[1,2], Mingda Han[3], Mingda Jia[4], Zehua Sun[1,2], Pengfei Hu[3], Yu Zhang[5], Tao Gu[5],, Weitao Xu[1,2]*. 2023. XGait: Cross-Modal Translation via Deep Generative Sensing for RF-based Gait Recognition . In *ACM Conference on Embedded Networked Sensor Systems (SenSys '23), November 12–17, 2023, Istanbul, Turkiye.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3625687.3625792

1 INTRODUCTION

1.1 Background and Motivation

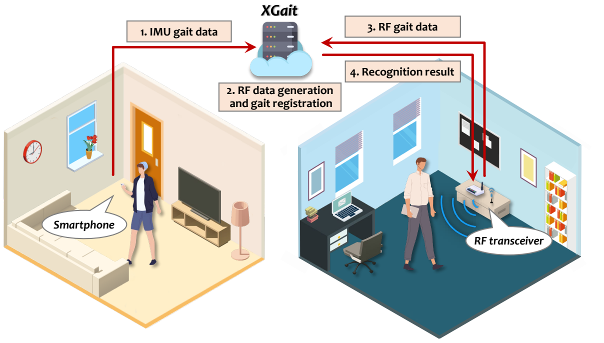
Radio Frequency (RF)-based gait recognition has attracted extensive attention due to its ability to authenticate individuals in a non-intrusive and device-free manner. This technology exploits the fact that each person has a unique walking pattern, and RF signals can capture these differences. The applications of RF-based gait recognition include smart buildings, personalized services, and monitoring/surveillance applications [10]. Numerous studies have explored the potential of different RF signals, such as Wi-Fi [70, 74, 81] and mmWave [42, 68].

Despite its promise, RF-based gait recognition, like many other RF-based sensing tasks, faces a fundamental limitation that necessitates extensive training using prior instances of individuals walking in the same area [51, 57, 70, 81]. This constraint poses challenges for the practical implementation of this technology in real-world applications. For instance, in monitoring or surveillance scenarios, an RF-based gait recognition system may struggle to accurately identify individuals who have not previously walked in the monitored area and undergone on-site registration. To address this issue, recent works have explored the use of few-shot learning and domain adaptation to alleviate the burden of training data, primarily focusing on minimizing the required training instances [11, 13, 21, 50, 58, 73]. Although they have shown promising results, it is important to note that they still require prior RF data collection with two key limitations: 1) the deployment of RF devices in the data collection area, and 2) users visiting the target area to pre-collect a few instances.

This paper explores a novel approach that fundamentally differs from existing RF-based sensing applications, particularly in gait recognition, as it eliminates the need for prior RF data collection. Inspired by the recent success of deep generative models (e.g., ChatGPT), we aim to explore the feasibility of utilizing alternative sensing modalities to eliminate the need for data collection. Traditional gait recognition methods primarily rely on three sensing modalities: RF [57, 68, 70], camera [6, 36, 38], and wearable sensors [17, 56, 62]. This motivates us to leverage video and wearable sensor gait signals to generate RF gait signals. While the idea of using video footage to generate RF gait signals has been proposed in XModal-ID [33], we found that wearable sensors offer several advantages over video in this context for the following reasons.

* indicates corresponding author.

SenSys '23, November 12–17, 2023, Istanbul, Turkiye
© 2023 Association for Computing Machinery.
ACM ISBN 979-8-4007-0414-7/23/11…$15.00
https://doi.org/10.1145/3625687.3625792



(a) Home: Gait registration using IMU (b) Office: Gait recognition using RF signal

Figure 1: An application scenario of XGait. Users can leverage their mobile devices for gait registration, while performing gait recognition in various contexts equipped with an RF-based gait recognition system.

Firstly, video-based methods still require the deployment of a camera in the data collection environment. Secondly, videos shot by cameras are known to be affected by several practical factors, such as occlusion, lighting conditions, and viewpoint, which can introduce instability in the generated RF signals. In addition, privacy concerns arise when implementing video recording for gait analysis. In contrast, the Inertial Measurement Unit (IMU) has been widely equipped in everyone's mobile devices, such as smartphones and smartwatches. Moreover, utilizing IMU sensors that are readily accessible essentially eliminates the cost of device deployment. Lastly, since walking is a daily activity, gait data can be measured in daily life without explicitly asking the users to participate in data collection. A recent study [3] involving 717,527 individuals across 111 countries indicates that people, on average, walk about 3,000 steps per day, presenting huge opportunities for data collection. Fig. 1 illustrates a practical application scenario in two parts. Initially, users can register their gait information by walking anywhere while holding their smartphones, enabling the collection of IMU data associated with their unique gait patterns. The approach of utilizing IMU embedded in mobile devices for gait data collection facilitates spontaneous, implicit, and low-effort data collection.

One may ask "Since IMU-based gait recognition has been well studied, why do we still need to translate IMU data into RF data, which seems unnecessary?" Indeed, there is a rich body of studies on IMU-based gait recognition [9, 64, 77], but they are mainly designed to authenticate the user who is using the wearable device. As such, they are not applicable to ubiquitous and device-free scenarios, such as user recognition in a smart space. This motivates us to harness the unique strengths of IMU-based solutions to offset the limitations intrinsic to RF-based methods. More specifically, our approach seeks to leverage IMU signals, simulating the RF signals that would be produced if the same person were to walk near RF devices, thereby facilitating a more efficient collection of RF data.

1.2 Challenges and Contributions

We need to address several key challenges to achieve the aforementioned goal.

Challenge 1: Diversity of RF devices. Various RF signals such as Wi-Fi, mmWave, and LoRa which operate at different frequencies and use different modulation technologies have been utilized for wireless sensing tasks, resulting in unique expressions of sensing results for gait information. Consistently extracting and representing essential gait features across different RF signals remains a challenge. To address this challenge, we examine the fundamental principles of different RF sensing approaches. We reveal that different RF sensing approaches exhibit distinct sensing indicators due to varying modulation methods. Despite these differences, the consistent principle underlying all approaches is the phase change caused by human walking, which can be uniformly represented across different RF signals. Following this idea, we fuse multiple novel signal processing techniques and a tailored spectrogram generation method to extract the essential aspects of RF gait data, ensuring consistency across various RF signals.

Challenge 2: Intrinsic difference between IMU and RF signals. IMU signals have intrinsically different properties than RF signals. Firstly, IMU signals are represented as real numbers, whereas RF signals are represented as complex numbers. Secondly, IMUs capture inertial forces and rotation of the body when the user is walking [14], while RF sensing leverages shadowing, diffraction, reflection, and scattering phenomena exerted by a walking person on wireless links [55]. To tackle this challenge, we delve into human gait's IMU and RF signals associated with human gait. We employ a theoretical model to investigate the inherent correlation between them. However, due to the complex nature of human walking patterns, it is difficult to derive corresponding RF data from IMU data using mathematical calculations. Thus, built on the strength of the deep generative model [23, 24], we develop a deep generative network with a spectrogram fusion module and a spectrogram translation module, enabling effective conversion of IMU data into RF data.

Challenge 3: Complexity of gait. Gait recognition, regardless of the sensing modality, poses a significant challenge. First, gait is the coordinated movement of various parts of the body during walking, which involves 2 phases (stance and swing phase), 8 events (heel strike, preswing, etc.), and 24 body parts (arm, shoulder, and leg, etc.) [44]. However, capturing the whole 3D dynamic and complex locomotion using a single sensing modality is difficult. Moreover, the high similarity of gait signals among different people further hampers the accuracy of gait recognition systems. To achieve accurate gait recognition, it is crucial to address the limitations of traditional hand-crafted features [33, 57, 78], which fail to handle the complexity of gait. While the transformer architecture excels in image-related tasks, its ability to RF gait spectrograms is limited due to the unique time-frequency representation of gait signals. Therefore, we propose a spectrogram transformer specifically designed to handle RF spectrograms for gait recognition. This model significantly improves accuracy by incorporating critical components such as shifted spectrogram patches and the locality self-attention mechanism.

By incorporating the above solutions, we design and implement XGait, a unified **Cross**-modal **Gait** recognition framework that eliminates the need for prior RF device deployment and explicit data collection. This advancement pushes the existing RF-based gait recognition system into practical applications, as demonstrated in the video [1]. Our extensive evaluation, involving thirty subjects, three environments, three RF devices, and seven mobile devices, demonstrates that XGait achieves an average accuracy of 96.32%, 97.13%, and 93.26% in indoor, outdoor, and through-wall settings, respectively. The Top-3 accuracy consistently exceeds 99% in cross-registration and recognition scenarios, demonstrating the robustness of XGait. Our contributions are summarized as follows:

- To the best of our knowledge, XGait is the first cross-modal RF gait recognition framework that requires no RF device deployment and no explicit data collection, significantly reducing the burden of data collection in different environments.
- XGait offers dedicated threefold approaches for accurate gait recognition with different RF signals (i.e., Wi-Fi, LoRa, and mmWave), including a unified spectrogram generation model, a spectrogram translation model, and a spectrogram transformer recognition model. These three models address the aforementioned challenges associated with different RF signals and ensure efficient and accurate gait recognition in cross-modal settings.
- We collect a large dataset to comprehensively evaluate the performance of XGait in various scenarios. Experimental results show XGait achieves an average accuracy of 95.11% across various RF modalities and mobile devices.

2 RELATED WORK

Video-based Gait Recognition. Video-based gait recognition has been widely explored due to their high accuracy [6, 35, 36, 38]. Examples of such methods include silhouette-based approaches, which involve extracting and analyzing the human silhouette from video frames to recognize distinctive gait patterns [38]; gait energy image (GEI) analysis, where gait features are represented as energy images that capture the spatial distribution of motion energy throughout a gait cycle [36]. Despite the success of these methods, they exhibit certain drawbacks, as they are easily influenced by lighting conditions and cannot work effectively under non-line-of-sight (NLoS) conditions, limiting their applicability.

Sensor-based Gait Recognition. To overcome the limitations of video-based methods, sensor-based gait recognition approaches using various sensors [47, 53, 54, 82] are proposed. For example, Ren et al.[47] proposed a wearable-based method that uses acceleration to extract gait patterns from the human body. Wang et al. [82] presented a gait recognition approach that leverages smartphones' built-in sensors to recognize individuals' gait patterns in unconstrained real-world environments. Vera-Rodriguez et al.[54] explored the use of piezoelectric sensors on the floor to capture a user's gait information for user identification. However, these methods require users to wear additional equipment or install extra sensors on the floor, which constrains their application.

RF-based Gait Recognition. RF-based gait recognition has garnered significant attention due to its ability to authenticate individuals in a non-intrusive and device-free manner. Most RF-based gait recognition methods are based on Wi-Fi channel state information (CSI) [33, 51, 57, 61, 70, 81]. For example, WiFiU [57] proposes a method to extract physical features from CSI spectrograms and uses autocorrelation methods to eliminate feature imperfections. XModel-ID employs video to generate Wi-Fi signals, which are then used for similarity comparison. However, the use of video can potentially impair user privacy, as it involves capturing and processing visual data. Moreover, mmWave Radar-based solutions have attracted a lot of attention. MU-ID [68] analyzes the mmWave signal of the range-Doppler domain and extracts features such as step length, period, and speed for multi-person identification. GaitCube [42] analyses the 3D joint-feature representation of mmWave signal over time to achieve gait recognition. In addition, LoGait [16] uses LoRa signals for user recognition both indoors and outdoors. RFID-based methods are also explored. For example, RF-Gait [30] identifies a person through unobtrusive gait sensing using COTS RFID, while RFPass [7] can achieve environment-independent gait-based user authentication. However, RFID-based methods necessitate extra tags, potentially impairing user convenience and impeding their widespread adoption. Although various solutions are currently available, as shown in Tab. 1, none can eliminate the user registration in the target area, which restricts the deployment of RF gait recognition systems. Thus, we design XGait to achieve unified RF-based gait recognition that does not require the prior deployment of RF devices or explicit data collection.

Table 1: Comparison with RF-based methods. (●=High, ◐=Moderate, ○=Low.)

System	Technology	Spectrum	Modulation	Target area data collect	Privacy Preserve
MU-ID [68]	mmWave	77-81 GHz	FMCW	Yes	◐
GaitCube [42]	mmWave	77 GHz	FMCW	Yes	◐
mmGaitNet [39]	mmWave	77 GHz	FMCW	Yes	◐
Wi-FiU [57]	Wi-Fi	5 GHz	OFDM	Yes	◐
Wi-Fi-ID [74]	Wi-Fi	5.19 GHz	OFDM	Yes	◐
GaitID [78]	Wi-Fi	5.19 GHz	OFDM	Yes	◐
GaitSense [79]	Wi-Fi	5.825 GHz	OFDM	Yes	◐
XModal-ID [33]	Wi-Fi	5.18 GHz	OFDM	No	○
GaitWay [61]	Wi-Fi	5.8 GHz	OFDM	Yes	◐
LoGait [16]	LoRa	Unknown	CSS	Yes	◐
XGait	Wi-Fi	5.18 GHz	OFDM	No	◐
	mmWave	77 GHz	FMCW	No	◐
	LoRa	915 MHz	CSS	No	◐

3 PRELIMINARIES

In this section, we first elucidate the sensing preliminary of three prevalent RF signals (i.e., LoRa, Wi-Fi, and mmWave). Then, we analyze the human gait principle and the correlation between the RF signal and the IMU signal.

3.1 RF Sensing Preliminary

Phase change demonstrates high sensitivity to target motion in various studies [26, 71, 72]. In contrast, the signal amplitude is less sensitive, making it less desirable. Thus, the phase change is used for RF gait sensing in our system.

3.1.1 LoRa Sensing. LoRa utilizes Chirp Spread Spectrum (CSS) modulation to encode data with frequency-varying chirps. The received LoRa signal can be categorized as static path signals H_s from static object reflections and line-of-sight (LoS) signals, and dynamic path signals H_d generated by target movement. As the target moves, the dynamic path signals H_d undergo a rotation relative to the static path signal H_s, leading to a phase change in the combined signal H_c. However, random phase shifts from factors like carrier frequency offsets (CFO) and sampling frequency offsets (SFO) can corrupt the signal's phase. To address the random phase offsets, we remove the varying random phase offsets by calculating the ratio of two signals received at the two antennas [72] as

$$RF_1(t) = \frac{H_1(t)}{H_2(t)} = \frac{e^{\phi_o}\ (H_{s1}(t) + H_{d1}(t))}{e^{\phi_o}\ (H_{s2}(t) + H_{d2}(t))}, \tag{1}$$

where $H_1(t)$ and $H_2(t)$ are the signals received at two antennas. The phase offsets e^{ϕ_o} are removed by the division.

3.1.2 Wi-Fi Sensing. Wi-Fi communication utilizes Orthogonal Frequency Division Multiplexing (OFDM) to encode digital data across multiple subcarriers. Additionally, CSI serves to characterize the attenuation, fading, and scattering effects experienced by OFDM signals as they propagate through the environment [28, 29]. Analogous to the LoRa signal, the received Wi-Fi CSI consists of both static path signals H_s and dynamic path signals H_d. Nevertheless, the inherent limitations of commercial Wi-Fi devices introduce certain imperfections in the retrieval of real-world CSI. Specifically, these imperfections manifest as random phase shifts within the acquired CSI data, thereby rendering the phase value unreliable and unusable. Therefore, the CSI ratio [71] is used to eliminate these random phase shifts, which is defined as the division of CSI between two antennas:

$$RF_2(t) = \frac{H_1'(t)}{H_2'(t)} = \frac{e^{\phi_o'}\left(H_{s1}'(t) + H_{d1}'(t)\right)}{e^{\phi_o'}\left(H_{s2}'(t) + H_{d2}'(t)\right)}, \tag{2}$$

where $H_1'(t)$ and $H_2'(t)$ are CSI received at two antennas, and $e^{\phi_o'}$ is the varying phase offsets.

3.1.3 mmWave Sensing. For Frequency Modulated Continuous Wave (FMCW) mmWave radar systems, the chirp signal is utilized, wherein the frequency of the transmitted signal increases linearly with time [8, 46]. The transmitted chirp signal can be expressed as $s_T(t) = A\cos\left[2\pi\left(f_0 t + \frac{St^2}{2}\right)\right]$, where f_0 is the starting frequency and S is the frequency modulation slope. The received chirp signal is essentially a time-delayed version of the transmitted signal, given by the expression $s_R(t) = \alpha A\cos\left[2\pi\left(f_0(t-\tau) + \frac{S(t-\tau)^2}{2}\right)\right]$, where α is the path loss, $\tau = 2d/c$ is the time delay, d is the target distance, and c is the speed of light. Then, the transmitted signal is mixed with the received signal to obtain the Intermediate Frequency (IF) signal. Though we process the IF signal subsequently to perform various sensing tasks, the movement target is essentially sensed by measuring the phase change between transmitted and received chirp signals:

$$\Delta\phi = \frac{4\pi\Delta d}{\lambda} = \frac{4\pi v T_c}{\lambda}, \tag{3}$$

where v is the target velocity, T_c is the chirp duration, and λ is the wavelength.

3.2 Understaning Human Gait

Gait is characterized by a combination of various locomotive patterns exhibited by distinct body segments. These diverse ambulatory characteristics originate from the unique interactions among specific body components, such as arms, torso, and legs, which operate at different velocities for each individual, leading to a wide range of walking patterns, such as variations in walking speed, stride length, and arm swing [41]. These factors collectively contribute to the distinctive gait characteristics of each individual. In the upper portion of Fig. 2, a moving wheel from left to right is applied to human gait. We imagine a cyclic pattern of movement that is repeated continuously, step by step. Descriptions of walking are typically confined to a single cycle, assuming that successive cycles are nearly identical. Although not strictly accurate, this assumption is reasonable for most individuals [44]. By convention, the cycle begins when one of the feet makes contact with the ground.

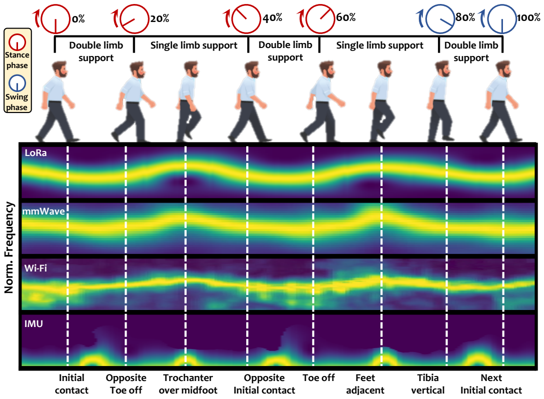


Figure 2: Gait cycle and the corresponding RF spectrograms calculated from different sensing modalities.

Phases. The gait cycle consists of two phases: the stance phase and the swing phase. During the stance phase, the foot is in contact with the ground, while in the swing phase, that same foot loses contact with the ground as the leg swings forward in preparation for the subsequent foot strike. Note that while Fig. 2 refers to the right body side, the same terms apply to the left side, which is typically half a cycle behind. Hence, the first double support for the right side corresponds to the second double support for the left side, and vice versa.

3.3 Correlation Analysis

IMU. As mentioned in Sec. 3.2, gait characteristics emerge from the distinct interactions among specific body components. Each individual operates these components at varying velocities, resulting in diverse walking patterns. Consequently, the IMU data from an on-body mobile device, denoted as $I(t)$, collected during walking, is a composite representation of the accelerations experienced by different human body segments. This can be expressed as

$$I(t) = \sum_n h_n(a_n(t)), \tag{4}$$

where $a_n(t)$ is the acceleration of the n^{th} body part at time t, and $h_n(\cdot)$ represents the transfer function that maps the acceleration caused by the movement of each body part to the on-body device.

RF. As discussed in Sec. 3.1.1, the movement of human walking can result in phase changes of the RF signal. Although different RF signals have different expressions, human movement is reflected in the phase change of the RF signal. Therefore, we can mathematically represent the RF signal changes caused by the user walking:

$$RF(f,t) = H_s(f,t) + \sum_n A_n(f,t)e^{-j(\frac{2\pi}{\lambda}(d_{n0}+a_n(t)\cdot t^2))}. \tag{5}$$

In this equation, $H_s(f,t)$ denotes the complex static path signal, $A_n(f,t)$ represents the amplitude of the dynamic path signal reflected off the n^{th} body part, d_{n0} is the signal propagation length at time $t = 0$, and $a_n(t)$ is the acceleration of the n^{th} body part. Let $R(t)$ represent the magnitude square of the baseband signal $RF(t)$. Assuming that $|H_s(f,t)| \gg |A_m(f,t)|$, we can express $R(t)$ as

$$R(t) = P + \sum_n B_n \cos\left(\frac{2\pi}{\lambda}(d_{n0} + a_n(t)\cdot t^2) - \phi_s\right), \tag{6}$$

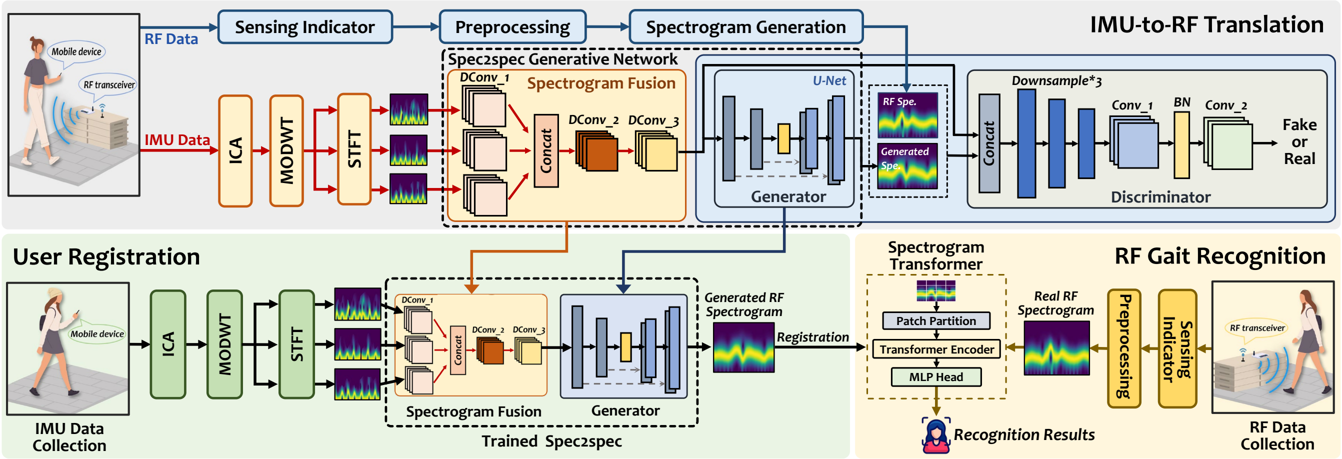


Figure 3: XGait overview. 1) Train an IMU-to-RF translation model for IMU-to-RF data conversion. 2) Register users by collecting, processing, and converting their IMU gait data to RF data, then use the data to train the recognition model. 3) Conduct RF gait recognition to identify users using on-site RF gait data.

where $P = |H_s(f,t)|^2 + \sum_n |A_n(f,t)|^2$ refers to the DC component of $R(t)$, $B_n = 2\,|H_s(f,t) \cdot A_n(f,t)|$, and ϕ_s is the phase of the complex static path signal. Eq. 6 shows that the received RF signal can be viewed as a superposition function, with the acceleration of each body part acting as the independent variable. As a result, the RF signal contains a variety of frequency components, stemming from the unique acceleration profiles of different body parts.

Correlation. As indicated by Eq. 4 and Eq. 6, $I(t)$ encompasses frequency components that either overlap or closely correspond to the frequencies present in $R(t)$. As demonstrated in Fig. 2, we present the processed RF and IMU spectrograms for two gait cycles of the same individual, showing that different gait events (e.g., initial contact, feet adjacent, and toe-off) induce correlated changes in RF and IMU spectrograms. This empirical evidence further substantiates the observed correlation between these two modalities. Given that both are defined by functions with acceleration as the independent variable, there exists a possibility of converting IMU data into RF data through a non-linear function $\mathcal{F}(\cdot)$, as represented by the following equation:

$$R(t) = \mathcal{F}(I(t)). \tag{7}$$

However, accurately obtaining the precise form of the function $\mathcal{F}(\cdot)$ using traditional mathematical approaches is difficult. This is due to the intrinsic complexity and non-linearity of the transformation between the IMU and RF data, along with the influence of factors, such as noise, signal attenuation, and multi-path effect [2, 77]. In this work, leveraging the remarkable non-linear fitting capabilities of deep learning, we train a deep generative model to establish this non-linear mapping relationship from the IMU data to the RF data.

4 SYSTEM OVERVIEW

Fig. 3 shows XGait's overview, comprising three phases.

User Registration. In the user registration phase, the user uses his/her on-body smart device such as a smartphone and smartwatch to collect IMU gait data when he/she is walking. The collected IMU data is processed by independent component analysis (ICA), maximal overlap discrete wavelet transform (MODWT), and short-time Fourier transform (STFT) to yield multiple time-frequency spectrograms. Subsequently, the trained spectrogram fusion network is utilized to fuse all spectrograms into a single feature map. The trained generator is then employed to reconstruct the RF spectrogram. Finally, the generated RF spectrograms are used to train the gait recognition model.

IMU-to-RF Translation. We design a deep generative model called Spec2spec to transform the IMU data into RF data. The IMU data is first processed to obtain corresponding time-frequency spectrograms using MODWT and STFT, and the specially designed spectrogram fusion module then combines multiple IMU spectrograms. For RF data, the raw data is initially converted into relevant sensing indicators and transformed into spectrograms through the spectrogram generation method. Finally, a spectrogram translation module is employed to convert the IMU data to RF data.

Gait Recognition. Registered users can use the RF-based gait recognition system directly without on-site registration. The system first processes the RF gait data to get the corresponding spectrogram and recognizes users using the designed spectrogram transformer.

5 SPECTROGRAM GENERATION

5.1 RF Spectrogram Generation

5.1.1 LoRa Signal. As discussed in Sec. 3.1.1, the raw LoRa phase is unsuitable for use due to random phase discrepancies. Therefore, the phase ratio is employed to generate the LoRa gait spectrogram. Initially, we employ the median filter and the Hampel identifier [43] to eliminate outliers that deviate significantly from the expected values. After that, recognizing the inherent advantages of multi-scale analysis and the attainment of optimal frequency and time domain resolution, the MODWT [45] is implemented to further mitigate residual noise present in the preprocessed denoised signal. Specifically, the preprocessed denoised signal is subjected to a decomposition process, resulting in the acquisition of approximation coefficients β and detail coefficients α. The approximation coefficients $\alpha(0,k)$ for scale 0 is calculated as

$$\alpha(0,k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) \cdot \psi(0, n-k), \tag{8}$$

where N is the length of input signal $x(n)$, and $\psi(0, n-k)$ represents the value of the wavelet function at time $(n-k)$ for scale 0. Then the

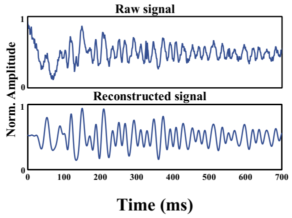


Figure 4: MODWT result of LoRa.

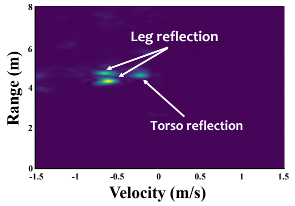


Figure 5: mmWave RDM.

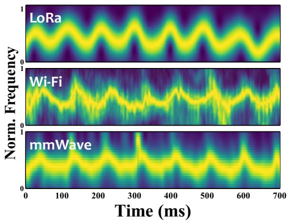


Figure 6: Extracted RF feature.

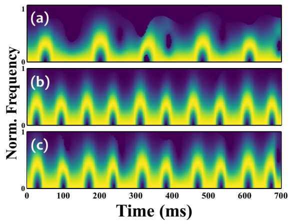


Figure 7: Extracted IMU feature.

approximation and detail coefficients are then calculated iteratively by the following equations:

$$\alpha(j,k) = \frac{1}{\sqrt{2^j}} \sum_{n=0}^{N-1} h(j,n) \cdot \alpha(j-1, 2k-n),$$
$$\beta(j,k) = \frac{1}{\sqrt{2^j}} \sum_{n=0}^{N-1} g(j,n) \cdot \alpha(j-1, 2k-n), \tag{9}$$

where $g(j,n)$, $h(j,n)$ represents the wavelet function of the high-pass and low-pass filter for scale j. We choose the Symlets 4 wavelet (sym4) and decompose the signal to four levels.

Thereafter, the detail coefficient threshold is applied to each level to discard the ambient clutter. Finally, a level-dependent reconstruction is performed using

$$x(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \alpha(J,k) \cdot \varphi(J, n-k) + \sum_{j=1}^{J} \sum_{k=0}^{N-1} \beta(j,k) \cdot \psi(j, n-k), \tag{10}$$

where $\varphi(J,n)$ represents the approximation wavelet function for scale J, and $\psi(j,n)$ represents the detail wavelet function for scale j. The reconstructed LoRa signal is shown in Fig. 4.

Next, we transform the reconstructed signal into a spectrogram using STFT, which involves dividing the reconstructed signal into short overlapping segments and calculating the Fourier transform for each segment [33, 67]. The STFT coefficient $X(n,k)$ at time index n and frequency bin k is given by

$$X(n,k) = \sum_{m=0}^{M-1} x(m) \cdot w(m - nH) \cdot e^{-j\frac{2\pi}{N}km}, \tag{11}$$

where $x(m)$ denotes the reconstructed signal at index m, $w(m)$ denotes the window function value at index m, M is the segment length, and H is the hop size.

5.1.2 *Wi-Fi Signal.*

As outlined in Sec. 3.1.2, the Wi-Fi signal encounters comparable instances of random phase shifts similar to the LoRa signal. Consequently, the utilization of the CSI ratio has been proven to be instrumental in mitigating these random phase shifts. Drawing upon the approach employed for LoRa phase processing detailed in Sec. 5.1.1, we first use the median filter and Hampel identifier for preliminary noise reduction. Since the Wi-Fi signal contains multiple subcarriers, the different subcarrier wavelengths lead to different responses to walking movements. Therefore, we leverage a subcarrier selection method [59] on the preprocessed data to obtain the subcarrier that best reflects the user's walking characteristics. Subsequently, we process the selected subcarriers using MODWT to further eliminate noise and reconstruct the signal, ensuring that it accurately represents the user's walking characteristics. Finally, we derive the spectrogram of the reconstructed signal using STFT.

5.1.3 *mmWave Signal.*

As illustrated in Sec. 3.1.3, the IF signal obtained by mixing is used for sensing human walking. In addition to the user's walking movements, the IF signal contains a lot of static noise generated by static objects such as walls, tables, and chairs. For each frame, we use the average of all IF signals as the static noise vector and subtract this static noise vector from each IF signal to obtain the denoised data [66]. Then, we use the range Fast Fourier Transform (FFT) and Doppler FFT to obtain a Range-Doppler Map (RDM) for each frame as shown in Fig. 5, which reflects the range and velocity information of the user while walking in the current frame. Finally, to obtain the velocity change information during the walking, we use the following equation [20] to transform the RDMs of all frames into a 2D time-velocity feature map:

$$\boldsymbol{V}_{(n,i)} = \frac{\sum_{j=1}^{N_R} [\boldsymbol{RDM}_{(n,i,j)} \cdot B_j]}{N_R}, \quad i \in [1, N_D], j \in [1, N_R], \tag{12}$$

where N_R and N_D are the Range FFT and Doppler FFT numbers, B_j is the range bin index, and $\boldsymbol{RDM}_{(n,i,j)}$ denotes the value corresponding to Doppler bin i and range bin j in the n^{th} RDM frame. As discussed in Sec. 3.1.3, the Doppler FFT responds to changes in phase difference, which are proportional to frequency. Therefore, we can derive the normalized frequencies by normalizing the extracted velocities.

Fig. 6 presents the extracted spectrograms for the three types of RF signals. Distinct and consistent gait patterns can be observed, indicating the effectiveness of capturing the user's gait information.

5.2 IMU Spectrogram Generation

During walking, accelerometer data from a single body location often combines accelerations from multiple locations (e.g., leg, waist, arm) [63]. Common mobile device placements include hip, chest, hand, and wrist. Devices on the body trunk primarily capture gait-related motion signals, allowing direct use of acceleration readings. However, devices on swing arms detect mixed gait and arm swing signals. To extract useful gait information, we must separate leg and arm swing motion signals. In this paper, we use ICA to separate signals from different body sources. Assuming IMU-measured acceleration $A(t)$ is a mix of hand-waving and trembling, our ICA model is expressed as $A(t) = A \cdot S(t)$ with mixing matrix A and independent sources $S(t)$. FastICA [25] is used to obtain unmixing matrix $W = A^{-1}$, estimating source signals with $S(t) = W \cdot A(t)$. We obtain denoised acceleration by selecting the independent component with the lowest dominant frequency [63]. To further mitigate residual noise, the acceleration data along the X, Y, and Z axes are decomposed into four levels using MODWT. At each level, a detail coefficient threshold is applied to discard ambient clutter. The resulting denoised signals are then reconstructed employing Eq. 10.

Finally, the three processed IMU signals are employed to derive three spectrograms using STFT, which are shown in Fig. 7(a)-(c), respectively. The obtained spectrograms from the three directions are utilized as the inputs of the spectrogram fusion module in Sec. 6.1, which integrates the gait information to generate RF spectrograms.

6 SPEC2SPEC GENERATIVE NETWORK

This section introduces Spec2spec, a deep generative model designed to convert IMU spectrograms to RF spectrograms. Fig. 3 shows the Spec2spec structure, which consists of a spectrogram fusion module and a spectrogram translation module.

6.1 Spectrogram Fusion

Conventional image learning networks [22, 52] are designed to process individual RGB images as input, operating on a single image basis. However, in spectrogram translation, relying solely on a single IMU spectrogram may result in the omission of crucial features, as 3D IMU data encompasses spectrograms from three directions, each containing gait information. To overcome this issue, we propose a spectrogram fusion module that combines the three IMU spectrograms for a more effective representation of gait data.

Standard Convolutional Neural Networks (CNNs) are commonly employed for image fusion tasks. The core of a CNN is the standard convolution operation, as illustrated in Fig. 8(b), which can be described as

$$Y(\sigma) = \sum_{\sigma' \in S} X(\sigma + \sigma') * K(\sigma'), \tag{13}$$

where $Y(\sigma)$ is the output feature map, $X(\sigma)$ is the input feature map, $K(\sigma')$ is the convolution kernel, and S is the neighborhood around the pixel σ. IMU data encompasses a wide range of information relating to signal frequency and motion intensity, which can vary rapidly across time. Moreover, the frequency spectrum shapes of different subjects' IMU data are highly diverse. These factors make effectively modeling and extracting features from IMU data [37, 69] a challenge.

To tackle the above challenge, we develop a Deformable Convolutional Network (DCN)-based spectrogram module, as illustrated in Fig. 8(a). This module combines the three IMU spectrograms by employing a deformable convolution operation, shown in Fig. 8(c). By introducing an offset $\Delta\sigma'$ to the regular grid sampling locations, the module allows for more flexible feature extraction and enhanced modeling of the features in the IMU spectrograms:

$$Y(\sigma) = \sum_{\sigma' \in S} X(\sigma + \sigma' + \Delta\sigma') * K(\sigma'), \tag{14}$$

where $\Delta\sigma'$ is the offset for the pixel σ'. The offsets are learned during the training, allowing the model to adapt to geometric variations and focus on frequencies at different timescales. The offsets are predicted by another convolutional layer:

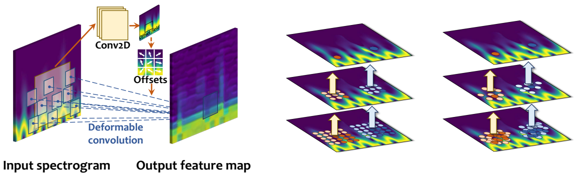
$$\Delta\sigma' = F_{offset}(X, K_{offset}), \tag{15}$$

where F_{offset} represents the offset layer, and K_{offset} is the kernel for the offset layer. To handle irregular grid sampling locations, a bilinear interpolation is used:

$$In(\sigma) = \sum_{\kappa \in N(\sigma)} X(\kappa) * max(0, 1 - |\sigma_x - \kappa_x|) * max(0, 1 - |\sigma_y - \kappa_y|), \tag{16}$$

where $In(\sigma)$ is the interpolated value at location σ, and $N(\sigma)$ is the set of nearest neighbor pixels around location σ. Combining the deformable convolution with the bilinear interpolation, we get the final formula:

$$Y(\sigma) = \sum_{\sigma' \in S} In(\sigma + \sigma' + \Delta\sigma') * K(\sigma'). \tag{17}$$



(a) Offsets learning of DCN. (b) CNN. (c) DCN.

Figure 8: Illustration of the deformable convolution.

The DCN-based spectrogram module takes three IMU spectrograms as input and the output is a processed spectrogram of the same size as the input. The use of learned offsets and bilinear interpolation to handle irregular grid sampling locations enhances flexibility and adaptation to geometric variations, which is well-suited for spectrogram fusion.

6.2 Spectrogram Translation

Based on the formulas in Sec. 3.3, it can be inferred that a correlation exists between IMU and RF sensing data for gait. However, the relationship between them is highly complex, which makes it challenging to use mathematical calculations to derive. Our goal is to convert the IMU spectrograms into a corresponding RF spectrogram. To achieve this objective, we design a spectrogram translation module. The design details are as follows.

Module design. The spectrogram translation module, as depicted in Fig. 3, relies on a conditional generative adversarial network (cGAN) architecture [40]. This module takes two inputs for training: the ground truth signal spectrogram p, which corresponds to RF signals, and the condition spectrogram w fused from three input IMU spectrograms. The generator network G combines a noise vector v with the condition w to yield a fake spectrogram $G(v \mid w)$, which is one of the inputs to the discriminator network D. Additionally, D receives a second input that combines p and w to represent the real spectrogram under the condition w. During the training process, D learns to distinguish between $G(v \mid w)$ and the ground truth spectrogram $p \mid w$, while G adjusts its parameters to generate a $G(v \mid w)$ that can fool D. G learns the mapping from IMU spectral features to RF spectral features by playing a zero-sum game with D. Once the training is complete, the generator G can accurately reconstruct an RF spectrogram using the IMU spectrograms, even when the sample was not included in the training set.

Generator. We employ the U-Net architecture [48] as the generator's network. The U-Net structure is symmetrical, featuring convolutional layers on the left and upsampling layers on the right. The upsampling layers decode features to predict pixel labels. The skip connection is employed to enhance the flow of information between layers, as illustrated by the gray dashed line in the upper panel of Fig. 3.

Discriminator. The discriminator is designed with three convolutional layers that work together to process the input spectrograms. The discriminator examines and classifies each patch within the spectrogram as either real or fake. The model continuously refines

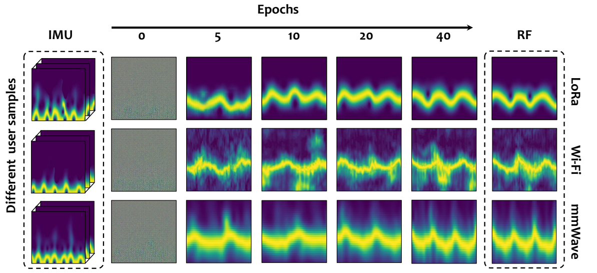
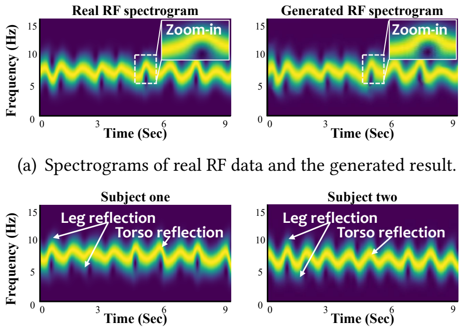


Figure 9: Training progress for various RF signal.



(a) Spectrograms of real RF data and the generated result.

(b) Spec2spec generated spectrograms of two different people.

Figure 10: Illustration of the generated results.

its ability to discriminate between real and fake patches. Once the training process is complete, the final output of the discriminator D is determined by taking the average of all responses obtained from a single convolutional pass across the entire spectrogram.

Loss function. To make the reconstructed RF spectrogram more similar to the ground truth, we define the loss function for the magnitude spectrogram of generated RF spectrograms and original RF spectrograms. It can be expressed as

$$L_M = \left\| M(t,f) - M_q(t,f) \right\|_1, t \in T, f \in F, \tag{18}$$

where $M(t,f)$ and $M_q(t,f)$ represent the magnitude spectrogram of the generated and original RF signals, and T, F denote the time and frequency domain, respectively.

Training. The spectrogram fusion module has one convolution layer and two deformable convolution layers with a kernel size of 3×3 with padding of one. The spectrogram translation module consists of a generator and a discriminator. The generator has three downsampling and upsampling operations with concatenation and the submodules of the generator use a kernel size of 4×4, stride of two, and padding of one. The discriminator has three convolutional layers with kernel size 1×1, followed by a leaky ReLU activation function and batch normalization. We train each model with 300 epochs with a learning rate of 0.0002 in the first 150 epochs and use Adam [31] to adaptively adjust the learning rate. Fig. 9 displays the training progress of various RF signals, demonstrating the efficacy of the Spec2spec model.

To demonstrate the effectiveness of Spec2spec, we conduct an experiment using LoRa as an example, as the spectrogram of different RF modalities can vary. A participant walks along a path that perpendicularly intersects the RF link, carrying a smartphone to capture IMU data. Fig. 10(a) displays the spectrogram obtained from actual RF data, along with the spectrogram generated from the IMU data. The obvious similarity between the generated spectrogram and the real RF spectrogram underscores the accuracy of our proposed approach. Furthermore, Fig. 10(b) presents two generated spectrograms corresponding to two different users, illustrating that their gait patterns are distinguishable in the figure.

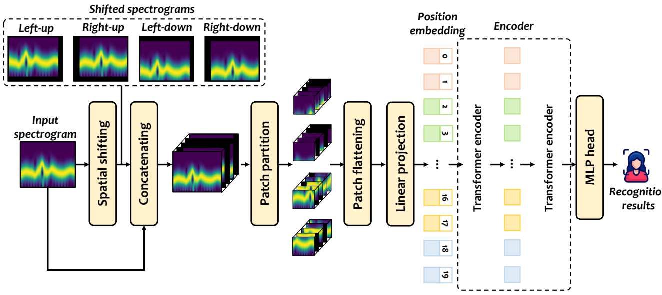


Figure 11: Spectrogram transformer.

7 GAIT RECOGNITION

As discussed in Sec. 1, accurate gait recognition is challenging because of the nature of human walking movements and the high similarity of different users' walking patterns. Conventional feature extraction for gait recognition is achieved manually using hand-designed operators [33, 57, 78], which are inherently limited compared to learnable features. Inspired by the transformer architecture [12, 34] in image-related tasks, we propose a spectrogram transformer specially designed to handle RF spectrograms for gait recognition. Using shifted spectrogram patches and a patch embedding layer, the model captures spatial information and inter-patch relationships, while the locality self-attention mechanism enhances gait feature extraction by focusing on local spatial relationships. In gait recognition, we aim to distinguish unique user-specific patterns, treating each user as a distinct class in a multi-class recognition problem. The details of our proposed method are as follows.

Shifted spectrogram patch. To overcome the low receptive field issue in vision transformers [12], we employ shifted spectrogram patches to enhance spatial information capture. As illustrated in Fig. 11, we first generate overlapping patches from the input spectrogram by shifting it diagonally in various frequency and time directions. We then concatenate these shifted versions with the original spectrogram, providing a comprehensive representation that includes multiple time-frequency perspectives. Next, we extract smaller, non-overlapping patches for subsequent processing.

Spectrogram patch embedding layer. This layer transforms patches into embeddings for subsequent model layers while incorporating positional information, enabling the model to comprehend patch positions within the input spectrogram. As depicted in Fig. 11, we first linearly project flattened spectrogram patches into a lower-dimensional space using a learnable weight matrix. This reduces computational complexity and memory requirements. Positional encodings can be generated using sinusoidal functions and learnable parameters. Combined through element-wise addition, the resulting embeddings incorporate both patch and positional information, which is crucial for gait recognition.

Locality self-attention mechanism. To effectively learn input spectrogram details, we use the locality self-attention mechanism [34]

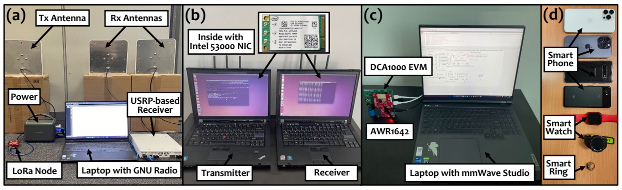


Figure 12: Experimental devices.

Table 2: Mobile device specifications.

Name	CPU Frequency	RAM	OS	IMU Model*
iPhone 13 Pro Max	3.23 GHz	6 GB	iOS 15	BS IMU
iPhone 14 Pro	3.46 GHz	6 GB	iOS 16	BS IMU
Samsung S10	2.84 GHz	8 GB	Android 11	STM LSM6DSO
Nexus 6P	1.95 GHz	3 GB	Android 9	IS ICM40604
Apple Watch S7	1.8 GHz	1 GB	Watch OS 9	Unknown
Huawei Watch GT2	200 MHz	32 MB	Lite OS 11	STM LSM6DSOWTR
Customized Smart Ring	64 MHz	64 kB	N/A	IS MPU9250

* BS: Bosch, STM: STMicroelectronics, IS: TDK InvenSense.

to selectively concentrate on the most relevant spatial information. Specifically, We first mask the dot product matrix diagonal in self-attention calculation. This forces the attention module to prioritize inter-token relations. Then we use a learnable scaling factor, providing greater flexibility in modulating the softmax function and adaptively sharpening attention score distribution.

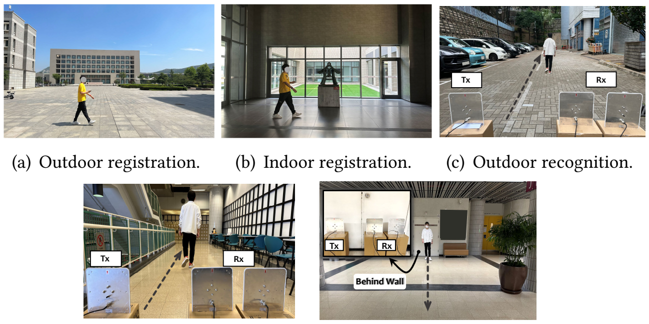
The model architecture consists of two transformer layers, each featuring 64-dimensional patch projections and four attention heads. The model is trained for 800 epochs using a learning rate of 0.001 and a weight decay of 0.0001. The transformer and MLP head layers have hidden units of $(128, 64)$ and $(2048, 1024)$, respectively. To sum up, our proposed method addresses the data-intensive and complex training demands of traditional transformer models by incorporating essential components, leading to enhanced recognition capability, as will be demonstrated in Sec. 8.3.

8 EVALUATION

8.1 Experimental Methodology

System implementation. As shown in Fig. 12 (a)–(c), we use three types of RF devices for data collection. Firstly, the Wi-Fi setup uses two laptops with Intel 5300 WLAN NICs, operating on channel 64 at 5.32 GHz. The transceivers are equipped with CSI Tool [19] for CSI data collection, and both transceivers have three omnidirectional antennas, making a total of 270 data streams. Secondly, the LoRa system consists of an Arduino Uno with Semtech SX1276 transmitting at 915 MHz with a 125 kHz bandwidth, and a USRP X310 and GNU Radio-based gateway connected to a data processing laptop. Lastly, the mmWave implementation features a 77 GHz FMCW AWR1642 radar and a DCA1000EVM real-time data-capture adapter. For IMU data collection, we utilize seven mobile devices as shown in Fig. 12 (d). The specifications of these devices are detailed in Tab. 2. The default sampling rates of IMUs are set to 100 Hz.

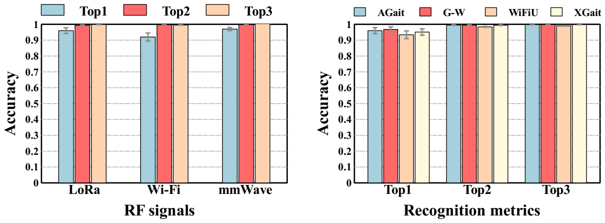
Data collection. To evaluate XGait, we recruited 30 participants (14 women and 16 men), aged from 15 to 64 and weighing between 45 kg and 85 kg. All participants are healthy and participate in a series of experiments [1]. To assess the spectrogram translation model, we randomly selected 15 participants, each walking 30 times in front of the three RF devices carrying mobile devices, both indoors and outdoors. To evaluate the final gait recognition results, we choose the remaining 15 participants, each performing two separate sets of walking trials: (1) for registration data collection, they walk 3 min carrying a mobile device in both indoor and outdoor environments as shown in Fig. 13, and (2) for gait recognition data collection, they walk 30 times in front of RF devices in indoor, outdoor, and through-wall as shown in Fig. 13.



(a) Outdoor registration. (b) Indoor registration. (c) Outdoor recognition.

(d) Indoor recognition. (e) Through-wall recognition.

Figure 13: Experimental scenarios.



(a) Overall accuracy. (b) Comparison with baselines.

Figure 14: Overall Performance.

8.2 Overall Performance

Overall accuracy. Fig. 14(a) shows the Top-1, Top-2, and Top-3 accuracy of LoRa, Wi-Fi, and mmWave for gait recognition with registration using IMU. Top-N accuracy measures how frequently the correct user appears within the top N predictions. The Top-1 accuracy for LoRa, Wi-Fi, and mmWave are 96.21%, 92.14%, and 96.97%, respectively. Similarly, the Top-2 accuracy values are 99.54%, 98.72%, and 99.98%, and the Top-3 accuracy values are 99.89%, 99.56%, and 100%. These results demonstrate the effectiveness of our system in recognizing users across various communication technologies. The high accuracy across all three categories indicates that XGait is adaptable and performs well regardless of the RF signal being used.

Comparison with baselines. We then compare our system with three state-of-the-art gait recognition systems, namely i) AGait [65], an RF-based gait recognition system; ii) Gait-Watch [64], an IMU-based gait recognition system; and iii) WiFiU [57], an RF-based recognition system with explicit features. The performance of each system is fine-tuned to achieve the best results on our dataset. For RF-based systems, we report the average results under three different RF conditions. As shown in Fig. 14(b), On average, XGait performs comparably with the state-of-the-art: it is 1.1% lower than AGait, 1.3% lower than Gait-Watch, and 2% higher than WiFiU. The

[1] Ethical approval has been granted by the corresponding organization.

Table 3: Cross registration and recognition scenario results. (● for chosen, ○ for unchosen, I.D.:indoor, O.D.: outdoor, T.W.: through-wall.)

	Reg. Sce.		Recognit. Sce.			Accuracy		
RF	I.D.	O.D.	I.D.	O.D.	T.W.	Top-1	Top-2	Top-3
LoRa	●	○	●	○	○	96.52%	99.78%	100%
	○	●	●	○	○	97.13%	99.85%	100%
	●	○	○	●	○	97.56%	99.92%	100%
	○	●	○	●	○	98.14%	99.92%	100%
	●	○	○	○	●	91.56%	98.31%	99.52%
	○	●	○	○	●	94.21%	99.45%	99.83%
Wi-Fi	●	○	●	○	○	92.48%	98.46%	99.41%
	○	●	●	○	○	92.97%	98.52%	99.48%
	●	○	○	●	○	93.47%	98.76%	99.63%
	○	●	○	●	○	94.11%	99.45%	99.91%
	●	○	○	○	●	89.56%	98.58%	99.45%
	○	●	○	○	●	89.97%	98.58%	99.52%
mmWave	●	○	●	○	○	97.52%	99.92%	100%
	○	●	●	○	○	98.12%	100%	100%
	●	○	○	●	○	98.51%	100%	100%
	○	●	○	●	○	99.23%	100%	100%
	●	○	○	○	●	94.53%	100%	100%
	○	●	○	○	●	94.99%	100%	100%

relatively lower accuracy of WiFiU can be attributed to its reliance on traditional machine learning classifiers and manually crafted RF features. These results demonstrate that XGait can achieve performance comparable to that of the state-of-the-art systems.

8.3 Micro-Benchmark Evaluation

Generalization ability. We evaluate the system's adaptability to various environments and RF modalities. Tab. 3 presents the recognition performance of XGait under various combinations of registration and recognition scenarios. The results show that XGait consistently maintains high accuracy across different scenarios, with all Top-3 accuracy values exceeding 99%. We note that mmWave achieves superior performance due to its short wavelengths resulting from the high frequency (i.e., 77 GHz). While Wi-Fi operates at a higher frequency than LoRa, its performance is inferior, mainly due to the use of omnidirectional antennas, which offer expansive coverage but sacrifice focus and precision. To evaluate the adaptability of XGait to different RF modalities, we choose a model of one RF modality as the baseline, then fine-tune new models with data from two other modalities through ten epochs of training. As observed in Fig. 15(a), every modality achieves over 90% accuracy, demonstrating the model's superior generalization ability of different RF modalities.

Performance of IMU to RF translation. As illustrated in Fig. 15(b), we evaluate the performance of the Spec2spec network, which converts IMU spectrograms to RF spectrograms. We employ the widely used Structural Similarity Index Measure (SSIM) [60] to quantify spectrogram similarity. The results reveal that the Spec2spec network consistently excels across various RF signals and environments, achieving 95.18%, 92.98%, and 96.01% similarity for three RF devices in indoor scenarios, and 97.12%, 93.78%, and 97.34% in outdoor scenarios. We also evaluate the information loss during the IMU to RF translation by comparing the Learned Perceptual Image Patch Similarity (LPIPS) [75] of the generated and true RF signals. LPIPS is a measure of perceptual similarity between two images, where a smaller value indicates less loss. The average LPIPS of the generated RF spectrograms is low at 10.05%, and the standard deviation is small at 0.94%. This suggests that the generative model tends to produce samples that closely resemble the true RF signals, resulting in a low and consistent level of information loss. These promising results highlight the effectiveness of our IMU to RF translation method.

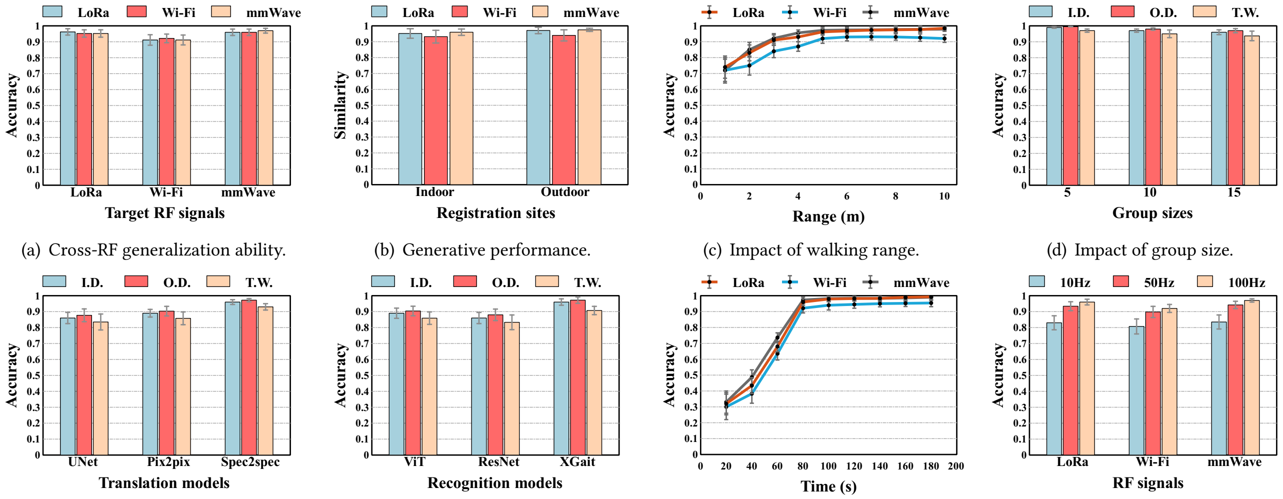
Impact of walking range. Fig. 15(c) demonstrates the effect of accumulated walking range on XGait's performance, showcasing accuracy from 1 m to 10 m. We notice that accuracy improves as the walking range increases. Specifically, the average accuracy rises by 29.85% and 5.34% when walking range expands from 1 m to 5 m and from 5 m to 10 m, respectively. This occurs because longer walking distances capture more information about a person's gait, resulting in higher recognition accuracy. Furthermore, we observe that the mmWave radar achieves the highest accuracy within 7 m, while LoRa excels beyond 7 m. This is due to the mmWave radar's superior short-range sensing capacity, which stems from its high frequency (i.e., 77 GHz) causing increased attenuation and susceptibility to objects. In contrast, LoRa performs better at longer ranges because of low frequency (i.e., 915 MHz) causing increased attenuation and longer sensing range. Wi-Fi achieves lower accuracy in gait recognition primarily due to the use of omnidirectional antennas, which offer wider coverage at the expense of reduced focus and precision. In summary, Fig. 15(c) emphasizes XGait's superior gait recognition performance across various walking ranges.

Impact of group size. We evaluate the impact of group size on average recognition accuracy using various RF signals in XGait. As shown in Fig. 15(d), we can see that the accuracy decreases with larger group sizes. Specifically, it drops by 2.33% and 3.41% as group size rises from 5 to 10 and 10 to 15, respectively. This is due to the increased difficulty of distinguishing individuals in larger groups. However, XGait achieves 95.11% accuracy even with a group size of 15, illustrating its robustness. In summary, the results reveal XGait's strong performance across diverse group sizes.

Impact of Spec2spec generative network. Figure 15(e) demonstrates the impact of the Spec2Spec generative network by comparing the accuracy of using Spec2Spec, UNet [80], and pix2pix [27] in different scenarios. It is evident that Spec2Spec outperforms the other models in all three settings. In particular, compared to UNet, accuracy increases by 9.87%, 8.31%, and 6.62% for indoor, outdoor, and through-wall scenarios, respectively. In contrast to pix2pix, accuracy increases by 7.67%, 6.51%, and 5.12%. These results emphasize Spec2Spec's efficacy in IMU-to-RF translation, resulting in enhanced performance in various environments.

Impact of spectrogram transformer. We evaluate the impact of the proposed spectrogram transformer model by comparing the accuracy of the vision transformer (ViT) [12], residual network (ResNet) [22], and XGait under indoor, outdoor, and through-wall settings. Fig. 15(f) illustrates the performance differences between different recognition models in various environments. Specifically, the average accuracy of XGait outperforms ViT and ResNet by 8.43% and 11.01%, respectively. This result indicates the XGait's efficacy in capturing features for improved recognition performance.

Impact of registration duration. We evaluate the impact of user registration duration on recognition accuracy. As shown in Fig. 15(g), we observe that accuracy improves with increasing duration. Specifically, the average accuracy increases by 64. 21% and 3. 68% when the duration extends from 20 s to 80 s and from 80 s to 180 s, respectively. This is due to the longer duration providing



(a) Cross-RF generalization ability.
(b) Generative performance.
(c) Impact of walking range.
(d) Impact of group size.
(e) Impact of Spec2spec model.
(f) Impact of spectrogram transformer.
(g) Impact of registration duration.
(h) Impact of sampling rate.

Figure 15: Experimental results.

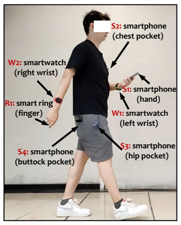


Figure 16: Device positions.

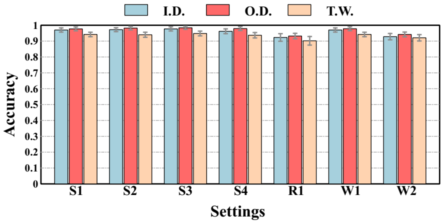


Figure 17: Impact of device positions.

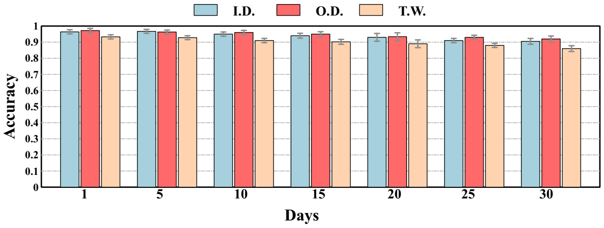


Figure 18: Impact of temporal variability.

more data for training the recognition model. We choose 80 s as it strikes a balance between performance and efficiency.

Impact of sampling rate. In the evaluation, we investigate the impact of sampling rate on the accuracy of XGait recognition results, showcasing results for 10 Hz, 50 Hz, and 100 Hz sampling rates. As shown in Fig. 15(h), the accuracy decreases by about 2.24% when the sampling rate is reduced from 100 Hz to 50 Hz, and it drops more significantly by approximately 8.89% when further lowered to 10 Hz. This substantial decline can be attributed to the loss of crucial human gait information at lower sampling rates, which compromises recognition performance. Based on these findings, it is recommended to use a sampling rate of 50 Hz or higher to ensure accurate results in gait recognition systems.

Impact of device positions. Then, we explore how device positions affect the performance of XGait. We present findings for seven distinct body parts, such as wrist, chest, hip, buttock, hand, and finger, with positions shown in Fig. 16. The left arm remains stationary to simulate the scenario of a person walking while holding a smartphone, while the right arm moves naturally. As shown in Fig. 17, devices on the right waving arm yield lower accuracy compared to devices on the body and the non-waving arm. Our results are consistent with prior research [64], which can be intuitively understood since devices on the body's core register acceleration more consistently than those on moving limbs. This is due to the predominant role of the torso in human gait, while arm movements primarily serve to maintain balance and can vary without significantly affecting a person's normal walking pattern. These results demonstrate promising accuracy across various device positions.

Impact of temporal variability. Human gait exhibits slight variations over time, necessitating the evaluation of XGait's performance in a time-varying context. For this study, we recruited three participants who registered with a smartphone on the first day. Subsequently, we tested the recognition accuracy of these participants every five days for a month. In each scenario, we collected 20 minutes of walking data from the users. As depicted in Fig. 18, the accuracy declines as time progresses. Specifically, the average accuracy decreases by 2.31% and 4.24% when the duration extends from day 1 to day 15 and from day 15 to day 30, respectively. This decline can be attributed to the slight changes in the biological factors affecting the users' gait over time. However, even after 30 days, the accuracy remains at 89.69%, indicating XGait's robustness despite the temporal variability in human gait. For continual performance, considering the evolving nature of gait data distribution, strategies such as periodic retraining or online learning techniques can be used [4].

8.4 Discussion

User feedback. We assessed XGait's usability with 100 subjects answering six SUS-derived questions [5] on ubiquity, security, privacy, efficiency, accuracy, and user-friendliness. Before proceeding with our usability assessment for XGait, we obtained informed consent from all 100 subjects. The questions were: (1) I believe the

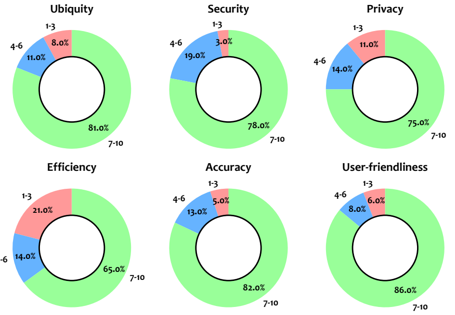


Figure 19: User feedback results. 100 participants joined the user study, consisting of six questions with responses from 1 to 10.

application of the method is ubiquitous; (2) I believe the method is secure; (3) I believe the method is privacy-preserving; (4) I believe the method is efficient; (5) I believe the method is accurate; and (6) I believe the method is user-friendly. Participants rated each aspect on a 1–10 scale. As shown in Fig. 19, XGait received average scores over eight for all questions. Specifically, the average scores for the six dimensions are 8.11 ± 2.18, 8.75 ± 1.66, 8.03 ± 2.26, 8.21 ± 1.79, 8.77 ± 1.05, and 8.58 ± 1.32, respectively. Some users raised efficiency concerns due to the computationally demanding deep generative model. Overall, XGait has favorable usability.

Privacy concern on IMU data. Despite the promising results of our system, it is important to address any potential privacy concerns associated with the use of IMU data. There are indeed privacy risks tied to the collection of IMU data during the registration process, such as potential exposure of user inputs like passwords or revealing speech contents [15, 76]. Nonetheless, our user feedback suggests an overall positive perception of this approach. Moreover, we aim to mitigate privacy risks by capturing and processing IMU data only during walking and with explicit user consent, ensuring our gait analysis respects user privacy. Furthermore, it is worth noting that IMU data from built-in sensors is widely regarded as reliable [18] and is processed locally within a trusted environment [49], without external sharing.

Handling of non-registered users. While our system is primarily aimed at recognizing registered users based on individual gait patterns, the handling of non-registered users introduces additional complexity. Future enhancements could consider incorporating spoofer detection techniques [32] to improve differentiation between registered and non-registered users, ensuring the system's robustness in various usage scenarios.

Range ability. The utilization of distinct RF antennas can lead to variations in sensing ranges due to discrepancies in signal-to-noise ratios (SNR) among different antennas. However, our proposed solution remains effective for long-range recognition tasks when combined with a suitable antenna, thereby expanding its potential application in a broader range of scenarios.

9 CONCLUSION

In this paper, we introduce XGait, the first RF-based gait recognition system that overcomes the limitations of the prior deployment of RF devices and explicit data collection by leveraging the built-in IMUs in mobile devices. This approach enables spontaneous, implicit, and low-effort data collection. XGait incorporates a unified spectrogram generation model, a spectrogram translation model, and a spectrogram transformer recognition model to effectively address the challenges associated with cross-modal settings and ensure efficient and accurate gait recognition. Our extensive evaluation demonstrates the remarkable performance of XGait, achieving over 99% Top-3 accuracy across diverse scenarios.

ACKNOWLEDGMENTS

The work described in this paper was substantially sponsored by project 62101471 supported by NSFC and was partially supported by the Shenzhen Research Institute, City University of Hong Kong. The work was also partially supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 21201420 and CityU 11201422), NSF of Shandong Province (Project No. ZR2021LZH010).

REFERENCES

[1] 2023. XGait demo video. https://youtu.be/6ZePQ1cP4Ho. Accessed: 2023-10-19.
[2] Fadel Adib and Dina Katabi. 2013. See through walls with WiFi!. In *ACM SIGCOMM*.
[3] Tim Althoff, Rok Sosič, Jennifer L Hicks, Abby C King, Scott L Delp, and Jure Leskovec. 2017. Large-scale physical activity data reveal worldwide activity inequality. *Nature* (2017).
[4] Terry Anderson. 2008. *The theory and practice of online learning*. Athabasca University Press.
[5] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability Evaluation in Industry* (1996).
[6] Hanqing Chao, Yiwei He, Junping Zhang, and Jianfeng Feng. 2019. Gaitset: Regarding gait as a set for cross-view gait recognition. In *AAAI*.
[7] Yunzhong Chen, Jiadi Yu, Linghe Kong, Yanmin Zhu, and Feilong Tang. 2022. RFPass: Towards environment-independent gait-based user authentication leveraging RFID. In *IEEE SECON*.
[8] Kaiyan Cui, Qiang Yang, Yuanqing Zheng, and Jinsong Han. 2023. mmRipple: Communicating with mmWave radars through smartphone vibration. In *ACM/IEEE IPSN*.
[9] Omid Dehzangi, Mojtaba Taherisadr, and Raghvendar ChangalVala. 2017. IMU-based gait recognition using convolutional neural networks and multi-sensor fusion. *MDPI Sensors* (2017).
[10] Fani Deligianni, Yao Guo, and Guang-Zhong Yang. 2019. From emotions to mood disorders: A survey on gait analysis methodology. *IEEE JBHI* (2019).
[11] Shuya Ding, Zhe Chen, Tianyue Zheng, and Jun Luo. 2020. RF-net: A unified meta-learning framework for RF-enabled one-shot human activity recognition. In *ACM SenSys*.
[12] 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).
[13] Di Duan, Huanqi Yang, Guohao Lan, Tianxing Li, Xiaohua Jia, and Weitao Xu. 2023. EMGSense: A low-effort self-supervised domain adaptation framework for EMG sensing. In *IEEE PerCom*.
[14] Eric Foxlin. 1996. Inertial head-tracker sensor fusion by a complementary separate-bias Kalman filter. In *IEEE VR*.
[15] Ming Gao, Yajie Liu, Yike Chen, Yimin Li, Zhongjie Ba, Xian Xu, and Jinsong Han. 2022. InertiEAR: Automatic and device-independent IMU-based eavesdropping on smartphones. In *IEEE INFOCOM*.
[16] Yao Ge, Wenda Li, Muhammad Farooq, Adnan Qayyum, Jingyan Wang, Zikang Chen, Jonathan Cooper, Muhammad Ali Imran, and Qammer H Abbasi. 2023. LoGait: LoRa sensing system of human gait recognition using dynamic time wraping. *IEEE Sens. J.* (2023).
[17] Tao Gu, Liang Wang, Hanhua Chen, Xianping Tao, and Jian Lu. 2011. Recognizing multiuser activities using wireless body sensor networks. *IEEE TMC* (2011).
[18] Le Guan, Jun Xu, Shuai Wang, Xinyu Xing, Lin Lin, Heqing Huang, Peng Liu, and Wenke Lee. 2016. From physical to cyber: Escalating protection for personalized auto insurance. In *ACM SenSys*.
[19] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. 2011. Tool release: Gathering 802.11n traces with channel state information. *ACM SIGCOMM Comput. Commun. Rev.* (2011).
[20] Mingda Han, Huanqi Yang, Tao Ni, Di Duan, Mengzhe Ruan, Yongliang Chen, Jia Zhang, and Weitao Xu. 2023. mmSign: mmWave-based few-shot online handwritten signature verification. *ACM TOSN* (2023).

[21] Harish Haresamudram, Irfan Essa, and Thomas Plötz. 2021. Contrastive predictive coding for human activity recognition. *ACM IMWUT* (2021).

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *IEEE CVPR*.

[23] Pengfei Hu, Yifan Ma, Panneer Selvam Santhalingam, Parth H Pathak, and Xiuzhen Cheng. 2022. Milliear: Millimeter-wave acoustic eavesdropping with unconstrained vocabulary. In *IEEE INFOCOM*.

[24] Pengfei Hu, Hui Zhuang, Panneer Selvam Santhalingam, Riccardo Spolaor, Parth Pathak, Guoming Zhang, and Xiuzhen Cheng. 2022. Accear: Accelerometer acoustic eavesdropping with unconstrained vocabulary. In *IEEE S&P*.

[25] Aapo Hyvarinen. 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE TNN* (1999).

[26] Cesar Iovescu and Sandeep Rao. 2017. The fundamentals of millimeter wave sensors. *Texas Instruments* (2017), 1–8.

[27] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *IEEE CVPR*.

[28] Sijie Ji and Mo Li. 2021. CLNet: Complex input lightweight neural network designed for massive MIMO CSI feedback. *IEEE Wireless Commun. Lett.* (2021).

[29] Sijie Ji, Yaxiong Xie, and Mo Li. 2022. SiFall: Practical online fall detection with RF sensing. In *ACM SenSys*.

[30] Shang Jiang, Jianguo Jiang, Siye Wang, Yanfang Zhang, Yue Feng, Ziwen Cao, and Yi Liu. 2022. RF-Gait: Gait-based person identification with COTS RFID. *WCMC* (2022).

[31] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[32] Hao Kong, Li Lu, Jiadi Yu, Yanmin Zhu, Feilong Tang, Yi-Chao Chen, Linghe Kong, and Feng Lyu. 2022. Push the limit of wifi-based user authentication towards undefined gestures. In *IEEE INFOCOM*.

[33] 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 *ACM MobiCom*.

[34] Seung Hoon Lee, Seunghyun Lee, and Byung Cheol Song. 2021. Vision transformer for small-size datasets. *arXiv preprint arXiv:2112.13492* (2021).

[35] Shuangqun Li, Wu Liu, and Huadong Ma. 2019. Attentive spatial–temporal summary networks for feature learning in irregular gait recognition. *IEEE TMM* (2019).

[36] Shuangqun Li, Wu Liu, Huadong Ma, and Shaopeng Zhu. 2018. Beyond view transformation: Cycle-consistent global and partial perception gan for view-invariant gait recognition. In *IEEE ICME*.

[37] Rex Liu, Albara Ah Ramli, Huanle Zhang, Erik Henricson, and Xin Liu. 2022. An overview of human activity recognition using wearable sensors: Healthcare and artificial intelligence. In *Springer ICIOT*.

[38] Yasushi Makihara, Atsuyuki Suzuki, Daigo Muramatsu, Xiang Li, and Yasushi Yagi. 2017. Joint intensity and spatial metric learning for robust gait recognition. In *IEEE CVPR*.

[39] 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 *AAAI*.

[40] Mehdi Mirza and Simon Osindero. 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* (2014).

[41] M Patricia Murray, Ross C Kory, and Bertha H Clarkson. 1969. Walking patterns in healthy old men. *J. Geront.* (1969).

[42] Muhammed Zahid Ozturk, Chenshu Wu, Beibei Wang, and KJ Ray Liu. 2021. GaitCube: Deep data cube learning for human recognition with millimeter-wave radio. *IEEE IoTJ* (2021).

[43] Ronald K Pearson. 2002. Outliers in process modeling and identification. *IEEE Trans. Control Syst. Technol.* (2002).

[44] Matjaž Perc. 2005. The dynamics of human gait. *European Journal of Physics* (2005).

[45] Donald B Percival and Andrew T Walden. 2000. *Wavelet methods for time series analysis.* Cambridge University Press.

[46] Sandeep Rao. 2017. Introduction to mmWave sensing: FMCW radars. *Texas Instruments (TI) mmWave Training Series* (2017).

[47] Yanzhi Ren, Yingying Chen, Mooi Choo Chuah, and Jie Yang. 2014. User verification leveraging gait recognition for smartphone enabled mobile healthcare systems. *IEEE TMC* (2014).

[48] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional networks for biomedical image segmentation. In *MICCAI*.

[49] Mohamed Sabt, Mohammed Achemlal, and Abdelmadjid Bouabdallah. 2015. Trusted execution environment: What it is, and what it is not. In *IEEE TrustCom*.

[50] Zhiyao Sheng, Huatao Xu, Qian Zhang, and Dong Wang. 2022. Facilitating radar-based gesture recognition with self-supervised learning. In *IEEE SECON*.

[51] Cong Shi, Jian Liu, Hongbo Liu, and Yingying Chen. 2017. Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT. In *ACM MobiHoc*.

[52] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[53] Wang Tianben, Zhu Wang, Daqing Zhang, Tao Gu, Hongbo Ni, Jiangbo Jia, Xingshe Zhou, and Jing Lv. 2016. Recognizing Parkinsonian gait pattern by exploiting fine-grained movement function features. *ACM TIST* (2016).

[54] Ruben Vera-Rodriguez, John SD Mason, Julian Fierrez, and Javier Ortega-Garcia. 2012. Comparative analysis and fusion of spatiotemporal information for footstep recognition. *IEEE TPAMI* (2012).

[55] Jie Wang, Qinhua Gao, Miao Pan, and Yuguang Fang. 2018. Device-free wireless sensing: Challenges, opportunities, and applications. *IEEE Network* (2018).

[56] Liang Wang, Tao Gu, Hanhua Chen, Xianping Tao, and Jian Lu. 2010. Real-time activity recognition in wireless body sensor networks: From simple gestures to complex activities. In *IEEE RTCSA*.

[57] Wei Wang, Alex X Liu, and Muhammad Shahzad. 2016. Gait recognition using WiFi signals. In *ACM UbiComp*.

[58] Xuan Wang, Tong Liu, Chao Feng, Dingyi Fang, and Xiaojiang Chen. 2023. RF-CM: Cross-modal framework for RF-enabled few-shot human activity recognition. *ACM IMWUT* (2023).

[59] Xuyu Wang, Chao Yang, and Shiwen Mao. 2017. PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity Wi-Fi devices. In *IEEE ICDCS*.

[60] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: From error visibility to structural similarity. *IEEE TIP* (2004).

[61] Chenshu Wu, Feng Zhang, Yuqian Hu, and KJ Ray Liu. 2020. GaitWay: Monitoring and recognizing gait speed through the walls. *IEEE TMC* (2020).

[62] Weitao Xu, Guohao Lan, Qi Lin, Sara Khalifa, Mahbub Hassan, Neil Bergmann, and Wen Hu. 2018. KEH-Gait: Using kinetic energy harvesting for gait-based user authentication systems. *IEEE TMC* (2018).

[63] Weitao Xu, Girish Revadigar, Chengwen Luo, Neil Bergmann, and Wen Hu. 2016. Walkie-talkie: Motion-assisted automatic key generation for secure on-body device communication. In *ACM/IEEE IPSN*.

[64] Weitao Xu, Yiran Shen, Yongtuo Zhang, Neil Bergmann, and Wen Hu. 2017. Gait-watch: A context-aware authentication system for smart watch based on gait recognition. In *ACM/IEEE IoTDI*.

[65] Yang Xu, Wei Yang, Min Chen, Sheng Chen, and Liusheng Huang. 2020. Attention-based gait recognition and walking direction estimation in Wi-Fi networks. *IEEE TMC* (2020).

[66] Huanqi Yang, Mingda Han, Shuyao Shi, Zhenyu Yan, Guoliang Xing, Jianping Wang, and Weitao Xu. 2023. Wave-for-safe: Multisensor-based mutual authentication for unmanned delivery vehicle services. In *ACM MobiHoc*.

[67] Qiang Yang, Kaiyan Cui, and Yuanqing Zheng. 2023. VoShield: Voice liveness detection with sound field dynamics. In *IEEE INFOCOM*.

[68] 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 INFOCOM*.

[69] Syed Usama Yunas, Abdullah Alharthi, and Krikor B Ozanyan. 2019. Multi-modality fusion of floor and ambulatory sensors for gait classification. In *IEEE ISIE*.

[70] Yunze Zeng, Parth H Pathak, and Prasant Mohapatra. 2016. WiWho: WiFi-based person identification in smart spaces. In *ACM/IEEE IPSN*.

[71] Youwei Zeng, Dan Wu, Jie Xiong, Enze Yi, Ruiyang Gao, and Daqing Zhang. 2019. FarSense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas. *ACM IMWUT* (2019).

[72] Fusang Zhang, Zhaoxin Chang, Kai Niu, Jie Xiong, Beihong Jin, Qin Lv, and Daqing Zhang. 2020. Exploring LoRa for long-range through-wall sensing. *ACM IMWUT* (2020).

[73] Jin Zhang, Zhuangzhuang Chen, Chengwen Luo, Bo Wei, Salil S Kanhere, and Jianqiang Li. 2022. MetaGanFi: Cross-domain unseen individual identification using WiFi signals. *ACM IMWUT* (2022).

[74] Jin Zhang, Bo Wei, Wen Hu, and Salil S Kanhere. 2016. WiFi-ID: Human identification using WiFi signal. In *IEEE DCOSS*.

[75] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE CVPR*.

[76] Shijia Zhang, Yilin Liu, and Mahanth Gowda. 2023. I spy you: Eavesdropping continuous speech on smartphones via motion sensors. *ACM IMWUT* (2023).

[77] Yuting Zhang, Gang Pan, Kui Jia, Minlong Lu, Yueming Wang, and Zhaohui Wu. 2014. Accelerometer-based gait recognition by sparse representation of signature points with clusters. *IEEE Trans. Cybern.* (2014).

[78] Yi Zhang, Yue Zheng, Guidong Zhang, Kun Qian, Chen Qian, and Zheng Yang. 2020. GaitID: Robust Wi-Fi based gait recognition. In *Springer WASA*.

[79] 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 TOSN* (2021).

[80] R ZHAO, J Yu, T Li, H Zhao, and CHE Ngai. 2022. Radio2Speech: High quality speech recovery from radio frequency signals. *INTERSPEECH* (2022).

[81] Han Zou, Yuxun Zhou, Jianfei Yang, Weixi Gu, Lihua Xie, and Costas Spanos. 2018. WiFi-based human identification via convex tensor shapelet learning. In *AAAI*.

[82] Qin Zou, Yanling Wang, Qian Wang, Yi Zhao, and Qingquan Li. 2020. Deep learning-based gait recognition using smartphones in the wild. *IEEE TIFS* (2020).