Gate-ID: WiFi-based Human Identification Irrespective of Walking Directions in Smart Home

Jin Zhang, Bo Wei, Fuxiang Wu, Limeng Dong, Wen Hu, Salil S. Kanhere, Chengwen Luo, Shui Yu, Jun Cheng, *Member, IEEE*

***Abstract*—Research has shown the potential of device-free WiFi sensing for human identification. Each and every human has an unique gait and prior works suggest WiFi devices are able to capture the unique signature of a person's gait. In this paper, we show for the first time that the monitored gait could be inconsistent and have mirrorlike perturbations when individuals walk through WiFi devices in different directions, provided that WiFi antenna array is horizontal to walking path. Such inconsistent mirrored patterns are to negatively affect the uniqueness of gait and accuracy of human identification. Therefore, we propose a system called Gate-ID for accurately identifying individuals' identities irrespective of different walking directions. Gate-ID employs theoretical communication model and real measurements to demonstrate that antenna array orientations and walking directions contribute to the mirrorlike patterns in WiFi signals. A novel heuristic algorithm is proposed to infer individual's walking directions. A set of methods are employed to extract and augment the representative spatial-temporal features of gait and enable the system performing irrespective of walking directions. We further propose a novel attention-based deep learning model that fuses various weighted features and ignores ineffective noises to uniquely identify individuals. We implement Gate-ID on commercial off-the-shelf devices. Extensive experiments demonstrate that our system can uniquely identify people with average accuracy of 90.7% to 75.7% from a group of 6 to 20 people, respectively, and improve the accuracy by 12.5%-43.5% compared with baselines.**

***Keywords*-WiFi, channel state information, human identification, neural networks**

I. Introduction

WIFi-integrated devices such as smart lamps, refrigerators, TV etc. are ubiquitous in modern urban indoor spaces like smart home and smart offices. These WiFi devices communicate with each others meanwhile invisibly fill the air with Radio Frequency (RF) signals as a WiFi spectrum. When in this RF space, a person's motions create perturbations that can be captured and utilized for recognizing daily activities such as standing, sitting, walking and running [1] and even hand gestures [2] and breathing [3].

Many existing works exploit the reception of WiFi signals from the ubiquitous WiFi-integrated devices for human identification. Everyone's natural walking style (i.e. gait) is unique which is characterized by the differences in the limb (hand and feet) movement patterns and velocity [4]. WiFi-ID [5] [6] first demonstrated the WiFi signal could be used to capture the unique signature of a person's gait in WiFi spectrum and thereafter identify the individual's identity. WiWho [7] extracted various features of each gait cycles (steps) in order to recognize identities. WifiU [8] analyzed the patterns of WiFi signals in frequency domain to estimate the impact of torso and legs on the WiFi spectrum. XModal-ID [9] exploited video and WiFi to realize cross-modal human identification. GateWay [10] extracted gait cycles and estimate each step's speed that used as the unique patterns to recognize identities.

While previous researches are promising, there exist several challenges that needed to be addressed. Prior works [7], [8], [9], [10] require an individual to be monitored to walk over a long distance (5m-10m) in a corridor or an open area and the recorded WiFi signals tend to contain periodical rhythm that treated as gait cycles for tracking walking speed and calculating unique patterns of gait. However, such common assumptions may not hold true in real-world environment. First, the smart home such as the modern apartment depicted in Fig. 1 need personalization or authentication services in each room but often not spacious enough to meet the requirement of the prior works. Second, the periodical rhythm of gait cycles may not exist in WiFi signals. As shown in Fig. 2, it is hard to observe periodical patterns in WiFi signals in a brief moment when individuals walk through WiFi transceivers. Third, prior works mostly employ traditional machine learning or shallow deep learning (DL) model for recognition that only support limited number ($<$11) of individuals to be classified. Moreover, our theoretical model and comprehensive experiments show that an individual's gait manifested as fluctuation of WiFi signals could be inconsistent and have mirrorlike patterns in different walking directions. This phenomenon will happen provided the antenna arrays of two WiFi devices are horizontal to individuals' walking trajectories. The consequent inconsistency negatively influence the uniqueness of gait.

The general problem of uniquely identifying an individual

Manuscript received September, 22, 2020; revised October, 21, 2020, November, 14, 2020; accepted November, 20, 2020.

Jin Zhang, Fuxiang Wu, Jun Cheng are with Guangdong-Hong Kong-Macao Joint Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China. Email: {jin.zhang, fx.wu1, jun.cheng}@siat.ac.cn; Bo Wei is with Computer and Information Sciences, Northumbria University, UK. Email: bo.wei@northumbria.ac.uk; Limeng Dong is with School of Information and Communications Engineering, Xi'an Jiaotong University, China. Email: dlm_nwpu@hotmail.com; Wen Hu, Salil S. Kanhere are with Computer Science and Engineering, University of New South Wales, Australia. Email: {wen.hu, salil.kanhere}@unsw.edu.au; Chengwen Luo is with the College of Computer Science and Software Engineering, Shenzhen University, China, Email: chengwen@szu.edu.cn; Shui Yu is with the School of Computer Science, University of Technology Sydney, Australia, Email: shui.yu@uts.edu.au; Corresponding author: Jun Cheng Email: jun.cheng@siat.ac.cn

This work was supported in part by National Natural Science Foundation of China (61772508, U1713213); part by Shenzhen Technology Project (JCYJ20170413152535587, JCYJ20180507182610734), CAS Key Technology Talent Program; part by Australia ARC DP180102828 and ARC DP200101374.

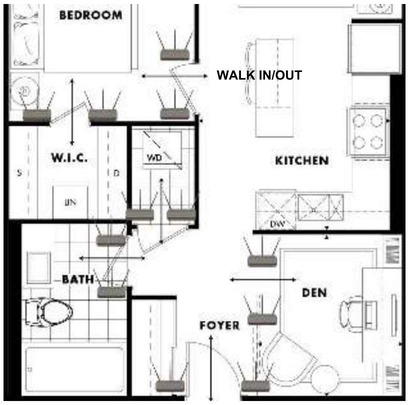


Fig. 1: A Gate-ID application scenario deployed in smart home.

in any physical setting is arguably very challenging. To make the problem more tractable and applicable in realistic environment, we consider a simple yet common application scenario as depicted in Fig. 1. Consider a smart home or smart office where the household may be recognized once entering a room and provided with personalized services (e.g. turning on TV in the den for kids), or selected staff may be prohibited to enter and leave offices, etc. The goal is to recognize individuals' identities from a group of people when they walk in or out a space (i.e. in two possible directions) through a door. A pair of WiFi transceiver devices could be deployed next to the door as a virtual gate to capture human gait for a brief moment once individuals cross the gate so as to achieve personalization or authentication services.

In this paper, we propose Gate-ID, a WiFi based human identification system that is deployed on two commercial off-the-shelf WiFi-integrated devices. The two devices are placed at the edged sides of a door about 1 meters away and form a virtual gate as shown in Fig. 1. When a person walks through the gate in either directions, the Gate-ID system is able to capture the impact of human body on WiFi spectrum and identify the individual's identity. As detailedly illustrated in Fig. 2, the system consists of two WiFi devices, one device as the transmitter continuously broadcasts packets to the other device as the receiver which passively records CSI data from the received packets. When a person in the vicinity of the gate, the walk style (i.e. gait) impacts WiFi spectrum in a unique manner, that in turn manifested as unique signatures of the person in the CSI data which allow recognizing this person's identity. There are several challenges in realizing the Gate-ID system as follows.

The CSI time series constitutes data from multiple transmit (N_t) and receive (N_r) antenna pairs each comprised of multiple subcarriers (e.g., 30 OFDM subcarriers for 802.11n). The first challenge is to analyze how human gait affect the fluctuations of WiFi signals. We theoretically model the WiFi propagation and demonstrate the signature of human gait reside in superposition of multiple radio phenomena, i.e. reflection, diffraction, scattering, absorption. Considering the placement of WiFi antenna array affect the propagation paths of WiFi packets, we employ the reflection model to study the impact of antenna array orientations and walking directions. Our simulation and real measurements demonstrate that a person's gait in two walking directions (back and forth) are pronounced as the symmetrical and mirrorlike perturbations in WiFi CSI, when WiFi antenna arrays in the horizontal orientation. In contrast, the person's gait in the two directions are identical when WiFi antenna arrays in the vertical placement. Additional efforts are made in the following to alleviate the inconsistent patterns of human gait due to varying walking directions and antenna array orientation. Moreover, we make best use of the mirrored patterns to infer the person's walking directions. We calculate the short time energy of CSI streams and employ a least-square optimization algorithm to accurately determine individuals' direction.

The second challenge is to calculate representative features of human gait. To capture the unique patterns of each individual, multiple spatial-temporal features are calculated from the CSI streams to extract the uniqueness of each individual's gait. Since a person's gait could be symmetrical in different walking directions, we employ a feature augmentation method that flipping and reversing data to generate synthetic dataset in opposite walking directions to encourage the identification model to become invariant of the mirrorlike patterns.

The third challenge is to accurately identify individuals, so that a novel attention-based deep learning model is designed for human identification. The DL model constitutes of a novel attention-based feature fusion mechanism and stack of neural networks. Various types of calculated features may not be equally useful, therefore we firstly adopt ReliefF, a feature weighting method to estimate the importance weights of each type of features. The attention mechanism exploits the ReliefF feature weights as a priori knowledge and employs the feature projecting network to ensure the DL model focus on most effective features and ignore ineffective ones. The DL model further employs residual neural network and bidirectional long short-term memory neural network to extract spatial-temporal patterns of gait for uniquely identifying individuals. The dropout method is used to against the overfitting problem when the DL model applied on the small-size dataset.

In summary, we made the following contributions:

- Design and implement a WiFi based human identification system that mitigate the impact of walking directions and WiFi antenna array orientations, and captures unique features of human gait to accurately identify people.
- Demonstrate by using theoretical communication model and real measurements that the WiFi CSI signatures of human gait could be inconsistent and have symmetrical fluctuations in WiFi spectrum due to different walking directions and antenna array orientations. To make best use of the symmetrical patterns of gait, a novel optimization algorithm is proposed to estimate the walking directions.
- Extract representative spatial-temporal features of human gait and design a feature augmentation method that generates synthetic CSI data in opposite walking directions to mitigate the mirrorlike inconsistency of gait.
- Design a novel attention-based deep learning model for weighting and fusing prior calculated features and uniquely identifying individuals. Extensive experiments

show that Gate-ID can identify a person in real-world environment with an average accuracy from 90.7% to 75.7% from a group size of 6-20 people, respectively, and improve the identification accuracy by 12.5%-43.5% compared with two baselines.

The remaining paper is organized as follows. Section II discusses related works in human identification, radio sensing and wifi-based human identification areas. Section III explains our proposed Gate-ID system. Section III-A discusses the WiFi CSI and analyzes the wireless communication phenomena impacted by human gait. Section III-B1 discusses the silence removal to segment individuals' walk motions. Section III-B2 presents the theoretical model to analyze the effect of walking directions and antenna array orientations on CSI patterns. Section III-B3 presents an optimization algorithms to predict walking directions. Section III-C explains the feature extractions and augmentation approaches. Section III-D proposes the attention-based feature fusion mechanism and the deep learning model. Section IV discusses the implementation and comprehensively evaluate the system. Section V concludes the paper.

II. Related Work

A. Human identification

Researchers have researched human identification for a decade and developed various types of methods. The common widely deployed method is the vision-based sensor. [11] and [12] use video cameras to record silhouette of people and extract gait patterns for identifications. While the camera-based approaches achieve well performance, they mostly are considered to be too intrusive (from the perspective of privacy) for use in residential place. Prior works also exploit wearable [13] [14] and floor sensors [15] [16] to achieve gait monitoring, in which the accelerometers are used to track human body motion and floor vibration. Nevertheless, wearable and floor sensors should be carried on the interest of subject or deployed in a large area which take much effort and cost. Recent researchers exploit acoustic sensors i.e. microphone and speaker to recognize individuals. The acoustic signals [17] [18] are mechanical waves and easily capture Doppler effect incurred by human gait. However, due to the high attenuation characteristic of mechanical waves, the clothing of monitored individuals and environment acoustic noise could significantly affect the performance, which limit its applications in realistic environment. Therefore, it is desired to explore a non-intrusive pervasive and high performance approach for human identification.

B. Radio sensing

Wireless radio has been used for monitoring human activity and developed various types of methods. The recent advanced microwave radars [19] have integrated radar front end and micro processing unit in a single chip which is portable and powerful enough for recognizing human motions. In [20], the micro Doppler radar is used for monitoring gait. The microwave radar is further used for tracking multiple persons in [21], and monitoring hand and finger gestures [22]. However, the application of microwave radars is still limited due to the high costs of radar front end and signal processing units and the strict radio regulations in high frequency bands applied in indoor environment. The CSI data contain rich information from every subcarrier of WiFi signal, and the perturbations in each subcarriers could be used to recognize various human activities such as sitting, walking and running [1] [23]. Prior works employ deep learning model and signal processing techniques to recognize hand gestures [24] and breathing [3]. Herein, our work show that WiFi CSI could be used to monitor the intrinsic patterns of human gait irrespective of walking directions and uniquely identify individuals.

C. WiFi-based human identification

We have witnessed a growing interest in exploiting WiFi signals for human identification and gait recognition. WifiU [8] analyzed WiFi spectrogram to estimate the torso and leg speeds and classifies the uniqueness of each individual. CrossSense [25] proposed to use transfer learning in achieving gait recognition across different sites. CrossSense and WifiU have decent identification performance, though they have a common assumption that the monitored individual should walk in a predefined single direction, which is not applicable in realistic environment. WiWho [7] calculated various patterns of gait cycles (steps) for recognizing human identities. XModal-ID [9] exploited the cross-modal (video and WiFi) method to realize human identification. GateWay [10] exploited the gait cycles' speed as unique features to identify people. These prior works require the monitored individuals to walk over a long distance (5m-10m) and expect the existence of periodical rhythm in WiFi CSI as discussed in Section I. The peak finding and autocorrelation approaches are applied either on CSI time series or spectrograms to segment gait cycles and estimate walking speed and calculate patterns of each step. However, it is only a brief moment when individuals cross a door to enter or leave a room in realistic indoor environment as shown in Fig. 1 and Fig. 2. In fact, it is hard to observe the periodical patterns in temporal and spatial domain of WiFi CSI signals as illustrated in Fig. 6. On the contrary, our work is able to effectively capture the unique signature of human gait in such brief moment i.e. a short period (around 3s) once individuals walk in/out a room through a door and pass WiFi transceivers. Recent works makes use of standard deep learning methods [26] [27] such as autoencoder and convolutional neural networks and convex tensor shapelet learning [28] to capture unique features of human gait. Our work combines feature extraction, augmentation and a novel attention based deep learning model to capture the uniqueness of gait irrespective of walking directions and accurately identify individuals out of a large group of people. Moreover, our work theoretically analyze how human gait impact WiFi spectrum and employ communication model to demonstrate human gait captured in WiFi CSI could be inconsistent on the impact of antenna array orientation and walking directions. Thus, we believe our proposed system is an important progress to WiFi based human identification.

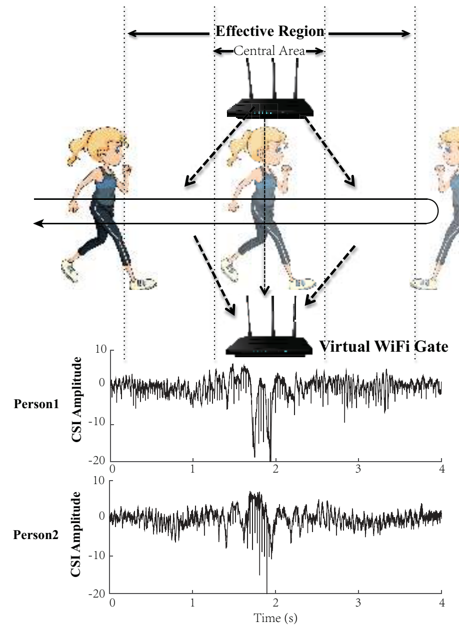


Fig. 2: Operational scenario for Gate-ID. Two WiFi devices form a virtual WiFi gate for recognizing individuals' identities who walk through the gate in two possible (i.e. back and forth) directions. The WiFi CSI of two persons' gait have unique patterns in the central area of the effective region.

III. Gate-ID System

The operational scenario of Gate-ID is depicted in Fig. 2. Gate-ID involves two WiFi devices, one transmitter broadcasts WiFi packets and one receiver captures the packets and extracts CSI data as mentioned in Section III-A1. In operations, the individuals are asked to cross the Line of Sight (LoS) of transceivers and walk through the virtual gate as shown in Fig. 2. The individuals are allowed to walk back and forth randomly in both two directions. The recorded CSI data are processed to extract a set of features and a deep learning mode is used to uniquely identify the monitored individuals.

Fig. 3 depicts the architecture of Gate-ID as follows.

- The preliminary study in Section III-A outlines the theoretical background in which Section III-A1 discusses the WiFi CSI and Section III-A2 presents the WiFi propagation model impacted by human gait constitutes of multiple radio phenomena.
- The gait uniqueness analysis in Section III-B segments WiFi CSI that effectively capture the human gait illustrated in Section III-B1, theoretically model the WiFi CSI of human gait and analyze the impact of antenna orientation and walking directions detailed in Section III-B2, and estimate the moving directions based on the prior gait model discussed in Section III-B3.
- Gate-ID calculates the representative spatial-temporal features from the CSI data and generates synthetic CSI dataset in reversed order to against the effect of different walking directions, as discussed in Section III-C.
- Gate-ID designs a novel attention-based deep-learning model with the built-in feature fusion mechanism for extracting representative features of human gait and uniquely identifying individuals as discussed in Section III-D.

A. Preliminary Study

1) Channel State Information (CSI): Modern off-the-shelf WiFi devices support the IEEE 802.11n/ac standard and equip over three antennas for MIMO communications. The WiFi NICs continuously monitor the frequency response of OFDM subcarriers as CSI at PHY layer [29] and in turn dynamically adapt to external radio environment. CSI capture information in each subcarriers between each transceiver antenna pairs, therefore the multiple wireless phenomena such as frequency selective fading, shadowing, multipath etc. could be captured and utilized for monitoring human body. Our WiFi MIMO system constitutes $N_t = 3$ transmit antennas and $N_r = 3$ receive antennas, thus $N_t \times N_r$ antenna pairs. The frequency response for subcarrier i and antenna pair p is denoted as Y_i^p and X_i^p, and the Channel Frequency Response (CFR) is denoted as H_i^p. Then,

$$Y_i^p = H_i^p \times X_i^p \quad i \in [1, C] \quad p \in [1, N_t \times N_r], \tag{1}$$

H_i^p is a complex value and $\|H_i^p\|$ simplified as h_i^p denote its amplitude. The time series of h_i^p are called CSI streams. The WiFi driver NIC [30] in our experiments contain 30 (i.e. C = 30) OFDM subcarriers between each antenna pair. Therefore, the dimension of the CSI streams is $30 \times N_t \times N_r$.

2) WiFi Propagation Model: As an individual in the vicinity of WiFi devices the presence of moving human body interfere the propagation channels between WiFi transceivers and incur fluctuations in WiFi CSI. As such the characteristics of WiFi propagation channel need to be theoretically modeled in order to capture the unique patterns of individuals' gait. The human body in an indoor radio environment is a strong reflector for electromagnetic waves and can be approximated by a conducting circular cylinder [31]. The different position of human body relative to the LoS of transceivers introduce multiple complex radio phenomena, i.e. reflection, diffraction, scattering, absorption, as follows,

Reflection occurs when the human body is 0.5m away from LoS [31], because the dimension of human body is significantly larger than the WiFi signal wavelength. Reflection mostly accounts for the fluctuation of WiFi signals in the *effective region* out of *central area* shown in Fig. 2. The power of reflected and LoS WiFi rays dissipate as the radio signal propagates in space. Thus, the following path loss model [32] could be used to simulate the reflection and LoS radio paths $H_{r|los}$,

$$H_{r|LoS} = \frac{P_r}{P_t} = \frac{\lambda^2}{16\pi^2 d^2} \cdot \frac{f_{max} - f_{min}}{B} \cdot G_{tx} \cdot G_{rx}, \tag{2}$$

where P_r is the received power, P_t is the transmit power, λ is the wavelength of light, B is WiFi bandwidth, f_{max} and f_{min} are the maximum and minimum frequency of WiFi signals respectively, G_{tx} and G_{rx} are transmit and receive antenna gains respectively. The prior parameters are all constants,

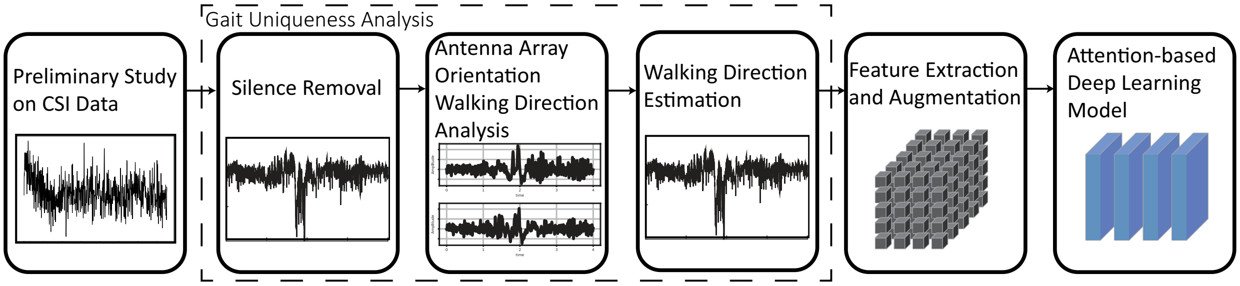


Fig. 3: Overview of the Gate-ID system.

therefore The attenuation of radio signals is a function of propagation distances d. The position of human body relative to the antenna of transceiver affect WiFi propagation that is to be discussed in Section III-B2.

Diffraction (often termed as shadowing) occurs when the human body block or is close to LoS, because the diffracted signals can reach receivers that shadowed by the human body [33]. Diffraction mostly corresponds to the large scale fluctuations of WiFi signals in the *central area* outlined in Fig. 2. The power of diffracted signal H_d is a function of distances from the obstacle to transmitter d_{ic} and receiver d_{jc} and the height of the obstacle d_c [34], as follows,

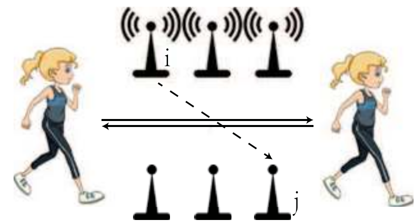
$$\begin{aligned} H_d &= \left| \int_v^\infty \exp(\frac{-\mathbf{J} \cdot \pi z^2}{2}) \mathrm{d}z \right|, \\ v &= d_t \sqrt{2(d_{ic} + d_{jc})/(\lambda \cdot d_{ic} \cdot d_{jc})}. \end{aligned} \tag{3}$$

Scattering occurs when WiFi signals impinge on rough surface of human body and clothing, and the scattered energy spread out in all directions [33]. **Absorption** occurs when the torso is close to transmitter. The peak energy absorption rate (SAR) in the torso at 5GHz using 100 mW transmit power is 0.399 mW kg^{-1} when an individual sits still 34cm away in front of a WiFi device [35]. Thus, scattering and absorption denoted as H_{sa} capture the characteristics of human body surface and weight which have moderate effect on the WiFi propagation.

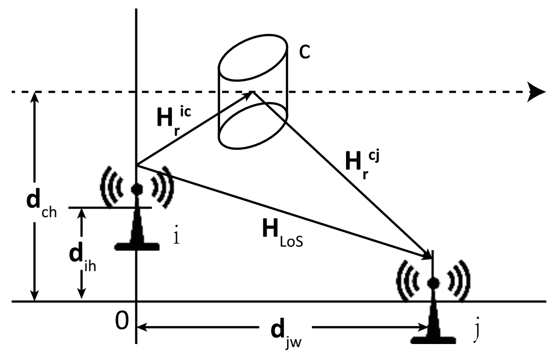
The aggregated WiFi CSI amplitude constitute of all radio phenomena mentioned above and the CSI amplitude h_i^p is expressed as follows,

$$h_i^p = H_{LoS} + H_r + H_d + H_{sa} + H_n, \tag{4}$$

where H_n is the embedded system and ambient noise. We assume the transceivers remain static and the distance between the two devices is unchanged, thus H_{LoS} a constant value if the LoS is not blocked. Therefore, the signature of human gait in CSI is carried in $H_r + H_d + H_{sa}$ and the methods introduced in the following are to remove noise and capture unique features of individuals' gait.



(a) Horizontal antenna array orientation.



(b) WiFi propagation model subject to human body reflection.

Fig. 4: An ideal operation scenario and corresponding propagation model.

B. Gait Uniqueness Analysis

1) Silence Removal and Segmentation: In applications such as authentication and personalization, the WiFi transmitter and receiver should operate continuously and the receiver keeps recording CSI data from the WiFi packets sent by the transmitter. In the absence of individuals around the devices, the CSI data are to capture the ambient noise in the vicinity which interfere the identification model and need to be discarded. Moreover, as shown in our application scenario in Fig. 1, the door of each room are quite close in the residential apartment, thus the CSI data captured over a long distance may not belong to the area of interest where WiFi devices deployed. The meaningful CSI data correspond to the short period when an individual cross the LoS of WiFi devices through the virtual gate. This particular monitored area locates in the *central area* and *effective region* shown in Fig. 2, where the limb motions mostly impact WiFi spectrum and generate unique signatures in H_r, H_d, and H_{sa}. Herein, the key challenge is to locate the *effective region* and determine the length (or duration in terms of time denoted as the parameter T) of this effective region. A large time period could include unintentional individuals who move around the monitored area and bring unrelated external interference. In contrast, a too short region may not capture sufficient features of individuals' gait for accurate recognition. To achieve this, we employ the silence removal algorithm mentioned in our prior work [5] to segment the CSI data of duration T denoted as $\mathcal{H}$ which represents walking observation dataset. Section IV-C evaluates the effect of varying duration of T.

2) Antenna Array Orientation and Walking Direction: Modern WiFi devices employ the linear antenna array to support MIMO mechanism. The orientation and placement of the WiFi antenna array could affect the length and directions of radio propagation paths and thus influence the sensing behavior of WiFi spectrum. On the other hand, when entering or leaving a space through a gate in smart home, an individual would walk in two arbitrary directions, i.e. in or out of room. In certain situations the distance from human body to WiFi antenna array in different walking directions may not be consistent and identical. Therefore, the both factors antenna array orientations and walking direction together could affect the uniqueness of human gait captured in WiFi CSI data.

To investigate the impact of the both factors, we assume a typical operation scenario where the linear antenna array of WiFi transceivers are in horizontal orientation and parallel with an individual's walking trajectories shown in Fig. 4(a). The signature of this individual's gait captured in WiFi CSI reside in the superposition of $H_r + H_d + H_{sa}$ mentioned in Section III-A2. As for the diffraction H_d, the height d_t and the distances to transceivers d_{it}, d_{jt} keep identical in different walking directions when the human body blocks LoS. The body shape, weight and clothing also remain unchanged in a limited time period. Therefore, H_d and H_{sa} are not affected by the antenna orientation and walking directions. As for the reflection H_r, we observe that the propagation distances of reflected paths could be inconsistent in different walking directions especially for the channels between the diagonal transmit antenna i and receive antenna j as shown in Fig. 4(a). A theoretical WiFi propagation model outlined in Fig. 4(b) is proposed in the following to study the impact of the human body reflection phenomenon on WiFi CSI.

Consider the horizontal antenna orientation scenario including one source (transmit antenna i), one reflector (conducting circular cylinder c referred to as human body), one receiver (receive antenna j) as depicted in Fig. 4(b). d_{jw} denotes the distance between first and third antenna in horizontal. d_{ih} denotes the distance between two linear antenna array in vertical. d_{ch} denotes the distance from the trajectory of c to antenna j in vertical. We assume maximal reflection by c without loss and phase shift, and the power of reflected signals two or more times is negligible. When antenna i transmits a message x, the received signals y_j at j constitutes of signals from i and reflection from c [36], illustrated as follows,

$$y_j = H_r^{ic} \cdot H_r^{cj} \cdot x + H_{LoS} \cdot x, \tag{5}$$

where H_r^{ic} denotes the i-to-c channel and H_r^{cj} denotes the c-to-j channel. For H_{LoS} is a constant value mentioned in Section III-A2, the reflection channel H_r can be simplified as follows,

$$H_r = H_r^{ic} \cdot H_r^{cj}. \tag{6}$$

Recall that the reflection paths follow the path loss model Eq. (2) mentioned in Section III-A2. Substituting Eq. (2) into Eq. (6) yields,

$$H_r = \phi \cdot \frac{1}{d_{ic}^2 \cdot d_{cj}^2}, \tag{7}$$

where ϕ represents the multiplication of all constant values. Assume the individual's walking distance is d_c and the coordinates of i, j and c are (0, d_{ih}), (d_{jw}, 0), (d_c, d_{ch}), respectively. Thus, we improve expression Eq. (7) to

$$H_r(d_c) = \phi \cdot \frac{1}{d_c^2 + (d_{ch} - d_{ih})^2} \cdot \frac{1}{(d_c - d_{jw})^2 + d_{ch}^2}. \tag{8}$$

The power of reflected CSI signal changes over time that could be expressed as follows,

$$\frac{\mathrm{d}|H_r(t)|}{\mathrm{d}t} = \frac{\mathrm{d}|H_r(d_c)|}{\mathrm{d}d_c} \cdot \frac{\mathrm{d}d_c}{\mathrm{d}t} = v \cdot \frac{\mathrm{d}|H_r(d_c)|}{\mathrm{d}d_c}, \tag{9}$$

where v stands for individuals' walking velocity which is a constant value in our scenario, and the derivation of Eq. (8) is omitted. The two WiFi devices are placed at side of a door and the width of door normally is $d_{ih} = 0.8$m. The distance between multi-antenna of laptop or router is $d_{jw} = 0.15$m, approximately. Assume the trajectories are straight in middle of transceivers $d_{ch} = d_{ih}/2 = 0.4$m and fixed in different directions. Fig. 5 simulates the power changes of reflected signals in two walking directions i.e. from left to right or vice versa. We observe the power of CSI reflections rapidly change as the person approaches WiFi transceivers and exhibit the mirrorlike patterns in different walking directions. To verify this mirrored pattern, we conduct an experiment in a setting that the WiFi antenna arrays of transceivers were in the horizontal orientations as shown in Fig. 4(a) and a student volunteer was asked to walk back and forth in between WiFi transceiver. We select all subcarriers of antenna pair between i and j and extract the 2nd principle component and use bandpass filter to remove noise. Fig. 6 shows the cleaned CSI streams and spectrograms in both directions. We find the overall CSI fluctuations in time and frequency domain are quite similar, which demonstrates the signature of this volunteer's gait do persist in varied walking directions. On the other hand, the magnitude variation in time and frequency domain also contain mirrorlike patterns. As shown in Fig. 6, the top figures have large variation in left side in both temporal and spectral domain and the bottom figures have mirrored effects. This symmetrical pattern proves the prior simulation of our proposed reflection model Eq. (9), depicted in Fig. 5.

Additionally, we investigate the vertical antenna array orientation scenario and one same individual walks back and forth as shown in the Fig. 7(a). The displacement between antennas in this situation is no longer existed ($d_{jw} = 0$). Thus, the CSI streams of certain antenna pair are almost identical as displayed in Fig. 7(b). It demonstrates that even though the individual walks into different rooms, the WiFi CSI is not affected by the static environment but in function of radio effects such as reflection, diffraction, etc. as shown in Eq. (4). In real-world environment, the antenna array orientation WiFi devices could be in between horizontal and vertical status and the random device placements may even exaggerate the mirrored patterns. The mirrorlike perturbations of CSI could negatively affect the uniqueness of individuals' gait and human

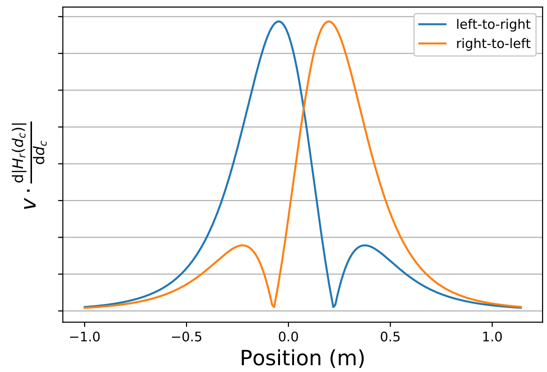
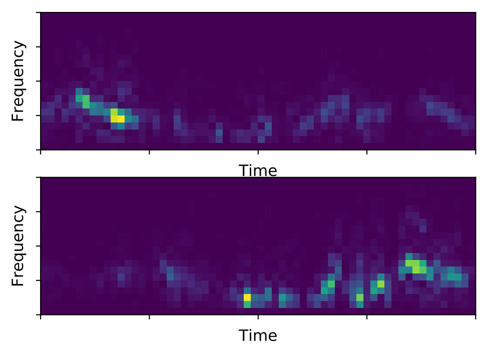
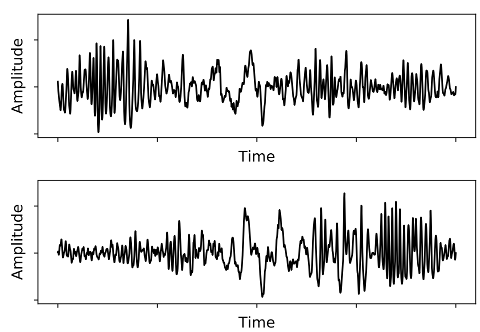


Fig. 5: The reflection power changes in different walking directions, i.e. left to right or vice versa.



(a) CSI spectrograms.

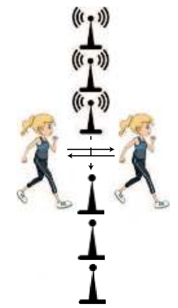


(b) CSI streams.

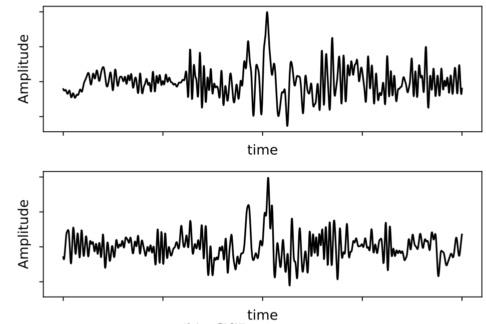
Fig. 6: An individual's CSI streams and spectrograms when WiFi antenna array in horizontal orientation. Each top and bottom sub-figures in Figs. 6(a) to 6(b) refer to walking directions from left to right and vice versa. The CSI streams and spectrograms contain mirrorlike patterns corresponding to the simulation shown in Fig. 5.

identification performance. To capture the persistent signature of gait resided in $H_r + H_d + H_{sa}$, we design a feature augmentation method to remove the gait signature inconsistency in WiFi CSI and mitigate the impact of walking directions and antenna array orientation as discussed in Section III-C.

3) Walking Direction Estimation: The mirrorlike patterns of gait is harmful to human identification but could be used to infer individuals' walking directions. Estimating the walking direction helps to understand and eliminate the influence of moving directions on the CSI data for accurately identifying individuals. Moreover, it is also beneficial to track the moving directions of the monitored persons, especially elders, for the security perspective [37]. Herein, we design a heuristic algorithm (outlined in Algorithm 1) to estimate the walking directions while monitoring gait.



(a) Vertical antenna array orientation.



(b) CSI streams.

Fig. 7: An individual's CSI streams when WiFi antenna array in vertical orientation. The top and bottom sub-figures in Fig. 7(b) refer to the mirrored walking directions and are almost identical.

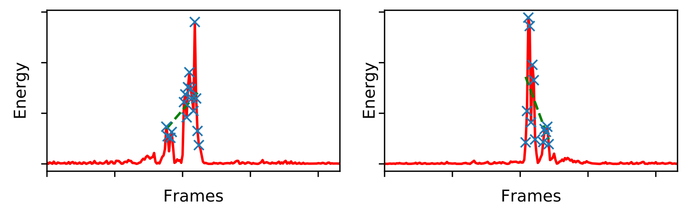


Fig. 8: Examples of different walking direction estimation. The red line are the short time energy of CSI gait. The slope of fitted dash lines indicate walking directions.

As mentioned in previous subsection, the power changes of CSI signals due to human body reflection contain symmetrical patterns in different walking directions. According to Eq. (9) depicted in Fig. 5, the short time energy of CSI signals in real measurement are to change over time in a similar

Algorithm 1 Walking Direction Estimation

Input: One walk observation dataset $\mathcal{H}$
select the 2nd principal component h
$h = |h - mean(h)|$
partitioned into a sequence of frames $S_j(n)$, $j \in [1, Z]$, $n \in [1, N]$, where Z is the total number of frames, each frame has N CSI values
for each frame $j = 1 : Z$ **do**
$$E_j = \frac{1}{N} \sum_{n=1}^{N} |S_j(n)|^2$$
end for
$E_j = MedianFilter(E_j)$
select frames which $E_j > mean(E_j)$
apply least-square optimization based curve fitting
objective function: $f(x) = ax + b$
Output: Estimate walking directions. The direction is forward (backward) if $a > 0$ ($a < 0$)

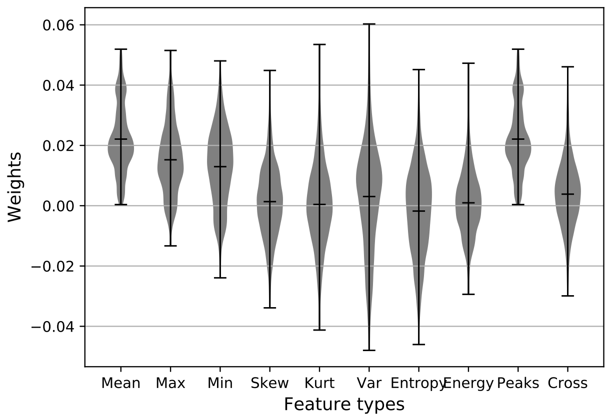


Fig. 9: The feature weights distribution.

symmetrical patterns. As such, we exploit the trends of the short time energy of CSI signals to infer walking directions. As discussed in Section III-A1, each antenna pair consists of 30 subcarriers. We select one certain ($N_t = 1, N_r = 3$) antenna pair from the walk observation $\mathcal{H}$ provided by the silence removal approach. We apply the principal component analysis (PCA) [38] and select the 2nd principal component indicated as h. The absolute value of h subtracted by its average is used for the following processing. Next, we partition the CSI principal component into short frames and calculate the short time energy E_j in each frame. A median filter is used to smooth E_j. A contiguous block of frames whose energy are over their mean are chosen, denoted by the cross in Fig. 8. Next, the least-square optimization based curve fitting method is applied on the contiguous block. A simple polynomial function $f(x) = ax + b$ is set as the objective. The parameter a to be estimated indicates the slope of the short time energy of CSI frames and represents walking directions, forward if $a > 0$ and backward if $a < 0$. Fig. 8 display two examples and the overall mirrorlike patterns have much similarities with the simulation shown in Fig. 5. The dash line in Fig. 8 denote the fitted polynomial and their slopes clearly indicate individuals' moving directions.

C. Feature Extraction and Augmentation

Our goal herein is to identify the features that are most representative of the individual's gait and thus help in achieving high accuracy in uniquely identifying people. In this section we outline how Gate-ID accomplishes feature extraction and augmentation.

Feature Extraction: The CSI streams from all $30 \times N_t \times N_r$ subcarriers are used to capture the comprehensive data. Prior work [5] shows the WiFi CSI in particular frequency band 20-80 Hz contains the most unique patterns that are able to differentiate individuals' gait. Thus, we extract the 20-80 Hz signal and further segment each WiFi CSI stream using a window size of 0.1 seconds. For each window, Gate-ID computes a comprehensive set of statistical features in both time and frequency domain. The time domain features are mean, max, min, skewness, kurtosis, variance and mean crossing rate which are able to capture patterns of CSI waveform. The frequency domain features are normalized entropy, normalized energy and FFT peaks which measures the energy distribution patterns. These similar 10 features have been used in WiFi-based activity recognition works [39].

Feature Augmentation: As mentioned in Section III-B2, the individuals' walking direction could be arbitrary when entering in the corridor and the CSI fluctuation in the back and forth directions have explicit differences. To mitigate the influence of the varied orientations, we generate an additional CSI feature training dataset with opposite orientations in the training phase. Specifically, the feature dataset of one certain category are rearranged into three dimension space - $30 \times T/t \times N_t \cdot N_r$, simplified as $N_c \times N_s \times N_a$. The feature dataset are to be flipped along its 2nd dimension resulting in the reversed orientations. We employ the label-preserving transformation method that the augmented training samples are to preserve their class labels (i.e. individuals' identities) [40][41][42]. The original feature dataset together with the augmented dataset form the synthetic training dataset which helps the following deep learning model to identify individuals in whichever orientations.

D. Attention Based Deep Learning Model

In this section, we design a novel deep neural network model for radio-based human identification. This model integrates an attention mechanism that combine the whole extracted features (see Section III-C) and focus on the certain effective features. Fig. 10 presents the architecture of the DNN model. Traditional method in [5] simply neglects features with low importance weights and concatenates the dataset of selected features which may still involve disturbance. Our proposed attention mechanism firstly calculates importance weights of different types of features in Section III-C, then exploits the feature projecting network to focus on effective features that help to discriminate individuals and become independent with walking directions. As such the overall feature dataset could be utilized while eliminating the impact of ineffective data and external interference. The signature of gait is mostly manifested as the spatial-temporal patterns in CSI. Thus, we adopt multiple Residual Networks (ResNets) [43] and Recurrent Neural Network (RNN) to extract the representative spectral and temporal features of gait, respectively. The CSI feature dataset indicated as $\mathcal{H}_{input}$ is processed by a series of neural models, i.e. feature projecting layers, multiple ResNets, a convolutional (Conv) network layer, a RNN layer, and a linear layer.

As mentioned in Section III-C, ten types of features are extracted from CSI streams, therefore each sample $\mathcal{H}_{input}$ lies in a high dimensional space - $10 \cdot (N_c \times N_s \times N_a)$. Inspired from the self-attention mechanism [44] that imitates the human sight to focus on interesting parts of an entity, we design an attention-based feature fusion mechanism that constitutes of two components, i.e. ReliefF [45] feature weighting and feature projecting neural network layers.

For not all features are equally beneficial for identification, we employ ReliefF [45], an efficient feature weighting algorithm with low-order polynomial time complexity, to calculate

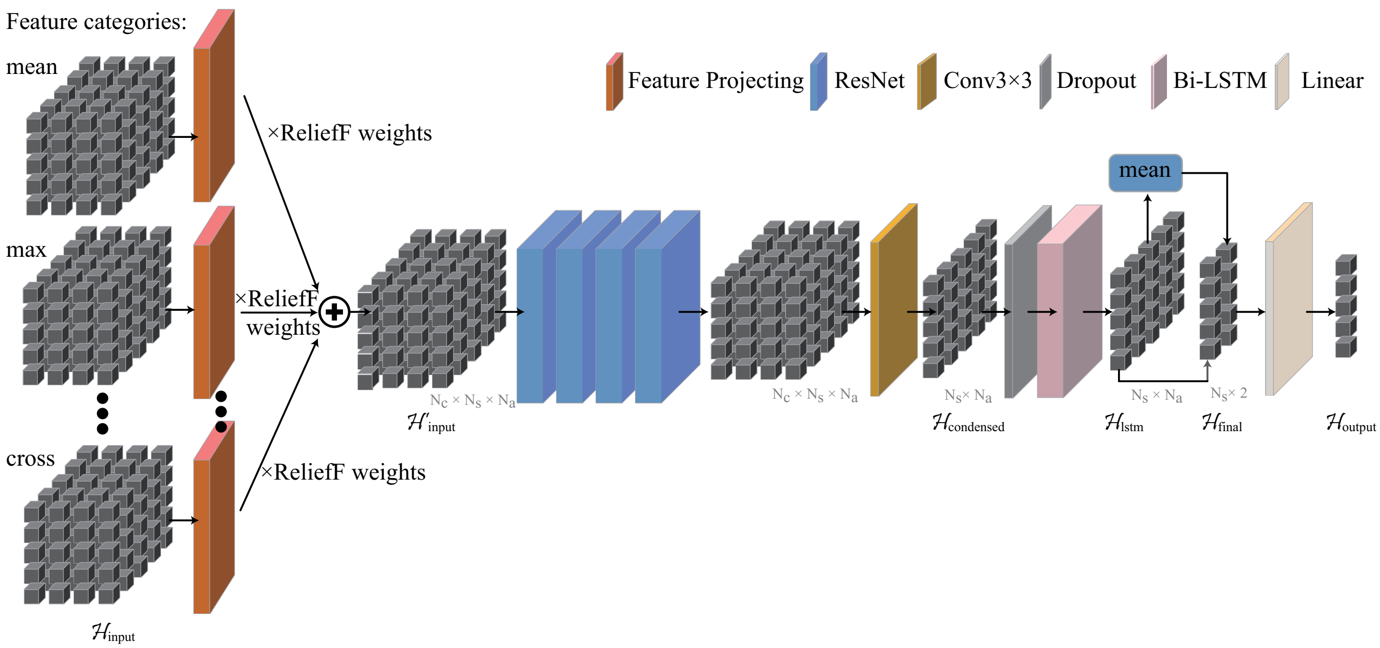


Fig. 10: The DL model is presented from left to right. The CSI feature attributes denoted as $\mathcal{H}_{input}$ (i.e. the black cubs) are fed into the attention-based feature fusion mechanism that has multiple feature projecting layers (i.e. dark red cuboids). $\mathcal{H}^{'}_{input}$ denote the weighted sum of the projected features. The 4-layers ResNet (i.e. blue cuboids) are used to extract spatial features. Next, a convolutional layer (Conv 3×3) indicated as a yellow panel is used to condense features. Next, a dropout layer (i.e. gray panel) is used to avoid overfitting. The Bi-LSTM RNN (i.e. purple cuboid) is to capture temporal patterns of gait. Finally, a linear layer is used to predict the human identities of $\mathcal{H}_{input}$.

weights of feature categories and select the useful ones for accurately recognizing individuals. ReliefF assigns the weights on each feature attributes based on how well they distinguish individuals. Fig. 9 displays the calculated weights' distribution of ten feature categories. The y-axis represents the weights while the x-axis represents the feature categories. We observe that feature categories have different weights, and specifically, the mean and FFT peaks have relative high weights. In Section IV-G, we demonstrate the effectiveness of the feature weighting on the identification accuracy.

The feature projecting neural network layers exploit ReliefF weights as the a priori knowledge and automatically focus on the effective features as training the identification model continuously. Specifically, it involves multiple feature projecting layers in parallel that transform the subspace of ten of feature categories into one unified feature space. The feature projecting layer constitutes one 1×1 Convolution network (Conv) and one Gated Linear Unit (GLU) [46]. The projected feature subspaces are multiplied with the weights of each corresponding feature categories (denoted as ReliefF weights in Fig. 10), which are calculated by averaging all feature attributes' weights as shown in Fig. 9. Consequently, the ten ($l = 10$) of feature dataset are converged as the unified feature space as following,

$$\mathcal{H}^{'}_{input} = [x_1, x_2, \cdots, x_l], \tag{10}$$

where $[x_1, x_2, \cdots, x_l]$ refers to the weighted sum of the feature-maps produced in feature projecting layers $1, \cdots, l$, and $\mathcal{H}^{'}_{input}$ denotes the unified feature space. Section IV-F evaluates the impact of the attention mechanism. The following methods are used to extract the spatial and temporal patterns of gait in $\mathcal{H}^{'}_{input}$.

Next, we employ ResNets [43] incorporated with Convolutional Neural Networks (CNNs) to extract distinguishing local features from $\mathcal{H}^{'}_{input}$. ResNets are chosen for its well stability without the vanishing gradient problem [47] when network and data are complex. We exploit 4 stacked layers of ResNets to capture the spatial patterns in CSI dataset. Section IV-E evaluates the impact of varied number of ResNets layers. Next, a single Conv 3×3 layer is applied to further compress and extract representative features on the N_f dimension, denoted as the yellow panel in Fig. 10. The generated feature space of $\mathcal{H}_{condensed}$ is $N_c \times N_s$. The CSI dataset collected in experiments mentioned in Section IV-A has limited amount due to the cumbersome collecton process. The small-size dataset often cause overfitting problem to the deep-learning techniques. It is proved that dropout could alleviate the overfitting issue [48], therefore we add one dropout layer and the dropout rate is set as 0.2, which evaluated in Section IV-D. Bidirectional Long Short-Term Memory (Bi-LSTM) [49] RNN is good at understanding long-term dependencies embedded in feature space. Therefore, an additional Bi-LSTM RNN layer is employed on the $\mathcal{H}_{condensed}$ to understand the temporal patterns of gait. To summarize the $\mathcal{H}_{lstm}$, the concatenation of its first column $\mathcal{H}^0_{lstm}$ and the mean of $\mathcal{H}_{lstm}$ along its first dimension $\overline{\mathcal{H}_{lstm}}$ become the final representation:

$$\mathcal{H}_{final} = [\overline{\mathcal{H}_{lstm}}; \mathcal{H}^0_{lstm}] \tag{11}$$

In the end, a linear layer projects features to the unnormalized probabilities $\mathcal{H}_{output}$ and a Softmax function calculates the probability vector to predict human identities. The learning rate of the model is fixed as 0.0008. For training the model, the negative log-likelihood loss (equiv. cross entropy function) is adopted to optimize the loss between predictions and ground truths.

IV. Performance Evaluation

In this section we present a comprehensive evaluation of the Gate-ID system. Section IV-A explains the setup of experiments. Section IV-B evaluates the walking direction estimation algorithm. Section IV-C evaluates the effect of the varying effective region T on the performance. In Section IV-D and Section IV-E, we study the effect of the dropout rate and varying number of stacked ResNet layers, respectively. Section IV-F presents the study of effect of the attention mechanism of the deep learning model. Section IV-G evaluates the effect of feature weighting and augmentation on the classification performance. Section IV-H evaluates the performance of Gate-ID in different group sizes. Finally, Section IV-I compares Gate-ID with two benchmarks.

A. Experiment setup

We use off-the-shelf devices to implement a prototype of Gate-ID. The system consists of WiFi-integrated devices: one HP 8530p laptop with built-in Intel WiFi link 5300 802.11n card as the receiver, and one WiFi router Netgear R7000 as the transmitter. Ubuntu 10.10 and modified Intel NIC WiFi driver were installed in the HP laptop. The data collection of Gate-ID system depicted in Fig. 2 was implemented on the laptop. Other than that the identification model and signal processing tasks were implemented on the server equipped with NVIDIA TESLA P40. We set the central frequency and bandwidth of WiFi devices as 5.19 GHz and 40 MHz respectively in our experiments. The sampling rate of CSI is 800 Hz to ensure the stable measurements of CSI data. The two WiFi devices were placed about 1.3 meters away on tables at a height of 1 meter next to a door in an office. The campus WiFi network coexisted in the corridor and was operational during experiments.

To evaluate Gate-ID, we recruit 20 college students both male and female for experiments. Each subject was asked to walk in/out a room through a door and cross WiFi transceiver back and forth, and repeated 5 times resulting in 10 of walking observations. Each observation lasts 4 seconds approximately, thus total data collection for each individuals last about 1 minute. Note this data collection only happen once for training the system. We choose 6, 2, and 2 out of 10 walking observations of each subject that correspond to 120, 40, and 40 data samples as training, validation and test dataset, respectively. The twofold augmented CSI dataset mentioned in Section III-C has 240 training data samples. Two evaluation metrics are adopted in experiments - (i) true detection rate (accuracy) and (ii) confusion matrix.

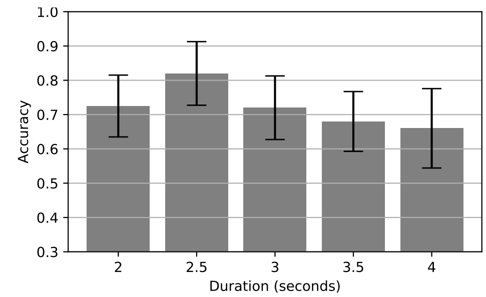


Fig. 11: Impact of duration of effective region

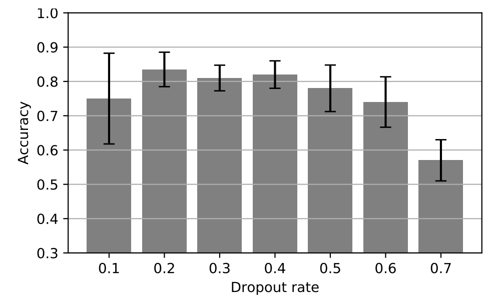


Fig. 12: Impact of dropout rate

B. Walking direction estimation

As discussed in Section III-B3, we take advantage of horizontal antenna setup and design an algorithm for walking direction estimation. We randomly select 5 individuals' walking observations to examine the effectiveness of the walking direction estimation algorithm outlined in Algorithm 1. The experiments show the average accuracy is of 92.17% and the standard deviation is of 4.2%. This proves the statement in Section III-B3 that the walking directions can be estimated when the individual walk along the WiFi antenna array. Moreover, this indicates that the varying walking directions indeed create different fluctuations in WiFi spectrum that causes the gait signature inconsistency. It illustrates necessary efforts such as feature weighting and augmentation together with the attention-based DL model are important to remove the gait signature inconsistency in WiFi CSI and extract unique features of human gait.

C. Effect of the duration of the effective region T

In this subsection we study the impact of the duration of the effective region T mentioned in Section III-B1 on the classification performance. We use the augmented CSI dataset and employ the proposed DL model as classifier. In Fig. 11, it is evident that the best-performing setting is $T = 2.5$ seconds. Our system only needs capture a short period (2.5s) of WiFi CSI to accurately recognize a person's identity. We keep T = 2.5 seconds in the following evaluations.

D. Effect of dropout rate

As discussed in Section III-D, the dropout method is used to alleviate the overfitting issue due to the small-size CSI dataset. Herein, we evaluate the effect of different dropout rate from 0.1 to 0.7 as shown in Fig. 12. Note we also use the augmented

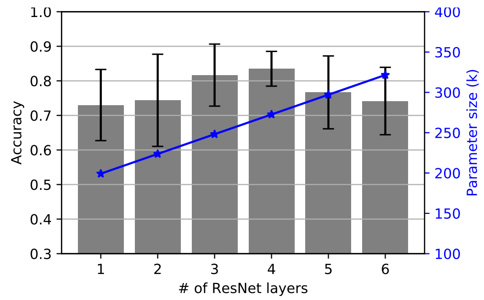
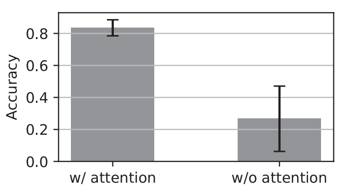
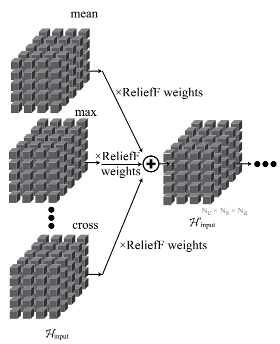


Fig. 13: Impact of number of ResNet layers



(a) Comparing the model w/ and w/o attention mechanism.



(b) The model without feature projecting layers.

Fig. 14: Impact of attention-based feature fusion mechanism.

CSI dataset in this experiment. We find the dropout rate of 0.2 is the best-performing setting. The lower dropout rate has high deviations and the larger rates could cause unnecessary noise and degrade performance. The dropout rate 0.2 is set as default in the following experiments.

E. Effect of number of ResNet layers

Recall that the attention-based deep learning model exploits a number of ResNets to extract spacial features. Herein, we evaluate the impact of number of ResNet layers on the recognition accuracy. We observe that the best-performing setting is to use 4 ResNets layers shown in Fig. 13. The increased stacked layers raise the model's parameter size and the extra representation capacity could cause overfitting issue on the small-size data, thus the performance rise and drop as the increase of ResNets layers. Therefore, we set the 4 ResNets layers as the default settings.

F. Effect of attention mechanism

As shown in Fig. 10, our proposed model exploits multiple feature projecting layers to form the attention mechanism for fusing and extracting effective features. Herein, we evaluate the effect of the attention mechanism on the performance. To achieve this we remove the feature projecting layers and concatenate features directly and the rest of model is unchanged (see Fig. 14(b)). The trimmed model is to treat each feature categories equally. Fig. 14(a) displays the comparison of the deep learning model with and without the attention mechanism. Note this experiment chose 10 subjects randomly.

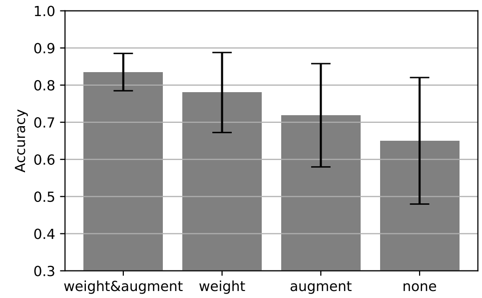


Fig. 15: Impact of feature weighting and augmentation. 'weight&augment' denotes the system implemented with both feature weighting and augmentation methods. 'none' denotes the system without both two methods.

It is obvious that the attention mechanism significantly improve the performance from 26% to 83.5% and increase the accuracy by 57.5%. It proves that the attention based feature fusion mechanism is able to focus on the effective features for satisfactory performance.

G. Effect of feature weighting and augmentation

As discussed in Section III-D, the attention-based feature fusion mechanism calculates the importance weights of varied categories of features and multiply the averaged weights with projected feature-maps for feature fusion. Moreover, an additional feature dataset with reversed walking direction are generated as the augmented training dataset. The synthetic dataset help the identification model to alleviate the influence of varied directions of individuals. Herein, we study the impact of both these two approaches on the human recognition accuracy. We consider four different cases that the system employs both feature weighting and augmentation methods, choose either weighting or augmentation method, or without any of them (see Fig. 15). We will assign the importance weights as 1 if not using feature weighting and only utilize the original dataset if not using data augmentation. Note we chose 10 subjects randomly and repeated the experiments for 20 times. In Fig. 15, we observe the data augmentation increases accuracy from 65.3% to 71.5% and effectively alleviate the influence of varying walking directions. In Fig. 15, it is obvious that the system with both two approaches achieves the best performance, average accuracy of 83.5% with the lowest deviation. Hence, it proves the benefits of feature weighting together with attention-based feature fusion mechanism, and the feature augmentation approach.

H. Effect of group sizes

In this subsection we evaluate the performance of Gate-ID in varying group sizes N. We vary N from 6 to 20 that is a typical small office or home setting. Note that, we do not consider the intruder who not belong the group. For each value of N we randomly select 20 combinations of groups of N people in the dataset. In Fig. 16, it is evident that the accuracy decreases with an increase in group size. It is consistent with other biometric authentication approaches such as video based face identification. Nevertheless, Gate-ID

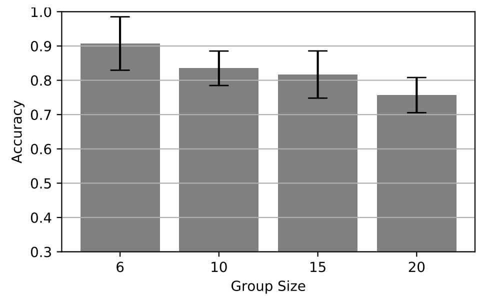


Fig. 16: Impact of group size

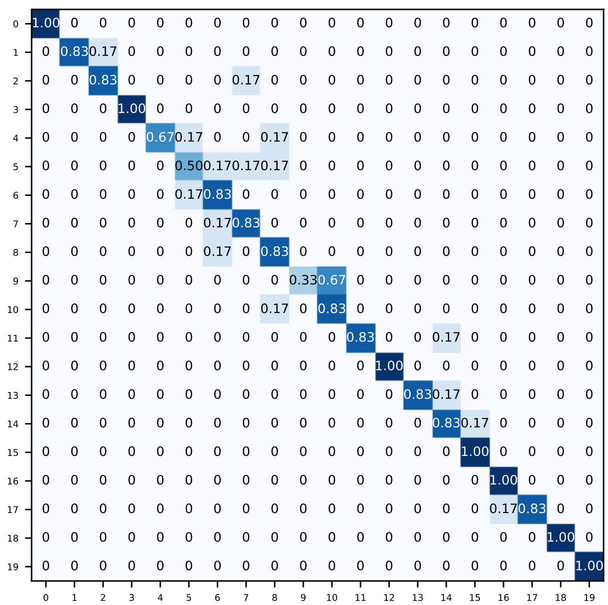


Fig. 17: Confusion matrices of group size 20. They x and y axis indicate the index of the participants.

achieves the identification accuracy of 90.7% to 75.7% for the group size from 6 to 20. WiFi-ID reported the average accuracy of 77% with the group size of 6, whereas Gate-ID significantly increases the accuracy up to 90.7% on average. Fig. 17 shows the confusion matrices for group size of 20, which could be seen as a subset of experiments shown in Fig. 16. The confusion matrices with varying group sizes and members have similar results, that is most individuals could be uniquely identified with high probability. Nevertheless, we also observe the individual 4, 5 and 9 in Fig. 17 have relative low recognition accuracy. The reason behind could be the recruited volunteers in our experiments are college students with similar age, the subtle difference between particular persons could not be captured by our current system. In future work mentioned in Section V, we plan to use multi-modal wireless sensors such as mmwave radar to further improve performance.

I. Comparing with WiFi-ID and SmartUserAuth

We finally compare Gate-ID against the two comparable works SmartUserAuth [26] and WiFi-ID [5]. SmartUserAuth proposes to track the gait when individuals walk between several locations in a room. SmartUserAuth does not rely on segmenting the gait cycles and asks individuals to walk back and forth in two directions which share the same scenarios with ours. Fig. 18 shows the identification accuracy of the three approaches in the group size of 10. SmartUserAuth can only achieve close to an average accuracy of 70% with large deviations. SmartUserAuth employs a three-layers stacked Autoencoder neural network model to extract representative features for identification. This simple network architecture is unable to understand temporal dependencies inside of time series. It does not consider the impact of variable walking directions and WiFi antenna array orientations and have no abilities in achieving feature fusion to eliminate influence of the gait signature inconsistency in WiFi CSI and ineffective features. As for WiFi-ID, it achieved around 40% identification accuracy within the group size of 10 which close to the accuracy of random guess. The WiFi-ID suffered distractions from the changes of walking directions and antenna array orientations. Besides that the employed classifier Sparse Approximation based Classification (SAC) is ineffective to extract unique features from the high-dimensional complex data. In summary, as illustrated in Fig. 18, Gate-ID significantly increases the identification accuracy by 12.5% and 43.5% compared with the two baselines SmartUserAuth and WiFi-ID respectively, which is attributed to the fact that (1) the feature extraction combined with the attention-based feature fusion mechanism of the DL model equips our system with the capability of capturing unique signatures of each person's gait; (2) the feature augmentation method help eliminate the human gait signature inconsistency in WiFi CSI in WiFi CSI and mitigate the impact of varying walking directions and WiFi antenna array orientations.

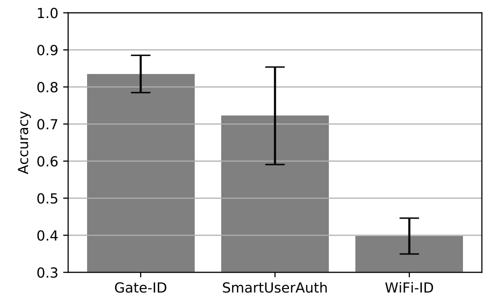


Fig. 18: Comparison of our system with WiFi-ID [5] and SmartUserAuth [26], which use SAC and Autoencoder neural network as classifiers, respectively.

V. Conclusion and Future Work

In this paper we present Gate-ID a WiFi-based device-free human identification system. Gate-ID exploits two WiFi devices to form a virtual gate and achieves accurate human identification irrespective of walking directions when individuals crossing the gate. Each individual has a unique walking style which in turn create unique perturbations in the WiFi spectrum. Gate-ID extracts unique features of human gait from the WiFi perturbations through a group of methods to identify individuals. First, we employ theoretical model and real measurements to demonstrate that human gait in different walking directions have mirrorlike fluctuations when the WiFi antenna array in the horizontal placement. While such gait signature inconsistency in WiFi CSI is harmful to human identification, a heuristic algorithm is designed to exploit this inconsistent mirrored pattern to infer walking directions. Next, we extract a set of spatial-temporal features and synthesize an extra dataset by flipping and reversing data to remove influence of the gait signature inconsistency in WiFi CSI. The novel attention-based DL model is proposed to fuse the synthetic weighted dataset and exploit the ResNet and Bi-LSTM neural networks to accurately identify individuals. A silence removal method is employed to segment the most pronounced CSI time series due to the motion of the person. Gate-ID achieves 90.7% to 75.7% human identification accuracy for 6 to 20 individuals in a group, respectively. Compared with the two baselines SmartUserAuth and WiFi-ID, Gate-ID significantly increases the identification accuracy by 12.5% and 43.5%, respectively. We envision that Gate-ID could be a generalized solution for wide applications in smart home and smart building to provide personalization services and interaction with smart devices.

In the future we plan to consider more generalized scenarios in real-world environment that participants could walk in multiple random directions around WiFi devices, and even may not walk in a straight line, instead, circle the WiFi transceivers. Under these circumstances, the WiFi CSI of a person's gait would be certainly twisted to some extent. We plan to retain the uniqueness of human gait in these natural settings to enable the wide personalized applications in smart home. In addition, we might incorporate the other wireless signals such as millimeter wave (mmwave) radar in 60GHz and visual information for monitoring gait. Compared with the WiFi CSI data, the mmwave signals can accurately obtain a person's speed, location, and Doppler shifts. Prior works [50] exploit WiFi and vision multimodal learning to further improve activity recognition accuracy. We plan to fuse the multi-modal wireless signals and visual images to capture human skeleton as a priori knowledge and reconstruct human gait from radio signals in aforementioned natural settings. To achieve this, the techniques like transfer learning and teacher student deep learning architecture might be utilized in fusing the multi-modal data and accurately monitoring human gait.

REFERENCES

[1] B. Wei *et al.*, "Radio-based device-free activity recognition with radio frequency interference," in *International Conference on Information Processing in Sensor Networks*. ACM, 2015, pp. 154–165.

[2] L. Sun *et al.*, "Widraw: Enabling hands-free drawing in the air on commodity wifi devices," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 77–89.

[3] J. Zhang, W. Xu, W. Hu, and S. S. Kanhere, "Wicare: Towards in-situ breath monitoring," in *Proceedings of the 14th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. ACM, 2017, pp. 126–135.

[4] M. S. Nixon *et al.*, *Human identification based on gait*. Springer Science & Business Media, 2010, vol. 4.

[5] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, "Wifi-id: Human identification using wifi signal," in *International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2016, pp. 75–82.

[6] J. Zhang *et al.*, "Human identification using wifi signal," in *International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. IEEE, 2016, pp. 1–2.

[7] Y. Zeng, P. H. Pathak, and P. Mohapatra, "Wiwho: wifi-based person identification in smart spaces," in *Proceedings of the 15th International Conference on Information Processing in Sensor Networks*. IEEE, 2016, p. 4.

[8] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proceedings of International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 363–373.

[9] B. Korany, C. R. Karanam, H. Cai, and Y. Mostofi, "Xmodal-id: Using wifi for through-wall person identification from candidate video footage," in *ACM International Conference on Mobile Computing and Networking (MobiCom)*, 2019.

[10] C. Wu, F. Zhang, Y. Hu, and K. R. Liu, "Gaitway: Monitoring and recognizing gait speed through the walls," *IEEE Transactions on Mobile Computing*, 2020.

[11] L. Wang *et al.*, "Silhouette analysis-based gait recognition for human identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1505–1518, 2003.

[12] C. Wang *et al.*, "Human identification using temporal information preserving gait template," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2164–2176, 2012.

[13] M. Sousa, A. Techmer, A. Steinhage, C. Lauterbach, and P. Lukowicz, "Human tracking and identification using a sensitive floor and wearable accelerometers," in *2013 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2013, pp. 166–171.

[14] F. Sun, W. Zang, R. Gravina, G. Fortino, and Y. Li, "Gait-based identification for elderly users in wearable healthcare systems," *Information Fusion*, vol. 53, pp. 134–144, 2020.

[15] S. Pan, N. Wang, Y. Qian, I. Velibeyoglu, H. Y. Noh, and P. Zhang, "Indoor person identification through footstep induced structural vibration," in *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, 2015, pp. 81–86.

[16] S. Pan, T. Yu, M. Mirshekari, J. Fagert, A. Bonde, O. J. Mengshoel, H. Y. Noh, and P. Zhang, "Footprintid: Indoor pedestrian identification through ambient structural vibration sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–31, 2017.

[17] Y. Wang, Y. Chen, M. Z. A. Bhuiyan, Y. Han, S. Zhao, and J. Li, "Gait-based human identification using acoustic sensor and deep neural network," *Future Generation Computer Systems*, vol. 86, pp. 1228–1237, 2018.

[18] W. Xu, Z. Yu, Z. Wang, B. Guo, and Q. Han, "Acousticid: gait-based human identification using acoustic signal," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–25, 2019.

[19] C. Li, Z. Peng, T.-Y. Huang, T. Fan, F.-K. Wang, T.-S. Horng, J.-M. Munoz-Ferreras, R. Gomez-Garcia, L. Ran, and J. Lin, "A review on recent progress of portable short-range noncontact microwave radar systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 5, pp. 1692–1706, 2017.

[20] I. Orović *et al.*, "A new approach for classification of human gait based on time-frequency feature representations," *Signal Processing*, vol. 91, no. 6, pp. 1448–1456, 2011.

[21] T. Fan, C. Ma, Z. Gu, Q. Lv, J. Chen, D. Ye, J. Huangfu, Y. Sun, C. Li, and L. Ran, "Wireless hand gesture recognition based on continuous-wave doppler radar sensors," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 11, pp. 4012–4020, 2016.

[22] Z. Peng, J. M. Muñoz-Ferreras, Y. Tang, C. Liu, R. Gómez-García, L. Ran, and C. Li, "A portable fmcw interferometry radar with programmable low-if architecture for localization, isar imaging, and vital sign tracking," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 4, pp. 1334–1344, 2016.

[23] J. Huang, B. Liu, C. Chen, H. Jin, Z. Liu, C. Zhang, and N. Yu, "Towards anti-interference human activity recognition based on wifi subcarrier correlation selection," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2020.

[24] J. Yang, H. Zou, Y. Zhou, and L. Xie, "Learning gestures from wifi: A siamese recurrent convolutional architecture," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 763–10 772, 2019.

[25] J. Zhang, Z. Tang, M. Li, D. Fang, P. Nurmi, and Z. Wang, "Crosssense: towards cross-site and large-scale wifi sensing," in *Proceedings of the 24th International Conference on Mobile Computing and Networking*. ACM, 2018, pp. 305–320.

[26] C. Shi, J. Liu, H. Liu, and Y. Chen, "Smart user authentication through actuation of daily activities leveraging wifi-enabled iot," in *Proceedings of the 18th International Symposium on Mobile Ad Hoc Networking and Computing*. ACM, 2017, p. 5.

[27] J. Wang, Y. Zhao, X. Fan, Q. Gao, X. Ma, and H. Wang, "Device-free identification using intrinsic csi features," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8571–8581, 2018.

[28] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. J. Spanos, "Wifi-based human identification via convex tensor shapelet learning," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[29] Z. Yang *et al.*, "From rssi to csi: Indoor localization via channel response," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, p. 25, 2013.

[30] D. Halperin *et al.*, "Tool release: Gathering 802.11 n traces with channel state information," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 53–53, 2011.

[31] M. Ghaddar *et al.*, "A conducting cylinder for modeling human body presence in indoor propagation channel," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 11, pp. 3099–3103, 2007.

[32] Q. C. Zhi Ren, Guangyu Wang and H. Li, "Modelling and simulation of rayleigh fading, path loss, and shadowing fading for wireless mobile networks," *Simulation Modelling Practice and Theory*, vol. 19, no. 2, pp. 626 – 637, 2011.

[33] B. Sklar, "Rayleigh fading channels in mobile digital communication systems .i. characterization," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 90–100, 1997.

[34] J. Wang, J. Xiong, H. Jiang, K. Jamieson, X. Chen, D. Fang, and C. Wang, "Low human-effort, device-free localization with fine-grained subcarrier information," *IEEE Transactions on Mobile Computing*, vol. 17, no. 11, pp. 2550–2563, 2018.

[35] R. P. Findlay and P. J. Dimbylow, "Sar in children from exposure to wireless local area networks (wlan)," in *2012 Asia-Pacific Symposium on Electromagnetic Compatibility*, 2012, pp. 733–736.

[36] H. Shen, W. Xu, S. Gong, Z. He, and C. Zhao, "Secrecy rate maximization for intelligent reflecting surface assisted multi-antenna communications," *IEEE Communications Letters*, vol. 23, no. 9, pp. 1488–1492, 2019.

[37] K. K. Peetoom, M. A. Lexis, M. Joore, C. D. Dirksen, and L. P. De Witte, "Literature review on monitoring technologies and their outcomes in independently living elderly people," *Disability and Rehabilitation: Assistive Technology*, vol. 10, no. 4, pp. 271–294, 2015.

[38] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular value decomposition and principal component analysis," in *A practical approach to microarray data analysis*. Springer, 2003, pp. 91–109.

[39] Y. Zeng *et al.*, "Analyzing shopper's behavior through wifi signals," in *Proceedings of the 2nd workshop on Workshop on Physical Analytics*. ACM, 2015, pp. 13–18.

[40] J. Schlüter and T. Grill, "Exploring data augmentation for improved singing voice detection with neural networks." in *International Society for Music Information Retrieval (ISMIR)*, 2015, pp. 121–126.

[41] j. zhang, f. wu, w. hu, q. zhang, w. xu, and j. cheng, "Wienhance: Towards data augmentation in human activity recognition using wifi signal," in *2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*, 2019, pp. 309–314.

[42] J. Zhang, F. Wu, B. Wei, Q. Zhang, H. Huang, S. W. Shah, and J. Cheng, "Data augmentation and dense-lstm for human activity recognition using wifi signal," *IEEE Internet of Things Journal*, pp. 1–1, 2020.

[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of International Conference on computer vision and pattern recognition*. IEEE, 2016, pp. 770–778.

[44] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[45] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of relieff and rrelieff," *Machine learning*, vol. 53, no. 1-2, pp. 23–69, 2003.

[46] A. Van den Oord, N. Kalchbrenner, L. Espeholt, O. Vinyals, A. Graves *et al.*, "Conditional image generation with pixelcnn decoders," in *Advances in neural information processing systems (NIPS)*, 2016, pp. 4790–4798.

[47] R. Pascanu, T. Mikolov, and Y. Bengio, "Understanding the exploding gradient problem," *CoRR, abs/1211.5063*, vol. 2, p. 417, 2012.

[48] M. Cogswell, F. Ahmed, R. Girshick, L. Zitnick, and D. Batra, "Reducing overfitting in deep networks by decorrelating representations," *arXiv preprint arXiv:1511.06068*, 2015.

[49] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[50] H. Zou, J. Yang, H. P. Das, H. Liu, Y. Zhou, and C. J. Spanos, "Wifi and vision multimodal learning for accurate and robust device-free human activity recognition," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 426–433.

Jin Zhang received his PhD degree in Computer Science and Engineering from University of New South Wales, Australia in 2017. He is a postdoctoral researcher in Shenzhen institutes of advanced technology, Chinese Academy of Sciences. His research interests include WiFi human sensing, Internet of Things and computer network.

Bo Wei has been a senior lecturer in the Department of Computer and Information Sciences at Northumbria University. He was a Postdoctoral research assistant in University of Oxford. He obtained his PhD degree in Computer Science and Engineering in 2015 from the University of New South Wales, Australia. His research interests are Mobile Computing, Internet of Things, and Wireless Sensor Networks.

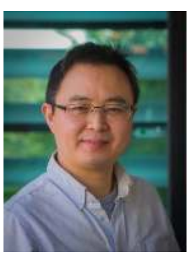
Fuxiang Wu received the Ph.D. degree in pattern recognition and intelligent system from Beihang University, Beijing, China, 2017. He works in Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China. His research interests include text-to-image synthesis, multimodal deep learning, and natural language processing. His current research focuses on text analysis and image editing in multimodal deep learning.

Limeng Dong received the Bachelor's degree in 2012, the Master's degree in 2015, and the Ph.D. degree in 2019 from the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, Shaanxi, 710072, China. He is now a Postdoctoral Researcher with the Ministry of Education Key Lab for Intelligent Networks and Network Security, School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China. During 2015 to 2017, he was once a visiting Ph.D. student at the School of Electrical Engineering and Computer Science, University of Ottawa, Canada. His research interests include multi-antenna communications, cognitive radio and physical layer security.

Wen Hu Wen Hu is currently an Associate Professor with the School of Computer Science and Engineering, University of New South Wales (UNSW). He has published regularly in the top-rated sensor network and mobile computing venues, such as IPSN, SenSys, MobiCom, UbiComp, TOSN, the TMC, TIFS and the PROCEEDINGS OF THE IEEE. His research interests focus on the novel applications, low-power communications, security and compressive sensing in sensor network systems and the Internet of Things (IoT). He is a Senior Member of ACM and IEEE. He is an Associate Editor of ACM TOSN. He is the General Chair of CPS-IoT Week 2020, and serves on the organizing and program committees of networking conferences, including IPSN, SenSys, MobiSys, MobiCom, and IoTDI.

Salil S. Kanhere received the M.S. and Ph.D. degrees from Drexel University, Philadelphia, USA. He is currently a Professor of Computer Science and Engineering with UNSW Sydney, Australia. His research interests include the Internet of Things, cyber physical systems, blockchain, pervasive computing, cybersecurity, and applied machine learning. Salil is also affiliated with CISRO's Data61 and the Cybersecurity Cooperative Research Centre. He is a recipient of the Friedrich Wilhelm Bessel Research Award (2020) and the Humboldt Research Fellowship (2014), both from the Alexander von Humboldt Foundation. He is a Senior Member of the IEEE and ACM and an ACM Distinguished Speaker. He serves as the Editor in Chief of the Ad Hoc Networks journal and as an Associate Editor of the IEEE Transactions On Network and Service Management, Computer Communications, and Pervasive and Mobile Computing. He has served on the organising committee of several IEEE/ACM international conferences and is the General Chair for the IEEE International Conference on Blockchain and Cryptocurrency (ICBC) 2021. He has co-authored a book titled Blockchain for Cyberphysical Systems released by Artech House in 2020.

Chengwen Luo received the PhD degree from the School of Computing, National University of Singapore (NUS), Singapore. He is currently an assistant professor in the College of Computer Science and Software Engineering, Shenzhen University (SZU), China. Before joining SZU, he was a postdoctoral researcher in CSE, The University of New South Wales (UNSW), Australia. He publishes regularly at top venues of mobile computing and networking such as IEEE TMC, IEEE/ACM ToN, ACM TOSN, ACM SenSys, ACM/IEEE IPSN, etc. His research interests include mobile and pervasive computing, indoor localization, wireless sensor networks, and security aspects of Internet of Things.



Shui Yu is a Professor of School of Computer Science, University of Technology Sydney, Australia. Dr Yu's research interest includes Big Data, Security and Privacy, Networking, and Mathematical Modelling. He has published three monographs and edited two books, more than 350 technical papers, including top journals and top conferences, such as IEEE TPDS, TC, TIFS, TMC, TKDE, TETC, ToN, and INFOCOM. His h-index is 49. Dr Yu initiated the research field of networking for big data in 2013, and his research outputs have been adopted by industrial systems. He is currently serving a number of prestigious editorial boards, including IEEE Communications Surveys and Tutorials (Area Editor), IEEE Communications Magazine, and IEEE Internet of Things Journal. He is a Senior Member of IEEE, a member of AAAS and ACM, and a Distinguished Lecturer of IEEE Communication Society.

Jun Cheng received the B.Eng. and M.Eng. degrees from the University of Science and Technology of China, Hefei, China, in 1999 and 2002, respectively, and the Ph.D. degree from the Chinese University of Hong Kong, Hong Kong, in 2006. He is currently a Professor with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, and the Director of the Laboratory for Human Machine Control. His current research interests include computer vision, robotics, machine intelligence, and control.