Toward Robust and Effective Behavior Based User Authentication With Off-the-Shelf Wi-Fi

Lei Zhang, *Member, IEEE*, Yunzhe Jiang, Yazhou Ma, Shiwen Mao, *Fellow, IEEE*, Wenyuan Huang, Zhiyong Yu, *Member, IEEE*, Xiao Zheng, *Member, IEEE*, Lin Shu, *Senior Member, IEEE*, Xiaochen Fan, *Member, IEEE*, Guangquan Xu, *Member, IEEE*, and Changyu Dong, *Member, IEEE*

***Abstract*— Behavior-based Wi-Fi user authentication has gained popularity in user-centered smart systems. However, its wide adoption has been hindered by certain critical issues, including significant performance degradation when the environment changes, the inability to handle unknown activities, and weak security due to basing authentication on the recognition of a single, one-off activity. In this paper, we propose Wi-Dist, which authenticates a user using a behavior password, i.e. a pre-chosen sequence of activities. Wi-Dist addressed the previously mentioned technical challenges through a cross-layer joint optimization framework. In particular, we address environment dependency by incorporating adversarial learning and optimizing both the signal layer and the domain adaptation layer. This enhances the performance of the learned model across various environments. To effectively handle unknown behaviors, we utilize an adversarial learning-based network. This network establishes a pseudo-decision boundary between samples from known and unknown sources, ensuring robust authentication. Additionally, for authentication using continuous activities, we employ double-sliding windows activity monitoring. This approach, coupled with activity state correction, partitions activities for accurate recognition. We also conducted extensive experiments in indoor environments to demonstrate that Wi-Dist is effective and robust.**

***Index Terms*— Wi-Fi, channel state information, action recognition, cross-environment.**

Manuscript received 9 January 2024; revised 12 May 2024 and 19 June 2024; accepted 7 July 2024. Date of publication 15 July 2024; date of current version 27 September 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB3102100; in part by the Natural Science Foundation of Tianjin under Grant 22JCYBJC00120; in part by the Open Fund of Anhui Engineering Research Center for Intelligent Applications and Security of Industrial Internet under Grant IASII22-01; in part by Guangdong Provincial Key Laboratory of Human Digital Twin under Grant 2022B1212010004; in part by the Key Laboratory of Spatial Data Mining and Information Sharing of Ministry of Education, Fuzhou University under Grant 2023LSDMIS06; in part by the National Science Foundation of China under Grant U22B2027, Grant 62172297, Grant 62102262, Grant 61902276, Grant 62272311, and Grant 62332014; and in part by Tianjin Intelligent Manufacturing Special Fund Project under Grant 20211097. The associate editor coordinating the review of this article and approving it for publication was Dr. Yan Wang. *(Lei Zhang and Yunzhe Jiang are co-first authors.) (Corresponding authors: Guangquan Xu; Changyu Dong.)*

Lei Zhang is with the College of Intelligence and Computing and Tianjin Key Laboratory of Advanced Network Technology and Application, Tianjin University, Tianjin 300050, China, and also with the State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China (e-mail: lzhang@tju.edu.cn).

Yunzhe Jiang, Yazhou Ma, Wenyuan Huang, and Guangquan Xu are with the College of Intelligence and Computing and Tianjin Key Laboratory of Advanced Network Technology and Application, Tianjin University, Tianjin 300050, China (e-mail: jiangyz@tju.edu.cn; yazhouma@tju.edu.cn; wy_huang1998@tju.edu.cn; Losin@tju.edu.cn).

Shiwen Mao is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849 USA (e-mail: smao@ieee.org).

Zhiyong Yu is with the College of Computer and Data Science, Fuzhou University, Fuzhou, Fujian 350108, China (e-mail: yuzhiyong@fzu.edu.cn).

Xiao Zheng is with the School of Computer Science and Technology, Anhui University of Technology, Maanshan, Anhui 243032, China, and also with Anhui Engineering Research Center for Intelligent Applications and Security of Industrial Internet, Maanshan 243032, China (e-mail: xzheng@ahut.edu.cn).

Lin Shu is with the School of Future Technology, South China University of Technology, Guangzhou 510641, China (e-mail: shul@scut.edu.cn).

Xiaochen Fan is with the Institute for Electronics and Information Technology in Tianjin, Tsinghua University, Tianjin 300467, China, and also with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: fanxiaochen33@gmail.com).

Changyu Dong is with the Institute of Artificial Intelligence, Guangzhou University, Guangzhou 510700, China (e-mail: changyu.dong@gzhu.edu.cn).

Digital Object Identifier 10.1109/TIFS.2024.3428367

I. Introduction

USER authentication aims to verify whether a user has the authority to access private resources, making it fundamental to ensure the security of a computer system. However, traditional authentication mechanisms often struggle to seamlessly integrate into smart homes, offices, and other "smart" environments due to their rigidity and inconvenience. There have been notable developments in user authentication [1], [2], [3], aiming to enhance security and improve user experiences. Among these developments, behavior-based user authentication [4], [5], [6], [7], which verifies a user's identity based on their daily activities, has garnered significant attention. These methods are appealing because they strike a balance between security and user experience. Typically, wearable devices are employed to achieve user authentication through human activities [1], [8], [9], [10]. However, carrying costly wearable devices may not be practical in real-world authentication scenarios. To achieve device-free behavior-based user authentication, computer vision is utilized [11], [12], allowing authentication through the analysis of videos captured by external cameras. Nonetheless, this approach also presents inherent issues, including lighting constraints and privacy violations [13], [14].

The innovation of Wi-Fi technology has garnered significant interest due to its noninvasive nature, ubiquity, and cost-effectiveness. In recent years, behavior-based Wi-Fi user authentication has emerged as an attractive research topic, propelled by the technology's predominant advantages, resulting in a surge of work in this area [5], [15], [16], [17], [18].

This approach utilizes ubiquitous wireless signals in typical indoor environments to extract inherent behavioral characteristics, enabling device-free authentication of users. However, existing Wi-Fi-based user behavior authentication encounters the following issues.

- Firstly, mitigating dependency on the environment proves challenging. As illustrated in [19] and [20], at the core of Wi-Fi-based behavior authentication systems lies an activity recognition model. This model extracts features from Wi-Fi signals and infers user activities. However, due to the low spatial resolution of Wi-Fi signals, an activity recognition model trained in one environment often experiences a significant decline in performance even with minor environmental changes. Moreover, the data-driven nature of the model renders it sensitive to activities not present in the training data, potentially misclassifying them and leading to decreased accuracy. Typically, constructing an effective data-driven model necessitates millions of samples, making it labor-intensive to gather extensive sensing data. Despite these challenges, existing Wi-Fi-based behavior authentication systems [3], [21], [22], [23], [24], [25] fail to address this environmental dependency. Consequently, collecting new data and training a new model for each environment becomes impractical in real-world scenarios.
- Secondly, the emergence of new activities presents a challenge to user authentication. Previous research in activity recognition [4], [19], [26] has employed domain adaptation methodologies to mitigate environmental dependency. However, these approaches often assume activities to be invariant and overlook the possibility of new activities. In reality, users can engage in various activities beyond those predefined in the system. Systems that only recognize a limited number of activities may encounter difficulty when confronted with new activities, as they must classify them into pre-existing categories, potentially leading to incorrect authentication outcomes. This poses a second challenge that existing Wi-Fi-based behavior authentication systems struggle to address.
- Lastly, effectively segmenting continuous activities into individual ones and implementing robust authentication mechanisms remains a significant challenge. Current Wi-Fi-based systems [4], [6], [25], [27] authenticate users through single, one-off activities, rendering them vulnerable to zero-effort attacks and imitation attacks. Continuous activities exhibit more substantial spatial and temporal dynamic relations [21], [22], [28], which can compensate for the inherent limitations of Wi-Fi and enable more accurate user authentication. However, partitioning continuous activities into a sequence of atomic activities presents challenges, as various activities or activities performed by different individuals have different durations.

To achieve Wi-Fi-based user behavior authentication and address the aforementioned issues, we propose Wi-Dist, a cross-layer optimization-based user authentication framework utilizing off-the-shelf Wi-Fi (as depicted in Fig. 1).

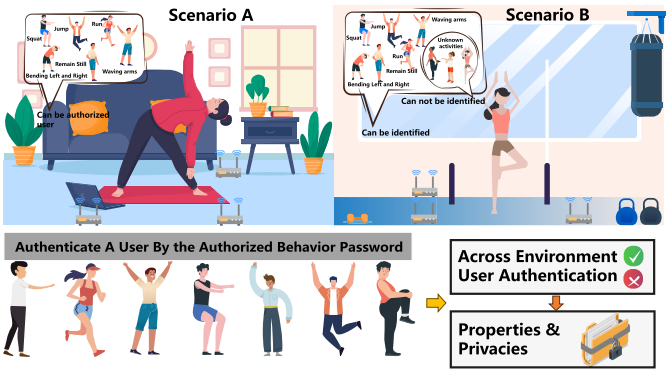


Fig. 1. Wi-Dist system.

The ultimate goal of Wi-Dist is to realize access control to specific areas through the identification and judgment of users' "behavioral passwords". When a user group registers, it needs to select several activities as the shared password library of the user group. During verification, identity verification is performed by executing a legal activity sequence, and then access rights to the area are obtained. The user's behavior password is composed of action sequences in the behavior password library unique to the user group and supports repeated variable-length actions. This design means that users do not need to remember complex character combinations but can perform familiar actions to prove their identity to the system and gain access to specific areas. Since attackers are not familiar with the actions in the cryptographic library, we can classify actions through open domain adaptation, identify unknown actions, and then deny dangerous access.

We also address the aforementioned challenges as follows: (1) The issue of the recognition model's environment dependency is tackled from both the signal layer and the domain adaptation layer. Specifically, at the signal layer, joint antenna selection and conjugate multiplication are performed to enhance signal quality. An adversarial learning algorithm is applied to calibrated signals to generate higher-quality training data. (2) At the domain adaptation layer, an open set domain adaptation model is employed to reduce environment dependency and differentiate between unknown activities and known ones in altered environments. The classification model consists of two opponents: the feature generator and the domain classifier. In the optimization process, the feature generator attempts to deceive the domain classifier by maximizing classification error, while the classifier aims for accurate classification. The domain classifier undergoes weak training to establish a pseudo-decision boundary between samples from known sources and those from unknown targets. (3) To address the issue of continuous activity segmentation, dynamic activity states are effectively captured and partitioned at the signal layer using double sliding windows activity monitoring and activity state correction. Real user authentication is then implemented at the authentication layer using binary logistic regression classification to determine the legitimacy of a user. As illustrated in Fig. 2, information flows from the signal layer to the authentication layer. Through joint optimization, effective Wi-Fi behavior-based user authentication is achieved.

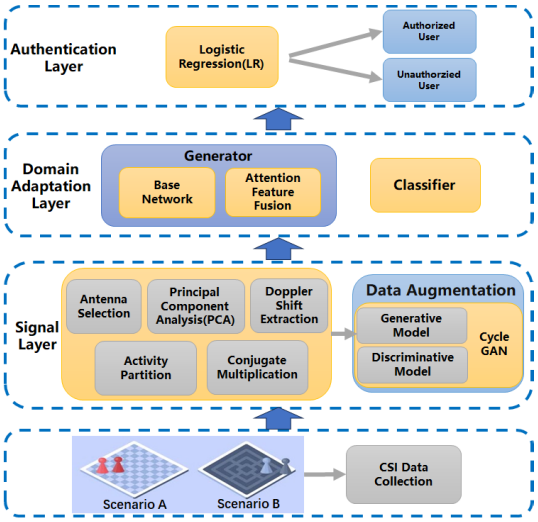


Fig. 2. System infrastructure. Wi-Dist contains three layers: signal layer, domain adaptation layer and authentication layer.

To the best of our knowledge, this is the first cross-layer framework to achieve cross-environment behavior-based user authentication. Wi-Dist is evaluated in typical indoor environments, showcasing its robustness in user behavior-based authentication. The contributions of this work can be summarized as follows:

- An adversarial learning-based data synthesizer is proposed to generate high-quality training data.
- An innovative adversarial network model is proposed to mitigate the model's environment dependency and distinguish unknown activities from known ones in the altered environment.
- A dynamic dual sliding window activity state tracker is proposed to partition continuous activities into single activities. Achieve user authentication through continuous activity recognition, thereby achieving access control in specific areas.
- Wi-Dist is implemented and extensive experiments are conducted to evaluate its performance. The results demonstrate the feasibility, robustness, and effectiveness of the system.

The remainder of the paper is structured as follows: Sec. II surveys and presents related work. Sec. III introduces Wi-Dist, a cross-layer, jointly optimized user authentication framework based on off-the-shelf Wi-Fi. Sec. IV outlines the experimental environment settings, and evaluation metrics, and details the comprehensive experiments conducted. In Sec. V, the limitations of Wi-Dist and future prospects are discussed. Finally, the conclusion of the entire text is provided in Sec. VI.

II. Related Work

A. Wi-Fi-Based User Authentication

In recent years, significant advancements have been made in contactless human behavior recognition using Wi-Fi, leading to the development of numerous valuable applications such as gestures [20], [27], gait [29], [30], and respiration [31]. Taking advantage of the rapid development of Wi-Fi technology, researchers have begun exploring Wi-Fi-based user authentication. Wi-Fi Channel State Information (CSI) is influenced by behavioral changes, embodying characteristic information that reflects these alterations while encapsulating fine-grained details regarding the propagation environment. Employing diverse data processing techniques, such as signal feature extraction and analysis, facilitates the effective detection of human presence and behavior identification. For instance, in [32], WiWho is proposed to implement user authentication through sensing human gaits. In [22], FingerPass leverages surrounding Wi-Fi signals' CSI to authenticate users continuously via finger gestures. Additionally, in [6], an adversarial learning-based model is proposed to identify individuals using Wi-Fi signals independently of gestures. However, when using Wi-Fi, changes in the environment where the activity is conducted can lead to the deterioration of the system's performance. In this research, we propose a novel framework to achieve cross-environment user authentication.

B. Data Augmentation

A practical approach to address the issue of data scarcity is through data augmentation. This deep learning technique expands the original dataset by generating new training data from the existing dataset. Yang et al. [33] utilize a limited subset of paired Kinect and RFID data to augment their dataset for subsequent model training. Building on this, Wang et al. [34] refine the previous work by introducing a forward kinematics layer, which generates synthetic Kinect poses using simulated human skeletal pose data and quaternion data. Patel et al. [35] propose the use of conditional GAN (CGAN) for data enhancement in the context of automatic modulation classification (AMC). They adopt a small portion of real data to generate high-quality, labeled radio modulation data conditioned on auxiliary information. Additionally, Yang et al. [36] address the issue of sensitivity diversity by mapping measurements of signal strength and utilizing the short-time Fourier transform (STFT) to construct generalized tensor data. Unfortunately, these approaches mainly apply data augmentation to the raw signals, limiting their effectiveness.

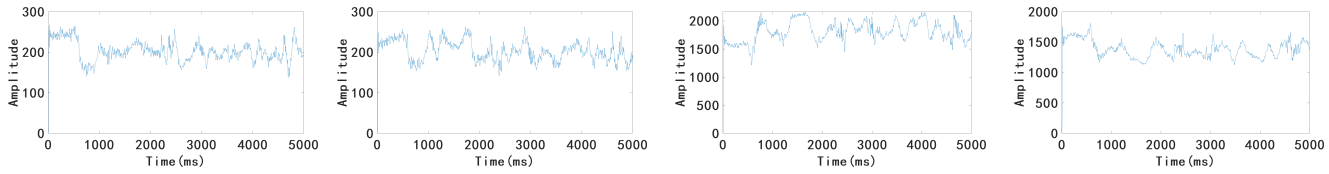
III. Attack Model & System Design

A. Attack Model

In real-world scenarios, Wi-Dist may encounter two types of attacks: brute force attacks and observation attacks.

1) Brute Force Attacks: The adversary does not know the user group's action password library. Brute force attacks entail the guessing of behavioral passwords through the trial of various activity combinations. However, unlike conventional passwords which are vulnerable to exhaustive methods, the Wi-Dist system needs a user to perform the activities continuously consuming considerable time and effort. Additionally, abnormal behavior by the attacker is more prone to detection, impeding the successful execution of brute force attacks.

2) Observation Attacks: The adversary attempts to observe user behavior to infer the activity base of user groups. In real-life environments such as home or office, systems are often isolated from the outside world. This closed nature makes



(a) Antenna Pair I (b) Antenna Pair II (c) Antenna Pair III (d) Antenna Pair IV

Fig. 3. Antenna selection.

it very challenging for an attacker to collect the password through observation, This will reduce the risks of observation attacks effectively.

Wi-Dist system aims to authenticate a user group. The users in the group share the same behavioral password library. Each one can randomly pick up the activities from the library and make a behavioral password with any length and combination. The typical scenario of Wi-Dist is the indoor environment, such as an office or household. There are two steps to authenticate a user in the group. Firstly, the user informs Wi-Dist that he belongs to the group. Secondly, Wi-Dist checks if he belongs to the group by his behavioral password. It is assumed that the first stage is completed. Wi-Dist system focuses on group-based user authentication. Wi-Dist effectively thwarts unauthorized user access to hazardous zones by scrutinizing user activity sequences.

B. System Overview

The schematic of the Wi-Dist system framework is depicted in Fig. 2, consisting of three layers: signal layer, domain adaptation layer, and authentication layer. Wi-Dist initially acquires CSI data either from the same domain or across domains (Scenario 1 and Scenario 2 depicted in the figure), which is subsequently processed at the signal layer. This layer comprises two components. The first component conducts preliminary processing on the incoming CSI, encompassing Antenna Selection, Conjugating Multiplication, PCA, Activity Partitioning, Data Augmentation, and Doppler Shift Extraction. These processes are responsible for data calibration and activity segmentation. The second component involves data augmentation, where the calibrated data is employed to synthesize high-quality virtual data. This procedural framework encompasses two pivotal models: the generative model and its counterpart, the discriminative model. The generative model gains knowledge of the underlying distribution of a subset of the real samples and synthesizes more virtual samples with the same distribution as the real ones. The discriminative model aims to determine whether a sample is genuine or not. Through the adversarial optimization between them, the generated samples progressively approach the characteristics of the real ones.

The domain adaptation layer comprises an adversarial neural network, which includes a feature generator and a domain classifier. The feature generator is composed of the base network and attentional feature fusion, used for feature extraction and subsequent feature fusion, respectively. The domain classifier strives for accurate classification, while the feature generator adopts an adversarial approach to trick the domain classifier. At the user authentication layer, a user authentication mechanism using logistic regression is leveraged to verify user legitimacy through continuous behavior recognition, and then determine whether to grant access rights to the corresponding area. There is an action library for each specific area, and there is a user group. All users in this user group know all the actions in the legal action library. Each user can select any number of repeatable actions from the library as the user's unique behavioral password.

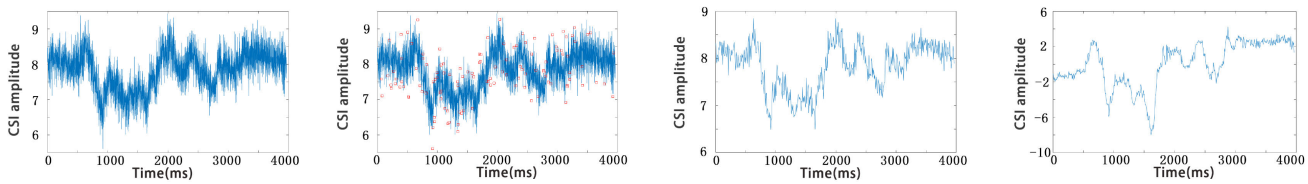
C. The Signal Layer

1) Antenna Selection, Conjugating Multiplication and Principal Component Analysis(PCA): The received CSI can be articulated as follows:

$$CSI(f,t) = A_{\text{noise}}(f,t)e^{-j\theta_{\text{offiet}}(f,t)}\left(H_s(f) + H_d(f,t)\right) \tag{1}$$

where A_{noise} represents the amplitude noise, θ_{offiet} is the random phase offset and $H_s(f) + H_d(f,t)$ represents static and dynamic components, respectively. In the experiment, each transmitter or receiver has two antennas, in total of four antenna pairs for one transmitter-receiver pair. Fig. 3 shows the amplitude distribution of four sets of CSI on one receiver of the same action. A larger amplitude is apt to result in a greater static response, and a more significant variance is apt to result in a greater dynamic response. Pair IV has the highest amplitude, but the dynamic information is not obvious, and Pair I and II have the lower amplitude but the most obvious dynamic information. Moreover, noise or errors in the raw CSI data may hinder the use of Wi-Fi signals for activity sensing. Therefore, the relative CSI is derived through the computation of conjugate multiplication applied to the CSI of two contiguous antennas.

$$\begin{aligned}
H_{(f,t)} &= \text{CSI}_1(f,t)\overline{CSI_2(f,t)} \\
&= A_{\text{noise}}(f,t)^2\left(H_{s1}(f) + H_{d1}(f,t)\right) \\
&\quad \times \left(\overline{H_{s2}(f)} + \overline{H_{d2}(f,t)}\right) \\
&= A_{\text{noise}}(f,t)^2(\underbrace{H_{s1}(f)\overline{H_{s2}(f)}}_{(1)} + \underbrace{H_{s1}(f)\overline{H_{d2}(f,t)}}_{(2)} \\
&\quad + \underbrace{\overline{H_{s2}(f)}H_{d1}(f,t)}_{(3)} + \underbrace{H_{d1}(f,t)\overline{H_{d2}(f,t)}}_{(4)}) \\
&\approx A_{\text{noise}}(f,t)^2(H_{s1}(f)\overline{H_{s2}(f)} + H_{s1}(f)\overline{H_{d2}(f,t)} \\
&\quad + \overline{H_{s2}(f)}H_{d1}(f,t)).
\end{aligned} \tag{2}$$



(a) Original CSI (b) After hampel filter (c) After butterworth filter (d) After PCA

Fig. 4. CSI waveforms.

Meanwhile, the antenna Pair I and II are chosen to retain more dynamic information [6], shown as Eq. (2), where (1) is a time-invariant term; (2) and (3) are time-varying terms.

Next, Hampel filtering is adopted to detect and remove outliers, as indicated by the red circles in Fig. 4 (b). Then, the utilization of a Butterworth filter is implemented for the purpose of attenuating high-frequency noise. At the carrier frequency of 5 GHz, the sampling rate of 1000 Hz and with a maximum speed of 3.2 m/s for the seven basic activities, the Butterworth filter's cutoff frequency is set to $\omega_s = 120/(f_n/2)$, where f_n represents the sampling rate. Fig. 4 (c) illustrates the result after applying the Butterworth filter. The results demonstrate the effective removal of significant noise from the raw CSI measurements using the Butterworth filter. Fig. 4 (d) illustrates the first principal component after PCA.

2) Consecutive Activity Partition: The consecutive activity partition aims to identify the initiation and termination points of individual activities within a time series of successive motions. In this section, the local and global sliding double windows are employed to locate the duration of the consecutive activity. This process is crucial for extracting features related to single movements. The CSI amplitude is the most sensitive to human motions. The amplitude fluctuates when there are human motions. Hence, the amplitude variance is chosen as a state change indicator for monitoring.

There is a global sliding window X_1 and a local sliding window X_2. The sampling frequency is f packet/s, and the size of X_1 is set to be five times of f. There is a 2-second overlap between each of the two adjacent global windows. Meanwhile, the size of the local window X_2 is set to be $1/2f$, and its sliding frequency is $1/4f$.

At the i_{th} slide of X_1, the variance $\sigma_i^2(X_2^n)$ (n represents the nth slide of X_2 in X_1) of time series in each X_2 is calculated, and get its average variance $\mu_i(\sigma)_i^2(X_2^n))$ and its standard deviation of variance $\sigma_i(\sigma)_i^2(X_2^n))$. The segmentation threshold ϕ_i is set to be $\alpha * \mu_i((\sigma)_i^2(X_2^n)) + \beta * \sigma_i((\sigma)_i^2(X_2^n))$, here the weight values are set to be $\alpha = 2/3$ and $\beta = 1/3$ according to the experience. When $\sigma_i^2(X_2^n) > \phi_i$, the lower bound of the current interval in the local sliding window C_1 is recorded. On the contrary, $\sigma_i^2(X_2^n) < \phi_i$ the upper bound of the local sliding window C_2 is recorded.

There is a problem that the current state A is identified as a motion, while its preceding state A^- and its succeeding states A^+ are identified as non-motion. At this moment, the current segment is converted to its neighbor's state. An enhanced method is proposed to tackle this problem while the pseudocode is provided in Algorithm 1. All the time series of segments are checked one by one and corrections are made when necessary.

As presented in Fig. 5, where X_1 and X_1' represent two adjacent global sliding windows, X_2 represents the local sliding window. There are five single activities distinguished.

Algorithm 1 Sequential Activity Partition

input : The time series H, the sampling frequency f, the size of X_1 $5f$, the size of X_2 $1/2f$, the sliding frequency of X_2 $1/4f$, $\alpha = 2/3$, and $\beta = 1/3$
output: The segmented segments $[C_1, C_2]$

```
for i = 1 to |H| By X_1 do
    for n = 1 to X_1 By X_2 do
        calculate the variance σ_i^2(X_2^n) //n represents the n_th slide of X_2 in X_1
    calculate average variance μ_i(σ)_i^2(X_2^n)) and standard deviation of variance σ_i(σ)_i^2(X_2^n))
    Threshold φ_i= α * μ_i(σ)_i^2(X_2^n)) + β * σ_i(σ)_i^2(X_2^n))
    for n = 1 to X_1 By X_2 do
        if σ_i^2(X_2^n) > φ_i then
            C_1 ← the lower bound of X_2
        else
            C_2 ← the upper bound of X_2
    for n = 2 to X_1-X_2 By X_2 do
        if (A == motion && A^- == non-motion && A^+ == non-motion) then
            current segment ← non-motion
            C_2 ← the upper bound of A
        if A == non-motion && A^- == motion && A^+ == motion then
            current segment ← motion
            C_1 ← the lower bound of A
```

3) Doppler Shift Extraction: The Doppler effect quantifies the alterations in the frequency of observed waves resulting from the relative motion between a transmitter and a receiver [37]. The Doppler frequency shift(DFS) is represented as follows:

$$f_D = -\frac{1}{\lambda}\frac{\mathrm{d}}{\mathrm{dt}}p(t) \tag{3}$$

where λ represents the signal's wavelength, and $p(t)$ represents the distance traveled by the reflected signals. However, in a real-world environment, signals undergo multiple reflections from the transmitter to the receiver, and reach the

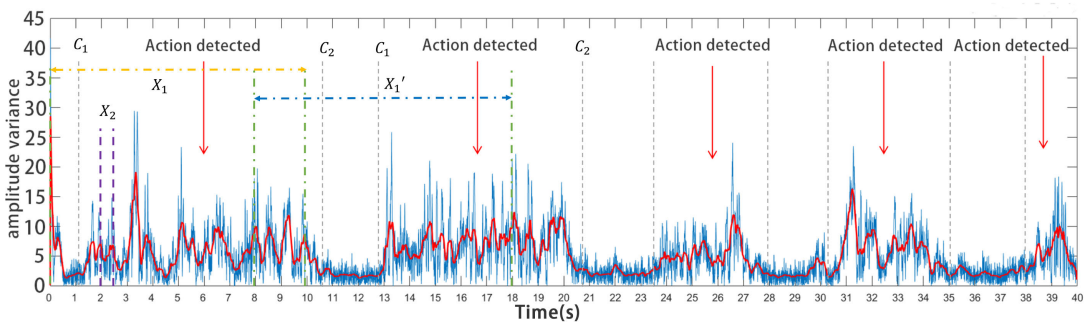


Fig. 5. Activity monitor.

destination through diverse pathways. Therefore, the response at the frequency f and time t received by the receiver is a superposition of the individual path responses, expressed as follows:

$$H(f,t) = \sum_{i=1}^{L} a_i(t) e^{-j2\pi f \tau_i(t)} \tag{4}$$

where $a_i(t)$ is the composite attenuation factor of the k path, $\tau_i(t)$ is the k_{th} path from the transmitter to the receiver, L is the number of paths.

The flight time $\tau_i(t)$ for the k_{th} path corresponds to the duration it takes for the signal to traverse the path length $d_i(t)$, i.e., $d_i(t) = c\tau_i(t)$, where c denotes the speed of light. The representation of the channel response involves the DFS associated with each path, formulated as follows:

$$H(f,t) = H_s(f) + \sum_{i \in N_d} a_i(t) e^{j2\pi \int_{-\infty}^{t} f_{D_i}(u)\mathrm{du}} \tag{5}$$

where $H_s(f)$ is the sum of CSI responses caused by static invariant paths, and N_d is a set representing dynamic CSI paths. In fact, $a_i(t)$ and $f_{D_i(t)}$ are almost the same in a short time interval so that the spectrum can be acquired in the following manner:

$$DFS(f,t) \approx H_s(f) + \sum_{i \in P_d} a_i(t) B\left(f_{D_i}(t)\right) \tag{6}$$

where $B(\cdot)$ represents a window function employed to extract the signal segment of interest.

The partitioned CSI values are obtained. Subsequently, the first principal component goes through an STFT to generate the spectrogram. A Gaussian window with a duration of 0.1 s is employed in the STFT to ensure a constant time-of-flight and attenuation factor within a short time interval. Zero padding is applied to enhance spectrogram resolution. Non-overlapping CSI fragments are then stitched to form a complete spectrogram ultimately. Fig. 6 presents the resulting Doppler frequency shift. Different body parts move at varying speeds during motion, leading to distinct Doppler frequency shifts. Consequently, the spectrum diagrams of different actions exhibit noticeable differences. The above results verify that each activity has its underlying unique signal representation, providing a solid foundation for behavior-based Wi-Fi user authentication.

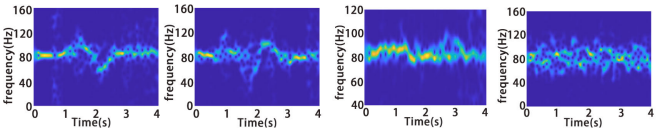


Fig. 6. Spectrograms of different actions. (a) Squat; (b) Jump; (c) Wave arms; (d) Run in place.

4) Data Augmentation: The learning models suffer from poor performance when there is insufficient training data. To address this issue, the sensing data after data calibration are utilized to do the data augmentation. This improves the quality of generated sensing data. The process can be regarded as a stack of multiple GAN networks, with each GAN network learning information at different scales or blocks. By applying this approach to the calibrated data, the generated CSI spectrograms with arbitrary time lengths can be synthesized. There are multiple generators, denoted as G_n, and multiple discriminators, denoted as D_n, n = 1, 2,. . . . . The generator G synthesizes virtual data that closely resembles the real training data, albeit with a different distribution. The discriminator D verifies if a sample is real or not. The adversary ensures that the generated data closely matches the distribution of real sensing data. The functions of the generator and discriminators are as follows:

$$\min_{G_n} \max_{D_n} \mathcal{L}_{\text{adv}}\left(G_n, D_n\right) + \alpha \mathcal{L}_{\text{rec}}\left(G_n\right) \tag{7}$$

where $\mathcal{L}_{\text{adv}}$ is the adversarial loss, which introduces a penalty that regulates the distribution misalignment between the authentic sample and the synthetic sample. $\mathcal{L}_{\text{rec}}$ represents the reconstruction loss, ensuring the existence of a specific set of noise maps capable of generating data.

In this scheme, the training data set is defined as $X^r = \{x^{r_1}, x^{r_2}, \ldots, x^{r_M}\}$, the corresponding label is defined as $Y^{k_i} \in Y = \left\{y^1, y^2, \ldots y^K\right\}$ where K is the action category number, the testing data set is defined as $X^u = \{x^{u_1}, x^{u_2}, \ldots, x^{u_N}\}$, and the testing data set has no data labels. The synthetic signal $X^g = \{x^{g_1}, x^{g_2}, \ldots, x^{g_L}\}$, the generator G is synthesized by the label of Y. In the above description, the values of M, N, and L are not necessarily equal; an entire data set is denoted as X, which is the union of X^u, X^r, and X^g, as follows.

$$X = X^r \cup X^u \cup X^g. \tag{8}$$

The model consists of multiple stacked generators $\{G_0, \ldots, G_N\}$ and multiple discriminators $\{D_0, \ldots, D_N\}$. Each generator G_n is trained to deceive the corresponding discriminator D_n, which aims to distinguish between real and synthetic signals. As shown in Fig 8, the generation process starts from the Nth generator and then passes through the generators of the previous layer to a finer scale. As the layers ascend, the receptive field size diminishes, resulting in composite signals that adhere to the desired specifications.

The objective is for the generator G_n to take a random noise vector z^g as input, along with a corresponding label y, and produce synthesized data that closely resembles real data, which is challenging to distinguish. In other words, we aim for the mapping $\mathrm{G_n} : (z^g, y) \rightarrow x^g$. However, the wireless signal synthesized by the signal synthesizer is susceptible to noise and might be distorted during the generation process, even though the sensing data are preprocessed.

The specific features can be extracted automatically with convolutional operations. The convolution operations for automatic feature extraction are incorporated to enhance the generator's ability to generate synthetic data consisting of the real data distribution. After the convolution operation, the synthetic data is enhanced by retaining the essential features and removing the noise interference. A least squares smoothing filter is then applied to eliminate residual noise further:

$$L_g = -\frac{1}{m} \sum_{(z^g, y)} \log \left(D \left(G \left(z^g, y \right) \right) \right). \tag{9}$$

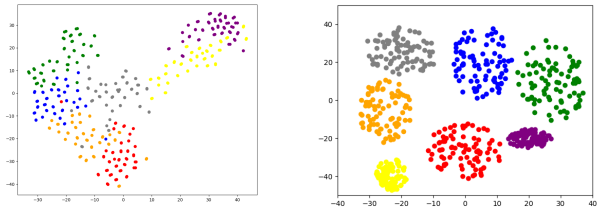
The discriminator D is presented with both authentic data in the form of $< x^r, y >$ and synthesized data represented as $< x^g, y >$ during its input phase, and calculates the likelihood that the input data is synthetic. In order to enable the discriminator to work more effectively, convolution operations are introduced to improve the accuracy of discrimination. In addition to training the discriminator to differentiate between genuine and synthetic data, batch signals are employed to enhance the antagonistic effect. The parameters are updated with the following:

$$\mathrm{L}_{d_g} = -\frac{1}{m} \left[\sum_{x^g, y} \log \left(1 - D \left(x^g, y \right) \right) + \sum_{x^r, y} \log D \left(x^r, y \right) \right] \tag{10}$$

where $D(x, y) = R/F$ represents the probability of the input data being a synthetic signal, and $1 - D(x, y)$ represents the probability of it being a real signal. R indicates that the data source is real data, and F indicates that the data source is synthetic data.

Finally, the network model iteratively updates its parameters in the training phase until the minimum overall loss function is obtained. The generation and discriminative adversarial process of the network model terminates when the discriminator can no longer distinguish between synthetic and real signals. The network parameters undergo optimization through stochastic gradient descent (SGD) employing the learning rate set at 0.0001.

The t-SNE [38] method is used to analyze the distribution of synthetic data. Fig. 7 presents the data visualization result after dimensionality reduction processing on the synthetic data generated from the real processed CSI measurements. Seven basic activities are conducted for the experiments; different colored dots represent different activities. Fig. 7 (a) represents the visualization result of spectrograms. Fig. 7 (b) is the visualization result of real data spectrograms. For each activity, the real and synthetic data are well aligned. As a result, the virtual data distribution is consistent with the real Wi-Fi data's distribution, and data diversity is expanded.



(a) (b)

Fig. 7. (a) The visualization result of the acquired partial processing doppler shift map after dimensionality reduction; (b) The visualization result of real data and an equivalent quantity of generated data after dimensionality reduction.

D. Domain Adaptation Layer

Due to the insufficient spatial resolution, the features derived from the movement induced CSI dynamics are specific to the environment where movement occurs. The classifier's performance may deteriorate when trained in one scenario and applied in the other. In addition, the unknown activities conducted in the new scenarios will be mistakenly regarded as known activities. These issues are addressed in this section.

A domain adaptation model based on adversarial networks is proposed to distinguish between known and unknown samples in a new environment. It comprises two primary components: a classifier C and a feature generator G_F. The Generator G_F consists of base network and attentional feature fusion(AFF). The base work model drives features from the spectrograms, and the attentional feature fusion model generates the corresponding feature enhancement. The classifier C is trained to set the pseudo decision boundary that separates known and "unknown" classes. The feature generator G_F is trained to generate features difficult for classifier C to classify accurately, thereby deceiving it. The classifier C maximizes its classification error by leveraging the ability of the generator G_F to manipulate the probability of the unknown class, either increasing or decreasing it. The proposed adversarial-based model trains the classifier and feature generator in this adversarial manner together to mitigate the negative impact caused by environmental changes and identify suspicious users based on their unauthorized activities.

The proposed AFF model combines local and global non-linear features and integrates multi-scale feature contexts within the attention module. Furthermore, the model puts forward the scale-related issue in channel attention through pointwise convolution. This approach promotes the fusion of features driven by attention. Each domain has two mutually perpendicular sensing links for the data collection. The inputs of generator G_F are spectrograms from the horizontal link

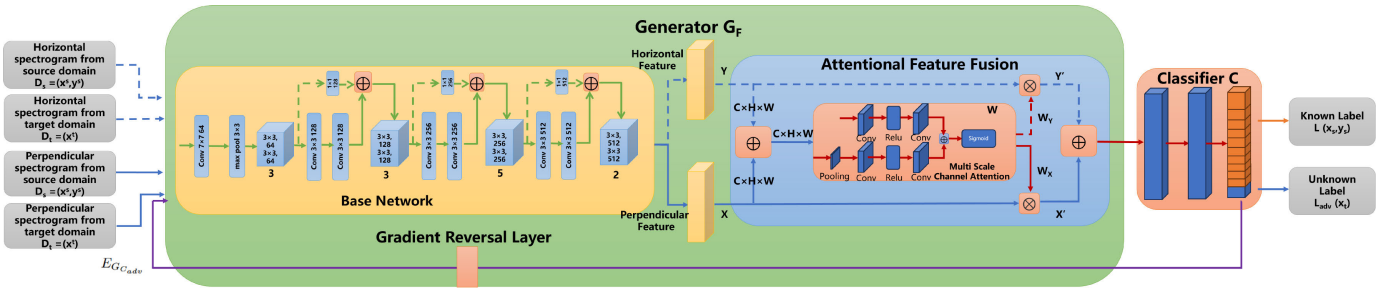


Fig. 8. Domain adaptation model, which is composed of two parts, the generator G_F and the classifier C, as depicted in Fig. 8. The generator contains the base network and attentional feature fusion model, and the inputs of the base network are spectrograms from the horizontal link (indicated by the solid blue line), as well as the perpendicular link (indicated by the dotted blue line). It also contains a gradient reversal layer for facilitating the simultaneous update of both classifier and generator parameter. The output of the model contains two parts: the known label and the unknown label.

(indicated by the solid blue line in Fig. 8), as well as the perpendicular link (indicated by the dotted blue line in Fig. 8).

1) Generator G_F: The generator G_F includes the base network module and the attention feature fusion module. In Wi-Dist, a deeper auto-regressive base network with a ResNet-34 encoder is leveraged to derive the fine-grained features from the spectrograms. The base network takes the spectrograms from source and target domains as inputs, pre-trained on the spectrogram network [20]. Residual learning is implemented in each stacking layer, and the residual block can be expressed as:

$$y = F(x, \{W_i\}) + x \tag{11}$$

where x denotes the input vector, and y represents the output vector of the layers. $F(x, W_i)$ signifies the residual mapping to undergo learning. Before feeding into Resnet-34, the spectrogram is preprocessed by arbitrarily sampling its shorter side in the range of [256, 480] for scale augmentation. A random 224 × 224 crop is sampled from either a spectrogram or its horizontally flipped counterpart, with per-pixel mean subtraction applied. Batch normalization (BN) is applied immediately after each convolution and prior to activation. The resulting spectrogram serves as input for the base network.

Resnet-34 is excellent at deriving non-linear features and is chosen as the basic feature extractor. The network takes a spectrogram input with three channels and a size of 224 × 224 from the input layer. Then, a convolutional layer with a 7 × 7 kernel size is employed for spectrogram enhancement. Next, the feature extractor captures the basic linear relationship and passes it through a 3 × 3 pooling layer. At this time, the extractor size becomes 56 × 56, and the number of channels is 64. The model consists of 4 residual parts in total. The first part contains three residual modules and does not use convolution for downsampling. Each residual block comprises 64 convolutional kernels, each with a size of 3 × 3, a stride of 1, and padding of 1. The remaining three parts comprise 4, 6, and 3 residual blocks, respectively, with downsampling occurring in the first residual block, as illustrated in Fig. 8. The initial convolutional layer in the residual block employs a 3 × 3 convolutional kernel with a stride of 2, a padding of 1, and doubles the channel count. The second convolutional layer within the residual block uses a 3 × 3 convolutional kernel with a stride of 1, and a padding of 1, and retains the same number of channels. To align the residual bypass component with the convolutional layer, a 1 × 1 convolution with a stride of 2 is employed in the downsampling layer, doubling the channel count. The remaining residual blocks cannot be downsampled.

After the feature extraction, the features derived are further refined by the attentional feature fusion model. It seeks to add attention weights to the high-level input features with the least overhead. The attentional feature fusion efficaciously combines high-level features, enhancing their overall quality. In the multi-scale channel attention module, complete global information is acquired through global average pooling, while more precise local information is obtained through pointwise convolution. Two branches channel attention to identifying global features as well as local features. Local features are derived as follows:

$$\mathbf{L}(\mathbf{X}) = \mathcal{B}(\mathrm{PWConv}_2(\delta(\mathcal{B}(\mathrm{PWConv}_1(\mathbf{X}))))) \tag{12}$$

where X is the input feature, $\mathcal{B}$ is the regularization layer, δ represents ReLU activation function.

In the attention feature fusion module, X corresponds to features in the horizontal direction, and Y corresponds to features in the perpendicular direction. W_X and W_Y represent the fusion weights $\mathbf{M}(\mathbf{X} \uplus \mathbf{Y})$ and $\mathbf{X} + (1 - \mathbf{M}(\mathbf{X} \uplus \mathbf{Y}))$, respectively. $\mathbf{X}'$ are derived as follows:

$$\mathbf{X}' = \mathbf{X} \otimes \mathbf{M}(\mathbf{X}) = \mathbf{X} \otimes \sigma(\mathbf{L}(\mathbf{X}) \oplus \mathbf{G}(\mathbf{X})) \tag{13}$$

where $G(X)$ represents channel attention, mainly used to calculate global features. First, the input feature X undergoes global average pooling, followed by a calculation similar to $L(X)$. $M(X)$ denotes the attention weight generated, $\oplus$ signifies broadcast addition, σ is the activation function, and $\otimes$ indicates element multiplication.

Then, the attentional feature fusion model fuses the features obtained from the aforementioned module, as shown below:

$$\mathbf{Z} = \mathbf{M}(\mathbf{X} \uplus \mathbf{Y}) \otimes \mathbf{X} + (1 - \mathbf{M}(\mathbf{X} \uplus \mathbf{Y})) \otimes \mathbf{Y} \tag{14}$$

where Z represents the fused feature, $\uplus$ represents the initial feature fusion. So far, the enhanced features can be provided to the classifier C for the action classification.

2) Classifier Module: The classifier C extracts features from the generator G_f and classifies them into $K + 1$ categories. Here, K represents the number of known or source

categories, and the $(K+1)_{th}$ index signifies the probability for the unknown category. A structure that comprises a fully connected layer followed by an activation function is employed to transform the feature matrix into a lower-dimensional vector. Then, each value of this vector is converted into a representation of probability. The final outputs are the logarithm $\{l_1, l_2, l_3 \ldots l_i \ldots, l_{K+1}\}$ for each data sample, where i denotes one class. The normalized exponential function. Softmax is employed to convert the logarithm into probabilities for each class. The probability of data sample x being assigned to category j can be formulated as:

$$P(y = j \mid x) = \frac{\exp\left(l_j\right)}{\sum_{i=1}^{K+1} \exp\left(l_i\right)} \tag{15}$$

where l is the logit vector, $P(y = j \mid x)$ is a $K+1$-dimensional probability associated with x.

As the target domain data are unlabeled during training, the proposed method endeavors to establish a pseudo-decision boundary between known source samples and target samples via the weak training of the domain classifier C. This results in assigning the target samples to the unknown category. Simultaneously, C is trained to output $P(y = N+1 \mid x_j^t) = T$ for the unknown class. Subsequently, G_F undergoes adversarial training to deceive C. It equips G_F with the capacity to manipulate the 'unknown' class probability, denoted as $P(y = N+1 \mid x_j^t)$, to maximize the error of classifier C and align target samples with either known or unknown classes. It is assumed that the generative model is guided in constructing an effective boundary between known and unknown samples by the empirical threshold value, with T set to 0.5.

E. Authentication Layer

In the domain adaptation layer, the result logarithm $\{l_1, l_2, l_3 \ldots l_i \ldots, l_{K+1}\}$ generated by the model is converted into the probability of each category through the softmax function. For the result of each behavior recognition in user authentication, The category with the top probability is chosen, a set of tuples (P_n, A_n) (n = 1,2... N) is derived, where P is the probability of the action, A is the action type, and N represents the number of authorized actions. This tuples (P_n, A_n) needs to be transformed to a triple P_n, T_n, S_n (n = 1,2... N), so that the activity identification results can be brought to the authentication layer. The variables can be explained as follows: P_n refers to the probability that user behavior is recognized as a certain type. This probability value is calculated by the adversarial network and indicates the confidence level for each action. T_n indicates whether the action recognized is consistent with the actual activity performed. If they are consistent, T_n is 1; otherwise, T_n is 0. S_n indicates whether the user is legitimate. This label is used to build a binary classification model to predict the legitimacy of a user. A user is considered legitimate when S_n is 1.

In order to set appropriate thresholds and achieve accurate user authentication at the authentication layer, a logistic regression linear classifier is employed for the threshold calculation and binary classification. The decision function of linear classification can be defined as:

$$y(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b \tag{16}$$

The logistic sigmoid function is utilized to compute the posterior probability of category C, as shown in the following expression:

$$P\left(C \mid \mathbf{x}\right) = y(\mathbf{x}) = \sigma\left(\mathbf{w}^{\mathrm{T}}\mathbf{x} + b\right) \tag{17}$$

where x represents the input vector, w denotes the weight vector, and b stands for the bias. Among them, $\sigma(.)$ is a logistic sigmoid function, characterized by the following expression:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{18}$$

During the training phase, S_n is a known label and is used to train the confidence threshold t. In the testing phase, we use P_n and T_n in the triples (P_n, T_n, S_n) as input features of logistic regression, and the final output is S_n. If $P\left(C \mid \mathbf{x}\right) \geq$ t, the user is verified to be a legitimate user, S_n is set to 1. Otherwise, the user will be an illegal user.

F. Training Phase

During training, the source domain samples are utilized using the following formulas:

$$E_C = \frac{1}{n_s} \sum_{i=1}^{n_s} L_C\left(G_F\left(x_i^s\right), y_i^s\right) \tag{19}$$

where L_C represents the standard cross-entropy loss function used to minimize classification errors. n_s signifies the feature count in the source domain. $G_F(x_i^s)$ represents the feature output of the generator G_F for the i_{th} feature in the source domain, and y_i^s corresponds to the class label of the respective i_{th} data sample in the target domain. The loss term E_C is fine-tuned to minimize the error of the classifier C. Subsequently, a binary cross-entropy loss is employed to maximize the error of C adversarially, thereby separating known and unknown samples, as shown in the following expression:

$$\begin{aligned} E_{C\,\mathrm{adv}} = &-\frac{1}{n_t}\sum_{j=1}^{n_t} T \log\left(P\left(y = K+1 \mid x_j^t\right)\right) \\ &-\frac{1}{n_t}\sum_{j=1}^{n_t}(1-T)\log\left(1 - P\left(y = K+1 \mid x_j^t\right)\right) \end{aligned} \tag{20}$$

where x_j^t represents the unlabeled instances of the target domain.

The classifier C aims to make the probability of the "unknown" class $P(y = K+1|x_i^t)$ identical to T. Conversely, differentiation of $P(y = K+1|x_i^t)$ from T is strived for by the generator G_F in order to maximize $E_{C_{adv}}$'s value. The ultimate training objective is defined as follows:

$$\theta_C = \underset{\theta_C}{\operatorname{argmin}} E_C + E_{C_{adv}}\theta_{G_F} \quad = \underset{\theta_{G_F}}{\operatorname{argmin}} E_C - E_{C_{adv}}. \tag{21}$$

To efficiently compute the gradient of $E_{C_{adv}}$, we employ a gradient reversal layer proposed by [39]. During the backward

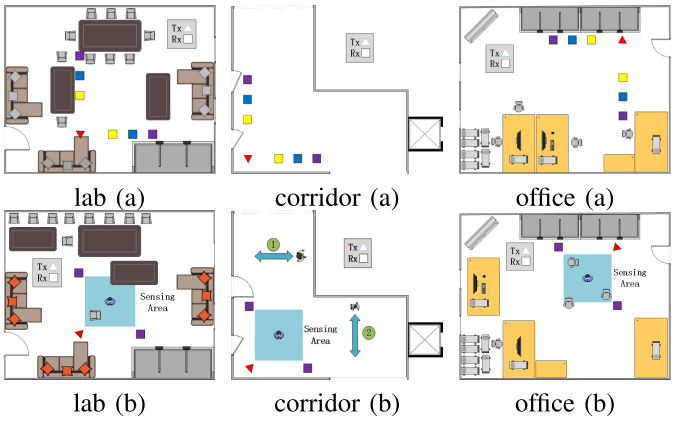


Fig. 9. Experiment scenarios.

process, the flipping of the gradient sign is enabled by this layer, facilitating the simultaneous update of both classifier C and generator G parameters.

IV. Experiments and Evaluation

A. Implementation

Currently, there are two kinds of off-the-shelf commercial network interface cards (NIC) available for Wi-Fi sensing: Intel 5300 and Intel AX210. AX210 NIC is optimized for modern, complex wireless transmission environments. It adopts new technologies as well as standards and offers better noise resistance than 5300 NIC. Therefore, AX210 NIC equipped with two antennas is utilized in this study. In Wi-Dist, the PicoScenes Toolbox Core [40] is employed on Ubuntu Desktop 20.04 LTS. Three desktop computers with Intel AX210 NICs are used, as shown in Fig. 13 (a), one transmitter Tx, and two receivers Rx1, Rx2 form a coordinate system. Tx is situated at the origin, while Rx1 and Rx2 are positioned 5 meters away from the transmitter along the x-axis and y-axis, respectively. The Tx-Rx1 line is perpendicular to the Tx-Rx2 line. All the devices are positioned at a height of 0.8 meters. The Wi-Fi channel frequency is set to 5 GHz on channel 165, with a bandwidth of 20 MHz and a packet transmission rate of 1000Hz. The antennas are arranged in a horizontal line, with a wavelength interval between them at each receiver. For each data stream, a total of 114 subcarriers are covered by two receiving antennas, with 57 subcarriers in each. MATLAB is employed for processing CSI data, and the Kinect 2.0 camera is used to capture ground truth.

Wi-Dist is assessed in three typical environments, lab, office, and corridor, as depicted in Fig. 9 (a). The lab measures 8.1 m $\times$ 7.2 m with several tables, chairs, sofas, and computers. The L-shaped corridor has a maximum width of 9.6 m, and a length of 10.2 m, and includes two doors. The office measures 8.8 m $\times$ 7.4 m and is furnished with four desks, four chairs, an air conditioner, nine computers, and two cabinets. Seven common actions are chosen: WA (wave arms), SQ (squat), JU (jump), RU (run), PU (push), LL (leg lift), and RH (raise the hand). Participants have different heights, weights, and body shapes. Participant's height ranges from 1.65m to 1.88m, and the weight ranges from 50-85kg.

The lab is considered the source domain, while the corridor and office are treated as the target domains. In total, 3500 samples (10 people $\times$ 50 samples $\times$ 7 activities) in the source domain and 1400 samples(10 people $\times$ 10 samples $\times$ 7 activities $\times$ 2 environments) are collected for the target domain. What's more, some UA (unknown activity) samples such as kick, walk, bend, and so on are collected in the target domain, a total of 2000 samples (10 people $\times$ 20 samples $\times$ 5 additional activities $\times$ 2 environments) to evaluate the cross-domain unknown activity identification. Moreover, synthetic samples($2 \times (3500 + 1400 + 2000)$) are generated to expand training data. In the training phase, samples from both the source and target domains are jointly used. The overall count of samples after data augmentation is 20700. From this pool, 18,630 samples are randomly allocated for the training set, while the remaining 2,070 samples are chosen for the testing set. Simultaneously, we conduct data augmentation after partitioning the original dataset into training and testing sets. This ensures that the augmented samples derived from each original sample are exclusively allocated for either training or testing, but not both. This practice is crucial for preventing data leakage and ensuring accurate evaluation results.

B. Overall Performance

1) Experimental Metrics: To evaluate Wi-Dist's performance comprehensively, True Positive Rate(TPR), False Positive Rate(FPR), Accuracy(ACC), Receiver Operating Characteristic Curve(ROC) and Area Under the Curve(AUC) are adopted for user authentication. TPR is denoted as the rate at which positive cases are correctly identified, which is a common method for user authentication [5], [22]:

$$TPR = \frac{TP}{TP + FN}. \tag{22}$$

FPR can be interpreted as the rate that negative cases that are incorrectly identified as positive:

$$FPR = \frac{FP}{TN + FP}. \tag{23}$$

ACC is defined as the overall correctness of the classifications, expressed as follows:

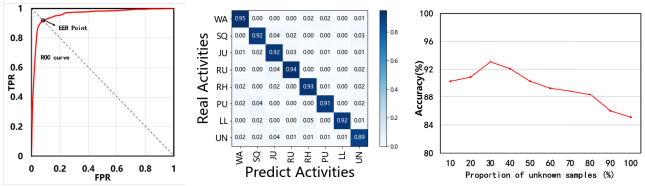
$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{24}$$

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively. Among the four indicators, two types of errors, FP and FN, need to be considered. These errors exhibit asymmetric costs and will impact the system and users differently. For FP, it mistakenly identifies an actual negative sample as a positive sample, which will have a direct threat to the security of the system, and unauthorized users may be mistakenly identified as legitimate users. For FN, positive samples are not correctly identified, which may result in false alarms or exception handling. While users may need to invest additional time and effort to address these issues, they generally do not affect the overall usability and security of the system.

The ROC curve represents the relationship between the True Positive Rate (TPR), plotted on the y-axis, and the False

TABLE I
THE OVERALL PERFORMANCE OF WI-DIST IN THE THREE TYPICAL ENVIRONMENTS

Environment	Accuracy Rate	TPR	FPR
Laboratory	94.72%	94.84%	7.80%
Corridor	93.03%	93.18%	9.11%
Office	91.63%	90.90%	12.83%



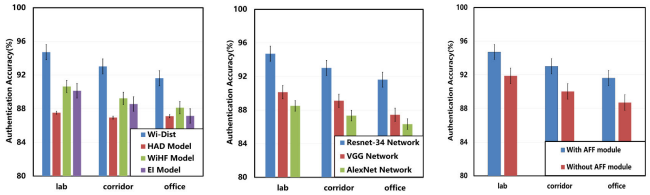
(a) (b) (c)

Fig. 10. (a) ROC curve of the average identification accuracy for one participant action recognition. (b) The confusion matrix of average action recognition rate in three environments; (c) The impact of unknown sample proportions.

Positive Rate (FPR), plotted on the x-axis, of the classifier at different classification thresholds. The curve closer to the upper-left corner (0,1) indicates better performance, which means the model achieves a high TPR while keeping the FPR low. AUC is an important performance metric, with an AUC closer to 1 indicating better model performance.

Cross-validation is used to evaluate Wi-Dist's accuracy in three environments. Initially, we randomly select half volunteers, with 2100 samples (5 participants $\times$ 7 activities $\times$ 60 samples per activity per participant $=$ 2100) in one specific environment. Subsequently, we perform k-fold cross-validation, with k $=$ 10 in the experiment. In order to ensure the impact of data segmentation, we conducted 10-fold cross-validation with multiple repetitions (250 times in the article), and finally obtained 2500 accuracy values to make the results more accurate. The average results in three typical environments are shown in Table I. Wi-Dist achieves an overall user authentication accuracy of 94.72%, 93.03%, and 91.63% in the lab, corridor, and office, respectively. What's more, Wi-Dist demonstrates satisfactory performance across all three scenarios, boasting a mean TPR of 92.97% and a mean FPR of 9.91%. However, distinctions emerge among these scenarios. Notably, Wi-Dist performs optimally in the lab setting, benefitting from consistent training and testing environments. Conversely, its performance dips in the office scenario, where intricate layouts and multipath effects pose challenges, resulting in a TPR of 90.90% and an FPR of 12.83%. Furthermore, the ROC curve delineates the tradeoff between TPR and FPR across various conditions, as depicted in Fig. 10 (a). At the equal error rate (EER) point, Wi-Dist achieves an average TPR and FPR of 92.34% and 8.63%, respectively, in practical applications. These findings underscore the viability of Wi-Dist for activity recognition.

In Fig. 10 (b), the confusion matrix illustrates the average recognition rates for seven typical and unknown activities in three different environments. The average accuracy of the framework is between 89% and 95%. The confusion matrix is consistent with the overall performance.



(a) (b) (c)

Fig. 11. (a) Impact of domain adaptation model; (b) Impact of basic networks;(c) Impact of AFF modules.

2) Impact of Unknown Sample Proportions: During the experiments, there are a few unknown activity samples to train the domain adaptation model. The influence of the ratio of unknown activity samples in the training data across three environments is examined. The results, presented as the average obtained from three environments, reveal in Fig. 10 (c) that with a 30% proportion of unknown samples, the maximum average accuracy reaches about 93.12%. With an increase in the proportion of unknown samples, there is a corresponding decrease in accuracy. This trend is due to the fact of overfitting during adversarial training. It's evident that the user authentication accuracy is affected by the proportion of unknown samples.

3) Comparison With Baseline Methods: We compare Wi-Dist with several cross-environment action recognition methods:

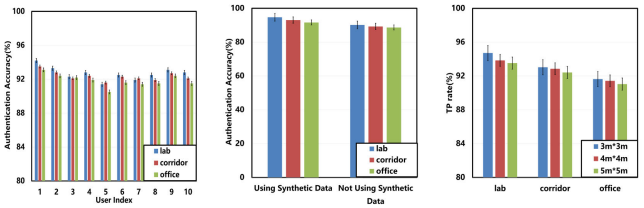
HDA: A cutting-edge unsupervised domain adversarial adaptation method is proposed in computer vision [41]. This approach effectively maintains the balance of achieving domain invariance while preserving domain-specific information through a heuristic search perspective.

WiHF: A split and stitching scheme is utilized to optimize collaborative learning under dual tasks for DNN and deduces motion change patterns of actions across different environments from Wi-Fi signals [27].

EI: EI stands as a state-of-the-art Unsupervised Domain Adaptation for Human Activity Recognition (UDA-HAR) model. It introduces a discriminator to learn domain-invariant features through adversarial training, minimizing the discrepancies between the source and target domains [19].

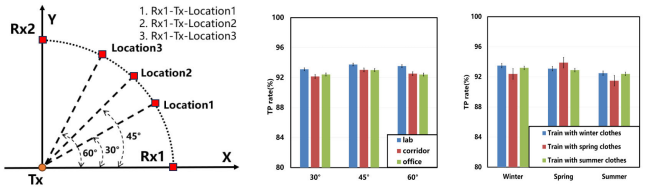
Wi-Dist undergoes a comparative analysis with three state-of-the-art baseline methods across three typical indoor environments, and the results are the averaged outcomes obtained from these settings, as shown in Fig. 11 (a). The accuracy for Wi-Dist, HDA, WiHF, and EI are 93.12%, 87.19%, 90.04%, and 90.34%, respectively. Wi-Dist performs the best since only Wi-Dist takes care of the issue of the cross-domain unknown activities identification.

4) Impact of Different Basic Networks: In this section, the basic network is replaced with the VGG network and the corresponding evaluation is made. In Fig. 11 (b), the depicted data illustrates the average user authentication accuracy across three environments for the three networks. Wi-Dist achieves an average accuracy of 93.12%, 90.13%, and 88.24% with the ResNet-34, VGG, and AlexNet, respectively. The results show that the basic network of Wi-Dist is the best.



(a) (b) (c)

Fig. 12. (a) Impact of user; (b) Impact of Syhthetic data; (c) Impact of sensing area.



(a) (b) (c)

Fig. 13. (a) Different orientations; (b) Impact of user orientation; (C) Impact of apparel.

5) Impact of AFF Module: The AFF module is used to refine extracted features. We assess its impact on Wi-Dist. Fig. 11 (c) shows user authentication accuracy in different environments with or without the AFF module. We can see that without the AFF model, the accuracy drops from 94.72% to 91.89% in the lab, from 93.03% to 90.04% in the corridor, and from 91.63% to 88.72% in the office. This attestation substantiates the performance enhancement achieved by the AFF module.

6) Impact of Users: Numerous physiological characteristics exhibited by the user, including height, body size, and weight, have the potential to exert influence on the precision of user authentication. How a user's physiological factors impact performance is studied. There are 207 randomly chosen samples as the testing set, about 50 samples for each subject, and none of them are in the training set. Of these, 105 samples are from the source domain and comprise seven known activities. The remaining 102 samples are selected randomly from the target domain. In Fig. 11 (a), the depicted data illustrates the average accuracy across the three environments for 10 users. The accuracy of all users is above 90%. This result can verify that Wi-Dist is strongly activity-related but user-agnostic.

7) Impact of Data Augmentation: Data augmentation is a technique of artificially increasing the training set by generating the data using the existing data. A virtual data synthesizer based on the calibrated data is proposed to reduce the labor-intensive data collection and obtain high-quality data. The evaluation of the overall performance encompasses an assessment of the impact wielded by the proposed data augmentation scheme. Fig. 12 (b) shows the impact of data augmentation on user authentication accuracy under three environments. Without using synthetic data, the accuracy dropped from 94.72% to 90.15% in the lab, from 93.03% to 89.27% in the corridor, and from 91.63% to 88.65% in the office. Obviously, the data augmentation scheme can generate effective training data and improve overall performance.

8) Impact of Sensing Area: To analyze the influence of sensing area size, three distinct sensing areas are configured across three scenarios, as depicted in Fig. 9 (a). Tx is located at a red triangle, and Rx is located at three different positions represented by yellow, blue, and purple squares, while the distances between Tx-Rx are 3m, 4m, and 5m, respectively. Fig. 12 (c) illustrates the results of the sensing area change. In the lab, Wi-Dist achieves an average TPR of 94.72%, 93.03%, and 91.63% at 3m $\times$ 3m, 4m $\times$ 4m, and 5m $\times$ 5m, respectively. In the corridor, Wi-Dist reaches an average TPR of 93.84%, 92.85%, and 91.42% at 3m $\times$ 3m, 4m $\times$ 4m, and 5m $\times$ 5m, respectively. In the office, Wi-Dist achieves an average TPR of 93.52%, 92.41%, and 91.04% at 3m $\times$ 3m, 4m $\times$ 4m, and 5m $\times$ 5m, respectively. These experiment results confirm that the sensing area can result in a slight decrease in TPR. Since the TPR of all the sensing areas is above 91%, Wi-Dist can be resilient to the changes.

9) Impact of User Orientation: In this section, we explore whether user orientation has an impact on authentication. As illustrated in Fig. 13 (a) within a 2D coordinate system, Tx is situated at the origin, whereas Rx1 and Rx2 are positioned 5 meters away from Tx along the x-axis and y-axis, respectively. During the experiment, the devices are fixed, and the user faces the line of sight (LoS) and switches his location from 1 to 3. The angles of Rx1-Tx-1, Rx1-Tx-2, and Rx1-Tx-3 are 30 degrees, 45 degrees, and 60 degrees, respectively. In each deployment, a total of 100 samples are amassed. There are some differences in user authentication results. The optimal location for achieving the highest accuracy in user authentication is within the vicinity of the perpendicular bisector of LoS, which is coincident with the conclusion in [42]. This confirms that Wi-Dist can be resilient to user orientation changes.

10) Impact of Apparel: The investigation delves into the influence of apparel across three typical environments. Wi-Dist is trained using samples from a user wearing winter, spring, and summer clothes, respectively. Subsequently, it undergoes testing and evaluation under three distinct seasonal attires. During the training phase, none of the testing data is accessible to the learning model. The outcomes, derived via a 10-fold cross-validation methodology as depicted in Fig. 13 (c), clearly indicate a substantial impact of clothing on the TP rate. When subjected to training with winter attire and subsequently tested with summer attire, Wi-Dist attains its lowest TP rate, approximately at 91.5%. Despite never being trained on samples of the other seasonal attires, Wi-Dist consistently maintains a TPR above 91%.

11) Impact of Environment Changes and Human Interference: To assess Wi-Dist's performance when there are environmental changes and human interference, we conduct experiments in three scenarios. As shown in Fig. 9 (b). Tx is represented as a red triangle, and Rx is represented as a purple square. The distance between Tx-Rx is 5m. The experimental results are derived from the identical settings outlined in the previous section (Impact of Sensing Area).

To assess the influence of environmental variations, adjustments are made in both office and laboratory settings. This involved opening the room door and repositioning tables and chairs. As illustrated in Fig. 9 (b), within the sensing area,

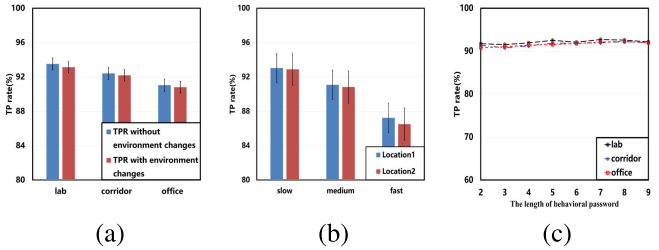


Fig. 14. (a) Impact of environmental changes; (b) Impact of moving subjects; (c) Impact of number of activities without attacks.

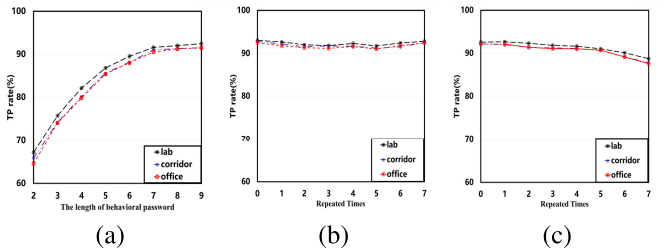


Fig. 15. (a) Impact of number of activities with attacks; (b) Impact of number of number of repeated activities without attacks; (c) Impact of number of number of repeated activities with attacks.

one chair with one table is placed in the lab and three chairs are placed in the office. As shown in Fig. 14 (a), TP rates with the environment changes and without environment changes are compared. The observed minimal impact of environmental changes on authentication suggests the robustness of Wi-Dist in cross-environment user authentication.

In accordance with the literature [43], the human comfortable walking speed (v) falls within the range of 1 to 1.5 m/s. To assess the influence of human movement interference, we design two tests in the corridor and set three distinct speed levels: slow (below (v − 0.3) m/s), medium ((v − 0.3) m/s ∼ (v + 0.3) m/s) and fast (above (v + 0.3) m/s). The participants are instructed to walk back-and-forth at different speeds (as mentioned above) along two paths, path 1 and path 2, as indicated by green circles 1 and 2 in Fig. 9, corridor (b). The approximate distances from both paths to the edge of the sensing area are 3m. In each test, we collect the same amount of samples. The results, depicted in Fig. 14 (b), clearly indicate that the moving subject interference has less impact on the authentication. Wi-Dist reaches an average TPR of 93%, 91%, and 87% at the low, medium, and high speed, respectively. In particular, when the speed is gradually increased from medium to high speed, there is a significant decrease in TPR. Overall, the authentication results are impacted by the speed of the subject.

12) Impact of Number Of Authorized Activities: To investigate the influence of the number of authorized activities on user authentication robustness, we conduct the following experiment. Prior to the experiment, each participant randomly chooses a specific number of activities from the activity library, which is unique to each user group. The experiments are carried out in three typical environments, lab, corridor, and office. The experimental results are presented in Fig. 14 (c). Since the authentication accuracy mainly depends on the authentication accuracy of the actions, we can see that in the absence of attackers, as the number of activities gradually increases, the certified TP rate is nearly stable.

Moreover, unknown users (attackers) who do not know the specific password types in the campaign participated in the experiment. They are only told the password length. The experiments are carried out in three typical environments, lab, corridor, and office. Fig. 15 (a) illustrates TP rates when behavioral password lengths are different. We observe that comparable authentication accuracy is achieved in all cases. When tested with 7 or more activities, the TP rate reached over 90%. This is because as the number of operations increases, it becomes more and more difficult for an attacker to crack the password through brute force attacks, and the probability of the behavioral code being cracked decreases. Additionally, when the number of selected activities exceeds 4, the authentication TP rate exceeds 80%, Wi-Dist can maintain relatively secure.

13) Impact of Number Of Number of Repeated Activities: To investigate the impact of the number of activity repetitions on user authentication accuracy, we conducted the following experiments. The behavioral password length is 7. The number of activity repetitions gradually changes from 0 to 7. The experiments are carried out in three typical environments, lab, corridor, and office. The results are presented in Fig. 15 (b). The figure shows that as the number of password repetitions increases, the overall authentication accuracies are comparable. Therefore, the number of activity repetitions does not influence the overall authentication accuracy.

Unknown users (attackers) who do not know the specific password types in the campaign participated in the experiment. They are only told the password length. The experiments are carried out in three typical environments, lab, corridor, and office. The results of these experiments are shown in Fig. 15(c). The figure indicates a gradual decrease in overall authentication accuracy as the number of password repetitions increases. This decline can be attributed to the heightened repetition rate, which renders passwords more susceptible to cracking and increases the likelihood of attackers being falsely authenticated as legitimate users. Conversely, when the number of password repetitions ranges from 0 to 5, the overall authentication accuracy remains relatively stable. This stability arises from the high complexity of the password, where authentication accuracy is predominantly influenced by the recognition accuracy of multiple actions. Moreover, minimal discrepancies in recognition accuracy between actions contribute to the steady overall accuracy within this range. At the same time, when the number of repetitions is less than 6, the overall accuracy of Wi-Dist is higher than 90%.

V. Limitation And Future Work

Wi-Dist system shows the feasibility of Wi-Fi-based user authentication that can authenticate a user by his authorized activities. However, it still has some limitations.

A. Multi-Subject Motion Sensing

One limitation of Wi-Dist lies in its inability to precisely differentiate the CSI dynamics induced by each subject's

activity within the sensing regions. Different frequency offsets and phase shifts are observed in the Wi-Fi signals induced by different moving subjects. This produces complex superpositions at the receiving end and is a tough issue for Wi-Fi-based activity recognition. The larger bandwidth [44] or the transceiver with more antennas [45] may help the signal separation. The optimal transceiver deployment and the sensing region extension are proposed to improve the sensing capability. However, it is challenging to leverage only one method to achieve multi-subject activity recognition.

B. Model Learning Cost

In the Wi-Dist system, a requisite quantity of training data is necessitated from both the source and target domains, though virtual data can be generated to achieve accurate cross-domain user authentication. To further reduce data collection efforts, future work will be leveraging more advanced algorithms such as Few-shot Learning [46] or Zero-shot Learning [47] to further reduce data collection efforts. The model is capable of being trained with an even smaller amount of labeled data in the target domain. Zero-shot Learning can be applied to the scenario without labeled training data, making the system more capable and adaptable.

VI. Conclusion

By combining Wi-Fi signals and activity sequences for user authentication, Wi-Dist not only streamlines the authentication process but also significantly strengthens system security. Unlike conventional authentication methods reliant solely on passwords, which are susceptible to theft or replication, the utilization of activity sequences adds an extra layer of protection. Moreover, the versatility of Wi-Dist extends beyond traditional authentication systems. Its innovative design allows for authentication across various indoor environments. Whether in an office setting, household, or other indoor spaces, Wi-Dist ensures consistent and reliable user authentication. In a scenario that is relatively closed and insensitive to observation attacks, Wi-Fi transceivers can be discreetly installed on the wall to monitor Wi-Fi signal changes while safeguarding privacy and security. Managers can utilize the intelligent detection system to ascertain the real-time safety of specific areas and adjust area size and authentication methods as needed. As users, individuals can enjoy a smarter and more convenient working environment without the need for additional identity verification or operations. Furthermore, Wi-Dist tackles numerous technical challenges through a sophisticated cross-layer joint optimization framework. This framework addresses complexities such as the accurate collection and analysis of Wi-Fi signal data, as well as the precise generation and identification of activity sequences. In summary, Wi-Dist's adaptability to diverse indoor environments and the successful resolution of technical challenges underscore its significant contributions to modern authentication systems.

References

[1] H. Kong, L. Lu, J. Yu, Y. Chen, and F. Tang, "Continuous authentication through finger gesture interaction for smart homes using WiFi," *IEEE Trans. Mobile Comput.*, vol. 20, no. 11, pp. 3148–3162, Nov. 2021.

[2] C. Shi, J. Liu, H. Liu, and Y. Chen, "WiFi-enabled user authentication through deep learning in daily activities," *ACM Trans. Internet Things*, vol. 2, no. 2, pp. 1–25, May 2021.

[3] S. Liu, Y. Chen, H. Wang, H. Liang, and L. Chen, "A low-calculation contactless continuous authentication based on postural transition," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 3077–3090, 2022.

[4] C. Shi, J. Liu, N. Borodinov, B. Leao, and Y. Chen, "Towards environment-independent behavior-based user authentication using WiFi," in *Proc. IEEE 17th Int. Conf. Mobile Ad Hoc Sensor Syst. (MASS)*, Dec. 2020, pp. 666–674.

[5] R. Ou, Y. Chen, and Y. Deng, "WiWalk: Gait-based dual-user identification using WiFi device," *IEEE Internet Things J.*, vol. 10, no. 6, pp. 5321–5334, Mar. 2023.

[6] H. Kong et al., "Push the limit of WiFi-based user authentication towards undefined gestures," in *Proc. IEEE Conf. Comput. Commun.*, May 2022, pp. 410–419.

[7] Y. Zhang, Y. Zheng, G. Zhang, K. Qian, C. Qian, and Z. Yang, "GaitSense: Towards ubiquitous gait-based human identification with Wi-Fi," *ACM Trans. Sensor Netw.*, vol. 18, no. 1, pp. 1–24, Oct. 2021.

[8] K. Choi, H. Ryu, and J. Kim, "Deep residual networks for user authentication via hand-object manipulations," *Sensors*, vol. 21, no. 9, p. 2981, Apr. 2021.

[9] H. Wang, T. Chen, X. Liu, and J. Chen, "Exploring the hand and finger-issued behaviors toward natural authentication," *IEEE Access*, vol. 8, pp. 55815–55825, 2020.

[10] H. Bi, Y. Sun, J. Liu, and L. Cao, "SmartEar: Rhythm-based tap authentication using earphone in information-centric wireless sensor network," *IEEE Internet Things J.*, vol. 9, no. 2, pp. 885–896, Jan. 2022.

[11] Z. Zheng et al., "Where are the dots: Hardening face authentication on smartphones with unforgeable eye movement patterns," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 1295–1308, 2023.

[12] S. Keykhaie and S. Pierre, "Lightweight and secure face-based active authentication for mobile users," *IEEE Trans. Mobile Comput.*, vol. 22, no. 3, pp. 1551–1565, Mar. 2023.

[13] S. Arun Kumar, R. Ramya, R. Rashika, and R. Renu, "A survey on graphical authentication system resisting shoulder surfing attack," in *Advances in Intelligent Systems and Computing*, N. N. Chiplunkar and T. Fukao, Eds., Singapore: Springer, 2021, pp. 761–770.

[14] B. Hettwer, S. Gehrer, and T. Güneysu, "Applications of machine learning techniques in side-channel attacks: A survey," *J. Cryptograph. Eng.*, vol. 10, no. 2, pp. 135–162, Jun. 2020.

[15] Y. Gu et al., "WiONE: One-shot learning for environment-robust device-free user authentication via commodity Wi-Fi in man–machine system," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 3, pp. 630–642, Jun. 2021.

[16] J. Liu, C. Xiao, K. Cui, J. Han, X. Xu, and K. Ren, "Behavior privacy preserving in RF sensing," *IEEE Trans. Dependable Secure Comput.*, vol. 20, no. 1, pp. 784–796, Jan. 2023.

[17] J. Zhang, Z. Chen, C. Luo, B. Wei, S. S. Kanhere, and J. Li, "MetaGanFi: Cross-domain unseen individual identification using WiFi signals," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 6, no. 3, pp. 1–21, Sep. 2022.

[18] C. Wu, F. Zhang, Y. Hu, and K. J. R. Liu, "GaitWay: Monitoring and recognizing gait speed through the walls," *IEEE Trans. Mob. Comput.*, vol. 20, no. 6, pp. 2186–2199, Jun. 2021.

[19] W. Jiang et al., "Towards environment independent device free human activity recognition," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, 2018, pp. 289–304.

[20] Y. Zhang et al., "Widar3.0: Zero-effort cross-domain gesture recognition with Wi-Fi," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 8671–8688, Nov. 2022.

[21] C. Shi, J. Liu, H. Liu, and Y. Chen, "Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT," in *Proc. 18th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.* New York, NY, USA: Association for Computing Machinery, Jul. 2017, pp. 1–10.

[22] H. Kong, L. Lu, J. Yu, Y. Chen, L. Kong, and M. Li, "FingerPass: Finger gesture-based continuous user authentication for smart homes using commodity WiFi," in *Proc. 20th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Jul. 2019, pp. 201–210.

[23] X. Wang, F. Li, Y. Xie, S. Yang, and Y. Wang, "Gait and respiration-based user identification using Wi-Fi signal," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 3509–3521, Mar. 2022.

[24] M. Shahzad and S. Zhang, "Augmenting user identification with WiFi based gesture recognition," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 1–27, Sep. 2018.
[25] R. Xiao, J. Liu, J. Han, and K. Ren, "OneFi: One-shot recognition for unseen gesture via cots WiFi," in *Proc. 19th ACM Conf. Embedded Netw. Sens. Syst.* New York, NY, USA: Association for Computing Machinery, 2021, pp. 206–219.
[26] Z. Wang, S. Chen, W. Yang, and Y. Xu, "Environment-independent Wi-Fi human activity recognition with adversarial network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 3330–3334.
[27] C. Li, M. Liu, and Z. Cao, "WiHF: Enable user identified gesture recognition with WiFi," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Jul. 2020, pp. 586–595.
[28] J. Liu, Y. Chen, Y. Dong, Y. Wang, T. Zhao, and Y.-D. Yao, "Continuous user verification via respiratory biometrics," in *Proc. IEEE Conf. Comput. Commun.*, Jul. 2020, pp. 1–10.
[29] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2015, pp. 65–76.
[30] L. Zhang, C. Wang, and D. Zhang, "Wi-PIGR: Path independent gait recognition with commodity Wi-Fi," *IEEE Trans. Mobile Comput.*, vol. 21, no. 9, pp. 3414–3427, Sep. 2022.
[31] F. Zhang et al., "From Fresnel diffraction model to fine-grained human respiration sensing with commodity Wi-Fi devices," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–23, Mar. 2018.
[32] Y. Zeng, P. H. Pathak, and P. Mohapatra, "WiWho: WiFi-based person identification in smart spaces," in *Proc. 15th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, Apr. 2016, pp. 1–12.
[33] C. Yang, Z. Wang, and S. Mao, "RFPose-GAN: Data augmentation for RFID based 3D human pose tracking," in *Proc. IEEE 12th Int. Conf. RFID Technol. Appl. (RFID-TA)*, Sep. 2022, pp. 138–141.
[34] Z. Wang, C. Yang, and S. Mao, "Data augmentation for RFID-based 3D human pose tracking," in *Proc. IEEE 96th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2022, pp. 1–2.
[35] M. Patel, X. Wang, and S. Mao, "Data augmentation with conditional GAN for automatic modulation classification," in *Proc. 2nd ACM Workshop Wireless Secur. Mach. Learn.*, Jul. 2020, pp. 31–36.
[36] C. Yang, X. Wang, and S. Mao, "TARF: Technology-agnostic RF sensing for human activity recognition," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 2, pp. 636–647, Feb. 2023.
[37] V. C. Chen, "The micro-doppler effect in radar," in *Electronics: A First Course*. Boston, MA, USA: Artech House, 2011, pp. 133–139.
[38] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[39] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, vol. 2, Jul. 2015, pp. 1180–1189.
[40] Z. Jiang et al., "Eliminating the barriers: Demystifying Wi-Fi baseband design and introducing the PicoScenes Wi-Fi sensing platform," *IEEE Internet Things J.*, vol. 9, no. 6, pp. 4476–4496, Mar. 2022.
[41] S. Cui, X. Jin, S. Wang, Y. He, and Q. Huang, "Heuristic domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 7571–7583.
[42] R. Gao et al., "Towards position-independent sensing for gesture recognition with Wi-Fi," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 2, pp. 1–28, Jun. 2021.
[43] R. Bohannon, "Comfortable and maximum walking speed of adults aged 20–79 years: Reference values and determinants," *Age Ageing*, vol. 26, pp. 9–15, Jan. 1997.
[44] S. Tan, L. Zhang, Z. Wang, and J. Yang, "MultiTrack: Multi-user tracking and activity recognition using commodity WiFi," in *Proc. CHI Conf.* New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–12.
[45] D. Wu et al., "WiTraj: Robust indoor motion tracking with WiFi signals," *IEEE Trans. Mobile Comput.*, vol. 22, no. 5, pp. 3062–3078, May 2023.
[46] J.-C. Su, S. Maji, and B. Hariharan, "When does self-supervision improve few-shot learning?" in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 12352, Glasgow, U.K., 2020, pp. 645–666.
[47] H. Huang, C. Wang, P. S. Yu, and C.-D. Wang, "Generative dual adversarial network for generalized zero-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 801–810.

Lei Zhang (Member, IEEE) received the Ph.D. degree in computer science from Auburn University, Auburn, AL, USA, in 2008. She was an Assistant Professor with the Computer Science Department, Frostburg State University, Frostburg, MD, USA, from 2008 to 2011. She is currently an Associate Professor with the College of Intelligence and Computing, Tianjin University, Tianjin, China. Her research interests include mobile computing and data mining. She is a member of ACM.

Yunzhe Jiang received the B.E. degree from Tianjin University, Tianjin, China, in 2022, where he is currently pursuing the master's degree. His research interests include wireless sensing and data mining.

Yazhou Ma received the B.E. degree from Shanxi Agricultural University, Shanxi, China, in 2021. He is currently pursuing the master's degree with Tianjin University, Tianjin, China. His research interests include wireless sensing and data mining.

Shiwen Mao (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Polytechnic University, Brooklyn, NY, USA, in 2004. He is currently a Professor and an Earle C. Williams Eminent Scholar of electrical and computer engineering with Auburn University, Auburn, AL, USA. His research interests include wireless networks, multimedia communications, and smart grids. He received the IEEE ComSoc TCCSR Distinguished Technical Achievement Award in 2019, Auburn University Creative Research and Scholarship Award in 2018, and the NSF CAREER Award in 2010. He was a co-recipient of the 2021 Best Paper Award of Elsevier/KeAi Digital Communications and Networks Journal, the 2021 IEEE Internet of Things Journal Best Paper Award, the 2021 IEEE Communications Society Outstanding Paper Award, the IEEE Vehicular Technology Society 2020 Jack Neubauer Memorial Award, the IEEE ComSoc MMTC 2018 Best Journal Paper Award, the 2017 Best Conference Paper Award, the Best Demo Award of IEEE INFOCOM 2022 and IEEE SECON 2017, the Best Paper Award of IEEE ICC 2022 and 2013, IEEE GLOBECOM 2019, 2016, and 2015, and IEEE WCNC 2015, and the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems. He was a Distinguished Lecturer and a Distinguished Speaker of the IEEE Vehicular Technology Society from 2014 to 2018 and from 2018 to 2021, respectively. He is a Distinguished Lecturer of the IEEE Communications Society and the IEEE Council of RFID. He serves on the Editorial Board of several journals, including IEEE TRANSACTIONS ON MOBILE COMPUTING.

Wenyuan Huang received the B.E. degree from Hainan University, Hainan, China, in 2020, and the M.D. degree from Tianjin University, Tianjin, China, in 2023. His research interests include wireless sensing and machine learning.

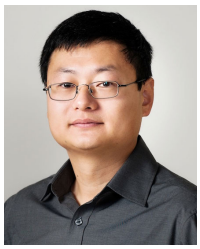
Zhiyong Yu (Member, IEEE) received the M.E. and Ph.D. degrees in computer science and technology from Northwestern Polytechnical University, Xi'an, China, in 2007 and 2011, respectively. He was a Visiting Student with Kyoto University, Kyoto, Japan, from 2007 to 2009, and a Visiting Researcher with the Institut Mines-Telecom, Telecom SudParis, Evry, France, from 2012 to 2013. He is currently a Full Professor with the College of Computer and Data Science, Fuzhou University, Fuzhou, China. His current research interests include pervasive computing, human–machine intelligence, and mobile crowdsensing.

Xiao Zheng (Member, IEEE) received the B.S. degree from Anhui University, China, the M.S. degree from Zhejiang University of Science and Technology, China, and the Ph.D. degree in computer science and technology from Southeast University, China. He is currently a Professor with the School of Computer Science and Technology, Anhui University of Technology (AHUT), China, and the Vice President of AHUT. Before that, he was the Dean of the School of Computer Science and Technology, AHUT. His research interests include the industrial Internet of Things, mobile cloud computing, and privacy-preserving computing. He is a Senior Member of CCF and a member of ACM.

Lin Shu (Senior Member, IEEE) received the B.E. degree in information engineering and the M.E. degree in communication and information system from South China University of Technology, Guangzhou, China, in 2005 and 2008, respectively, and the Ph.D. degree from the Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hong Kong, in 2012. He is currently a Professorate Senior Engineer with the School of Future Technology, South China University of Technology. His research interests include flexible sensors, wearable sensor systems, and virtual reality.

Xiaochen Fan (Member, IEEE) received the B.E. degree from Beijing Institute of Technology in 2013 and the Ph.D. degree from the University of Technology Sydney in 2021, respectively. He is currently a Post-Doctoral Researcher with the Institute for Electronics and Information Technology in Tianjin and the Department of Electronic Engineering, Tsinghua University. His research interests include mobile computing, urban science, and deep learning.

Guangquan Xu (Member, IEEE) is currently with the Department of Intelligence and Computing, Tianjin University, the Deputy Director of the Institute of Software and Information Security Engineering, the Director of China Information Security Evaluation Center (Tianjin) Branch Center–Tianjin University Network Security Joint Laboratory, Saining-Tianjin University Network Attack, and the Defense Joint Experiment Director. His interests include intelligent (smart society) social collaborative governance, research directions include trust management, network and information security, security privacy and trust, trusted computing, and artificial intelligence security.



Changyu Dong (Member, IEEE) received the Ph.D. degree from Imperial College London. He is currently a Professor with the Institute of Artificial Intelligence, Guangzhou University. He has authored over 70 publications in international journals and conferences. His research interests include applied cryptography, security of AI, data privacy, and security policies. His recent work focuses mostly on designing practical secure computation protocols with applications to large scale privacy preserving data processing. The application domains include for example privacy preserving machine learning, secure cloud computing, and privacy preserving data mining.