WiAU: An Accurate Device-free Authentication System with ResNet

Chi Lin*,†, Jiaye Hu*,†, Yu Sun*,†, Fenglong Ma†, Lei Wang*,†, and Guowei Wu*†
*School of Software Technology, Dalian University of Technology, Dalian, 116023, China
†Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian, 116621, China
‡Department of Computer Science and Engineering, University at Buffalo, NY, 14221, USA
Email: c.lin@dlut.edu.cn, hujiac@ieee.org, fenglong@buffalo.edu, lei.wang@ieee.org, wgwdut@dlut.edu.cn

Abstract—The ubiquitous and fine-grained features of WiFi signals make it promising for achieving device-free authentication. However, traditional methods suffer from drawbacks such as sensitivity to environmental dynamics, low accuracy, long delay, etc. In this paper, we introduce how to validate human identity using the ubiquitous WiFi signals. We develop WiAU, a device-free authentication system which only utilizes a Commodity Off-The-Shelf (COTS) router and a laptop. We describe the constitutions of WiAU and how it works in detail. Through collecting channel state information (CSI) profiles, WiAU automatically segments coherent activities and walking gait using an automatic segment algorithm (ASA). Then, a ResNet algorithm with two dedicated loss functions is designed to validate legal users and recognize illegal ones. Finally, experiments are conducted from different scenes to highlight the superiorities of WiAU in terms of high accuracy, short delay and robustness, revealing that WiAU has an accuracy of over 98% in recognizing human identity and human activities respectively.

I. INTRODUCTION

Authenticating human identity is a key technique in pervasive computing and human-computer interactions. To correctly validate human identity, many methods are developed in recent decades from different aspects, such as behavioral biometrics [1], fingerprint [2], gait [3], and face recognition [4]. Though these approaches indeed realize human authentication with high accuracy and low latency, they still have two fundamental limitations: the one is requiring extra specific devices (e.g., sensors, cameras, and RFID readers), and the other is requiring line of sight and destroying human privacy potentially.

To tackle the aforementioned limitations, the ubiquitous signal, WiFi, has already been used for authentication and identification [5]–[7]. To authenticate users through gait, Wang et al. [5] proposed WiFiU, which utilized principal components analysis (PCA) and support vector machine (SVM) to validate identities. Later, Xin et al. [6] developed an authentication approach called FreeSense, which applied both PCA and k-Nearest Neighbor (KNN). Shi et al. [7] utilized deep neural network (DNN) technique to identify human based on analysing features of activity variances. However, these state-of-the-art techniques still have some disadvantages as follows:

1) Low accuracy for illegal recognition. Previous methods [5], [6] are usually developed for authenticating legal users, but weak in judging the validity of a stranger.
2) Segmentation activities based on experience. Traditional schemes [6] mostly rely on a dynamic threshold when extracting human movements, which is assigned in advance based on researchers' experience. Sometimes, an improper assignment may degrade the recognition accuracy.
3) Unsuitable for environmental dynamics [7]. As WiFi signal is sensitive to environment and noise, a slight change in the deployment places may lead to catastrophic authentication failure [6]. Therefore, an authentication system with careful design in robustness for environmental dynamics is a necessity.
4) Long recognition time delay. Approaches in [5]–[7] spend a long time in authentication, which hinders their widespread popularization in practical applications.

In order to overcome these disadvantages and design an accurate authentication system, in this paper, we develop WiAU, a WiFi based device-free authentication system which utilizes a two-loss-function based Residual Network (ResNet) technique. The benefit of applying ResNet is that it can effectively obtain the accurate representative features of human activities and human walking gaits from the collected WiFi CSI data. To fully make use of these features, we design two loss functions: the one guarantees the human identification accuracy, and the other achieves a high judgment accuracy of illegal users. Based on the well trained model and the short-cut connection design of the ResNet, WiAU is able to recognize identity fast. Meanwhile, taking the advantages of a great self-learning ability of the ResNet, it processes transferring learning capability which is robust enough to alleviate environmental dynamics.

In summary, the contributions of our work are listed as follows:

1) We design a new device-free human authentication system, named WiAU, which is developed upon a two-loss-function based ResNet algorithm and able to accurately identify authorized users and recognize illegal users with an accuracy of 98% and 92%, respectively. Besides, comparing with the state-of-the-art WiFi CSI based recognition/authentication methods (e.g. FreeSense [6] and WiFiU [5]), WiAU has a shorter authentication delay.

978-1-5386-4281-8/18/$31.00 ©2018 IEEE
2) A novel partition algorithm (called ASA) is developed to discretize the continuous fluctuating WiFi signals caused by movements of people. The advantage of ASA is that it can automatically segment continuous signals, which dramatically simplifies the data collecting and preprocessing work.

3) Transfer learning technology is implemented to alleviate interferences caused by environment, making the proposed WiAU system feasible for popularizing in practical applications and robust against environmental dynamics.

The rest of this paper is organized as follows. Section II briefly reviews the state-of-the-art literature. Section III introduces related background in terms of WiFi CSI and ResNet. Section IV describes our system in detail. Section V details the key components in ResNet model. In Section VI, experiments are conducted to demonstrate the outperformed features of WiAU. Section VII concludes this paper and points out the future work.

II. RELATED WORK

In recent years, related methods are generally in two folds: activity recognition [8], [9] and identity recognition [5], [7], [10].

A. Gesture/Activity Recognition

Gesture recognition concentrates on WiFi signals reflected by human activities. When a gesture or activity is taken nearby WiFi devices, it will cause phase rotation and amplitude variation potentially. Pu et al. [9] proposed WiSee, which used USRP-N210s to recognize gestures. Wang et al. [11] developed the E-eyes system to monitor activities in the indoor environment. Ali et al. [12] implemented a keystroke recognition system, namely WiKey. It successfully distinguished keystroke behaviors with an accuracy of 96.4%. Melgarejo et al. [13] presented a recognition scheme for fine-grain gestures. They leveraged directional antenna and short-range wireless propagation properties to recognize a vocabulary of action-oriented gestures from the American Sign Language. Tan et al. [14] presented a fine-grained finger gesture recognition scheme named WiFinger using a single commodity WiFi device, and it can work with WiFi beacon signals and provide accurate gesture recognition under non-line-of-sight (NLOS) scenarios. Wang et al. [15] developed a WiHear system, which enabled WiFi signals to “hear” people talks without deploying any devices by detecting and analyzing fine-grained radio reflections from mouth micro-movements. Han et al. [16] employed the time variability and special diversity of CSI as the indicator of human activities and enabled automatic fall detection in their WiFall system.

B. Identity Recognition/Authentication

To achieve accurate device-free human authentication, Wang et al. [5] presented WiFiU, which utilized COTS WiFi devices to capture fine-grained gait patterns as indicators to recognize people. Shi et al. [7] extracted WiFi signal features from both walking and activities, and proved that human activities can be used to identify human. Zeng et al. [10] proposed a framework that can identify a person from a small group of people in a device-free manner using WiFi. They verified that walk gait can be detected and used to recognize a person’s identity.

To achieve a higher detection rate and wider sensing coverage in dense multi-path indoor scenarios, Zhou et al. [17] designed a lightweight subcarrier and path configuration scheme, which made full use of frequency and spatial diversities. Statistically, the detection rate reached 92.0%.

In general, the above mentioned methods all suffer from low recognition accuracy problem. In this paper, we develop a ResNet based device-free authentication system, namely WiAU. It has two sophisticated loss functions which guarantee higher accuracy with shorter delay. Moreover, WiAU is able to accurately recognize illegal users and is robust to environmental dynamics.

III. PRELIMINARIES

In this section, we briefly introduce the channel state information (CSI) and ResNet as background.

A. Channel State Information

WiFi has been evolving from the thought of providing network connectivity to all kinds of mobile and smart devices with higher speed and more advanced technologies. A prominent feature for using WiFi signal as a sensory medium is its convenience to obtain the channel state information (CSI) from the physical layer, which is actually the channel attribution of the communication link [18]. CSI describes the attenuation factor of the signal on each communication path, which is the value of the elements of channel gain matrix $H$, e.g. signal scattering, environmental fading and power decay with distance.

When sampling WiFi signal, each transmit-receive (TX-RX) antenna pair of a transmitter and receiver has 30 sub-carriers. Let $MTx$ and $MRx$ represent the number of transmitting and receiving antennas. Every time, $30 \times MTx \times MRx$ CSI streams can be timely collected, providing fundamental basis for our recognizing research.

B. CNN and ResNet

Convolution Neural Network (CNN) [19] is an easy training feed-forward neural network, which has been widely applied in image classification [20], image recognition, speech recognition and so on. It has several benefits: sharing weights, multiple layers, scalable to high-dimensional data, and so on.

Residual Network (ResNet) [21] has a great performance on training deep layer networks. Comparing with the plain network which piles convolution layers up, residual neural inserts shortcut connections, leading to a faster convergence on training networks.

In our work, the collected CSI streams are high dimensional and sensitive to environment, furthermore, they contain deeper level features for sensed movements. A straightforward method to extract such features is to use deep learning
methods (i.e. CNN or ResNet). However, simply piling up convolution layers suffers from problems of slow convergence, easy over-fitting, information missing, and so on [22]. As a result, we novelly combine CNN and ResNet together, where CNN downsamples original data and ResNet extracts the high-level features of activities and guarantees the integrity of information. With such combination, deep level features in CSI streams can be extracted accurately and classified automatically.

IV. SYSTEM DESIGN

To build an easily deployed and device-free solution for fine-grained human identity recognition, we develop a system that authenticates/identifies human via WiFi signals. First of all, we discuss the goals, system overview, and core components in detail.

A. Design Goals and Requirements

Our objective is to develop a device-free authentication system that accurately identifies people in a short delay with the ubiquitous WiFi signals. Our design, implementation and requirement of system involve a number of challenges:

- **Ubiquitous identification without additional equipment.** Our system should be easily deployed on existing COTS WiFi without introducing any dedicated or specialized hardware or requiring users to wear/carry additional devices. To highlight its ubiquity, it should only utilize the existing WiFi traffic or beacons at the deployed AP/router without dedicated user generated traffic.

- **Flexible and scalable for device-free recognizing/authenticating.** Our system should not be limited to recognition/authentication with only a single data type as gait. Other types of CSI data, such as gestures, movements, or a mixture of continuous behaviors should also be applicable, making such a system more flexible and scalable to practical applications.

- **Fast & accurate authentication with noise avoidance.** Our system should be able to authenticate users and recognize illegal people in a quick manner with a high accuracy. Moreover, when processing the collected WiFi signals, our system should be able to mitigate invariance and noise.

B. System Overview and Design

The basic idea of our system is to automatically collect and extract features of WiFi signals caused by dynamic behaviors of people, and then authenticate them with a high accuracy and low latency via deep learning technologies.

As shown in Figure 1, WiAU is made up of four key modules: data collection, CSI pool, preprocessing, and identification. First, continuous raw CSI streams are collected and then recorded into the CSI pool. Afterwards, a preprocessing module is developed to denoise, segment, select, and mark the featured CSI data. Next, authentication/recognition process is carried out, in which ResNet is used to handle preprocessed CSI data for accurate identity classification. A prominent feature in our model is two sophisticated loss functions: loss1 and loss2, are developed, through which WiAU performs well in terms of identity authentication, activity recognition, illegal recognition, and robust to environmental dynamics.

Obviously, preprocessing and identification modules are key components in WiAU. Next, we illustrate both of them in detail.

C. Preprocessing

As signals collected by our equipment contain interferences and noise, which cannot be directly used for extracting features before authenticating [14], an essential step is to preprocess the WiFi CSI data. Taking a deep investigation, such interferences are mainly caused by high frequency noise or disturbance from the open environment. Therefore, our first step is to utilize Butterworth low-pass filter for obtaining denoised CSI data (i.e. removing high frequency noise). Then we mainly focus on how to extract features from the remaining signals.

1) **Data Transformation:** Although the denoised CSI amplitude data can be obtained by filtering, it is still hard to segment or locate the exact variations caused by certain activities, we give a simple example here. As shown in Figure 2, we respectively depict the 2nd and 10th subcarrier variances collected from our devices.

We observe that, when nobody passes through, the corresponding CSI stream is an approximately stable line with a small variance and disturbance (see Figure 2(a)). Whenever someone walks through or a movement is detected, significant variations can be sensed by almost all sub-carriers (see Figure 2(b)).

Since it is hard to detect and segment these mixed variations, we should enlarge the partial difference in every small scale
Observation 1: Variations reflected by activities are larger than intervals. Moreover, intervals’ variances have distinguished features, i.e., they approximate to zero and always maintain stable.

Observation 2: Activities are caught by all subcarriers simultaneously.

2) Automatic Segmentation Algorithm (ASA): Observation 1 indicates that, if the distinguished feature of transformed data can be located, activity intervals can be accurately discovered.

We hereby define a dynamic threshold \( l \) to locate intervals so as to distinguish activities automatically. Firstly, we segment a transformed data stream \( Y_i \) evenly as:

\[
C_v = \{ y \in Y_i, y_1 + (v - 1) \cdot l < y \leq y_1 + v \cdot l \}, \tag{2}
\]

where \( v \in [1, \frac{\text{max}(Y_i) - \text{min}(Y_i)}{l}], y_1 = \text{min}(Y_i). \)

At this moment, CSI stream data has been roughly segmented into several sets based on value \( l \), and the first set \( C_1 \) represents the smallest elements which contain the activity intervals. According to observation 1, the influences caused by time intervals are stable and dense. In other words, \( l \) should be automatically adjusted until \( C_1 \) contains the most stable and smallest elements. Therefore, when splitting transformed data streams, two conditions should be satisfied simultaneously: 1) the first set \( C_1 \) contains the most number of elements and 2) values of the transformed data should be the most stable.

To find the accurate activity intervals, we need to design constraints as described in Equation (3). When Equation (3) is satisfied, \( C_1 \) will contain the accurate activity intervals and the corresponding value \( l \) will be the ideal dynamic threshold value.

\[
\begin{cases}
|C_1| \geq |C_v| \\
\sigma(C_1) \leq \sigma(C_v).
\end{cases} \tag{3}
\]

Here, \(|C_v|\) indicates the number of elements in set \( C_v \), \( \sigma(C_v) \) is the standard deviation of set \( C_v \). Finally, we record the activity intervals as:

\[
S : C_1 \rightarrow T_i, S(c) = t + w, \forall c \in C_1, t \in T_i, \tag{4}
\]

where \( S \) is a bijection, which maps the elements in \( C_1 \) to the corresponding time domain \( T_i \), \( t \) is the corresponding time of \( c \), and \( w \) is the moving window size. As only a small difference \( w \) exists in the transformed data \( Y_i \) and original data \( X_i \) on time domain (see Equation (1)), this mapping method can find out the time of activity intervals in the original denoised CSI data stream \( X_i \).

A distinct feature here is that there is no need to initialize the value of the threshold, because the value \( l \) obtained by ASA only depends on the CSI amplitude variation. Detailed descriptions are given below (see Algorithm 1).

3) Selecting Subcarriers: To reduce the residual noise, we aim to select a small fraction of representative subcarriers. According to observation 2, we combine the time of activity intervals on different subcarriers to get an accurate detection of the total activity intervals.

Accordingly, we perform the singular value decomposition (SVD) on transformed data, which represents the magnitude degree of the waveform:

\[
\Sigma = T(Y \times (Y^\top)), \tag{5}
\]
Algorithm 1 Automatic Segmentation Algorithm (ASA)

1: **Input:** Transformed data matrix $Y$, a moving window $w$.  
2: **Output:** Segmented activity interval set $C_1$ and its corresponding time set $T_i$.  
3: **for** $\forall Y_i \in Y$ **do**  
4: Use variable $l$ to distinguish $Y_i$ into several sets $C_v$ as Equation (2).  
5: Calculate values of $|C_v|$ and $\sigma (C_v)$ as Equation (3).  
6: $l \leftarrow 0$.  
7: **while** $|C_1| < |C_v|$ or $\delta (C_1) > \delta (C_v)$ **do**  
8: Set a sufficient small variable $\Delta$.  
9: $l \leftarrow l + \Delta$.  
10: **end while**  
11: **end for**  
12: $S : C_1 \rightarrow T_i$ as Equation (4).  
13: **Return:** Set $T_i$.

where $Y$ is the transformed data matrix, $T$ denotes the eigenvalue of matrix $Y \times (Y^T)$, and each transformed data stream $Y_i$ has its corresponding eigenvalue $\Sigma_i$. The sum of $\Sigma$ indicates the total energy of this matrix. If there exits a subset $Y'$ constructed by some CSI streams, and their corresponding eigenvalue set $\Sigma' \subset \Sigma$ can be found while satisfying:

$$\sum (\Sigma') = \mu \times \sum (\Sigma),$$

these CSI streams $Y'$ can represent more than half of total streams. Here, $\sum (\Sigma')$ indicates the energy of the selected streams, which should be able to represent over half of the CSI streams, consequently, we set $\mu \geq 50\%$.

4) **Overall Process:** To process all of the CSI amplitude data streams, we combine above operations together and find a union set of the activity intervals in each stream. Then, it can represent the total detected activity intervals of a CSI amplitude data.

We use Algorithm 1 to automatically segment each selected CSI stream and calculate the union set of these segmented time interval sets. Then, we have:

$$U = \cup_{k=1}^{\lfloor Y' \rfloor} T_k, \forall Y_k \in Y',$$

where $Y_k$ denotes the selected CSI stream and $T_k$ indicates their corresponding time in activity intervals, $\lfloor Y' \rfloor$ denotes the number of selected CSI streams and $U$ indicates the union time of detected time in activity intervals of each data streams.

After obtaining a set $U$ containing interval time for the amplitude data, we preserve all CSI streams and highlight the variances which reflect human activities. We hereby assign values for each denoised CSI amplitude stream $X_i$ as:

$$X_i (m) = 0, \forall m \in U.$$

The above operations have two advantages: 1) it removes the majority of the residual noise and 2) it generates a sparse matrix through filling 0’s, which are suitable input for ResNet algorithm in the future calculations.

### V. Methodology in Authentication Process

In this section, we introduce the recognition/authentication model in detail. We utilize ResNet to extract various levels of features and learn a better presentation of them. It is composed of three modules: convolution module, global average pooling and softmax module, and loss function module (see Figure 4).

After using the global average pooling and softmax module, features are transformed into the predicted labels. To satisfy application requirements in Section IV-A, we design different loss functions to accomplish the recognition tasks (see Algorithm 2).

#### A. Convolution Module

The convolution module used here is to extract the high-level features of the preprocessed CSI data. It has two parts: CNN units and ResNet units (see Section III-B). CNN units downsample the original data and ResNet units extract deeper level features.

Originally, our collected raw CSI data have different lengths. After preprocessing, they are converted into a sparse matrix with the same length.

The input matrix $X$ is denoted as $X = (X_1, X_2, \ldots, X_i, \ldots, X_n)$, where $n$ is the number of all subcarriers between transmitter and receiver pairs, $X \in \mathbb{R}^{o \times n}$, where $o$ denotes the length of each input $X_i$. Then we input $X$ into ResNet module to generate a new feature vector as:

$$R = \text{Res}(X), R \in \mathbb{R}^{k \times n}.$$  

Here, $\text{Res}()$ is the combination of CNN units and ResNet units. After executing the convolution module, the dimension of feature vector can be reduced to $k$. To downsample and tailor data for later ResNet calculations, each convolution unit is designed to contain a convolution layer and a max-pooling layer, as shown in Table I.

<table>
<thead>
<tr>
<th>FILTER</th>
<th>STRIDE</th>
<th>FORMAT</th>
<th>FORMAT</th>
<th>FORMAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv.</td>
<td>Pooling</td>
<td>Block 1</td>
<td>Block 2</td>
<td>Block 3</td>
</tr>
<tr>
<td>7 x 7</td>
<td>3 x 3</td>
<td>$1 \times 1$</td>
<td>32</td>
<td>$1 \times 1$</td>
</tr>
<tr>
<td>3 x 3</td>
<td>3 x 3</td>
<td>$1 \times 1$</td>
<td>128</td>
<td>$1 \times 1$</td>
</tr>
<tr>
<td>1 x 1</td>
<td>1 x 1</td>
<td>$1 \times 1$</td>
<td>128</td>
<td>$1 \times 1$</td>
</tr>
<tr>
<td>1 x 1</td>
<td>1 x 1</td>
<td>$1 \times 1$</td>
<td>128</td>
<td>$1 \times 1$</td>
</tr>
<tr>
<td>1 x 1</td>
<td>1 x 1</td>
<td>$1 \times 1$</td>
<td>128</td>
<td>$1 \times 1$</td>
</tr>
</tbody>
</table>

A convolution module has 1 CNN layer and 15 ResNet layers. The kernel of CNN is $7 \times 7$, and convolution layers in ResNet have kernels of $1 \times 1$ and $3 \times 3$, respectively. It can not only achieve a high accuracy of recognition/authentication, but also guarantee a fast recognition process.

#### B. Global Average Pooling and Softmax Module

Since the dimensions of the output feature maps are still large, global average pooling is used to reduce the dimension. We firstly calculate global average on the feature vector, so that each feature vector obtains an output as:

$$A = \text{Avg}(R), A \in \mathbb{R}^{d \times n},$$  

$$h = [A_1, A_2, \ldots, A_1, \ldots, A_n].$$

### TABLE I

**STRUCTURE OF THE CONVOLUTION MODULE**
Here, $A_i$ is the average value of each feature map $R_i$, $h$ represents their union, and $d$ indicates the classes of input $X$. Then, we use softmax regression as a classifier to assign probabilities to each input:

$$\hat{y} = \text{softmax}(W^T \times h + b).$$

(11)

Here, $W \in \mathbb{R}^{d \times n}$ is the weight matrix and $b \in \mathbb{R}^{d \times 1}$ means the bias vector. The softmax function is given by:

$$\text{softmax}(z)_{\beta} = \frac{e^{z_{\beta}}}{\sum_{\alpha} e^{z_{\alpha}}},$$

(12)

where $e^{z_{\beta}}$ is the probability of the $\beta$th category, $\sum_{\alpha} e^{z_{\alpha}}$ represents the overall probability of all categories.

C. Loss Function Module

The functionality of loss function module is to evaluate the differences between expected results and actual ones. In our design, we respectively develop two loss functions to achieve two goals: 1) recognizing the identity of legal users and 2) distinguishing the illegal users.

To achieve the first goal, we utilize cross-entropy, see Equation (13). In the meanwhile, to prevent overfitting, we perform the $L_2$ regular operator:

$$\text{Loss}_1 = -\frac{1}{d} \sum_{m=1}^{d} y_{m} \log (\hat{y}_{m}) + \lambda \Omega,$$

(13)

where $\Omega$ is the $L_2$ regularizer and $\lambda$ is the regularization parameter. $y_m$ is the true label of input and the $\hat{y}_m$ is the expected value.

To achieve the second goal, we divide the training data and label them into two classes. We randomly choose a percentage $p \in [0, 1]$ of the users labeled as the illegal class, and others as legal. Then, the second loss is designed as:

$$\text{Loss}_2 = \frac{1}{d} \sum_{m=1}^{d} \max(0, 1)(y_m - \hat{y}_m + \frac{\hat{y}_m}{p \times d}),$$

(14)

where $\hat{y}_m$ means the predicted label and $y_m$ indicates the true label. $\mathbb{I}(y_m)$ is an indicator function which indicates whether the (il)legal user is labeled as (il)legal class, and defined as follows:

$$\mathbb{I}(y_m) = \begin{cases} 1, & \text{if } y_m \text{ is (il)legal and its label is (il)legal;} \\ 0, & \text{otherwise}. \end{cases}$$

(15)

Then we describe the detailed authentication/identification process in Algorithm 2. When we intend to authenticate legal people ($p = 0$), loss 1 is used to calculate the difference between the true label and predicted one. Otherwise, besides loss 1, we additionally use loss2 to make a two-class prediction.

Algorithm 2 Double Loss Algorithm (DLA)

1: **Input**: CSI raw data $X$, CSI label $y_m$, illegal volunteer percentage $p$.
2: **Output**: Loss value.
3: $y_m \leftarrow \text{softmax}(\text{Avg}(\text{Res}(X_i)))$.
4: if $p = 0$ then
5: $\text{Loss} \leftarrow -\frac{1}{d} \sum_{m=1}^{d} y_m \log (\hat{y}_m) + \lambda \Omega$.
6: else
7: $\text{Loss}_1 \leftarrow -\frac{1}{d} \sum_{m=1}^{d} y_m \log (\hat{y}_m) + \lambda \Omega$.
8: $\text{Loss}_2 \leftarrow \frac{1}{d} \sum_{m=1}^{d} \max(0, 1)(y_m - \hat{y}_m + \frac{\hat{y}_m}{p \times d})$.
9: **end if**
10: **Return** Loss value.

D. Transfer Learning

To guarantee the robustness of system against environmental dynamics, the transfer learning capability is also our concern. Given two similar recognition tasks, $T_1$ and $T_2$, our solution is to tune the model of task $T_2$ based on parameters trained by $T_1$ instead of initializing new parameters.

Suppose we have two similar learning tasks: $T_1$ and $T_2$, e.g. identifying the same people through data collected from different floors. After training task $T_1$, we can get parameters $\theta(T_1)$ satisfying:

$$\theta(T_1) = \arg\min_{\theta \in \Theta} \sum_{i} \text{Loss}_1(C_i^{(T_1)}, \theta).$$

(16)

Here, $\Theta$ and $C_i^{(T_1)}$ respectively indicate the parameter domain and input data of $T_1$ and $\text{Loss}_1$ is the loss function. Then, to train the task $T_2$, we do not need to initialize parameters but directly tune the parameters obtained from $\theta(T_1)$ as:

$$\theta(T_2) = \arg\min_{\theta \in \Theta} \sum_{i} \text{Loss}_1(C_i^{(T_2)}, \theta(T_1) + \theta),$$

(17)

where $\theta(T_1) + \theta$ means task $T_1$ is the training base and $C_i^{(T_2)}$ indicates a small part of the data collected from task $T_2$.

Consequently, WiAU owns the transfer learning ability, it can deal with environmental changes (i.e. it possesses a high accuracy when completing task $T_2$). Corresponding verifications are given in the Section VI-D1.
VI. EXPERIMENTS

To demonstrate the performance of our scheme, experiments are conducted.

A. Experiment Setup

Our WiAU system consists of two COTS devices: a TP-Link TL WR886N router for continuously emitting WiFi signals and a common laptop equipped with Intel5300 network adapter, which has three antennas for receiving signals. WiAU is developed based on Linux 802.11n CSI tool [18] running on the Ubuntu 12.04 operating system.

In WiAU, each transmit-receive (TX-RX) antenna pair has 30 sub-carriers. Therefore, 90 continuous CSI streams can be received all the time. To display CSI streams in a real-time manner on the screen, we develop a visualization platform using python. The transmitting frequency here is 0.03 second, whenever our laptop receives a packet, a new CSI curve will be displayed. We aggregate and select all data of subcarriers as the training dataset and record into the CSI pool. When processing data, a PC is used which is configured with CPU: Intel Core i5-6500, GPU: NVIDIA GeForce GTX 750 Ti, RAM: DDR4 8Gb×3.

To demonstrate the advantages of WiAU, two experimental scenes, an office and three corridors from different floors are selected as shown in Figure 5. A number of volunteer students are enrolled in the experiments. A distinct feature of our experiment is that, two volunteers, labeled as V6 and V7, are twins, because we intend to testify the capability of WiAU in distinguishing them although their appearance are nearly the same.

Besides, we also demonstrate WiAU’s capability in identifying activities, we collected 16 different activities (see Table II) to construct the CSI pool. Moreover, two state-of-the-art recognition algorithms: FreeSense [6] and WiFiU [7] are implemented as baselines for comparison.

B. Identity Authentication with Continuous Activities

At first, we test the authentication performance through dealing with continuous movement signals taken in the office. After that, we further exploit its capability when confronting with a stranger or illegal person (i.e. whose CSI data are not recorded in the CSI pool).

1) Authenticate Legal People: In this experiment, 12 volunteers (9 male and 3 female students), labeled as $P_1, P_2, \ldots, P_{12}$, are required to complete two tasks in the office for constructing the CSI pool: (1) work on the desk (i.e. $1 \rightarrow 2 \rightarrow 3 \rightarrow 2 \rightarrow 1$ in Figure 5(a)) and (2) open the window (i.e. $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1$ in Figure 5(a)). In this process, we do not ask them to make specified gestures or walk in deterministic route as FreeSense, because WiAU is able to segment continuous activities into pieces, and then authenticate based on them.

We respectively select training data, validation data and testing data with the ratio of 60%, 20%, and 20% from the CSI pool. The recognition accuracy is calculated as:

$$\eta = \frac{n_{\text{recognized}}}{n_{\text{actual}}}$$

where $n_{\text{actual}}$ represents the number of labels referring to the target people and $n_{\text{recognized}}$ indicates the number of CSI streams correctly matching with labels.

The authentication confusion matrix are shown in Figure 6(a). We observe that, the average accuracy is over 98%, indicating that WiAU can precisely recognize authorized people. Another feature appealing us is its ability in distinguishing twin brothers. As shown in 6th and 7th columns, WiAU is able to successfully identify them with an accuracy of 100%.

Then we compare WiAU with baseline algorithms. As shown in Figure 7(a), we observe that, WiAU outperforms others with the highest accuracy. FreeSense’s accuracy is much lower than that of WiAU, the reason is that, through using ResNet, WiAU is able to extract deeper features from WiFi streams, which are absent in FreeSense and WiFiU.

Then, to demonstrate WiAU fits environmental changes, similar experiment is taken in the corridor with 14 volunteers (see Figure 5(b)). The only difference here is, at this time, only the gait CSI streams are collected and used for authentication. The confusion matrix and comparison results are respectively shown in Figure 6(b) and Figure 7(b). Consequently, we get identical conclusions as before.

2) Identify Illegal People: As mentioned in our requirements in Section IV, the recognition system should also be capable of distinguishing strangers or illegal people. We hereby conduct an experiment in our office and the identifying results are presented in Figure 7(c).

We randomly mark a fraction of volunteers as legal users and others as strangers. The comparison results among three
methods are shown in Figure 8(a), we note that WiAU is more accurate than others. The reason is that, during the loop iteration process of minimizing the loss function, the differences between illegal and legal people are enlarged, making distinguishing illegal people more easily and precisely.

C. Activity Recognition

Another feature of WiAU is it can recognize activities as we expected in Section IV. We collect CSI data of 16 activities (i.e. marked as $A_1$, $A_2$, ..., $A_{16}$, see Table II) from volunteers and show the recognition results in Figure 6(c). We note that, WiAU can successfully distinguish activity with an over 90\% accuracy, indicating that WiAU is flexible and scalable to be extended into recognition application. We also compare such performance with two baselines. As depicted in Figure 7(c), WiAU’s accuracy is higher than the others by at least 10\%. The accuracy of WiFiU is smaller than 50\%, because it cannot precisely fragment continuous CSI stream. FreeSense outperforms a little better than WiFiU, but still cannot surpass WiAU.

D. Other Characteristics

Besides authentication and recognition, WiAU still has some other features deserved to be deeply explored.

1) Transfer Learning Ability: As a special requirement of our system, the proposed system should have the transfer learning ability, which indicates the robustness when dealing with environmental changes.
We hereby respectively installed WiAU in the corridors on three different floors (2nd, 3rd, and 4th floor in Figure 5). 14 volunteers are requested to walk through the LOS monitoring area. To show its robustness, we collected data from 2nd and 3rd floor as the training set, and testify using data from 4th floor. The recognition results are shown in Figure 8(b), we note that, WiAU achieves an accuracy over 95%, which indicates the transfer learning module can mitigate environmental changes, demonstrating well performance in scalability and flexibility. We can conclude that, no matter where WiAU is deployed, its accuracy can always be guaranteed.

E. Recognition Latency

At last, we compare the recognition speed in Table III. We observe that, recognition/authentication in WiAU can be completed within 9 seconds, the quickest of which from from three algorithms. The reason is that ResNet technique can load the recorded model for recognizing new subjects, preserving coherent relations and spatial locality for data.

<table>
<thead>
<tr>
<th>TABLE III</th>
<th>COMPARISON OF RECOGNITION DELAY AMONG ALGORITHMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>WiFiU</td>
<td>FreeSense</td>
</tr>
<tr>
<td>Authentication in the office</td>
<td>13.2s</td>
</tr>
<tr>
<td>Activity recognition</td>
<td>10.8s</td>
</tr>
<tr>
<td>Authentication in the corridor</td>
<td>11.59s</td>
</tr>
</tbody>
</table>

VII. CONCLUSIONS AND FUTURE WORKS

In this paper, we design and implement WiAU, an accurate device-free authentication system based on ResNet. First, we deeply demonstrate the constitutions of WiAU and its working process. Then, a ResNet based authentication algorithm with two dedicated loss functions is designed to validate legal users and recognize illegal ones. At last, experiments are conducted to demonstrate the advantages of WiAU. Results indicate that WiAU can achieve fast and accurate authentication in terms of recognizing human identity and human activities.

As part of our future works, we will improve the performance of WiAU and extend it to authenticate multiple users simultaneously.

ACKNOWLEDGMENTS

We thank Linlin Guo, Binxian Lu and other anonymous reviewers for their advice. This research is sponsored in part by the National Natural Science Foundation of China (61402078, 61602080, 61502071, 61772113, 61733002, 61432003, 61720106005).

REFERENCES


