IEEE INFOCOM 2022 - IEEE Conference on Computer Communications | 978-1-6654-5822-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/INFOCOM48880.2022.9796740

Push the Limit of WiFi-based User Authentication towards Undefined Gestures

Hao Kong[*], Li Lu[†], Jiadi Yu[*§], Yanmin Zhu[*], Feilong Tang[*], Yi-Chao Chen[*], Linghe Kong[*], and Feng Lyu[‡]
[*]Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, P.R.China
Email: {hao.kong, jiadiyu, yzhu, tang-fl}@sjtu.edu.cn, yichao@utexas.edu, linghe.kong@sjtu.edu.cn
[†]School of Cyber Science and Technology and Key Laboratory of Blockchain
and Cyberspace Governance of Zhejiang Province, Zhejiang University, Hangzhou, P.R.China
Email: li.lu@zju.edu.cn
[‡]School of Computer Science and Engineering, Central South University, Changsha, P.R.China
Email: fenglyu@csu.edu.cn
[§]Coppresonding Author

***Abstract*—With the development of smart indoor environments, user authentication becomes an essential mechanism to support various secure accesses. Although recent studies have shown initial success on authenticating users with human activities or gestures using WiFi, they rely on predefined body gestures and perform poorly when meeting undefined body gestures. This work aims to enable WiFi-based user authentication with undefined body gestures rather than only predefined body gestures, i.e., realizing a gesture-independent user authentication. In this paper, we first explore physiological characteristics underlying body gestures, and find that statistical distributions under WiFi signals induced by body gestures can exhibit invariant individual uniqueness unrelated to specific body gestures. Inspired by this observation, we propose a user authentication system, which utilizes WiFi signals to identify individuals in a gesture-independent manner. Specifically, we design an adversarial learning-based model, which suppresses specific gesture characteristics, and extracts invariant individual uniqueness unrelated to specific body gestures, to authenticate users in a gesture-independent manner. Extensive experiments in indoor environments show that the proposed system is feasible and effective in gesture-independent user authentication.**

***Index Terms*—User authentication, gesture independence, WiFi signals, adversarial learning**

I. Introduction

Recent years have witnessed the surge of user authentication deploying in various infrastructures, including typical, mobile, and Internet of Things (IoT) devices, to provide critical guard for user privacy. Traditional authentication approaches either depend on knowledge (e.g., password and PIN), or rely on inborn biometric uniqueness of human (e.g., fingerprint and facial information), both of which require extra interactions to interrupt ongoing operations. Recently, the behavior-based user authentication, i.e., verifying the identity of a person from human daily activities or gestures, attracts more attentions. Such an authentication balances the trade-off between security requirements and non-intrusive user experiences. For example, allowing legitimate users to freely access privacy information along with current activities, or preventing unauthorized users from stealing confidential documents with malicious actions.

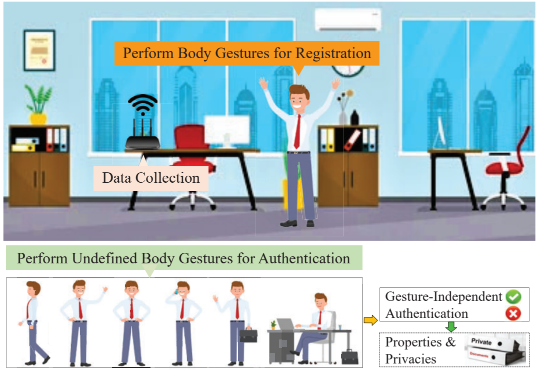

Perform Body Gestures for Registration
Data Collection
Perform Undefined Body Gestures for Authentication
Gesture-Independent Authentication
Properties & Privacies
Private

Fig. 1: System Illustration.

To achieve user authentication with human daily activities or gestures, some works [1], [2] utilize wearable devices to extract user behavioral features, which requires users' involvements and induces additional costs. To achieve non-intrusive authentication, other researches [3], [4] investigate vision to realize behavior-based authentication. But they present a similar problem to other visual applications, i.e., they depend on lighting conditions and also raise privacy concerns. To address these, recent works [5]–[8] exploit widely-existed WiFi signals to sense specific daily activities or gestures for authentication.

However, these approaches rely on predefined activities and gestures, i.e., they must be previously learned in the registration process. In practical scenarios, if user authentication can be carried out in any body gestures (i.e., activities and gestures), it is able to support security protections for a wide range of real-world situations. For example, in an IoT environment, to meet higher security requirements, a safety guard should authenticate a user whenever the user performs an arbitrary body gesture (such as daily activity, human-computer interaction, etc.), so as to provide real-time secure access for the IoT environment. Also, in the current COVID-19 epidemic [9], for a comprehensive and safe epidemiological investigation, a monitoring system should continuously track a person's identity through the person's any potential behaviors in the indoor environment. Towards this end, our goal is to realize user authentication under not only defined body gestures but also undefined body gestures, i.e., realizing gesture-independent user authentication.

Due to the wide deployment of WiFi infrastructures and the contact-free manner of WiFi-based sensing, we consider leveraging WiFi signals to achieve user authentication. To realize such an authentication system, we face several challenges in practice. First, we need to extract fine-grained features caused by human body gestures from commodity WiFi signals. Second, we need to characterize the invariant individual uniqueness that is in depth embedded underlying various body gestures. Finally, we should accurately identify individuals without the restriction of predefined body gestures.

In this paper, we first investigate the nature of human body gestures, and find that inborn physiological characteristics have a significant impact on body gestures, which induce invariant individual uniqueness independent of specific body gestures (i.e., unique physiological characteristics underlying various body gestures). To extract the invariant individual uniqueness, we further explore statistical distributions under Channel State Information (CSI) of WiFi signals induced by body gestures, and observe that different individuals exhibit individual differences in statistical distributions underlying various body gestures. Based on the observation, we propose a user authentication system, $FreeAuth$, which identifies individuals in a gesture-independent manner, i.e., the authentication does not rely on predefined gestures. Specifically, we design an adversarial learning-based model to enable the gesture-independent authentication. $FreeAuth$ first utilizes Convolutional Neural Network (CNN) to extract fine-grained features from CSI of WiFi signals induced by body gestures. Based on the extracted features, $FreeAuth$ employs Recurrent Neural Network (RNN) to extract specific gesture characteristics through sequential relationships under CSI sequences, which aims to suppress behavioral interferences of body gestures. Meanwhile, $FreeAuth$ uses Gaussian Mixture Model (GMM) to characterize individual uniqueness unrelated to specific body gestures through statistical distributions under CSI sequences, which aims to enhance the capability of extracting unique physiological characteristics. Through optimizing the two opposed objectives with adversarial learning, i.e., minimizing the specific gesture characteristics extracted from RNN and maximizing the individual uniqueness characterized from GMM, $FreeAuth$ can identify individuals in a gesture-independent manner. We evaluate the performance of the proposed system in real indoor environments, and the results show that $FreeAuth$ effectively authenticates users in a gesture-independent manner. Fig. 1 illustrates a typical scenario of the system.

We highlight our contributions as follows.

- We explore the physiological characteristics underlying human body gestures, and observe that different individuals exhibit individual differences in statistical distributions underlying various body gestures.
- We propose a user authentication system, $FreeAuth$, which can identify individuals in a gesture-independent manner using WiFi signals. To the best of our knowledge, this work is the first research to enable gesture-independent authentication for indoor environments.
- We design an adversarial learning-based model, which can suppress behavioral interferences of body gestures and extract invariant individual uniqueness unrelated to specific body gestures for gesture-independent authentication.
- We evaluate the performance of the system in real environments, and the results show that $FreeAuth$ can effectively identify users in a gesture-independent manner.

II. Preliminary

In this section, we present the insight of invariant individual uniqueness underlying body gestures, and then explore the feasibility of leveraging WiFi signals to extract invariant individual uniqueness for gesture-independent authentication.

A. Insight of invariant Individual Uniqueness Underlying Body Gestures

To explore the insight of invariant individual uniqueness underlying body gestures, we first analyze the generative process of a body gesture. Usually, a body gesture is performed by a person's limbs and torso in a manner that suits the person's physiology. Hence, human body gestures are always constrained by human physiological characteristics (e.g., the length of limbs, the power generated by limb movements). These physiological characteristics intrinsically induce the behavioral uniqueness for different people. For example, people with different muscle masses perform gestures in different accelerations and velocities, resulting in their behavioral uniquenesses. Hence, the behavioral uniqueness is determined by the distinct physiological characteristics of each person. An existing work [10] demonstrates that human body gestures are shaped by the inborn physiological characteristics, which encourages us to extract unique physiological characteristics underlying body gestures for authentication.

Different from the extrinsic behavioral characteristics, the intrinsic physiological characteristics are gesture-independent, i.e., such features remain static for a specific individual regardless of body gesture kind. The physiological characteristics relate more to the inborn physical and biochemical functions of people [11], so they hardly change in different body gestures, which induce the invariant individual uniqueness. Towards this end, we are motivated to investigate the feasibility of extracting unique physiological characteristics underlying various body gestures to realize gesture-independent user authentication.

B. Feasibility Study of Gesture-Independent Individual Identification Using WiFi

To study the feasibility of gesture-independent user authentication, we conduct an experiment to extract unique physiological characteristics underlying body gestures using WiFi signals. The Channel State Information (CSI) of WiFi signals describes the channel properties of propagation paths, which provides fine-grained sensing information of human movement [12]. Thus, we employ WiFi signals to conduct the experiment and analyze the CSI induced by body gestures. In the experiment, two participants perform three body gestures,

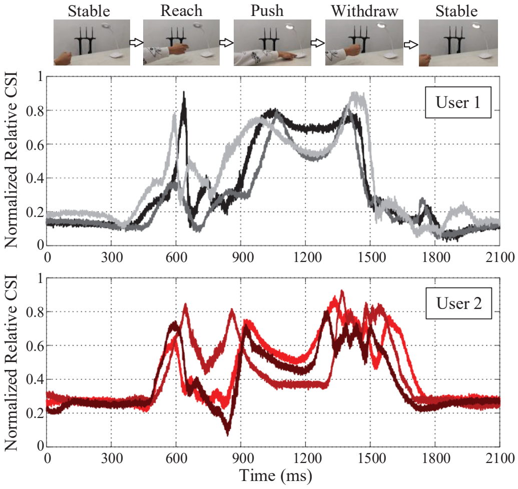
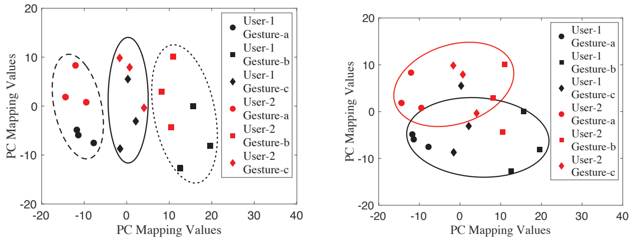


Fig. 2: Illustration of normalized relative CSI for two users.



(a) Distinguish body gestures.

(b) Distinguish users.

Fig. 3: Distributions of the three gestures performed by the two users in two principal components.

i.e., turning on a light, fetching a cup, and plug a power adapter, respectively. We employ a laptop equipped with an Intel WiFi Link 5300 NIC and 3 external receive antennas to capture WiFi signals. Fig. 2 illustrates the normalized relative CSI phase induced by a body gesture of two users respectively. The relative CSI is derived by the CSI from consecutive antennas, which is described in Section III-B. It can be observed from Fig. 2 that the signal patterns of the two users are different when they perform the same body gesture. Such differences result from unique behavioral characteristics of each user, demonstrating an effective and robust WiFi-based gesture-dependent user authentication similar to previous studies. Similarly, the other two gestures of fetching a cup and plug a power adapter also exhibit such differences caused by behavioral uniqueness of the users.

However, from Fig. 2, it is difficult to directly observe individual differences of physiological characteristics underlying body gestures from CSI of WiFi signals. In order to extract the unique physiological characteristics of each individual for gesture-independent authentication, inspired by text-independent speaker verification [13], we ignore the temporal sequences of CSI, and model the underlying statistical distributions of physiological observations to extract invariant uniqueness of each individual. To model the statistical distributions from CSI induced by body gestures, we employ the Principle Component Analysis (PCA) method, which derives the correlations between different CSI sequences and exhibits the principal components with minimum correlation for eliminating redundant information. Such a method can exhibit the statistical distributions under CSI sequences. Fig. 3 shows the distributions of two principal components from CSI induced by two users performing the three different gestures. In the figure, the x-axis is the first dominant component of PCA results, indicating the major movement information of each gesture. The y-axis is a specially screened component of PCA results. It can be observed from Fig. 3(a) that the three different body gestures are distinctly separated, which supports the accurate gesture recognition using WiFi. This result is consistent with existing researches of WiFi-based gesture recognition. Moreover, from another perspective shown in Fig. 3(b), we can observe that although performing different body gestures, the two users are still roughly distinguishable. The result demonstrates that the invariant individual uniqueness exists in statistical distributions underlying the CSI induced by body gestures. With the encouraging experiment result, we are motivated to design a gesture-independent user authentication using commodity WiFi.

III. System Design

In this section, we present the design details of the gesture-independent user authentication system, $FreeAuth$.

A. System Overview

Fig. 4 shows the system architecture, which is divided into a model construction stage and a user authentication stage.

In the model construction stage, $FreeAuth$ requires users to perform several body gestures, and employs WiFi signals to sense the body gestures as training data for one-off model construction. $FreeAuth$ first preprocesses the received signals through calculating relative CSI, and then segments the signals into episodes of each body gesture based on the changing rate of relative CSI. Then, $FreeAuth$ constructs an adversarial neural network, including a Convolutional Neural Network (CNN)-based feature extractor, a Recurrent Neural Network (RNN)-based gesture suppressor, and a Gaussian Mixture Model (GMM)-based user authenticator. In the adversarial neural network, the feature extractor extracts fine-grained features from the input signals, and the features are fed into the gesture suppressor and user authenticator respectively. After that, the gesture suppressor extracts specific gesture characteristics, while the user authenticator characterizes individual uniqueness unrelated to specific body gestures. Through training the neural network in an adversarial learning way, $FreeAuth$ finally obtains a trained feature extractor which can extract features independent of specific body gestures, and a trained user authenticator which is able to identify individuals through the extracted features.

In the user authentication stage, $FreeAuth$ authenticates a user based on the body gestures performed the user. $FreeAuth$ first preprocesses and segments the CSI of received WiFi signals induced by the user's body gestures, which is the same as that in the model construction stage. Then, $FreeAuth$ uses the trained feature extractor to extract features from the

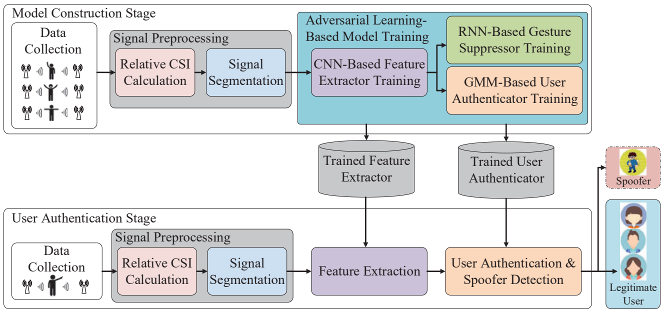

Model Construction Stage
Data Collection
Signal Preprocessing
Relative CSI Calculation
Signal Segmentation
Adversarial Learning-Based Model Training
CNN-Based Feature Extractor Training
RNN-Based Gesture Suppressor Training
GMM-Based User Authenticator Training
Trained Feature Extractor
Trained User Authenticator
Spoofer
User Authentication Stage
Data Collection
Signal Preprocessing
Relative CSI Calculation
Signal Segmentation
Feature Extraction
User Authentication & Spoofer Detection
Legitimate User

Fig. 4: System architecture of $FreeAuth$.

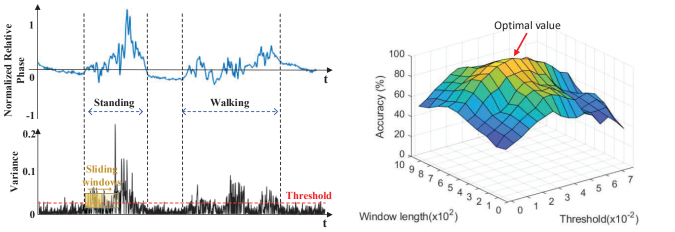
body gesture performed by the user, and further applies the trained user authenticator to identify the user.

B. Signal Preprocessing

$FreeAuth$ first preprocesses raw CSI by calculating relative CSI and segmenting body gesture sequences.

Relative CSI Calculation. Since commodity WiFi has a centimeter-level wavelength (e.g., $5.17cm$ for $5.8GHz$ band), CSI can capture centimeter-level movement of human. However, the errors in raw CSI of WiFi signals hinder the direct use of CSI for human movement sensing. To eliminate errors in CSI, we employ relative CSI, which is calculated by conjugate multiplication between the CSI of two adjacent antennas in the same receiver. The reason is that although different WiFi cards are not time-synchronized, all the transceiver chains on a single WiFi card share the same sampling clock, which have the same hardware errors [14]. In addition, to eliminate low (e.g., $< 5Hz$) and high (e.g., $> 100Hz$) frequency noises, we also employ Butterworth filter on the received CSI. The passband of the filter is set between $5Hz - 100Hz$ which covers the frequency range of most body gestures [5]. Hence, $FreeAuth$ first mitigates the impact of errors and noises on raw CSI of WiFi signals.

Signal Segmentation. To effectively extract features from body gestures, $FreeAuth$ segments consecutive signal series into episodes of each independent body gesture. Since the errors in CSI relative phase are eliminated through conjugate multiplication, the variance of relative phase only results from human movement. Hence, to separate each body gesture from adjacent ones in a consecutive signal series, we propose to detect variance of relative phase and compare the variance with a threshold for signal segmentation. The variance of relative phase is calculated by the difference between adjacent elements. However, in practice, some variances within a body gesture are also below the threshold. Fig. 5(a) shows the variances of two body gestures, i.e., standing up and walking. We can observe that some variances in walking are below the threshold, which may result in a false segmentation and discard useful behavioral information. To accurately segment body gestures, we employ sliding windows to measure the overall variance condition within each time window. Specifically, if half of the variances are above the threshold, the signal in the sliding window is considered as the component of a body gesture, and the first point higher than the threshold is judged as the start of the body gesture. On the contrary, half of the variances lower than the threshold is considered as interval, which the last point higher than the threshold is judged as the end of the body gesture. The length of sliding windows and the value of threshold are empirically studied. Fig. 5(b) shows the gesture segmentation accuracy under different threshold values and window lengths. It can be observed that there is a trade-off between threshold and window length, where segmentation accuracy could reach 99.4% with a threshold of 0.04 and window length of 600. Hence, through sliding windows with the empirically studied parameters, $FreeAuth$ effectively segments signals into episodes of each body gesture.



(a) Normolized relative phase and variances for continous gestures.

(b) Accuracy of signal segmentation under different settings.

Fig. 5: Illustration of signal segmentation.

C. Adversarial Learning-based Model Construction

The CSI of WiFi signals induced by body gestures contains specific gesture characteristics and the underlying physiological characteristics of each individual. Hence, we propose an adversarial learning-based authentication model, i.e., an adversarial neural network [15], to suppress the behavioral interference of body gestures and extract invariant individual uniqueness unrelated to specific body gestures for gesture-independent authentication.

1) Designing the Adversarial Neural Network for Gesture-Independent Authentication: Fig. 6 shows the architecture of the adversarial neural network designed for gesture-independent authentication.

Feature Extractor. The feature extractor extracts fine-grained features from CSI of WiFi signals induced by body gestures to characterize individual uniqueness (i.e., physiological characteristics of each individual). The input of feature extractor is the preprocessed CSI of WiFi signals. In order to make full use of the CSI of all communication links and all subcarriers, $FreeAuth$ reshapes the received CSI under t communication links and m subcarriers in each communication link to $(m \cdot t) \times n$-dimension, where n is the length of each segmented relative CSI amplitude or phase. Then, $FreeAuth$ integrates the relative CSI amplitude and phase as a two-channel input I with $2 \times (m \cdot t) \times n$-dimension, which embeds non-linear features induced by body gestures. Since the Convolutional Neural Network (CNN), especially the convolutional operation, is specialized in well abstracting non-linear features, we select it as the basis of feature extractor.

The proposed CNN-based feature extractor consists of six layers, i.e., three convolutional layers and three pooling layers. The convolutional layer abstracts the input I as a compressed representation through the convolutional operation, and the pooling layer further reduces the dimension of the compressed

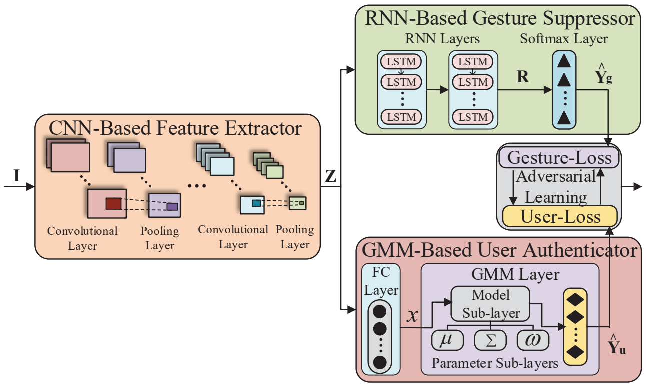


Fig. 6: Architecture of the adversarial neural network.

representation. Specifically, the input I is first fed to the first convolutional layer with 32 convolutional kernels of 5×5-dimension to derive a compressed feature map. After that, the compressed feature map is normalized by the batch normalization operation, and further activated by a Rectified Linear Unit (ReLU) function for accelerating the convergence velocity during model training. Then, the activated feature map is further fed to the first max pooling layer of 2×2-dimension to reduce the feature dimension. By analogy, the reduced feature map is fed to the second convolutional and pooling layers, as well as the third convolutional and pooling layers in turn. The second and third convolutional layers are with 64 4×4-dimension and 128 3×3-dimension convolutional kernels respectively. Both convolutional layers are normalized with the batch normalization and activated by the ReLU function. Also, the second and third pooling layers both conduct the max pooling operation of 2×2-dimension. Through the stack of the 6 layers, the feature extractor finally extracts a feature map Z, which is further fed to the gesture suppressor and user authenticator respectively.

Gesture Suppressor. Since the fine-grained features are extracted from the CSI induced by body gestures, specific gesture characteristics are inevitably contained underlying the extracted features. Hence, we need to suppress the behavioral interferences of body gestures to extract features independent of specific body gestures. We thus develop a gesture suppressor to suppress the behavioral interferences of body gestures. The behavioral interferences of body gestures are embedded underlying the sequential relationship of input representations, which depict the content of body gestures. Therefore, we employ the Recurrent Neural Network (RNN) that explores the sequential relationship [16] in the inputs to construct the gesture suppressor.

The gesture suppressor consists of two RNN layers with Long Short-Term Memory (LSTM) units and a softmax layer. Specifically, the gesture suppressor first partitions the feature map Z from the feature extractor into N small fragments, i.e., $Z = [Z_1, Z_2, ..., Z_N]$. Then, the fragmented inputs Z are fed to the two stacked RNN layers with LSTM units successively. Based on the two RNN layers, feature R is extracted as the output to represent the sequential relationships of body gestures underlying the extracted feature Z. The output R is further activated by the softmax layer to derive the probability $\hat{Y_g}$ which represents the probability of recognizing body gestures, i.e.,

$$\hat{Y_g} = softmax(W_g R + b_g), \tag{1}$$

where W_g and b_g are the weight and bias respectively, and $\hat{Y_g} = \{\hat{Y_g^1}, \cdots, \hat{Y_g^n}\}$. The probability is under the constraints that $0 \leqslant \hat{Y_g^k} \leqslant 1$ and $\sum_{k=1}^{n} \hat{Y_g^k} = 1$. This probability serves as the basis of gesture-loss for model construction, which aims to suppress the behavior interferences of body gestures.

User Authenticator. As mentioned in Section II-B, the physiological characteristics exist in statistical distributions underlying the CSI induced by body gestures. Inspired by text-independent speaker identification [13], we employ Gaussian Mixture Model (GMM) to construct a user authenticator to identify individuals in a gesture-independent manner. GMM can utilize multiple Gaussian distributions to fit the statistical distributions of an arbitrary input, which indicates that the statistical distributions of CSI can be extracted from GMM to characterize physiological characteristics for gesture-independent authentication.

The GMM-based user authenticator consists of a Fully Connected layer (i.e., FC layer) and a GMM layer [17]. In the fully connected layer, the feature map Z is first linearly combined with the weight W_1 and bias b_1, and then activated by Rectified Linear Unit (ReLU) function to derive an intermediate representation x, i.e., $x = ReLU(W_1 Z + b_1)$. Afterwards, the intermediate representation x is further fed into the GMM layer. The GMM layer consists of a model sub-layer to fit the statistical distributions of input features, three parameter sub-layers to store relative parameters, and an output sub-layer to calculate identity probability. In particular, assume there are n users registering in the system. The model sub-layer thus contains n nodes, each of which employs a GMM model to fit the distributions of each registered user under input features. Each node s (which corresponds to a registered user) is a likelihood function, which consists of g Gaussian functions, i.e.,

$$p(x|s) = \sum_{i=1}^{g} \omega_{si} \mathcal{N}(x, \mu_{si}, \Sigma_{si}), \tag{2}$$

where $\mathcal{N}(x, \mu_{si}, \Sigma_{si})$ is the i^{th} Gaussian function for node s, ω_{si}, μ_{si} and Σ_{si} are the weight, mean and covariance for the i^{th} Gaussian function respectively. To support the model sub-layer, the three additional parameter sub-layers store the weights $\omega_s = [\omega_{s1}, \cdots, \omega_{sg}]$, means $\mu_s = [\mu_{s1}, \cdots, \mu_{sg}]$, and covariances $\Sigma_s = [\Sigma_{s1}, \cdots, \Sigma_{sg}]$ for all the nodes s ($s \in [1, \cdots, n]$) of the model sub-layer, respectively. After calculating the likelihoods, the GMM layer further derives the logarithm joint distribution in the output layer, i.e.,

$$\log(p(x, s)) = \log(p(s)) + \log(p(x|s)), \tag{3}$$

where $p(s)$ is the prior of each registered user. The posterior probability $p(s|x)$ thus can be derived as

$$p(s|x) = \frac{p(x, s)}{\sum_s p(x, s)}. \tag{4}$$

Based on the posterior probabilities, the GMM-based user authenticator finally derives the user identity probability $\hat{Y_u} = \{\hat{Y_u^1}, \cdots, \hat{Y_u^n}\}$, where $\hat{Y_u^s} = p(s|x), s \in [1, n]$. The probability is under the constraints that $0 \leqslant \hat{Y_u^s} \leqslant 1$ and $\sum_{s=1}^{n} \hat{Y_u^s} = 1$. This probability servers as the basis of user-loss for model construction, which aims to enhance the capability of extracting individual uniqueness.

2) Training the Authentication Model Based on Adversarial Learning: Although the three sub-networks has seemingly collaborative function, the authentication model can achieve gesture independence only when they collaborate effectively. $FreeAuth$ trains the authentication model through adversarial learning for gesture-independent authentication.

To train the authentication model, $FreeAuth$ first initializes the structure and parameters of the designed adversarial neural network. Since the number for registered users only determines the structure of GMM layer, we initial the structure of the GMM layer by configuring it with n nodes same with the number of registered users. Then, we initialize the parameters of the neural network. The weights and biases of the CNN-based feature extractor are initialized as random values from normal distribution and a constant of 0.1 respectively. The weights and bias of the RNN-based gesture suppressor are initialized as orthogonal matrices (which are derived from singular value decomposition of a normal distribution matrix) and a constant of 0.1 respectively to avoid gradient vanishing and explosion. For the GMM-based user authenticator, the μ-layer and Σ-layer are set as random values from normal distribution and unit matrices respectively, and the ω-layer is initialized with uniform values of $1/g$.

After initialization, the adversarial neural network is then trained to enable gesture-independent authentication based on adversarial learning. Given the input I, the feature extractor abstracts a feature map Z. Then, the feature map Z is fed to the gesture suppressor and user authenticator to derive gesture probability vector $\hat{Y_g}$ and user identity probability vector $\hat{Y_u}$ respectively. After that, we derive gesture-loss and user-loss from the two kinds of probabilities. Specifically, given the identified identity probability vector $\hat{Y_u}$ from the user authenticator and the ground truth Y_u, the user-loss is defined as:

$$L_u = -\sum_{i=0}^{|U|-1} Y_u^i \log(\hat{Y}_u^i), \tag{5}$$

where Y_u^i and $\hat{Y}_u^i$ are the i^{th} entries in the corresponding probability vectors, and $|U|$ is the length of the two probability vectors determined by the number of registered users. Similarly, given the recognized gesture probability vector $\hat{Y}_g$ and the encoded gesture probability vector of ground truth Y_g, the gesture-loss is defined as:

$$L_g = -\sum_{k=0}^{|G|-1} Y_g^k \log(\hat{Y}_g^k), \tag{6}$$

where Y_g^k and $\hat{Y}_g^k$ are the k^{th} entries in the corresponding probability vectors, and $|G|$ is the length of the two probability vectors determined by the number of gesture kinds in the training data.

With the two loss functions (i.e., Eq. (5) and Eq. (6)), $FreeAuth$ can be optimized to extract gesture-independent individual uniqueness through maximizing the gesture-loss while minimizing the user-loss. However, since the specific gesture characteristics are more significant than the underlying physiological characteristics in the initial inputs, the optimization objective at the beginning of model training needs to lay emphasis on suppressing the behavioral interferences of body gestures. With the gradual optimization of suppressing behavioral interferences, the model training should turn to focus on characterizing invariant individual uniqueness for a gesture-independent authentication. Hence, linear combination of the two losses with subtract operations cannot balance the above optimization priority. Based on the analysis above, we employ a non-linear function (i.e., the exponential function) to combine the gesture-loss and user-loss, i.e.,

$$\min L = \min(\alpha(L_u + b) + \beta e^{-L_g + c}), \tag{7}$$

where α and β are weights of the user-loss and gesture-loss respectively, and b and c are the biases for user-loss and gesture-loss respectively. The non-linear function takes advantage of the property of the exponential function to gradually lower the priority of suppressing behavioral interferences of body gestures, and thereby relatively raises the priority of characterizing individual uniqueness with the gradual optimization. Using the combined optimization objective above, the adversarial neural network-based authentication model can be gradually trained with the capability of extracting invariant individual uniqueness from the feature extractor, and identifying individuals through the extracted individual uniqueness from the user authenticator.

IV. Evaluation

In this section, we evaluate the system performance of $FreeAuth$ in indoor environments.

A. Experimental Setup & Methodology

We use laptops quipped with Intel WiFi Link 5300 NIC and Linux 802.11n CSI Tool [12] as WiFi transmitter and receivers for extracting CSI of WiFi signals. The transmitter continuously emits $5GHz$ WiFi signals at 2000 sampling rate, and the two receivers with external antennas capture the WiFi signals and extract CSI through the tool. The distance between adjacent antennas is half the wavelength of the WiFi signal. To evaluate the performance of $FreeAuth$ under various environmental backgrounds, we conduct the experiment in three indoor environments, i.e., a meeting room, a lab, and an apartment. Fig. 7 shows the layout of the three indoor environments, where users are asked to perform body gestures within the rectangular region formed by the transmitter and receivers. The size of the sensing area is $2m \times 2m$. A video camera is placed in each environment to record the ground truth of body gestures and user identities.

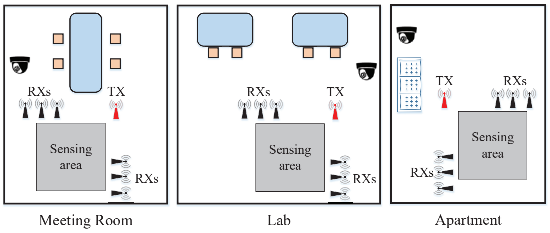


Fig. 7: Experimental environments.

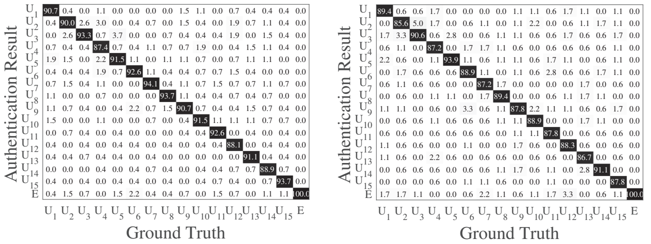
A total of 30 volunteers including 18 males and 12 females, aged between 21 and 45, are recruited in the experiment. These volunteers are recruited by offline advertisements in the university. We divide them equally into a legitimate user group and spoofer group, each of which contains 9 males and 6 females with the similar age distributions. The legitimate users register on $FreeAuth$ through performing body gestures, and the spoofers attempt to deceive the system.

We totally design 20 body gestures that are commonly used in daily life for the experiment. To avoid the impact of subjective selection of these gestures, we conduct 5-fold cross validations, in which different combinations of all the gestures are used as training and testing sets respectively. Specifically, the 20 gestures are divided into 5 groups, each of which contains 4 gestures, as shown in Table I. During each cross validation, 4 groups' gestures are set as *known gestures* for model training, i.e., 16 gestures are provided by each legitimate user as training data for the system. And the left one group's gestures are set as *unknown gestures* where the 4 gestures are out of use for training but only for evaluation. Finally, we obtain final evaluation performance by averaging the 5 cross validations. During each cross validation, the legitimate users are required to perform each known gesture 12 times to train the authentication model. For evaluation, a user performs 15 times for each body gesture, including the known and unknown gestures.

Overall Performance. Fig. 8 shows the confusion matrices of legitimate user identification under known and unknown gestures respectively. The confusion matrix exhibits which legitimate identity or not a legitimate identity (i.e., an empty identity represented by 'E') is identified for each user. We can observe that for known gestures, $FreeAuth$ can achieve overall 91.3% authentication accuracy with a deviation of 2.1% for identifying legitimate users. The result indicates that $FreeAuth$ can achieve satisfactory performance on traditional user authentication. On the other hand, for unknown gestures, $FreeAuth$ achieves the overall accuracy of 88.5% in legitimate user identification with a deviation of 2.4%. The difference of authentication accuracies between known and unknown gestures is only 2.8%, which is insignificant. The results demonstrate that $FreeAuth$ is effective in identifying users in gesture-independent manner.

TABLE I: Body gestures and corresponding groups.

Group 1	Stand up, Turn on the light, Pick up the phone, Type keyboard
Group 2	Walk, Wear the coat, Open the door, Clean the desk
Group 3	Sit down, Throw the rubbish, Hang out the cloth, Wear glasses
Group 4	Turn around, Pick up a cup, Plug the power adapter, Sweep the floor
Group 5	Open the drawer, Take off the hat, Pick up the tableware, Turn the book



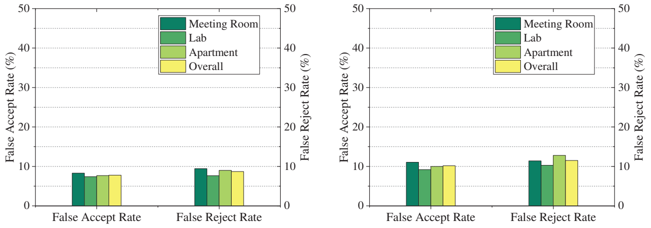
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	E
U1	90.7	0.4	0.0	1.1	0.0	0.0	0.0	0.0	1.5	1.1	0.0	0.7	0.4	0.4	0.4	0.0
U2	0.4	90.0	2.6	3.0	0.0	0.4	0.7	0.0	1.5	0.4	0.0	1.9	0.7	1.1	0.4	0.0
U3	0.0	2.6	93.3	0.7	3.7	0.0	0.0	0.7	0.4	0.4	0.4	1.9	0.0	1.5	0.4	0.0
U4	0.7	0.4	0.4	87.4	0.7	0.4	1.1	0.7	0.7	1.9	0.0	0.4	1.5	1.1	0.4	0.0
U5	1.9	1.5	0.0	2.2	91.5	1.1	0.0	1.1	1.1	0.7	0.0	0.7	1.1	1.1	0.0	0.0
U6	0.4	0.4	0.4	1.9	0.7	92.6	1.1	0.4	0.4	0.7	0.7	1.5	0.4	0.0	0.0	0.0
U7	0.7	1.5	0.4	1.1	0.0	0.0	94.1	0.4	1.1	0.7	1.5	0.7	0.7	1.1	0.7	0.0
U8	1.1	0.0	0.0	0.7	0.0	0.0	0.0	93.7	1.1	0.4	0.4	0.7	0.4	0.7	0.0	0.0
U9	1.1	0.7	0.4	0.0	0.4	2.2	0.7	1.5	90.7	0.7	0.4	0.4	1.5	0.7	0.4	0.0
U10	1.5	0.0	0.0	1.1	0.0	0.0	0.7	0.0	0.4	91.5	1.1	1.1	1.1	0.7	0.7	0.0
U11	0.0	0.7	0.4	0.0	0.4	0.0	0.0	0.0	0.4	0.4	92.6	0.0	0.4	0.0	0.4	0.0
U12	0.4	0.0	0.4	0.0	0.4	0.4	0.0	0.4	0.0	0.4	0.4	88.1	0.0	0.4	0.4	0.0
U13	0.4	0.4	0.7	0.4	0.4	0.0	0.4	0.0	0.0	0.0	0.0	0.0	91.1	0.4	0.4	0.0
U14	0.0	0.0	0.0	0.4	0.0	0.4	0.7	0.4	0.0	0.4	0.4	0.7	0.7	88.9	0.7	0.0
U15	0.4	0.0	0.4	0.0	0.4	0.4	0.0	0.4	0.0	0.4	0.7	0.4	0.0	0.7	93.7	0.0
E	0.4	1.5	0.7	0.0	1.5	2.2	0.4	0.4	0.7	0.0	1.5	0.7	0.0	1.1	1.1	100.0

(a) Known gestures.

	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	E
U1	89.4	0.6	0.6	1.7	0.0	0.0	0.0	0.0	1.1	1.1	0.6	1.7	1.1	0.0	1.1	0.0
U2	0.0	85.6	5.0	1.7	0.0	0.6	1.1	0.0	1.1	2.2	0.0	0.6	1.1	1.7	1.1	0.0
U3	1.7	3.3	90.6	0.6	2.8	0.0	0.6	1.1	1.1	0.6	0.6	1.7	0.6	0.6	1.7	0.0
U4	0.6	1.1	0.0	87.2	0.0	1.7	1.7	1.1	0.6	0.0	0.6	0.6	0.6	0.6	1.1	0.0
U5	2.2	0.6	0.0	1.1	93.9	1.1	0.6	1.1	1.1	0.6	1.7	1.1	0.0	0.6	1.1	0.0
U6	0.0	1.7	0.6	0.6	0.6	88.9	1.1	0.0	1.1	0.6	2.8	0.6	0.6	1.7	1.1	0.0
U7	0.6	1.1	0.6	0.6	0.0	0.0	87.2	1.7	0.0	0.0	0.0	0.0	1.7	0.0	0.6	0.0
U8	0.0	0.6	0.0	1.1	1.1	0.0	1.7	89.4	0.0	0.0	0.6	1.1	1.7	1.1	0.0	0.0
U9	1.1	1.1	0.0	0.6	0.0	3.3	0.6	1.1	87.8	2.2	1.1	1.1	0.6	0.6	1.7	0.0
U10	0.0	0.6	0.6	1.1	0.6	0.6	0.6	1.1	1.1	88.9	0.0	0.0	1.7	1.1	0.0	0.0
U11	0.6	0.6	0.0	0.6	0.6	0.6	0.6	0.0	1.1	1.1	87.8	0.0	0.6	0.6	0.6	0.0
U12	1.1	0.6	0.6	0.0	0.0	1.1	0.0	0.0	1.7	0.6	0.0	88.3	0.6	0.0	0.6	0.0
U13	1.1	0.6	0.0	2.2	0.6	0.0	0.6	0.6	0.0	0.6	0.6	0.0	86.7	0.0	0.6	0.0
U14	0.0	0.6	0.6	0.6	0.0	0.6	0.6	1.1	1.7	0.6	1.1	0.0	2.8	91.1	0.0	0.0
U15	0.0	0.0	0.0	0.6	0.0	1.1	1.1	0.6	0.0	0.0	1.1	0.0	0.0	0.0	87.8	0.0
E	1.7	1.7	1.1	0.0	0.0	0.6	2.2	1.1	0.6	1.1	1.7	3.3	0.0	0.6	1.1	100.0

(b) Unknown gestures.

Fig. 8: Confusion matrices of authentication accuracy under known and unknown gestures.



(a) Known gestures.

(b) Unknown gestures.

Fig. 9: FAR and FRR for unknown gestures.

Fig. 9(a) and Fig. 9(b) show the FAR and FRR under known and unknown gestures in the three environments. It can be observed that for known gestures, $FreeAuth$ achieves an overall FAR of 7.8% and FRR of 8.7%. As for unknown gestures, $FreeAuth$ achieves overall FAR and FRR of 10.1% and 11.5% respectively, which are only slightly higher than that of known gestures. This result further demonstrates that $FreeAuth$ can authenticate users in the gesture-independent manner. Also, we observe that the accuracy deviation among different environments is not significant, i.e., only 0.9% in average. This result indicates that $FreeAuth$ is robust in authenticating users under different indoor environments with various room sizes and layouts.

Comparison with Baseline Approaches. By authenticating users in the gesture-independent manner, $FreeAuth$ enables more general and flexible authentication capability compared with existing gesture-dependent approaches. We further evaluate the effectiveness of our system on user authentication by comparing with 3 state-of-the-art gesture-dependent approaches, i.e., $Smart$ [5], $WiID$ [8], and $FingerPass$ [16], which act as the baseline for comparison. For variable control, the process of training and evaluation for the three approaches follows the guideline of $FreeAuth$ described in the evaluation setup, so all the four approaches are with the same training data and evaluation methodology.

Fig. 10 shows the authentication performance of the four approaches under known and unknown gestures respectively. For known gestures, we can see that the three baseline approaches achieve over 92% accuracy, and $FreeAuth$ achieves a similar authentication accuracy of 90.7%, which demon-

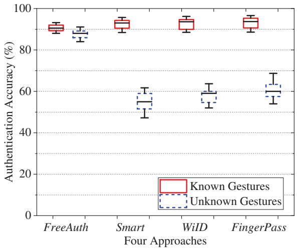


Fig. 10: Authentication accuracy of different approaches.

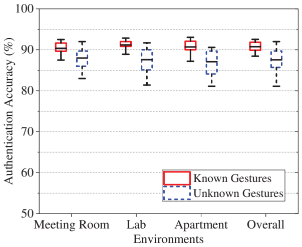


Fig. 11: Authentication accuracy under cross validations.

strates an effective traditional authentication capability of the proposed system. On the other hand, for unknown gestures, the three baseline approaches suffer from significant performance decline for about 30% compared with $FreeAuth$. This is because the existing approaches do not involve gesture-independent techniques to their systems. Moreover, we can see that the deviation of authentication accuracies of $FreeAuth$ is smaller than other approaches. This is because the adversarial learning can avoid the impact of different gestures on model training, and $FreeAuth$ extracts more generalized features from different gestures to enable robust user authentication. The above result demonstrates the improvement of $FreeAuth$ over baseline approaches.

Robustness Performance. To evaluate the robustness performance of $FreeAuth$ under different gesture sets, we analyze the authentication accuracy among different rounds of cross validation. Since the cross validation mechanism conducts several rounds of evaluation where different combinations of gesture sets are used as training and testing samples, the deviation among different validation rounds is able to exhibit the robustness of $FreeAuth$ among different gesture sets. Fig. 11 shows the authentication accuracy among the cross validations in three environments with known and unknown gestures. It can be first observed that the differences between different environments are insignificant, which is consistent with previous results. Besides, the performances among cross validations for the known gestures are more stable than that of unknown gestures. Specifically, the deviation of authentication accuracies under known gestures for cross validations is only 2.06%, while that under unknown gestures is 6.13%. This result indicates that different gesture sets could affect the authentication performance under unknown gestures.

Impact of Loss Function. Loss function is a key design of the adversarial neural network in $FreeAuth$. In this experiment, in addition to the exponential function combining the user-loss and gesture-loss (i.e., L = Eq. (7)), we employ three other loss functions, i.e., only user-loss function ($L_1 = \alpha_1 L_u$), linear function for loss combination ($L_2 = \alpha_2 L_u + \beta_2(-L_g) + c_2$), and logarithmic function for loss combination ($L_3 = \alpha_3 L_u + \beta_3 \log(-L_g + c_3)$).

Fig. 12 shows the authentication accuracy of $FreeAuth$ with the four different loss functions. It can be observed that $FreeAuth$ with the loss function L has the best authentication performance among all proposed loss functions. For other three loss functions, the authentication accuracies under unknown gestures decrease below 70%. This is because L_1 only minimizes the user-loss without suppressing the body gesture interferences, which lefts abundant gesture-specific information in the extracted features. The linear function L_2 takes both user-loss and gesture-loss into consideration, but the simple linear combination of the two losses cannot well balance the priority of interference suppression and uniqueness extraction. As for the logarithmic function L_3, it optimizes the model following to gradually strengthen the suppression of gesture interferences, and thereby occupies the priority of extracting individual uniqueness, which is not an appropriate optimization way to extract the gesture-independent features.

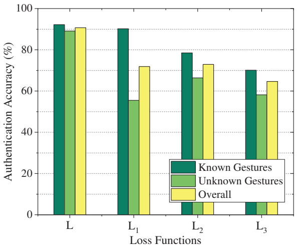


Fig. 12: Authentication accuracy with different loss functions.

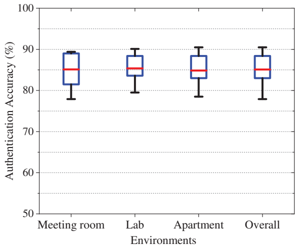


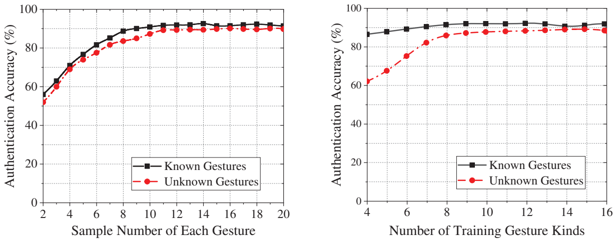
Fig. 13: Authentication accuracy for in-the-wild evaluation.

In-the-wild Evaluation. Since $FreeAuth$ aims to realize gesture-independent user authentication, the evaluation that releases the restriction of performing specific gestures can comprehensively evaluate the system performance. Hence, we conduct an in-the-wild evaluation where users perform body gestures without any gesture kind restriction, to mirror the reality of actual gesture-independent authentication scenarios.

Fig. 13 shows the authentication accuracy of in-the-wild experiment. We can see from the figure that the mean authentication accuracies are 85.1%, 85.4%, and 84.8% for the three environments respectively, which demonstrates an insignificant impact of environment layout. The overall authentication accuracy of in-the-wild experiment is 85.1%. Compared with the overall authentication accuracy (i.e., 89.9%) for all designed gestures, the authentication accuracy of in-the-wild scenario does not decline much. This result indicates that our system can achieve an acceptable performance for practical gesture-independent authentication. However, it can be also observed that the variances among different users are significant. For example, in the apartment, the best authentication performance for a user is 90.5% while the worst is only 78.5%. By analyzing the behavior contents of these users from the video, we conclude that the user who performs more complex and conspicuous body gestures tends to be identified more precisely. This is because these gestures provide more information about behavioral and physiological characteristics of the individual.

Impact of Training Data Size. The training data size depends on the number of samples for each body gesture, and the number of body gesture kinds for training. A large training data size requires frequent performing of body gestures during registration, while a small training data size could not support the model with sufficient generalization ability.

We first evaluate the performance of $FreeAuth$ under different numbers of samples for each body gesture. Fig. 14(a) shows authentication accuracies with different numbers



(a) Sample number of each gesture. (b) Number of gesture kinds.

Fig. 14: Authentication accuracy under different training sizes.

of samples for each gesture in model training under known and unknown gestures respectively. We can see that as the number of samples for each gesture increases, the authentication accuracies first increase rapidly, and then go stable. When the number of samples for each gesture increases to 12, $FreeAuth$ could achieve around 91% and 88% authentication accuracies under known and unknown gestures respectively. More training samples would not contribute to an improvement in system performance.

We also evaluate the performance of $FreeAuth$ under different numbers of body gesture kinds. In this experiment, we involve 12 samples for each body gesture, which follows the results of the previous experiment. The numbers of gesture kinds for testing under known and unknown gestures are both set as 4 for variable control of the experiment. Fig. 14(b) shows the authentication accuracy of $FreeAuth$ under different numbers of body gesture kinds in model training under known and unknown gestures. It can be observed from the figure that under known gestures, the authentication accuracy remains stable at around 90% as the number of body gesture kinds increases. But for unknown gestures, the authentication accuracy first increases rapidly and then tends to be stable with the increase of body gesture kinds. This is because as the number of body gesture kinds increases, $FreeAuth$ is capable with rich prior knowledge of physiological characteristics from various body gestures, which helps to improve the capability of $FreeAuth$ on gesture-independent user authentication. It can be also observed that as the number of body gesture kinds increases to 10 for model training, $FreeAuth$ approaches 90% authentication accuracy for both known and unknown gestures. Such a number of body gesture kinds during registration is acceptable for most users.

V. Related Work

In this section, we review some existing works that are related to $FreeAuth$.

Biometrics-based User Authentication. Biometrics-based authentication approaches are widely investigated and realized, which exploits the physiological uniqueness of individuals for user authentication. Early studies utilize fingerprint [18], face recognition [19], and voiceprint [20] for authentication. However, these approaches either are vulnerable to replay attacks or require specialized infrastructures for attack resistance. To handle these problems, some works [21]–[24] achieve liveness verification using low-cost approaches. However, all these approaches are one-off user authentications and usually induce intrusive user experiences.

Gesture-based User Authentication. To enable user authentication, some studies explore the individual uniqueness from human behaviors. Existing works require users wearing on-body devices [1], [2] to capture body behaviors for authentication. Other researches use special vision devices [3], [4] to capture human behaviors to identify individuals. However, the wearable-based approach are intrusive for users, and vision-based methods require customized devices. To employ low-cost and widely deployed infrastructures, other studies explore WiFi signals to sense human movements for user authentication. Specifically, some researches implement user authentication through sensing human gaits [25], [26]. Following works extend to realize user authentication with coarse-grained activities [5] or fine-grained gestures [6], [8], [16]. However, these work can only authenticate users when they perform predefined activities and gestures.

Wireless Radio-based Application. WiFi infrastructures are widely deployed in indoor environments recently, which realizes the WiFi-based sensing to support various applications. Previous studies utilize WiFi signals to realize wireless sensing applications, such as crowd counting [27], indoor localization [28], [29], activity recognition [30]–[34], gesture recognition [35]–[37], human tracking [38]–[40], and breathing rate monitoring [41], etc. Following works [15], [42]–[44] utilize machine learning or model translation methods to realize cross-domain WiFi sensing. All of these works demonstrate the surge of the wireless radio-based applications.

VI. Conclusion

In this paper, we propose a user authentication system, $FreeAuth$, which leverages WiFi signals to identify individuals in a gesture-independent manner. First, we explore the physiological characteristics underlying body gestures, and find that different individuals exhibit individual differences in the statistical distributions under WiFi signals induced by various body gestures. We propose an adversarial learning-based model, which can suppress the behavioral interferences of body gestures, and extract invariant individual uniqueness unrelated to specific body gestures. With the model, $FreeAuth$ can continuously identify individuals through arbitrary body gestures. Experiment results in real indoor environments demonstrate that $FreeAuth$ is effective in gesture-independent user authentication.

Acknowledgement

This work is supported by National Key Research and Development Program of China (No. 2018YFC1900700), National Natural Science Foundation of China (No. 62172277, 62102354), China Computer Federation-Ant Financial Services Group Research Fund (No. CCF-AFSG RF20200013), Science and Technology Commission of Shanghai Municipality (STCSM) AI project (No. 19511120300), and Fundamental Research Funds for the Central Universities (No. 2021FZZX001-27).

REFERENCES

[1] J. Ranjan and K. Whitehouse, “Object hallmarks: Identifying object users using wearable wrist sensors,” in *Proc. ACM Ubicomp’15*, Osaka, Japan, 2015.

[2] Y. Li and M. Xie, “Understanding secure and usable gestures for realtime motion based authentication,” in *IEEE INFOCOM WKSHPS’18*, Honolulu, HI, USA, 2018, pp. 13–20.

[3] J. Wu, J. Konrad, and P. Ishwar, “Dynamic time warping for gesture-based user identification and authentication with kinect,” in *Proc. IEEE ICASSP’13*, Vancouver, BC, Canada, 2013.

[4] X. Wang and J. Tanaka, “Gesid: 3d gesture authentication based on depth camera and one-class classification,” *Sensors*, vol. 18, no. 10, p. 3265, 2018.

[5] C. Shi, J. Liu, H. Liu, and Y. Chen, “Smart user authentication through actuation of daily activities leveraging wifi-enabled iot,” in *Proc. ACM MobiHoc’17*, Chennai, India, 2017, p. 5.

[6] C. Li, M. Liu, and Z. Cao, “Wihf: Enable user identified gesture recognition with wifi,” in *IEEE INFOCOM’20*. Toronto, ON, Canada: IEEE, 2020, pp. 586–595.

[7] H. Kong, L. Lu, J. Yu, Y. Chen, and F. Tang, “Continuous authentication through finger gesture interaction for smart homes using wifi,” *IEEE Transactions on Mobile Computing*, 2020.

[8] M. Shahzad and S. Zhang, “Augmenting user identification with wifi based gesture recognition,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 3, p. 134, 2018.

[9] WHO, “Coronavirus disease (covid-19) pandemic,” [Online]. Available: https://www.who.int/en/emergencies/diseases/novel-coronavirus-2019, 2021.

[10] B. B. Mjaaland, P. Bours, and D. Gligoroski, “Walk the walk: attacking gait biometrics by imitation,” in *Proc. ISC’10*, FL, USA, 2010, pp. 361–380.

[11] Wikipedia, “Physiology.” [Online]. Available: https://https://en.wikipedia.org/wiki/Physiology/, 2019.

[12] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, “Tool release: Gathering 802.11 n traces with channel state information,” *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 53–53, 2011.

[13] H. Gish and M. Schmidt, “Text-independent speaker identification,” *IEEE signal processing magazine*, vol. 11, no. 4, pp. 18–32, 1994.

[14] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, “Spotfi: Decimeter level localization using wifi,” in *Proc. ACM SIGCOMM’15*, London, United Kingdom, 2015, pp. 269–282.

[15] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas *et al.*, “Towards environment independent device free human activity recognition,” in *Proc. ACM MobiCom’18*, New Delhi, India, 2018, pp. 289–304.

[16] 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. ACM MoboHoc’19*, Catania, Italy, 2019.

[17] E. Variani, E. McDermott, and G. Heigold, “A gaussian mixture model layer jointly optimized with discriminative features within a deep neural network architecture,” in *Proc. IEEE ICASSP’15*, South Brisbane, 2015, pp. 4270–4274.

[18] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, “Local contrast phase descriptor for fingerprint liveness detection,” *Pattern Recognition*, vol. 48, no. 4, pp. 1050–1058, 2015.

[19] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” in *Proc. IEEE CVPR’15*, Boston, MA, USA, 2015, pp. 815–823.

[20] L. G. Kersta, “Voiceprint identification,” *Nature*, vol. 196, no. 4861, pp. 1253–1257, 1962.

[21] Y. Li, L.-M. Po, X. Xu, L. Feng, and F. Yuan, “Face liveness detection and recognition using shearlet based feature descriptors,” in *Proc. IEEE ICASSP’16*, Shanghai, China, 2016, pp. 874–877.

[22] L. Lu, J. Yu, Y. Chen, H. Liu, Y. Zhu, Y. Liu, and M. Li, “Lippass: Lip reading-based user authentication on smartphones leveraging acoustic signals,” in *IEEE INFOCOM’18*, Honolulu, HI, USA, 2018, pp. 1466–1474.

[24] Y. Cao, Q. Zhang, F. Li, S. Yang, and Y. Wang, “Ppgpass: Nonintrusive and secure mobile two-factor authentication via wearables,” in *IEEE INFOCOM’20*. Toronto, ON, Canada: IEEE, 2020, pp. 1917–1926.

[23] Y. Cao, F. Li, Q. Zhang, S. Yang, and Y. Wang, “Towards nonintrusive and secure mobile two-factor authentication on wearables,” *IEEE Transactions on Mobile Computing*, 2021.

[25] Y. Zeng, P. H. Pathak, and P. Mohapatra, “Wiwho: wifi-based person identification in smart spaces,” in *Proc. IEEE IPSN’16*, Vienna, Austria, 2016, p. 4.

[26] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, “Wifi-id: Human identification using wifi signal,” in *Proc. IEEE DCOSS’16*, Hangzhou, China, 2016, pp. 75–82.

[27] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, “Freecount: Device-free crowd counting with commodity wifi,” in *Proc. IEEE GLOBECOM’17*, Singapore, 2017, pp. 1–6.

[28] D. Li, J. Xu, Z. Yang, Y. Lu, Q. Zhang, and X. Zhang, “Train once, locate anytime for anyone: Adversarial learning based wireless localization,” in *IEEE INFOCOM’21*. Vancouver, BC, Canada: IEEE, 2021, pp. 1–10.

[29] X. Wang, L. Gao, S. Mao, and S. Pandey, “Csi-based fingerprinting for indoor localization: A deep learning approach,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, 2016.

[30] Z. Chen, C. Cai, T. Zheng, J. Luo, J. Xiong, and X. Wang, “Rf-based human activity recognition using signal adapted convolutional neural network,” *IEEE Transactions on Mobile Computing*, pp. 1–13, 2021.

[31] S. Ding, Z. Chen, T. Zheng, and J. Luo, “Rf-net: A unified meta-learning framework for rf-enabled one-shot human activity recognition,” in *Proc. IEEE SenSys’20*, 2020, pp. 517–530.

[32] J. Yu, L. Lu, Y. Chen, Y. Zhu, and L. Kong, “An indirect eavesdropping attack of keystrokes on touch screen through acoustic sensing,” *IEEE Transactions on Mobile Computing*, 2019.

[33] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, “Understanding and modeling of wifi signal based human activity recognition,” in *Proc. ACM MobiCom’15*, New York, USA, 2015, pp. 65–76.

[34] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, “E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures,” in *Proc. ACM MobiCom’14*, Maui, Hawaii, USA, 2014, pp. 617–628.

[35] S. Tan and J. Yang, “Wifinger: leveraging commodity wifi for fine-grained finger gesture recognition,” in *Proc. ACM MobiHoc’16*, Paderborn, Germany, 2016, pp. 201–210.

[36] L. Zhang, Y. Zhang, and X. Zheng, “Wisign: Ubiquitous american sign language recognition using commercial wi-fi devices,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 11, no. 3, pp. 1–24, 2020.

[37] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu, “Design and implementation of a csi-based ubiquitous smoking detection system,” *IEEE/ACM Transactions on Networking*, vol. 25, no. 6, pp. 3781–3793, 2017.

[38] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, “Widar: Decimeter-level passive tracking via velocity monitoring with commodity wi-fi,” in *Proc. ACM MobiHoc’17*, Chennai, India, 2017, p. 6.

[39] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu, “Widar2. 0: Passive human tracking with a single wi-fi link,” in *Proc. ACM MobiSys’18*, Munich, Germany, 2018, pp. 350–361.

[40] S. Tan, L. Zhang, Z. Wang, and J. Yang, “Multitrack: Multi-user tracking and activity recognition using commodity wifi,” in *Proc. ACM CHI’19*, S. A. Brewster, G. Fitzpatrick, A. L. Cox, and V. Kostakos, Eds. Glasgow, Scotland, UK: ACM, 2019, p. 536.

[41] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, “Tracking vital signs during sleep leveraging off-the-shelf wifi,” in *Proc. ACM MobiHoc’15*, Hangzhou, China, 2015, pp. 267–276.

[42] J. Zhang, Z. Tang, M. Li, D. Fang, P. Nurmi, and Z. Wang, “Crosssense: towards cross-site and large-scale wifi sensing,” in *Proc. ACM MobiCom’18*, New Delhi, India, 2018, pp. 305–320.

[43] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, “Zero-effort cross-domain gesture recognition with wi-fi,” in *Proc. ACM MobiSys’19*, Seoul, South Korea, 2019, pp. 313–325.

[44] A. Virmani and M. Shahzad, “Position and orientation agnostic gesture recognition using wifi,” in *Proc. ACM MobiSys’17*, NY, USA, 2017, pp. 252–264.