NeuralWave: Gait-based User Identification through Commodity WiFi and Deep Learning

Akarsh Pokkunuru[1], Kalvik Jakkala[1], Arupjyoti Bhuyan[2], Pu Wang[1] and Zhi Sun[3]
[1]Dept. of Computer Science, University of North Carolina at Charlotte, NC, 28223 USA
[2]Idaho National Laboratory (INL), Idaho Falls, ID, 83402 USA
[3]Dept. of Electrical Engineering, University at Buffalo, Buffalo, NY, 14260 USA
Email: {apokkunu, kjakkala, pu.wang}@uncc.edu[1], arupjyoti.bhuyan@inl.gov[2], zhisun@buffalo.edu[3]

***Abstract*—This paper proposes *NeuralWave*, an intelligent and non-intrusive user identification system based on human gait biometrics extracted from WiFi signals. In particular, the channel state information (CSI) measurements are first collected from commodity WiFi devices. Then, a collection of data preprocessing schemes are applied to sanitize and calibrate the noisy and erroneous CSI data samples to manifest and augment the gait-induced radio-frequency (RF) signatures. Next, a 23-layer deep convolutional neural network, namely RadioNet, is developed to automatically learn the salient features from the preprocessed CSI data samples. The extracted features constitute a latent representation for the gait biometric that is discriminative enough to distinguish one person from another. Using the latent biometric representation, a softmax multi-class classifier is adopted to achieve accurate user identification. Extensive experiments in a typical indoor environment are conducted to show the effectiveness of our system. In particular, NeuralWave can achieve $87.76 \pm 2.14\%$ user identification accuracy for a group of 24 people. To the best of our knowledge, NeuralWave is the first in the literature to exploit deep learning for feature extraction and classification of physiological and behavioral gait biometrics embedded in CSI signals from commodity WiFi.**

I. Introduction

As a critical security enforcement method, user identification aims to verify the identity of a user before granting access to the data, services, devices, and facilities. Recently, the ubiquitously available WiFi devices have been exploited to capture human gait biometrics for user identification [1]–[3]. Gait is a person's natural walking style and a complex biological process that involves nervous and musculo-skeletal systems [4]. Medical studies have shown that gait patterns are unique to each person [5]–[7]. As shown in Figure 1, when a person walks, the movement of his/her body parts, such as arms, legs, and torso, cause unique variations in the WiFi signals, which can be captured by the fine-grained channel state information (CSI) available at the WiFi devices. Then, these CSI data, which serve as the radio-frequency (RF) gait biometrics, can be exploited for user identification. The user identification systems based on RF gait biometrics have three key advantages. First, they are of low cost leading to wide deployment of WiFi infrastructure. In addition, with the prosperity of the Internet of Things (IoTs) and the proliferation of WiFi-enabled smart devices, such as smart refrigerator, smart TV, smart thermostat, and home security system, everyone is surrounded by an invisible system of WiFi signals. Second, they are non-intrusive and non-disruptive because the gait-modulated WiFi signals can be captured from the distance without user awareness and cooperations, which greatly increases the usability of the authentication procedure. Third, they are robust to gait spoofing attacks, where a person tries to imitate the walking style of someone else in order to gain illegitimate access and advantages. The existing research shows that gait is potentially difficult to spoof because it is behavioural and encompasses the whole body [8]. What is more counterintuitive, as the spoofers devote more training efforts into gait mimicking, the results of the gait spoofing attacks become worse because our physiological habits work against us when we are trying to modify something as fundamental as the way we walk [9].

Work supported through the INL Laboratory Directed Research & Development (LDRD) Program under DOE Idaho Operations Office Contract DE-AC07-05ID14517.

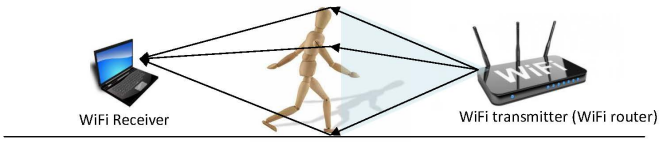

Fig. 1. User identification through gait biometrics embedded in WiFi signals

Despite aforementioned advantages, the performance of existing RF-biometric identification systems [1]–[3] is far from satisfactory. They can achieve high identification accuracy (maximum 93%) only for a very small group of people (i.e., between 2 to 10). Their identification accuracy decreases drastically from around 93% to around 78% as the number of users increases. These limitations stem from the fact that all the existing solutions use hand-crafted features to train a shallow classifier for recognizing different people. The commonly-adopted hand-crafted features include (1) time-domain features, such as the maximum, minimum, mean, skewness, kurtosis and standard deviation of the CSI amplitude and phase, and (2) the frequency-domain features, such as entropy, energy and spectrogram signatures of the CSI waveforms. These time-frequency domain features are generally selected in a heuristic and suboptimal manner, and thus fail to characterize the representative patterns in RF gait biometrics, which are invariant (or robust) to irrelevant variations in the

978-1-5090-6684-1/18/$31.00 ©2018 IEEE

WiFi signals and are discriminative enough to distinguish a large number of people from each other.

To address aforementioned limitations, we propose *NeuralWave*, an intelligent and non-intrusive user identification system that offers significantly higher accuracy and scales well for a large group of human subjects. Our system employs a sequence of data preprocessing schemes, which sanitize and calibrate the noisy and erroneous CSI data samples to augment and manifest the gait-induced radio-frequency (RF) signatures. Then, a 23-layer deep convolutional neural network (ConvNet), called RadioNet, is designed to automatically extract the salient gait features from the preprocessed CSI data samples. The extracted features constitute a latent representation for the gait biometric that is discriminative enough to distinguish one person from another. Using the latent biometric representation, a softmax multi-class classifier is adopted to achieve accurate user identification. Experiments in a typical indoor environment are conducted to show the effectiveness of our system. In particular, NeuralWave can achieve $87.76 \pm 2.14\%$ user identification accuracy for a group of 24 people. To the best of our knowledge, NeuralWave is the first in the literature to exploit deep learning for feature extraction and classification of the physiological and behavioral gait biometrics embedded in CSI signals from commodity WiFi.

The rest of this paper is organized as follows. Section II introduces the overall system design and Section III presents the data preprocessing schems. Section IV provides comprehensive details about the proposed RadioNet. In Section V, we show the experimental validation results and conclude the paper in Section VI.

II. System Design

In this section, we first introduce some preliminaries of CSI, then present the system design challenges, and finally provide an overview of *NeuralWave* system.

A. Preliminaries

The multipath propagation of wireless channel can be characterized by channel impulse response (CIR) as followers

$$h(\tau;t) = \underbrace{\sum_{n \in p^s} a_n\, \delta(\tau - \tau_n)}_{h_{static}(\tau)} + \underbrace{\sum_{m \in p^d} a_m(t)\, \xi(f)\, \delta(\tau - \tau_m(t))}_{h_{dynamic}(\tau;t)} \tag{1}$$

where $h(\tau;t)$ is the received signal at the receiver if the transmitter sends a Dirac pulse signal $\delta(\tau)$. Due to the presence of the multiple signal propagation paths, at the receiver, more than one pulse will be received, and each one of them travels along a different length. When a person walks, some paths $\forall m \in p^d$ will experience a time-varying length change caused by the movements of body parts. This leads to time-varying propagation delay $\tau_m(t)$ and signal attenuation $a_m(t)$ on each path $m \in p^d$. The other paths $\forall n \in p^s$ are not affected by the moving person and thus have constant path lengths. This leads to time-invariant delay τ_n and attenuation a_n. Furthermore, $\xi(f)$ is the frequency dependent absorption cross section coefficient (ACS), which is related to the body specific signal absorption [10]. For the time-varying impulse response $h(\tau;t)$, we can define a time-varying channel frequency response (CFR) $H(f;t) = \int_{-\infty}^{\infty} h(\tau;t) e^{-2j\pi f \tau} d\tau$, which is also known as the channel state information (CSI). Similar to $h(\tau;t)$, $H(f;t)$ contains both static and dynamic components, i.e.,

$$H(f;t) = \underbrace{\sum_{n \in p^s} a_n\, e^{-2j\pi f \tau_n}}_{H_{static}(f)} + \underbrace{\sum_{m \in p^d} a_m(t)\, \xi(f)\, e^{-2j\pi f \tau_m(t)}}_{H_{dynamic}(f;t)} \tag{2}$$

where the dynamic component $H_{dynamic}(f;t)$ contains the wireless channel perturbations determined by every individual's physiological characteristics such as body shape, height and natural walking style, i.e. gait. Therefore, $H_{dynamic}(f;t)$ can be treated as the RF biometric, which is unique for each person and can be exploited for user identification.

B. CSI Data Sample

The CSI measurements can be collected at commercial off-the-shelf WiFi devices. In particular, current WiFi standards, e.g., IEEE 802.11n/ac, exploit orthogonal frequency division modulation (OFDM) along with the multiple-input and multiple-output (MIMO) technology, where signals are transmitted and received over multiple subcarriers using multiple antennas at transmitter and receiver sides. Thus both spatial and frequency diversity features are exploited. In this case, the CSI between a pair of transmitting antenna x and receiving antenna y at the subcarrier i with the central frequency f_i can be presented by

$$H_{x,y}(f_i;t) = \|H_{x,y}(f_i;t)\|\, e^{j\angle H_{x,y}(f_i;t)} \tag{3}$$

where $\|H_{x,y}(f_i;t)\|$ and $\angle H_{x,y}(f_i;t)$ represent the amplitude and phase of $H_{x,y}(f_i;t)$, respectively. Let N_{tx}, N_{rx} and N_C represent the number of transmitter antennas, receiver antennas and subcarriers, respectively. The time series of the CSI measurements between each antenna pair over one particular subcarrier constitute one CSI waveform. Each CSI sample consists of $N_{wave} = N_{tx} \times N_{rx} \times N_C$ CSI waveforms and each waveform consists of N_T CSI measurements over the measurement time duration T. Moreover, since each CSI measurement contains both amplitude and phase attributes, we can structure each CSI data by a 2D matrix, i.e.,

$$\mathbf{X}_{CSI} \in \mathbb{R}^{2N_{wave} \times N_T} \tag{4}$$

which has a dimension of $N_T \times 2N_{wave}$.

In this research, we adopt the 802.11n WiFi transceiver based on Intel 5300 NIC, which has 3 transmitter antennas ($N_{tx} = 3$), 3 receiver antennas ($N_{rx} = 3$) and 30 subcarriers ($N_C = 30$). Therefore, each CSI observation has $N_{wave} = N_{tx} \times N_{rx} \times N_C = 270$ CSI waveforms. While a person is walking, our system only takes the CSI measurements for 4 seconds, i.e. $T = 4$ second, and the WiFi transmitter sends 2000 packets/second. Thus, each CSI waveform consists of $N_T = 8000$ CSI measurements. We structure each CSI data sample $\mathbf{X}_{CSI}$ by horizontally concatenating the CSI

amplitude matrix $\mathbf{X}_{amp} \in \mathbb{R}^{N_{wave} \times N_T}$ and CSI phase matrix $\mathbf{X}_{ph} \in \mathbb{R}^{N_{wave} \times N_T}$.

$$\mathbf{X}_{CSI} = [\mathbf{X}_{amp} \quad \mathbf{X}_{ph}] \tag{5}$$

where $\mathbf{X}_{amp}$ and $\mathbf{X}_{ph}$ are shown in Fig. 2. Each column i of the $\mathbf{X}_{CSI}^{amp}$ is called the CSI amplitude waveform, which contains the amplitude values of a particular CSI waveform i. Similarly, each column i of the $\mathbf{X}_{CSI}^{ph}$ constitutes the CSI phase waveform i, which contains the phase values of a particular CSI waveform i.

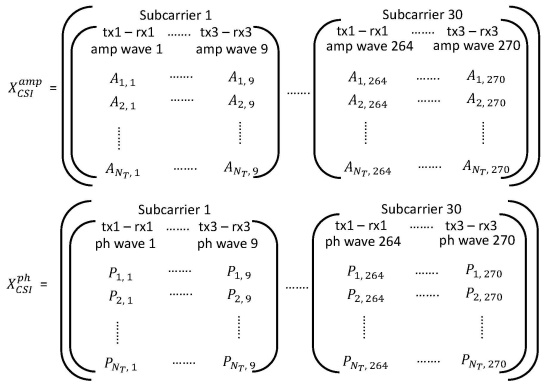

Fig. 2. CSI amplitude and phase matrices. tx i - rx i indicates the antenna pair between transmitting antenna i and receiving antenna j. The CSI amplitude and phase time series between each antenna pair on one particular subcarrier forms a CSI amplitude and phase waveforms

To visualize the gait-signature captured by $\mathbf{X}_{CSI}$, we apply principle component analysis to reduce the number of columns or waveforms in $\mathbf{X}_{CSI}$ while retaining most information. This is achieved by exploiting the high correlation among the 270 amplitude/phase waveforms to combine them into several principle component CSI (PC-CSI) waveforms that keep the maximum amount of variation or information about how the original CSI data is distributed. Fig. 3 (top) shows the PC-CSI waveform of principle component 3. The signal variations are caused by the movements of body parts (e.g., torso, arms, and legs). Different body parts move at different speed, which causes different Doppler frequency shift. To manifest such effects, we convert the PC-CSI waveform into time-frequency domain to generate the corresponding spectrogram. As shown in Fig. 3 (bottom), the high-energy (hot colored) components correspond to torso reflections and the low-energy components are caused by the swing of legs and arms during walking.

C. Challenges

The key idea of NeuralWave system is to achieve intelligent and accurate user identification by exploiting ConvNets to automatically learn high-level discriminative features of human gait biometrics embedded in noisy and error-prone WiFi signals. ConvNets have shown to be astoundingly effective for a wide range of tasks, such as image classification, object recognition, speech recognition, and machine translation. However, we are facing several challenges that greatly degrade the robustness and generalization capability of ConvNets.

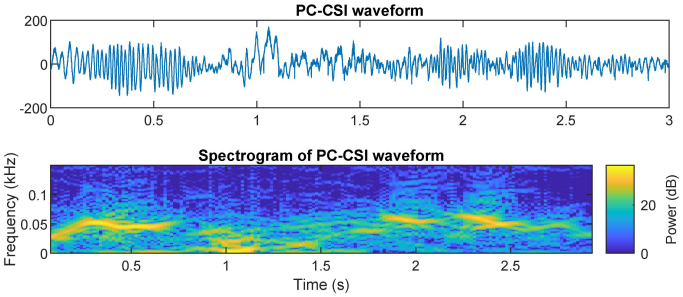

Fig. 3. Gait signatures captured by WiFi signals

- **Noisy and erroneous CSI measurements:** CSI data samples contain profound irrelevant variations, missing values, and errors due to noises, device imperfections, and internal state transitions of WiFi transmitter and receiver. The resulting random fluctuations in the training data can be picked up and learned as the features by the network, which significantly degrades the network robustness.
- **Limited number of CSI training samples**. The publicly-available large datasets are critical for training deep neural networks to learn internal structures and distinctive features from the raw data samples, which are essential for high-performance classification and recognition tasks. For example. ImageNet [11], which is a large visual database that contains 16 million labeled images, is the key for the success of applying ConvNets to realize superior object recognition accuracy. However, the CSI datasets that contain gait biometrics are not publicly available. Collecting such datasets is time/cost consuming. The size of such datasets is further limited because of privacy concerns. In our case, 24 human subjects are recruited and each of them has 40 CSI samples, 70% of which are used for training. This leads to a training dataset of 672 CSI samples. Such small CSI training dataset would prevent the state-of-the-art ConvNets, e.g., VGG-16/-19 [12] and ResNet-101/ResNet-152 [13], from learning useful features because they need sufficient data samples to tune tens of millions of learnable parameters.
- **High-dimensional CSI sample:** Each CSI sample is of high dimension in the sense that the number of feature elements in each sample is much larger than the number of available training samples. For example, a 4-second CSI sample could contain over $4,320,000$ feature elements, which is way above the size of our CSI training dataset that only contains around 672 CSI samples. Such high-dimensional data could easily make the network overfit the training dataset so that it wont generalize well to unseen and untrained data samples.

D. System Architecture Overview

As shown in Figure 4, NeuralWave system first collects CSI measurements to construct CSI data samples, each of which consists of 270 CSI phase waveforms and 270 CSI amplitude waveforms. Different from previous research [1]–[3] that only exploit CSI amplitude information, we employ

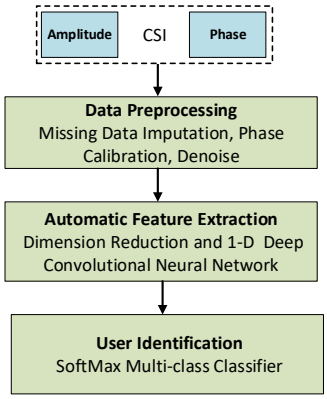

Fig. 4. NeuralWave System Architecture Overview

both CSI amplitude and phase to capture more particularities of human gait patterns. Then, a set of data preprocessing schemes are developed to sanitize and calibrate the raw CSI data samples so that more robust features can be learnt by deep neural networks. In particular, the number of active transmitter antennas of community WiFi is changing over time, and the CSI measurements coming from these inactive antennas will be unavailable. Therefore, data imputation is first applied to fill in the missing values in the CSI data samples. Then, we calibrate the CSI phase values to correct the errors caused by the sampling and carrier frequency offsets. These offsets are due to the lack of clock and frequency synchronization between transmitter and receiver. Next, wavelet denoising technology is adopted to preserve the essential information in both CSI phase and amplitude measurements, while suppressing the profound noises. With the preprocessed CSI data samples as inputs, we develop and train a 23-layer deep convolutional neural network, called RadioNet, which acts as an effective feature extractor to automatically extract the salient and distinctive features, which characterize the physiological and behavioral gait patterns. Then, the extracted features constitute the high-level latent representation of each gait-modulated CSI sample. The learnt latent representations are used by a softmax multi-class classifier for user identification. To mitigate over-fitting problem and improve network generalization capability, we first reduce the dimension of the preprocessed CSI data samples by employing principle component analysis (PCA) technology. Then, the hyperparameters of RadioNet, e.g., network depth, layer types, number of filers, and filter size, are fine-tuned to make sure the network architecture has sufficient capacity to characterize the complexity of CSI data, while the number of learnable parameters is not excessively large to ensure the network generalizes well even through the small training dataset is available to train the network.

III. CSI Data Preprocessing

The CSI measurements come from off-the-shelf WiFi devices, which are generally contaminated due to noises and device imperfections. In particular, lots of irrelevant variations appear in the CSI measurements because of the ambient noise, measurement noise, and impulse noise induced by internal state transitions at the transmitter and receiver. Moreover, the hardware imperfections, e. g., clock and frequency offsets, lead to the error-prone CSI phase estimations. To learn the salient features of CSI tensors for user identifications, it is of significant importance to perform effective data sanitization to recover clean and genuine CSI samples from contaminated and erroneous ones.

A. Missing Data Imputation

The WiFi devices can dynamically change the operation modes between SIMO (single-input-multiple-output) and MIMO. When SIMO is adopted, only one or two transmission antennas are used for data transmissions. As a result, the amplitude and phase matrices can have some missing elements for certain antenna pairs. For example, if antenna 1 is not used during transmission at time instant 1, $A_{1,1}$, $A_{1,2}$, and $A_{1,3}$ in matrix X_{CSI}^{amp} (shown in Fig. 2) will be missing. To fill in or impute missing values, a variety of imputation approaches can be used, such as mean, KNN, and regression imputation [14]. In NeuralWave system, we adopt row-wise mean imputation, where the missing values are replaced by the mean of the available ones, which are in the same row as the missing one. The motivation of performing mean imputation is that each row of the amplitude or phase matrix contains $N_{tx} \times N_{rx} \times N_C$ elements, which represent the CSI measurements of the CSI waveforms at a particular time instance. Since the CSI waveforms are correlated at each time instance, the mean imputation will not introduce much bias. Moreover, mean imputation is computationally light, which is suitable for real-time implementation. An example of imputed CSI waveform is shown in Fig. 5.

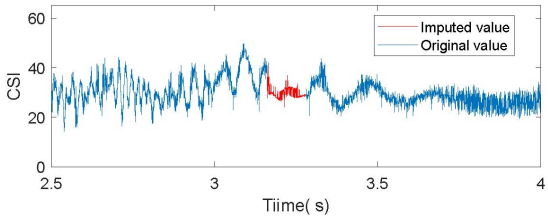

Fig. 5. Mean Imputation of CSI waveform. The red points are the imputed data that do not exist in the original CSI measurements

B. Phase Calibration

Besides the CSI amplitude, the CSI phase information is also exploited by NeuralWave system to extract the unique gait features of each person. However, the CSI phase measurements are highly random and inaccurate. The phase errors come from two sources: sampling frequency offset (SFO) and carrier frequency offset (CFO). SFO and CFO are due to the lack of clock and frequency synchronization between transmitter and receiver, respectively. As a result, the raw CSI phase measurement $\widehat{H_{x,y}(f_i;t)}$ consist of four parts:

$$\angle\widehat{H_{x,y}(f_i;t)} = \angle H_{x,y}(f_i;t) + \varepsilon_s \times i + \varepsilon_c + Z \qquad (6)$$

where $\angle H_{x,y}(f_i;t)$ is the genuine phase, ε_s is the error caused by SFO, i is the index of subcarrier f_i, ε_c is the

error caused by CFO, and Z is the measurement noise. To estimate the true phase information, we adopt the linear transformation method [15]. In particular, let $\angle\widetilde{H_{x,y}(f_i;t)}$ denote the estimation of the true phase $\angle H_{x,y}(f_i;t)$ and we have $\angle\widetilde{H_{x,y}(f_i;t)} = \angle\widehat{H_{x,y}(f_i;t)} - \alpha \times i - \beta$, where α and β are the slope and offset of the phase change across all subcarriers. For example, Intel 5300 WiFi NIC has the phase measurements of 30 subcarriers and correspondingly we have the slope $\alpha = \frac{\angle\widehat{H_{x,y}(f_{30};t)} - \angle\widehat{H_{x,y}(f_1;t)}}{30-1}$ and the phase offset $\beta = \frac{1}{30}\sum_{i=1}^{30}\angle\widehat{H_{x,y}(f_i;t)}$. It can be seen in Fig. 6 that the raw phase measurements (as red dots) are distributed randomly over all feasible angles between $0°$ and $360°$, while the calibrated phases after linear transformation are much more stable and reside within a small sector.

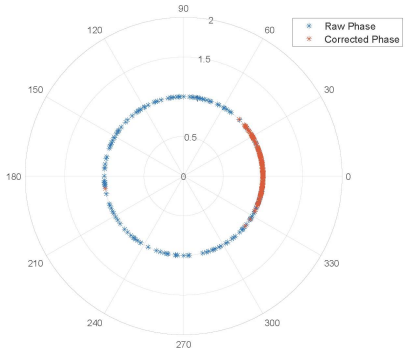

Fig. 6. Raw vs corrected phase of a CSI waveform that consists of 200 consecutively received packets

C. Wavelet Denoising

To mitigate the profound noise in CSI samples, we apply wavelet denoising for each CSI sample X_{CSI}. The key principle of wavelet denoising technique is to apply wavelet transform on the original noisy data, which concentrates data features in a few large-magnitude wavelet coefficients. The small wavelet coefficients are typically noise and those coefficients that can be shrunk to remove noise, while preserving the important signal data features. After thresholding the coefficients, we apply the inverse wavelet transform to reconstruct the denoised data. In NeuralWave system, we adopt the maximal overlap discrete wavelet transform to denoise each CSI sample X_{CSI} with Donoho and Johnstone's universal threshold and level-dependent thresholding of level 2 [16]. Figure 7 shows that by applying wavelet denoising, the high-frequency noise has been removed from the CSI waveform without distorting the trends of the waveform.

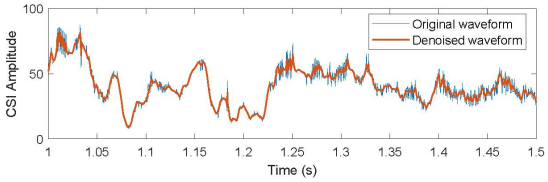

Fig. 7. Original CSI waveform vs denoised waveform

IV. Deep Convolutional Radio Network

A. Why Deep Convolutional Neural Networks?

Deep neural network uses a multilayer stack of non-linear processing layers, where each layer's output is the successive layer's input. As a special type of deep learning architecture, deep convolution neural network (ConvNet) is mainly composed by two types of layers: convolutional layers and pooling layers. The convolutional layers are organized in feature maps. Each unit of a feature map is connected to local patches in the feature maps of the previous layer through a set of weights called filter bank [17]. The filtered results are then passed through a non-linearity unit such as ReLU. The convolutional layer aims to detect local conjunctions of features from the previous layer, while the pooling layer aims to reduce the dimension of the feature maps by merging multiple feature units into one [17]. ConvNet is the well-suited feature extractor for RF biometrics, i.e., gait-modulated CSI waveforms. First, by stacking multiple non-linear layers, ConvNet can learn hierarchical features, where the lower layers extract low-level features, which are passed through higher layers to generate high-level features. The learnt high-level features are generally discriminative enough to distinguish different classes, e.g., the RF biometrics of different human subjects. Second, the features extracted by ConvNet are translation-invariant, which means they are insensitive to small transformations, distortions and translations of the input data. For example, signal variations caused by human movement are the raw gait features, which can happen at any time in a CSI waveform. It is highly desirable that the learnt features are shift-invariant, i.e., insensitive to the time shift of raw features in the CSI waveform.

B. Normalization and PCA Dimension Reduction

The preprocessed CSI sample is of large size/dimension. Such high-dimensional CSI data can easily lead to overfitting problem for ConvNet. Take the Intel 5300 WiFi device as an example. For a CSI observation of $T = 4$ seconds, it consists of $N_{tx} \times N_{rx} \times N_C = 3 \times 3 \times 30 = 270$ CSI amplitude and phase waveforms, respectively. Each waveform is made up of $N_T = 8000$ CSI measurements. Thus, each CSI sample, defined in eq. (5), consists of $270 \times 8000 \times 2 = 4,320,000$ feature elements. To reduce the dimension of the CSI data samples, principle component analysis (PCA) is adopted. PCA allows the data matrix X to be described using a small number of uncorrelated variables (the principal components), while retaining as much information as possible. As shown in Fig. 2, each CSI sample X_{CSI} has a shape of 8000×540, where the first 270 columns represent amplitude, the next 270 columns represent phase and the rows represent each time step. We reshape each 2D sample to a 1 dimensional vector containing all $8000 \times 540 = 4,320,000$ features and split our dataset into training and testing sets, i.e., X_{train} and X_{test}. We normalize the training data set so that the features are centered around 0 with a standard deviation of 1, i.e.,

$$X_{train}^{norm} = \frac{X_{train} - mean(X_{train})}{std(X_{train})} \tag{7}$$

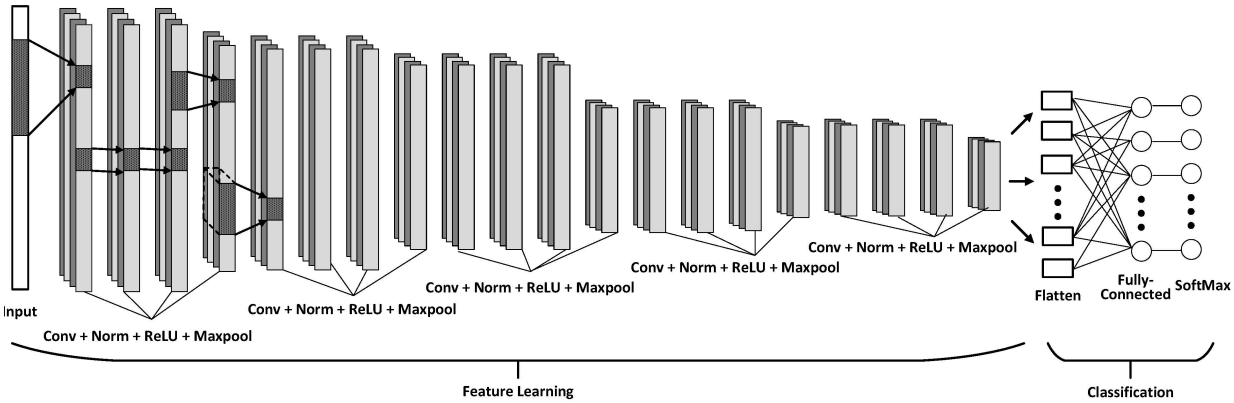

Fig. 8. RadioNet Architecture with 6 stages. The first 5 stages are for feature learning and the last stage is for classification. There are 6 types of layer in RadioNet including convolutional layer (Conv), batch normalization layer (Norm), ReLu activation layer (RelU), Maxpooling layer (Maxpool), Flatten layer, and fully-connected layer

The mean and standard deviation of the training set are used to obtain the normalized testing set X_{test}^{norm}. Ensuring normalized feature values implicitly weights all features equally in their representation, which makes convergence faster while training the network. Moreover, normalization also subtracts mean from the data set, which implicitly removes the impact of the CSI static component, i.e., $H_{static}(f)$ in eq. (2), which arises from the propagation paths reflected by the static objects in the environment, e.g., walls and furnitures.

Next, we apply PCA on the normalized training set. In particular, we first calculate the covariance matrix C of X_{train}^{std}, then compute its eigenvectors $W_{train} = [\mathbf{w}_1\ \mathbf{w}_2\ ,...,\ \mathbf{w}_{4,320,000}]$ and arrange the eigenvectors in descending order of the eigenvalues $V = (\lambda_1, \lambda_2, ..., \lambda_{4,320,000})$, where $\lambda_1 > \lambda_2 >, ..., > \lambda_{4,320,000}$. Now, we obtain the representation of X_{train}^{std} in the principal component (PC) space, i.e., $X_{train}^{PCA} = X_{train}^{std} W_{train}$, and each row of X_{train}^{PCA} can be considered as a PC-CSI sample. We take the first $k = 354$ eigenvectors that capture 95% of the total variance. This transforms the high-dimensional CSI sample with $4,320,000$ features into the low-dimensional PC-CSI sample with 354 features that contain 95% original information. We apply the same transformations on the testing set by using the eigenvectors W_{train} derived from the training set so that the testing sample also has the same number of features as the training sample. Moreover, since noise is also distributed among all eigenvectors, PCA dimension reduction can further reduce noise.

C. RadioNet Overview

Some celebrated ConvNet architectures already exist in the literature, e.g., VGG-16/-19 [12] and ResNet-101/ResNet-152 [13]. However, they are designed for computer vision applications based on 2-D images and thus are not suitable to handle 1-D CSI data samples. What is more important, these existing ConvNets have tens of millions of learnable parameters and thus can learn more complicated structures and patterns of input data. However, it makes the network more difficult to generalize, specially when our CSI training dataset is small. To address these issues, we develop a customized 1-D ConvNet, called RadioNet, which follows the generic multilayer stack design of ConvNets with fine-tuned hyperparameters, e.g., network depth, layer types, filter size, and number of filters. As shown in Fig. 8, the RadioNet architecture has 23 layers with $13,976$ learnable parameters. The specific configuration of each layer is shown in Table I. In particular, the first layer represents the input layer, which has the same size, i.e., 354×1, as the preprocessed CSI sample. The remaining 22 layers is organized into six stages. The first five convolutional stages, which are used for feature learning and extraction, follow the same pipeline pattern that comprises one convolution layer, one batch normalization layer [18], one ReLU activation layer [19], and one maxpooling layer [20]. The last stage, which is used for user classification, consists of one flatten layer,

TABLE I
THE CONFIGURATION OF RADIONET

No.	Layer	Dimension	Filter #	Filter Size	Stride	Padding	Parameters #
0	Input	354 × 1					0
1	Convolution	354 × 16	16	17	1	SAME	288
2	Batch Normalization	354 × 16					64
3	ReLU	354 × 16					0
4	MaxPool	177 × 16		2	2		0
5	Convolution	177 × 16	16	11	1	SAME	2832
6	Batch Normalization	177 × 16					64
7	ReLU	177 × 16					0
8	MaxPool	88 × 16		2	2		0
9	Convolution	88 × 16	16	11	1	SAME	2832
10	Batch Normalization	88 × 16					64
11	ReLU	88 × 16					0
12	MaxPool	44 × 16		2	2		0
13	Convolution	44 × 16	16	7	1	SAME	1808
14	Batch Normalization	44 × 16					64
15	ReLU	44 × 16					0
16	MaxPool	22 × 16		2	2		0
17	Convolution	22 × 16	16	7	1	SAME	1808
18	Batch Normalization	22 × 16					64
19	ReLU	22 × 16					0
20	MaxPool	11 × 16		2	1		0
21	Flatten	176 × 1					0
22	Fully-Connected	24 × 1					4248
	Total Parameters	14,136					
	Trainable Parameters	13,976					
	Non-Trainable Parameters	160					

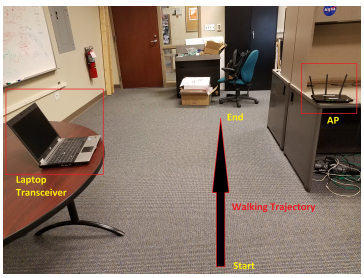

Fig. 9. NeuralWave System Setup

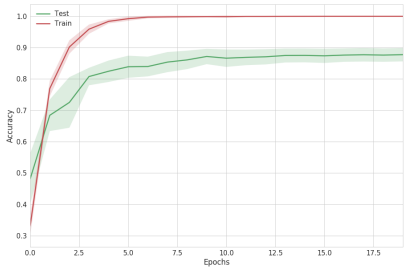

Fig. 10. Convergence of test and training accuracies

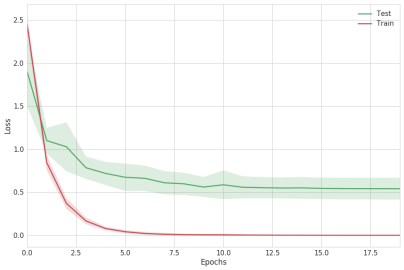

Fig. 11. Convergence of loss function for test and training datasets

one fully-connected layer, and one output layer. The flatten layer gets the feature maps of the last convolutional stage and flattens them to create a single long feature vector. This feature vector is used by the fully-connected layer to generate a score vector or logits Z and each element in Z is the score for a particular class, given a data sample i. The score vector constitutes the latent representation for each CSI data sample. The output layer implements the softmax activation, which uses exponentiation and normalization to transfer the score vector into a class probability vector $\mathbf{y}_i = [y_{i1}, y_{i2}, ..., y_{iC}]$, where C is the number of classes (number of users). Each element y_{ic} of $\mathbf{y}_i$ is the probability that a CSI sample i is predicted to be class c, i.e.,

$$y_{ic} = P(c|i) = \frac{\exp^{Z(c)}}{\sum_{j=1}^{C} \exp^{Z(j)}} \tag{8}$$

Then, the class c^* with the maximal prediction value is the predicted class for unknown CSI sample i, i.e.,

$$c^* = \underset{c \in C}{\operatorname{argmax}}\, P(c|i) \tag{9}$$

The above classifier is called softmax classifier. RadioNet is trained to maximize its classification accuracy by minimizing the cross-entropy loss function, i.e.,

$$L(t, y) = -\sum_{i=1}^{N} \sum_{c \leq C} t_{ic} log(y_{ic}) \tag{10}$$

where t_{ic} is 1 if and only if sample i belongs to class c. N is the number of samples in each training batch.

V. Experimental Validation

In this section, we present the experimental results of our proposed *NeuralWave* human identification system.

A. NeuralWave System Setup

The data collection campaign was conducted in a typical indoor laboratory environment as shown in Fig. (9). The experimental NeuralWave system consists of two WiFi devices. In particular, the WiFi transmitter is based on Gateworks GW5400 single board computer, which is equipped with a compex WLE900VX-I 802.11a/b/g/n/ac WiFi card and 3x3 MIMO antennas. The WiFi receiver is a HP 6930p laptop with Intel 5300 3x3 MIMO WiFi card. The transmitter node runs OpenWRT embedded linux operating system and operates in 5 GHz frequency band with 40MHz bandwidth. The data rate of the transmitter is fixed at 2000 packets/second. The receiver laptop runs Ubuntu 16.04 operating system with kernel version 3.34 with a modified Intel WiFi driver [21], which is capable of collecting the CSI measurements from the Intel 5300 WiFi card. Each CSI measurement consists of 30 subcarrier groups covering 114 OFDM subcarriers across 40Mhz channel.

B. Datasets and Training Methodology

During data collection campaign, we recruited 24 volunteers with the age range bettween 21 and 40. As shown in the Fig. 9, each participant was asked to walk in his/her natural way between the start and end points, which is counted as one walk instance. During each walk instance, the laptop records the received CSI measurements to generate one CSI data sample. We collected 40 CSI samples for each human subject. The CSI dataset is randomly split into a training dataset and validation dataset (70% / 30%) and Monte-Carlo cross-validation is performed, where the results are averaged over 20 splits. We divide the training dataset into mini-batches that are used to update network parameters. Each mini-batch contains 32 training samples. Adam optimizer [22] is applied to train RadioNet with an initial learning rate of 0.001 with a decay rate of 0.0001 for 20 epochs.

C. Evaluation Results

1) User Identification Accuracy: We first evaluate the performance of our system on user identification accuracy (i.e., test accuracy), which is the percentage of correct identity predications over all test samples. Formally, it is defined as

$$Test\ Accuracy = \frac{\text{Total number of correct predications}}{\text{Total number of test samples}}$$

Since the test samples are not seen by the network during training, the test accuracy measures the generalization capability and predictive power of the trained network. Besides test accuracy, training accuracy measures the descriptive power of the network, which indicates how well the network fits the training dataset and estimates the likely performance of the network on out-of-sample (unseen) data. Formally, it is defined as

$$Training\ Accuracy = \frac{\text{Total number of correct predications}}{\text{Total number of training samples}}$$

Our system achieves $87.76 \pm 2.14\%$ test identification accuracy and 100% training accuracy. Moreover, as seen in Figure 10 and 11, both classification accuracy and loss converge after 10 epochs, which means our network can be quickly trained. Note that the curves in Figure 10 and 11 are plotted with the mean accuracy or loss with the shaded areas that represent a standard deviation above and below the mean for all cross-validations

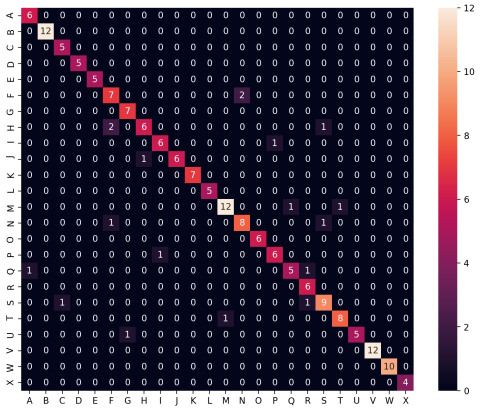

Fig. 12. Confusion matrix for the validation dataset that achieves 90% identification accuracy

2) Confusion matrix: Confusion matrix is a widely-adopted tool to summarize the prediction results on classification problems. The number of correct and incorrect predictions are summarized with count values and broken down by each class. Confusion matrix contains four performance metrics: true positives (TPs), true negatives (TNs), false positives (FPs) and false negatives (FNs). TPs are instances classified as positive by the network, which actually are positive. TNs are instances classified as negative by the network, which actually are negative. FPs are instances classified as positive, which actually are negative. FNs are instances classified as negative, which actually are positive. Each row of the matrix represents the instances in a predicted class, while each column represents the instances in an actual class. The diagonal elements of the matrix represent the TPs for each class. The TNs for a class are obtained by subtracting the true positives of this class from the total true positives of all classes. After removing the diagonal elements, the remaining elements in each row show the FPs and the remaining elements in each column count the FNs. As shown in Figure 12, NeuralWave system achieves very high TPs along with very low FPs and FNs for almost every user.

VI. Conclusion

In this paper, we developed an user identification system based on gait biometrics extracted from WiFi signals. Fundamentally different from existing solutions that rely on hand-crafted features and shallow classifiers, our system exploited deep learning to automatically extract the salient features from gait-modulated CSI waveforms, which are discriminative enough to distinguish one person from another. In particular, a 23-layer 1-D deep convolutional neural network, called RadioNet, was proposed, whose hyperparameters were fine-tuned to mitigate the over-fitting problem cased by the high-dimensional CSI data samples. As a result, for a group of 24 human subjects, our system can achieve $87.76 \pm 2.14\%$ identification accuracy. For the future work, we will improve the identification accuracy by exploiting different data pre-processing schemes, autoencoder-based dimension reduction approaches, and different ConvNet architectures (such as deep residual networks). Moreover, we will recruit more volunteers to increase the size of training data set.

References

[1] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proc. ACM Ubicomp*, 2016, p. 363373.

[2] Y. Zeng, P. Pathak, and P. Mohapatra, "Wiwho: Wifi-based person identification in smart spaces," in *Proc. IEEE IPSN*, 2016.

[3] J. Zhang, B. Wei, W. Hu, and S. Kenhere, "Wifi-id: Human identification using wifi signal," in *Proc. IEEE DCOSS*, 2016.

[4] C. Vaughan, B. Davis, and J. OCononor, *Dynamics of human gait*, 2nd ed. South Africa: Kiboho Publishers, 1999.

[5] M. P. Murray, A. B. Drought, and R. C. Kory, "Walking patterns of normal men," *Journal of Bone and Joint Surgery*, 1964.

[6] M. P. Murray, "Gait as a total pattern of movement," *American Journal of PhysicalMedicine*, pp. 290 – 332, 1967.

[7] J. E. Cutting and L. T. Kozlowski, "Recognizing friends by their walk: Gait perception without familiarity cues," *Bulletin of the psychonomic society*, 1977.

[8] D. Gafurov, E. Snekkenes, and P. Bours, "Spoof attacks on gait authentication system," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 491 – 502, 2007.

[9] S. Wang, J. Zhang, and L. Tong, "Walk the walk: Attacking gait biometrics by imitation," in *Proc. International Conference on Information Security*, 2010, pp. 361 – 380.

[10] G. Melia, *Electromagnetic Absorption by the Human Body from 1-15 GHz*. University of York, 2013. [Online]. Available: https://books.google.com/books?id=rFn6oAEACAAJ

[11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vision*, vol. 115, no. 3, pp. 211–252, Dec. 2015. [Online]. Available: http://dx.doi.org/10.1007/s11263-015-0816-y

[12] "Vgg16," available: https://gist.github.com/ksimonyan/211839e770f7b538e2d8#file-readme-md.

[13] "Resnet," available: https://github.com/KaimingHe/deep-residual-networks.

[14] P. Schmitt, J. Mandel, and M. Guedj, "A comparison of six methods for missing data imputation," vol. 06, Jan. 2015.

[15] X. Wang, L. Gao, and S. Mao, "Csi phase fingerprinting for indoor localization with a deep learning approach," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1113–1123, Dec 2016.

[16] D. Donoho and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425 – 455, 2007.

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–44, May 2015.

[18] S. Ioffe and C. Szegedy, "Gait recognition using wifi signals," in *Proc.ICML 15*, 2015.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of NIPS 2012*, USA, 2012, pp. 1097–1105.

[20] D. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in *Proc. of IJCAI 2011*, 2011.

[21] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, pp. 53–53, Jan. 2011. [Online]. Available: http://doi.acm.org/10.1145/1925861.1925870

[22] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. International Conference for Learning Representations (ICLR)*, 2015.