IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 3, 2019, accepted November 14, 2019, date of publication November 20, 2019, date of current version December 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2954651

WiRIM: Resolution Improving Mechanism for Human Sensing With Commodity Wi-Fi

XINBIN SHEN, LINGCHAO GUO, ZHAOMING LU, XIANGMING WEN, AND ZHIHONG HE

Beijing Key Laboratory of Network System Architecture and Convergence, Beijing University of Posts and Telecommunications, Beijing 100876, China
Beijing Laboratory of Advanced Information Networks, Beijing University of Posts and Telecommunications, Beijing 100876, China
School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Zhaoming Lu (lzy0372@bupt.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0810204, and in part by the National Natural Science Foundation of China under Grant 61671073.

ABSTRACT The growing physical (PHY) layer capabilities of Wi-Fi have made it possible to use Wi-Fi signals for both communication and human sensing. Wi-Fi channel state information (CSI) in PHY layer can be obtained from commodity Wi-Fi devices. As CSI can detect the minute environment changes that alter signal propagation, it is thus capable of capturing the subtle human activities to provide cost-effective and easy-to-use human sensing. However, the limited bandwidth of each individual Wi-Fi channel fundamentally constrains the resolution of signals, resulting in poor performance of human sensing. In this paper, we present WiRIM, a resolution improving mechanism for Wi-Fi based human sensing. We design a channel switching and aggregation algorithm to extend the effective bandwidth of commodity Wi-Fi signals and improve the performance and efficiency of human sensing applications. With aggregated CSI, WiRIM constructs feature images which contain rich frequency, temporal and spatial characteristics, and then uses a deep learning method to process the extracted features. We propose a cross-location human activity recognition (CLHAR) scenario as a case study. The CLHAR scenario requires a high enough resolution of the Wi-Fi signals to accurately recognize different activities under the interference of tiny changes in human location. The experiments demonstrate the generality and effectiveness of the proposed mechanism.

INDEX TERMS Human sensing, resolution improvement, Wi-Fi, CSI, cross-location human activity recognition.

I. INTRODUCTION

With the rapid development of Internet of Things (IoT), such IoT applications as smart home and motion-controlled video games have attracted increasing attention. Recently, Wi-Fi based human sensing has become a promising technique in human activity recognition [1]–[15], objects tracking [16]–[21], physiological indicators detection [22]–[27] and human detection [28], etc. This prosperity benefits from some properties of Wi-Fi, including its low cost, extensive use and device-free experience. The proposed DP-DFLAGR system in [29], [30] estimates target's location, activity, and gesture based on the received signal strength (RSS). The designed DFLAR system in [31] collects channel state information (CSI) and estimates the location and activity of a person using a machine learning approach. Unfortunately, due to the limited resolution of current Wi-Fi signals, the above systems cannot always achieve good performance especially when performing some activities or at some locations.

The basic principle of human sensing based on Wi-Fi signals is: due to reflection, scattering and diffraction introduced by objects in space, Wi-Fi signals transmit through a multipath channel from the transmitter to the receiver [32]. When a person is around the Wi-Fi devices, his body will affect the travel-through signals. The received signals contain the fine-grained physical (PHY) layer measurement of CSI, which characterizes the propagation space. Using CSI obtained from commodity Wi-Fi devices [33], power delay profile (PDP) can be calculated through IFFT (Inverse Fast Fourier Transform) [6], [34]–[36]. Through analyzing a high resolution PDP, tiny human activities can be detected by

The associate editor coordinating the review of this manuscript and approving it for publication was Guan Gui.

differentiating subtle multipath channel changes [6], [35], [37], [38].

However, the time resolution of PDP is determined by the bandwidth of Wi-Fi signals. Currently, the bandwidth of a single channel in commodity Wi-Fi NICs (network interface cards) only spans 20 *MHz*, which provides a time resolution of 50 *nses* for distinguishing different signal propagation paths in PDP (these will be explained in detail in III-A). So the resolution provided by the limit bandwidth [38], [39] limits the performance of Wi-Fi based human sensing. For higher performance and efficiency, how to improve the resolution of signals with commodity Wi-Fi devices is the basic problem to be solved.

To address the above challenges, we present WiRIM, a resolution improving mechanism for human sensing with commodity Wi-Fi. We design a channel switching and aggregation algorithm to realize fast switch of adjacent channels in Wi-Fi NICs and aggregate these channels into an extended channel. CSI from the extended channel has a wider bandwidth and provides higher resolution for human sensing. Leveraging both CSI amplitude and phase, we construct feature images, which contain rich frequency, temporal and spatial characteristics of the aggregated CSI. Finally, WiRIM merges the features into a deep learning module for human sensing.

We propose a cross-location human activity recognition (CLHAR) scenario as a case study. Since human activity and location information is particularly important in human sensing, the CLHAR scenario has important practical significance. Existing studies use some techniques to minimize the influence of location difference on activity-related feature extraction, so as to realize Wi-Fi based human sensing. However, this approach is limited in CLHAR scenarios where the location difference is tiny.

Because of this, we conduct two controlled experiments. Firstly, we use CSI to estimate human activity and location information. The results of WIRIM using the aggregated channel are compared with those of WIRIM using a single channel. Secondly, we use CSI to recognize human activities when there is a slight difference in human location. And we compared WIRIM with a benchmark approach [13] of human activity recognition (HAR).

Experiment results demonstrate the effectiveness and generality of the proposed mechanism.

The main contributions of this paper are summarized as follows:

1) We prove that the time resolution of PDP is limited by Wi-Fi channel bandwidth, and the high resolution provided by CSI can detect minute human activities through distinguishing subtle changes in multipath channels.
2) We present WiRIM, a resolution improving mechanism that is universally applicable to Wi-Fi based human sensing. And we implement WiRIM with commodity Wi-Fi.
3) We use a cross-location human activity recognition system as a case study and conduct a series of comparison experiments. Experimental results verify the effectiveness and generality of WiRIM.

The rest of the paper is organized as follows. Related work and some preliminaries are discussed in Section II and Section III, respectively. Then, the mechanism design of WiRIM is elaborated in Section IV. The detailed experiment procedures are presented in Section V and the results are analyzed in Section VI followed by a conclusion in Section VII.

II. RELATED WORK

A. HUMAN SENSING WITH COMMODITY WI-FI

There are numerous human sensing applications with commodity Wi-Fi.

Here we first introduce some representative human activity recognition applications based on Wi-Fi. CARM [4] proposes a CSI-activity model for a CSI based human activity recognition. WiDance [5] extracts complete information of motion-induced Doppler shifts and prototypes a contactless dance-pad exergame. WiHear [6] analyzes fine-grained radio reflections from mouth movements and detects pre-defined vocabulary. WiKey [7] is a Wi-Fi based keystroke recognition system with an accuracy of 93.5%. WiFall [8] can achieve fall detection with high accuracy. Wi-Chase [10] fully utilizes all available subcarriers of the Wi-Fi signals and recognizes corresponding human activities. PCA-Kalman [11] is a method for detecting indoor activity using ubiquitous Wi-Fi. PCA-Kalman uses the principal component analysis (PCA) algorithm to reduce the dimension of the data, Kalman filter algorithm to remove some outlier values, and support vector machine (SVM) algorithm to classify data. WiFit [12] recognizes the exercise type and count the number of exercise for diverse population even in different environments. Reference [13] proposes an attention based bi-directional long short term memory (ABLSTM) for Wi-Fi based HAR and achieves the best recognition performance for all activities when compared with other approaches. WiAct [14] utilizes Doppler shift correlation value and Extreme Learning Machine (ELM) for activity data classification. Wri-Fi [15] detects the writing action persistently by the Fast Fourier Transform (FFT) based energy indicator and realizes the air-write recognition.

These HAR systems are all based on CSI data from a single channel, use some technologies to eliminate the influence of human location and use statistical methods to extract features reflecting activity characteristics. However, tiny changes in activities, as well as tiny changes in a person's location can have an impact on Wi-Fi signals. Especially in our proposed CLHAR scenario, tiny changes in human location will interfere with signals and affect the extraction of human activity features. If Wi-Fi sensing technology is to be applied in the real world, such interference cannot be ignored. This requires a high enough resolution of the Wi-Fi signals to

accurately recognize different activities, even if disturbed by small changes in human locations.

Meanwhile, there are still many advanced applications in Wi-Fi based human sensing. Widar [16], [17] enables tracking of both human locations and velocities with CSI dynamics. Widar2.0 [18] devises an efficient algorithm for estimating Angle-of-Arrival (AoA), Time-of-Flight (ToF), and Doppler shifts together and designs a pipeline to translate the erroneous raw parameters into precise locations. IndoTrack [19] estimates the target speed and position through spatial-temporal Doppler and AoA information, and determines the absolute trajectory of the target. MobileRF presented in [20] can estimate the location of a person via RSS from radio frequency (RF). References [22]–[25] utilize the First Fresnel Zone (FFZ) to achieve higher than 98% accuracy of respiration rate monitoring. WiVit [26] is a Wi-Fi based sensing platform hosted on Wi-Fi devices. WiVit accurately captures human vitality information and achieves 98% accuracy of vitality detection. Reference [28] exploits the MUSIC algorithm to estimate the phase differences caused by a moving person and proposes a Wi-Fi based human detection system.

However, these systems have to face the challenge of a single channel when using commodity Wi-Fi devices. In contrast, we address this challenge by a resolution improving mechanism for human sensing with commodity Wi-Fi.

B. METHODS TO IMPROVE THE RESOLUTION OF WIRELESS SIGNAL

Prior works have also focused on improving the resolution of wireless signals. Leveraging the spatial and frequency diversity in MIMO transceiver, [40] introduces MIMO smoothing and shows that MIMO smoothing can increase the accuracy of multipath resolution. SWAN presented in [41] builds an array of stitched antennas extended from the radio chains of commodity Wi-Fi and increases Wi-Fi throughput by more than 30%, resulting in much finer resolution and higher accuracy. ToneTrack in [42] combines data from a WARP hardware radio platform as it hops across different channels in a frequency band. However, the systems above require customized hardware with high cost and special equipment, which limits the practical use of IoT applications. Compared with other prior works using special equipment, we only need commodity Wi-Fi devices, which makes Wi-Fi based human sensing more convenient and applicable.

C. APPLICATIONS USING MULTIPLE CHANNELS

A few papers use multiple channels to locate Wi-Fi APs. A recent paper [43] performs channel impulse response (CIR) estimation by splicing measured CSI over multiple Wi-Fi bands. ToneTrack [42] is an indoor location system implemented on the WARP hardware radio platform. ToneTrack hops across different channels and uses six of the platforms served as APs to localize Wi-Fi clients. Splicer [44] is a software-based system that achieves single-AP localization by splicing the CSI measurements from multiple Wi-Fi frequency bands. Leveraging different Wi-Fi bands, Chronos [45] measures the absolute ToF and enables a single Wi-Fi AP to localize another. However, unlike WIRIM which improves the performance and efficiency of human sensing applications, these papers focus their attention on Wi-Fi device-to-device localization.

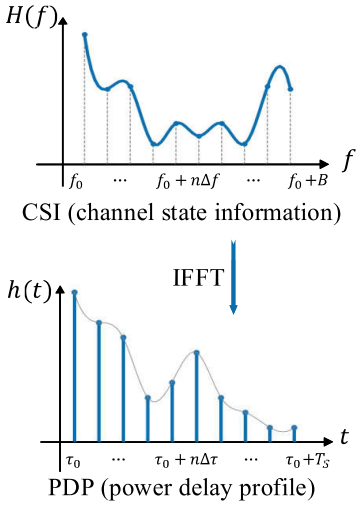


FIGURE 1. The time domain PDP is calculated by the frequency domain CSI through IFFT.

III. PRELIMINARIES

A. PRINCIPLE OF RESOLUTION IMPROVING

The channel frequency response (CFR) characterizes the effects of multipath in the frequency domain. CSI is a set of the discrete samples of CFR:

$$H(f) = \sum_{n=1}^{N} \alpha_n e^{-j2\pi f \tau_n}, \tag{1}$$

where n denotes the sequence number of total N multipaths, α_n and τ_n are the amplitude and propagation time delay corresponding to the n *th* path. With the channel bandwidth B and the sampling resolution Δf in the frequency domain, we can deduce the number of CFR samples:

$$N = B/\Delta f. \tag{2}$$

As shown in FIGURE 1, the time domain PDP is calculated by the frequency domain CSI through IFFT. The corresponding PDP of CSI is:

$$h(t) = \sum_{n=1}^{N} \alpha_n \delta\left(t - \tau_n\right), \tag{3}$$

where $\delta(\cdot)$ is the Dirac delta function, Since PDP is able to characterize multipath channel features, it is widely used in the field of human sensing based on Wi-Fi. In an indoor multipath environment, a person's activities can be detected more accurately by using a finer PDP with a much higher time resolution. In IFFT, the time resolution is defined as:

$$\Delta\tau = T_s/N, \tag{4}$$

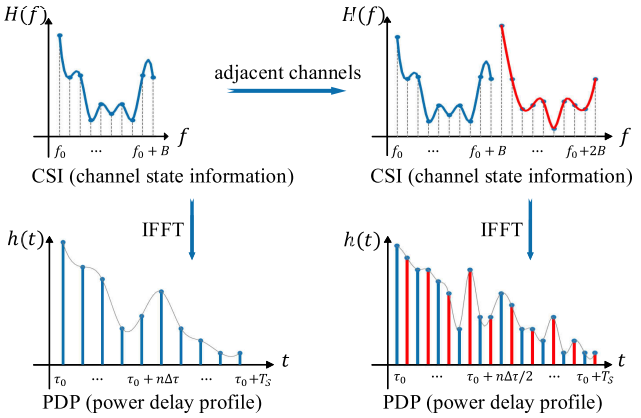


FIGURE 2. The time resolution of the PDP obtained from CSI is determined by the bandwidth of Wi-Fi signals.

where T_s is the window length of the time series and N is the number of data points used in IFFT. The smaller the value of $\Delta\tau$, the higher the time resolution. The sample frequency Δf can be expressed as [46], [47]:

$$\Delta f = 1/T_s. \tag{5}$$

Refer to (2), (4) and (5), we have:

$$\Delta\tau = 1/B. \tag{6}$$

Hence, the time resolution of the PDP obtained from CSI is determined by the bandwidth of Wi-Fi signals. As shown in FIGURE 2, it makes clear that a wider bandwidth leads to a higher PDP resolution, which is easier to distinguish subtle multipath channel changes caused by minute activities [6], [37]. Therefore, a large channel bandwidth is crucial for fine-grained Wi-Fi human sensing applications. However, on one hand, single channel in Wi-Fi NICs only spans 20 *MHz*, which limits the performance of human sensing, on the other hand, the total bandwidth of the Wi-Fi signals is large enough. When switching multiple channels rapidly, we can obtain CSI from multiple adjacent Wi-Fi channels in a very short time interval. Then, by aggregating CSI from these channels, the Wi-Fi signal bandwidth is extended to the entire bandwidth of these adjacent channels, as shown in FIGURE 3.

As the signal bandwidth is extended, the resolution of Wi-Fi signals can be accordingly improved.

B. ANGLE-OF-ARRIVAL ESTIMATION

CSI contains the amplitude and phase. The activities of humans change the multipath characteristic of the wireless channel, which is reflected on the amplitude and phase of CSI. Thus, the amplitude and phase can be used to extract the activity-related features. As shown in FIGURE 4, the CSI amplitudes are in different ranges when the environment is static and when the user is performing activities, respectively. Therefore, different thresholds of CSI amplitudes, such as variance, mean absolute deviation (MAD),

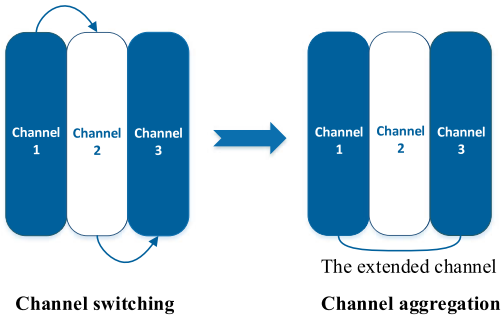


FIGURE 3. Channel switching and aggregation algorithm.

finite difference, can be used to determine human status. Threshold-based methods are widely used in Wi-Fi based human activity recognition [6], [48]–[51]. Unlike the raw amplitudes which can be directly used, the raw phases obtained by commodity Wi-Fi devices distribute randomly as FIGURE 5. In real-world Wi-Fi systems, the received phases are often deteriorated by some sources such as hardware and software errors.

According to [47], [52], [53],the measured phase $\angle\widehat{CSI}_i$ of the *i th* subcarrier can be expressed as:

$$\angle\widehat{CSI}_i = \angle CSI_i + 2\pi T_s\left(\Delta f + i\frac{\zeta}{T_u}\right) + 2\pi i\frac{\tau_0}{N} + \gamma + n, \tag{7}$$

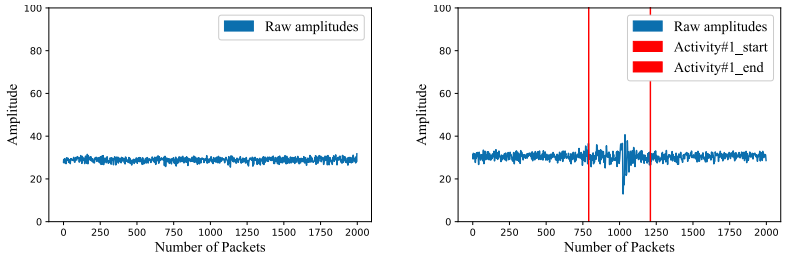
where $\angle CSI_i$ is the true phase of the *i th* subcarrier, T_s and T_u are the total length of the OFDM symbol and the length of the data symbol, respectively, Δf is the carrier frequency offset (CFO), ζ is the sampling frequency offset (SFO) (formulated in Equation (8)), T and T' are the sampling time at the transmitter and receiver, respectively, τ_0 is a time shift in the time-domain samples, N is the number of data points used in IFFT, γ is the oscillator phase offset in phase locked loops (PLL), n is the Additive White Gaussian Noise (AWGN) seen by the receiver.

$$\zeta = (T' - T)/T. \tag{8}$$

Three antennas of a Wi-Fi NIC have the same carrier and sampling clock frequency offset, sample clock time offset [53]. Thus the phase difference of the *i th* subcarrier can be approximated as:

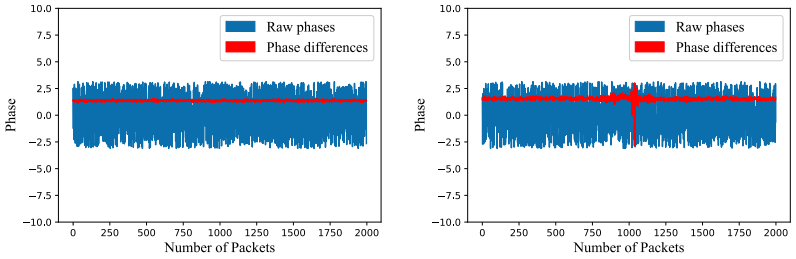
$$\Delta\angle\widehat{CSI}_i = \Delta\angle CSI_i + \Delta\gamma + \Delta n, \tag{9}$$

where $\Delta\gamma$ is a difference of constant phase shifts, Δn is Gaussian noise. According to [53], [54], $\Delta\gamma$ is constant over time between antennas on one NIC, which does not change the phase information of human activities. Compared to the phases, the phase differences become more stable by removing CFO, SFO and time offset. As shown in FIGURE 5, on one hand, the phase differences maintain on a stable level in the static environment, on the other hand, fluctuations appear on phase differences when there are human activities. Therefore, the raw phases can be calibrated by phase differences, verifying our derivation. FIGURE 6 plots the



(a) Static environment. (b) Human performs an activity.

FIGURE 4. CSI amplitudes affected by human activities: static environment, human performs an activity. And we can use the variance of the CSI amplitudes to detect the beginning and end of each activity.



(a) Static environment. (b) Human performs an activity.

FIGURE 5. CSI phases and phase differences affected by human activities: static environment, human performs an activity.

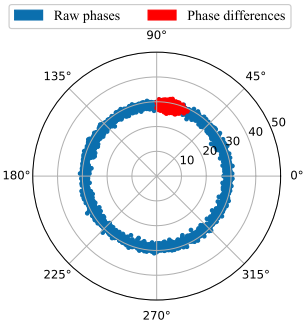


FIGURE 6. CSI phases and phase differences in the polar coordinate.

raw CSI values against polar coordinates, which includes the amplitude, phase and phase difference information. The phases are distributed between 0° and 360°, while phase differences are distributed between 65° and 90°.

In our experiments, we set the distance between two adjacent antennas to $\lambda/2$, where λ is the wavelength. Let θ_i be the AoA of the *i th* subcarrier, representing the angle at which the signal of the *i th* subcarrier arrives at the antenna. As shown in FIGURE 7, before reaching the second antenna, the Wi-Fi signal travels an extra distance, which can be approximated as:

$$\Delta d_i = \lambda/2 \cos(\theta_i). \tag{10}$$

The distance difference Δd_i will result into a phase difference:

$$\Delta \angle \widehat{CSI}_i = 2\pi \Delta d_i/\lambda. \tag{11}$$

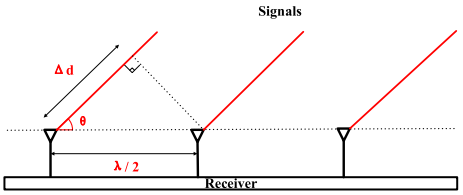


FIGURE 7. The signal travels an extra distance, which can be calculated by the AoA.

We calculate the AoA value of subcarrier *i*:

$$\theta_i = \arccos(\Delta \angle \widehat{CSI}_i/\pi). \tag{12}$$

The AoA values of the multiple subcarriers contain rich information of the signals in multipath channels, including both information from static paths in the environment and information caused by human activities. Although the phase of CSI is affected by several error sources, the phase difference is relatively stable, we can use AoA values obtained from the phase differences to accurately distinguish locations.

IV. MECHANISM DESIGN

The architecture of WiRIM mainly consists of the following modules: a channel switching and aggregation algorithm which collects CSI with higher resolution, a feature extraction module and a deep learning module, as shown in FIGURE 8.

A. CHANNEL SWITCHING AND AGGREGATION ALGORITHM

By leveraging the Intel 5300 NICs with modified driver and firmware [55], CSI is extracted from the received packets.

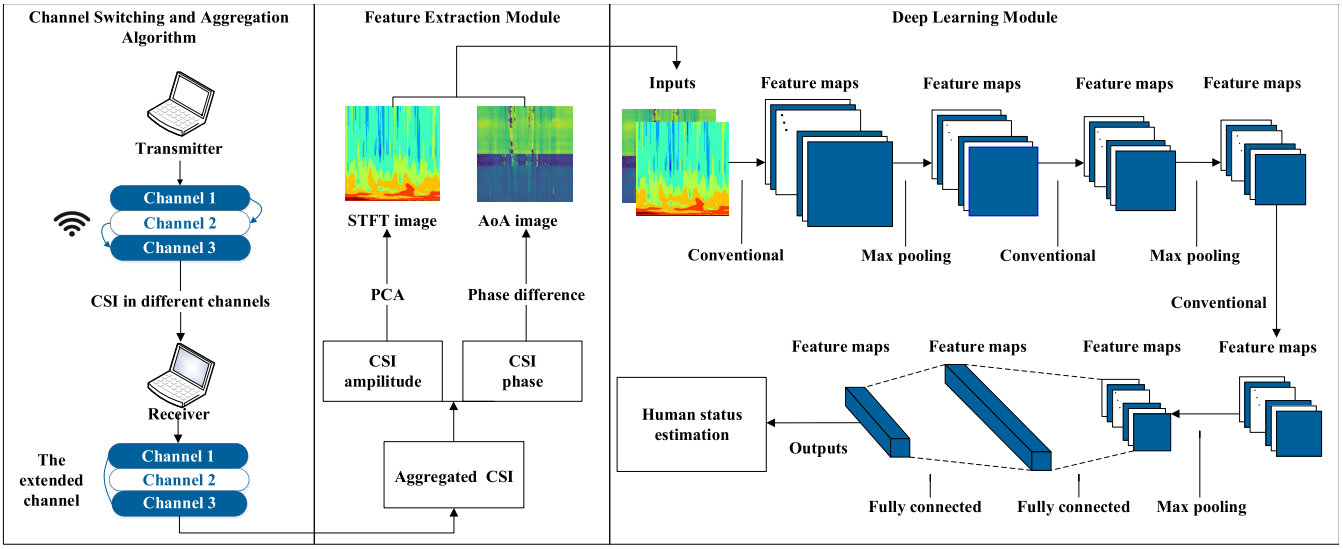


FIGURE 8. An illustration of WiRIM, the resolution improving mechanism for Wi-Fi based human sensing.

Algorithm 1 Channel Switching and Aggregation Algorithm

Input:

Number of packets len

Number of channels n

Output:

Aggregated CSI $C_1, C_2, \ldots, C_{len/n}$ from the extended channel

1: Select n adjacent channels $h_1, h_2, \ldots, h_n$;
2: Add $h_1, h_2, \ldots, h_n$ to *channelList*;
3: Both transmitter TX and receiver RX switch to the first channel h_1;
4: **for** each $i \in [1, len]$ **do**
5: $j = i \,\%\, n + 1$;
6: TX transmits a packet p_i;
7: **while** Both TX and RX is listening to the socket **do**
8: **if** RX receives p_i **then**
9: RX replies with an acknowledgement ACK;
10: RX switches to the next channel h_j;
11: break;
12: **end if**
13: **if** TX does not receive ACK with given time-out duration 1 ms **then**
14: TX retransmits the packet p_i;
15: **end if**
16: **end while**
17: TX switches to the next channel h_j;
18: **if** $i \,\%\, n == 0$ **then**
19: Aggregate CSI $c_1, c_2, \ldots, c_n$ from $h_1, h_2, \ldots, h_n$, and obtain the aggregated CSI $C_{i/n}$;
20: **end if**
21: **end for**
22: **return** All aggregated CSI $C_1, C_2, \ldots, C_{len/n}$;

As aforementioned in III-A, CSI, which is derived from a single Wi-Fi channel with limited bandwidth, is of low quality. Thus, we design a method for switching multiple channels by modifying the iwlwifi driver and the open source IEEE 802.11 packet injection library – LORCON (Loss Of Radio CONnectivity) [56], and finally achieve multiple channels switching and packet processing in millisecond granularity. To ensure that the channel switching process does not affect the extraction of activity-related features, the switching time should not be greater than the sampling time.

In the process of channel switching, it must be ensured that the transmitter and the receiver always work synchronously in the same channel. In an unstable wireless environment, packets may be lost. Once the packet is lost, an error will occur in CSI aggregation.

To solve this problem, we have carried out a large number of verification experiments. Finally, WIRIM uses an acknowledgement and retransmission mechanism in synchronous transmission, which avoids the phenomenon of packet loss in the process of fast channel switching. As shown in Algorithm 1, before each round of data transmission, the receiver and the transmitter are in the same channel. At first, the transmitter sends a packet (step 6). After receiving the packet, the receiver replies with an acknowledgement and then switches to the next channel (step 8-12). When the transmitter receives this acknowledgement, it also switches to the next channel (step 17). However, once a packet is lost after a given time-out duration, the transmitter will retransmit the packet to ensure that the receiver receives a unique packet in each channel (step 13-15). After one cycle, we aggregate CSI from these channels (step 18-20). Based on the process, WiRIM realizes fast channel switching between receiver and transmitter. And we can obtain CSI from the extended channel with wider bandwidth.

B. FEATURE EXTRACTION MODULE

1) AoA IMAGE CONSTRUCTION

As aforementioned in III-B, AoA values are robust even in complex indoor environments due to the stability of the phase differences. Thus we can use AoA values to accurately estimate a given location. With aggregated CSI, WiRIM calculates phase differences between adjacent antennas, and then obtains AoA values calculated by Equation (12). We take time as x-axis and subcarrier as y-axis and transform AoA values on multiple subcarriers into AoA images.

2) STFT IMAGE CONSTRUCTION

In order to use CSI amplitude to classify different activities, we should first extract the representative features of different activities. WiRIM first applies PCA to the CSI streams so that we can capture all relevant human activities and greatly reduce the data dimensions [17]. WiRIM uses the first principal component for further processing, as it clearly captures human activities. Given the correlation between the frequency and the reflected path length change rate caused by human activities, we use STFT (Short-Time Fourier Transform) to extract the temporal-frequency features [57] from the first principal component. Based on this, WiRIM extracts the STFT spectrum of each activity.

AoA values are stable enough for a given location, and STFT can well distinguish the temporal-frequency characteristics of different activities. By constructing AoA spectrum images and STFT spectrum images, we obtain rich frequency, temporal and spatial characteristics of the aggregated CSI.

C. DEEP LEARNING MODULE

WiRIM uses a CNN (Convolutional Neural Network) [58] to learn the recognition features from the constructed images in IV-B. The proposed deep learning model consists of several convolution layers and maximum pooling layers, as well as multiple fully connected layers.

1) CONVOLUTION LAYER

For the convolution layer, we extract time-frequency features from the STFT spectrum images and extract time-space features from the AoA spectrum images. Each convolution layer consists of multiple convolution filters with different weights and biases. The filter weights of the convolution layer are connected to local areas in the previous layer's feature maps. The convolution layer function in WIRIM can be expressed as:

$$x_i^l = f\left(\sum_{j\in M_i} x_j^{l-1} * w_{ij}^l + b_i^l\right), \tag{13}$$

where x_i^l is the i th feature map generated in layer l, M_i is the set of feature maps in layer $l-1$, w_{ij}^l is the filter weight to generate x_i^l, b_i^l is the bias of x_i^l, $f(\cdot)$ is the rectified linear unit (ReLU) function [59]. It can be inferred from (13) that the operation of a feature map in the convolution layer is a discrete convolution. Through the convolution layer, local features can be obtained by sharing the same weights at different locations of the previous layer's feature maps.

2) MAX POOLING LAYER

After obtaining local features from the convolution layer, WIRIM uses the max pooling layer to merge similar features into one. The max pooling layer function can be written as:

$$x_i^l = \max x_i^{l-1} u(n, n), \tag{14}$$

where x_i^l is the j th feature map generated in layer l, $u(n, n)$ is a window function. As can be seen, a max pooling layer computes the maximum value within the pooling window from the feature maps generated by the convolution layer. To be specific, the max pooling layer achieves spatial invariance by reducing the resolution of the feature maps. Because of this, WIRIM can use the max pooling function to effectively optimize the memory usage and reduce the training time.

3) FULLY CONNECTED LAYER

At the end of the WIRIM, there are two fully connected layers. The first fully connected layer conducts feature dimension reduction and the second one performs the classification task. The fully connected layer consists of multiple perceptrons, which represent the nonlinear relationship between the weighted sum of inputs and outputs. The first fully connected layer function can be expressed as:

$$X^{L-1} = f\left(Dense(X^{L-2})W^{L-1} + B^{L-1}\right), \tag{15}$$

where L is the number of layers of the proposed CNN, X^{L-1} is the feature maps in layer $L-1$, $Dense(\cdot)$ compresses the outputs of the previous layer into the inputs of the fully connected layer, W^{L-1} and B^{L-1} are the weights and biases of the perceptrons in layer $L-1$, respectively.

The last fully connected layer has a node for each class and the outputs of the nodes can be expressed as:

$$Z^L = X^{L-1}W^L + B^L. \tag{16}$$

The fully connected layer then uses a softmax activation function to output the probability value of each class (Equation (17)). The class with the highest probability is estimated as the correct class.

$$y_i = \frac{e^{z_i}}{\sum_{z_k\in Z^L} e^{z_k}}. \tag{17}$$

Here y_i is the probability of class i, which is the output of WIRIM, z_k indicates the weighted sum of inputs to the k th unit in the output layer. WIRIM then uses a cross entropy loss function, defined as:

$$C = -\frac{1}{n}\sum_{i=1}^{n}[(\hat{y}_i \log(y_i) + (1-\hat{y}_i)\log(1-y_i)], \tag{18}$$

where $\hat{y}_i$ is the true value of i th class, which can only be 0 or 1. The higher the accuracy of the model, the smaller the value of the loss function. By minimizing the value of the

loss function with the BP (error Back Propagation) algorithm, we optimize the weights and biases with the Adam (Adaptive Moment Estimation) algorithm [60], [61]. Adam algorithm can compute adaptive learning rates for different parameters to avoid overfitting. In addition, dropout mechanism is also used in WIRIM to reduce the risk of overfitting. Note that dropout has proven to be a very efficient way to reduce overfitting by preventing complex co-adaptations on training data [62], [63].

V. CASE STUDY

When a person locates at different locations or performs different activities, the influence of the person on Wi-Fi signals will be different. Since this paper focuses on the impact of resolution improving mechanism on Wi-Fi based human sensing, it is of great importance whether the system using the proposed mechanism can accurately estimate the human activity and location, even if the distance between these selected locations is small (0.6 m). Based on this, we propose a particular scenario: a cross-location human activity recognition scenario. Because the distances between these locations are small in this scenario, higher Wi-Fi signal resolution is required. Otherwise, low resolution makes it impossible for the classifier to extract complete activity-related signal features. Therefore, this scenario is of great practical significance and more challenging to HAR.

We conduct two controlled experiments. Firstly, we verify the accuracy of human activity and location estimation by using CSI data sets in the aggregated channel and in a single channel, respectively. Secondly, we use CSI to recognize human activities when there is a slight difference in human location. And we compared WIRIM with a benchmark approach of Wi-Fi based HAR.

A. EXPERIMENTAL SETUP

In our experiments, we use two mini-PCs as the transmitter and receiver, both equipped with the Intel Wireless Link 5300 NIC. The transmitter uses one antenna while the receiver is equipped with three antennas. Given the reported subcarrier number as 30, then $3 \times 1 \times 30$ CSI streams are captured in one channel. As aforementioned in IV-A, we implement the channel switching and aggregation algorithm based on LORCON. We use 5 GHz Wi-Fi channels with 20 MHz bandwidth and select three channels (Channel 132, Channel 136, and Channel 140). After channel switching and aggregation, we then have $3 \times 1 \times 90$ CSI streams. The switching time from one channel to another is about 2 ms. Given the channel switching overhead, the sampling rate is 200 Hz in our system.

B. DATA COLLECTION

In this particular CLHAR scenario, we collect 8405 samples in the room of size $8m \times 5m$, as shown in FIGURE 9. We select four representative adjacent locations for experiments. There are four adjacent locations which labeled as the star shape in FIGURE 9, and the target performs one of five activities

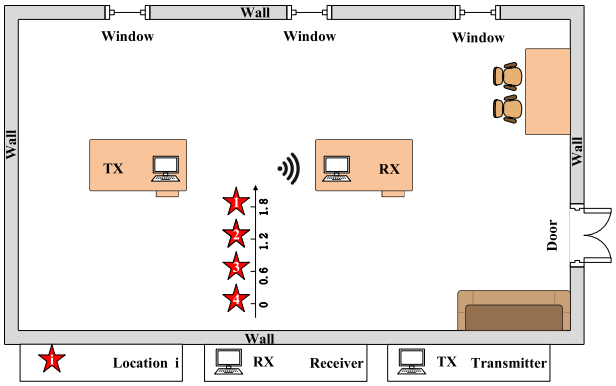


FIGURE 9. A cross-location human activity recognition scenario with four representative adjacent locations.

TABLE 1. Number of samples collected in the system with the resolution improving mechanism.

	Walk	Run	Stand	Sit	Pick
Location1	500	500	500	300	302
Location2	500	499	500	300	302
Location3	499	498	500	300	302
Location4	500	500	500	300	303

TABLE 2. Number of samples collected in the system using a single channel, as most off-the-shelf studies do.

	Walk	Run	Stand	Sit	Pick
Location1	499	499	499	301	301
Location2	499	499	500	301	301
Location3	499	500	500	301	301
Location4	500	500	500	301	301

continuously for one hour at each location. Because ''Walk'', ''Run'' and ''Stand'' activities are continuous, we divide a long sequence of packets into multiple sequences. For ''Sit'' and ''Pick'' activities, we first need to use the variance of the CSI amplitudes to detect the beginning and end of each activity. The number of samples for different activities at different locations is described in TABLE 1. We randomly select 100 samples from each label, and then use the 2000 samples as the test sets. The remaining 6405 are used as the training sets.

To show our effectiveness of the channel switching and aggregation module, we keep the experimental environment unchanged and then collect CSI from only a single channel, as most off-the-shelf studies do, and carry out the same data processing and model training. TABLE 2 describes the number of samples collected in the system using a single channel.

C. IMAGE CONSTRUCTION

We calculate 90 phase differences between antenna 1 and 2, and 90 phase differences between antenna 2 and 3. Thus, we obtain 180 AoA values in each packet. When we take

TABLE 3. Architecture of the CNN for WIRIM.

Layer	Name	Size	Activation
1	Convolution layer	$5 \times 5(16)$	ReLU
2	Max pooling layer	2×2	None
3	Convolution layer	$5 \times 5(32)$	ReLU
4	Max pooling layer	2×2	None
5	Convolution layer	$5 \times 5(64)$	ReLU
6	Max pooling layer	2×2	None
7	Fully connected layer	$400(1)$	ReLU
8	Fully connected layer	$20(1)$	Softmax

180 consecutive packet samples, we can construct an AoA spectrum image with the size of 180×180. With a sampling rate of 200 *Hz*, we can extract the frequency range up to 100 *Hz*. When 128 frequency components are extracted by STFT, the corresponding frequency granularity is 0.78 *Hz*. Such frequency ranges and frequency granularity are sufficient to cover common human activities. In this way, AoA spectrum images and STFT spectrum images are constructed to train our models.

D. TRAINING SETUP

We implement the deep learning module of WiRIM using tensorflow [64] and the workstation for our experiments is equipped with 48 core CPUs of Intel Xeon(R) 2.50GHz, 4 GPUs of NVIDIA GeForce GTX1080Ti and 48GB memory.

As the input data of the CNN, the constructed images are then processed in convolution layers and max pooling layers. The first convolution layer contains 16 convolution filters with size 5×5 and the max pooling layer with size 2×2 is used to obtain 16 feature maps with size 88×88. Then, by successively implementing the other two convolution layers and two max pooling layers as shown in TABLE 3, we obtain 64 feature maps of size 19×19. Through two fully connected layers, we calculate the cross entropy of every activity at every location and choose the one with the largest likelihood as the state of the human. Other parameters of the CNN for WIRIM are presented in TABLE 4.

To verify that the images from the feature extraction module can also use other deep learning method to extract activity-related features and accurately classify them, we compare the performance of CNN with that of long short term memory (LSTM) [65], which can hold temporal state information [66]. In addition, we can also prove whether the deep learning method we choose is suitable through comparison. TABLE 5 presents the parameters of the LSTM for WIRIM. We use the Adam algorithm to reduce the risk of over-fitting.

E. A BENCHMARK APPROACH OF HUMAN ACTIVITY RECOGNITION

To show the feasibility and generality of WIRIM in human sensing, we also perform a comparison with a benchmark approach of Wi-Fi based HAR. Note that the proposed ABLSTM in [13] was shown to be effective for HAR. As aforementioned above, in the particular CLHAR scenario of our case study, the distance between each location is small. This scenario is very common in the real world, but is often overlooked in Wi-Fi based activity recognition. Such human activity recognition system with close location is a difficult problem to be solved. It is not just small changes in activities, small changes in a person's location can also have an impact on Wi-Fi signals. Thus it is essential that the resolution of the system is high enough to accurately distinguish different activities under the influence of location. And that's why we use [13] as a benchmark approach of Wi-Fi based human activity recognition, and compare it with WIRIM.

TABLE 4. Parameters of the CNN for WIRIM.

Name	Value	Name	Value
Dropout rate (η)	50%	Exponential decay rates (β_1)	0.9
Learning rate (α)	0.05	Exponential decay rates (β_2)	0.999
Batch size (n)	128	Infinitesimal parameter(ϵ)	10^{-8}
Maximum epoch (e)	200		

TABLE 5. Parameters of the LSTM for WIRIM.

Name	Value	Name	Value
Number of hidden units	600	Exponential decay rates (β_1)	0.9
Learning rate (α)	0.01	Exponential decay rates (β_2)	0.999
Batch size (n)	128	Infinitesimal parameter(ϵ)	10^{-8}
Maximum epoch (e)	200		

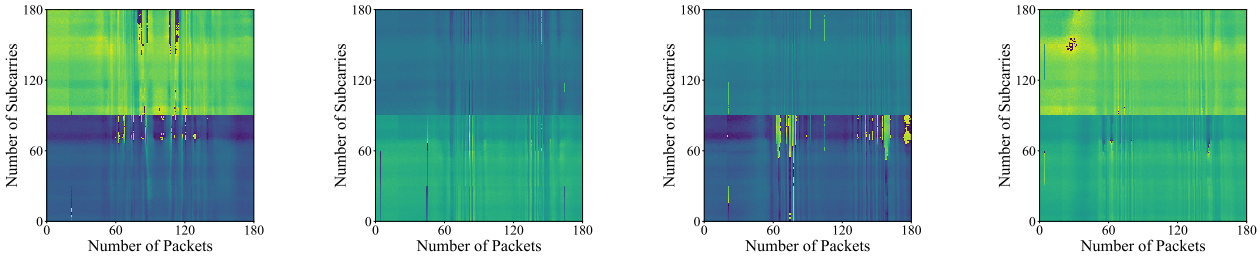
VI. PERFORMANCE EVALUATION

A. WIRIM PERFORMANCE

We first evaluate the performance of the system with the resolution improving mechanism.

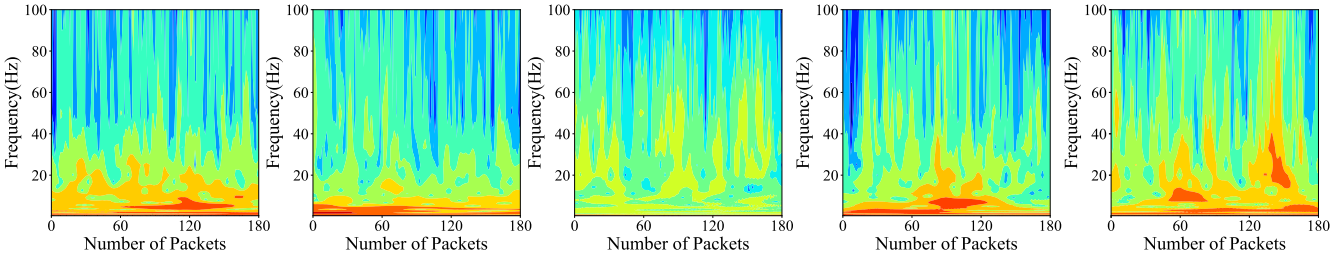
With aggregated CSI, WiRIM uses CSI phase and amplitude to construct AoA spectrum images and STFT spectrum images, respectively. FIGURE 10 shows the AoA images at four different locations. It can be seen that the distribution of AoA spectrums from four adjacent locations are significantly different, so the AoA spectrum images can be used as the fingerprint of indoor localization. FIGURE 11 shows the extracted STFT spectrum images of typical activities. The frequency band concentrated by high energy ("hot" colored) represents the frequency band at the corresponding time of the activity. It is clear that the STFT images characterize the influence of human activities on Wi-Fi signals. As can be seen from the figures, the feature images constructed by WiRIM for different locations and different activities are intuitively easy to classify, which is due to the high resolution provided by WiRIM.

To explore a fine-grained performance, we show the confusion matrix of accuracy at four locations in FIGURE 12. The recognition accuracy of each activity at each location ranges from 87% to 100%. To be specific, the accuracy at



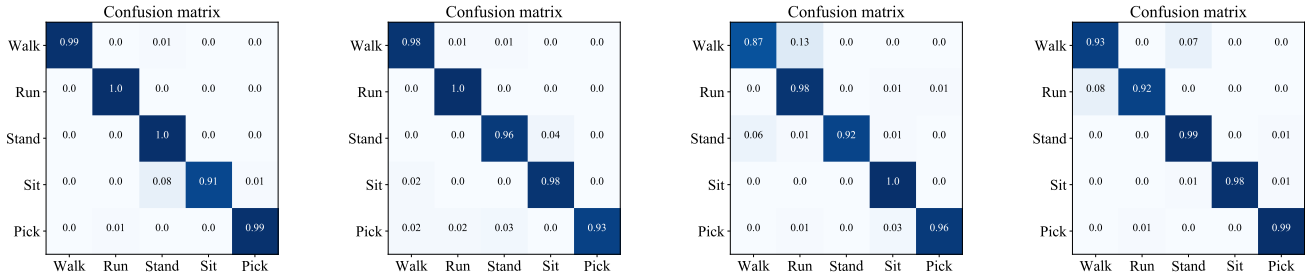
(a) Location 1 (b) Location 2 (c) Location 3 (d) Location 4

FIGURE 10. AoA spectrum images of walking at different locations, constructed in the system with the resolution improving mechanism. The images constructed by WiRIM for different locations are intuitively easy to classify, which is due to the high resolution provided by WiRIM.



(a) Run (b) Walk (c) Stand (d) Sit (e) Pick

FIGURE 11. STFT spectrum images of different activities at location 2, constructed in the system with the resolution improving mechanism. The images constructed by WiRIM for different activities are intuitively easy to classify, which is due to the high resolution provided by WiRIM.



(a) Location 1. (b) Location 2. (c) Location 3. (d) Location 4.

FIGURE 12. Confusion matrix of accuracy at different locations when using WIRIM with the aggregated channel.

four locations is 97.8%, 97%, 94.6% and 96.2%, respectively. It is clear that even if the four locations we selected are closely adjacent to each other, WiRIM can still achieve a good recognition performance at any location, rather than being confused by different activities at different locations, which proves that WiRIM is able to achieve high resolution.

To achieve a comprehensive analysis, we employ the following metrics widely used in statistics [67], [68]: 1) precision (PR), defined as $\frac{TP}{TP+FP}$, where TP (true positive) means the correctly predicted positive value and FP (false positive) means the wrongly predicted positive value, 2) recall (RE), defined as $\frac{TP}{TP+FN}$, where FN (false negative) means the wrongly predicted negative value, 3) F1-score (F1), defined as $\frac{2*PR*RE}{PR+RE}$, is a metric that combines the PR and RE, 4) false positive rate (FPR), defined as $\frac{FP}{FP+TN}$, where TN (true negative) means the correctly predicted negative value. FIGURE 13 shows the statistical results of the four metrics, describes the performance of human activity and location recognition in our experiments. The 20 labels on the x-axis correspond to five activities at four locations. In FIGURE 13, we can observe that all the PR, RE and F1 are above 87% with an average precision of 95.4%, while all FPR are below 0.6%, indicating that WiRIM with the aggregated channel can not only accurately but also comprehensively classify the overwhelming majority of corresponding activity and location.

Besides, we calculate the average localization error (ALE) in meters by converting the location labels into distance:

$$ALE = \frac{1}{N}\sum_{i=1}^{N} \|g(l_i) - g\left(\hat{l}_i\right)\|_2^2, \tag{19}$$

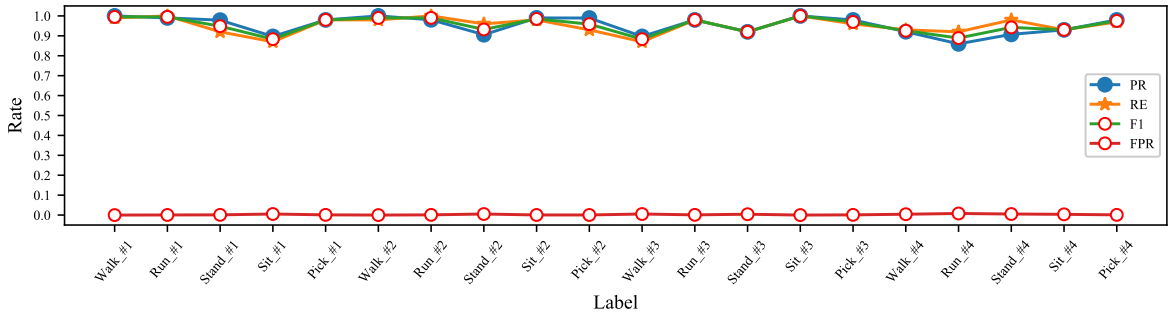
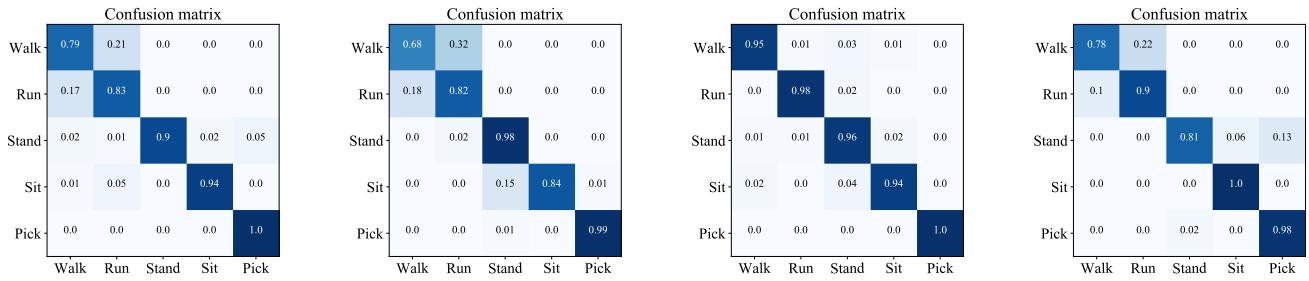


FIGURE 13. PR, RE, F1 and FPR of corresponding activity and location when using WIRIM with the aggregated channel. The performance is excellent and stable.



(a) Location 1. (b) Location 2. (c) Location 3. (d) Location 4.

FIGURE 14. Confusion matrix of accuracy at different locations when using WIRIM with only a single channel.

where N is the number of test sets ($N = 2000$), $\|\cdot\|_2^2$ is to compute the distance, l_i and $\hat{l}_i$ are the predicted location label and the true location label, respectively, $g(\cdot)$ is to convert the location label to the spatial coordinates according to FIGURE 9. According to Equation (19), ALE of WIRIM is 0.043 m, which is accurate enough.

B. COMPARING AGGREGATED CHANNEL WITH SINGLE CHANNEL

In this section, to demonstrate the effectiveness of the aggregated CSI from multiple channels, we will compare the aggregated channel with a single channel. As aforementioned in V-B, without changing the environment, we carry out controlled experiments in WIRIM with a single channel. It should be noted that in off-the-shelf studies only a single channel was considered, while aggregated CSI from multiple channels is used in our paper.

FIGURE 14 shows the confusion matrix of WIRIM with a single channel at four locations. It can be seen that, except at the third position, the performance is poor. This is because the selected locations are relatively close to each other, so the limited resolution of the system can easily cause confusion. At some locations, "Walk" and "Run" are not easy to distinguish, which may be due to the fact that they are both continuous activities with high frequency. Therefore, when the distance between adjacent locations is small or when several activities are similar, the system with limited resolution will have difficulty distinguishing these differences. However, our proposed system is not confused by these differences, which proves the effectiveness of the aggregated channel.

Additionally, we use PR, RE, F1 and FPR to measure the classification performance of WIRIM with a single channel. FIGURE 15 shows that at some locations, the PR, RE and F1 can be as high as 100%, while at some other locations, it can only achieve an accuracy of nearly 60%. Compared to FIGURE 13, the performance in FIGURE 15 fluctuates significantly. That is to say, when obtaining human activity and location information at the same time, the performance of the system that uses only a single channel is unstable. The results demonstrate that the accuracy decreases obviously when using only a single channel. In detail, the overall accuracy of the system using the aggregated channel is as high as 95.4% while the limited resolution system drops to 86.5%. Via Equation (19), the average localization error of the system is 0.069 m, which is coarser than the system using the aggregated channel.

This result suggests that the system with higher resolution can achieve better performance in Wi-Fi based human sensing, which confirms our analysis. Although the system using only a single channel can achieve reasonable accuracy, the performance of the system using aggregated CSI from multiple channels will be remarkably better.

C. COMPARING CNN WITH LSTM

And then we design experiments by using different deep learning methods to evaluate their effect on the performance.

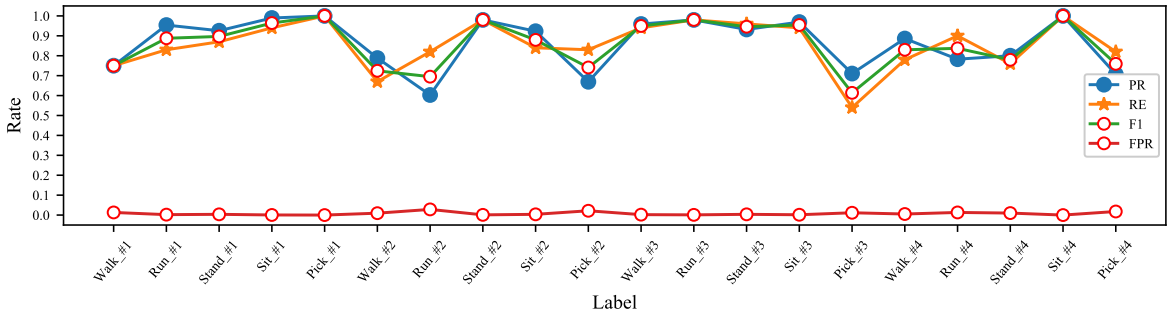


FIGURE 15. PR, RE, F1 and FPR of corresponding activity and location when using WIRIM with only a single channel. The performance is poor and unstable.

TABLE 6. The training time and accuracy of WIRIM using the two approaches.

	Training time	Accuracy
CNN	8834(s)	95.4%
LSTM	40216(s)	95.1%

The results are shown in TABLE 6. The overall accuracies of the system using the CNN and the LSTM are 95.4% and 95.1%, respectively. It should be mentioned that both CNN and LSTM have been successfully applied to learn deep features from the images. WIRIM leverages the amplitude and phase of the aggregated CSI to construct feature images, and exploits a deep learning based image processing module for extracting deep image features from the constructed images, so that both CNN and LSTM can learn informative features from the images and achieve excellent performance in our system. This further indicates the effectiveness of the proposed WIRIM for Wi-Fi based human sensing.

Note that the computational cost of the deep learning method is a common concern. To evaluate the time complexity of these two approaches, we measure the training time required to achieve a human status estimation. As TABLE 6 shows, the training time of LSTM is much larger than that of CNN. This is because LSTM can hold and encode the temporal state information of the activity, which takes more time in training. On one hand, the major drawback of LSTM is that it has very high computational cost, on the other hand, WIRIM converts the raw time varying signals into feature images, which maximizes CNN's ability to learn features from feature images. Therefore, considering the approximate accuracy of the two approaches, we choose CNN with faster training speed to train the constructed images.

D. COMPARING WIRIM WITH A BENCHMARK APPROACH

In this section, to demonstrate the effectiveness of our proposed mechanism, we will compare the activity recognition accuracy of WIRIM with a state-of-the-art approach of HAR.

FIGURE 16 demonstrates the accuracy of all the activities under WIRIM and the benchmark approach ABLSTM [13]. The overall accuracy of the system using ABLSTM is 88.6%

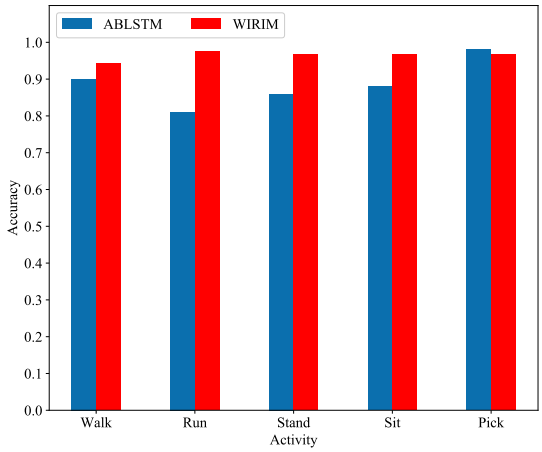


FIGURE 16. The accuracy of all the activities under the two mechanisms.

while WIRIM can achieve 96.4% in HAR. We can observe that WIRIM is superior to ABLSTM in most categories. Since the proposed cross-location human activity recognition scenario requires high resolution, ABLSTM no longer performs well in this scenario. In particular, some "Run" activities are recognized as "Walk", while "Stand" and "Sit" are not easily distinguished. This is why ABLSTM keeps a relatively low accuracy in most activities. The comparative experiments show that WIRIM is more capable of distinguishing the different activities at different locations and can remove such location impacts more effectively. This is due to the fact that WIRIM uses the aggregated channel, which improves the resolution of Wi-Fi signals and facilitates the classifier to learn more abundant and complete features.

VII. CONCLUSION

In this paper, we present WiRIM, a resolution improving mechanism generally applicable to Wi-Fi based human sensing. As most Wi-Fi based human sensing technologies aim to fundamentally understand the governing law on how a human's activity and location impact the received signals, we use a cross-location human activity recognition system as a case study.

We conduct two controlled experiments. Firstly, we estimate the human activity and location information.

The experiment results suggest that the channel switching and aggregation module, the feature extraction module and the deep learning module all improve the performance, and their combination enables the largest performance gain.

Secondly, we do human activity recognition when there is a slight difference in human location. We conduct comparison experiments with the state-of-the-art approach of HAR. The system using WiRIM reaches an activity recognition accuracy of 96.4%, compared with 88.6% for the system using the state-of-the-art approach without the resolution improving mechanism.

Experiment results demonstrate that the resolution improving mechanism can immediately benefit activity recognition. As our future work, we plan to make WIRIM feasible to various applied environments.

REFERENCES

[1] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Commun. Surveys Tuts.*, to be published.

[2] M. A. Al-Qaness, M. Abd Elaziz, S. Kim, A. A. Ewees, A. A. Abbasi, Y. A. Alhaj, and A. Hawbani, "Channel state information from pure communication to sense and track human motion: A survey," *Sensors*, vol. 19, no. 15, p. 3329, 2019.

[3] J. Wang, L. Zhang, Q. Gao, M. Pan, and H. Wang, "Device-free wireless sensing in complex scenarios using spatial structural information," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2432–2442, Apr. 2018.

[4] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2015, pp. 65–76.

[5] K. Qian, C. Wu, Z. Zhou, Y. Zheng, Z. Yang, and Y. Liu, "Inferring motion direction using commodity Wi-Fi for interactive exergames," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2017, pp. 1961–1972.

[6] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with Wi-Fi!" *IEEE Trans. Mobile Comput.*, vol. 15, no. 11, pp. 2907–2920, Nov. 2016.

[7] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using WiFi signals," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 90–102.

[8] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2016.

[9] Y. Zou, W. Liu, K. Wu, and L. M. Ni, "Wi-Fi radar: Recognizing human behavior with commodity Wi-Fi," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 105–111, Oct. 2017.

[10] S. Arshad, C. Feng, Y. Liu, Y. Hu, R. Yu, S. Zhou, and H. Li, "Wichase: A WiFi based human activity recognition system for sensorless environments," in *Proc. IEEE 18th Int. Symp. A World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Jun. 2017, pp. 1–6.

[11] X. Dang, Y. Huang, Z. Hao, and X. Si, "PCA-Kalman: Device-free indoor human behavior detection with commodity Wi-Fi," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, p. 214, 2018.

[12] S. Li, X. Li, Q. Lv, and D. Zhang, "WiFit: A bodyweight exercise monitoring system with commodity Wi-Fi," in *Proc. ACM Int. Joint Conf. Int. Symp. Pervas. Ubiquitous Comput. Wearable Comput.*, 2018, pp. 396–399.

[13] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "WiFi CSI based passive human activity recognition using attention based BLSTM," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2714–2724, Nov. 2019.

[14] H. Yan, Y. Zhang, Y. Wang, and K. Xu, "WiAct: A passive WiFi-based human activity recognition system," *IEEE Sensors J.*, to be published.

[15] Z. Fu, J. Xu, Z. Zhu, A. X. Liu, and X. Sun, "Writing in the air with WiFi signals for virtual reality devices," *IEEE Trans. Mobile Comput.*, vol. 18, no. 2, pp. 473–484, Feb. 2019.

[16] K. Qian, C. Wu, Z. Yang, C. Yang, and Y. Liu, "Decimeter level passive tracking with WiFi," in *Proc. 3rd Workshop Hot Topics Wireless*, 2016, pp. 44–48.

[17] 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. 18th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2017, Art. no. 6.

[18] 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. 16th Annu. Int. Conf. Mobile Syst., Appl., Services*, 2018, pp. 350–361.

[19] X. Li, D. Zhang, Q. Lv, J. Xiong, S. Li, Y. Zhang, and H. Mei, "IndoTrack: Device-free indoor human tracking with commodity Wi-Fi," *ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, p. 72, Sep. 2017.

[20] A. S. Paul, E. A. Wan, F. Adenwala, E. Schafermeyer, N. Preiser, J. Kaye, and P. G. Jacobs, "MobileRF: A robust device-free tracking system based on a hybrid neural network HMM classifier," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput.*, 2014, pp. 159–170.

[21] K. Qian, C. Wu, Z. Yang, Y. Liu, F. He, and T. Xing, "Enabling contactless detection of moving humans with dynamic speeds using CSI," *ACM Trans. Embedded Comput. Syst. (TECS)*, vol. 17, no. 2, 2018, Art. no. 52.

[22] F. Zhang, D. Zhang, J. Xiong, H. Wang, K. Niu, B. Jin, and Y. Wang, "From fresnel diffraction model to fine-grained human respiration sensing with commodity Wi-Fi devices," *ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 1, 2018, Art. no. 53.

[23] K. Niu, F. Zhang, Z. Chang, and D. Zhang, "A fresnel diffraction model based human respiration detection system using COTS Wi-Fi devices," in *Proc. ACM Int. Joint Conf. Int. Symp. Pervas. Ubiquitous Comput. Wearable Comput.*, 2018, pp. 416–419.

[24] Y. Zeng, E. Yi, D. Wu, R. Gao, and D. Zhang, "A full human respiration detection system using commodity Wi-Fi devices," in *Proc. ACM Int. Joint Conf. Int. Symp. Pervas. Ubiquitous Comput. Wearable Comput.*, 2018, pp. 480–483.

[25] D. Zhang, H. Wang, and D. Wu, "Toward centimeter-scale human activity sensing with Wi-Fi signals," *Computer*, vol. 50, no. 1, pp. 48–57, Jan. 2017.

[26] Y. Zhang, X. Li, S. Li, J. Xiong, and D. Zhang, "A training-free contactless human vitality monitoring platform using commodity Wi-Fi devices," in *Proc. ACM Int. Joint Conf., Int. Symp. Pervas. Ubiquitous Comput. Wearable Comput.*, 2018, pp. 488–491.

[27] Q. Xu, Y. Han, B. Wang, M. Wu, and K. J. R. Liu, "Indoor events monitoring using channel state information time series," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4977–4990, Jun. 2019.

[28] T. Xin, B. Guo, Z. Wang, P. Wang, J. C. K. Lam, V. Li, and Z. Yu, "FreeSense: A robust approach for indoor human detection using Wi-Fi signals," *ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 3, 2018, Art. no. 143.

[29] J. Wang, X. Zhang, Q. Gao, H. Yue, and H. Wang, "Device-free wireless localization and activity recognition: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6258–6267, Jul. 2017.

[30] X. Zhang, J. Wang, Q. Gao, X. Ma, and H. Wang, "Device-free wireless localization and activity recognition with deep learning," in *Proc. IEEE Int. Conf. Pervas. Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2016, pp. 1–5.

[31] Q. Gao, J. Wang, X. Ma, X. Feng, and H. Wang, "CSI-based device-free wireless localization and activity recognition using radio image features," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10346–10356, Nov. 2017.

[32] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, Dec. 2013, Art. no. 25.

[33] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, Jan. 2011.

[34] S. Sen, J. Lee, K.-H. Kim, and P. Congdon, "Avoiding multipath to revive inbuilding WiFi localization," in *Proc. 11th Annu. Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, Taipei, Taiwan, Jun. 2013, pp. 249–262, doi: 10.1145/2462456.2464463.

[35] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, and Y. Liu, "Omnidirectional coverage for device-free passive human detection," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 7, pp. 1819–1829, Jul. 2014.

[36] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni, "FILA: Fine-grained indoor localization," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2210–2218.

[37] 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. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 617–628.

[38] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, H. Cai, Y. Liu, and L. M. Ni, "WiFi-based indoor line-of-sight identification," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6125–6136, Nov. 2015.

[39] Z. Zhou, Z. Yang, C. Wu, W. Sun, and Y. Liu, "LiFi: Line-of-sight identification with WiFi," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr./May 2014, pp. 2688–2696.

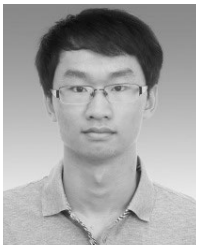
[40] E. Soltanaghaei, A. Kalyanaraman, and K. Whitehouse, "Poster: Improving multipath resolution with MIMO smoothing," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw.*, 2017, pp. 585–587.
[41] Y. Xie, Y. Zhang, J. C. Liando, and M. Li, "SWAN: Stitched Wi-Fi ANtennas," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, 2018, pp. 51–66.
[42] J. Xiong, K. Sundaresan, and K. Jamieson, "ToneTrack: Leveraging frequency-agile radios for time-based indoor wireless localization," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 537–549.
[43] M. B. Khalilsarai, S. Stefanatos, G. Wunder, and G. Caire, "WiFi-based indoor localization via multi-band splicing and phase retrieval," 2019, *arXiv:1903.02367*. [Online]. Available: https://arxiv.org/abs/1903.02367
[44] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity Wi-Fi," *IEEE Trans. Mobile Comput.*, vol. 18, no. 6, pp. 1342–1355, Jun. 2019.
[45] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single WiFi access point," in *Proc. 13th Symp. Networked Syst. Design Implement. (NSDI)*, 2016, pp. 165–178.
[46] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[47] J. K. Tan, "An adaptive orthogonaly division multiplexing baseband modem for wideband wireless communication," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2006.
[48] X. Wu, Z. Chu, P. Yang, C. Xiang, X. Zheng, and W. Huang, "TW-See: Human activity recognition through the wall with commodity Wi-Fi devices," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 306–319, Jan. 2019.
[49] X. Guo, B. Liu, C. Shi, H. Liu, Y. Chen, and M. C. Chuah, "WiFi-enabled smart human dynamics monitoring," in *Proc. 15th ACM Conf. Embedded Netw. Sensor Syst.*, 2017, Art. no. 16.
[50] 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. 16th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2015, pp. 267–276.
[51] T. Xin, B. Guo, Z. Wang, M. Li, Z. Yu, and X. Zhou, "FreeSense: Indoor human identification with Wi-Fi signals," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–7.
[52] M. Speth, S. A. Fechtel, G. Fock, and H. Meyr, "Optimum receiver design for wireless broad-band systems using OFDM. I," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1668–1677, Nov. 1999.
[53] J. Gjengset, J. Xiong, G. McPhillips, and K. Jamieson, "Phaser: Enabling phased array signal processing on commodity WiFi access points," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 153–164.
[54] X. Wang, C. Yang, and S. Mao, "PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2017, pp. 1230–1239.
[55] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 159–170, 2010.
[56] Foudnstone. (2012). *Getting Started With LORCON*. [Online]. Available: http://blog.opensecurityresearch.com/2012/09/getting-started-with-lorcon.html/
[57] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using WiFi signals," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput.*, 2016, pp. 363–373.
[58] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
[59] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.
[60] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[61] H. Huang, Y. Peng, J. Yang, W. Xia, and G. Gui, "Fast beamforming design via deep learning," *IEEE Trans. Veh. Technol.*, to be published.
[62] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, *arXiv:1207.0580*. [Online]. Available: https://arxiv.org/abs/1207.0580
[63] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019.
[64] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. Design Implement. (OSDI)*, 2016, pp. 265–283.
[65] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.
[66] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using WiFi channel state information," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 98–104, Oct. 2017.
[67] K. Ohara, T. Maekawa, and Y. Matsushita, "Detecting state changes of indoor everyday objects using Wi-Fi channel state information," *ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, 2017, Art. no. 88.
[68] Z. Wu, S. Zhao, Y. Peng, X. He, X. Zhao, K. Huang, X. Wu, W. Fan, F. Li, and M. Chen, "Studies on different CNN algorithms for face skin disease classification based on clinical images," *IEEE Access*, vol. 7, pp. 66505–66511, 2019.

XINBIN SHEN received the B.S. degree in communication engineering from the Beijing University of Posts and Telecommunications, China, where he is currently pursuing the M.S. degree in electronic and communication engineering. His current research interests include Wi-Fi sensing, machine learning, and deep learning.

LINGCHAO GUO received the B.S. degree in communication engineering from the China University of Petroleum, Qingdao, China, in 2017. She is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications. Her research interests include Wi-Fi sensing and Wi-Fi imaging.

ZHAOMING LU received the Ph.D. degree from the Beijing University of Posts and Telecommunications (BUPT), in 2012. He joined the School of Information and Communication Engineering, BUPT, in 2012. His research interests include open wireless networks, QoE management in wireless networks, software-defined wireless networks, and cross-layer design for mobile video applications.

XIANGMING WEN received the M.Sc. and Ph.D. degrees in information and communication engineering from the Beijing University of Posts and Telecommunications (BUPT). He is currently the Director of the Beijing Key Laboratory of Network System Architecture and Convergence, where he is managing several projects related to open wireless networking. He is also the Vice President of BUPT. His current research interests include radio resource and mobility management, software-defined wireless networks, and broadband multimedia transmission technology.



ZHIHONG HE received the B.S. degree in communication engineering from the Beijing University of Posts and Telecommunications, China, where he is currently pursuing the M.S. degree in communications engineering. His current research interests include Wi-Fi sensing, vital sign monitoring, machine learning, and deep learning.