On Spatial Diversity in WiFi-Based Human Activity Recognition: A Deep Learning-Based Approach

Fangxin Wang, *Student Member, IEEE*, Wei Gong, *Member, IEEE*, and Jiangchuan Liu, *Fellow, IEEE*

***Abstract*—The deeply penetrated WiFi signals not only provide fundamental communications for the massive Internet of Things devices but also enable cognitive sensing ability in many other applications, such as human activity recognition. State-of-the-art WiFi-based device-free systems leverage the correlations between signal changes and body movements for human activity recognition. They have demonstrated reasonably good recognition results with a properly placed transceiver pair, or, in other words, when the human body is within a certain *sweet zone*. Unfortunately, the sweet zone is not ubiquitous. When the person moves out of the area and enters a *dead zone*, or even just the orientation changes, the recognition accuracy can quickly decay. In this paper, we closely examine such spatial diversity in WiFi-based human activity recognition. We identify the dead zones and their key influential factors, and accordingly present WiSDAR, a WiFi-based spatial diversity-aware device-free activity recognition system. WiSDAR overshadows the dead zones yet with only one physical WiFi sender and receiver. The key innovation is extending the multiple antennas of modern WiFi devices to construct multiple separated antenna pairs for activity observing. Profiling activity features from multiple spatial dimensions can be more complicated and offer much richer information for further recognition. To this end, we propose a deep learning-based framework that integrates the hidden features from both temporal and spatial dimensions, achieving highly accurate and reliable recognition results. WiSDAR is fully compatible with commercial off-the-shelf WiFi devices, and we have implemented it on the commonly available Intel WiFi 5300 cards. Our real-world experiments demonstrate that it recognizes human activities with a stable accuracy of around 96%.**

***Index Terms*—Deep learning, human activity recognition, spatial diversity.**

I. Introduction

As a cornerstone service in such important Internet of Things applications as smart home, health diagnosis, and intrusion detection, human activity recognition has attracted great attention in both academia and industry. Among many forms of sensing technologies, e.g., camera [1], wearable sensor [2], [3], and RFID [4], WiFi-based activity recognition is of particular interest given its ubiquity, low cost, device-free experience, and low dependence [5].

Manuscript received May 1, 2018; revised August 15, 2018; accepted September 12, 2018. Date of publication September 20, 2018; date of current version May 8, 2019. This work was supported in part by the Canada Technology Demonstration Program and in part by the Canada NSERC Discovery Grant. *(Corresponding author: Jiangchuan Liu.)*

F. Wang and J. Liu are with the School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada (e-mail: fangxinw@sfu.ca; jcliu@cs.sfu.ca).

W. Gong is with the School of Computer Science and Technology, University of Science and Technology of China, Hefei 230000, China (e-mail: weigong@ustc.edu.cn).

Digital Object Identifier 10.1109/JIOT.2018.2871445

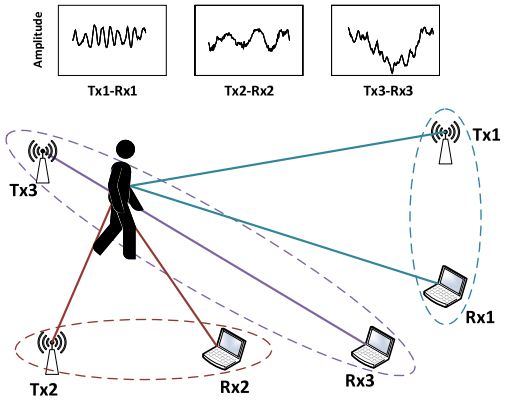

Fig. 1. Three Tx–Rx pairs observe different patterns in CSI amplitude for the same walking activity due to the spatial diversity. The dashed ellipses are their IAs, respectively.

Intuitively, when a person resides in the surrounding area of a WiFi transceiver pair, his/her body movement will affect the travel-through WiFi signals. Through analyzing such signal characteristics as coarse-grained received signal strength indicator (RSSI) and fine-grained channel state information (CSI), different activities can be recognized [5]–[10]. Existing solutions have demonstrated reasonably good recognition results with a properly placed transceiver pair [8], [10], or, in other words, when the human body is within a certain *sweet zone*. Unfortunately, our observations show that the sweet zone [which we refer to as the effective area (EA) of a recognition algorithm] is not ubiquitous. When the person moves out of the area, or even just the orientation changes, the accuracy of recognition can quickly decay.

To understand the impact of such *spatial diversity*, we consider a simple case of the walking activity in Fig. 1, which is observed by three different transceiver pairs. The Tx1–Rx1 pair is from a vertical angle, observing a fast waving shape in the CSI amplitude based on the state-of-the-art solutions [8], [11], whereas the Tx2–Rx2 pair is from a horizontal angle, observing a much slower waving shape. Clearly, the different patterns of changes in wireless channel metrics may lead to different recognition results, even though they are observing the same activity of the same person. The third pair, Tx3–Rx3, which is largely blocked by the target, sees an even worse result: a significant drop in the CSI amplitude. Such faded power is hardly useful for recognition; in other words, the person is in the *dead zone* of the Tx3–Rx3 pair. In fact, our observations suggest that any transceiver pair has a non-negligible dead zone [refer to as an ineffective area (IA), as outlined by a dashed ellipse in Fig. 1].

In this paper, through extensive field experiments and analysis, we closely examine the spatial diversity in WiFi-based

human activity recognition. We identify the IAs and their key influential factors. Motivated by the shadowless lamp design in surgery, we develop a WiFi-based spatial diversity-aware device-free activity recognition (WiSDAR) system, which overshadows the IAs yet with only one physical WiFi sender and receiver. The key innovation is extending the multiple antennas of modern WiFi devices to construct multiple separated antenna (SA) pairs and obtain features from multiple spatial dimensions. A target area determination scheme is also applied to select those ineffective pairs (referring to those affected pairs due to the existence of the target in their IAs) and filter out the corresponding *dirty* features. Different from existing solutions [9], [11] that use multiple WiFi links to obtain extra features, WiSDAR is cost-effective since it only requires two transceivers and one-time initial deployment. For example, if we want to construct three senders and three receivers for observing, traditional approaches like [9] and [11] require six WiFi devices while WiSDAR only requires two devices.

Profiling activity features from multiple spatial dimensions is more complicated and also offers much richer information for further activity recognition. Conventional classification tools used in existing systems, e.g., hidden Markov model (HMM) [8] and k-nearest neighbors (kNNs) [11], however, are not powerful enough to mine the hidden temporal and spatial relationships from such data. WiSDAR employs an advanced deep learning model to analyze their patterns through supervised learning. In particular, both convolutional neural network (CNN) [12] and long short term memory (LSTM) [13] network are applied to integrate the features from both temporal and spatial dimensions, and achieves highly accurate and reliable activity recognition results.

WiSDAR is fully compatible with commercial off-the-shelf (COTS) WiFi devices. We have implemented WiSDAR on the commonly available Intel WiFi Link 5300 cards. Our real-world experiment results demonstrate that it recognizes human activities with a stable accuracy of around 96%. This greatly surpasses the state-of-the-art solutions [8] (around 75% when not carefully considered the spatial diversity, especially the IA).

The rest of this paper is organized as follows. Section II introduces the reflection model and the impact of the spatial diversity. Section III outlines the system overview. We describe the preprocessing scheme and the deep learning-based recognition approach in Sections IV and V, respectively. The implementation and evaluation are presented in Section VI. We review the related works in Section VII, with the discussion and conclusion in Sections VIII and IX.

II. Understanding Spatial Diversity in Reflection Model

We start from some necessary background information, followed by examining the existence and impact of the spatial diversity in WiFi reflection models.

A. Reflection Model for Activity Recognition

As shown in Fig. 2, in an indoor environment, radial signals can be reflected by many objects, e.g., walls and human bodies, and thus arrive at a receiver through multiple paths. CSI is commonly used to characterize the channel frequency response (CFR) of a communication link in WiFi systems. Let $H(f, t)$ represent the CFR measured for frequency f at time t, we have

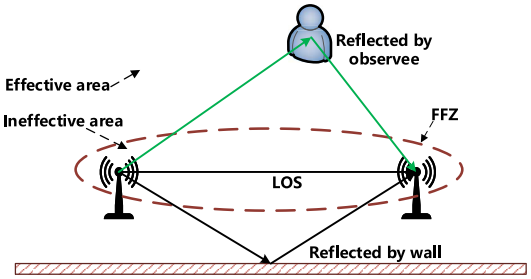

Fig. 2. Illustration of the reflection model as well as the IA and EA of a transceiver pair.

$$H(f,t) = \sum_{k=1}^{K} \alpha_k(t) e^{-j2\pi f \tau_k(t)} \tag{1}$$

where K is the total number of multipaths, $\alpha_k(t)$ and $\tau_k(t)$ are the complex channel attenuation and the time of flight for path k, respectively.

The relative movement between transceivers and a reflector (human body in our context) will change the frequency observed at the receiver, i.e., the *Doppler effect*. Given λ, the wavelength, and $d(t)$, the change of reflected path length, the frequency shift of signals bounced off is $f_D = -(1/\lambda)(d/dt)d(t)$ [14], and the total CFR then can be represented as

$$H(f,t) = e^{-j2\pi \Delta f t}\left(H_s(f) + \sum_{k \in P_d} \alpha_k(t) e^{j2\pi \int_{-\infty}^{t} f_{D_k}(u)du}\right) \tag{2}$$

where Δf is the carrier frequency offset (CFO), $H_s(f, t)$ is the sum of static paths, and P_d is the set of dynamic paths.

The magnitude of the combined CFR changes with the dynamic component, which can be explored for human activity recognition. The relationship however is obscured by the unknown CFO. Earlier works [8], [11] make use of the CFR power [i.e., multiplication of $H(f, t)$] to eliminate the impact of unknown CFO, building up the correlation between the wave frequencies of CFR power and the dynamic path length changes. The reflection model approximates the velocity of the body movement as a fixed function of the path length change rate, and utilizes the extracted features in the time-frequency domain for human activity recognition.

B. Spatial Diversity of the Observed Target Area

Observation 1: The CFR power can be largely attenuated when the target is located in a certain area (the IA) of a transceiver pair, thereby affecting the activity recognition accuracy.

Fig. 3(a) plots the CSI amplitude of 30 subcarriers from a pair for an observed target. From 1 to 3.2 s, the CSI exhibits a high amplitude when the target is moving. Yet the amplitude attenuates significantly from 3.2 s, which is hardly reliable for activity recognition. Fig. 3(b) shows the CSI amplitude change when the target passes through the IA. The CSI amplitude drop lasts about 0.26 s, which coincides with the time of passing through the area.

The power fading can be caused by many effects, such as absorption, diffraction, and interference, not just directly blocking the line-of-sight (LOS) path [15]. The *Fresnel zones*

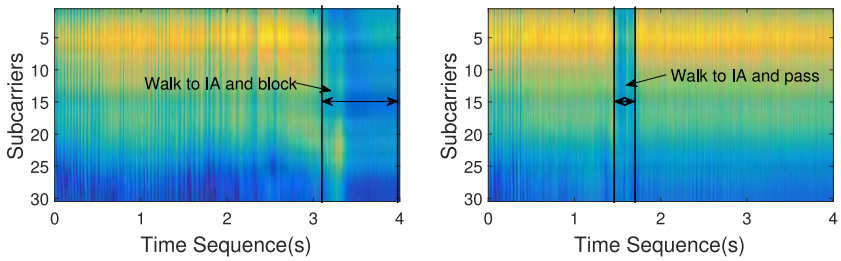

Fig. 3. Amplitude of 30 CSI subcarriers of a transceiver pair. When the target is in the IA, the amplitude attenuates obviously. (a) Person walks to IA and stays inside. (b) Person passes through IA.

are a series of concentric ellipsoidal regions whose foci are the pair of transceivers [15]. The first Fresnel zone (FFZ) is the innermost ellipsoid where the difference between major axis length and the foci length is half of the wavelength. It is known that most of the RF energy is transmitted through the FFZ, instead of through the LOS path only [16]. Obstruction in the FFZ will negatively affect the power of signal transmission. Our experiment shows that the recognition accuracy for activities in the FFZ is only 50% on average (see more details in Section VI). That said, the FFZ outlines an IA for activity recognition, and otherwise the EA, as illustrated in Fig. 2. In this paper, we refer to the area inside the FFZ as an IA. The size of the FFZ depends on the distance between the transceivers and the frequency of the radios. The maximum diameter of the FFZ can be calculated as $F_1 = \sqrt{(cD/f)}$ [15], where c is the light speed, D is the distance between the transceivers, and f is the radio frequency. For example, when we use 5 GHz frequency and set the distance to 3 m, the maximum diameter of the FFZ is 0.42 m. The impact of the FFZ has been examined in such applications as localization, tracking, and respiration detection [17]–[20]. We, however, mainly focus on investigating the impact of the FFZ on activity recognition and how to eliminate such impact.

Note that the energy is not equally distributed among all the subcarriers due to the frequency selective fading [21], which is a normal phenomenon. Also, some subcarriers may be affected by multipath effect and experience an amplitude rise when the target is in the IA; see, for example, from the 20th to the 27th subcarrier in Fig. 4. Nevertheless, we observe that most subcarriers are normally affected by power fading, so that the overall amplitude still attenuates dramatically (e.g., from the 1st to the 18th subcarrier).

C. Spatial Diversity of the Observing Transceivers

Besides the spatial diversity of the observed target in different areas, the spatial diversity of observing transceivers also affects observations and hence recognition results.

Observation 2: For the same activity, a pair of transceivers can observe quite different CFR power characteristics when they are placed at different locations.

Fig. 5 compares the spectrograms of the CFR power for a body falling activity that is observed from three different transceiver pairs. The three pairs are of different locations and orientations. Given the correlation between the wave frequency and the reflected path length change rate, we can use short-time Fourier transform (STFT) or discrete wavelet transform

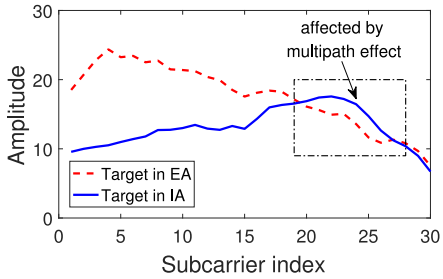

Fig. 4. CSI comparison when the target is in IA and in EA.

(DWT) to separate the frequency components across the time domain.

The Tx1–Rx1 pair observes a high energy rising from 10 Hz to about 60 Hz from time 4.5 to 5 s, and an energy drop to very low frequency (near static) [Fig. 5(a)]. According to the reflection model, this spectrogram implies that the target accelerates from a low speed to high speed in a short time and then suddenly stops. Such characteristics match the falling activity well, i.e., the target falls very fast and then keeps stationary on the floor.

This falling activity however is not well captured by the Tx2–Rx2 pair, which shows very different spectrogram [in Fig. 5(b)]. This spectrogram has a very small crest at about 4.7 s, only achieving at 25 Hz in the frequency domain. It is likely to be classified as some other activities due to the low frequency. The observation from Tx3–Rx3 pair is even worse, since the spectrogram keeps at a low frequency, as if there is nothing happened.

In short, the relative locations and orientations between the observing transceivers and the target matter for activity recognition. There has been efforts toward estimating the location and orientation [10]. Yet given the existence of the IA, using a single pair of transceivers will simply fail if the target unluckily moves there.

III. System Overview

Our observations in the previous section reveal that the spatial diversity in the reflection model seriously undermines activity recognition. Our WiSDAR, a spatial diversity-aware device-free human activity recognition system, seeks to address this issue yet with little extra hardware overhead. The key innovation is utilizing the MIMO feature and multiple extended antennas of existing WiFi devices to construct multiple separated observing pairs. As such, WiSDAR

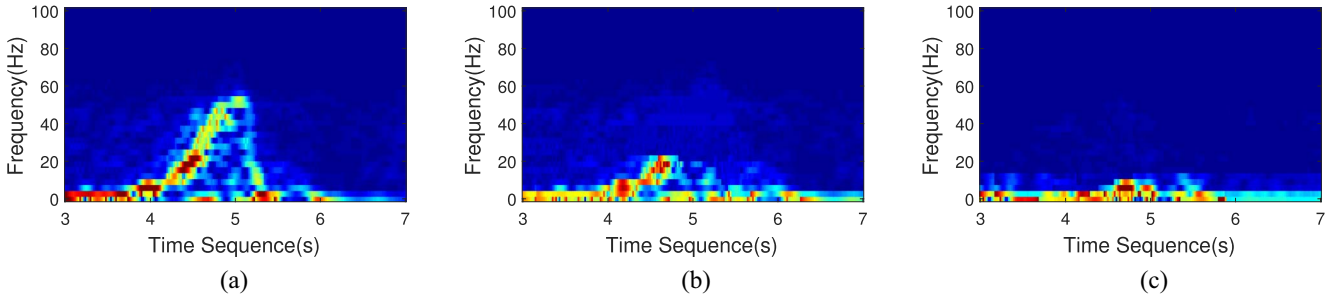

Fig. 5. Comparison of spectrograms of the same falling activity from different observing pairs. (a) Observation from Tx1–Rx1. (b) Observation from Tx2–Rx2. (c) Observation from Tx3–Rx3.

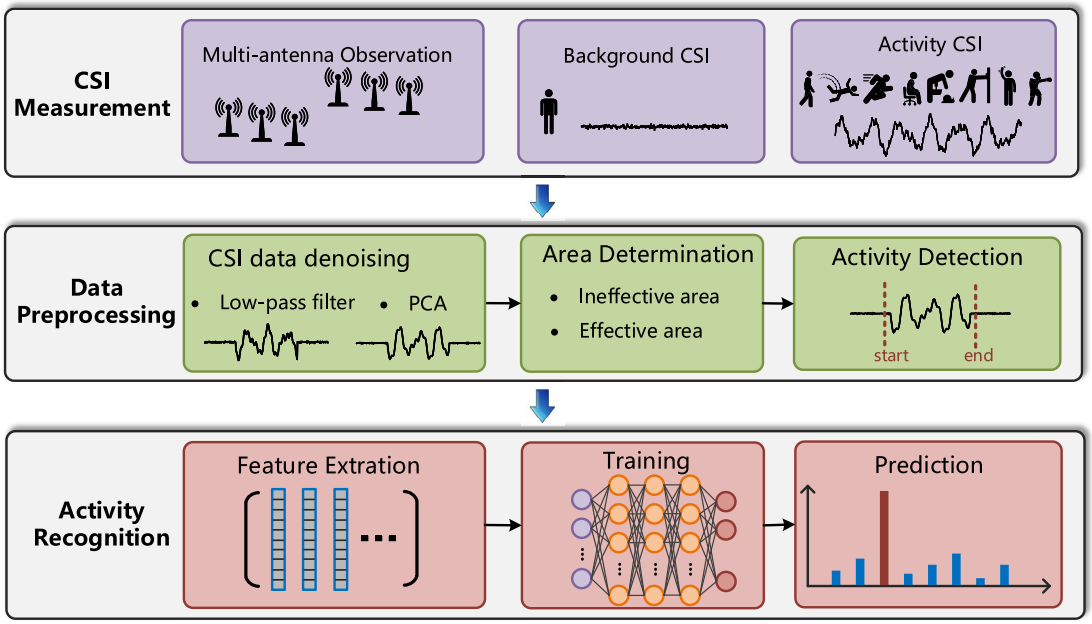

Fig. 6. WiSDAR system framework.

obtains more diverse features from multiple spatial dimensions, which not only minimizes the impact of IAs like a shadowless lamp does in surgery but also offers richer spatial and temporal information that works for the advanced learning tools. As our later experiments show, three antennas are generally good enough to construct separated pairs for activity recognition, which are readily available in today's WiFi devices. We emphasize that WiSDAR still uses only one pair of physical WiFi devices with no extra NICs or APs needed. It is fully compatible with the current WiFi standards and we have implemented it with the commonly available Intel 5300 NICs.

The WiSDAR system framework is illustrated in Fig. 6, which consists of three modules as follows.

A. CSI Measurement

Our system collects CSI as input from COTS wireless devices. The background CSI is first collected as baseline data. When there is a person performing an activity, the CSI of such activity is then collected for further processing. Different from the state-of-the-art solution that only has one effective transceiver pair for monitoring, we extend the multiple antennas and collect CSI from every observing pair.

B. Data Preprocessing

WiSDAR denoises the collected raw signal by low-pass filtering and principal component analysis (PCA). Through a target area determination scheme, WiSDAR detects whether the target is located in the IA of a pair or not. We accordingly discard the features of such ineffective pairs and remain the effective features for further processing. WiSDAR detects the existence of an activity based on the CSI changes compared to the baseline data.

C. Activity Recognition

WiSDAR uses STFT to extract features on both the time domain and the frequency domain to generate the spectrogram. We stack all the generated spectrograms of each observing antenna pairs as the initial input for further training and learning. We then use a deep learning model consisting of CNN and LSTM to integrate the multidimensional features and classify different activities.

IV. Data Collection and Preprocessing

A. CSI Collection and Denoising

We separate the antennas of WiFi devices for CSI collection. With N_{Tx} transmitting antennas, N_{Rx} receiving antennas, and

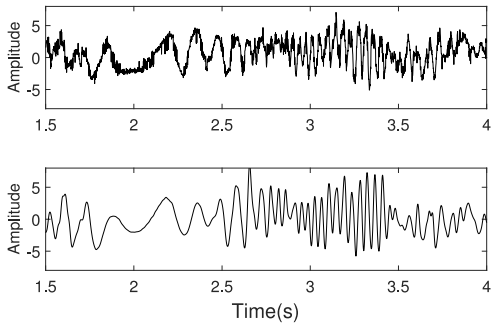

Fig. 7. CSI amplitude before and after our denoising scheme.

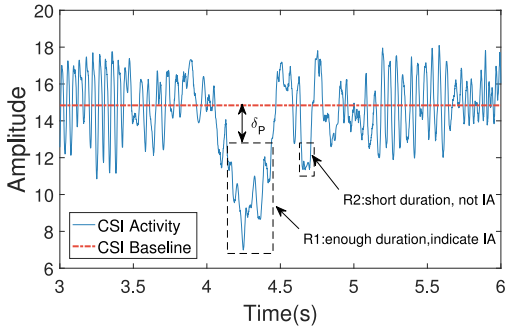

Fig. 8. Case of target area determination. We indicate R1 as an period of IA due to its big attenuation with enough duration, while we ignore R2 because its duration is too short.

N_{sc} CSI subcarriers (e.g., Intel 5300 NIC driver reports 30 subcarriers), we can obtain a total of $N_{sc} * N_{Tx} * N_{Rx}$ CSI streams[1] for each measurement. We call the N_{sc} streams for every transceiver pair as a stream group.

The raw CSI provided by COTS WiFi devices is very noisy and cannot be directly used for recognition. Hence, WiSDAR denoises the collected raw CSI data and then extracts effective features. Since the CFR power changes caused by human movement are mostly low-frequency components, we first let all CSI streams pass a low-pass filter (e.g., the Butterworth filter) to remove any high-frequency noise. Due to the correlations in CSI streams among one stream group [8], we also apply PCA on a stream group to capture such correlations and abstract multiple principal components. Given the first principal component captures too much noise, we use the average of the second and third principal components for further processing, which we refer to as *p-stream* in the remaining part of this paper.

Fig. 7 shows the comparison between a randomly selected raw CSI stream from a stream group of a Tx–Rx pair and the denoised *p*-stream of this pair. We can see that the raw CSI stream contains a lot of high-frequency components, such as impulse and burst noises. After the denoising process, the *p*-stream contains little high-frequency noise and conserves the activity features effectively.

B. Target Area Determination

Since the collected signals of activities in the IA of a Tx–Rx pair are largely affected by power fading, we need to filter out these related pairs in case the abnormal features affect the recognition. There are two criteria to determine that a target is located in the IA. First, the amplitude of most subcarriers will have an obvious drop. Even frequency selective fading [21] and multipath effect may cause an amplitude increase of some subcarriers, the amplitude drop dominates and we can use the average amplitude of all subcarriers as an indicator. Second, the amplitude drop lasts for a relatively long duration. This is intuitive because a person is impossible to pass through the IA in a moment. Note that human activities can also cause the CSI amplitude to fluctuate to a low value, whereas the duration of such amplitude drop is quite short and the amplitude should keep fluctuating up and down.

Based on these two criteria, we develop a target area determination scheme to judge whether a target is located in the IA of a pair. We first need to collect the baseline CSI data when no obstruction is located in the IA and no target is moving. Using the data denoising method introduced before, we obtain the *p*-stream of baseline data as C_b. We intercept this baseline data for a time period and calculate the mean value as $\overline{C_b}$. When a person is performing activities, we then obtain the *p*-stream value of a Tx–Rx pair as C_a. WiSDAR empirically selects a *power threshold* δ_P and a *time threshold* δ_T. We judge that a target is located in the IA if the following conditions are both satisfied:

$$\forall\ t_i \in [t_p, t_q],\ \text{s.t.}$$

$$\overline{C_b} - C_a(t_i) \geq \delta_P \tag{3}$$

$$t_q - t_p \geq \delta_T. \tag{4}$$

According to this method, we can find out those large amplitude drops with long durations and consider them as ineffective parts. We discard the CSI features of the corresponding antenna pair during the time rage $[t_p, t_q]$.

Fig. 8 illustrates a case that a person is walking around a pair of transceivers and passing through the IA. In region R1, the CSI amplitude experiences a large drop with enough time duration. Since it satisfies the two conditions, we can infer that the target is located in the IA. Yet in region R2, even the amplitude value falls below the power threshold δ_P, the short time duration indicates that this drop is more likely a normal fluctuation rather than caused by the power fading. According to our extensive measurement, WiSDAR can achieve a high determination rate when we set the power threshold δ_P as 2.5 dB and the time threshold δ_T as 300 ms. We use this setting throughout the rest of this paper.

C. Activity Detection

Before we begin to recognize human activities, an important step is to detect in which period an activity exists. To detect the start and the end of an activity, we consider the wave patterns of the *p*-stream for each pair. We have two key observations here. First, during an activity, the CSI amplitude has a large variance according to the reflection model, whereas the amplitude keeps steady or has only very small variance in the absence of an activity. Second, an activity usually lasts a relatively long time duration instead of a short burst (e.g., walking and standing up even falling activities last at least 0.2 s). Then a short wave burst can be viewed as noise.

Based on these two observations, we develop an effective approach to determine the start and the end of an activity automatically. We first collect the CSI and obtain the denoised

[1]A CSI stream is the time-series CFR value of an OFDM subcarrier of a particular antenna pair.

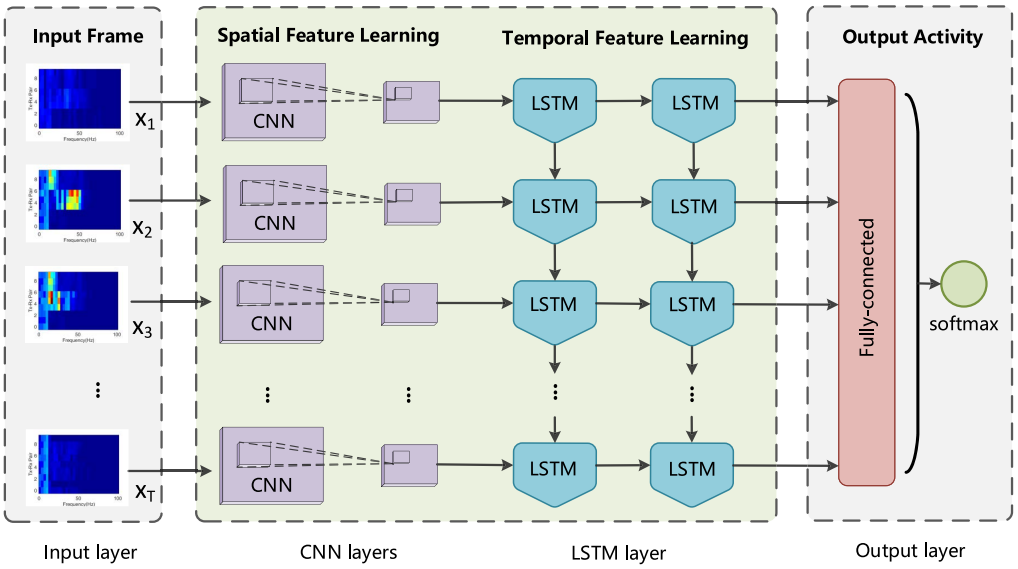

Fig. 9. Learning architecture of WiSDAR.

p-stream C_b and C_a when a person is stationary and is performing an activity, respectively. Then we take the mean value of 3 highest peak values from $|C_b|$ and $|C_a|$, denoted as P_b and P_a, respectively. We can easily find all the peaks since a wave peak has to satisfy two criteria, i.e., its value is larger than the mean value and its neighbors' values. Our WiSDAR system sets the *activity amplitude threshold* as $\theta_P = (P_b + P_a)/2$. Besides the activity amplitude threshold, WiSDAR also sets an *activity duration threshold* θ_L, representing the monitoring time duration. We determine the start and end of an activity according to the following method.

- *Claim 1:* When current state is *no activity*, we consider the peak point t whose value exceed θ_P. If during the following time duration θ_L there is still other peak values larger than θ_P, then t is the start of an activity and we change current state to *in activity*.
- *Claim 2:* When current state is *in activity*, we consider the peak point t whose value exceed θ_P. If during the following time duration θ_L there is no other peak values larger than θ_P, then t is the end of an activity and we change current state to *no activity*.

Since we have separated multiple Tx–Rx pairs monitoring from different location and orientation, we coherently combine these observations together. We determine the start of an activity if *Any* one pair satisfy the claim 1, whereas we determine the end of an activity only if *All* pairs satisfy claim 2 (except those ineffective pairs affected by the target in their IAs).

From an empirical study of our dataset, we set the θ_L as 200 ms. And the detection result in our dataset shows that WiSDAR is able to detect 97% of the start and end of activities with no false positive result. Note that our detection approach is able to automatically adjust according to the environment change. WiSDAR continuously collects the baseline data C_b when there is no activity and the activity data C_a in presence of activities. In our experiment, we collect the new radio signals for *no activity* every 60 s and update the P_b accordingly in case of the environment change. Once we detect an activity, we set a new P_a based on the collected radio signals of the current activity, and update the amplitude threshold θ_P using the new P_a. In this way, the threshold θ_P can be iteratively updated to better fit the environment change.

V. Deep Learning for Activity Recognition

In this section, we describe the main components of the WiSDAR design. Fig. 9 illustrates the learning architecture of WiSDAR, which consists four main layers, including an input layer, CNN layers, LSTM layers, and an output layer. These layers stack together to form a deep neural network for activity recognition. We describe this architecture as follows.

A. Feature Extraction

We first consider to feeding the learning engine with a rich set of distinguishable and representative features. Although the p-stream of each antenna pair extracts the inner wave patterns from the correlated CSI streams, it is not a good feature representation since it only reflects the time and amplitude of a waveform, not revealing the frequency domain characteristics explicitly. Two activities may exhibit similar waveform shapes but have different frequencies (e.g., running and walking). As such, p-streams are not good choices to be directly used in activity recognition.

WiSDAR applies STFT on the p-stream of each pair to extract frequency component. As one of the most popular time-frequency analysis tool, STFT divides a long-time signal into shorter segments of equal length and then compute the Fourier transform separately on each shorter segment. Compared to DWT that uses different resolution on different frequency level, STFT can achieve more fine-grained resolution on all frequencies. This allows detailed exhibition of CFR frequency when it changes over time (as illustrated in previous Fig. 5).

In our system, the sampling rate for each antenna pair is 500 Hz so that we can extract frequency range up to 250 Hz. The STFT algorithm extracts a total of 125 frequency components. Then the frequency granularity is 2 Hz. Such a frequency range and granularity are sufficient to cover all common activities.

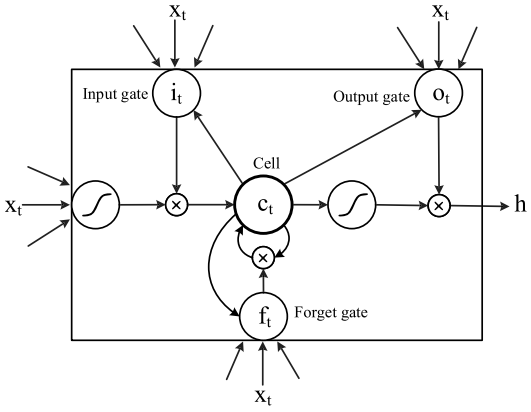

Fig. 10. Architecture of an LSTM cell.

B. Deep Learning Architecture

We first describe how we design the input to feed the deep neural network. The feature extraction stage has output the spectrogram of received signals for each Tx–Rx pair. Each spectrogram is a matrix and each element can be represented as m_{ij}, where i is the time scale index and j indicates a frequency component. Recall that we have $N = N_{\mathrm{Tx}} * N_{\mathrm{Rx}}$ observing antenna pairs. Here, we extract the spectrogram of one time slot and all the observing pairs to form an input matrix, which we refer to as an *input frame*. In our settings, the size of an input frame is $125 * N$, where 125 is the total frequency components extracted by STFT. The input layer then takes all the input frames $\boldsymbol{x} = \{x_1, x_2, \ldots, x_T\}$ of each time slot and further serves for the hidden layer.

The hidden layer in our deep learning architecture includes a CNN structure and an LSTM structure for activity identification. CNN is powerful in extracting the implicit spatial patterns, e.g., object location relationships and textures, and thus has been extensively used in the field of computer vision for activity recognition [22], [23]. We first use CNN to process each input frame since the CNN structure can effectively consider the spatial features from different antenna pairs. In our system, the input features of multiple antenna pairs in a time slot is a 2-D data frame, where a window filter in the CNN layer is applied to slide over the data frame to reduce the data frame into a 1-D vector. This can effectively integrate the observed features from multiple antenna pairs and reduce the data dimensions. For example, a falling activity may cause obvious frequency changes in some antenna pairs while leading to implicit frequency changes in the other pairs, where CNN can keep these changes for future processing. In our system, we use a one-layer CNN structure.

We stacked the outputs of the CNN layer across time as the input of the following LSTM layers [13]. LSTM is a variant of recurrent neural networks and is specifically designed for sequence processing of temporal data. Fig. 10 shows the structure of an LSTM cell. The input i_t and output gate o_t incorporate the incoming and outgoing signals to the memory cell, and the forget gate f_t controls whether to forget the previous state of the memory cell. Each LSTM cell maintains a floating point value c_t, which may be diminished or erased through a multiplicative interaction with forget gate f_t or be additively updated by the current input multiplied by the activation of input gate i_t. The final emission of the memory value from the LSTM cell is determined by output gate o_t. The calculation process can be represented as follows:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \tag{5}$$

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right) \tag{6}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \tag{7}$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{8}$$

$$h_t = o_t \tanh(c_t) \tag{9}$$

where the b terms denote the bias vector, the W terms denote weight matrices, σ is the sigmoid function, and tanh is the hyperbolic tangent function. The cell output h_t is used to predict the label of the current training instance.

The LSTM architecture reveals many unique advantages compared to conventional prediction methods used in activity recognition, such as HMM [8] and kNNs [10]. First, LSTM is capable of mining the hidden relationships of time series. It can combine the current inputs and the past states stored in the memory cell to exploit the time scale relationships and achieve a comprehensive classification. Specially, it is able to learn long-term dependencies, which aligns well with our activity recognition context since some activities may last several seconds. Besides, LSTM's sophisticated network structure enable itself with strong representation ability from raw data input, thus requires little efforts on feature extraction. Conventional models [8] rely on manually selected features, such as speed, acceleration, etc., which not only exerts extra overhead but also can cause low accuracy due to incomprehensive extraction. Moreover, LSTM can easily support the input with various length, which is important since different activities can have different time durations.

The extracted features from the LSTM layers are then fed to a fully connected layer, which is widely used to avoid overfitting [24]. Besides, we also use dropout mechanism in our network to further avoid overfitting. The last layer of our deep learning architecture is an output layer, which receives the outputs from the last LSTM layer and normalizes them with a softmax function. This function computes the distribution probabilities of each activity and the one with the highest probability is finally labeled as the predicted activity.

In this learning process, we mainly use a CNN+LSTM architecture. CNN can effectively integrate the spatial features of different antenna pairs together. LSTM is capable of integrating the temporal features together for recognition. With sufficient data support, such learning architecture applies well in our context for activity recognition.

VI. Implementation and Evaluation

A. Implementation

We have implemented the WiSDAR system with COTS hardware that is readily available. We use two Dell Latitude D820 laptops both equipped with an Intel 5300 WiFi card as the transmitter and receiver. Since each WiFi card has three antennas, we can construct at most nine SA pairs. Note that we use the SIMO mode (1 antenna sends and 3 antennas receive) instead of using the MIMO mode (3 antennas send and 3 antennas receive) in our experiment. This is because we find that different antenna pairs have very strong coherence in MIMO mode, while in the SIMO mode the features of different antennas are independent. Therefore, to fully utilize the three antennas in the transmitter, we develop a time division transmitting mechanism that every antenna sends a packet and

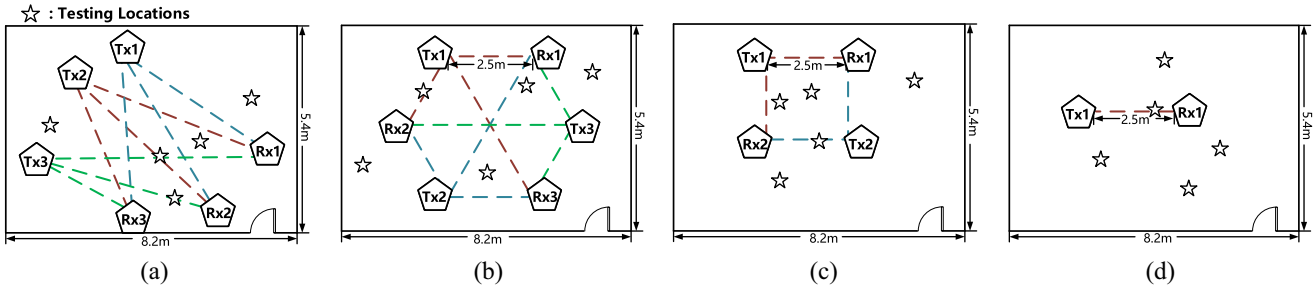

Fig. 11. Different antenna placement topologies. Stars are the testing locations. (a) Random shape, 3 × 3 pairs. (b) Hexagon shape, 3 × 3 pairs. (c) Square shape, 2 × 2 pairs. (d) Line shape, 1 × 1 pair.

TABLE I
ACTIVITY DATASET AND THE COLLECTED SAMPLES

Activity	# Samples	Activity	# Samples
Walking (Wa)	850	Picking (Pk)	710
Falling (Fa)	480	Pushing (Ps)	760
Running (Rn)	850	Waving (Wv)	760
Sitting (St)	590	Boxing (Bx)	760

then switch to the next antenna. When the last antenna finishes sending, the system jumps to the first antenna to begin a new cycle. The time division mechanism is equivalent to the MIMO mode and allows us to construct Tx × Rx pairs. Both the transmitter NIC and receiver NIC are working in the monitor mode so that we can distinguish different transmitting antennas by sending self-defined packets.

Our measurement shows that the switching antenna time is less than 1 ms, which is of little overhead to our 500 Hz sampling rate. At the receiver, we use the CSI tool [21] to collect CSI values from WiFi frames. We use 5 GHz WiFi channels with 20 MHz bandwidth carriers throughout our experiment.

B. Evaluation Setup

We collect 5760 training samples for eight activities (in Table I) in our laboratory (8.2 m × 5.4 m) under different antenna topologies as illustrated in Fig. 11. In our experiment, we selected eight representative activities, which can be divided into two categories, i.e., torso-based activities and gesture-based activities. The torso-based activities mostly reflect the radio signals using the human torso, which is a relatively large area. Yet the gesture-based activities mostly reflect the radio signals using hands and arms, which are relatively small areas. Our settings comprehensively considered all the two categories. Besides, these activities are common and representative activities in our daily life, e.g., walking, sitting down, and waving hands. We select such representative activities also following those state-of-the-art works [8], [11].

The topology includes a *line shape*, a *hexagon shape*, a *square shape*, and a *random shape*. The line shape confines all transceiver pairs to a single line, which essentially reduces to the case with only one effective transceiver pair. It therefore serves as a baseline for comparison between our WiSDAR and state-of-the-art solutions, in particular, CARM [8]. The hexagon shape and the square shape are regular topologies, which can completely eliminate IAs with strategically placed transceivers in our system. The random shape is likely to eliminate the impact of IAs, and can be the most convenient deployment in practice. The random deployment distance are 4.2 m for Tx1–Rx1, 4.8 m for Tx2–Rx2, and 2.7 m for Tx3–Rx3, respectively.

Typically, it is assumed that, in free space radio propagation, there is no obstruction or reflection in the first 8–12 Fresnel zones [18], [25]. To reduce the impact of the radio reflection by the ground, we keep the first 12 Fresnel zones clear and set the height of antenna as 0.8 m above the ground. The data collection spots are randomly distributed in the laboratory and we ask the target to perform activities toward different orientations. For walking and running, the target moves for a short distance (about 2 m) in a straight line, where the starting locations are randomly selected in the experiment environment.

Note that a large amount of data is helpful to improve the classification accuracy. Hence, we utilize *data augmentation* to expand our effective dataset. For example, since the walking and running activities are continuous, we can divide a long data sequence into multiple sequences with different length. Most activities are symmetric so that we can also reverse features in the time domain for augmentation. For example, the walking and running activities are conducted along a straight line so that the reversed features of an activity toward one direction is similar to the features of an activity toward the reversed direction. Similarly, picking, waving, and boxing are all reciprocating movement, e.g., a person will first stretch out the arm and then retract the arm for boxing, which can be treated as a symmetric activity.

Once the model is trained, we test it for six volunteers (both male and female students varying in height, weight, and age) in four locations under these topologies. We train our model in the laboratory, while we test the accuracy in three other environments besides the laboratory, including a 40 m × 10 m big hall, an apartment with an area of 35 m^2 and a small office with an area of 18 m^2, respectively. The testing spots of the activities are marked as stars in Fig. 11 considering both the IA and the EAs. In our evaluation, the default parameter settings are using hexagon shape with deep learning methods, and LSTM cell number is 128. We train our deep neural network model using a testbed equipped with an NVIDIA Geforce GTX 1060 GPU card. The training time is about 10 min (the detailed time for different setting can be found in Table III) and the inference time is only 0.1 s for an activity, which do not incur much system overhead.

C. Evaluation on Target Area Determination and Activity Detection

We first evaluate the target area determination mechanism described in Section IV-B. We vary the power threshold δ_P and

TABLE II
ACCURACY AND FPR OF DIFFERENT SETTINGS ON POWER THRESHOLD δ_P AND TIME THRESHOLD δ_T

δ_P	2dB	2.5dB	2.5dB	3dB	3dB	3.5dB	3.5dB
δ_T	0.35s	0.35s	0.3s	0.3s	0.25s	0.25s	0.2s
Accuracy	1	0.74	1	0.85	0.91	0.65	0.81
FPR	0.21	0.05	0.02	0	0.02	0	0.01

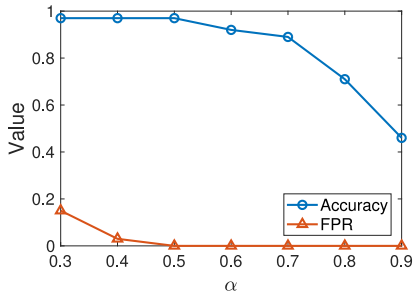

Fig. 12. Activity detection accuracy and FPR in different amplitude threshold settings.

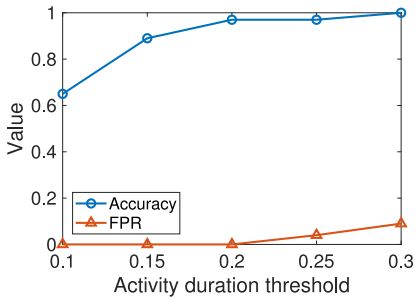

Fig. 13. Activity detection accuracy and FPR in different duration threshold settings.

time threshold δ_T to evaluate the impact on target area determination accuracy and false positive rate (FPR). As illustrated in Table II, we can find that the best setting is $\delta_P = 2.5$ dB and $\delta_T = 0.3$ s, where our system can detect all the activities in the IA and keep the FPR in a low level. When δ_P is set as 2 dB and δ_T is 0.35 s, the accuracy still achieves 100%, while the FPR is quite large. This is because low-frequency activities can also cause the amplitude falling below the threshold for a low time.

We then consider the impact of the activity amplitude threshold θ_P and activity duration threshold θ_L on the activity detection described in Section IV-C. We set $\theta_P = \alpha P_a + \beta P_b$, where $\alpha+\beta = 1$. Fig. 12 shows the activity detection accuracy with a range of α values when $\theta_L = 0.2$ s. We can observe that when $\alpha = 0.5$, the detection accuracy achieves 97% with no false positive result. If α is set as a large value (e.g., 0.9), the system will miss many activities since the amplitude threshold can be too large. Similarly, Fig. 13 varies θ_T when setting $\alpha = 0.5$. We can find that the detection accuracy increases as the threshold becomes more relaxed. Yet if we set a large θ_T, the FPR also rises a lot. Thus in our experiment, we set $\alpha = 0.5$ and $\theta_T = 0.2$ s.

As to the EA outside the FFZ, not all the activities therein can be detected and recognized. The detection range is determined by the distance between the target person and the transceivers as well as the radio signal strength. Since the radio signals will attenuate in air, the received reflected radio

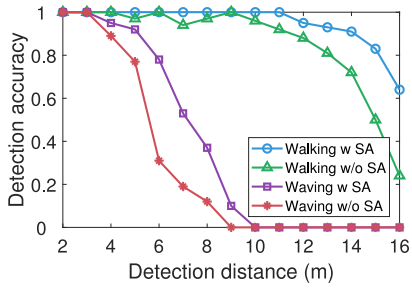

Fig. 14. Detection distance of different activities.

Actual Label	Predicted Label							
	Wa	Fa	Rn	St	Pk	Ps	Wv	Bx
Wa	1	0	0	0	0	0	0	0
Fa	0	1	0	0	0	0	0	0
Rn	0.03	0	0.97	0	0	0	0	0
St	0	0.02	0	0.93	0.05	0	0	0
Pk	0	0.01	0	0.04	0.95	0	0	0
Ps	0	0	0	0	0	0.92	0.02	0.06
Wv	0	0	0	0	0	0.02	0.97	0.01
Bx	0	0	0	0	0	0.03	0.01	0.96

Fig. 15. Confusion matrix of activity recognition.

signal can be too weak for recognition when the person is far away from the antenna pairs. Obviously, antenna pairs with higher power can have a longer detection range. In this paper, we mainly focus on using COTS WiFi devices. We consider the activity detection accuracy of our scheme, i.e., SAs and the baseline scheme, i.e., combined antenna (CA), in different distances to the center of the transceivers as illustrated in Fig. 14. We can find that the walking activity has an obvious longer detection distance than the waving hand activity, no matter using what kind of antenna settings. This is because walking is a torso-based activity with a large reflection area, while waving is a gesture-based activity with relatively small reflection area. For both the torso-based activity and the gesture-based activity, our SA scheme outperforms the CA scheme. When the detection distance is 6 m, SA can still achieve about 80% accuracy for waving activity detection, while CA only has 30% accuracy. For the walking activity, our SA scheme has more than 82% detection accuracy when the distance is as large as 15 m, while the accuracy of CA scheme falls below 50%. From these comparisons, we can observe that the SA scheme has a much longer effective detection distance than CA scheme.

D. Evaluation on Activity Recognition

1) Overall Evaluation: To comprehensively evaluate the classification result, the following metrics have been widely used: 1) false positive rate (FP) indicates the ratio of falsely selected activities as another activity; 2) precision (PR) is defined as [TP/(TP + FP)], where TP is the ratio of a correctly labeled activity; 3) recall (RE) is [TP/(TP+FN)], where FN is the false negative rate; and 4) $F1$-score ($F1$) is another evaluation metric, defined as $[(2 * \mathrm{PR} * \mathrm{RE})/(\mathrm{PR} + \mathrm{RE})]$.

Fig. 15 lists the confusion matrix of the eight activities under the hexagon topology in the laboratory with the target area detection scheme, where each row represents the actual activity and each column indicates the predicted activity. We find that the walking and falling activities can achieve 100% accuracy

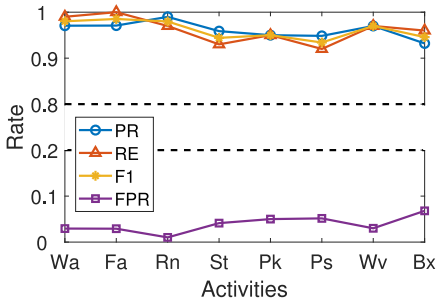

Fig. 16. Precision, recall, and FPR of all activities.

TABLE III
DIFFERENT STATISTICS METRICS AND TRAINING TIME WHEN USING DIFFERENT LSTM CELL NUMBER

LSTM cells	16	32	64	128	256	512
False-positive rate	0.249	0.114	0.08	0.042	0.039	0.057
Precision	0.751	0.886	0.92	0.962	0.961	0.943
Recall	0.792	0.853	0.912	0.958	0.956	0.931
F1-score	0.769	0.872	0.916	0.96	0.958	0.937
Training time	564s	589s	597s	602s	604s	609s

because these two activities have obvious unique features. The average recognition accuracy is 96% with a 2.3% standard deviation. The result shows that WiSDAR system achieves a high recognition accuracy among all activities.

To further understand the recognition result from a statistical view, we examine the FPR, precision, and recall as illustrated in Fig. 16. The FPR of all activities are below 10% with a mean value of 3%, and the precision and recall are all above 90% with both mean values of 96%. The result indicates that our WiSDAR can not only accurately but also comprehensively classify these different activities with low miss and error rates.

We next conduct fine-grained evaluations to examine the impact of different LSTM cell numbers, different topologies, the area detection scheme, different environments, different people, and different antenna distances, respectively.

2) Impact of Different LSTM Cell Numbers: Table III shows some statistics metrics of average activity recognition and the network training time when we set different LSTM cell numbers. We can find that when the total LSTM cell number is relatively low (less than 128), the value of precision, recall and $F1$-score all improve as the number of LSTM cells increase. When the LSTM cell number is 1024, the average statistics metrics achieve the highest value. This result indicates that with more LSTM cells, the deep learning model can better extract the inner features and achieve a high accuracy. Yet when the LSTM cell number keeps increasing, the precision, recall, and $F1$-score begin to drop (e.g., the precision drops from 0.962 to 0.943 when the LSTM cell number increases from 128 to 512). This is because too many nodes in the hidden layer can easily cause overfitting, which degrades the accuracy of activity recognition. In our training process, we set the epoch as 20. The total training time for the collected activity when using different LSTM cells does not show significant difference. The 10-min training time is not a high computational overhead given it is a one-time preparation.

3) Impact of Different Recognition Methods: We further examine the contribution of the SA scheme and the deep learning method, respectively. Fig. 17 considers four combinations, i.e., deep learning with SAs (DL+SA), deep learning with CAs (DL+CA), HMM with SAs (HMM+SA), and HMM with CAs (HMM+CA). The last one is essentially the state-of-the-art CARM solution. CARM can achieve an accuracy of 96% [8], however, only when the activities were measured in the EA. In this paper, we argue that the activity recognition accuracy can be largely affected when the target person is in the IA, while CARM did not consider this situation. We can observe that the HMM+CA has the lowest recognition accuracy among all the activities, and both SA scheme and the deep learning method can improve the accuracy. Our WiSDAR approach, i.e., DL+SA, has the highest recognition accuracy, achieving more than 92% accuracy for every activity. In contrast, HMM+CA only has a recognition accuracy of around 75% without considering the impact of spatial diversity.

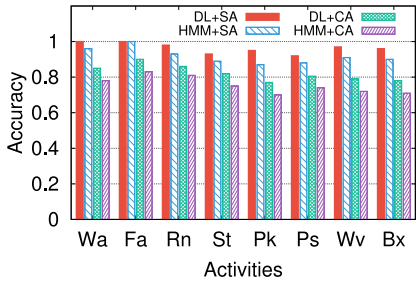

Fig. 17. Recognition accuracy using different recognition methods.

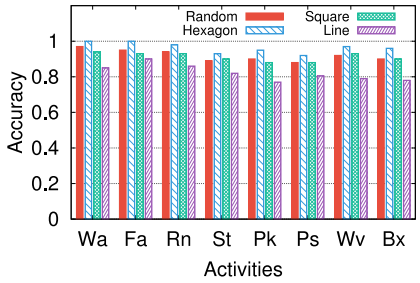

Fig. 18. Recognition accuracy under different antenna placement topologies.

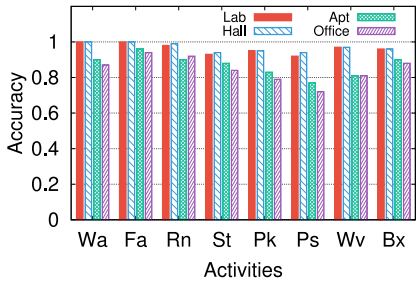

Fig. 19. Recognition accuracy in different environments.

4) Impact of Different Antenna Topologies: Fig. 18 shows the recognition results under different antenna topologies in the laboratory. The hexagon shape outperforms other topologies with fewer antenna pairs, which indicates that multiple observing pairs can better mitigate the impact of spatial diversity and obtain more useful features. Compared to the random shape with the same antenna pairs, the hexagon shape also has a higher accuracy. This is probably because the random shape has some spots that cannot be well observed by all pairs. The line shape has the worst performance with only 82% average accuracy over all activities. This is because its single effective feature fails in the IA and is largely affected by the spatial

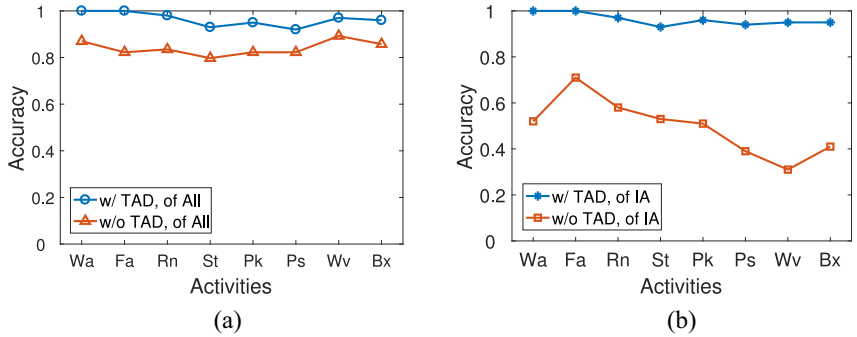

Fig. 20. Accuracy when with or without target area determination mechanism. (a) Accuracy relevant to all activities. (b) Accuracy relevant to activities in IA.

diversity, as described in Section II. From this comparison, we can know that deploying more observing pairs is helpful to improve the recognition accuracy. Besides, the antenna pairs should be distributed as evenly as possible to avoid the common IA and expend the EA.

5) Impact of Different Environments: Fig. 19 shows the recognition accuracy of activities under the hexagon topology in different environments. Among these locations, only the laboratory is trained while other locations are not trained. From this comparison, we can find that our WiSDAR system achieves an average accuracy of 96% in the laboratory and hall, which indicates that our approach has quite good recognition performance in the environment with weak multipath effects. For the environment with rich multipath effects, such as the apartment and the office, the general recognition accuracy is lower than that of the other two locations. Specifically, the recognition results are relatively good for most of the activities (e.g., Wa, Fa, Rn, St, and Bx), with an accuracy of around 90%, while for Pk, Ps, and Wv, the accuracy is around 80% in the apartment and laboratory. This is because in the very narrow indoor environment, the strong multipath effect can affect the received signal and further undermine the recognition accuracy. Furthermore, the activities with similar frequency features will be more noticeably affected. For example, pushing and waving are both gesture-based activities with similar frequency features, which can be easily misidentified as each other, leading to a relatively low accuracy.

6) Impact of Target Area Determination: Fig. 20 compares the different recognition accuracy under the hexagon topology. Fig. 20(a) shows the accuracy rate of recognized activities to all the performed activities. We can see that WiSDAR achieves an average of 96% accuracy with the target area detection scheme, while this accuracy falls to only an average of 82% when the IAs are neglected. Fig. 20(b) further explains this result by plotting the fraction rate of recognized activities to activities performed in IAs. From this figure, we know that in the IA of one antenna pair the recognition accuracy is only about 50% on average without target area detection, whereas it achieves 97% on average with target area detection. These results demonstrate that our target area detection scheme is necessary and effective.

7) Impact of Different People: We also examine the impact of human diversity on activity recognition in our experiment. We have six volunteers for testing and they vary in gender, height, and weight. Fig. 21 shows the recognition accuracy under hexagon topology in the laboratory for these people. We find that the recognition result does not show noticeable diversity among different people. WiSDAR achieves over

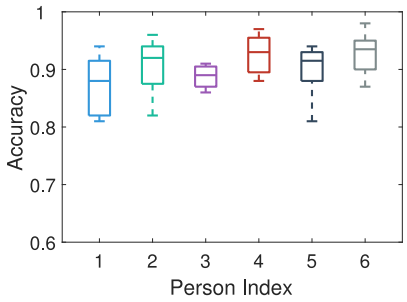

Fig. 21. Recognition accuracy of different people.

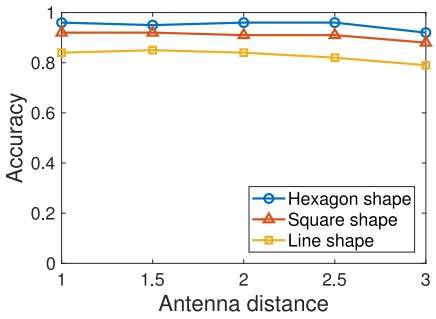

Fig. 22. Recognition accuracy with different antenna deployment distance.

85% of average recognition accuracy for strange people, which indicates that WiSDAR is resilient to different people.

8) Impact of Different Antenna Distances: We next consider the impact of the distance between the transmit antenna and the receive antenna on the activity recognition accuracy. Fig. 22 shows the recognition accuracy of three different topologies with a range of antenna distances. We can find that as the antenna distance increases from 1 m to 2.5 m, there is no obvious accuracy change for these settings. This result indicates that the antenna distance does not affect the accuracy much when the distance is not too long. Yet when the antenna distance becomes very long (e.g., more than 3 m), the recognition begins to decrease. This is because a very long antenna extended cable can cause the power attenuation during the transmission, which undermines the recognition accuracy.

VII. Related Work

State-of-the-art device-free WiFi-based human activity recognition systems can be broadly classified into two categories according to their hardware demands, i.e., the specialized hardware-based systems and the COTS-based systems.

A. Specialized Hardware-Based

Many systems use dedicated hardware with software defined radios to capture more fine-grained wireless signal metrics for activity recognition and other related applications. WiSee [5] uses USRP to measure the Doppler effects of WiFi signals caused by body movement to classify different motions and achieves an average accuracy of around 95% using two wireless sources. AllSee [26] uses an analog envelope-detection for gesture recognition by profiling the different pattern changes. With specialized hardware, micro-Doppler information can also be measured [27], [28], achieving very high recognizing resolution. The cost of such hardware in these systems however can be high, and their availability and compatibility are generally not good, either. Different from these specialized hardware-based systems, WiSDAR uses COTS devices for activity recognition.

B. COTS-Based

The other systems mainly use commodity laptops and WiFi NICs to measure the changes of CSI. Since these COTS devices are readily available in the market, their costs and compatibility are quite good, though the captured CSI can be limited and coarse-grained. WiGest [6] leverages the change of patterns in RSSI to sense in-air gestures. WiFall [7] uses fine-grained CSI changes to detect a single activity of falling. E-eye [9] profiles the CSI changes across multiple subcarriers to recognize both in-home activities and walking movements. CARM [8] builds up the correlation between movement velocity and CSI dynamics to recognize activities. Virmani and Shahzad [10] explored the relationship between CSI features and target location and orientation, and translates CSI measurements to other virtual samples for recognition. Besides the activity recognition, COTS-based WiFi sensing has also been widely used in other aspects, such as localization [29], tracking [14], [20], in-air drawing [30], vital sign monitoring [31], and recognizing typing [32], speaking [33], and dancing steps [34].

These COTS-based activity recognition systems neglect the impact of the spatial diversity, especially in the IAs where recognition can fail. Different from existing systems, WiSDAR proposes to extend the WiFi antennas to capture features from multiple spatial dimensions, and leverage an advanced deep learning tool to integrate features from both temporal and spatial dimensions to achieve highly accurate and reliable recognition.

VIII. Discussion

A. Benefits of Using Separated Antennas

Compared with using multiple transceiver pairs, using SAs within one physical sender and receiver reveals two key advantages. First, using a single transceiver pair can reduce the equipment cost effectively. It is not cost-effective and convenient to deploy too many WiFi APs in a small indoor environment. Moreover, our approach provides easy synchronization among the transceivers, where the received signals from different antennas are automatically synchronized. It not only provides great convenience for subsequent signal processing and further real-time online activity prediction but also is easy to manage. Yet using multiple transceivers for observation inevitably incurs much overhead on the synchronization among all the receivers. A small synchronization error can have a dramatic impact on signal processing and further undermine the activity recognition accuracy.

B. Attenuation With Extended Antennas

Since WiSDAR separates the WiFi antennas by extended cables, the signal power can have an attenuation. In our deployment, we use low loss coaxial cable such as LMR400 cable. Even we use 3 m extended cable for each antenna, the total power attenuation is only 2.2 dB [35], which has little impact on the CSI measurement as well as the recognition results.

C. Scalability and Multiperson Recognition

WiSDAR can be easily extended to support using multiple WiFi devices for activity recognition. For the transmitters, we just need to combine each transmitter and schedule each transmitting antenna to send packets one after another repeatedly. For the receivers, since each antenna receives wireless signal independently, we just need to collect all the features and integrate them together for recognition. WiSDAR now only supports activity recognition for one person. We leave multiperson activity recognition as the future work.

IX. Conclusion

In this paper, we proposed WiSDAR, a WiFi-based spatial diversity-aware device-free activity recognition system. Due to the spatial diversity of the target areas and the observing transceivers, the CSI characteristics can be largely affected and lead to inaccurate activity recognition results. The time domain and the frequency domain features have been considered individually in the literature of WiFi-based activity recognition; in this paper, we considered them jointly and examined their specific impacts on human activity recognition. Our key innovation is extending the multiple antennas of modern WiFi devices to construct multiple SA pairs and obtain features from multiple spatial dimensions. We have also proposed a deep learning-based architecture to effectively process the derived rich information. We implemented our system with commercial WiFi cards and conducted real-world evaluation to examine the performance. The result demonstrated that WiSDAR achieved an average of 96% activity recognition accuracy.

References

[1] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," *ACM Comput. Surveys (CSUR)*, vol. 43, no. 3, p. 16, 2011.

[2] E. Ertin *et al.*, "AutoSense: Unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field," in *Proc. 9th SenSys*, 2011, pp. 274–287.

[3] K. Yatani and K. N. Truong, "BodyScope: A wearable acoustic sensor for activity recognition," in *Proc. 14th UbiComp*, 2012, pp. 341–350.

[4] H. Ding *et al.*, "FEMO: A platform for free-weight exercise monitoring with RFIDS," in *Proc. 13th SenSys*, 2015, pp. 141–154.

[5] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th MobiCom*, 2013, pp. 27–38.

[6] H. Abdelnasser, M. Youssef, and K. A. Harras, "WiGest: A ubiquitous WiFi-based gesture recognition system," in *Proc. IEEE INFOCOM*, 2015, pp. 1472–1480.

[7] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput. (TMC)*, vol. 16, no. 2, pp. 581–594, Feb. 2017.

[8] 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. (MobiCom)*, 2015, pp. 65–76.
[9] Y. Wang *et al.*, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. 20th MobiCom*, 2014, pp. 617–628.
[10] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using WiFi," in *Proc. 15th MobiSys*, 2017, pp. 252–264.
[11] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Areas Commun. (JSAC)*, vol. 35, no. 5, pp. 1118–1131, May 2017.
[12] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[14] 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 MobiHoc*, 2017, p. 6.
[15] A. F. Molisch, *Wireless Communications*. Hoboken, NJ, USA: Wiley, 2012.
[16] D. D. Coleman and D. A. Westcott, *CWNA: Certified Wireless Network Administrator Official Study Guide: Exam Pw0-105*. Hoboken, NJ, USA: Wiley, 2012.
[17] C. Liu *et al.*, "RSS distribution-based passive localization and its application in sensor networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2883–2895, Apr. 2016.
[18] H. Wang *et al.*, "Human respiration detection with commodity WiFi devices: Do user location and body orientation matter?" in *Proc. UbiComp*, 2016, pp. 25–36.
[19] F. Zhang *et al.*, "From Fresnel diffraction model to fine-grained human respiration sensing with commodity Wi-Fi devices," in *Proc. UbiComp*, vol. 2, p. 53, 2018.
[20] D. Wu, D. Zhang, C. Xu, Y. Wang, and H. Wang, "WiDir: Walking direction estimation using wireless signals," in *Proc. UbiComp*, 2016, pp. 351–362.
[21] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 159–170, 2010.
[22] J. Donahue *et al.*, "Long-term recurrent convolutional networks for visual recognition and description," in *Proc. CVPR*, 2015, pp. 2625–2634.
[23] J. Y.-H. Ng *et al.*, "Beyond short snippets: Deep networks for video classification," in *Proc. IEEE CVPR*, 2015, pp. 4694–4702.
[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.
[25] H. D. Hristov, *Fresnel Zones in Wireless Links, Zone Plate Lenses and Antennas*. Boston, MA, USA: Artech House, 2000.
[26] B. Kellogg, V. Talla, and S. Gollakota, "Bringing gesture recognition to all devices," in *Proc. 11th USENIX NSDI*, vol. 14, 2014, pp. 303–316.
[27] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," in *Proc. 11th USENIX NSDI*, vol. 14, 2014, pp. 317–329.
[28] F. Adib, Z. Kabelac, and D. Katabi, "Multi-person localization via RF body reflections," in *Proc. 12th USENIX NSDI*, 2015, pp. 279–292.
[29] W. Gong and J. Liu, "Robust indoor wireless localization using sparse recovery," in *Proc. IEEE 37th ICDCS*, 2017, pp. 847–856.
[30] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "WiDraw: Enabling hands-free drawing in the air on commodity WiFi devices," in *Proc. 21st MobiCom*, 2015, pp. 77–89.
[31] X. Wang, C. Yang, and S. Mao, "PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices," in *Proc. IEEE 37th ICDCS*, 2017, pp. 1230–1239.
[32] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using WiFi signals," in *Proc. 21st MobiCom*, 2015, pp. 90–102.
[33] 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.
[34] K. Qian *et al.*, "Inferring motion direction using commodity Wi-Fi for interactive exergames," in *Proc. CHI*, 2017, pp. 1961–1972.
[35] *Lmr400 Cable*. Accessed: Sep. 12, 2018. [Online]. Available: https://www.timesmicrowave.com/documents/resources/LMR-400.pdf

Fangxin Wang (S'15) received the B.S. and M.S. degrees from the Department of Computer Science and Technology, Beijing University of Post and Telecommunication, Beijing, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree at the School of Computing Science, Simon Fraser University, Burnaby, BC, Canada.

His current research interests include wireless networks, big data analysis, and machine learning.

Wei Gong (M'14) received the B.E. degree in computer science from the Huazhong University of Science and Technology, Wuhan, China, and the M.E. degree in software engineering and the Ph.D. degree in computer science from Tsinghua University, Beijing, China.

He is a Professor with the School of Computer Science and Technology, University of Science and Technology of China, Hefei, China. He had also conducted research with Simon Fraser University, Burnaby, BC, Canada, and the University of Ottawa, Ottawa, ON, Canada. His current research interests include wireless networks, Internet-of-Things, and distributed computing.

Jiangchuan Liu (S'01–M'03–SM'08–F'17) received the B.Eng. degree (*cum laude*) in computer science from Tsinghua University, Beijing, China, in 1999, and the Ph.D. degree in computer science from the Hong Kong University of Science and Technology, Hong Kong, in 2003.

He is currently a Full Professor (with University Professorship) with the School of Computing Science, Simon Fraser University, Burnaby, BC, Canada.

Dr. Liu was a co-recipient of the Test of Time Paper Award of IEEE INFOCOM in 2015, the ACM TOMCCAP Nicolas D. Georganas Best Paper Award in 2013, and the ACM Multimedia Best Paper Award in 2012. He is a Steering Committee member of the IEEE TRANSACTIONS ON MOBILE COMPUTING and an Associate Editor of the IEEE/ACM TRANSACTIONS ON NETWORKING, the IEEE TRANSACTIONS ON BIG DATA, and the IEEE TRANSACTIONS ON MULTIMEDIA. He is an NSERC E. W. R. Steacie Memorial Fellow.