WiFi-based Activity Recognition using Activity Filter and Enhanced Correlation with Deep Learning

Zhenguo Shi, J. Andrew Zhang, Richard Yida Xu, Qingqing Cheng
Global Big Data Technologies Center
University of Technology Sydney, Sydney, Australia
{zhenguo.shi, qingqing.cheng}@student.uts.edu.au, {andrew.zhang, yida.xu}@uts.edu.au

***Abstract*—Device-free WiFi sensing utilizing channel state information (CSI) is attractive for human activity recognition (HAR). However, several challenging problems are yet to be resolved, e.g., difficulty in extracting proper features from input signals, susceptibility to the phase shift of CSI and difficulty in identifying similar behaviors (e.g., lying and standing). In this paper, we aim to tackle these problems by proposing a novel scheme for CSI-based HAR that uses activity filter-based deep learning network (HAR-AF-DLN) with enhanced correlation features. We first develop a novel CSI compensation and enhancement (CCE) method to compensate for the timing offset between the WiFi transmitter and receiver, enhance activity-related signals and reduce the dimension of inputs to DLN. Then, we design a novel activity filter (AF) to differentiate similar activities (e.g., standing and lying) based on the enhanced CSI correlation features obtained from CCE. Extensive simulation results demonstrate that our proposed HAR-AF-DLN scheme outperforms state-of-the-art methods with significantly improved recognition accuracy (especially for similar activities) and notably reduced training time.**

***Index Terms*—Human activity recognition, WiFi, Device free sensing, Channel state information, Deep learning networks.**

I. Introduction

Recent advances in device-free WiFi sensing demonstrate its great potential for human activity recognition (HAR) [1]. Different from conventional device-based sensing techniques that require assisted devices (e.g., wearable devices), WiFi-based HAR exploits the information in received WiFi signals modulated by human activities. As a result, WiFi-based HAR has attractive properties such as low-cost, privacy protection and convenience. Both received signal strength (RSS) and channel state information (CSI) can be used for WiFi HAR [2]. The performance of RSSI-based HAR is readily influenced by noise and shadowing fading [3]. In contrast, CSI can provide more fine-grained information, e.g., the phase information and frequency diversity. Hence, CSI-based HAR can generally lead to higher recognition accuracy [4].

Recently, applying deep learning (DL) to CSI-based HAR receives strong research interest. Assisted by signal processing techniques, DL-based HAR methods are able to automatically extract deep features from input signals, significantly improving recognition performance [5], [6]. The authors in [7] used recurrent neural networking (RNN) to extract hidden features from the raw CSI. To improve the quality of CSI, the authors in [8] proposed a feature enhancement scheme and employed RNN for feature extraction. In [9], the authors adopted an Autoencoder Long-term Recurrent Convolutional Network framework (AE-LRCN) to extract high-level representative features in CSI. However, these two methods are susceptible to the phase shift in CSI caused by timing offset between the WiFi transmitter and receiver. A small mismatch in the phase domain of CSI can result in notable performance degradation. To estimate and compensate for the timing offset, a few solutions have been proposed. For example, a linear fitting method is developed in [10] and a phase calibration approach is proposed in [11], but both schemes have high computational complexity. In addition, most of the existing algorithms are not very effective in identifying similar activities, such as standing and lying.

In this paper, we propose a novel scheme for CSI-based HAR using activity filter-based deep learning network (HAR-AF-DLN) with enhanced correlation features. Our scheme can effectively solve the phase mismatch problem caused by timing offset and significantly improve the identification accuracy for similar activities. Two major innovations in HAR-AF-DLN include CSI compensation and enhancement (CCE) and activity filter (AF). The CCE method is proposed to compensate for the timing offset between the WiFi transmitter and receiver, so as to improve the quality of CSI. Besides, CCE can also enhance activity-related signals and reduce the dimension of signals input to DLN, thereby increasing recognition accuracy with less complexity. The AF method is designed to distinguish similar activities (e.g., lying and standing) using the enhanced CSI correlation features obtained from CCE. Extensive experimental results demonstrate that our proposed HAR-AF-DLN scheme is superior to state-of-art HAR schemes, with less training time and higher recognition accuracy (especially for similar activities).

II. The HAR-AF-DLN Scheme

We illustrate the diagram of the proposed HAR-AF-DLN scheme in Fig. 1, including three main modules: CSI Collection, CSI Preprocessing and Activity Recognition. Next, we briefly review each module, and detail the last two modules in Section III and Section IV, respectively.

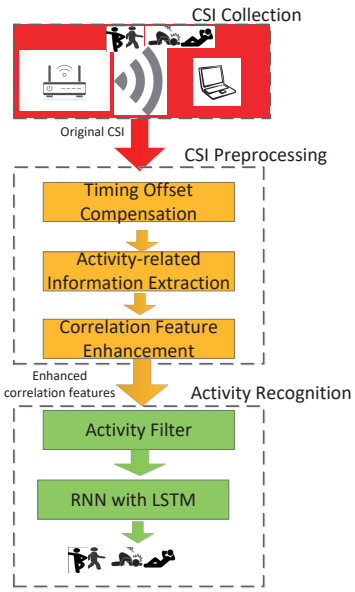

Fig. 1. Main processing modules of the HAR-AF-DLN Scheme.

A. CSI Collection

In this module, we apply the WiFi receiver to collect the CSI reflecting variations that human activities induce on wireless channels. As shown in Fig. 1, a person performs different activities in an indoor environment covered by a WiFi network, causing various changes to wireless channels, e.g., channel amplitudes, phase shift, and the number of multiple paths. Note that, these variations can be distinctive for different human behaviors, which can be utilized to facilitate activity classification. To do so, the Intel 5300 network interface card (NIC), a widely adopted commercial off-the-shelf (COTS) WiFi device, is utilized to acquire CSI. According to the protocol of IEEE 802.11n, we extract the CSI from 30 subcarriers for each pair of transmitter-receiver antennas using the CSI tools [12] (refer to Section V-A for the detailed experimental setup).

Let $\mathbf{h}(i)$ represent the CSI vector obtained from the ith packet

$$\mathbf{h}(i)=[h_{1,1}(i),\ldots,h_{1,m}(i),\ldots,h_{n,m}(i),\ldots,h_{N,M}(i)]^T, \quad (1)$$

where $h_{n,m}(i)$ means the CSI measurement at the mth subcarrier in the nth wireless link; M stands for the total number of available subcarriers in each wireless link; N denotes the total number of wireless links, and $N = N_t \times N_r$, where N_t and N_r represent the number of antennas at the transmitter and receiver, respectively; T is the transpose operation. The CSI matrix $\mathbf{H}$, acquired CSI vectors from I packets, is written as

$$\mathbf{H} = [\mathbf{h}(1),\ldots,\mathbf{h}(i)\ldots,\mathbf{h}(I)]. \quad (2)$$

B. CSI Preprocessing

The core tasks of this module include phase calibration for CSI, reduction of activity-unrelated information, compression of input signal to AF-DLN, and enhancement of feature signals. Note that it is not suitable to directly input the matrix $\mathbf{H}$ to DLN for HAR for the following two reasons. First, $\mathbf{H}$ contains a lot of activity-unrelated and environment-dependent information, e.g., multiple channel paths from static background objects. Such information poses a severe influence not only on the feature extraction in the subsequent processing but also the robustness of the recognition system. Second, since the size of $\mathbf{H}$ is very large, it is computationally expensive and time-consuming if applying $\mathbf{H}$ directly for HAR. To tackle these problems, we develop the CCE method with three key steps: *timing offset compensation*, *activity-related information extraction* and *correlation feature extraction*.

In the first two steps, we estimate the phase shift caused by the timing offset and make a compensation to improve the CSI quality. We then construct the CSI for activity-unrelated information (i.e., static objects) by utilizing a linear recursive operation, followed by subtracting it from the received signal. Ideally, activity-unrelated information can be largely reduced. In the third and last step, we compute the correlation feature matrix (CFM) of the output matrix obtained from Step 2. In comparison with the CSI matrix $\mathbf{H}$, CFM contains the condensed activity-related information with notably reduced dimension.

C. AF-DLN based Activity Recognition

This module intends to realize HAR with high sensing accuracy (especially for similar activities), by automatically extracting the hidden features from CFM obtained from Module 2.

To effectively recognize similar activities, we first propose an AF method to divide different activities into two groups (i.e., "light activity" and "intensive activity"). Next, for each group, we use one exclusive RNN network to automatically learn and extract hidden features from CFM, which can effectively differentiate different behaviors in the same group. To facilitate a successful HAR, we first train DLNs in an offline manner using the training data; Then we apply the well-trained networks to distinguish different human behaviors in an online manner.

III. CCE for CSI Preprocessing

In this section, we provide details of the designed CCE for CSI preprocessing. We will first discuss the first two steps, i.e., *timing offset compensation* and *activity-related information extraction*. Then we will describe the last step: *correlation feature extraction*.

A. Timing Offset Compensation and Activity-related Information Extraction

As aforementioned, these two steps aim to calibrate the CSI phase, reduce activity-unrelated information while retaining activity-related information. Consequently, it is possible to extract feature signals which are more activity-related and

less environment-dependent. To achieve that, $\mathbf{h}(i)$ in (2) is partitioned into two parts: dynamic CSI and static CSI

$$\mathbf{h}(i) = \mathbf{h}^{st}(i) + \mathbf{h}^{dy}(i), \tag{3}$$

where $\mathbf{h}^{st}(i)$ stands for the static CSI vector which is activity-unrelated; and $\mathbf{h}^{dy}(i)$ represents the dynamic CSI vector that is related to human activities. Notably, $\mathbf{h}^{st}(i)$ is the dominating component in $\mathbf{h}(i)$, and has much larger impact on $\mathbf{h}(i)$ than $\mathbf{h}^{dy}(i)$. The reason is that the influence of human behavior on the whole environment is generally limited, which is especially true when a person performs some minor actions, e.g., raising hands, sitting, standing, etc. Under such a situation, applying $\mathbf{h}(i)$ directly to HAR could degrade the recognition accuracy (refer to Fig. 6). Therefore, it is worthwhile to subtract the static information $\mathbf{h}^{st}(i)$ from $\mathbf{h}(i)$. To that end, we develop a recursive approach by referring to the exponentially weighted moving average (EWMA) algorithm [13].

However, one major problem here needs to be first resolved: the timing offsets between the WiFi transmitters and receivers are not clock-wise synchronized, which can vary over packets and cause linear phase shift of CSI. Therefore, before applying the recursive operation, estimation and compensation for the timing offset are required.

Let the recursive static CSI estimation at the i-th packet be $\widehat{\mathbf{h}}^{st}(i)$. The recursive operation over consecutive packets can be written as

$$\widehat{\mathbf{h}}^{st}(i) = \delta(\widehat{\mathbf{\Phi}}^{*}(i) \otimes \mathbf{I}_N)\mathbf{h}(i) + (1-\delta)\widehat{\mathbf{h}}^{st}(i-1), \tag{4}$$

where δ is the forgetting factor and is set with small value such as 0.01, the superscript $*$ represents conjugate of a matrix/vector, $\mathbf{I}_N$ stands for an $N \times N$ identity matrix, $\otimes$ denotes the Kronecker product, $\widehat{\mathbf{\Phi}}(i) = \text{diag}\{\exp(j\widehat{\phi}_{m,i})\}$ depicts a diagonal matrix with the m-th diagonal element $\exp(j\widehat{\phi}_{m,i})$, and $\widehat{\phi}_{m,i}$ is an estimate of the actual $\phi_{m,i}$ associated with the timing offset. Note that a common local clock is typically used for signals from/to all antennas, therefore, the timing offset due to clock offset is the same for all antennas. The phase shift $\phi_{m,i}$ can be represented as

$$\phi_{m,i} = m\psi_i + \theta_i, \tag{5}$$

where ψ_i and θ_i stand for phase shifts related to the timing offset.

To estimate ψ_i and θ_i, we first compute the dot product $\odot$ between $\mathbf{h}(i)$ and $(\widehat{\mathbf{h}}^{st}(i-1))^*$, by

$$\begin{aligned}\mathbf{r}(i) \triangleq& \mathbf{h}(i) \odot (\widehat{\mathbf{h}}^{st}(i-1))^* \\ =& (\mathbf{h}^{st}(i) + \mathbf{h}^{dy}(i)) \odot (\widehat{\mathbf{h}}^{st}(i-1))^* \\ \approx& \mathbf{h}^{st}(i) \odot (\widehat{\mathbf{h}}^{st}(i-1))^* \\ \approx& (\mathbf{\Phi}(i) \otimes \mathbf{I}_N)|\widehat{\mathbf{h}}^{st}(i-1)|^2,\end{aligned} \tag{6}$$

where $|\widehat{\mathbf{h}}^{st}(i-1)|^2$ means element-wise square of the absolute value. In (6), the first approximation is obtained based on the fact that the power of static paths are typically much more significant than dynamic ones, and the second approximation is based on the assumption that the estimate $\widehat{\mathbf{h}}^{st}(i-1)$ is close to the actual static CSI.

Then, $\mathbf{r}(i)$ is stacked into an $M \times N$ array, and each column contains CSI for one antenna. The mean over each row is computed, getting a new $M \times 1$ vector $\bar{\mathbf{r}}(i)$. Next, we compute the cross-correlation for neighbouring elements with equal spaced subcarrier indices in $\bar{\mathbf{r}}(i)$ and then compute the mean of the output, obtaining a sample denoted by γ_i. Then the estimate for ψ_i is given by

$$\widehat{\psi}_i = \angle(\gamma_i)/K_s, \tag{7}$$

where K_s represents the index intervals between the selected subcarriers which are equally spaced. In this paper, we use the Intel NIC5300 card in the experiments, $K_s = 2$.

Let $\bar{r}_{m,i}$ denote the m-th element in $\bar{\mathbf{r}}(i)$, then we estimate the parameter θ_i in (5) by

$$\widehat{\theta}_i = \angle\left(\sum_m \bar{r}_{m,i} e^{-jm\hat{\psi}_i}\right), \tag{8}$$

where the sum operation is conducted over a selected number of samples with larger energy for mitigating the noise.

Then, the estimate $\widehat{\mathbf{\Phi}}(i)$ and the recursive output $\widehat{\mathbf{h}}^{st}(i)$ can be obtained, respectively. Notably, we obtain the initial value of $\widehat{\mathbf{h}}^{st}(1)$ using (4) in a quiet environment.

As a result, the estimated value of dynamic CSI $\widehat{\mathbf{h}}^{dy}(i)$, obtained at the ith packet, can be expressed as

$$\widehat{\mathbf{h}}^{dy}(i) = (\widehat{\mathbf{\Phi}}^{*}(i) \otimes \boldsymbol{I}_N)\mathbf{h}(i) - \widehat{\mathbf{h}}^{st}(i). \tag{9}$$

The whole estimated dynamic CSI matrix $\widehat{\mathbf{H}}^{dy}$, over I packets, is represented as

$$\widehat{\mathbf{H}}^{dy} = [\widehat{\mathbf{h}}^{dy}(1), \ldots, \widehat{\mathbf{h}}^{dy}(i), \ldots, \widehat{\mathbf{h}}^{dy}(I)]. \tag{10}$$

Note that, the information contained in $\widehat{\mathbf{H}}^{dy}$ is mostly activity-related, hence, it can be utilized to extract more distinctive features that are less environment-dependent for HAR.

B. Correlation Feature Extraction

It is important to note that a person's activity can be divided into different stages. We can use a feature signal to represent each stage, and different stages are mutually dependent. Take the activity "stands up" as an example: a series of stages are involved during this process, from static, standing with accelerating, standing up with decelerating to standing still. The features of different stages, e.g., speed and spatial positions, are different but mutually correlated. Notably, all the feature signals for such an activity are hidden in $\widehat{\mathbf{H}}^{dy}$. Besides, there are also correlations between $\widehat{\mathbf{H}}^{dy}$ in different subcarriers, which provides additional information for HAR.

We can hence compute the correlation between signals at all subcarriers from all wireless links, given by

$$\mathbf{D}^{dy} = \widehat{\mathbf{H}}^{dy} \times (\widehat{\mathbf{H}}^{dy})^T, \tag{11}$$

where $\mathbf{D}^{dy}$ represents the correlation feature matrix (CFM). Note that, the size of $\mathbf{D}^{dy}$ $(MN \times MN)$ is significantly

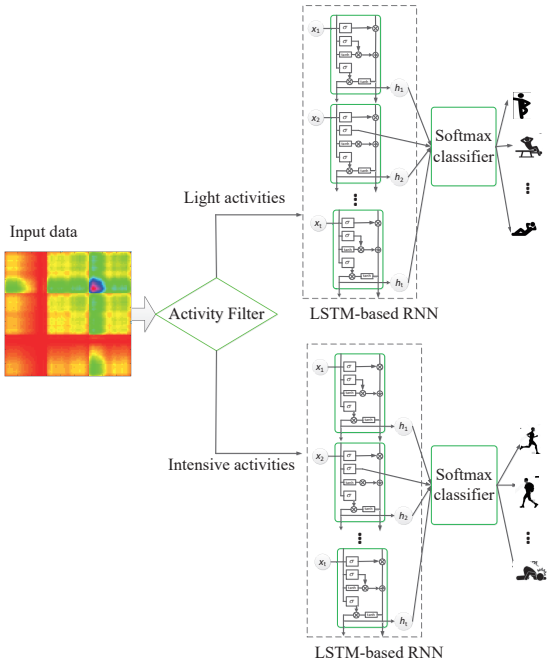

Fig. 2. Structure of AF-DLN based activity recognition using CFM as input.

smaller than the size of $\widehat{\mathbf{H}}^{dy}$ $(MN \times I)$. Consequently, inputting $\mathbf{D}^{dy}$, instead of $\widehat{\mathbf{H}}^{dy}$, into DLN for training process can considerably reduce the computational complexity. We provide details of performance assessment for CCE in Section V.

IV. AF-DLN based Human Activity Recognition

In this section, we present details of the proposed AF-DLN method, as depicted in Fig. 2. The method includes two main steps: activity filter (AF), and deeper feature extraction and classification. Note that, the first step divides similar activities into the same group, which allows DLNs (in the second step) to focus on the feature extraction of similar motions. Consequently, more distinctive characteristics of similar movements can be extracted, compared with obtaining features from all the activities, which is beneficial to classify these similar behaviors. The details of performance assessment for AF-DLN is provided in Fig. 7.

A. Step 1: Activity filter (AF)

According to the intensity and range of motions, human activities can be divided into two main groups: light activity and intensive activities. The former group refers to the activities with low intensity and small movement range, including e.g. *lying, standing, empty room, sitting*, which cause less CSI variation. The latter group refers to the activities with high intensity and large movement ranges, including, e.g., *walk, fall, running*, which cause larger CSI variation. The core task of AF is to determine which group the input signals $\mathbf{D}^{dy}$ belongs to (i.e., "light activity" or "intensive activity"). To achieve that, we apply a singular value decomposition (SVD) to $\mathbf{D}^{dy}$ and obtain its singular values, by

$$\Lambda = \text{svd}(\mathbf{D}^{dy}), \tag{12}$$

where svd(.) stands for the SVD operation, and $\Lambda \triangleq [\lambda_1, \lambda_2, \ldots, \lambda_{MN}]$ represents the vector containing singular values of $\mathbf{D}^{dy}$ in the descending order. Since the first two singular values (i.e., λ_1 and λ_2) contain most environment-dependent information [14], we adopt λ_3 as the metric for dividing human activities into two groups. To be specific, if λ_3 is smaller than a threshold β that is obtained empirically, the signal $\mathbf{D}^{dy}$ is divided into the "light activity" group, otherwise the "intensive activity" group.

B. Step 2: Deeper Feature Extraction and Classification

In this step, for each group (i.e., "light activity" or "intensive activity"), we train one DL architecture to distinguish human activities. For each DL architecture, we first apply RNN with LSTM to automatically learn and extract the hidden features from $\mathbf{D}^{dy}$. We then utilize the softmax regression algorithm for classification using the extracted deeper features from $\mathbf{D}^{dy}$. The process of the proposed AF-DLN is illustrated in Fig. 2.

It is noteworthy that conventional RNN-based sensing methods generally have time-consuming training processes due to the large size of training data. In contrast, our proposed HAR-AF-DLN scheme can effectively complete the training process with significantly less training overhead by using $\mathbf{D}^{dy}$ with largely reduced size of input data.

V. Implementation and Evaluation

In this section, we carry out extensive experiments to evaluate the performance of the proposed HAR-AF-DLN scheme.

A. Experimental Setup

To implement our proposed HAR-AF-DLN, two computers equipped with Intel WiFi 5300 network card are adopted as the transmitter and receiver, respectively. The transmitter continuously sends its packets with its single antenna ($N_t = 1$) at 5.32GHz frequency band, which follows the protocol of IEEE 802.11n. The receiver, which uses the CSI tools [12], collects and stores CSI with three antennas ($N_r = 3$) for 30 subcarriers ($S = 30$). A person performs six activities in total: stand up, standing, walk, fall, lying, and empty room. Since the rate of samples is 1kHz, the CSI matrix ($\mathbf{H}$) has size of 90×1000. We use a 3.4GHz PC with Nvidia P5000 graphic card (16GB memory) to train the presented HAR-AF-DLN. The total number of training iterations is 2000. We use LSTM-RNN with three hidden layers, and the hidden units for each layer is 200. We set the batch size and learning rate as 64 and 0.001, respectively. We empirically set the value of the threshold β of AF method to 0.6.

We conduct the experiments of the designed HAR-AF-DLN in two indoor configurations with different environmental complexities. Fig. 3 illustrates the layout of each indoor configurations. The first, with several obstacles between the transmitter and receiver, is a 4m$\times$ 6m meeting room. The second, with many obstacles between the transmitter and receiver, is a 8m$\times$ 10m laboratory room. Both the training

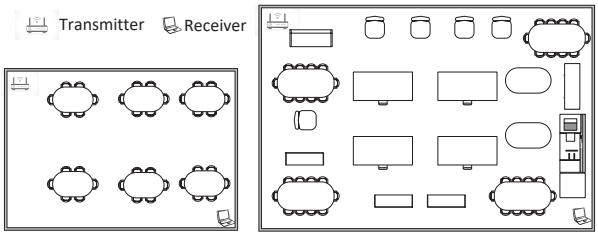

Fig. 3. Layout of two indoor experimental areas:(a) $4m \times 6m$ meeting room. (b) $8m \times 10m$ laboratory.

TABLE I
AVERAGE SENSING ACCURACY OF THE THREE METHODS IN THE TWO INDOOR CONFIGURATIONS

Methods \ Experiment	1st Exp.	2nd Exp.
Proposed HAR-AF-DLN	98.4%	93.4%
RNN [7]	87.5%	78.4%
DLN-eCSI [8]	94.2%	90.7%
AE-LRCN [9]	92.4%	89.7%

and testing data sets of each indoor configuration include six activities, with 300 times for each activity.

B. Performance Evaluation

In this section, we first compare the performance of our proposed HAR-AF-DLN scheme with other three state-of-the-art methods (i.e., RNN [7], DLN-eCSI [8] and AE-LRCN [9]), taking various parameters and configurations into consideration. We then provide in-depth evaluations of the effect of CCE and AF-DLN on our proposed scheme, respectively.

1) Performance Comparison for Different Methods: Table I illustrates the average recognition accuracy of three methods for the six activities with different configurations and parameters. As can be seen, the proposed HAR-AF-DLN clearly outperforms the other two methods in both indoor configurations. Take the second configurations as an instance, our proposed HAR-AF-DLN can achieve the average accuracy at 93.4%. In contrast, the corresponding sensing accuracies for the other three methods (i.e., RNN [7], DLN-eCSI [8] and AE-LRCN [9]) are lower than 91%.

To examine the performance of each method in detail, we present the confusion matrix for six activities in the first configuration in Fig. 4. The proposed HAR-AF-DLN performs much better than the other three methods in identifying these activities, particularly in differentiating similar activities such as lying and standing.

Fig. 5 demonstrates the impact of the number of subcarriers on average sensing accuracy. The six activities are performed in the second experimental configuration. Clearly, with an increasing number of subcarriers, each sensing method can achieve better average recognition accuracy. In all the cases with different numbers of subcarriers, HAR-AF-DLN achieves higher sensing accuracy than the others.

We provide Table II to compare the training time for the four methods. The DLNs are trained using a 3.4GHz workstation with Nvidia P5000 graphic card (16GB memory).

Predicted activity

Actual activity	Empty	Standing	Fall	Walk	Stand up	Lying
Empty	1	0	0	0	0	0
Standing	0.014	0.973	0	0	0	0.013
Fall	0	0	0.992	0.004	0.004	0
Walk	0	0	0.01	0.982	0.018	0
Stand up	0	0	0.026	0.017	0.957	0
Lying	0	0	0	0	0	1

(a) Proposed HAR-AF-DLN

Predicted activity

Actual activity	Empty	Standing	Fall	Walk	Stand up	Lying
Empty	0.774	0.207	0	0.019	0	0
Standing	0.009	0.791	0.014	0.027	0.011	0.148
Fall	0	0	0.939	0.053	0.008	0
Walk	0	0.027	0.006	0.931	0.032	0.004
Stand up	0	0.009	0.043	0.017	0.897	0.034
Lying	0	0.058	0	0.007	0.015	0.92

(b) RNN [7]

Predicted activity

Actual activity	Empty	Standing	Fall	Walk	Stand up	Lying
Empty	0.971	0.01	0.007	0	0	0.012
Standing	0.014	0.938	0	0.002	0.003	0.043
Fall	0	0.001	0.922	0.032	0.044	0.001
Walk	0	0.007	0.01	0.962	0.021	0
Stand up	0.001	0.006	0.045	0.031	0.917	0
Lying	0.005	0.042	0.01	0	0.001	0.942

(c) DLN-eCSI [8]

Predicted activity

Actual activity	Empty	Standing	Fall	Walk	Stand up	Lying
Empty	0.83	0.132	0	0.038	0	0
Standing	0.063	0.896	0.002	0.014	0.002	0.023
Fall	0	0	1	0	0	0
Walk	0	0.01	0.044	0.944	0.002	0
Stand up	0	0.043	0.026	0.026	0.853	0.052
Lying	0	0	0.007	0	0.014	0.979

(d) AE-LRCN [9]

Fig. 4. Confusion matrix for different human activity recognition methods.

TABLE II
TRAINING TIME FOR DIFFERENT METHODS

Methods \ Hidden units	300	500
Proposed HAR-AF-DLN	632.2s	1753.1s
RNN [7]	3822.8s	7837.1s
DLN-eCSI [8]	657.3s	1831.4s
AE-LRCN [9]	5591.4s	8956.1s

The numbers of training samples and iterations are 1200 and 2000, respectively. Our proposed HAR-AF-DLN scheme is shown to have much less training time. This largely credits to the notably reduced dimension of the input matrix $\mathbf{D}^{dy}$.

2) Impacts of CCE and AF-DLN on Sensing Performance: We present Fig. 6 to show the impact of the proposed CCE method on sensing performance in both experimental configurations. The average sensing accuracy of the proposed method with CCE is significantly higher than that of the method without CCE. This is because the CCE method can compensate for the timing offset, reduce activity-unrelated

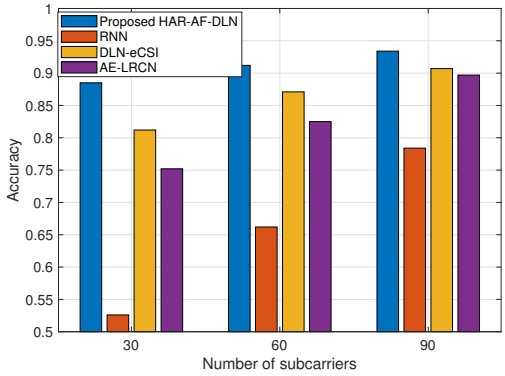

Fig. 5. Impact of the number of subcarriers on the recognition accuracy.

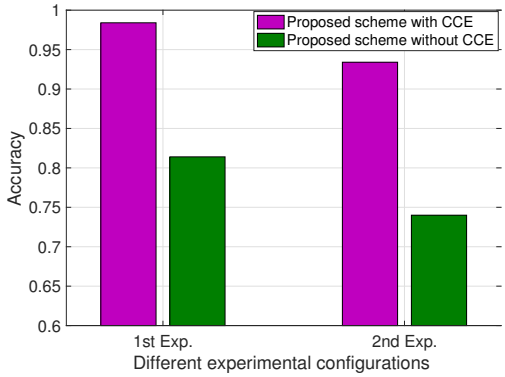

Fig. 6. Impact of CCE on recognition accuracy.

information and enhance activity-related information, which is beneficial to extract proper features for HAR.

The impact of the proposed AF method on recognition performance is shown in Fig. 7, under the second experimental configuration. From this figure, the sensing accuracy of the proposed method with AF in recognizing similar behaviors (e.g., standing and lying) is notably higher than that without AF. This is because the AF method can effectively distinguish similar activities, improving the recognition performance.

VI. Conclusion

In this work, we developed a HAR-AF-DLN scheme for human activity recognition, which consists of novel CCE and AF methods. The CCE method can compensate for the timing offset, enhance the activity-related signals and reduce the dimension of input signals to DLN. The AF method is able to distinguish similar activities based on the enhanced CSI correlation features achieved from CCE. Through extensive experimental results, we validate that our proposed HAR-AF-DLN scheme is superior to state-of-the-art methods in terms of recognition accuracy and training complexity.

References

[1] J. Yang, H. Zou, H. Jiang, and L. Xie, "Device-free occupant activity sensing using wifi-enabled iot devices for smart homes," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3991–4002, Oct 2018.

[2] B. Tan, Q. Chen, K. Chetty, K. Woodbridge, W. Li, and R. Piechocki, "Exploiting wifi channel state information for residential healthcare informatics," *IEEE Communications Magazine*, vol. 56, no. 5, pp. 130–137, May 2018.

Predicted activity

Actual activity	Empty	Standing	Fall	Walk	Stand up	Lying
Empty	1	0	0	0	0	0
Standing	0	0.954	0	0	0	0.046
Fall	0	0	0.95	0.031	0.019	0
Walk	0	0	0.086	0.901	0.013	0
Stand up	0	0	0.05	0.05	0.9	0
Lying	0	0.117	0	0	0	0.883

(a) Proposed method with AF-DLN

Predicted activity

Actual activity	Empty	Standing	Fall	Walk	Stand up	Lying
Empty	1	0	0	0	0	0
Standing	0.125	0.848	0.002	0	0.006	0.019
Fall	0	0	0.855	0.066	0.072	0.007
Walk	0	0.025	0.035	0.852	0.088	0
Stand up	0	0.006	0.063	0.063	0.806	0.062
Lying	0	0.047	0.047	0.014	0.061	0.831

(b) Proposed method without AF-DLN

Fig. 7. Impact of AF-DLN on the recognition accuracy.

[3] J. Wang, X. Zhang, Q. Gao, X. Ma, X. Feng, and H. Wang, "Device-free simultaneous wireless localization and activity recognition with wavelet feature," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1659–1669, Feb 2017.

[4] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, "Non-invasive detection of moving and stationary human with wifi," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2329–2342, Nov 2015.

[5] 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 Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 306–319, Jan 2019.

[6] F. Wang, W. Gong, and J. Liu, "On spatial diversity in wifi-based human activity recognition: A deep learning based approach," *IEEE Internet of Things Journal*, pp. 1–1, 2018.

[7] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, Oct 2017.

[8] Z. Shi, J. A. Zhang, R. Xu, and G. Fang, "Human activity recognition using deep learning networks with enhanced channel state information," in *2018 IEEE Globecom Workshops (GC Wkshps)*, Dec 2018, pp. 1–6.

[9] H. Zou, Y. Zhou, J. Yang, H. Jiang, L. Xie, and C. J. Spanos, "Deepsense: Device-free human activity recognition via autoencoder long-term recurrent convolutional network," in *2018 IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–6.

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

[11] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "Rt-fall: A real-time and contactless fall detection system with commodity wifi devices," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 511–526, Feb 2017.

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

[13] S. W. Roberts, "Control chart tests based on geometric moving averages," *Technometrics*, vol. 1, no. 3, pp. 239–250, 1959.

[14] J.-Y. Chang, K.-Y. Lee, Y.-L. Wei, K. C.-J. Lin, and W. Hsu, "Location-independent wifi action recognition via vision-based methods," in *Proceedings of the 24th ACM International Conference on Multimedia*, ser. MM '16. New York, NY, USA: ACM, 2016, pp. 162–166. [Online]. Available: http://doi.acm.org/10.1145/2964284.2967203