Noname manuscript No.
(will be inserted by the editor)

Device-free Cross Location Activity Recognition via Semi-supervised Deep Learning

Rui Zhou · Ziyuan Gong · Kai Tang · Bao Zhou · Yu Cheng

Received: date / Accepted: date

Abstract Human activity recognition plays an important role in a variety of daily applications. There has been tremendous work on human activity recognition based on WiFi Channel State Information (CSI). Although achieving reasonable performance in certain cases, they are yet faced with a major challenge: location dependence. An activity recognition model trained at one location does not perform properly at other locations, because the human location also has significant influence on WiFi signal propagation. In this paper, we aim to solve the location dependence problem of CSI-based human activity recognition and propose a device-free Cross Location Activity Recognition (CLAR) method via semi-supervised deep learning. We regard the locations with labeled activity samples as the source domains and the locations with unlabeled activity samples as the target domains. By exploiting pseudo labeling and feature mapping, CLAR trains an activity recognition model working across the source and the target domains as well as the unseen domains which have no training samples. CLAR first extracts the trend component from the activity samples by Singular Spectrum Analysis (SSA), then annotates the unlabeled samples with the pseudo labels through a dual-score multi-classifier labeling model. The activity recognition model is trained using the labeled samples from the source domains and the pseudo-labeled samples from the target domains. Both the labeling and the recognition models are based on Bidirectional Long Short Term Memory (BLSTM). Evaluations in real-world environments demonstrate the effectiveness and generalization of the method CLAR, which performs well for both the source and the target domains, and generalizes well to the unseen domains.

Keywords Bidirectional Long Short Term Memory (BLSTM) · Channel State Information (CSI) · Cross location activity recognition · Pseudo labeling · Semi-supervised deep learning

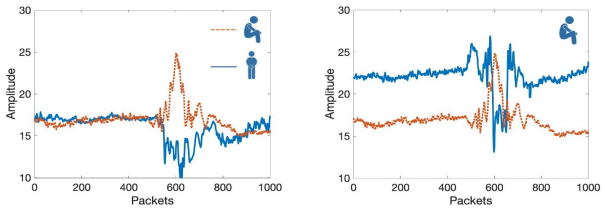
University of Electronic Science and Technology of China
E-mail: ruizhou@uestc.edu.cn

1 Introduction

Human activity recognition plays an important role in a variety of daily applications. Conventional approaches to human activity recognition are mainly based on vision [6,28] and wearables [8,20]. Vision-based solutions can achieve accurate recognition, but require Line of Sight (LOS) and luminous environments to work properly. In addition, cameras impose privacy concerns, thus are not appropriate in private areas such as bedrooms and bathrooms. Activity recognition based on wearables eliminates the limitations of LOS and does not impose privacy concerns. However, wearable-based solutions require the targets to wear dedicated devices, which is an extra burden on human bodies, causing inconvenience and reluctance. To overcome these shortcomings, human activity recognition has shed on wireless signals, such as radar [16,18], radio frequency [17,22], and WiFi [13,15]. In an indoor environment, the human body alters the propagation of WiFi signals, causing them to carry rich information of human, which can be utilized to infer human activities. Due to the ubiquity of wireless signals, WiFi-based device-free activity recognition avoids the inconvenience brought by wearables and eliminates privacy invasion from cameras. There has been tremendous work on human activity recognition based on WiFi Channel State Information (CSI), ranging from macro activities [1,4,26,29,35]

to micro activities [12, 24, 33, 36] and vital signs [11, 14, 19, 25, 31, 32]. However, most existing works require the target to perform activities at a specific location. If the target performs activities at different locations, the recognition performance will degrade dramatically. To deploy human activity recognition in real-world applications, a pair of transmitter and receiver should cover a relatively large area, e.g. a whole room. Collecting the training samples of all the activities at all the locations in a whole room is labor-intensive and impractical. It is greatly desired to build an activity recognition model working across all the locations with partially collected activity samples. This is a major challenge for CSI-based activity recognition on a large scale.

When a person performs an activity at a location, the WiFi signals are influenced not only by the activity but also by the person's location. Fig. 1(a) plots the CSI amplitude waves of one subcarrier during a person performing different activities at the same location, while Fig. 1(b) plots the CSI amplitude waves during a person performing the same activity at different locations. The variations of WiFi signals are the combined consequence of the activity and the location. To address this challenge, we propose a CSI-based device-free Cross Location Activity Recognition (CLAR) method. The aim is that the activity recognition model trained at some locations can be generalized to other locations. To mitigate the labor-intensive effort of collecting and labeling activity samples, CLAR makes use of both labeled samples and unlabeled samples (which are much easier to obtain), and further generalizes to the locations where no training samples are collected. The locations where the labeled samples are collected are regarded as the *source domains*, the locations where the unlabeled samples are collected are the *target domains*, and the locations where no training samples are collected are the *unseen domains*. To achieve the goal of activity recognition across all the locations/domains, we propose a dual-score multi-classifier labeling model to assign the pseudo labels to the unlabeled samples from the target domains. The recognition model is then trained using the labeled samples from the source domains and the pseudo-labeled samples from the target domains, to capture the cross domain features. Both the labeling model and the recognition model are based on Bidirectional Long Short Term Memory (BLSTM). All the activity samples are preprocessed by Singular Spectrum Analysis (SSA) to extract the trend components. The samples from the source and the target domains are feature-mapped before being classified by the recognition model. Evaluations in two real-world environments show that the proposed method CLAR is able to achieve the recognition accuracy (i.e. correct rate) of more than 0.86 across all the locations, more than 0.83 across the target domains and more than 0.74 across the unseen domains.



(a) Different activities same location (b) Same activity different locations

Fig. 1 CSI amplitude waves.

The main contributions of this paper are:

1. Solves the location dependence problem of device-free activity recognition based on WiFi CSI via semi-supervised learning using labeled and unlabeled samples, and generalizes well to unseen locations;
2. Proposes a dual-score multi-classifier labeling model based on BLSTM, which annotates the unlabeled samples with pseudo labels and solves the problem of insufficient labeled samples;
3. Captures cross domain activity features by training the BLSTM-based recognition model with the labeled samples from the source domains and the pseudo-labeled samples from the target domains, which are feature-mapped via autoencoders before entering the BLSTM classifier;
4. Extracts the trend components of the activity samples via SSA during data preprocessing, which proves more effective than other commonly used filters and enhances the recognition performance.

The remainder of the paper is organized as follows. Section 2 reviews the related work. Section 3 provides the overview of the proposed method CLAR. Section 4 elaborates on each part of CLAR. Evaluations are reported in Section 5. Section 6 concludes the paper.

2 Related work

There has been a great deal of work on WiFi-based activity recognition [13, 15]. The issue of domain independence is attracting more attention, as it is the prerequisite to large-scale application of activity recognition. However, the research on cross domain activity recognition is still in the exploration stage.

2.1 WiFi-based activity recognition

Researchers exploited WiFi signals to recognize macro activities, such as standing up, sitting down and falling.

CARM [29] built the CSI-speed model quantifying the correlation between CSI dynamics and human movement speeds, and the CSI-activity model quantifying the correlation between the movement speeds of different human body parts and specific human activities. Wi-Chase [1] utilized the variations in phase and magnitude of all available subcarriers to recognize daily activities. RT-Fall [26] exploited CSI phase and amplitude to segment and detect the falls in real-time. MAIS [4] identified multiple activities performed by multiple people in the same environment. Zhang et al. [35] proposed a diffraction-based sensing model to quantitatively determine the signal change wrt. a target's motions and linked signal variation patterns with human activities.

Researchers also exploited WiFi signals to recognize micro activities, such as gestures, mouth motions and emotions. WiHear [24] introduced mouth motion profile that leveraged partial multipath effects and wavelet packet transformation to recognize people talks. WiFinger [12] recognized finger-grained gestures to realize continuous text input in WiFi devices, based on the patterns in the time series of CSI values. EQ-Radio [36] transmitted radio frequency signals and analyzed the reflections off a person's body to recognize the emotional state. FingerDraw [33] sensed finger drawing trajectories based on the CSI-quotient model, which used the channel quotient between two antennas to cancel out the noise in CSI amplitude and the random offsets in CSI phase and quantified the correlation between CSI value dynamics and object displacement.

Researchers exploited WiFi signals to monitor vital signs as well. Liu et al. [14] used CSI to monitor respiration of a person under different sleeping positions by extracting rhythmic patterns associated with respiration and abrupt changes due to body movement. Khan et al. [11] used CSI to monitor the respiration rate of a patient during sleep. Wang et al. [25] leveraged the Fresnel model to develop the theory relating one's breathing depth, location and orientation to the detectability of respiration. WiHealth [19] could recognize and count breath and heartbeat with different postures. PhaseBeat [31] exploited CSI phase difference to monitor breath and heartbeat. TensorBeat [32] employed CSI phase difference to estimate breath rates of multiple persons.

2.2 Cross domain activity recognition

DFLAGR [27] designed a sparse autoencoder to learn discriminative features from RSS and used the softmax regression algorithm to realize location, activity and gesture recognition. In the experiments, DFLAGR deployed 8 Zigbee nodes forming 56 wireless links and collected training data at each location. Gao et al. [5] transformed CSI amplitude and phase into images, extracted optimized deep features, and estimated the location and activity using the softmax regression algorithm. The method also collected training data at each location. Chen et al. [3] proposed an attention based BLSTM for activity recognition with CSI, which leveraged an attention mechanism to assign different weights to all the learned features. In the experiments involving 7 persons in 2 rooms, the method collected training data for each person and each room. The above methods belonged to supervised learning, collecting training data for each room, location or person. Different from them, our work is based on semi-supervised learning with partially labeled samples and can generalize to new locations with no training samples.

Chang et al. [2] transformed CSI into images for activity recognition. They proposed a location-dependency removal method based on Singular Value Decomposition (SVD) to eliminate the background CSI and extract the channel information of signals reflected by human bodies. CARM [30] achieved activity recognition by developing the CSI-speed model and the CSI-activity model. To address human diversity, they used the Hidden Markov Model (HMM) to describe the speed features of human activities. To address environment diversity, they performed data fusion on multiple links. These works aimed for room diversity or human diversity, thus their experiments were conducted in a small area in the rooms. In comparison, our work focuses on location diversity and the deployment scale can cover a whole room.

EI [10] realized environment independent activity recognition by exploiting adversarial networks to remove environment and human specific information and learn transferable features of activities. Its goal was room diversity and human diversity. The testing area in the room was very small and all the volunteers were involved in the training. CsiGAN [34] enabled activity recognition adaptive to user changes based on semi-supervised Generative Adversarial Networks (GAN) to meet the scenarios that unlabeled data from left out users were very limited. Its goal was user change. We compare our work with the two methods in the evaluations and outperform them.

WiSDAR [23] achieved spatial diversity-aware activity recognition by extending the multiple antennas of WiFi devices to construct multiple separated antenna pairs. It employed Convolutional Neural Network (CNN) and LSTM to integrate the temporal and spatial features. The method was trained in one environment and tested in three others. All the testing environments were free space with the same antenna layout.

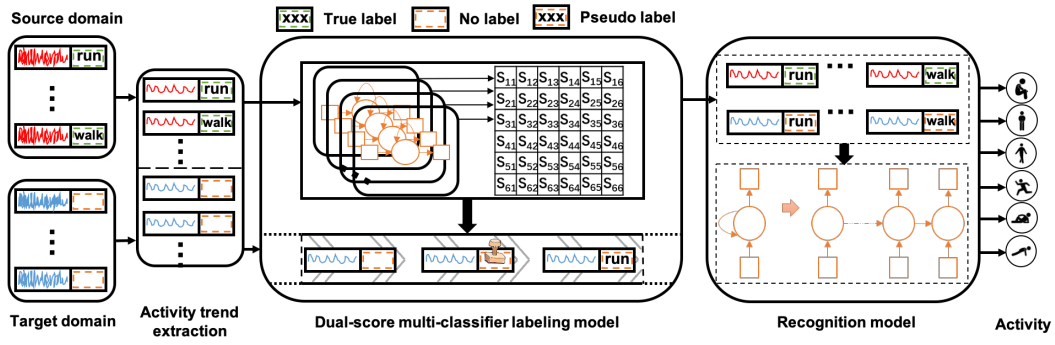


Fig. 2 Overview of the proposed method CLAR.

Sheng et al. [21] realized cross-scene activity recognition by integrating spatial features learned from CNN into BLSTM. To cope with environmental changes, the original model was fine-tuned in the new environment. But to achieve acceptable accuracy, fine-tuning needed a relatively large set of labeled data in the new environment. Both works were for room diversity and human diversity, while our work focuses on location diversity in a relatively large space.

3 Overview

The proposed method CLAR is composed of four components: data collection, activity trend extraction, labeling model and recognition model, as illustrated in Fig. 2. When a person performs activities, the CSI amplitudes are extracted to constitute the activity samples. The locations are categorized to source domains, target domains and unseen domains. The samples from the source domains are labeled with activities, while the samples from the target domains are unlabeled. The labeled samples are used to train the labeling model, which annotates the unlabeled samples with pseudo labels. The labeled and the pseudo-labeled samples are used together to train the recognition model, which is to recognize the activities across all the domains, including the unseen domains.

Data collection. During an activity, the CSI amplitude series of all the subcarriers are extracted from the packets sent from the WiFi transmitter to the receiver, which are used as an activity sample. Assume $(\bar{\boldsymbol{r}}, a)$ represents a labeled sample, where $\bar{\boldsymbol{r}} = (\bar{\boldsymbol{r}}_{\boldsymbol{1}}, \bar{\boldsymbol{r}}_{\boldsymbol{2}}, \cdots, \bar{\boldsymbol{r}}_{\boldsymbol{t}})$ represents the series of CSI amplitudes, t represents the time length, and a represents the activity label. $\bar{\boldsymbol{r}}_{\boldsymbol{i}} = (\bar{r}_{i1}, \bar{r}_{i2}, \cdots, \bar{r}_{il})$ represents the CSI amplitudes at time i, in which $\bar{r}_{ij}$ represents the amplitude of subcarrier j and l is the number of subcarriers.

Activity trend extraction. When a person performs activities, the CSI amplitude waves will fluctuate drastically. Different activities incur different patterns of CSI amplitude waves, the trends of which are closely related to the corresponding activities. We apply SSA on the CSI amplitude series to extract the trend components, meanwhile remove the noise and the periodic components. The samples after SSA, denoted as $\boldsymbol{r}$, are used in the labeling model and the recognition model.

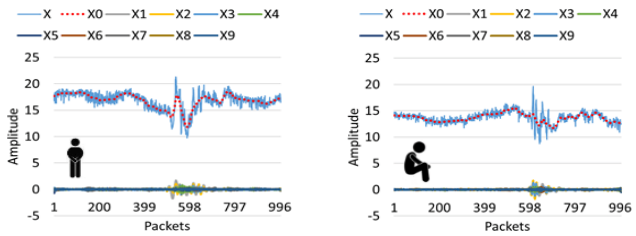
Labeling model. In supervised learning, to train an activity recognition model working across all the locations, we require the labeled samples of all the activities at all the locations. This is impractical due to the huge amount of effort. As unlabeled samples are much easier to obtain, we propose a dual-score multi-classifier labeling model to annotate the unlabeled samples with pseudo labels. The labeling model consists of multiple BLSTM classifiers and a dual-score table.

Recognition model. The activity recognition model is a classification model based on BLSTM, which is trained using the labeled samples from the source domains and the pseudo-labeled samples from the target domains. To minimize the domain difference and capture the cross domain activity features, feature mapping is conducted via autoencoders before the BLSTM classifier. Thus, the activity recognition model performs well for both the source and the target domains as well as the unseen domains.

4 Methodology

4.1 Activity trend extraction

When a person performs activities, the CSI amplitude waves will fluctuate dramatically. It is a combination of three kinds of components: the trend of the wave, the periodic components, and the noise. The trend is a slowly varying component, mainly caused by the activity. The periodic components are mainly composed of multipath oscillating wireless signals. We apply SSA on the CSI amplitude waves to extract the trend com-



(a) Standing up (b) Squatting down

Fig. 3 Components of activity samples. $\boldsymbol{X}$ represents raw activity sample, $\boldsymbol{X}_0$ represents trend component, $\boldsymbol{X}_1 - \boldsymbol{X}_9$ represent the other components.

ponent and remove the noise and the periodic components. SSA [7] is a time series analysis method that can decompose a time series into a set of summable components that are interpreted as trend, periodicity and noise. It can separate periodicities that occur on different time scales. The original time series can be recovered by summing together all the components. Through SSA, the raw activity samples are decomposed and the trend components can be extracted.

To extract the trend component of an activity sample $\bar{\boldsymbol{r}}$, we first construct a trajectory matrix $\boldsymbol{X}$ from the time series of CSI amplitudes $\bar{\boldsymbol{r}} = (\bar{\boldsymbol{r}}_1, \bar{\boldsymbol{r}}_2, \cdots, \bar{\boldsymbol{r}}_t)$:

$$\boldsymbol{X} = \begin{pmatrix} \bar{\boldsymbol{r}}_1 & \cdots & \bar{\boldsymbol{r}}_{t-w+1} \\ \vdots & \ddots & \vdots \\ \bar{\boldsymbol{r}}_w & \cdots & \bar{\boldsymbol{r}}_t \end{pmatrix} \tag{1}$$

where $\bar{\boldsymbol{r}}_i$ represents the CSI amplitudes at time i, t represents the time length of the activity, and w represents the length of the extraction window. We then decompose the trajectory matrix $\boldsymbol{X}$ with Singular Value Decomposition (SVD), and obtain w eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_w \geq 0$ with

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{2}$$

where $\boldsymbol{U}$ is a $w \times w$ unitary matrix containing the orthonormal set of left singular vectors of $\boldsymbol{X}$ as columns. $\boldsymbol{\Sigma}$ is a $w \times (t-w+1)$ rectangular diagonal matrix containing w singular values of $\boldsymbol{X}$ in descending order. $\boldsymbol{V}$ is a $(t-w+1) \times (t-w+1)$ unitary matrix containing the orthonormal set of right singular vectors of $\boldsymbol{X}$ as columns. The SVD of the trajectory matrix $\boldsymbol{X}$ can then be expressed as:

$$\boldsymbol{X} = \sum_{i=0}^{d-1} \boldsymbol{X}_i \tag{3}$$

where $d = max(i, \lambda_i)$ is the rank of the trajectory matrix $\boldsymbol{X}$, $\boldsymbol{X}_i$ is the i-th elementary matrix of $\boldsymbol{X}$.

Two examples of activity samples after decomposition by SSA are shown in Fig. 3. Fig. 3(a) shows activity trend extraction of standing up and Fig. 3(b) shows squatting down, where $\boldsymbol{X}$ represents the raw activity sample, $\boldsymbol{X}_0$ represents the trend component, and $\boldsymbol{X}_1 - \boldsymbol{X}_9$ represent the noise and the periodic components. It can be seen from the figures that the trend components are dominant in the activity samples. SSA extracts the trend but does not reduce the dimension.

There are other commonly used preprocessing methods, such as Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA). DWT is a time-frequency analysis tool and can remove the high frequency noise by decomposition and reconstruction, but it is not for component separation and trend extraction. PCA is basically a dimension reduction tool and can be used for dimension reduction and principal component extraction, but it is not for frequency analysis in time series. We compare SSA with DWT and PCA on the same amplitude wave in the evaluations, which shows that SSA extracts the trend more accurately.

4.2 Pseudo labeling

The activity samples in the target domains are unlabeled. To make the recognition model learn the features across different domains and enlarge the training set, we propose to annotate the samples in the target domains with pseudo labels. For such a purpose, we construct a labeling model to generate pseudo labels for the samples in the target domains. The labeling model is a dual-score multi-classifier model, consisting of multiple classifiers and a dual-score table to rate the performance of the classifiers, as illustrated in Fig. 4. The activity samples are time series data, therefore, we adopt BLSTM as the base classifier in the labeling model. The dual-score table contains the classifier scores and the category scores. A classifier score reflects the overall classification ability of a classifier, and a category score reflects the classification ability of a classifier on an activity category. We combine the classifier scores and the category scores together as the dual-scores to evaluate the performance of the classifiers.

4.2.1 Training of labeling model

Training of the labeling model is to train the multiple BLSTM classifiers and to construct the dual-score table. The labeled activity samples in the source domains are used to train the labeling model. The activity samples first go through SSA to extract the trend components, which are input to each BLSTM classifier in the labeling model to extract the time series features and be classified to the activities. Suppose m is the number of the classifiers in the labeling model, we use $BLSTM_i$

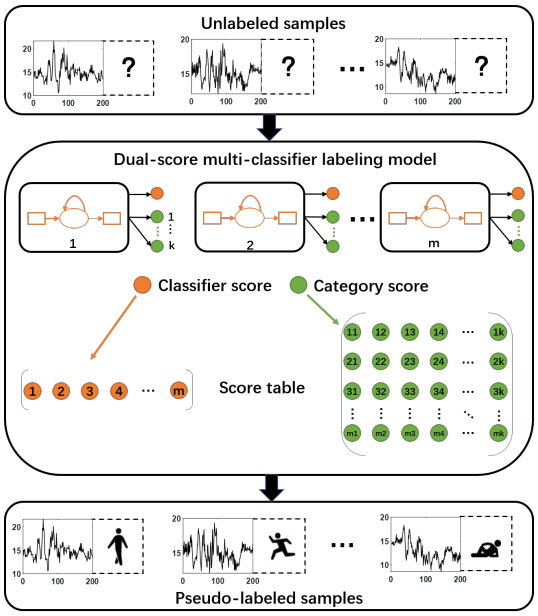


Fig. 4 The dual-score multi-classifier labeling model.

$(i = 1, 2, \cdots, m)$ to represent classifier i. The features extracted by $BLSTM_i$ can be expressed as:

$$\boldsymbol{F}_i = BLSTM_i(\boldsymbol{r}; \boldsymbol{\Theta}_i) \tag{4}$$

where $\boldsymbol{r}$ denotes an activity sample after SSA, $\boldsymbol{\Theta}_i$ denotes the parameter set of $BLSTM_i$. The last layer of each BLSTM classifier uses *Softmax* to classify the samples. Assume k is the number of activity categories, we use $\boldsymbol{P}_i = (p_{i1}, p_{i2}, \cdots, p_{ik})$ to denote the classification result of $BLSTM_i$, which is calculated as:

$$\boldsymbol{P}_i = Softmax(\boldsymbol{W}_{\boldsymbol{F}_i} \times \boldsymbol{F}_i + \boldsymbol{b}_{\boldsymbol{F}_i}) \tag{5}$$

where $\boldsymbol{W}_{\boldsymbol{F}_i}$ and $\boldsymbol{b}_{\boldsymbol{F}_i}$ are the weights and the biases of $BLSTM_i$. The activity with the highest probability p_{ij} is regarded as the classified activity.

Training of the labeling model needs to construct the dual-score table. For m classifiers and k activity categories, let s_i^r $(i = 1, 2, \cdots, m)$ denote the classifier score of $BLSTM_i$, let s_{ij}^c $(i = 1, 2, \cdots, m, j = 1, 2, \cdots, k)$ denote the category score of $BLSTM_i$ on activity category j. They are calculated as:

$$\begin{aligned} s_i^r &= \hat{n}_i^r / n_i^r \\ s_{ij}^c &= \hat{n}_{ij}^c / n_{ij}^c \end{aligned} \tag{6}$$

where n_i^r represents the number of all predicted activity samples by classifier i, $\hat{n}_i^r$ represents the number of correctly predicted activity samples by classifier i, n_{ij}^c represents the number of all predicted activity samples of category j by classifier i, and $\hat{n}_{ij}^c$ represents the number of correctly predicted activity samples of category j by classifier i.

The dual-score table $\boldsymbol{S} = (\boldsymbol{S}^r, \boldsymbol{S}^c)$, composed of the classifier score table $\boldsymbol{S}^r$ and the category score table $\boldsymbol{S}^c$, is constructed as:

$$\boldsymbol{S}^r = \begin{pmatrix} s_1^r & \cdots & s_{m-1}^r & s_m^r \end{pmatrix}$$

$$\boldsymbol{S}^c = \begin{pmatrix} s_{11}^c & \cdots & s_{1(k-1)}^c & s_{1k}^c \\ \vdots & \ddots & \vdots & \vdots \\ s_{(m-1)1}^c & \cdots & s_{(m-1)(k-1)}^c & s_{(m-1)k}^c \\ s_{m1}^c & \cdots & s_{m(k-1)}^c & s_{mk}^c \end{pmatrix} \tag{7}$$

4.2.2 Generation of pseudo labels

The activity samples in the target domains are unlabeled. We assign the pseudo labels to the unlabeled activity samples, so that the recognition model can capture the cross domain features from both the source and the target domains and enlarge the training set. Suppose $\boldsymbol{r}^u$ is an unlabeled activity sample. We input $\boldsymbol{r}^u$ to the labeling model. Each BLSTM classifier in the labeling model predicts the probability that the unlabeled sample $\boldsymbol{r}^u$ belongs to each activity category. Suppose p_{ij} $(i = 1, 2, \cdots, m,\ j = 1, 2, \cdots, k)$ represents the probability that $BLSTM_i$ classifies $\boldsymbol{r}^u$ into activity category j, suppose s_{ij} $(i = 1, 2, \cdots, m,\ j = 1, 2, \cdots, k)$ represents the score that $BLSTM_i$ classifies $\boldsymbol{r}^u$ into activity category j, s_{ij} is calculated as:

$$s_{ij} = p_{ij} s_{ij}^c s_i^r \tag{8}$$

where s_{ij}^c and s_i^r can be obtained from the dual-core table. s_i^r denotes the classifier score of $BLSTM_i$, reflecting the overall classification ability of $BLSTM_i$. s_{ij}^c denotes the category score of $BLSTM_i$ on activity category j, reflecting the classification ability of $BLSTM_i$ on activity category j. The total score s_j that $\boldsymbol{r}^u$ is classified into activity category j is calculated as the sum of the scores of all the classifiers in the labeling model:

$$s_j = \sum_{i=1}^{m} s_{ij} \tag{9}$$

The pseudo label of $\boldsymbol{r}^u$ is finally generated as the activity category having the maximal total score:

$$pLabel = \arg\max_j s_j \tag{10}$$

4.3 Activity recognition

The activity recognition model is based on BLSTM, which is trained using the labeled samples from the source domains and the pseudo-labeled samples from the target domains. To minimize the domain divergence and capture the cross domain features, feature mapping is conducted before the BLSTM classifier, as illustrated in Fig. 5.

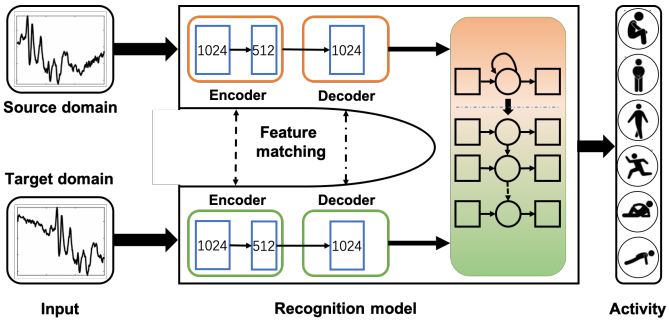
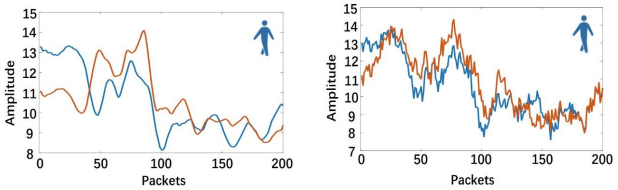


Fig. 5 Activity recognition model.



(a) Before mapping (b) After mapping

Fig. 6 Feature mapping of activity samples.

4.3.1 Feature mapping

Feature mapping is based on Stacked Autoencoders (SAE). The labeled samples of the source domains and the pseudo-labeled samples of the target domains are input to their encoders and restored by their decoders respectively. As the reason of recognition degradation is that the probability distributions of activity samples between the source and the target domains mismatch, we convert the activity samples in the target domains to the source domains, so that the probability distributions of the converted activity samples of the target domains are similar to the source domains. To this end, we build SAE to realize the conversion. The input data to the autoencoders are the activity samples in the source and the target domains, and the output data are the converted activity samples. The goal of SAE is to minimize the divergence of the probability distributions of the converted activity samples between the source and the target domains. We utilize Euclidean distance as the measurement in the training of the feature mapping model.

Assume $\boldsymbol{r}^s$ represents a sample of the source domains, $\boldsymbol{r}^s_{ed}$ represents the sample after encoding-decoding, defined as:

$$\boldsymbol{r}^s_{ed} = Decoder^s(Encoder^s(\boldsymbol{r}^s; \boldsymbol{\Theta}^s_{en}); \boldsymbol{\Theta}^s_{de}) \tag{11}$$

where $\boldsymbol{\Theta}^s_{en}$ is the parameter set of $Encoder^s$ and $\boldsymbol{\Theta}^s_{de}$ is the parameter set of $Decoder^s$. Assume D^s represents the training set of the source domains and $|D^s|$ represents the number of training samples, the loss of the source domain autoencoder is defined as:

$$L^s = \frac{1}{|D^s|}\sum_{i=1}^{|D^s|}((\boldsymbol{r}^s_{ed})_i - \boldsymbol{r}^s_i)^2 \tag{12}$$

Assume $\boldsymbol{r}^t$ represents a sample of the target domains, $\boldsymbol{r}^t_{ed}$ represents the sample after encoding-decoding, defined as:

$$\boldsymbol{r}^t_{ed} = Decoder^t(Encoder^t(\boldsymbol{r}^t; \boldsymbol{\Theta}^t_{en}); \boldsymbol{\Theta}^t_{de}) \tag{13}$$

where $\boldsymbol{\Theta}^t_{en}$ is the parameter set of $Encoder^t$ with $\boldsymbol{\Theta}^t_{en} = \boldsymbol{\Theta}^s_{en}$, and $\boldsymbol{\Theta}^t_{de}$ is the parameter set of $Decoder^t$. Assume D^t represents the pseudo-labeled training set of the target domains and $|D^t|$ represents the number of pseudo-labeled training samples, the loss of the target domain autoencoder is defined as:

$$L^t = \frac{1}{|D^t|}\sum_{i=1}^{|D^t|}((\boldsymbol{r}^t_{ed})_i - (\boldsymbol{r}^s_{ed})_i)^2 \tag{14}$$

where $(\boldsymbol{r}^s_{ed})_i$ represents the sample in the source domains with the same activity label as $(\boldsymbol{r}^t_{ed})_i$, to minimize the difference between the source and the target domains. The total loss of feature mapping via SAE is then defined as:

$$L_{ed} = L^s + \alpha L^t \tag{15}$$

where α is the balance factor.

Fig. 6 shows two samples of the same activity in two different domains (locations). Before feature mapping, the activity samples in different domains are quite different, while they can be unified by the SAE.

4.3.2 Activity classification

The feature-mapped activity samples are input to the BLSTM classifier in the recognition model for activity recognition. The prediction $\hat{a}$ of the BLSTM classifier is defined as:

$$\hat{a} = BLSTM(\boldsymbol{r}_{ed}; \boldsymbol{\Theta}) \tag{16}$$

where $\boldsymbol{r}_{ed}$ represents a feature-mapped activity sample from the source or the target domain, and $\boldsymbol{\Theta}$ is the parameter set of the BLSTM classifier. As activity recognition is a multi-class classification problem, we use softmax and cross entropy to define the loss of the BLSTM classifier in the recognition model. Assume D represents the set of feature-mapped training samples from both the source and the target domains, $|D|$ represents the number of samples, $\boldsymbol{a}_i = (a_{i1}, a_{i2}, \cdots, a_{iK})$ represents the ground truth of a sample in the form of one-hot encoding, $\hat{\boldsymbol{a}}_i = (\hat{a}_{i1}, \hat{a}_{i2}, \cdots, \hat{a}_{iK})$ represents

(a) Laboratory (b) Laboratory setup (c) Lobby (d) Lobby setup

Fig. 7 The testing environments.

the prediction of the sample in the form of probabilities over classes, K is the number of classes, the loss is then defined as:

$$L_a = -\frac{1}{|D|}\sum_{i=1}^{|D|}\sum_{k=1}^{K} \boldsymbol{a}_{ik} log(\hat{\boldsymbol{a}}_{ik}) \quad (17)$$

5 Evaluations

5.1 Experimental setup

We conducted activity recognition experiments in two different environments in our office building: the laboratory and the lobby. In each environment, two laptops equipped with Intel wireless link (IWL) 5300 were employed as the transmitter (TX) and the receiver (RX), each with three antennas. The firmware and driver of IWL5300 were modified to export CSI from each packet to the upper layers [9]. The transmitter sent packets to the receiver at a rate of 100Hz. The experimental environments and their setup are shown in Fig. 7. The laboratory and its setup are shown in Fig. 7(a)-7(b). It has the size of 6m×3m. The transmitter and the receiver were placed 3m apart. Six activities were performed at 8 locations (marked as dark circles in Fig. 7(b)), which were squatting down, standing up, walking, running, falling and climbing. Each activity was performed 10 times at each location. Thus 60 activity samples were collected for each location and 480 activity samples were collected totally in the laboratory. The lobby and its setup are shown in Fig. 7(c)-7(d). It has the size of 5m×5m. The transmitter and the receiver were placed 7m apart. Three activities were performed at 20 locations (marked as dark circles in Fig. 7(d)), which were squatting down, standing up and jumping. Each activity was performed 10 times at each location. Thus 30 activity samples were collected for each location and 600 activity samples were collected totally in the lobby. Each location is regarded as a domain. We category the locations into source domains, target domains, and unseen domains. The training samples in the source domains have labels, the training samples in the target domains are unlabeled, there are no training samples in the unseen domains. Half of the activity samples in the source domains were used as training samples and the other half as validation samples. Half of the activity samples in the target domains were used as training samples and the other half as testing samples. All the activity samples in the unseen domains were used as testing samples.

5.2 Model training

We first trained the labeling model using the training samples in the source domains, then annotated the training samples in the target domains with the pseudo labels generated by the labeling model. The recognition model was trained using the labeled training samples from the source domains and the pseudo-labeled training samples from the target domains.

The training of the labeling model was to train the BLSTM-based multi-classifiers and construct the dual-score table. The multi-classifiers were trained using the training samples from the source domains and validated using the validation samples from the source domains. We used bootstrap sampling to select the samples for each classifier. Each time a sample was selected randomly from the training set and then returned. So repeated multiple times to select multiple training samples for the classifier. The unselected samples were used as the validation samples for the classifier. In this way the m classifiers were trained and validated with different samples. The dual-score table was constructed based on the validation results. The learning rate was 10^{-4}, with Adam as the optimizer. The constructed labeling model had 6 BLSTM classifiers, each having 1 hidden layer with 128 units.

Through the labeling model, the unlabeled training samples of the target domains were assigned the pseudo labels. The recognition model was trained using the labeled training samples from the source domains and the pseudo-labeled training samples from the target domains. The recognition model contained a feature mapping model and a BLSTM classifier. The feature

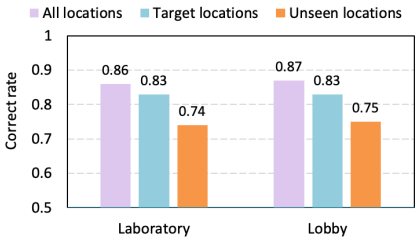
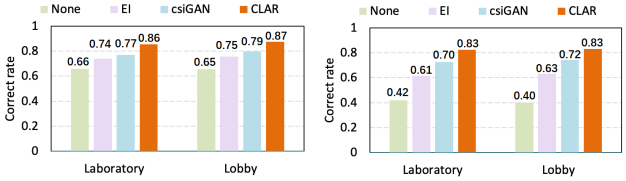


Fig. 8 Performance of cross location activity recognition.



(a) All locations (b) Target locations

Fig. 9 Comparison of cross location activity recognition.

mapping model had 2 pairs of encoder-decoders, one pair for source domains and the other for target domains. Each encoder had 2 hidden layers with (1024, 512) neurons and each decoder had 1 hidden layer with 1024 neurons. To train the BLSTM classifier, the samples from the source domains and the target domains were input to the classifier alternately. The learning rate was set 10^{-4}, with Adam as the optimizer. The BLSTM classifier had 1 hidden layer with 128 units.

5.3 Performance evaluation

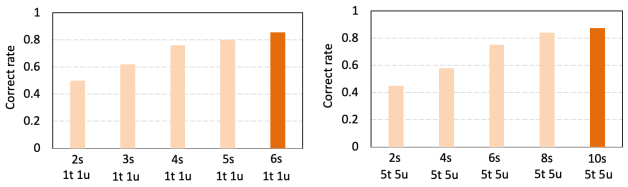
5.3.1 Performance of activity recognition

We conducted experiments to evaluate the performance of CLAR on cross location activity recognition. In the laboratory, we performed 6 activities at 8 locations. We chose 6 locations randomly as the source domains, 1 location as the target domain and 1 location as the unseen domain. As each activity was performed 10 times at each location, each location had 60 activity samples and the laboratory had 480 activity samples, in which the source domains had 360 samples, the target domain had 60 samples and the unseen domain had 60 samples. In the lobby, we performed 3 activities at 20 locations. We chose 10 locations randomly as the source domains, 5 locations as the target domains and 5 locations as the unseen domains. As each activity was performed 10 times at each location, each location had 30 activity samples and the lobby had 600 activity samples, in which the source domains had 300 samples, the target domains had 150 samples and the unseen domains had 150 samples. The experimental results are shown in Fig. 8. In the laboratory, CLAR achieved the recognition correct rate of 0.86 across all the locations (domains), 0.83 for the target locations and 0.74 for the unseen locations. In the lobby, CLAR achieved the recognition correct rate of 0.87 across all the locations (domains), 0.83 for the target locations and 0.75 for the unseen locations. Highly accurate activity recognition can be achieved with a large labeled dataset across all the domains, but this requires huge and intimidating effort. With a small and partially labeled dataset, CLAR achieved the recognition correct rate of more than 0.86 across all the locations, including unlabeled locations and unseen locations. For the unlabeled locations alone, CLAR achieved the correct rate of 0.83. For the unseen locations without any training samples, CLAR achieved the correct rate of more than 0.74.

5.3.2 Comparison with existing work

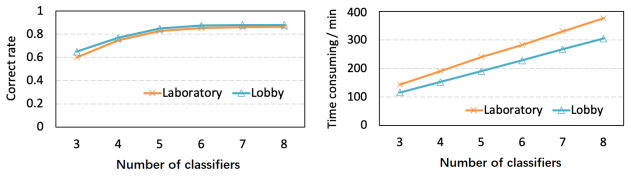
To demonstrate the effect of CLAR, we compared it with the state of the art on cross domain activity recognition: EI [10] and CsiGAN [34]. EI [10] was an environment independent activity recognition system, which could remove environment and subject specific information and learn transferable features of activities by exploiting adversarial networks. CsiGAN [34] was an activity recognition method based on semi-supervised GAN, which enabled activity recognition adaptive to user changes. We also compared them with the method of applying the source recognition model directly to the target domains, termed as None. As the methods aimed for cross domain activity recognition, we compared them on all the locations and on the target locations respectively, in order to compare their recognition and generalization abilities across different domains. EI and CsiGAN did not consider unseen domains, hence we did not compare them on the unseen locations.

Fig. 9 shows the comparisons of CLAR, EI, CsiGAN and None using the same datasets in the laboratory and the lobby. Fig. 9(a) shows the recognition correct rate of them across all the locations (domains), Fig. 9(b) shows the recognition correct rate of them on the target locations (domains). CLAR achieved the best performance of activity recognition across all the locations and on the target locations in both the laboratory and the lobby. Compared with None, CLAR improved the correct rate by more than 30% on all the locations, and by more than 98% on the target locations. The results show that the recognition model trained with only the labeled samples from the source domains was unable to achieve good performance on the target domains. Utilizing the unlabeled samples from



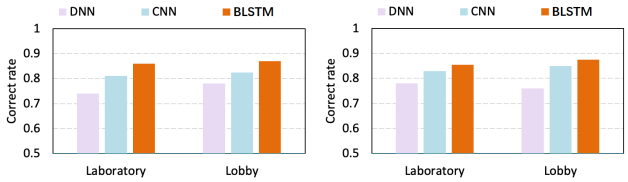
(a) Laboratory (b) Lobby

Fig. 10 Impact of number of source domains.



(a) Recognition correct rate (b) Training time

Fig. 11 Impact of number of classifiers in labeling model.



(a) Labeling model (b) Recognition model

Fig. 12 Classifier types in labeling model and recognition model.

the target domains helped learn better classifiers and enhanced the generalization ability on the target domains. CLAR, EI and CsiGAN were able to extract the common features shared by the source and the target domains, thus achieved better performance than None. Among them, CLAR achieved the best performance by exploiting pseudo labeling and feature mapping.

5.4 Parameter study

5.4.1 Number of source domains

We trained CLAR with different numbers of source domains, to observe its impact on the performance. In the laboratory, the number of source domains increased from 2 to 6, the number of target domains was 1 and the number of unseen domains was 1. In the lobby, the number of source domains increased from 2 to 10 by step 2, the number of target domains was 5 and the number of unseen domains was 5. As shown in Fig. 10, with the number of source domains increasing, the correct rate of cross domain activity recognition improved. With 6 source domains in the laboratory, the recognition correct rate reached 0.86. With 10 source domains in the lobby, the correct rate reached 0.87. When more labeled samples from more different domains were available, the classifier could capture the cross domain features more accurately. Meanwhile, the cost of data collection and model training increased.

5.4.2 Number of classifiers

The performance of the labeling model is critical to the performance of CLAR. To investigate the impact of the number of classifiers on cross domain activity recognition, we set the number of classifiers in the labeling model from 3 to 8, and compared the recognition correct rate and the training time of CLAR. Fig. 11 shows the results. In the laboratory, as the number of classifiers in the labeling model increased from 3 to 8, the recognition correct rate increased from 0.6 to 0.86, meanwhile the training time increased from 143 minutes to 376 minutes. In the lobby, as the number of classifiers in the labeling model increased from 3 to 8, the recognition correct rate increased from 0.65 to 0.88, meanwhile the training time increased from 115 minutes to 305 minutes. The classifier number 6 was a turning point, therefore, we used 6 classifiers in the labeling model.

5.4.3 Type of classifiers

The performance of classification depends on the type of classifiers. Hence the performance of CLAR depends on the types of classifiers in the labeling model and the recognition model. We employed BLSTM, Deep Neural Networks (DNN) and CNN as the classifiers respectively and compared their performance on activity recognition. Fig. 12(a) shows the comparison of using BLSTM, DNN and CNN as the classifiers in the labeling model and using BLSTM as the recognition model. Fig. 12(b) shows the comparison of using BLSTM, DNN and CNN as the recognition model and using BLSTM in the labeling model. As considering the time series features of activity samples, BLSTM outperformed DNN and CNN wrt. activity recognition in both cases.

5.5 Ablation study

5.5.1 Effect of SSA

As data preprocessing, we used SSA to extract the trend component of the raw activity samples. To prove its effectiveness, we compared SSA with other commonly used filtering methods, including moving average (MA),

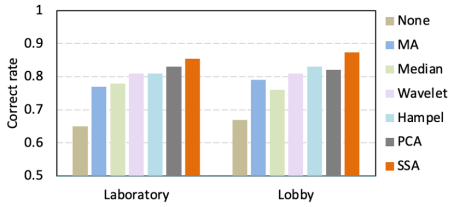


Fig. 13 Effect of SSA.

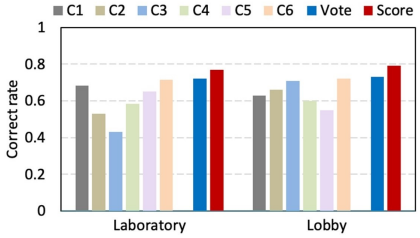


Fig. 14 Effect of pseudo-labeling.

median filter, DWT, hampel filter and PCA. Fig. 13 shows the comparison. SSA achieved the best performance, as SSA removed not only the noise but also the periodic components and extracted the trend which reflected the activity regardless of the background wireless information. All the filters outperformed without filtering.

5.5.2 Effect of pseudo-labeling

We evaluated the effect of pseudo-labeling in CLAR and compared it with single-classifier labeling and max-voting multi-classifier labeling. In the laboratory, we chose 6 locations randomly as the source domains and 1 location as the target domain to label. In the lobby, we chose 10 locations randomly as the source domains and 5 locations as the target domains to label. The labeling results are shown in Fig. 14. In the laboratory, CLAR (Score) achieved the labeling correct rate of 0.77, better than max-voting (Vote) of 0.72 and each single classifier (C1–C6). In the lobby, CLAR achieved the labeling correct rate of 0.79, better than max-voting of 0.73 and each single classifier. The dual-score pseudo-labeling method in CLAR made use of ensemble learning to predict the pseudo labels, thus outperformed a single classifier. The dual-score method considered not only the classification results of the samples but also the generic classification ability and the classification ability on a specific activity of the classifiers, thus achieved more reliable pseudo labels.

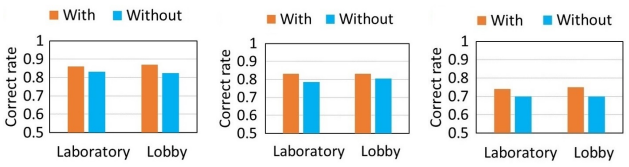


Fig. 15 Effect of feature mapping.

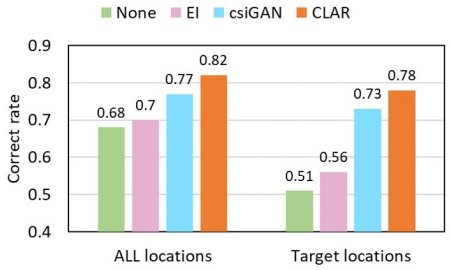


Fig. 16 Performance on public dataset.

5.5.3 Effect of feature mapping

We also evaluated the effect of feature mapping in CLAR, which is shown in Fig. 15. In both the laboratory and the lobby, CLAR with feature mapping outperformed without feature mapping, for all the locations (Fig. 15(a)), the target locations (Fig. 15(b)) and the unseen locations (Fig. 15(c)). Feature mapping minimized the divergence of the probability distributions between the source and the target domains, hence achieved higher recognition accuracy than without feature mapping.

5.6 Evaluation on public dataset

We also evaluated our method CLAR on the public dataset from [37], which included WiFi gesture samples at different locations. We regarded gestures as micro activities and compared CLAR with EI, CsiGAN and None on this dataset. The setup had one transmitter with one antenna and six receivers with three antennas each. The volunteer performed 6 gestures at 5 locations in a 2m×2m sensing area. The gestures were pushing and pulling, sweeping, clapping, sliding, drawing circle and drawing zigzag. Each gesture was performed 20 times per location. We randomly chose 4 locations as the source domains and 1 as the target domain. The results and the comparison are illustrated in Fig. 16. CLAR achieved better performance than EI and CsiGAN. Compared with None, CLAR improved the correct rate by a large margin.

6 Conclusions

Human activity recognition plays an important role in many daily applications such as health management and security protection. Yet for real-world applications, location independent activity recognition is still a challenge. By exploiting semi-supervised deep learning, we propose a device-free cross location activity recognition method CLAR based on CSI using one pair of WiFi transmitter and receiver, which achieves device-free location independent activity recognition on a room scale. CLAR first applies SSA to extract the trend components of the activity samples, which effectively reduces the influence of uncorrelated factors. To learn the across location features and enlarge the training set, we propose a labeling model based on ensemble learning with multiple BLSTM classifiers and construct a dual-score table as the criteria to annotate the unlabeled samples in the target domains with pseudo labels. The BLSTM-based activity recognition model is trained using the labeled samples from the source domains and the pseudo-labeled samples from the target domains, which go through feature mapping via autoencoders before entering the BLSTM classifier. Evaluations in two real-world environments show that the proposed method CLAR can recognize activities accurately across all the locations and can generalize well to the unseen locations in the environment.

The current method still has difficulty working across different sites, such as different rooms or buildings. Different sites usually have different layouts and WiFi setup, causing significantly different feature spaces and probability distributions, thus making pesudo-labeling and feature mapping difficult. Cross site activity recognition is our next step work.

7 Declarations

The manuscript is our original work. It it not published or under review elsewhere. No conflict of interest exists in the submission. The manuscript is approved by all the co-authors for publication in this journal.

References

1. Arshad, S., Feng, C., Liu, Y., Hu, Y., Yu, R., Zhou, S., Li, H.: Wi-Chase: A WiFi based human activity recognition system for sensorless environments. In: WoWMoM'2017, pp. 1–6 (2017)
2. Chang, J.Y., Lee, K.Y., Wei, Y.L., Lin, K.C.J., Hsu, W.: Location-independent WiFi action recognition via vision-based methods. In: MM'2016, pp. 162–166. ACM (2016)
3. Chen, Z., Zhang, L., Jiang, C., Cao, Z., Cui, W.: WiFi CSI based passive human activity recognition using attention based BLSTM. IEEE Transactions on Mobile Computing **18**(11), 2714–2724 (2019)
4. Feng, C., Arshad, S., Liu, Y.: MAIS: Multiple activity identification system using channel state information of WiFi signals. In: International Conference on Wireless Algorithms, Systems, and Applications, pp. 419–432 (2017)
5. Gao, Q., Wang, J., Ma, X., Feng, X., Wang, H.: CSI-based device-free wireless localization and activity recognition using radio image features. IEEE Transactions on Vehicular Technology **66**(11), 10346–10356 (2017)
6. Gkioxari, G., Girshick, R., Dollar, P., He, K.: Detecting and recognizing human-object interactions. In: CVPR'2018, pp. 8359–8367 (2018)
7. Golyandina, N., Nekrutkin, V., Zhigljavsky, A.: Analysis of time series structure: SSA and related techniques. Monographs on Statistics and Applied Probability 90 (2001)
8. Guan, Y., Plötz, T.: Ensembles of deep LSTM learners for activity recognition using wearables. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) **1**(2) (2017)
9. Halperin, D., Hu, W., Sheth, A., Wetherall, D.: Predictable 802.11 packet delivery from wireless channel measurements. In: SIGCOMM'2010, pp. 159–170. ACM (2010)
10. Jiang, W., Miao, C., Ma, F., Yao, S., Wang, Y., Yuan, Y., Xue, H., Song, C., Ma, X., Koutsonikolas, D., Xu, W., Su, L.: Towards environment independent device free human activity recognition. In: MobiCom'2018, pp. 289–304. ACM (2018)
11. Khan, M.I., Jan, M.A., Muhammad, Y., Do, D.T., ur Rehman, A., Mavromoustakis, C.X., Pallis, E.: Tracking vital signs of a patient using channel state information and machine learning for a smart healthcare system. Neural Computing and Applications (2021)
12. Li, H., Yang, W., Wang, J., Xu, Y., Huang, L.: WiFinger: Talk to your smart devices with finger-grained gesture. In: UbiComp'2016, pp. 250–261. ACM (2016)
13. Liu, J., Liu, H., Chen, Y., Wang, Y., Wang, C.: Wireless sensing for human activity: A survey. IEEE Communications Surveys & Tutorials **22**(3), 1629–1645 (2020)
14. Liu, X., Cao, J., Tang, S., Wen, J., Guo, P.: Contactless respiration monitoring via off-the-shelf WiFi devices. IEEE Transactions on Mobile Computing **15**(10), 2466–2479 (2016)
15. Ma, Y., Zhou, G., Wang, S.: WiFi sensing with channel state information: A survey. ACM Computing Surveys **52**(3), 1–36 (2019)
16. Markopoulos, P., Zlotnikov, S., Ahmad, F.: Adaptive radar-based human activity recognition with L1-norm linear discriminant analysis. IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology **3**(2), 120–126 (2019)
17. Orphomma, S., Swangmuang, N.: Exploiting the wireless RF fading for human activity recognition. In: 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, pp. 1–5 (2013)
18. Qi, F., Li, Z., Liang, F., Lv, H., An, Q., Wang, J.: A novel time-frequency analysis method based on HHT for finer-grained human activity using SFCW radar. In: 2016 Progress in Electromagnetic Research Symposium (PIERS), pp. 2536–2539 (2016)

19. Shang, J., Wu, J.: Fine-grained vital signs estimation using commercial Wi-Fi devices. In: 8th Wireless of the Students, by the Students, and for the Students Workshop, pp. 30–32. ACM (2016)
20. Shen, S., Wang, H., Roy Choudhury, R.: I am a smartwatch and I can track my user's arm. In: Mobisys'2016, pp. 85–96 (2016)
21. Sheng, B., Xiao, F., Sha, L., Sun, L.: Deep spatial–temporal model based cross-scene action recognition using commodity WiFi. IEEE Internet of Things Journal **7**(4), 3592–3601 (2020)
22. Sigg, S., Shi, S., Ji, Y.: RF-based device-free recognition of simultaneously conducted activities. In: UbiComp'13 Adjunct, pp. 531–540. ACM (2013)
23. Wang, F., Gong, W., Liu, J.: On spatial diversity in WiFi-based human activity recognition: A deep learning-based approach. IEEE Internet of Things Journal **6**(2), 2035–2047 (2019)
24. Wang, G., Zou, Y., Zhou, Z., Wu, K., Ni, L.: We can hear you with Wi-Fi! IEEE Transactions on Mobile Computing **15**, 2907–2920 (2016)
25. Wang, H., Zhang, D., Ma, J., Wang, Y., Wang, Y., Wu, D., Gu, T., Xie, B.: Human respiration detection with commodity wifi devices: Do user location and body orientation matter? In: UbiComp'2016, pp. 25–36. ACM (2016)
26. Wang, H., Zhang, D., Wang, Y., Ma, J., Wang, Y., Li, S.: RT-Fall: A real-time and contactless fall detection system with commodity WiFi devices. IEEE Transactions on Mobile Computing **16**(2), 511–526 (2017)
27. Wang, J., Zhang, X., Gao, Q., Yue, H., Wang, H.: Device-free wireless localization and activity recognition: A deep learning approach. IEEE Transactions on Vehicular Technology **66**(7), 6258–6267 (2017)
28. Wang, M., Ni, B., Yang, X.: Recurrent modeling of interaction context for collective activity recognition. In: CVPR'2017, pp. 7408–7416 (2017)
29. Wang, W., Liu, A.X., Shahzad, M., Ling, K., Lu, S.: Understanding and modeling of WiFi signal based human activity recognition. In: MobiCom'2015, pp. 65–76. ACM (2015)
30. Wang, W., Liu, A.X., Shahzad, M., Ling, K., Lu, S.: Device-free human activity recognition using commercial WiFi devices. IEEE Journal on Selected Areas in Communications **35**(5), 1118–1131 (2017)
31. Wang, X., Yang, C., Mao, S.: PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices. In: ICDCS'2017, pp. 1230–1239 (2017)
32. Wang, X., Yang, C., Mao, S.: TensorBeat: Tensor decomposition for monitoring multi-person breathing beats with commodity WiFi. ACM Transactions on Intelligent Systems and Technology **9** (2017)
33. Wu, D., Gao, R., Zeng, Y., Liu, J., Wang, L., Gu, T., Zhang, D.: Fingerdraw: Sub-wavelength level finger motion tracking with wifi signals. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) **4**(1), 1–27 (2020)
34. Xiao, C., Han, D., Ma, Y., Qin, Z.: CsiGAN: Robust channel state information-based activity recognition with GANs. IEEE Internet of Things Journal **6**(6), 10191–10204 (2019)
35. Zhang, F., Niu, K., Xiong, J., Jin, B., Gu, T., Jiang, Y., Zhang, D.: Towards a diffraction-based sensing approach on human activity recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) **3**(1), 1–25 (2019)
36. Zhao, M., Adib, F., Katabi, D.: Emotion recognition using wireless signals. Communications of the ACM **61**, 91–100 (2018)
37. Zheng, Y., Zhang, Y., Qian, K., Zhang, G., Liu, Y., Wu, C., Yang, Z.: Zero-effort cross-domain gesture recognition with wi-fi. In: MobiSys'2019, pp. 313–325. ACM (2019)