

CSI-based Device-free Wireless Localization and Activity Recognition Using Radio Image Features

Qinhua Gao, Jie Wang, *Member, IEEE*, Xiaorui Ma, Xueyan Feng, Hongyu Wang, *Member, IEEE*

Abstract—Device-Free wireless Localization and Activity Recognition (DFLAR) is an emerging technique which could estimate the location and activity of a person without equipping him/her with any device. It deduces the state of a person by analyzing his/her influence on surrounding wireless signals. Therefore, how to characterize the influence of human behaviors is the key question. In this paper, we explore and exploit a radio image processing approach to better characterize the influence of human behaviors on WiFi signals. Traditional methods deal with channel state information (CSI) measurements on each channel independently. However, CSI measurements on different channels are correlated, and thus lots of useful information involved with channel correlation may lose. This motivates us to look on CSI measurements from multiple channels as a radio image and deal with it from the two-dimensional perspective. Specifically, we transform CSI measurements from multiple channels into a radio image, extract color and texture features from the radio image, adopt a deep learning network to learn optimized deep features from image features, and estimate the location and activity of a person using a machine learning approach. Benefit from the informative and discriminative deep image features, experimental results in two clutter laboratories confirm the excellent performance of the proposed system.

Index Terms—Localization, device-free, activity recognition, radio image, feature extraction

I. INTRODUCTION

With the development of wireless technique, wireless signals have already emerged as the most pervasive signals. No matter where we are, there are always many wireless signals around us. Recent research has shown that the influence of human behaviors on surrounding wireless signals can be exploited as a powerful sensing tool to locate and recognize the activity of the person in the vicinity of a wireless network, which enables radio transceivers with the capability of acquiring an accurate human-scale understanding of space and motion. This emerging technique is called Device-free wireless Localization and Activity Recognition (DFLAR) since it does not require equipping a person with any device [1], [2].

DFLAR is the key technique that enables many potential applications, *e.g.*, smart homes, fitness tracking systems, and

This work was supported by National Natural Science Foundation of China under grant 61671102 and 61401059, Fundamental Research Funds for the Central Universities under grant DUT17JC03, DUT17RC(3)024, and DUT16QY33, Scientific Research Starting Foundation for the Returned Overseas Chinese Scholars, and Xinghai Scholars Program. Jie Wang is the corresponding author.

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Q. Gao, J. Wang, X. Ma, X. Feng, and H. Wang are with the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116023, China. e-mail: {wangjie@, qhao@, maxr@, fengxy@mail., whyu@}dlut.edu.cn.

search-and-rescue systems. In these applications, traditional approaches mainly use wearable sensors, cameras, or radars, to sense the location and activity of a person. However, all these approaches have certain disadvantages. Wearable sensor based methods are inconvenient and we always forget to wear sensors. Camera based solutions have strict requirements on lighting and line-of-sight conditions, and may cause privacy leakage. Radar based approaches are expensive and have limited sensing range. DFLAR technique overcomes the above disadvantages. It deduces the state of a person by analyzing his/her influence on surrounding wireless signals. Therefore, it does not require lighting, has better coverage, does not need equipping users with any device, and can preserve user identity privacy. Due to the aforementioned advantages and the pervasive of WiFi signals, WiFi based DFLAR becomes a promising candidate technique for future mobile computing applications.

As an emerging technique, WiFi based DFLAR still faces many challenges. The fundamental and essential problem is how to better characterize the influence of human behaviors on WiFi signals, *i.e.*, extract features from WiFi signals. Channel state information (CSI) extracted from WiFi physical layer can provide fine-grained information on multiple channels [3]. Therefore, CSI is a preferable choice to realize DFLAR. Pioneer work has done valuable exploration on how to utilize CSI measurements to realize device free localization [4]–[6] and activity recognition [7]–[9]. They develop methods to extract statistical features in both time-domain and frequency-domain from CSI measurements on every channel independently. As we know, subcarrier frequencies of different WiFi channels are adjacent to each other, and thus CSI measurements on multiple channels must be correlated. If we extract features from each channel independently, lots of useful information involved with channel correlation may loss. This motivates us to explore new methods to extract features from CSI measurements on multiple channels simultaneously from the two-dimensional perspective. In this paper, we look on CSI measurements from multiple channels as a radio image with time as x-axis and channel as y-axis, and develop a deep learning based image processing framework to extract optimized deep features from CSI measurements, so as to achieve better DFLAR performance.

The main contributions of the paper can be summarized as follows.

- 1) We propose a deep learning based image processing framework for extracting discriminative deep image features from radio images. Specifically, we transform CSI measurements from multiple channels into radio images,

extract color and texture features from the radio images, and develop a deep learning network to learn optimized deep image features from the raw image features.

- 2) We leverage both the amplitude and phase information of the CSI measurements on multiple channels to construct radio images, and deal with them from the two-dimensional perspective which preserves the channel correlation information between multiple channels.
- 3) We conduct experiments with hardware testbed in two clutter indoor laboratories, and evaluate the proposed schemes extensively.

The rest of the paper is structured as follows. Section II reviews the related work on device-free wireless localization and activity recognition. Section III presents the motivation behind the proposed schemes. Section IV introduces the detailed methodology of the proposed radio image processing based system. Section V validates the proposed schemes with extensive experimental evaluations. Finally, the conclusion is drawn in section VI.

II. RELATED WORK

Device free wireless sensing technique [1], [2] dates from the idea of device free localization (DFL). Zhang *et al.* [10] explore methods to realize DFL using the received signal strength (RSS) of wireless network signals with a geometric approach. Wang *et al.* [4], [5] reveal the advantage of time-of-flight and envision its applications in DFL. Wang *et al.* [6] adopt the amplitude of CSI measurements on multiple channels to realize DFL. Zhao and Patwari [11] utilize signal variance as the feature to realize DFL using radio tomography method. Wang *et al.* [12], [13], Guo *et al.* [14], and Yang *et al.* [15] adopt an ellipsoid model, an exponential-rayleigh model, and a saddle surface model to characterize the shadowing effect of a person on wireless links to realize DFL. Liu *et al.* [16] and Wang *et al.* [17] use diffraction theory to depict the shadowing effect and achieve reasonable DFL performance. Hong and Ohtsuki [18] extract signal eigenvectors from wireless signal amplitudes acquired by an antenna array and adopt them to realize DFL. Most of the aforementioned methods adopt time-domain mean or variance of signal amplitudes to realize DFL. These signal features are sufficient for realizing location estimation. However, to realize activity recognition, we may need to extract more elaborate and informative features from wireless signals or WiFi signals.

Sigg *et al.* [7] explore different approaches to extract features from WiFi signals. They extract mean, variance, and peak-to-peak values of the signal amplitudes in time-domain, and the corresponding energy and entropy in frequency-domain, to characterize the influence of human activities on WiFi signals. Wang *et al.* [8] leverage the moving variance of the CSI amplitude measurements to differentiate walking activities and in-place activities, and adopt raw CSI amplitude measurement and CSI amplitude distribution as features to recognize walking activities and in-place activities, respectively. Wang *et al.* [9] analyse the relationship between the moving speed of a person and the variation of CSI amplitudes, and adopt time-frequency analysis method to extract features from

CSI amplitude measurements to realize activity recognition. Wang *et al.* [19], [20] extract wavelet and deep features from RSS measurements on multiple wireless links using the wavelet transformation and the deep learning approach to realize localization and activity recognition simultaneously. Different with above work, we transform CSI measurements into radio images, so as to preserve the channel correlation information. In this paper, although we also use deep learning networks to extract features, our goal is to optimize the radio image features, which is different with the scheme in [20] that tries to extract features from the raw RSS measurements. Chang *et al.* [21], [22] transform CSI measurements into image and investigate the feasibility of processing CSI using vision-based methods. Their interesting work open a new window for the DFLAR technique.

Device free wireless sensing is a promising technique that develops rapidly. Many interesting applications are emerging, such as human detection [23], fall detection [24], respiration detection [25], gait recognition [26], gesture recognition [27], among many others. The key problem behind all these applications is how to better characterize the influence of human behaviors on WiFi signals, *i.e.*, extract features from WiFi signals. Although the proposed deep learning based image processing framework is for realizing DFLAR in this paper, it can be utilized by the aforementioned applications as well.

III. MOTIVATION

Traditional methods deal with CSI measurements on each WiFi channel independently. However, CSI measurements on different channels are correlated, and thus lots of useful information involved with channel correlation may loss if we do not deal with multiple channels jointly. This motivates us to look on CSI measurements from multiple channels as a radio image and deal with it from the two-dimensional perspective. To evaluate whether this idea is valid, we carry out some preliminary tests using two embedded industrial computers equipped with Intel 5300 network interface card [3]. We place the two computers 10m apart, and let one transmit packets continuously to another. The receiver records all the CSI measurements while a person performs different activities and stands at different locations. Suppose we have C channels, after recording the CSI measurements on multiple channels for consecutive T times, we can get a two dimensional CT measurement matrix. We take time as x-axis and channel as y-axis and transform the CSI measurements on multiple channels into radio images. Fig. 1 illustrates some radio images acquired while a person performs different activities at the same location and performs the same activity at different locations. From the figures, we can see that the radio images for different activities and different locations are intuitively easy to classify, no matter using amplitude or phase measurement to construct the radio image. **If we can extract discriminative features from the visually distinguishable radio images, we will be sure to achieve better DFLAR performance.** Therefore, this paper explores and exploits a deep learning based image processing framework for extracting discriminative deep image features from radio images, so as to achieve better DFLAR performance. Specifically, we leverage both the amplitude and phase

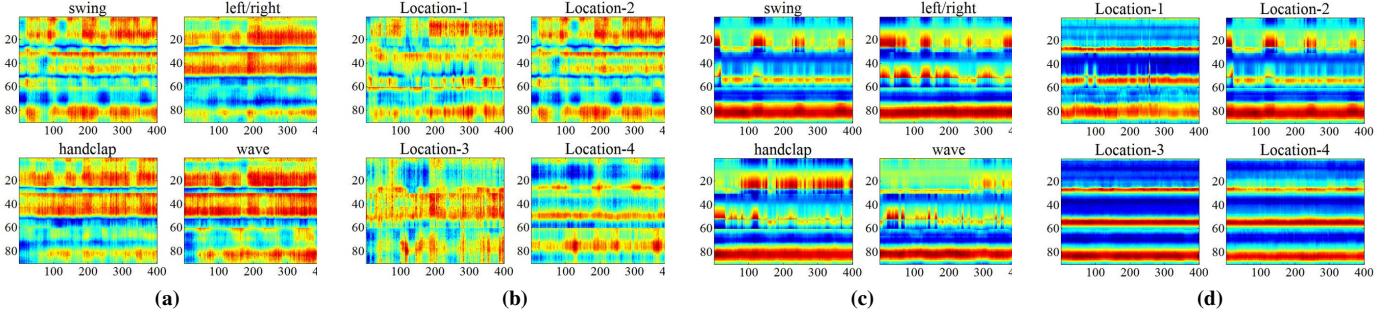


Fig. 1 Radio images of different activities and locations using amplitude and phase. (a) Performing different activities at location 2 using amplitude. (b) Performing swing arms activity at different locations using amplitude. (c) Performing different activities at location 2 using phase. (d) Performing swing arms activity at different locations using phase.

information of the CSI measurements on multiple channels to construct radio images, and propose a deep learning based image processing framework for extracting discriminative deep image features from radio images, so as to better characterize the influence of human behaviors on wireless signals.

IV. RADIO IMAGE PROCESSING BASED DFLAR SYSTEM

A. System Architecture

The system architecture of the proposed radio image based DFLAR system is shown in Fig. 2. The system mainly consists of the following modules: Radio image construction module which collects CSI amplitude and phase measurements, preprocesses the measurements, and transforms them into radio images. Image feature extraction module that extracts the color and texture features from radio images. Deep feature extraction module which extracts optimized deep image features from the raw image features using a deep learning network. Classification module that utilizes softmax regression algorithm [28], [29] to estimate the location and activity of the person. The parameters of the deep learning network and the softmax regression model should be learned offline using the training data sets. In the online phase, the system outputs the location and activity estimation directly using the currently measured radio image information. We will present the detailed implementation of each module in the following sections.

B. Radio Image Construction

Commercial WiFi infrastructures, such as Intel 5300 wireless network interface card [3], could provide physical layer channel frequency response information of multiple channels (subcarriers) to upper layers in the format of CSI as follows

$$\mathbf{H}_t = [H_t^1, \dots, H_t^c, \dots, H_t^C]^T, \quad (1)$$

where \mathbf{H}_t denotes the CSI measurement vector acquired at time t , C indicates the total number of channels, T denotes the transposition operation, and H_t^c represents the CSI of channel c at time t which is defined by

$$H_t^c = A_t^c e^{j\phi_t^c}, \quad (2)$$

where A_t^c and ϕ_t^c denote the amplitude and phase of channel c at time t .

We will utilize both amplitude and phase measurements to realize DFLAR. For the amplitude measurements, we first remove the influence of surrounding environment by subtracting the reference amplitudes acquired when the scenario is vacant from the currently measured amplitudes, then, use a simple median filter with the width of 5 to remove measurement noise. After pre-processing, we achieve the preprocessed amplitude measurement matrix $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_t, \dots, \mathbf{A}_T]$, where $\mathbf{A}_t = [A_t^1, \dots, A_t^c, \dots, A_t^C]^T$.

For the phase measurements, due to the asynchronous and hardware difference between the transmitter and receiver, the relationship between the measured phase ϕ_t^c and the ground earth true phase $\hat{\phi}_t^c$ is as follows [30]

$$\phi_t^c = \hat{\phi}_t^c + 2\pi f_c \Delta t + \beta + Z_t^c, \quad (3)$$

where f_c denotes the frequency of channel c , Δt represents the unknown time lag between the transmitter and receiver, $2\pi f_c \Delta t$ indicates the phase shift incurred by the propagation of signals, β denotes the unknown initial phase of the transmitted packet, and Z_t^c represents the measurement noise.

Since we feed phase measurements to a classifier to realize DFLAR, thus, we should make sure that the phase measurements are invariable with Δt and β . To eliminate the impact of these two random parameters, we define two calibration parameters a and b based on the phase measurements on multiple channels as follows

$$a = \frac{\phi_t^C - \phi_t^1}{C} = \frac{\hat{\phi}_t^C - \hat{\phi}_t^1}{C} + \frac{2\pi \Delta t}{C} (f_C - f_1), \quad (4)$$

$$b = \frac{1}{C} \sum_{c=1}^C \phi_t^c = \frac{1}{C} \sum_{c=1}^C \hat{\phi}_t^c + \frac{2\pi \Delta t}{C} \sum_{c=1}^C f_c + \beta. \quad (5)$$

The calibrated phase $\bar{\phi}_t^c$ can be calculated by

$$\bar{\phi}_t^c = \phi_t^c - af_c - b. \quad (6)$$

According to the 802.11n wireless LAN physical layer specification, the subcarrier frequencies for WiFi are symmetric. Therefore, the calibrated phase measurements are invariable with Δt and β anymore. After the calibration, we adopt a median filter with the width of 5 to remove measurement noise.

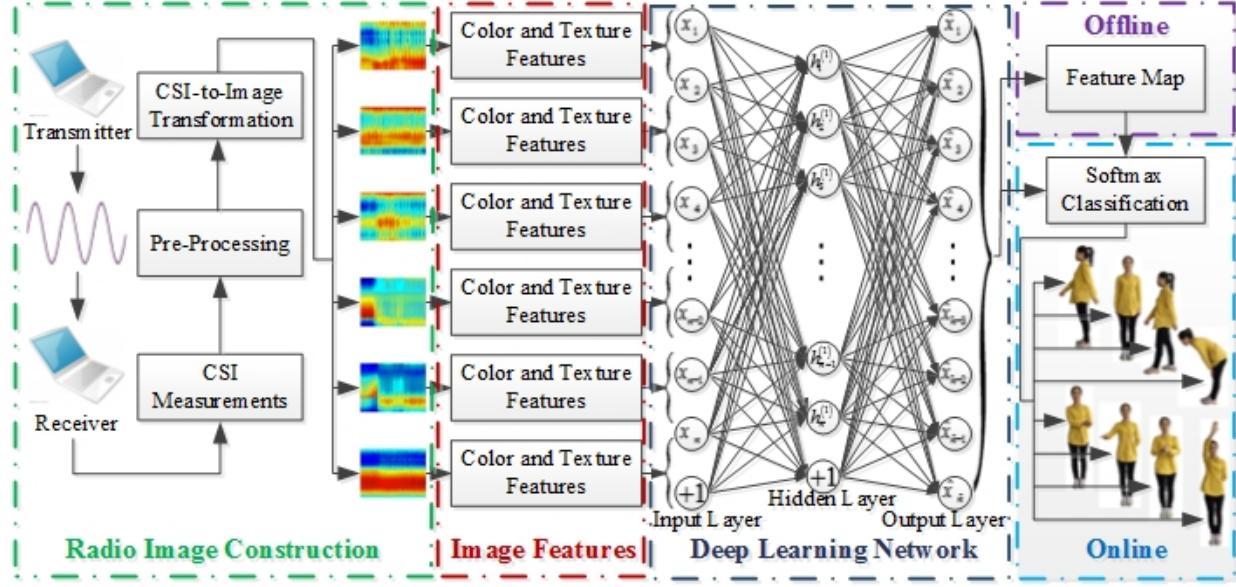


Fig. 2 System architecture of the radio image based DFLAR system.

After calibration and pre-processing, we achieve the preprocessed phase measurement matrix $\Phi = [\Phi_1, \dots, \Phi_t, \dots, \Phi_T]$, where $\Phi_t = [\bar{\phi}_t^1, \dots, \bar{\phi}_t^c, \dots, \bar{\phi}_t^C]^T$.

The preprocessed amplitude measurement matrix \mathbf{A} and phase measurement matrix Φ can be looked as matrix form of radio images, with row (time) as x-axis and column (channel) as y-axis. For every physical wireless link from a transmitting antenna to a receiving antenna, we can acquire T consecutive CSI measurements, and transform the amplitude and phase measurements into two radio images.

C. Image Feature Extraction

Let \mathcal{I} be the radio image transformed by the amplitude or phase measurements. We quantize the colors in \mathcal{I} into N_c colors. For a pixel $p = (x, y) \in \mathcal{I}$, let $\mathcal{I}(p)$ denotes its color. The pixel set with color n is represented by $\mathcal{I}_n \triangleq \{p | \mathcal{I}(p) = n\}$. For simplicity, we adopt the L_∞ -norm to measure the distance between pixels, *i.e.*, for pixels $p_i = (x_i, y_i)$ and $p_j = (x_j, y_j)$, we have

$$|p_i - p_j| = \max \{|x_i - x_j|, |y_i - y_j|\}. \quad (7)$$

We will explore methods to extract color and texture features from radio images. Specifically, we extract color features using color coherence vectors, color correlograms, and color autocorrelograms, and extract texture features using gabor filter and gray level co-occurrence matrix [31].

For any pixel in an image \mathcal{I} , the probability that the color of the pixel is n can be defined as follows

$$h_n = \frac{|\mathcal{I}_n|}{|\mathcal{I}|}, \quad (8)$$

where $|\mathcal{I}_n|$ and $|\mathcal{I}|$ denote the number of pixels within the pixel set \mathcal{I}_n and within the image \mathcal{I} , respectively. The color histogram is defined as follows

$$H = [h_1, \dots, h_n, \dots, h_{N_c}], \quad (9)$$

where N_c indicates the total number of colors.

To evaluate the spatial correlation between the pixels within the same pixel set, we further define coherent region and incoherent region. For a given coherent distance d , if the distance between any two pixels in a subset of the pixel set \mathcal{I}_n is less than or equal to d , we term this subset as a coherent region, and the pixels within this region are coherent pixels. With the aforementioned scheme, we can divide every pixel set \mathcal{I}_n into two subsets, *i.e.*, coherent set and incoherent set, and construct the color coherence vector as follows

$$\mathcal{V} = [(h_1^c, h_1^i) \dots, (h_n^c, h_n^i) \dots, (h_{N_c}^c, h_{N_c}^i)], \quad (10)$$

where h_n^c denotes the probability that a pixel is not only with color n but also a coherent pixel, and h_n^i indicates that of an incoherent pixel, respectively.

Given any pixel of color m in the radio image \mathcal{I} , color correlogram defines the probability that a pixel at distance d away from the given pixel is of color n as follows

$$\mathbf{C}_{m,n}^d = \Pr_{p_i \in \mathcal{I}_m, p_j \in \mathcal{I}} [p_j \in \mathcal{I}_n | |p_i - p_j| = d]. \quad (11)$$

The color correlogram with distance d is defined as

$$\mathbf{C}^d = [\mathbf{C}_{1,1}^d, \dots, \mathbf{C}_{m,n}^d, \dots, \mathbf{C}_{N_c,N_c}^d]. \quad (12)$$

If we adopt N_d different distances, the color correlogram with multiple distances is defined as

$$\mathcal{C} = [\mathbf{C}^{d_1}, \dots, \mathbf{C}^{d_{N_d}}]. \quad (13)$$

If we consider only the correlation between identical colors, the autocorrelogram is defined by

$$\overline{\mathbf{C}}_n^d = \mathbf{C}_{n,n}^d. \quad (14)$$

Using the similar method, we can construct the color autocorrelogram with multiple distances $\overline{\mathcal{C}}$. Note that the size of \mathcal{C} is $1 \times N_c N_c N_d$, while the size of $\overline{\mathcal{C}}$ is reduced to $1 \times N_c N_d$,

Different from color features which try to characterize an image by depicting every pixel, texture features can quantify the perceived texture of an image on the whole. They can provide information about the spatial arrangement of colors and intensities in an image.

Gabor filter [32] is a good texture feature extractor which can be utilized to achieve the spatial arrangement information of an image in a specific wavelength and orientation by convolving the image with a Gabor filter at that wavelength and orientation. It is a complex sinusoidally modulated Gaussian function as follows

$$\mathbf{G}(x, y, \lambda, \theta, \sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left\{ -\frac{1}{2} \left[\frac{R_x^2}{\sigma_x^2} + \frac{R_y^2}{\sigma_y^2} \right] \right\} \times \exp \left[i \cdot \frac{2\pi R_1}{\lambda} \right], \quad (15)$$

where

$$R_1 = x \cos \theta + y \sin \theta,$$

$$R_2 = -x \sin \theta + y \cos \theta,$$

λ and θ represent the wavelength and orientation of the wave, and σ_x and σ_y denote the standard deviation of the Gaussian envelope along the x-axis and y-axis, respectively.

By setting λ and θ as different values, we can construct gabor filters with different frequency response and directional selectivity, thus, enable the filter to extract orientation-dependent frequency contents information. For a given λ and θ , we can extract the gabor feature by convolving the filter with the radio image as follows

$$\bar{\mathbf{G}}(x, y) = \mathcal{I}(x, y) * \mathbf{G}(x, y). \quad (16)$$

The gabor feature can be viewed as the response of the gabor filter located at a sampling point. After the convolution operation, the gabor feature $\bar{\mathbf{G}}(x, y)$ has the same size as the radio image $\mathcal{I}(x, y)$, which may be a little large. To reduce the computational complexity, we adopt the mean μ and variance σ as metrics to represent $\bar{\mathbf{G}}(x, y)$ as follows

$$\mu = \frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y \bar{\mathbf{G}}(x, y), \quad (17)$$

$$\sigma = \sqrt{\frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y (\bar{\mathbf{G}}(x, y) - \mu)^2}, \quad (18)$$

where X and Y represent the length and width of the radio image.

Suppose we have N_λ wavelengths and N_θ orientations, the final gabor features \mathcal{G} is a $1 \times 2N_\lambda N_\theta$ vector as

$$\mathcal{G} = [(\mu_1, \sigma_1), \dots, (\mu_{N_\lambda}, \sigma_{N_\theta})]. \quad (19)$$

Similar to color correlogram, the Gray level co-occurrence matrix (GLCM) [31] is also an estimator of the joint probability density function of gray level pairs in an image. The difference is that to extract texture information, GLCM takes also the orientation between the two pixels into consideration. Given any pixel of color m in the radio image \mathcal{I} , GLCM defines the probability that a pixel in the orientation o with a distance d away from the given pixel is of color n as follows

$$\mathbf{M}_{m,n}^{d,o} = \frac{P_r}{p_i \in \mathcal{I}_m, p_j \in \mathcal{I}} [p_j \in \mathcal{I}_n | |p_i - p_j| = d, \overrightarrow{p_i p_j} = o], \quad (20)$$

where $\overrightarrow{p_i p_j}$ indicates the orientation from p_i to p_j .

Consider every possible color, we can achieve the $N_c \times N_c$ GLCM. We adopt the following 4 metrics to extract features from GLCM. The angular second moment metric which measures the homogeneity of the radio image as

$$\text{ASM}_{d,o} = \sum_{m=1}^{N_c} \sum_{n=1}^{N_c} (\mathbf{M}_{m,n}^{d,o})^2. \quad (21)$$

The entropy metric which measures the disorder, randomness, and complexity of the image as

$$\text{ENT}_{d,o} = - \sum_{m=1}^{N_c} \sum_{n=1}^{N_c} \mathbf{M}_{m,n}^{d,o} \lg \mathbf{M}_{m,n}^{d,o}. \quad (22)$$

The inverse difference moment metric which is inversely proportional to the image energy as follows

$$\text{IDM}_{d,o} = \sum_{m=1}^{N_c} \sum_{n=1}^{N_c} \frac{\mathbf{M}_{m,n}^{d,o}}{1 + (m - n)^2}. \quad (23)$$

The correlation metric which illustrates the spatial arrangement of gray levels. It indicates the linearity between pixel pairs as follows

$$\text{COR}_{d,o} = \frac{\left[\sum_{m=1}^{N_c} \sum_{n=1}^{N_c} mn \mathbf{M}_{m,n}^{d,o} \right] - \mu_1 \mu_2}{\sigma_1^2 \sigma_2^2}, \quad (24)$$

where

$$\mu_1 = \sum_{m=1}^{N_c} m \sum_{n=1}^{N_c} \mathbf{M}_{m,n}^{d,o},$$

$$\mu_2 = \sum_{n=1}^{N_c} n \sum_{m=1}^{N_c} \mathbf{M}_{m,n}^{d,o},$$

$$\sigma_1 = \sqrt{\sum_{m=1}^{N_c} (m - \mu_1)^2 \sum_{n=1}^{N_c} \mathbf{M}_{m,n}^{d,o}},$$

$$\sigma_2 = \sqrt{\sum_{n=1}^{N_c} (n - \mu_2)^2 \sum_{m=1}^{N_c} \mathbf{M}_{m,n}^{d,o}}.$$

With the aforementioned metrics, the GLCM feature for distance d and orientation o is defined as

$$\mathcal{M}_{d,o} = [\text{ASM}_{d,o}, \text{ENT}_{d,o}, \text{IDM}_{d,o}, \text{COR}_{d,o}]. \quad (25)$$

If we adopt N_d different distances and N_o different orientations, the GLCM feature \mathcal{M} is a $1 \times 4N_d N_o$ vector as follows

$$\mathcal{M} = [\mathcal{M}_{1,1}, \dots, \mathcal{M}_{N_d, N_o}]. \quad (26)$$

D. Deep Feature Extraction

With the aforementioned radio image features, *i.e.*, color coherence vector \mathcal{V} , color correlogram \mathcal{C} , color autocorrelogram $\bar{\mathcal{C}}$, gabor features \mathcal{G} , and GLCM features \mathcal{M} , we will further feed them into a deep learning network to learn optimized deep image features. It should be mentioned that the proposed framework is a general solution for extracting features for DFLAR. Therefore, we term the radio image features as a vector x , which may be one of the aforementioned features, the combination of multiple features, *e.g.*, the fusion of color autocorrelogram $\bar{\mathcal{C}}$ and GLCM features \mathcal{M} as $[\bar{\mathcal{C}}, \mathcal{M}]$, or the

combination of multiple features from multiple radio images, *e.g.*, the fusion features from 3 amplitude radio images and 3 phase radio images that transformed by the CSI measurement from 3 wireless links.

Since x is an one-dimensional vector, we adopt the sparse auto-encoder (SAE) network [33], [34], which is a high performance deep learning network suitable for dealing with one-dimensional data, to extract optimized deep image features. Compared with traditional neural networks, SAE can learn the initial network parameters with an unsupervised manner, which overcomes the initialization problem and simplifies the network training procedure. Taking a SAE network with three layers, the input layer, hidden layer, and output layer, as an example which is shown in Fig.2 and Fig. 3. The number of units in the above three layers are n , \bar{n} , and \hat{n} , respectively. The network parameters are $(W, b) = (W^{(1)}, b^{(1)}, W^{(2)}, b^{(2)})$, where $W^{(1)}$ is a $\bar{n} \times n$ matrix, $b^{(1)}$ denotes a $\bar{n} \times 1$ vector, $W^{(2)}$ indicates a $\hat{n} \times \bar{n}$ matrix, and $b^{(2)}$ represents a $\hat{n} \times 1$ vector. $W_{ij}^{(l)}$ represents the weighting parameter associated with the connection between unit j in layer l and unit i in layer $l+1$, and $b_i^{(l)}$ indicates the bias item of unit i in layer $l+1$. For a radio image feature vector $x(m)$, the activation (output) of the hidden layer $a^{(2)}(m)$ can be calculated as

$$a^{(2)}(m) = f(W^{(1)}x(m) + b^{(1)}), \quad (27)$$

where $f(\cdot)$ represents the activation function which is a sigmoid function in this paper. Using the similar method, we can calculate the output of every layer, thus, achieve the optimized deep image features.

Now, the problem is how to efficiently determine the network parameters using a training set $\{x(m)|m=1, \dots, M\}$. As illustrated in Fig. 3, SAE leverages two schemes to facilitate the parameter learning procedure, *i.e.*, layer-by-layer unsupervised learning and the sparse constraint imposed on the averaged activation.

The principle of the layer-by-layer unsupervised learning procedure is try to optimize the parameters to ensure that the decoder output should be equal to the input. Taking the hidden layer output $a^{(2)}(m)$ as an example, if we feed $a^{(2)}(m)$ into the decoder, the decoder output $r^{(2)}(m)$ is

$$r^{(2)}(m) = f(W^{(1)\top}a^{(2)}(m) + b^{(1)}), \quad (28)$$

where $W^{(1)\top}$ indicates the transposed matrix of $W^{(1)}$, and $b^{(1)}$ denotes the bias item.

With the aforementioned encoding and decoding process, we can calculate the output and decoded output for each layer. The cost function can be formulated as

$$\begin{aligned} \widehat{J}(W, b) = & \left[\frac{1}{M} \sum_{m=1}^M \left(\frac{1}{2} \|r^{(l)}(m) - a^{(l-1)}(m)\|^2 \right) \right] \\ & + \frac{\alpha}{2} \sum_{j=1}^{u_{l-1}} \sum_{i=1}^{u_l} \left(W_{ij}^{(l-1)} \right)^2, \end{aligned} \quad (29)$$

where the first term represents the reconstruction error of the encoding and decoding procedure, the second term indicates the weight decay, α controls the relative importance of the two terms, and u_l denotes the number of units in layer l .

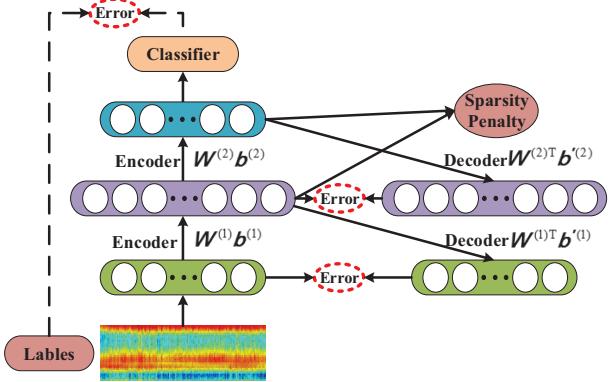


Fig. 3 Illustration of the sparse auto-encoder network.

Since the network is used to extract features for classification, thus, only a small number of units are useful. We want to keep the useful units to be active while others stay inactive. Therefore, we impose a sparsity penalty on the output of each layer using the Kullback-Leibler divergence metric [35] as follows

$$\text{KL}(\rho || \tilde{\rho}_j^{(l)}) = \rho \log \frac{\rho}{\tilde{\rho}_j^{(l)}} + (1 - \rho) \log \frac{1 - \rho}{1 - \tilde{\rho}_j^{(l)}}, \quad (30)$$

where ρ is the sparse degree parameter with a value close to 0, and $\tilde{\rho}_j^{(l)}$ denotes the average activation of unit j in layer l which is defined as

$$\tilde{\rho}_j^{(l)} = \frac{1}{M} \sum_{m=1}^M \left[a_j^{(l)}(m) \right], \quad (31)$$

where $a_j^{(l)}(m)$ represents the activation of unit j in layer l with an input sample $x(m)$.

Taking both the sparse constraint and the reconstruction error into consideration, the cost function is

$$J(W, b) = \widehat{J}(W, b) + \gamma \sum_{j=1}^{u_l} \text{KL}(\rho || \tilde{\rho}_j^{(l)}), \quad (32)$$

where γ indicates the weight of the sparsity penalty.

By minimizing $J(W, b)$ iteratively with backpropagation algorithm [36], we can calculate the network parameters (W, b) . For an input radio image feature vector x , we feed it into the SAE network and adopt the output of the last layer as the optimized deep image features.

E. Classification

DFLAR is a typical multi-class classification problem. Therefore, we utilize the softmax regression algorithm [28], [29], which could estimate the probabilities of an input sample belongs to each class, to realize the classification task.

In the offline phase, we first train the parameters of the SAE network with an unsupervised manner using the training radio image data set. Then, we merge the fine-tuning procedure of the SAE network with the parameter learning procedure of the softmax regression model, and optimize the whole system with a supervised manner.

In the online phase, the system measures the CSI on multiple channels, constructs radio images, extracts image features

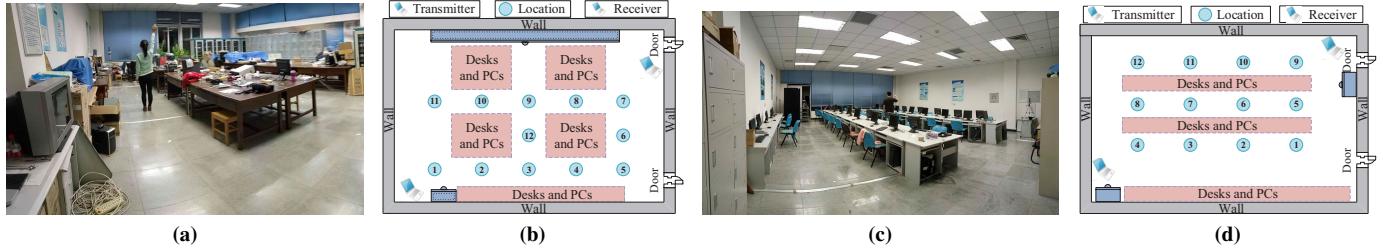


Fig. 4 Photographs and layouts of the two experiments. (a) Experimental scenario of the 1st experiment. (b) Layout of the 1st experiment. (c) Experimental scenario of the 2nd experiment. (d) Layout of the 2nd experiment.

TABLE I Performance of DFLAR systems using different single features.

Experiment	Features	T-Mean	T-Var	T-Peak	F-Energy	F-Entropy	C-CV	C-CC	C-AC	T-Gabor	T-GLCM
1st	Accuracy(%)	82.7	65.1	59.2	74.9	75.3	83.4	87.2	83.5	86.8	90.0
2nd	Accuracy(%)	88.5	67.7	57.8	84.7	85.2	88.6	92.0	89.8	92.1	94.8

from the images, achieves optimized deep image features from the image features, and then, feeds the deep features into softmax regression model and estimates the state of the person.

V. EXPERIMENTAL EVALUATION

A. Experiment Setting

We establish a prototype to evaluate the system performance in two challenging laboratories with the same size of $10.8m \times 7.8m$. The system consists of a transmitter and a receiver, which are two embedded industrial computers equipped with Intel 5300 network interface card and Ubuntu 11.04 operating system. We run the 802.11 CSI Tool [3] on the receiver to acquire and record CSI measurements on 30 channels from 3 antennas. The photographs and layouts of the two experiments are illustrated in Fig. 4. In the experiments, a person may perform one of eight feasible activities, *i.e.*, stand still, bow, swing arms, walk, arm up and down, arm left and right, handclap, and wave hand, on 12 possible locations. The receiver acquires CSI with a sampling frequency of 500Hz, and adopts the CSI measurements within a consecutive 4s to construct radio images. Since there are 1 transmitting antenna and 3 receiving antennas, there will be 3 wireless links. We can construct an amplitude image and a phase image for each antenna, therefore, we can construct 6 radio images with one data set. For every location or activity, we adopt 14 data sets for training and 14 data sets for testing.

In the following experiments, radio images are quantized into 20 colors. The coherence distant of the color coherence vector is 5. Color correlogram and autocorrelogram adopt 5 different distances, *i.e.*, 6, 7, 8, 9, and 10, to calculate image features. Gabor filter has a size of 17×17 , it has 5 different wavelengths, *i.e.*, $4, 4\sqrt{2}, 8, 4\sqrt{2}$, and 16, and 8 different orientations, *i.e.*, $0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}$, and $\frac{7\pi}{8}$. GLCM uses 3 different distances, *i.e.*, 1, 2, and 3, and adopts 4 different orientations, *i.e.*, $0, \frac{\pi}{4}, \frac{\pi}{2}$, and $\frac{3\pi}{4}$. We adopt a SAE with 3 layers. The number of units in the hidden layer and output layer are 500 and 600, the weight parameter of the cost function $\alpha = 3e - 6$, the sparsity parameter $\rho = 0.01$, and the weight of sparsity penalty term $\gamma = 1e - 2$. We use

TABLE II Performance of DFLAR systems using different fusion features.

Experiment	Features	M-TF	Wavelet	M-CT
1st	Accuracy(%)	85.6	86.5	91.4
	DP-Accuracy(%)	88.5	89.6	92.9
2nd	Accuracy(%)	90.8	91.2	95.2
	DP-Accuracy(%)	92.1	92.3	95.5

the percentage of correctly estimated states as the accuracy metric to evaluate the systems.

B. Performance Overview

To evaluate the performance of the proposed radio image based system, we compare the performance of DFLAR systems using our proposed features with several state-of-the-art time-domain features [6]–[8], [10]–[17], frequency-domain features [7], [9], and wavelet-domain features [19]. Specifically, we select three commonly used time-domain features, *i.e.*, mean, variance, and peak-to-peak value of CSI signals, and two widely used frequency-domain features, *i.e.*, spectrum energy and entropy. For simplicity, we abbreviate DFLAR system with mean value feature as T-Mean, variance feature as T-Var, peak-to-peak value feature as T-Peak, spectrum energy feature as F-Energy, spectrum entropy feature as F-Entropy, color coherence vector feature as C-CV, color correlogram feature as C-CC, color autocorrelogram feature as C-AC, gabor feature as T-Gabor, and GLCM feature as T-GLCM, respectively. Table I summarizes the performance of DFLAR systems using different single features. From the results, we can see that the proposed image features, such as GLCM, could achieve better performance than the traditional features.

We also compared the performance of difference fusion features. We abbreviate DFLAR system with the aforementioned three time-domain and two frequency-domain features as M-TF, wavelet features as Wavelet, and the fusion of color autocorrelogram and GLCM features as M-CT, respectively. Table II summarizes the performance of DFLAR systems using different fusion features. In the table, the DP-Accuracy represents the performance achieved by optimizing the fusion features using the deep learning network, while the Accuracy

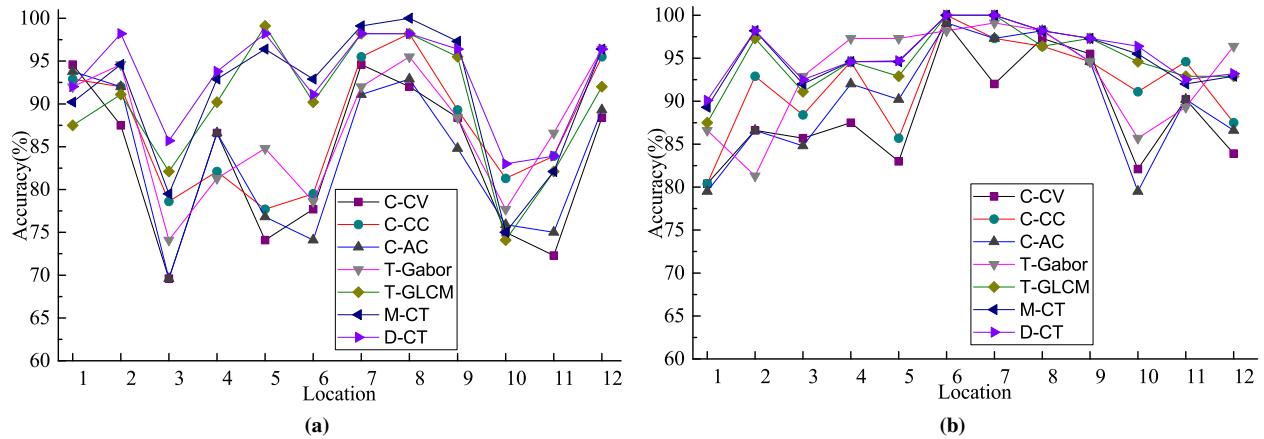


Fig. 5 Accuracy at different locations. (a) 1st experiment. (b) 2nd experiment.

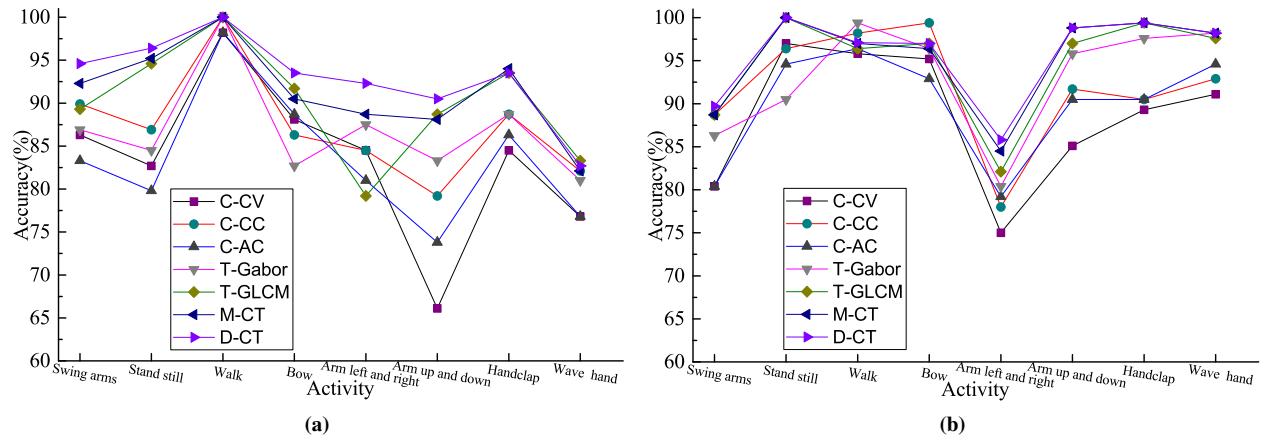


Fig. 6 Accuracy of different activities. (a) 1st experiment. (b) 2nd experiment.

indicates the performance using the original fusion features directly. From the table, we can see that our proposed image features outperform other fusion features, which confirms the effectiveness of the proposed radio image. Meanwhile, we also discover that DP-Accuracy is better than Accuracy, which confirms that the deep learning network can optimize the fusion features effectively. Specifically, if the performance of the original fusion features is relatively lower, the improvement will be more significant.

C. Performance Analysis

DFLAR system utilizes the influence of a person on wireless links to realize localization and activity recognition. When a person locates at different locations or performs different activities, the influence will be different. Intuitively, when the person is close to the wireless link or when the movement range of an activity is large, the influence will be more observably, and thus a better accuracy can be achieved. The results in Fig. 5 and Fig. 6 confirm the aforementioned analysis, where D-CT denotes the performance using the deep learning optimized fusion feature of color autocorrelogram and GLCM features. Fig. 5 illustrates that all methods achieve better activity recognition performance in the locations 1, 2,

7, 8, and 9 in the 1st experiment, and 4, 6, 7, and 9 in the 2nd experiment. By analysing the deployment of the experiment in Fig. 4, we discover that all these locations are near to the transmitter or receiver, and close to the wireless link that connects the transmitter and receiver. Fig. 6 shows the averaged activity recognition accuracy of different activities in the 12 locations. We can see that activities with larger movement range, such as walk and bow, could achieve better performance than activities with smaller movement range, such as arm up and down and arm left and right.

The proposed DFLAR system utilizes CSI as observation information, which could provide amplitude and phase measurements. To analyse the contribution of different type of measurements, we evaluate the system performance when using amplitude measurements, phase measurements, and both the amplitude and phase measurements. Fig. 7 reveals that the system could achieve reasonable localization and activity recognition accuracy even using only one type of measurement, the system performance will be remarkably better if using both the amplitude and phase measurements. The receiver could acquire CSI from 3 antennas, if we use measurements from different number of antennas, the performance must be different. Intuitively, if we use more number of antennas, we

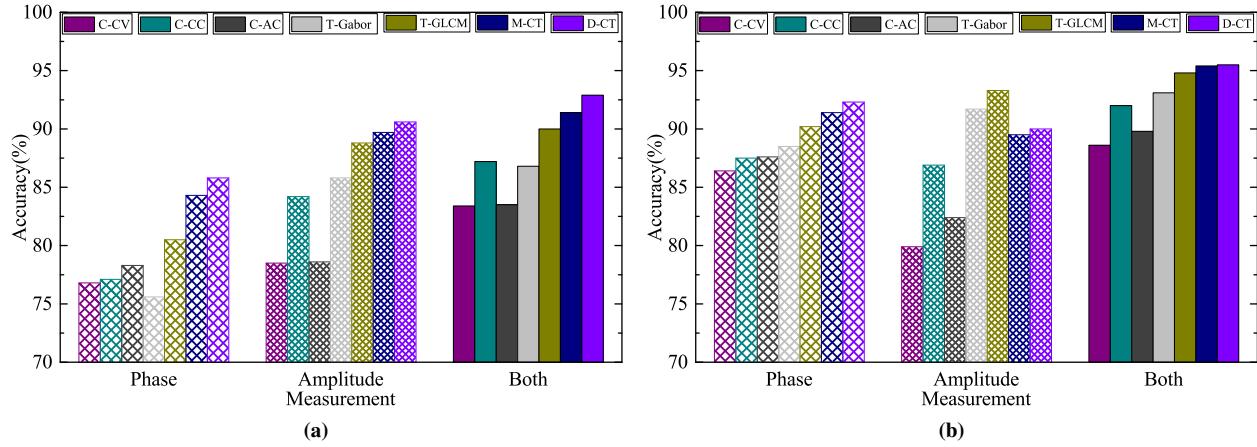


Fig. 7 Accuracy using different signals. (a) 1st experiment. (b) 2nd experiment.

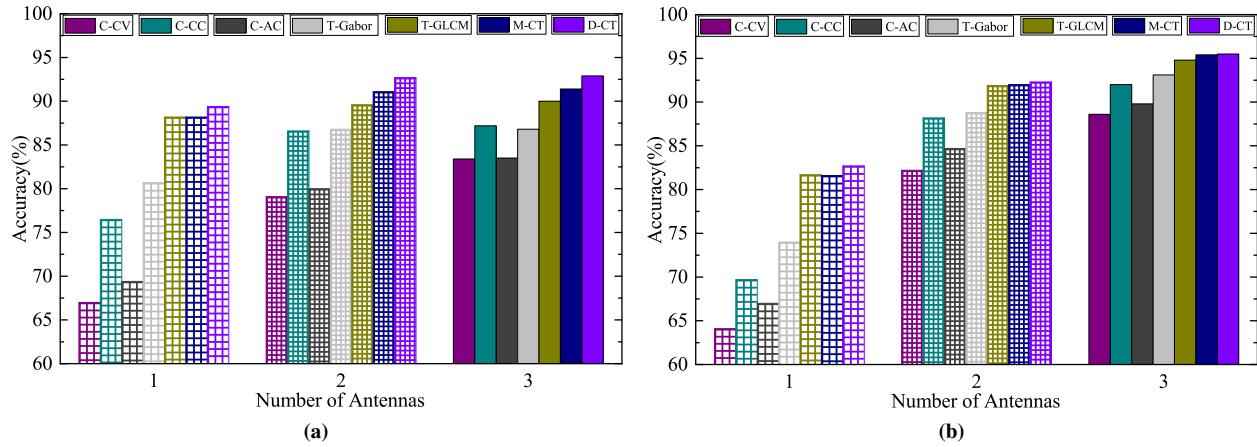


Fig. 8 Accuracy using different number of antennas. (a) 1st experiment. (b) 2nd experiment.

will acquire more observation information, and thus the localization and activity recognition performance will be better. The results in Fig. 8 confirm our analysis. With the mature of massive MIMO technique, wireless terminals will be equipped with more and more number of antennas. We can leverage multiple antennas to improve the performance of DFLAR systems.

D. Discussion

It should be mentioned that the proposed method of extracting radio image features is a general solution for extracting discriminative features. In this paper, we adopt it to solve the DFLAR problem. It is suitable for solving other problems, such as DFL as well. Taking the experimental datasets in the aforementioned experiments as examples, the performance of DFL systems using different features is summarized in Table III. From the result, we can see that the proposed M-CT scheme achieves better performance than other schemes. Meanwhile, since DFL needs only to estimate location information, thus the performance of DFL is better than DFLAR systems remarkably.

To evaluate the robustness of the proposed features when using different classifiers, we feed the extracted deep features

TABLE III Performance of DFL systems using different fusion features.

Experiment	Features	M-TF	Wavelet	M-CT
1st	Accuracy(%)	93.5	93.6	95.5
	DP-Accuracy(%)	94.0	94.2	96.2
2nd	Accuracy(%)	93.8	94.0	97.6
	DP-Accuracy(%)	94.5	94.8	98.2

into three different classifiers [18, 37, 38], i.e., Softmax, SVM, and K-binary classifiers, and evaluate the performance. For the M-CT scheme using Softmax classifier, the accuracy is 92.9% and 95.5% for the 1st and 2nd experiments. If SVM classifier is utilized, the performance will be 92.3% and 95.2%. If K-binary classifier is adopted, the performance will be 92.4% and 95.3%, respectively. We can see that all the classifiers could achieve reasonable performance. Meanwhile, due to the excellent multi-class classification ability, the performance of Softmax is a little better than other two schemes.

To evaluate the computational cost of the proposed scheme, we measure the average running time required to achieve a location and activity estimation for DFLAR systems using different features. We run different algorithms on a 2.4GHz PC with 8GBytes memory. The running times of DFLAR

TABLE IV Running times of DFLAR systems using different features.

Features	T-Mean	T-Var	T-Peak	F-Energy	F-Entropy	C-CV	C-CC	C-AC	T-Gabor	T-GLCM	M-TF	Wavelet	M-CT
Times(ms)	0.4	1	1.3	3.5	3.5	11	33	15	21	23	11	11.5	41

systems using different features are summarized in Table IV. It should be noted that although the running time of the proposed scheme is larger than other schemes, it is still acceptable for DFLAR systems which generally estimate user location and activity in the scale of seconds.

VI. CONCLUSION

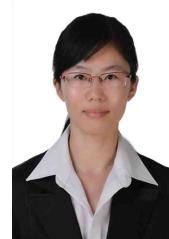
In this paper, we have developed a deep learning based image processing framework to characterize the influence of human behaviors on surrounding WiFi signals. Due to the abundant color and texture feature information involved in the radio image, and the feature optimization ability of the deep learning network, the proposed scheme can provide discriminative information for the classifier, thus, improve the performance of the device-free localization and activity recognition system. Extensive experiments confirm the excellent performance of the proposed schemes. The designed CSI-based DFLAR system could achieve an accuracy of more than 90% when using only one pair of transmitter and receiver in clutter laboratories with a size of 10.8m×7.8m.

It should be mentioned that the proposed deep learning based radio image processing framework is a general solution for extracting discriminative features from CSI measurements. It can be utilized by many other device free wireless sensing applications, such as emotion analysis, gait recognition, keystroke identification, among many others.

REFERENCES

- [1] Z. Zhou, C. Wu, Z. Yang, and Y. Liu, "Sensorless sensing with WiFi," *Tsinghua Science and Technology*, vol. 20, no. 1, pp. 1-6, Jan. 2015.
- [2] S. Savazzi, S. Sigg, M. Nicoli, V. Rampa, S. Kianoush, and U. Spagnolini, "Device-free radio vision for assisted living: Leveraging wireless channel quality information for human sensing," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 45-58, Mar. 2016.
- [3] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 53-53, Jan. 2011.
- [4] J. Wang, Q. Gao, H. Wang, Y. Yu, and M. Jin, "Time-of-flight-based radio tomography for device free localization," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2355-2365, May 2013.
- [5] J. Wang, Q. Gao, Y. Yu, X. Zhang, and X. Feng, "Time and energy efficient TOF-based device-free wireless localization," *IEEE Trans. Industrial Informatics*, vol. 12, no. 1, pp. 158-168, Jan. 2016.
- [6] J. Wang, H. Jiang, J. Xiong, K. Jamieson, X. Chen, D. Fang, and B. Xie, "LiFS: Low human-effort, device-free localization with fine-grained subcarrier information," in *Proc. 22th ACM Int. Conf. Mobile Computing Networking*, New York, USA, Oct. 2016, pp. 243-256.
- [7] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl, "RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals," *IEEE Trans. Mobile Computing*, vol. 13, no. 4, pp. 907-920, Apr. 2014.
- [8] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. 20th Int. Conf. Mobile Computing and Networking*, Hawaii, USA, Sep. 2014, pp. 617-628.
- [9] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21th Int. Conf. Mobile Computing and Networking*, Paris, France, Sep. 2015, pp. 65-76.
- [10] D. Zhang, K. Lu, R. Mao, Y. Feng, Y. Liu, Z. Ming, and L. M. Ni, "Fine-grained localization for multiple transceiver-free objects by using RF-based technologies," *IEEE Trans. Parallel and Distributed Systems*, vol. 25, no. 6, pp. 1464-1475, Jun. 2014.
- [11] Y. Zhao and N. Patwari, "Robust estimators for variance-based device-free localization and tracking," *IEEE Trans. on Mobile Computing*, vol. 14, no. 10, pp. 2116-2129, Oct. 2015.
- [12] J. Wang, D. Fang, Z. Yang, H. Jiang, X. Chen, T. Xing, and L. Cai, "E-HIPA: An energy-efficient framework for high-precision multi-target-adaptive device-free localization," *IEEE Trans. on Mobile Computing*, vol. 16, no. 3, pp. 716-729, Mar. 2017.
- [13] J. Wang, X. Chen, D. Fang, C. Wu, Z. Yang, and T. Xing, "Transferring compressive sensing based device-free localization across target diversity," *IEEE Trans. on Industrial Electronics*, vol. 62, no. 4, pp. 2397-2409, Apr. 2015.
- [14] Y. Guo, K. Huang, N. Jiang, X. Guo, Y. Li, and G. Wang, "An exponential-rayleigh model for RSS-based device-free localization and tracking," *IEEE Trans. Mobile Computing*, vol. 14, no. 3, pp. 484-494, Mar. 2015.
- [15] Z. Yang, K. Huang, X. Guo, and G. Wang, "A real-time device-free localization system using correlated RSS measurements," *EUROASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 186, pp. 1-12, Dec. 2013.
- [16] C. Liu, D. Fang, Z. Yang, H. Jiang, X. Chen, W. Wang, T. Xing, and L. Cai, "RSS distribution-based passive localization and its application in sensor networks," *IEEE Trans. on Wireless Communications*, vol. 15, no. 4, pp. 2883-2895, Apr. 2016.
- [17] Z. Wang, H. Liu, S. Xu, X. Bu, and J. An, "A diffraction measurement model and particle filter tracking method for RSS-based DFL," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2391-2403, Nov. 2015.
- [18] J. Hong and T. Ohtsuki, "Signal eigenvector-based device-free passive localization using array sensor," *IEEE Trans. Vehicular Technology*, vol. 64, no. 4, pp. 1354-1363, Apr. 2015.
- [19] 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 Trans. Vehicular Technology*, vol. 66, no. 2, pp. 1659-1669, Feb. 2017.
- [20] J. Wang, X. Zhang, Q. Gao, H. Yue, and H. Wang, "Device-free wireless localization and activity recognition: A deep learning approach," *IEEE Trans. Vehicular Technology*, vol. 66, no. 7, pp. 6258-6267, Jul. 2017.
- [21] J. Y. Chang, K. Y. Lee, K. C. Lin, and W. Hsu, "WiFi action recognition via vision-based methods," in *Proc. 41st IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Shanghai, China, Mar. 2016, pp. 2782-2786.
- [22] J. Y. Chang, K. Y. Lee, Y. L. Wei, K. C. Lin, and W. Hsu, "Location-independent WiFi action recognition via vision-based methods," in *Proc. ACM Multimedia 2016*, Amsterdam, Netherlands, Oct. 2016, pp. 162-166.
- [23] 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.
- [24] 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 Trans. on Mobile Computing*, vol. 16, no. 2, pp. 511-526, Feb. 2017.
- [25] H. Wang, D. Zhang, J. Ma, Y. Wang, Y. Wang, D. Wu, T. Gu, and B. Xie, "Human respiration detection with commodity wifi devices: do user location and body orientation matter?", in *Proc. ACM Int. Joint Conf. Pervasive and Ubiquitous Computing*, Heidelberg, Germany, Sep. 2016, pp. 25-36.
- [26] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proc. ACM Int. Joint Conf. Pervasive and Ubiquitous Computing*, Heidelberg, Germany, Sep. 2016, pp. 363-373.
- [27] Q. Zhou, J. Xing, J. Li, and Q. Yang, "A device-free number gesture recognition approach based on deep learning," in *Proc. IEEE Int. Conf. Computational Intelligence and Security*, Wuxi, China, Dec. 2016, pp. 1-7.

- [28] M. Alsheikh, S. Lin, D. Niyato, and H. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1996-2018, Fourth Quarter 2014.
- [29] D. Heckerman and C. Meek, "Models and selection criteria for regression and classification," in *Proc. 13th Int. Conf. Uncertainty in Artificial Intelligence*, San Francisco, USA, Aug. 1997, pp. 223-228.
- [30] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, "You are facing the mona lisa: Spot localization using PHY layer information," in *Proc. 10th ACM Int. Conf. Mobile Systems, Applications, and Services*, Ambleside, UK, Jun. 2012, pp. 183-196.
- [31] J. M. Patel and N. C. Gavit, "A review on feature extraction techniques in content based image retrieval," in *Proc. IEEE Int. Conf. Wireless Communications, Signal Processing and Networking*, Chennai, India, Mar. 2016, pp. 2259-2263.
- [32] X. Wang, X. Ding, and C. Liu, "Gabor filters-based feature extraction for character recognition," *Pattern Recognition*, vol. 38, no. 3, pp. 369-379, Mar. 2005.
- [33] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504-507, Jul. 2006.
- [34] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: a deep learning approach," *IEEE Trans. Vehicular Technology*, vol. 66, no. 1, pp. 763-776, Jan. 2017.
- [35] S. Kullback, "A lower bound for discrimination information in terms of variation," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 126C127, Jan. 1967.
- [36] H. Zhang, W. Wu, and M. Yao, "Boundedness and convergence of batch back-propagation algorithm with penalty for feedforward neural networks," *Neurocomputing*, vol. 89, pp. 141-146, Jul. 2012.
- [37] H. Wymeersch, S. Marano, W. M. Gifford, and M. Z. Win, "A machine learning approach to ranging error mitigation for UWB localization," *IEEE Trans. Communications*, vol. 60, no. 6, pp. 1719-1728, Jun. 2012.
- [38] V. Savic and E. G. Larsson, "Fingerprinting-based positioning in distributed massive MIMO systems," in *Proc. 82nd IEEE Int. Conf. Vehicular Technology*, Boston, USA, Sep. 2015, pp. 1-5.



Xiaorui Ma received her B.S. degree in Applied Mathematics from Lanzhou University, China, in 2008, and Ph.D degree in Electronic Engineering from Dalian University of Technology, Dalian, China, in 2017.

She is currently an Assistant Professor at Dalian University of Technology. Her research interests include machine learning, hyperspectral image classification, and computer vision.



Xueyan Feng received her B.S. degree in Electronic Engineering from Dalian University of Technology, China, in 2015. She has been working towards the M.S. degree in the Department of Electrical Engineering at Dalian University of Technology. Her research interests include device free wireless localization and recognition, wireless networks, and machine learning.



Qinghua Gao received the B.S., M.S., and Ph.D degrees in Electronic Engineering from Dalian University of technology in 2003, 2007, and 2012, respectively.

She is currently an Associate Professor at Dalian University of Technology. Her research interests are in the areas of wireless networks, including tracking, localization, and information processing.



Hongyu Wang received the B.S. degrees from Jilin University of Technology in 1990, and M.S. degrees from Graduate School of Chinese Academy of Sciences in 1993, both in Electronic Engineering. He received the Ph.D. in Precision Instrument and Optoelectronics Engineering from Tianjin University in 1997.

He is currently a Professor at Dalian University of Technology. His research interests include algorithmic, optimization, and performance issues in wireless ad hoc, mesh and sensor networks.



Jie Wang (M'12) received his B.S. degree from Dalian University of Technology, Dalian, China, in 2003, M.S. degree from Beihang University, Beijing, China, in 2006, and Ph.D degree from Dalian University of Technology, Dalian, China, in 2011, all in Electronic Engineering.

He is currently an Associate Professor at Dalian University of Technology. From January 2013 to January 2014, he was a visiting researcher with University of Florida. His research interests include wireless localization and tracking, radio tomography, wireless sensing, wireless sensor networks, and cognitive radio networks.

Dr. Wang served as Technical Program Committee member for many international conferences, e.g., IEEE GLOBECOM and ICC. He serves as an Associate Editor for IEEE Transactions on Vehicular Technology from January 2017. He served as a Guest Editor for International Journal of Distributed Sensor Networks for a special issue on wireless localization in 2015.