Wireless Networks (2020) 26:3495–3506
https://doi.org/10.1007/s11276-020-02274-7

Device-free crowd counting with WiFi channel state information and deep neural networks

Rui Zhou[1] · Xiang Lu[1] · Yang Fu[1] · Mingjie Tang[1]

Published online: 14 February 2020

Abstract
Crowd counting is of great importance to many applications. Conventional vision-based approaches require line of sight and pose privacy concerns, while most radio-based approaches involve high deployment cost. In this paper, we propose to utilize WiFi channel state information (CSI) to infer crowd count in a device-free way, with only one pair of WiFi transmitter and receiver. The proposed method establishes the statistical relationship between the variation of CSI and the number of people with deep neural networks (DNN) and thereafter estimates the people count according to the real-time CSI through the trained DNN model. Evaluations demonstrate the effectiveness of the method. For the crowd size of 6, the counting error was within 1 person for 100% of the cases. For the crowd size of 34, the counting error was within 1 person for 97.7% of the cases and within 2 persons for 99.3% of the cases.

Keywords Crowd counting · Channel state information · Device-free · Deep neural networks

1 Introduction

Crowd counting is of great importance to a number of applications in various scenarios, such as retail stores, shopping plazas, visitor centers, libraries and museums. It can help a retailer determine the percentage of visitors who actually make purchases and help optimize the usage of staff resources. It can be used to ensure the safe level of occupancy and give proper emergence response in public areas. Heating and cooling systems can be optimized automatically according to the level of occupancy for the purpose of energy saving. Different applications may require different counting accuracy. The applications in a small area with a small number of people, e.g. in a small retail store or a small museum room, require high accuracy, for which the counting error should not exceed one person. For the applications in a medium area with a medium number of people, e.g. in a visitor center or a library hall, errors around 2 persons are acceptable. For the applications with a large number of people, a rough estimation or density estimation is usually adequate.

Approaches of crowd counting are mainly classified into vision-based and radio-based. Vision-based crowd counting [1, 2] has been widely deployed in many places. However, cameras do not work well under dim lights and require line of sight (LOS), leading to degradation of monitoring quality or blind areas. More seriously, cameras pose privacy concerns. Radio-based solutions have attracted considerable attention in the recent years, either device-based or device-free. Device-based approaches [3, 4] require people to carry mobile devices for surveillance, thus may not be feasible in certain circumstances. Several device-free approaches have been proposed to tackle the problem, utilizing wireless sensor networks [5, 6], radio signal strength [7], ultra wideband [8], or pulsed radar [9]. However, they all require to set up dedicated infrastructure for surveillance. The high cost hinders their wide deployment.

As wireless local area networks are almost pervasive nowadays, device-free monitoring based on WiFi has emerged as a research focus in the recent years [10]. Approaches utilizing received signal strength (RSS) have been proposed to tackle the problem of crowd counting [11, 12]. However, complexity of indoor environments

✉ Rui Zhou
ruizhou@uestc.edu.cn

[1] School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China

and multipath fading may cause instability and declination of monitoring. Channel state information (CSI) is a fine-grained measurement from the physical layer, which describes the amplitude and phase of each orthogonal subcarrier in a channel. Researches [13, 14] show that CSI is sensitive to environmental variations and provides the capability to benefit from the multipath effect, thus suitable for accurate crowd counting. A few pioneer researches have been conducted in the field of crowd counting [10, 15–17] using WiFi CSI with different configurations and methods.

In this paper, we propose to treat device-free crowd counting as a regression problem and employ deep neural networks (DNN) to model the relationship between WiFi CSI variations and crowd counts. We utilize CSI amplitude to infer CSI variations through matrix dilation and acquire the percentage of non-zero elements (PEM) of each subcarrier as the features. The DNN model is trained with the CSI variations labelled with the corresponding crowd counts. For crowd counting, the input to the trained DNN model is the CSI variation and the output is the people count. Evaluations in two representative scenarios achieved the mean counting error of 0.11 and 0.14 person, respectively, outperforming the state of the art. The main contributions of the paper are summarized as follows:

1. The solution is based on commodity WiFi devices, requiring only one pair of transmitter and receiver;
2. Crowd counting is conducted in a device-free way without involvement of mobile devices;
3. Apply DNN regression to establish the relationship between CSI variations and people counts, thus to infer the crowd count from the real-time CSI variation through the established DNN regression model;
4. Evaluations were conducted in two different scenarios with different numbers of people;
5. Extensive experiments and comparisons were conducted demonstrating high performance of the proposed method.

The rest of the paper is organized as follows. Section 2 reviews the state of the art briefly. Section 3 presents the preliminaries of CSI and the rationales behind crowd counting. Section 4 proposes the method of device-free crowd counting with CSI and DNN. Evaluations are reported in Sect. 5. Section 6 concludes the paper.

2 Related work

There have been some researches on crowd counting with radio-based solutions. We focus on crowd counting using WiFi and review the state of the art briefly.

Nakatsuka et al. [11] proposed a crowd counting system using received signal strength indication (RSSI) of WiFi networks, by linear regression. The evaluation used a pair of transmitter and receiver and achieved the average counting error of 1.5 person.

Depatla et al. [12] proposed counting the number of people walking in an area using WiFi RSSI between a pair of transmitter and receiver. By characterizing the impact of the crowd on blocking the LOS, and the impact of the total number of people on the scattering effects, they developed a mathematical expression for the probability distribution of RSSI as a function of the people count, which was the base for the estimation using Kullback–Leibler divergence. Experiments showed an error of 2 or less 96% and 63% of the outdoor and indoor cases, when using omnidirectional antennas, and an error of 2 or less 100% of the cases when using directional antennas.

Xi et al. [15] presented a device-free crowd counting approach based on WiFi CSI in a space, called frog eye. They proposed a metric PEM, the percentage of non-zero elements in the dilated CSI matrix, representing the variation of wireless channels and formulated the monotonic relationship between PEM and people count by the grey Verhulst model (GVM). They conducted experiments both indoors and outdoors using one WiFi transmitter and three WiFi receivers. More than 98% errors are less than 2 persons indoors and 70% outdoors.

Di Domenico et al. [16] focused on crowd counting in a room using CSI, which did not require dedicated training in new environments, by identifying a set of differential CSI feature candidates and selecting the most effective ones via minimization of the summation of the Davies–Bouldin indexes. They assessed the proposed approach by training once for all the system in a room, and testing the system in two different rooms. The results show that more than 91% and 81% errors are less than or equal to 2 persons in the two different rooms. Di Domenico et al. [17] presented another crowd counting system without requiring dedicated training in new environments. The proposed approach analyzed the shape of the Doppler spectrum of the received WiFi signal which was correlated to the number of people moving in the monitored environment. Experiments showed that the errors were less than or equal to 2 persons for 99% and 92% of the cases in two different rooms.

Cianca et al. [10] proposed a crowd counting system, based on the received raw signal samples (RRSS) extracted from beacon messages, using a pair of WiFi transmitter and receiver. With a linear discriminant classifier, the system achieved the classification accuracy of 72% for office room and 69% for the meeting room, up to 5 persons.

3 Preliminaries and rationales

3.1 Preliminaries of CSI

The infrastructure of device-free passive crowd counting is composed of a wireless access point (AP) for data transmission, a monitoring point (MP) for data retrieval, and a server for data processing. The AP can be a commodity wireless router supporting 802.11n, i.e. supporting orthogonal frequency division multiplexing (OFDM) and multiple-input multiple-output (MIMO). The MP is a commodity wireless adapter supporting 802.11n, with modified firmware and driver to retrieve CSI [18]. For crowd counting, the AP–MP pair is placed at corners or edges, for complete coverage of the space, as shown in Fig. 1.

CSI is fine-grained information from the physical layer that describes channel frequency response (CFR) from the transmitter to the receiver. Leveraging commodity network interface card (NIC) with modified firmware and driver, the amplitude and phase of each subcarrier within a channel can be revealed to the upper layers for each packet in the format of CSI [18]. The raw data contain the number of transmitting antennas N_{tx}, the number of receiving antennas N_{rx}, and the CSI matrix H, which is a $N_{tx} \times N_{rx}$ matrix:

$$H = \left(H_{ij}\right)_{N_{tx} \times N_{rx}} \tag{1}$$

H_{ij} is the CSI of the channel formed by transmitting antenna i and receiving antenna j, containing the information of N_s subcarriers, expressed as

$$H_{ij} = (h_1, h_2, \ldots, h_{N_s})^T,\ i \in [1, N_{tx}],\ j \in [1, N_{rx}] \tag{2}$$

Subcarrier k in H_{ij} can be expressed as

$$h_k = |h_k| e^{j\angle h_k},\ k \in [1, N_s] \tag{3}$$

$|h_k|$ is the amplitude and $\angle h_k$ is the phase.

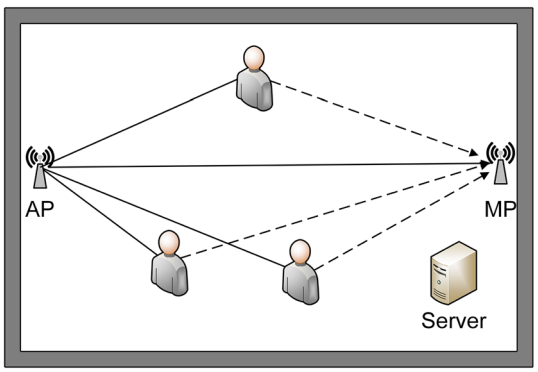


Fig. 1 Scenario and rationale of crowd counting

3.2 Rational of crowd counting

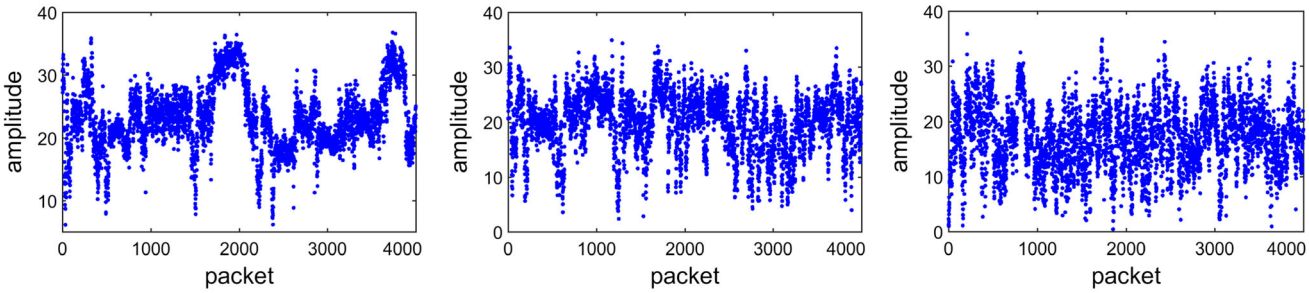
The scenario of crowd counting in a space is shown in Fig. 1. CSI portrays a fine-grained temporal and spectral structure of wireless channels. Separate subcarriers experience different multipath fading, thus when small movements have altered the environment, the individual subcarrier measurements are very likely to change. The rationale behind crowd counting is that CSI exhibits different patterns when different number of people are in the space, as presence of people and movements of them influence signal propagation and each person may block or reflect the signals in certain ways. Figure 2 illustrates the properties of CSI of one subcarrier over time, in which the packet-axis represents the packets over time, and the amplitude-axis represents the CSI amplitude of the subcarrier. Figure 2 illustrates the amplitude of one subcarrier over time/packets when 1, 3, 5 persons are walking casually in a room. Figure 3 illustrates the amplitudes of 30 subcarriers over time when 1, 3, 5 persons are walking casually in a room. It can be seen that when no person is present in the room, the CSI amplitudes are quite stable. When more people appear in the room, the CSI amplitudes vary more drastically. Thus we can infer the count of people according to CSI amplitudes and their variations.

4 Methodology of crowd counting

We propose a device-free crowd counting method using deep neural networks and WiFi CSI, aiming to count the number of people in a space. The overview of the method is illustrated in Fig. 4, which is composed of CSI data collection, CSI variation acquisition, DNN training, and crowd counting. From the collected CSI data, the percentage of non-zero elements PEM of each subcarrier is calculated to form the features, which indicate CSI variations due to different number of people. During training, the relationship between the features, i.e. PEMs of all the subcarriers, and the crowd counts is established through DNN, which will be used afterwards to infer the crowd count according to real-time CSI features, i.e. CSI variations.

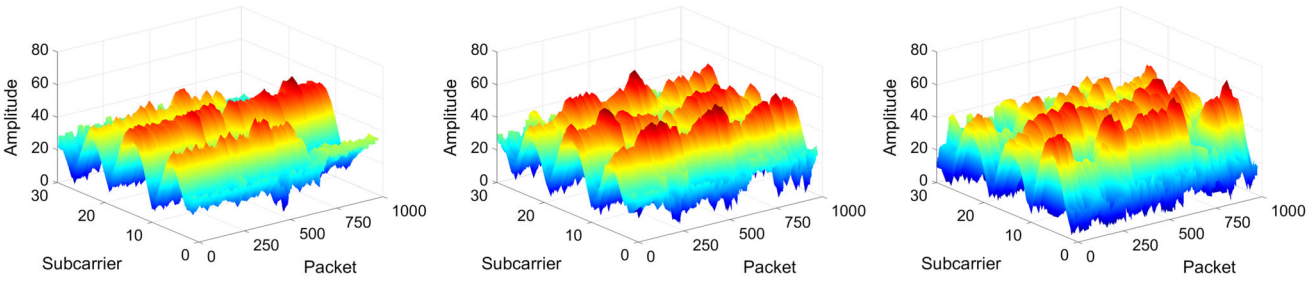
4.1 Acquisition of CSI variation

We use PEM, calculated from the dilated CSI matrix, proposed in [15], to indicate the CSI variation caused by different number of people. Assume M_0 represents a matrix with M rows and P columns, where P denotes the number of packets used to calculate PEM and M represents the



(a) 1 person **(b)** 3 persons **(c)** 5 persons

Fig. 2 CSI amplitude of one subcarrier with different numbers of people



(a) 1 person **(b)** 3 persons **(c)** 5 persons

Fig. 3 CSI amplitude of 30 subcarriers with different numbers of people

matrix resolution. Assume C represents the CSI amplitude matrix, with C_{ij} representing the amplitude of subcarrier i and packet j, C_{min} and C_{max} representing the minimal and maximal amplitude values in C. To calculate the PEM of subcarrier i, we first generate the two-dimensional matrix M_0 from the CSI matrix C. Firstly initialize each element in matrix M_0 to 0. Secondly calculate k as

$$k = \left\lfloor \frac{C_{ij} - C_{min}}{C_{max} - C_{min}} (M - 1) \right\rfloor + 1 \tag{4}$$

where M represents the number of rows in matrix M_0 indicating the matrix resolution. Thirdly set the element in row k and column j of matrix M_0 to 1, i.e. $M_0[k][j] = 1$. Thus in matrix M_0, there is one "1" in each column and the rests are "0"s. Then dilate matrix M_0 according to the dilation coefficient I and set the elements in M_0 around "1" within a distance of I to "1". After dilation we obtain the dilated matrix M_c. Count the number of non-zero elements in matrix M_c and obtain the PEM of subcarrier i, denoted as q_i. Repeating the previous steps for all the subcarriers produces the PEM vector $(q_1, q_2, \ldots, q_L)$, which is regarded as the features for DNN model training and crowd counting.

Figure 5(a) shows the PEM values of the subcarriers when 0, 1, 2, 3, 4, 5 persons walking casually in a room. Each curve corresponds to a crowd count. It can be seen from the figure that there is certain dependency between crowd counts and PEM vectors. Figure 5(b) shows the PEM values of a few subcarriers over time with the same number of people walking casually in the room. Each subcarrier fluctuates around a fixed value, thus is relatively stable over time. Figure 5(c) shows the PEM vectors at different times with the same number of people. Each curve represents a PEM vector at a different time. The curves aggregate meaning that they are similar. Figure 5(b, c) demonstrate that the PEM values are relatively stable over time if the number of people does not change. The PEM values capture the CSI variations caused by different number of people. Therefore, it is possible to infer the crowd count from the PEM vector. By analyzing the impact of the parameters: packet number P, matrix resolution M and dilation coefficient I, their values are chosen as $P = 50$, $M = 50$, and $I = 20$, for the two evaluation scenarios.

4.2 Crowd counting through DNN

We treat crowd counting as a regression problem and propose to apply DNN, which can model arbitrary relationships, to model the relationship between crowd counts and PEM vectors, thus able to count the people according to the real-time PEM values through the trained DNN model. For regression each training sample contains one target value and several features. Regression is to establish

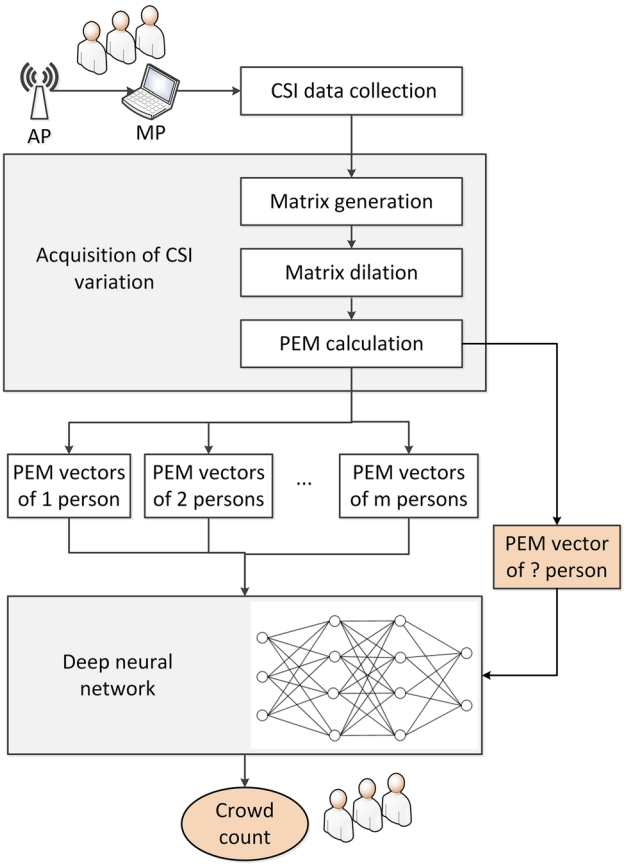


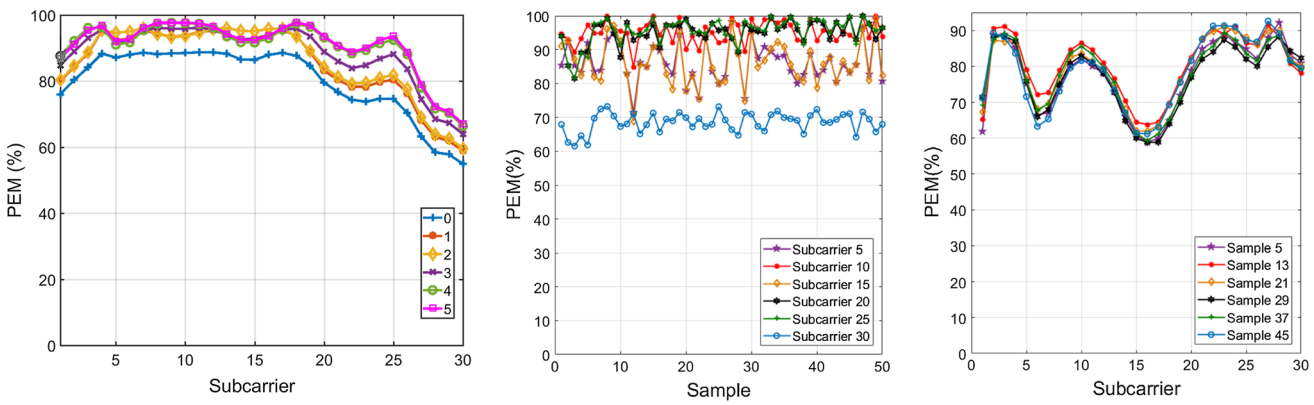
Fig. 4 The method of crowd counting

the functional dependency between the features and the target values based on the training samples, and thus able to determine the target values of the testing samples according to their features. For the problem of crowd counting, the target values are people counts, denoted as c, and the features are PEM vectors in the form of $(q_1, q_2, \ldots, q_L)$, in which q_i represents the PEM value of subcarrier i and L is the number of subcarriers, equal to $N_{tx} \times N_{rx} \times N_s$. Assume n is the number of training samples. Each training sample consists of a pair (r_i, c_i), where $r_i = (q_1, q_2, \ldots, q_L) \in R^L$ represents a PEM vector, and $c_i \in R$ denotes the crowd count. DNN model training is to establish the functional dependency $f : r \to c$ between PEM vectors and people counts using the dataset $\{(r_i, c_i)|i = 1, 2, \ldots, n\}$. Assume (r, c) is a testing sample, with $r \in R^L$ representing the real-time PEM vector, crowd counting is to determine the value of c with $f(r)$, expressed by the established DNN model.

4.2.1 Structure of DNN

We employ a DNN with N fully connected hidden layers with K_i $(i = 1, 2, \ldots, N)$ neurons on layer i to fulfill the regression task, as shown in Fig. 6. During model training, the CSI PEM vectors with corresponding crowd counts are fed into the network as a batch, and the weight of each neuron on each layer is trained, thus the DNN model is established. During crowd counting, the real-time CSI PEM vector is fed into the trained neural network, and the output is the crowd count c.

Assume $a_i^{(l)}$ represents the output of the neuron i on the hidden layer l, $w_{ij}^{(l)}$ represents the weight between the neuron j on the layer $l-1$ and the neuron i on the layer l and $b_i^{(l)}$ represents the bias of the neuron i on the layer l. The calculation is as follows: For the first hidden layer:



(a) Crowd counts and PEM vectors. Each curve corresponds to a crowd count.

(b) PEM values over time. Each curve corresponds to a subcarrier.

(c) PEM vectors at different times. Each curve corresponds to a different time.

Fig. 5 The PEM values

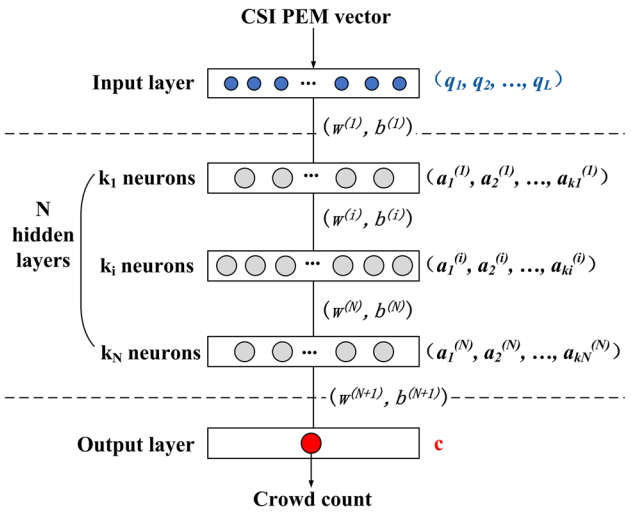


Fig. 6 Structure of DNN for crowd counting

$$\begin{aligned} a_1^1 &= f(w_{11}^1 q_1 + w_{12}^1 q_2 + \cdots + w_{1L}^1 q_L) + b_1^1 \\ a_2^1 &= f(w_{21}^1 q_1 + w_{22}^1 q_2 + \cdots + w_{2L}^1 q_L) + b_2^1 \\ &\vdots \\ a_{K_1}^1 &= f(w_{K_1 1}^1 q_1 + w_{K_1 2}^1 q_2 + \cdots + w_{K_1 L}^1 q_L) + b_{K_1}^1 \end{aligned} \tag{5}$$

For the other hidden layers, e.g. layer i:

$$\begin{aligned} a_1^i &= f(w_{11}^i a_1^{i-1} + w_{12}^i a_2^{i-1} + \cdots + w_{1L}^i a_L^{i-1}) + b_1^i \\ a_2^i &= f(w_{21}^i a_1^{i-1} + w_{22}^1 a_2^{i-1} + \cdots + w_{2L}^i a_L^{i-1}) + b_2^i \\ &\vdots \\ a_{K_i}^i &= f(w_{K_i 1}^i a_1^{i-1} + w_{K_i 2}^i a_2^{i-1} + \cdots + w_{K_i L}^i a_L^{i-1}) + b_{K_i}^i \end{aligned} \tag{6}$$

For the output layer:

$$c = f(w_{11}^{N+1} a_1^N + w_{12}^{N+1} a_2^N + \cdots + w_{1L}^{N+1} a_L^N) + b_1^{N+1} \tag{7}$$

4.2.2 Loss function

The loss function is employed to measure the difference between the true crowd count and the output of DNN, which is defined as

$$f_{loss} = \frac{1}{n_b} \sum_{i=1}^{n_b} |\hat{c}_i - c_i| \tag{8}$$

in which, n_b is the number of training samples in a batch, c_i is the true crowd count and $\hat{c}_i$ is the estimated crowd count. By minimizing the value of the loss function f_{loss} with the back propagation (BP) algorithm, the DNN weights are updated with adaptive moment estimation (Adam) algorithm, until the value of f_{loss} converges. The learning rate is set as 10^{-4}.

4.2.3 Activation function

The activation function introduces nonlinearity into the neural networks and is an important factor for performance. We choose rectified linear units (ReLU) as the activation function, which can be expressed as:

$$f(x) = max(0, x) \tag{9}$$

5 Evaluations

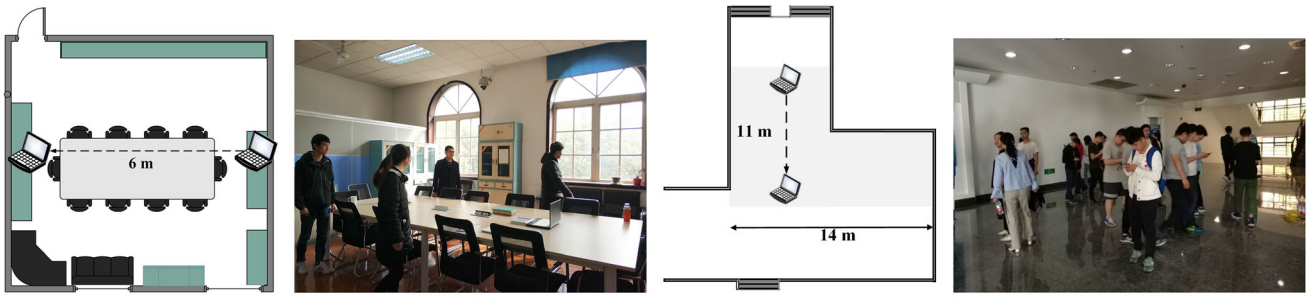
5.1 Experimental setup

We conducted evaluations of crowd counting in two representative scenarios in our university. The first testbed was in a rectangle meeting room with up to 5 people, and the second testbed was in an irregular-shaped hall in a teaching building with up to 34 people, as illustrated in Fig. 7. In each testbed, two laptops equipped with intel wireless link (IWL) 5300 NIC formed a pair of transmitter and receiver, each having 3 antennas, working in 5G band, at a height of 1.2 m.

During collection of training data, the testers walked casually in the WiFi covered area, without control on their movements. The raw CSI data were collected for each crowd number for a few minutes, with the sampling rate of 20 Hz, which were split to raw samples. Each raw sample consisted of 50 CSI packets. The raw CSI samples went through feature extraction to obtain PEM values, generating the training samples. Collection of the testing samples followed the same procedure. As listed in Table 1, in the meeting room we collected 30×6 (for 0–5 persons) training samples and 20×6 testing samples; in the hall we collected 60×26 (for 0–16, 18, 20, 22, 24, 26, 28, 30, 32, 34 persons) training samples and 40×26 testing samples. The time for data collection increased with the number of people, as each crowd number required samples. It was 1800 seconds for up to 5 people, and 7800 seconds for up to 34 people in the evaluation scenarios. The model training time increased with the number of people as well due to the increase of training samples. It was 300.5 seconds for up to 5 people and 631.3 seconds for up to 34 people.

5.2 DNN training

During the training of the DNN model, we analyzed the effect of various parameters on performance by experiment



(a) Room **(b)** Room **(c)** Hall **(d)** Hall

Fig. 7 The testbeds. **a**, **b** Rectangle meeting room with up to 5 people; **c**, **d** Irregular-shaped hall with up to 34 people

Table 1 Experimental configuration for the two testbeds

Configuration	Room	Hall
AP	IWL 5300	IWL 5300
MP	IWL 5300	IWL 5300
AP–MP distance	6 m	11 m
#TX	3	3
#RX	3	3
Frequency band	5 GHz	5 GHz
Packet frequency	20 Hz	20 Hz
Max people count	5	34
#Training samples	30×6	60×26
#Testing samples	20×6	40×26
Collection time	1800 s	7800 s
Training time	300.5 s	631.3 s

and identified an optimal set of parameters. Two metrics were used to analyze the crowd counting performance: (1) *accuracy* represents the mean counting error between the ground truth and the estimated people counts; (2) *precision* represents the cumulative distribution function (CDF) of the counting errors.

We used the training samples to train the DNN crowd counting models of the two testbeds. The training process stopped when the training error converged. The batch size was set as 100 and the learning rate was set as 10^{-4}. Figure 8 illustrates the training errors over training epochs for the two testbeds. For the testbed of the meeting room, as the red curve illustrates, the training error started at about 2 persons, as the circle on the left marks, and decreased quickly with the epochs. After about 400 epochs the decrease slowed down, and finally converged on 0.088 person training error after 22,000 epochs, as the circle on the right shows. For the testbed of the hall, as the blue curve illustrates, the training error started at about 9 persons, as the square on the left marks, and declined quickly with the epochs. After about 200 epochs the declination slowed down, and eventually converged on 0.065 person training error after 27,000 epochs, as the square on the right shows. After training, the chosen DNN model for the meeting room has 4 hidden layers with the neurons of [1000,500,100,10], while the DNN model for the hall has 6 hidden layers with the neurons of [1000,500,100,100,500,1000]. The parameters are listed in Table 2.

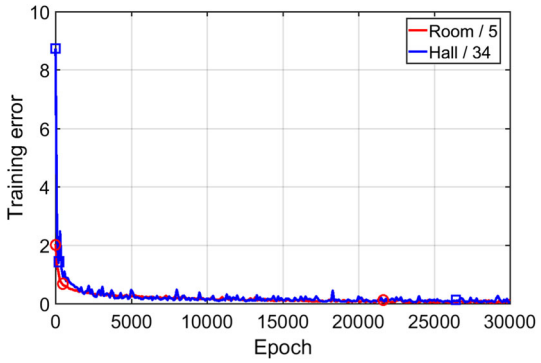


Fig. 8 Training errors over epochs

5.3 Results of crowd counting

The evaluation results of crowd counting in the two testbeds are shown in Table 3 and Fig. 9, in terms of mean crowd counting error $\overline{e}$ and cumulative distribution of crowd counting errors $P(e)$. As using regression to infer crowd counts, the results were often decimals, which were rounded to integers as the final people counts to calculate $\overline{e}$ and $P(e)$ in Table 3, while Fig. 9 illustrates CDF using the original regression results. The mean crowd counting error was 0.11 person in the meeting room and 0.14 person in the hall. In the meeting room, the correct counting was 89.2% of the cases, and the counting error was within 1 person 100%. In the hall, the correct counting was 90.6% of the cases, the counting error was within 1 person 97.7% and within 2 persons 99.3%. The proposed method achieved the precision of 100% within 1 error for a small crowd size,

Table 2 Model parameters for the two testbeds

Parameter	Room	Hall
#Layers of DNN	4	6
#Neurons of DNN	[1000,500,100,10]	[1000,500,100,100,500,1000]
Learning rate	10^{-4}	10^{-4}
Batch size	100	100
Optimizer	Adam	Adam
Activation function	ReLU	ReLU
P	50	50
M	50	50
I	20	20

Table 3 Results of crowd counting

Testbed	$\bar{e}$	$P(e=0)$ (%)	$P(e\leq 1)$ (%)	$P(e\leq 2)$ (%)
Room/5	0.11	89.2	100	100
Hall/34	0.14	90.6	97.7	99.3

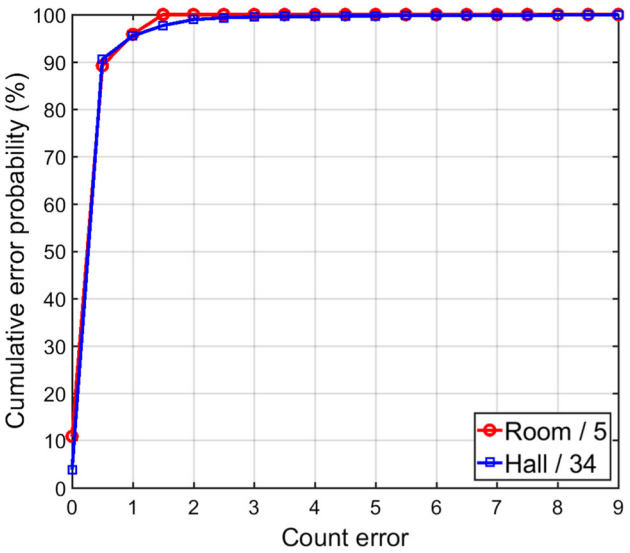


Fig. 9 Precision of crowd counting

and more than 97% within 1 error and more than 99% within 2 errors for a medium crowd size. The detailed confusion between estimated crowd counts versus ground truth is illustrated in Fig. 10.

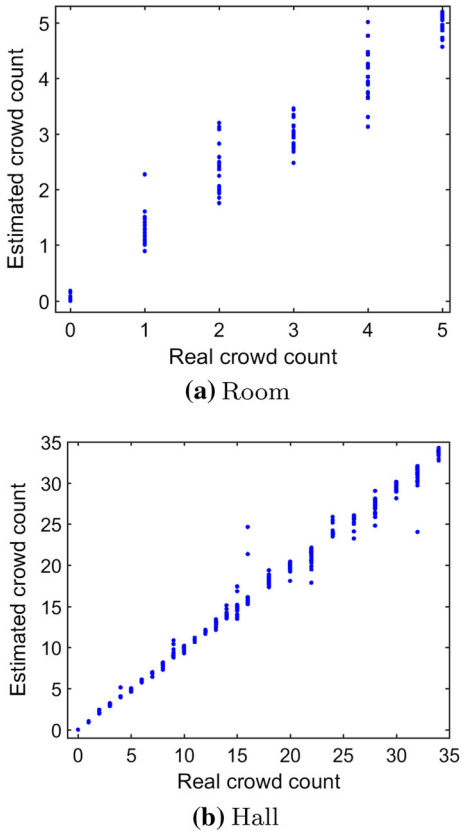


Fig. 10 Estimated crowd counts versus ground truth

5.4 Impact of parameters

5.4.1 Number of antennas

We analyzed the impact of the number of antennas. We tested one pair, two pairs and three pairs of antennas. Their evaluation accuracies in the meeting room with up to 5 people and in the hall with up to 34 people are listed in Table 4. One pair of antennas achieved the mean counting error of 0.36 person in the meeting room and 0.47 person in the hall, while two pairs improved the accuracy to (0.17, 0.14), and three pairs improved further to (0.11, 0.14). The evaluation precisions illustrated in Fig. 11(a) show that more antenna pairs achieved better performance.

5.4.2 Maximal crowd count

We evaluated the accuracy and precision of crowd counting with different maximal people counts. Figure 11(b) shows the precision when the maximal people

Table 4 Accuracy with different parameters and methods

Parameter	Value	$\overline{e}$ in room	$\overline{e}$ in hall
Antenna	1	0.36	0.47
	2	0.17	0.14
	3	0.11	0.14
Max. Count	5	–	0
	15	–	0.04
	24	–	0.07
	34	–	0.14
ML tool	DNN	0.11	0.14
	SVM	0.23	0.45
	KNN	0.28	0.16
	CART	0.32	0.20
	BN	0.37	0.17
Training	Set 1	–	1.39
	Set 2	–	1.19
Method	CSI	0.11	–
	Vision	0.20	–

counts were 5, 15, 24, 34 in the hall. With the increase of the maximal people count from 5 to 15 to 24 and to 34, the precision degraded and the mean counting error increased from 0 to 0.04 to 0.07 and to 0.14 person, as shown in Table 4. This result indicates that more people bring more confusion.

5.4.3 Layers and neurons of DNN

We tested different numbers of hidden layers N and different numbers of neurons $K = (K_1, K_2, \ldots, K_N)$ to evaluate the impact of DNN structure on crowd counting. We tested 3, 4, 5 hidden layers with different numbers of neurons in the meeting room, and tested 4, 5, 6 hidden layers with different numbers of neurons in the hall. The best accuracies achieved are shown in Table 5, the precisions are illustrated in Fig. 11(c) for the meeting room and Fig. 11(d) for the hall.

5.4.4 Training set

To demonstrate the generalization of the method, we trained the DNN model with some crowd counts and tested the trained model with the other crowd counts. The testbed in the hall had a larger crowd count, hence we used the dataset in the hall to conduct the evaluation. We trained the model with 0, 1, 2, 4, 6, 8, 10, 12, 14, 15, 16 people, and tested the model with 3, 5, 7, 9, 11, 13 people, denoted as set 1. We then trained the model with 0, 1, 2, 3, 5, 7, 9, 11, 13, 15, 16 people, and tested the model with 4, 6, 8, 10, 12, 14 people, denoted as set 2. The accuracies of set 1 and set 2 are shown in Table 4 and their precisions are illustrated in Fig. 11(g). Compared with using all the crowd counts, the accuracy and precision of using part of the crowd counts degraded. The reason was that part of the labels (crowd counts) were missing, which were important for accurate model training.

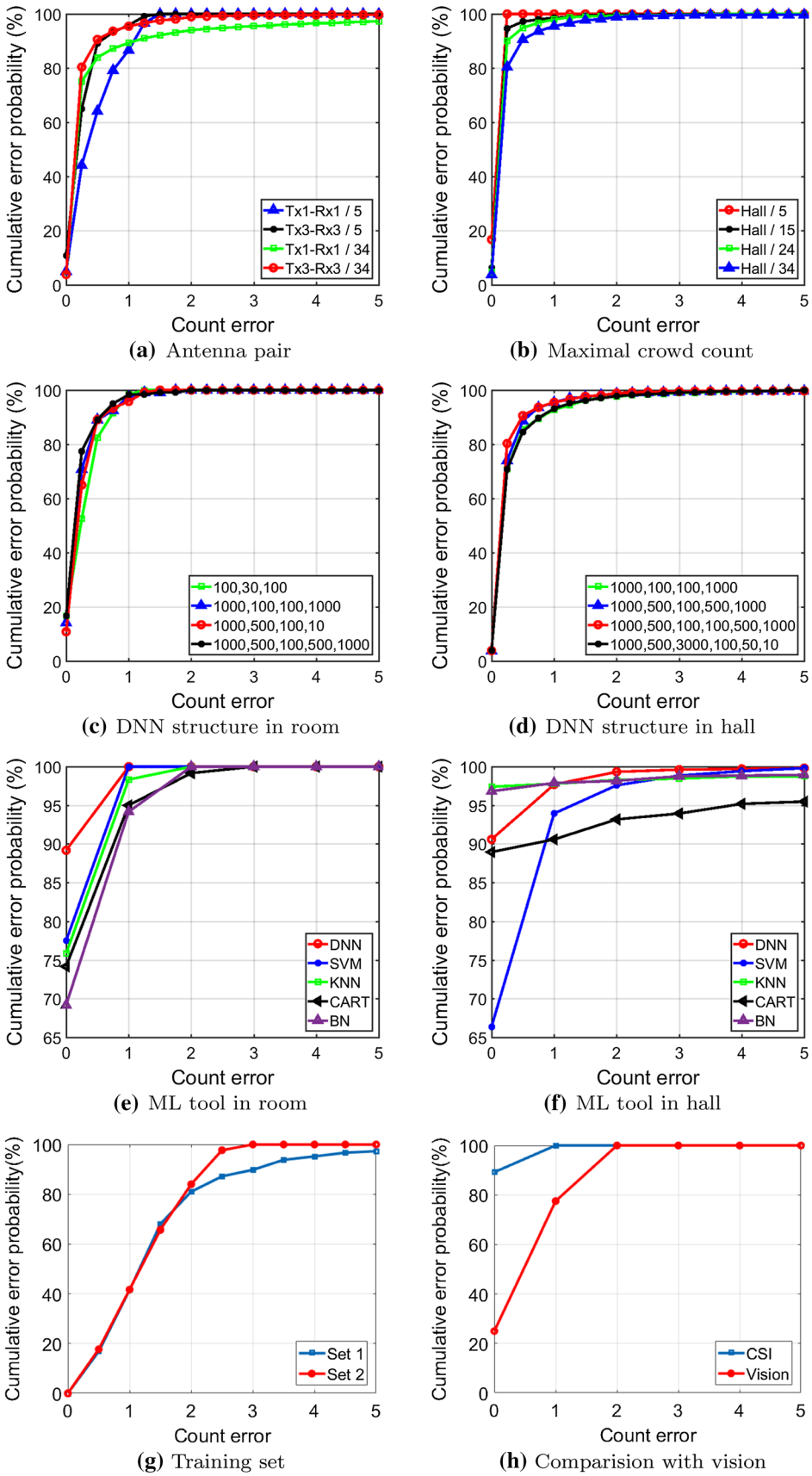
5.5 Comparison with other methods

5.5.1 Comparison with other machine learning tools

To prove the effectiveness of DNN on crowd counting, we compared it with several other machine learning (ML) tools: K-nearest neighbors (KNN), support vector machines (SVM), classification and regression tree (CART), and Bayesian network (BN). All the crowd counting results were rounded to integers for comparison. Their evaluation precisions in the meeting room and the hall are illustrated in Fig. 11(e, f), demonstrating that the proposed DNN method outperformed SVM, KNN, CART and NB. Table 4 shows that the proposed DNN method achieved the least mean counting error of 0.11 person in the meeting room and 0.14 person in the hall, outperforming SVM with (0.23, 0.45), KNN with (0.28, 0.16), CART with (0.32, 0.20) and BN with (0.37, 0.17).

5.5.2 Comparison with vision-based approaches

To demonstrate the effectiveness of crowd counting with WiFi CSI, we compared it with vision-based approaches. We conducted experiments with vision-based approaches using YOLO [19] and SSD [20] on 40 images collected under normal lights (20 images) and dim lights (20 images) in the same environment. Each image contained 5 people and some people were blocked by others from LOS. Under normal lights, out of 77 persons in LOS, 76 persons could be detected, and out of 23 persons in NLOS (blocked), only 5 persons could be detected. Under dim lights, out of 84 persons in LOS, 78 persons could be detected, and out of 16 persons in NLOS (blocked), only 4 persons could be detected. The experimental results show that the blocked people could not be detected for most of the cases, and the performance under dim lights degraded compared with under normal lights. This is a major disadvantage of vision-based approaches. The mean counting error of the 40 images was 0.20. The correct rate was 25%, error within 1 person was 77.5% and error within 2 persons was 100%. The evaluation accuracy and precision is illustrated in Fig. 11(h) and Table 4, demonstrating that the proposed WiFi CSI method outperformed vision-based approaches under NLOS and dim lights conditions.



(a) Antenna pair **(b)** Maximal crowd count

(c) DNN structure in room **(d)** DNN structure in hall

(e) ML tool in room **(f)** ML tool in hall

(g) Training set **(h)** Comparision with vision

Fig. 11 Precision of crowd counting with different parameters and methods

Table 5 Accuracy with different DNN layers and neurons

Testbed	#Layers	#Neurons	$\bar{e}$
Room	3	[100,30,100]	0.18
	4	[1000,100,100,1000]	0.12
	4	[1000,500,100,10]	0.11
	5	[1000,500,100,500,1000]	0.12
Hall	4	[1000,100,100,1000]	0.21
	5	[1000,500,100,500,1000]	0.16
	6	[1000,500,100,100,500,1000]	0.14
	6	[1000,500,300,100,50,10]	0.22

6 Conclusions

This paper proposes a device-free crowd counting method based on CSI with one pair of WiFi transmitter and receiver. The method first calculates PEM values of all the subcarriers as the features, and then establishes the DNN model representing the relationship between PEM values and crowd counts, thus able to count the crowd according to the real-time CSI variations. Evaluations in two different scenarios with different number of people show that the mean counting error was 0.11 person and 0.14 person and for more than 99% of the cases the counting errors were within 2 persons for a medium crowd size and 100% for a small crowd size. Such an accuracy can meet the requirements of most crowd count aware applications. When there are only a few people, e.g. in a small shop, higher accuracy is required and an error of 1 person is usually acceptable, whereas when there are many people, e.g. in a large hall, a rough estimation is adequate and an error of 2 persons is acceptable. There is no real limitation on the size of the room and the number of people, while the area is required to be covered with WiFi infrastructure. The major limitation of the current approach is that the model needs retraining if it is deployed in a new environment, which will be researched in our future work.

References

1. Li, M., Zhang, Z., Huang, K., & Tan, T. (2008). Estimating the number of people in crowded scenes by MID based foreground segmentation and head-shoulder detection. In *2008 19th international conference on pattern recognition* (pp. 1–4).
2. Kim, M., Kim, W., & Kim, C. (2011). Estimating the number of people in crowded scenes. *Proceedings of SPIE*, *7882*(23), 78 820L–78 820L-8.
3. Kannan, P. G., Venkatagiri, S. P., Chan, M. C., Ananda, A.L., & Peh, L.-S. (2012). Low cost crowd counting using audio tones. In *Proceedings of the 10th ACM conference on embedded network sensor systems (SenSys'12)* (pp. 155–168). ACM.
4. Weppner, J., & Lukowicz, P. (2013). Bluetooth based collaborative crowd density estimation with mobile phones. In *2013 IEEE international conference on pervasive computing and communications (PerCom)* (pp. 193–200).
5. Yuan, Y., Zhao, J., Qiu, C., & Xi, W. (2013). Estimating crowd density in an RF-based dynamic environment. *IEEE Sensors Journal*, *13*(10), 3837–3845.
6. Doong, S. H. (2016). Spectral human flow counting with RSSI in wireless sensor networks. In *2016 international conference on distributed computing in sensor systems (DCOSS)* (pp. 110–112).
7. Xu, C., Firner, B., Moore, R. S., Zhang, Y., Trappe, W., Howard, R., Zhang, F., & An, N. (2013). SCPL: Indoor device-free multi-subject counting and localization using radio signal strength. In *2013 ACM/IEEE IPSN* (pp. 79–90).
8. Lv, H., Liu, M., Jiao, T., Zhang, Y., Yu, X., Li, S., Jing, X., & Wang, J. (2013). Multi-target human sensing via UWB bio-radar based on multiple antennas. In *TENCON 2013* (pp. 1–4).
9. He, J., & Arora, A. (2014). A regression-based radar-mote system for people counting. In *2014 PerCom* (pp. 95–102).
10. Cianca, E., Sanctis, M. D., & Domenico, S. D. (2017). Radios as sensors. *IEEE Internet of Things Journal*, *4*(2), 363–373.
11. Nakatsuka, M., Iwatani, H., & Katto, J. (2008). A study on passive crowd density estimation using wireless sensors. In *2008 international conference on mobile computing and ubiquitous networking*
12. Depatla, S., Muralidharan, A., & Mostofi, Y. (2015). Occupancy estimation using only wifi power measurements. *IEEE Journal on Selected Areas in Communications*, *33*(7), 1381–1393.
13. Abdel-Nasser, H., Samir, R., Sabek, I., & Youssef, M. (2013). MonoPHY: Mono-stream-based device-free WLAN localization via physical layer information. In *2013 IEEE WCNC* (pp. 4546–4551).
14. Wu, K., Xiao, J., Yi, Y., Chen, D., Luo, X., & Ni, L. M. (2013). CSI-based indoor localization. *IEEE Transactions on Parallel and Distributed Systems*, *24*(7), 1300–1309.
15. Xi, W., Zhao, J., Li, X. -Y, Zhao, K., Tang, S., Liu, X., & Jiang, Z. (2014). Electronic frog eye: Counting crowd using WiFi. In *2014 IEEE INFOCOM* (pp. 361–369).
16. Di Domenico, S., De Sanctis, M., Cianca, E., & Bianchi, G. (2016). A trained-once crowd counting method using differential wifi channel state information. In *2016 WPA* (pp. 37–42). ACM.
17. Domenico, S. D.. Pecoraro, G., Cianca, E., & Sanctis, M. D. (2016). Trained-once device-free crowd counting and occupancy estimation using WiFi: A doppler spectrum based approach. In *2016 WiMob* (pp. 1–8).
18. Halperin, D., Hu, W., Sheth, A., & Wetherall, D. (2010). Predictable 802.11 packet delivery from wireless channel measurements. In: *2010 ACM SIGCOMM* (pp. 159–170). ACM.
19. Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv:1804.02767.
20. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2015). SSD: Single shot multibox detector. arXiv:1512.02325.

Rui Zhou received the Ph.D. degree from the University of Freiburg, Germany, in 2010. She is currently an associate professor with the University of Electronic Science and Technology of China. Her research interests include pervasive computing, Internet of Things, and artificial intelligence.

Xiang Lu received the M.S. degree from the University of Electronic Science and Technology of China in 2019. His research interests include pervasive computing and machine learning.

Yang Fu received the B.S. degree from the University of Electronic Science and Technology of China in 2019. His research interests include wireless sensing and artificial intelligence.

Mingjie Tang received the B.S. degree from the University of Electronic Science and Technology of China in 2019. His research interests include wireless sensing and artificial intelligence.