

No Need of Data Pre-processing: A General Framework for Radio-Based Device-Free Context Awareness

BO WEI*, Northumbria University, UK

KAI LI, Research Centre in Real-Time and Embedded Computing System (CISTER), Portugal

CHENGWEN LUO, Shenzhen University, China

WEITAO XU, City University of Hong Kong, China

JIN ZHANG, Shenzhen University, China

KUAN ZHANG, University of Nebraska-Lincoln, USA

Device-free context awareness is important to many applications. There are two broadly used approaches for device-free context awareness, i.e. video-based and radio-based. Video-based approaches can deliver good performance, but privacy is a serious concern. Radio-based context awareness applications have drawn researchers' attention instead because it does not violate privacy and radio signal can penetrate obstacles. The existing works design explicit methods for each radio based application. Furthermore, they use one additional step to extract features before conducting classification and exploit deep learning as a classification tool. Although this feature extraction step helps explore patterns of raw signals, it generates unnecessary noise and information loss. The use of raw CSI signal without initial data processing was, however, considered as no usable patterns. In this paper, we are the first to propose an innovative deep learning based general framework for both signal processing and classification. The key novelty of this paper is that the framework can be generalised for all the radio-based context awareness applications with the use of raw CSI. We also eliminate the extra work to extract features from raw radio signals. We conduct extensive evaluations to show the superior performance of our proposed method and its generalisation.

CCS Concepts: • **Networks** → **Sensor networks**; • **Computer systems organization** → **Sensor networks**;

Additional Key Words and Phrases: Device-free, channel state information, deep learning, context awareness

ACM Reference Format:

Bo Wei, Kai Li, Chengwen Luo, Weitao Xu, Jin Zhang, and Kuan Zhang. 2019. No Need of Data Pre-processing: A General Framework for Radio-Based Device-Free Context Awareness. *ACM Trans. Internet Things* 1, 1, Article 1 (January 2019), 27 pages. <https://doi.org/10.1145/3338026>

1 INTRODUCTION

Context awareness, e.g. activity recognition, gesture recognition, etc., is an essential component in many fields, such as health monitoring [29], emergency response [31], etc. Many approaches based on cameras and wearable sensors have been designed [7, 15, 35]. Camera-based solutions

*Corresponding Author

Authors' addresses: Bo Wei, Northumbria University, Newcastle upon Tyne, Tyne and Wear, UK, bo.wei@northumbria.ac.uk; Kai Li, Research Centre in Real-Time and Embedded Computing System (CISTER), Porto, Portugal, kaili@isep.ipp.pt; Chengwen Luo, Shenzhen University, Shenzhen, China, chengwen@szu.edu.cn; Weitao Xu, City University of Hong Kong, Hong Kong, China, weitaoxu@cityu.edu.hk; Jin Zhang, Shenzhen University, Shenzhen, China, jin.zhang@szu.edu.cn; Kuan Zhang, University of Nebraska-Lincoln, Lincoln, USA, kuan.zhang@unl.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Association for Computing Machinery.

2577-6207/2019/1-ART1 \$15.00

<https://doi.org/10.1145/3338026>

use informative images and show excellent performance for context recognition, especially when integrating additional depth sensors, e.g. Microsoft Kinect [15]. Motion-based context awareness applications are also prevalent by taking advantage of widely used motion sensors in mobile phones and smartwatches. Unfortunately, these solutions still have their limitations. Camera-based solutions have a privacy concern, especially when the captured images have to be stored and processed in the cloud. Although there is no privacy concern for motion sensor-based approaches, wearable devices must be carried and worn in proper manners to collect data by built-in sensors. This is inconvenient for older adults who may often forget to carry wearable devices. The wireless radio signal has broadly been investigated for *device-free* context awareness to resolve these issues. Halperin et al. [5] introduced a custom modified firmware based on a standard off-the-shelf WiFi card, Intel 5300 802.11n MIMO radio, which can provide informative Channel State Information (CSI) from Physical Layer. CSI measurements significantly increase resolution compared with the traditional Received Signal Strength (RSS), which helps realise many radio-based device free context awareness applications. The WiFi devices are widely deployed in an indoor environment, such as office and home, etc., and its ubiquity further stimulates its large-scale device-free application. CSI measurements collected from multiple radio devices deployed in the area of interests (AoI) can capture radio interference caused by monitored people. The patterns from CSI measurements due to the multi-path effect are mapped to relevant contexts. Recently, emerging radio-based context awareness applications (such as activity recognition [30–32], gesture recognition [13], “lip-read” [25], identity recognition [40], etc.) have been studied and can have equivalent performance as the traditional applications using wearable devices and cameras. The radio-based device-free context awareness applications gain a competitive advantage without requiring people to carry any device and breaking the privacy.

CSI based context awareness applications widely use supervised machine learning techniques which require the training stage to analyse labelled CSI measurements and derive a context awareness model. Both classic machine learning techniques (e.g., k-Nearest Neighbour [20], Sparse Representation Classification[31, 32, 40], etc.) and deep learning techniques (e.g. [13, 24, 27, 28, 41, 42]) have also been introduced for CSI based applications. One customised method is particularly designed for each application. An initial step for CSI data pre-processing further extracts features for classification. Raw complex-valued CSI measurements have been commonly considered as no usable and recognisable patterns without any data pre-processing due to their asynchronous phase. The existing research works all use the initial data processing as the first step[13, 20, 24, 27, 28, 31, 40–42]. As a result, recognition performance is highly relevant to these data processing methods. In contrast, the other successful fields of state-of-the-art machine learning applications, such as computer vision, use an entirely different strategy for pattern recognition. Without using any initial data processing methods, most computer vision techniques use raw images to train deep neural network (DNN) models and still significantly outperform traditional methods. They also use a general framework for computer vision-based applications [21]. They only need to modify the configurations or architectures for each particular application, but the whole DNN structure is still based on that general framework. Motivated by this, we design a novel DNN based general framework using raw CSI measurements without any additional feature selection and data pre-processing. We aim to generalise our framework to radio-based device-free context awareness applications. Our proposed method is capable of (1) preserving all the information without any pattern loss caused by an initial data processing stage¹, which equips itself with the ability of the generalisation and (2) revealing more information and patterns contained in raw CSI to achieve better performance. Capable of both

¹The information loss will be detailed in Section 3 and further quantifiably confirmed in Section 5.1.3 with the performance improvement compared with the existing method using initial data processing.

signal processing and classification, our proposed DNN framework attains excellent performance. As far as we know, we are the first to propose a general framework for radio-based application and directly use the *raw* CSI measurements to train the DNN model without any pre-processing.

To summarise, the contributions of this paper are as follows:

- To the best of our knowledge, this paper is the first to propose an innovative deep learning based *framework* using raw CSI data for context awareness without any feature selection and initial data pre-processing. This technical novelty of the model is the capability of using raw CSI data and, thus, being generalised to radio-based context awareness applications.
- We show the intuition and theoretical explanations in detail for the application of the raw CSI on training the DNN model. We also reveal the reason why the proposed general framework from raw CSI measurements can improve performance.
- We conduct extensive evaluations on various datasets to demonstrate the generalisation of our framework. We first design the DNN structure based on the general framework for radio-based device-free gesture recognition. We evaluate our proposed structure by using public CSI gesture recognition datasets that include 270 gestures and achieve approximately 100% accuracy.
- We further take advantage of our proposed general framework to design a DNN structure for activity recognition. We perform evaluations on CSI based activity recognition datasets in radio frequency interfered environment, which achieves equivalent performance as the previous work without any data processing. This confirms that the framework can be generalised for radio-based context awareness applications.

The rest of the paper is organised as follows. Section 2 is related work. Section 3 shows the background of CSI and the motivation of our method. The proposed DNN model is illustrated in Section 4 and evaluated in Section 5. Section 6 concludes the paper.

2 RELATED WORK

In this section, we discuss related research works in context awareness from radio data. The objects and people in one environment can reflect and diffract radio waves due to the multiple-path effect. This radio frequency perturbations can be detected by both the coarse-grained and fine-grained radio signal characteristics.

The radio signal has been used for context awareness applications, such as gesture recognition. WiSee designed by Adib et al. [2] and WiVi designed by Pu et al. [18] used software-defined radio to extract the Doppler effect caused by the gesture. In order to reduce energy consumption, Kellogg et al. [9] built AllSee that used RFID tags and power-harvesting sensors for gesture recognition. To further improve the resolution of the radio signal based localisation and gesture recognition, Adib et al. [1] designed WiTrack and obtained time-of-flight from the Frequency Modulated Carrier Wave (FMCW) technology for localisation in 3 dimensions.

Instead of using special software-defined radio equipment for those applications, RSS, which can be readily obtained from off-the-shelf radio cards, have been broadly used in device-free localisation applications[8, 33, 34, 37, 45, 46]. The coarse RSS was also successfully applied for gesture recognition [14].

However, sufficient characteristic information cannot be acquired from RSSI. To fulfil the gap, Halperin et al. [5] and Xie et al. [36] modified drivers of off-the-shelf WiFi cards, i.e. Intel 5300 and Atheros 9390, and obtained fine-grained Channel State Information (CSI) from the Physical layer of the customised hardware. The acquaintance of fine-grained features motivates many radio-based applications. Sen et al. used modified WiFi 5300 cards and collected CSI for indoor localisation [20]. Zhou et al. [47] also applied CSI information for omnidirectional passive human detection.

Gesture recognition application and activity recognition applications have been proposed by using fine-grained CSI information [13, 14, 24, 30, 31]. CSI based lip-read, emotion recognition applications and identity recognition applications were designed in [25, 26, 38, 40, 44]. There are also other context awareness research works using CSI, such as breath detection [39], sleep monitoring [12] and motion detection [4]. Moreover, phase information from CSI is used by research works for radio-based device-free applications as additional supplementary information [13, 19, 28, 29].

Deep learning, a branch of machine learning, is introduced for radio-based context awareness applications. Wang et al. [27, 28] designed indoor localisation systems by using deep learning techniques along with CSI phase information. [24, 42] and [41] also designed deep learning models for CSI based activity recognition and human identification, respectively. Ma et al. [13] took advantage of deep learning techniques and showed the feasibility of using 20 MHz radio band to recognise 276 gestures. These research works process CSI data initially and use deep learning as a classification tool. A specialised method is designed for each application. Different from these works, we propose a general framework for radio-based device-free applications. Our proposed framework conducts context awareness with raw CSI and outperforms the existing work without requiring any pre-processed CSI.

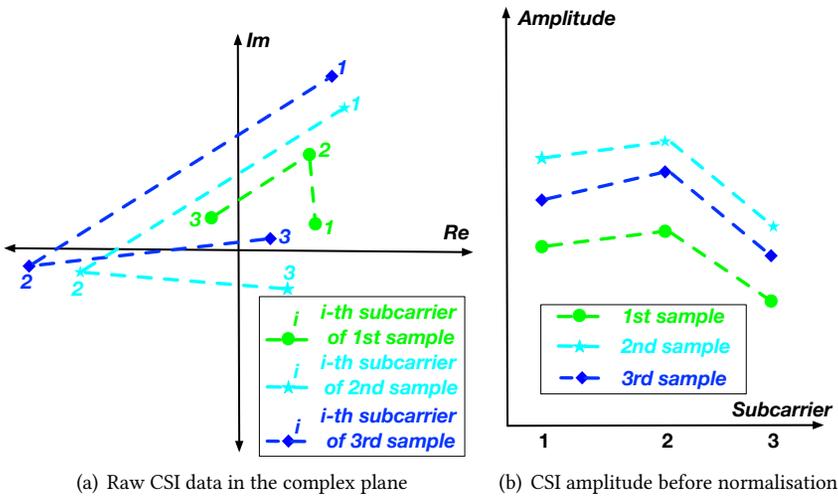


Fig. 1. Examples of data processing - Part 1 (a) Raw data for three samples in the complex plane; (b) CSI amplitudes before normalisation with similar shapes but different values for the corresponding samples

3 BACKGROUND AND MOTIVATION

The WiFi physical layer uses a multi-carrier modulation method called Orthogonal Frequency-Division Multiplexing (OFDM). At the common 2.4GHz WiFi band, the bandwidth of one channel is 20 MHz, and each channel has 52 subcarriers². The frequency response and phase shift of the centre frequency of each subcarrier are contained in CSI. The context variation significantly differentiates the multipath propagation of wireless signals and can be well reflected by CSI. The sensitivity of CSI to surrounding environments is broadly explored for context awareness by using supervised machine learning techniques.

²30 of the total 52 subcarriers in each channel at the 2.4GHz WiFi band can be reached when using the modified firmware of Intel 5300 WiFi card.

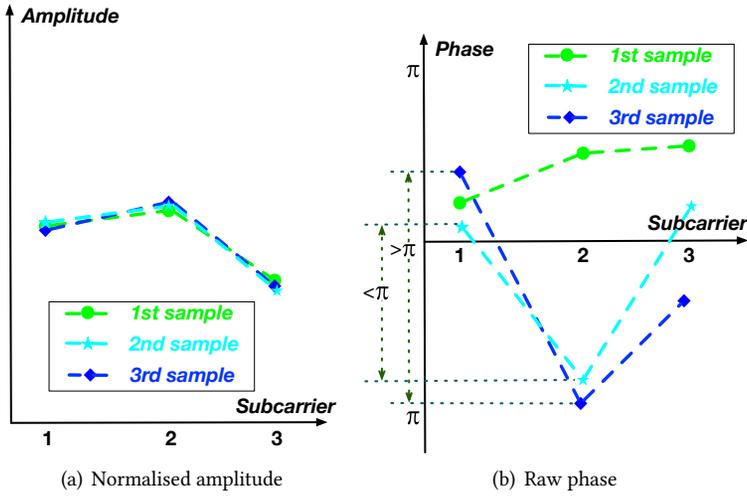


Fig. 2. Examples of data processing - Part 2 (a) CSI amplitudes after normalisation with similar shapes and values for the corresponding samples; (b) Raw phase data without any processing

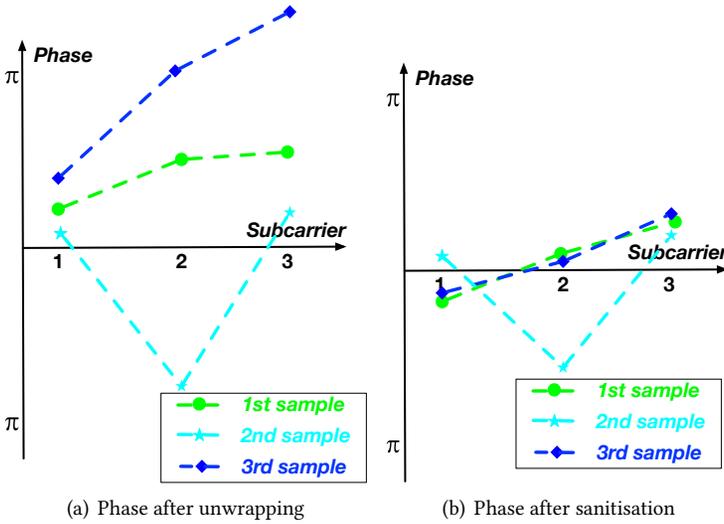


Fig. 3. Examples of data processing - Part 3 (a) Raw phase data after unwrapping with the default threshold π ; (b) Raw phase data after sanitisation proposed by [20]

Specifically, to make a prediction, supervised machine learning techniques require an explicit trained model obtained at the training stage using labelled data. The good quality of training data is exceptionally beneficial for increasing prediction accuracy. Two main factors prominently affect the quality of training data, i.e. the distance between classes and within one class. The high-quality training data further enlarge the distance between classes and narrow the distance within one class[23, 32], which plays a vital role in improving recognition accuracy.

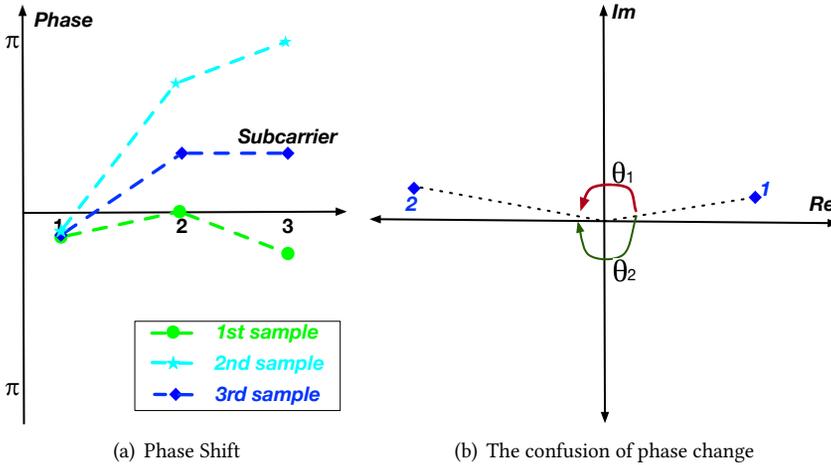


Fig. 4. Examples of data processing - Part 4 (a) Slope changes of CSI phases due to the asynchronous clock; (b) Different rotation directions in the complex domain

One CSI measurement is contained in a complex-valued vector $C = [C_1, C_2, \dots, C_n]$, where n is the number of subcarriers. To better illustrate the steps of CSI signal processing, we select 3 neighbouring subcarriers among them as an example. 3 CSI samples are continuously collected in a very short period during a static gesture is performed in order to ensure no changes in the surrounding environment. The raw CSI measurements in the complex domain are shown in Figure 1(a). Figure 1(b) shows the amplitudes of these 3 CSI samples.

Many radio-based applications conduct context prediction only using CSI normalised amplitude information [30, 31] and phase-based or complex-valued CSI [13, 27, 32]. We first discuss these data processing methods here and their limitations.

Amplitude information is the absolute values of each CSI subcarrier, able to provide relatively good performance for many radio context awareness applications. Normalisation is also applied because of the RSS variation as the example shown in Figure 1(b). Figure 2(a) demonstrates the consistency of normalised CSI amplitude information for a static context. However, the lack of phase information degrades recognition performance in many complicated scenarios, such as context awareness in radio frequency interference [32] or with a large number of classes [13], where complementary phase information needs to be explored. It has been shown that the use of phase information can improve the true detection rate up to 10% [32]. *Due to a random phase shift in OFDM from the unsynchronised time between the transmitter and receiver, it is, however, challenging to directly use the phase information without explicitly processing in these existing applications.* Sen et al. [20] proposed a CSI data sanitisation method to calculate the phase information, which is massively used by phase-based or complex-valued CSI context awareness applications [13, 27, 32]. The sanitisation method firstly unwraps the raw phase and calculates the phase slope and offset to eliminate the phase shift. Complex-valued vectors can be constructed using the normalised amplitude and phase vectors. The step of unwrapping corrections aims to correct radian phase angles to avoid the jump between consecutive subcarriers. In the unwrapping process, when the phase difference between neighbouring subcarriers is above an artificial threshold (the default threshold is π), a compensation 2π is added/subtracted to the phases of its following subcarriers.

CSI based context awareness commonly relies on supervised machine learning techniques. Different contexts derive multiple radio transmitting paths resulting in distinguishable CSI patterns, while the same context should be indicated by similar CSI patterns. However, improper signal processing may introduce unnecessary errors. Taking unwrapping CSI phases as an example, the phase difference of many consecutive subcarriers is extremely close to the artificial unwrapping threshold. Unwrapping method is only applied to the samples of which the phase difference of these consecutive subcarriers is above the threshold, while the rest samples stay unchanged. The use of the artificial threshold may introduce notably inconstant situations for these consecutive CSI samples within one context because not all of those neighbouring subcarriers in samples are treated by unwrapping. Figure 2(b) and Figure 3(b) are examples to further imply the unnecessary unreliability of CSI measurements caused by the unwrapping step. Phases for the three samples in Figure 1(a) without any processing are shown in Figure 2(b). The phase difference between subcarriers 1 and 2 of sample 3 is above the artificial threshold π . Therefore, after applying the unwrapping approach, the phases of sample 3 are unwrapped, and the phases of samples 1 and 2 stay unchanged, as shown in Figure 3(a). Please note that the similar shapes of the raw samples 2 and 3 become completely different after unwrapping step. Except for the unavoidable noise, another reason to result in this unwrapping issue is the changes in slope due to the asynchronous clock, which is demonstrated by another example shown in Figure 4(a). The phases of these 3 samples have the same shape with different orientations. Please note that the phase difference between subcarriers 1 and 2, its difference in sample 2, above the default artificial threshold π , is much greater than that in sample 1. As a result, the phase from sample 2 is the only case of being unwrapped when processing their raw phases with the unwrapping method. The unwrapping step produces unnecessary inconsistency and errors when applying supervised machine learning techniques on those samples because supervised machine learning resorts to the repeatable patterns from training data. The third reason for the instability at the unwrapping step is the different rotation directions in the complex domain as shown in Figure 4(b). θ_1 and θ_2 indicate phases for the same vector with anticlockwise and clockwise rotation in the complex domain, but θ_2 is above the default threshold for unwrapping. In other words, unwrapping in the CSI sanitisation method makes the phase information feasible for context awareness but introduces unnecessary unreliability.

Similar to CSI, time series inertial measurements have multiple dimensions. Deep learning has been applied on inertial measurements for context awareness [3]. However, raw CSI measurements and IMU measurements have fundamental differences. Figure 5 shows examples of CSI measurements and IMU measurements, which shows continuous 100 samples of CSI and IMU measurements. We collected these measurements using Intel IWL5300 [5] and iPhone X, respectively. From Figures 5(c) and 5(d), it is clearly seen that time series inertial measurements (acceleration measurements and gyroscope measurements) are continuous. Different from inertial measurements, CSI measurements in Figure 5(a) and Figure 5(b) are not as smooth as IMU measurements, especially the phase information. The phase information fluctuates significantly considering its range between $-\pi$ and π . Another difference between inertial measurements and IMU measurements is the measurement dimension. Different from the fact that acceleration and gyroscope measurements are 3 dimension data, CSI have 52 dimensions (i.e. sub-carriers) for each 2.4 GHz or 5 GHz channel with 20 MHz bandwidth and 114 dimensions (i.e. sub-carriers) for each 5 GHz channel with 40 MHz bandwidth³. Image processing related applications [21] also often use DNN. However, different from CSI measurements, the neighbouring pixels in raw images have clear relations where DNN can learn these patterns. In contrast, CSI raw data do not have strong and obvious patterns without pre-processing.

³When using Intel IWL5300 [5], CSI of 30 sub-carriers out of all are extracted.

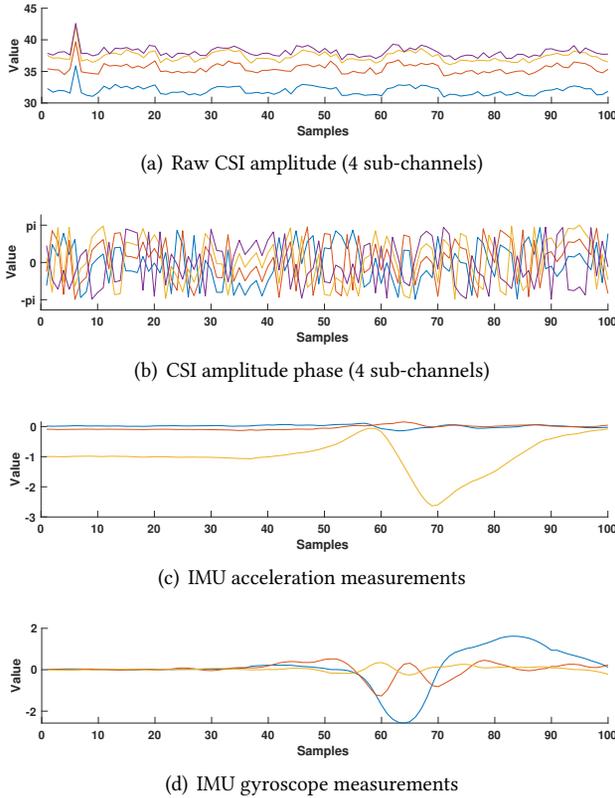
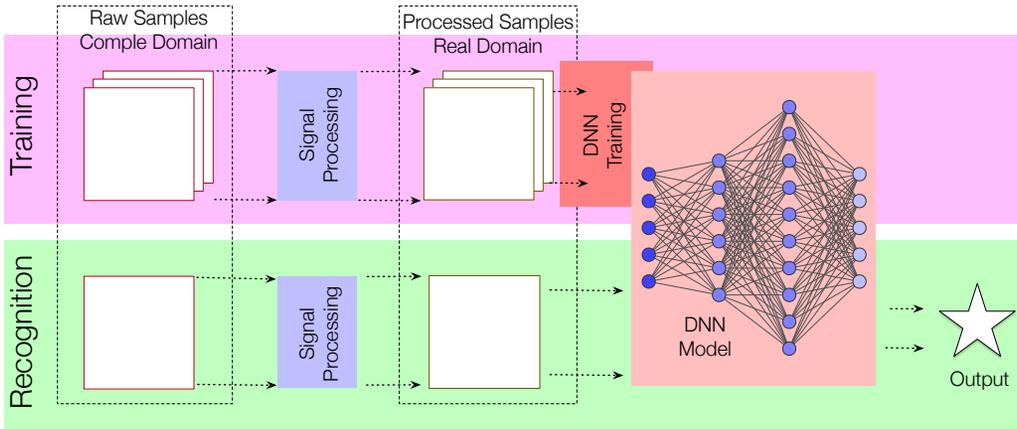


Fig. 5. Inertial measurements and raw CSI measurements

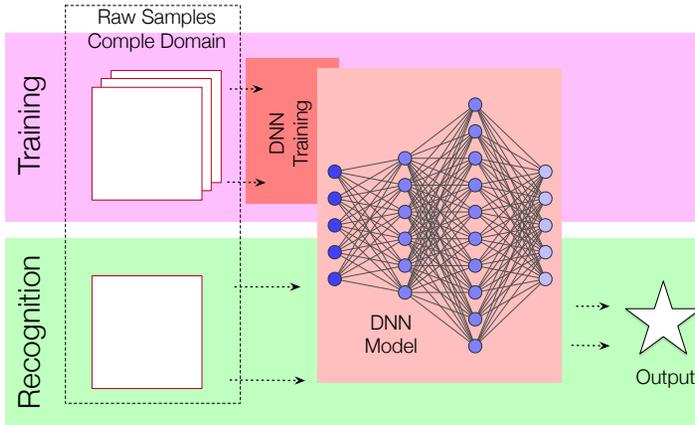
These differences make the use of deep learning on CSI more challenging than those on IMU data and image data in these aspects, requiring more explicit effort for data analysis and model design.

To summarise, there are two main challenges for the use of raw CSI to improve the performance of context awareness. First, the pattern of raw CSI, especially the phase information, is hidden without the use of proper processing and recognition framework. Second, the current processing mechanism can explore the patterns but accompany along potentially generated unavoidable unreliability due to the noise. To address these issues, we, therefore, propose a deep learning based general framework which considers the additional informative phase information along with the amplitude information to improve CSI based context recognition performance and avoid unnecessary unreliability by pruning any initial data pre-processing stage. Even though the unavoidable noise can involve inconsistency to the processing method, deep learning is used to equip the proposed model with the ability to recognise distinguishable patterns within noisy CSI measurements, which avoids inconsistency mentioned above as well as preserves CSI patterns within raw CSI. Deep learning has been applied on CSI measurements in the existing work [13], but its performance on raw CSI is not satisfying⁴. Since the raw CSI does not contain any visible repeated patterns, an explicit design of a general novel DNN structure to facilitate automatic pattern learning is of great importance.

⁴Results can be found in Section 5.1.3.



(a) Traditional DNN model for CSI based context awareness using signal processing



(b) The proposed Model for CSI based context awareness with any initial processing

Fig. 6. DNN models for CSI based context awareness

4 GENERAL FRAMEWORK: CSI BASED CONTEXT AWARENESS USING DEEP LEARNING

The sanitisation process in [20] tries to regularise the phase and constructs complex valued CSI measurements for further classification and clustering. Traditionally, computer vision recognition algorithms also use similar strategies. Data representation is explicitly designed for specific classification methods to improve the recognition performance, such as features in the face recognition field: Eigenface[43], Fisherface[10] and Laplacianface[6]. Recently, researchers use deep learning techniques to train DNN models using raw images without any pre-processing instead. Deep learning has significantly enhanced performance in many domains, such as face recognition[17], object recognition[11], etc. It can preserve all the features and avoid unnecessary information loss by directly processing raw images. Researchers in those domains also propose a framework which can be generalised to other datasets [21].

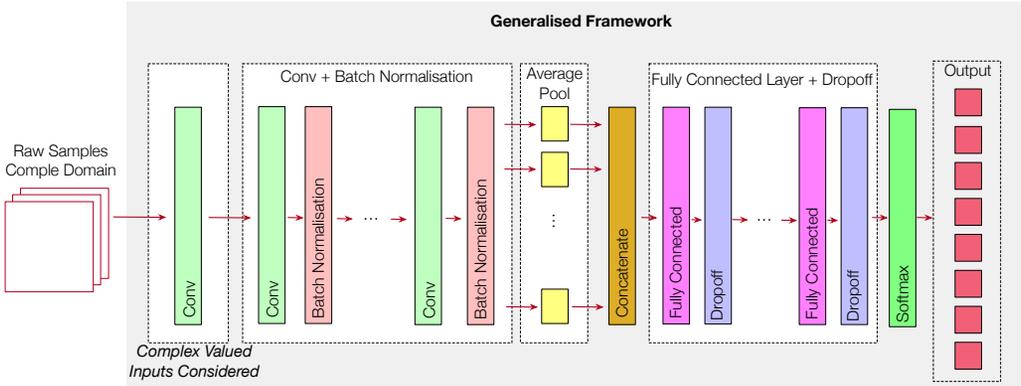


Fig. 7. The general framework for radio-based device-free context awareness

Motivated by the research works in other domains, we propose a general framework and investigate the application of raw CSI measurements for radio-based context awareness applications without any signal processing. The advantages are two-fold. Firstly, the unmodified CSI measurements preserve all the original and relevant information, and the DNN helps select the desirable features for context awareness instead of designing feature representation. Secondly, the framework can be applied to all the CSI based context awareness applications because our proposed algorithm does not design features for specific application scenarios. Several research works have already applied deep learning on radio-based context awareness[13, 24, 27, 28], but they have an additional step to process raw CSI measurements for training a DNN model. Furthermore, they have not investigated the feasibility of adapting their models to other applications or datasets. Figure 6(a) shows the structures of those traditional applications. Firstly, they apply a data pre-processing method to prepare the raw complex-valued CSI samples in the complex domain for further classification. There are two common signal processing methods for CSI data: (1) abstract the amplitude of complex-valued CSI measurements by calculating absolute values of that in each subcarrier; or (2) calculate the phase using the sanitisation method introduced in [20], and use phase information to construct sanitised complex vectors along with the amplitude information or as complimentary of the amplitude information. The processed measurements are taken into consideration for training a DNN model.

Different from the traditional methods, we propose a DNN model based general framework by using the raw CSI measurements in the complex domain directly. To achieve this, our proposed framework uses the structure shown in Figure 6(b). Instead of initially processing raw CSI measurements, we explore them directly for DNN model training and inference process. As illustrated in Figure 7, our proposed general framework has the following important components in the neural network architecture: (1) convolutional layers, (2) batch normalisation layers, (3) average pooling layers (4) dropout layer (5) fully connected layer and (6) softmax layer. Here are more details of our proposed architecture.

Raw CSI measurements as Input: For CSI based context awareness, the shape of each instance is 3 dimensions, i.e. $2m \times n \times c$. m is the number of CSI measurements in one CSI instance. We use real and imaginary parts together for training our deep model, so the size of this dimension is $2m$. n and c are the numbers of subcarriers and antenna pairs, respectively. Figure 8 shows the example of one instance.

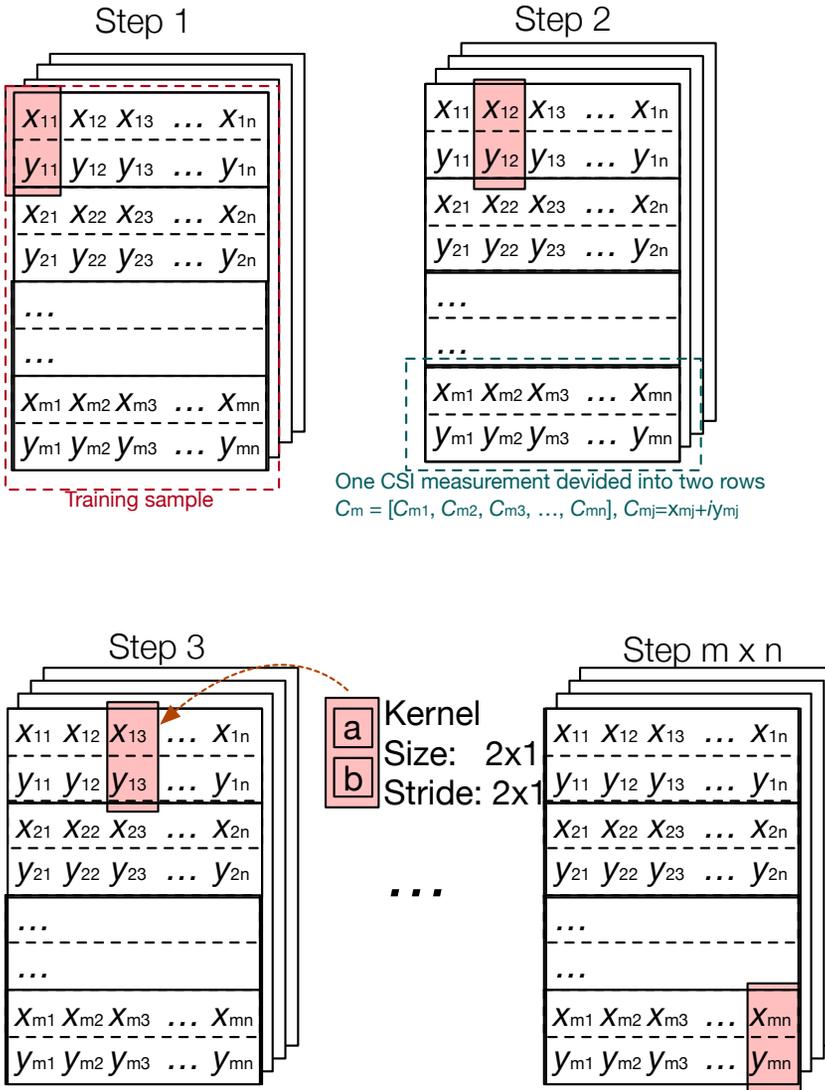


Fig. 8. first layer kernel for each step

Convolutional Layer and Batch Normalisation Layers: We use multiple convolutional layers in our framework. The first layer considers the complex-valued inputs. Figure 8 shows each step when using Kernel whose size and stride are both 2×1 . This layer takes both the complex and real parts of the CSI measurements into accounts. The stride 2×1 ensures that this layer only considers the relevant subcarrier of each sample at one time. The followed convolutional layers use different size of kernels to explore the diversity of the features. Rectified Linear Unit(ReLU) non-linearity activation functions are used with convolutional layers. Batch normalisation layers are also used along with convolutional layers. Since the inputs of DNN are raw CSI instances without any initial data processing, the batch normalisation layers can scale the activations in order to accelerate the

Table 1. Radio Based Gesture Recognition Dataset Overview

Dataset ID	Environment	Number of Gestures	Number of Users	Number of Instances
D1	Home	276	1	2,760
D2	Lab	276	1	5,520
D3	Lab	276	1	5,520
D4	Lab	150	5	7,500

training process and improve the performance. The batch normalisation layers are also important in managing extremely diverse radio signal strength.

Average Pooling Layer and Concatenate Layer: After extracting features from convolutional layers, we take the neighbouring subcarriers and samples into consideration and use average pooling layers. These layers play an essential role in recognising various degrees of movements. As shown in Table 2, in this layer, we used concurrent multiple pool sizes in order to consider different ranges of subcarriers generated from different contexts. A concatenate layer is appended to combine features from the outputs of these average pooling layers.

Dropout Layer: Overfitting is a common issue in the machine learning domain when the evaluation performance decreases with the increase of the training performance [22]. The dropout layer is a simple but effective technique to address the overfitting issue in deep learning based applications, which randomly drops a certain percentage of the outputs from the previous layer. Dropout layers are also employed in our framework to avoid the overfitting in the training stage.

Fully Connected Layer: Fully connected layers, along with ReLU non-linearity activation, are introduced to assemble the features from the above layers and determine patterns for further decision making.

Softmax Layer: The final softmax layer is constructed to recognise context based on the hidden state from fully connected layers.

This paper aims to propose a general framework for radio-based device-free context awareness. The proposed framework prunes the initial processing of CSI data, which has two advantages. First, the removal of the pre-processing helps generalise CSI based context awareness without the consideration of data processing for one particular application. This would equip our proposed framework with the capability of its application on more application scenarios. Additionally, we also find that CSI without pre-processing contains more usable patterns, as discussed in Section 3, which is capable of improving the recognition performance. The following Section 5 will evaluate our proposed general frameworks with specific DNN structures in two application scenarios.

5 PERFORMANCE EVALUATION: THE APPLICATIONS OF THE PROPOSED GENERAL FRAMEWORK

In this section, we validate our proposed deep neural network based framework based on two applications to show its generalisation and performance: (1) *SignFi*: the public datasets from SignFi [13] for gesture recognition⁵; and (2) *Activity*: the datasets for radio based activity recognition with radio frequency interference (RFI) in [32]. The key characteristics and fundamental difference of these two datasets are (1) *SignFi* with 276 classes has far more classes than other datasets. Therefore, the performance in *SignFi* can provide experience to other real application scenarios. (2) To show the generalisation of the proposed DNN framework, we also use our collected datasets *Activity*. Different from *SignFi*, one dataset in *Activity* is collected under radio interference, including

⁵The datasets can be downloaded from <https://yongseng.github.io/SignFi/>

more challenges for signal processing. The good performance in that dataset can further support the generalisation of the proposed framework. In this section, we show the DNN structures for each application based on the general framework and compare our proposed model with existing methods. To expedite the use of the proposed framework, we show the effect of the convolutional layer, the batch normalisation layer, the average pooling layer and the fully connected layer. The effect of those layers will ease the use of the proposed framework in the real application scenario and provide valuable experience to researchers and engineers. More details of the datasets in these two applications can be found in Section 5.1.1 and Section 5.2.1.

We use the true detection rate as the metric that is defined as a ratio of the number of correctly recognised samples and the total number of data samples. In this section, we evaluate our proposed method using k -fold cross-validation [16]. Specifically, when using k -fold cross-validation, the dataset is divided into k disjoint subsets, and k times validations are conducted. In each validation, one of k disjoint subsets is used as the validation set, and the rest ones are used as the training set. The overall true detection rate is its average in k validations, and the standard deviation of the true detection rates in k validations is shown as the error bar with the overall true detection rate in the corresponding figures to show the uncertainty.

5.1 The DNN Structure and Performance in SignFi Datasets

5.1.1 Dataset Description. There are four public datasets in SignFi obtained from experimental measurements for radio-based gesture recognition in a laboratory and a home scenario. The datasets contain raw CSI measurements with regards to different room sizes, distances between the transmitter and receiver, and orientations of their antennas. In these datasets, the sampling rate of CSI measurements is 200Hz, and the gesture duration in one instance is between 0.5 seconds and 2 seconds. The instances with a gesture have been segmented. In other words, an instance only contains a gesture. When using this dataset, we use the same strategy in [13] that samples with one gesture are used for training. Continuous activities in different instances are considered individually, which increases the generalisation without the assumption of correlation of neighbouring instances. The detailed information on the four datasets of SignFi is summarised in Table 1. In D1, 276 gestures are performed by one user in the home scenario, and 2,760 instances are collected. The CSI instances from D2 and D3 are simultaneously collected from the receiver and transmitter, respectively. The receiver and transmitter are deployed in the lab, where 5,520 CSI instances are collected. The CSI instances in D4 are also collected in the lab scenario, where 7,500 instances of gestures are performed by 5 users. Please note, due to multiple users in D4, each user's gestures are recognised by using his or her own CSI instances so that the DNN can be trained more accurately.

5.1.2 Architecture of DNN model. To achieve the best performance, we use the general framework and the configuration of the model shown in Table 2. Specifically, we use 4 convolutional layers each with batch normalisation layer, and 5 average pooling layers, and 1 fully connected layer with dropout layer.

In this section, we evaluate the proposed method 5-fold cross-validation with all the four datasets, as the same validation method is used in SignFi [13] as well. The datasets used by SignFi contains much more classes than other radio-based context awareness research works, which can be fully exploited to evaluate the proposed general framework. In [13], the authors proposed the deep learning model along with data pre-processing for radio-based gesture recognition. In this paper, we investigate the feasibility of using raw CSI without any data pre-processing instead.

When discussing the effect of the convolutional layers, the batch normalisation layers and average pooling layer, we only change the configuration of that layer category, and the rest layers

Table 2. Configuration of DNN used in SignFi datasets.

	Kernel Size	Stride
Conv_1	2 x 1	2 x 1
Conv_2	3 x 3	1 x 1
Conv_3	5 x 5	1 x 1
Conv_4	10 x 10	1 x 1
Pool Size		
AP_1	3 x 3	
AP_2	5 x 5	
AP_3	10 x 3	
AP_4	20 x 3	
AP_5	40 x 3	
Dropout Rate		
Dropout	0.8	
Number of Unit		
Fully connected	1000	

use the default model as shown in Table 2. Conv and AP stand for convolutional layer and average pooling layer, respectively.

5.1.3 Comparing with SignFi. We compare the proposed DNN model with that in SignFi[13] in terms of the true detection rate⁶. SignFi[13] also uses a deep learning model but hand-crafted features, i.e. CSI amplitude values. Different from the model from SignFi[13], we explore the raw CSI without any processing, which reserves all the information, as explained in Section 3. Figure 9 shows the performance of our proposed method, and our DNN model achieves 99.89%, 99.98%, 99.93% true detection rates in the home environment (D1) and the lab environment (D2 and D3), respectively⁷. SignFi achieves 98.91% and 98.01% true detection rates in the home environment (D1) and the lab environment (D2 and D3 together), respectively. Please note SignFi has used feature extraction introduced in [20] which is discussed in Section 3 and results of SignFi shown in this section are obtained with feature selection if without notation. Differently, our proposed deep learning model outperforms the existing model without any data pre-processing. SignFi also discussed the performance of its deep learning model without pre-processing in the home environment (D1) and the lab environment (D2 and D3 together). Without data pre-processing, the true detection rates drop to 93.98% and 95.72% using their method, while our proposed method can still achieve nearly 100% true detection rate. The results confirm that the proposed method without any data pre-processing can achieve equivalent or even better performance than handcraft feature. The performance decrease in the existing method can indicate the information loss due to the data processing. The significance of the performance in this evaluation is the confirmation of the feasibility of context awareness using the proposed framework without any data pre-processing. The unnecessary of pre-processing of our proposed framework readily enables its flexibility of extension to many CSI based applications without much consideration of feature selection for a special application scenario, which is one of the main contributions in the paper. This will be further confirmed by the results in Section 5.2 that uses our proposed general framework for another application: radio-based device-free project recognition in radio frequency interfered environment.

⁶The results for SignFi with and without signal processing in this section are from [13].

⁷The performances of SignFi are from [13], so error bars are not drawn due to the missing of the relevant data in [13].

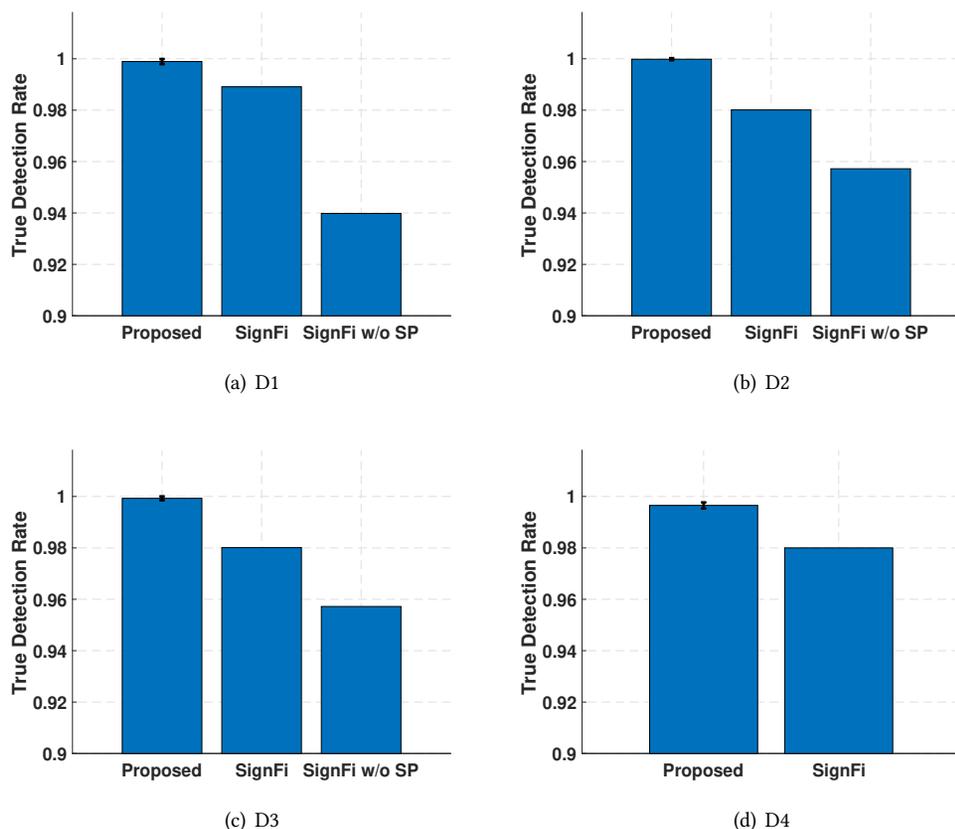


Fig. 9. The performance of true detection rate with regards to the four datasets

5.1.4 Effect of the Convolutional Layer. Next, we investigate the effect of the number of convolutional layers. Figure 10 shows the true detection rates of our proposed DNN model with regards to the number of convolutional layers, where the proposed model applies 4 convolutional layers. With using less convolutional layers, the performance drops dramatically except dataset D3. In D3, the true detection rate only drops from 99.93% to 93.59%. The performances using different layers (i.e. the proposed, 3 layers, 2 layers) are very close. However, the proposed clearly outperforms that using fewer layers in the other datasets. When using one less layer than the proposed model, the performance stays similar except the true detection rate drops 7.89% from 99.65% to 91.79% in D4. For D2 and D4, the true detection rates are only 73.24% and 87.23% with 2 convolutional layers, and it further drops to 36.81% and 72.88% when only using 1 convolutional layer. These results confirm that the increasing number of convolutional layers can significantly improve accuracy. Furthermore, our proposed model with 4 convolutional layers is more robust to all the datasets and able to achieve a sufficiently good and stable performance.

5.1.5 Effect of The Batch Normalisation Layer. Because of the changing received signal strength, it is impossible to learn a feasible model without batch normalisation layer.

Figure 11 depicts the true detection rates achieved by applying the proposed DNN model with and without the batch normalisation layer. In D2 and D3, the true detection rates are no more than

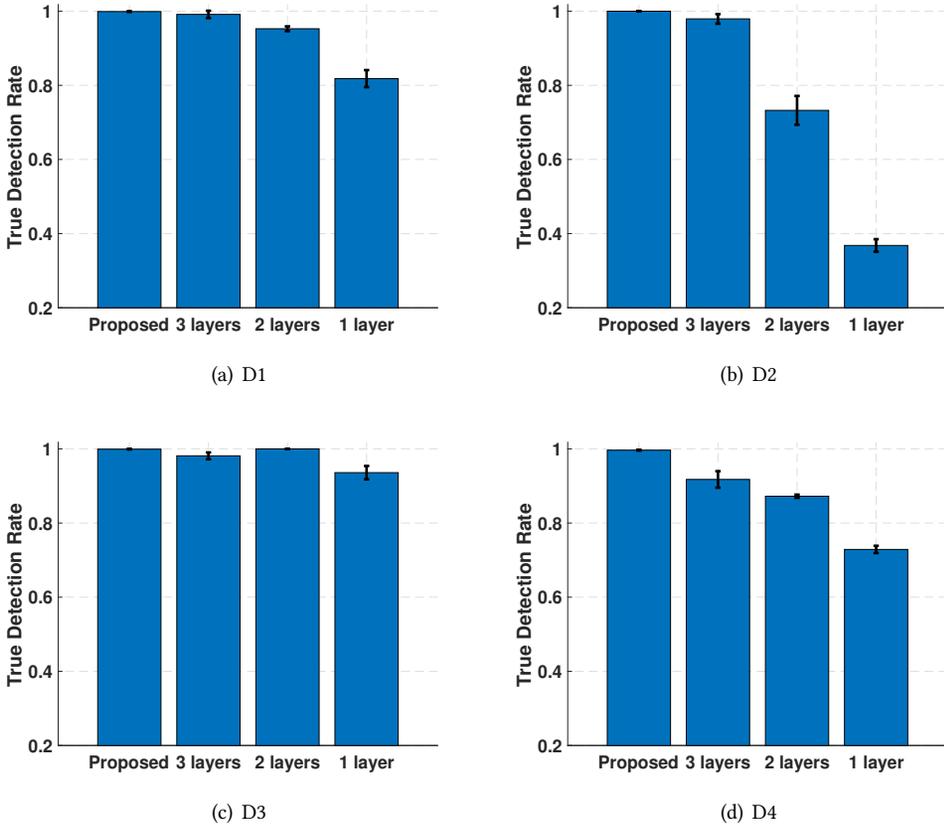


Fig. 10. The true detection rate of the proposed deep learning model with different number of Convolutional Layers

2% without the batch normalisation layer. Moreover, the performance of our model without the batch normalisation layer achieves 20.51% and 50.75% in D1 and D4, respectively. These results again demonstrate the importance of the batch normalisation layer in the DNN model for CSI based context awareness applications.

5.1.6 Effect of The Average Pooling Layer. Figure 12 shows the true detection rates of the proposed model with and without the average pooling layer. In the D1, D2 and D4, the true detection rates drop 13.64% (to 86.25%), 5.94% (to 94.04%) and 13.4% (to 86.25%), respectively. The performances are similar in D3. These results show that average pooling layers play a vital role in the performance stability of the radio-based context awareness.

5.1.7 Effect of Deep Learning. In Section 5.1.3, we have compared our proposed framework with the existing state-of-the-art feature based deep learning method. To further confirm the capacity of deep learning, we show the effect of our proposed DNN in SignFi datasets in this section. We compare the proposed method with the classical machine learning methods random forest. For a fair comparison, we do not conduct any initial processing before using the random forest method. Figure 13 clearly shows the proposed method outperforms the random forest. In all the datasets,

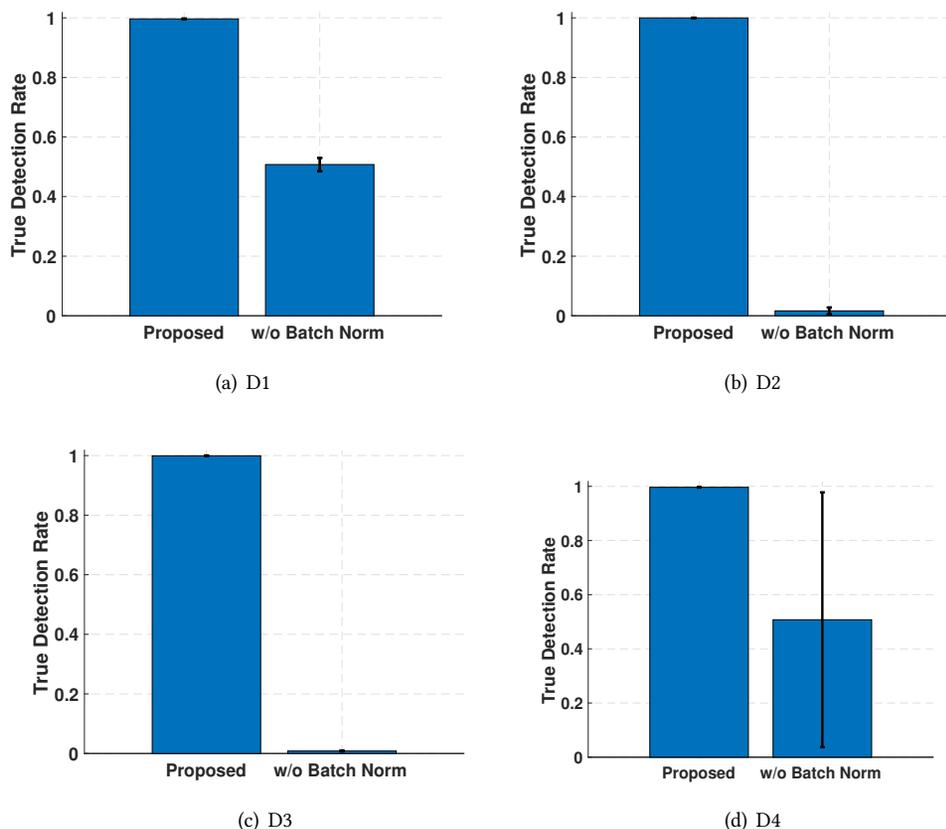


Fig. 11. The true detection rate of the proposed deep learning model with/without the batch normalisation layer.

the random forest only achieves the performance as a random guess, while the proposed method can achieve approximately 100% true detection rate.

5.1.8 The Performance in Each Class. Figure 14 shows the confusion matrix for D1, D2, D3, and D4. In previous sections, we see the true detection rates are 99.89%, 99.98%, 99.93% and 99.65% when using our proposed method. When considering a large number of classes and the small number of mismatch items, it, if a common heat map colour scheme is used, is not clearly finding incorrectly labelled classes in the confusion matrix. Therefore, a high contrast colour scheme is used instead to highlight mislabelled items. To further support the illustration items, the non-zeros elements not in the main diagonal of a matrix (meaning labelled mislabelled percentage) are circled for D1, D2, and D3 where the relevant points are still not very clearly visible due to low percentage of mismatched items. Figure 14 further confirms that our proposed method has a good true detection rate in all classes, and most of them have 100% true detection rates.

5.2 The Performance in Activity Dataset

5.2.1 Dataset Description. The Activity dataset is from the research work [32], where software-defined radio WAPS nodes are used to simulate various 802.11 protocols for radio-based device-free

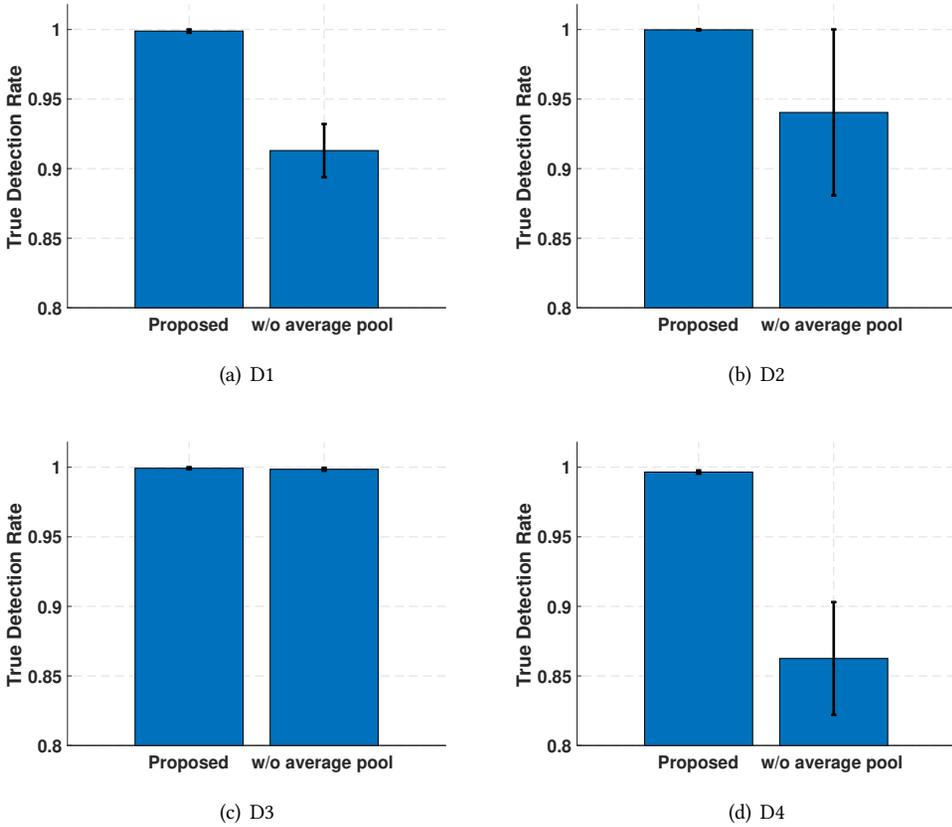


Fig. 12. The effect of average pooling

activity recognition. Two WASP nodes were deployed in a one-bedroom apartment. One WASP node was the transmitter, which sent one packet every 0.1 second. The other WASP node, as the receiver, received packets and transferred CSI data to a connected PC through a USB cable for further processing.

Two primary datasets have been collected in two radio frequency conditions, i.e. with and without radio frequency interference (RFI). When collecting the CSI data, WAPS operated in 5.8 GHz and used the entire 125 MHz industrial, scientific and medical (ISM) radio band. 8 common location-oriented activities were investigated in [32]. In the RFI environment, one router was deployed in the middle of the apartment and communicated with the PC to generate radio interference in Channel 157 whose bandwidth is 20 MHz. Two secondary datasets in Channel 157, with and without RFI, were selected to explore the performance of the radio-based device-free activity recognition in different RFI conditions. To evaluate our proposed general framework, we use these two secondary datasets, denoted as AD1 (Channel 157 without RFI) and AD2 (Channel 157 with RFI). Each dataset has 952 instances. In each instance, there are 5 samples, 52 subcarriers and 1 antenna pair. We also use the same setting 10-fold cross-validation in [32] to evaluate our proposed framework. As the same use of SignFi datasets illustrated in Section 5.1.3, we use the same method in [32] that neighbouring instances are considered individually for recognition.

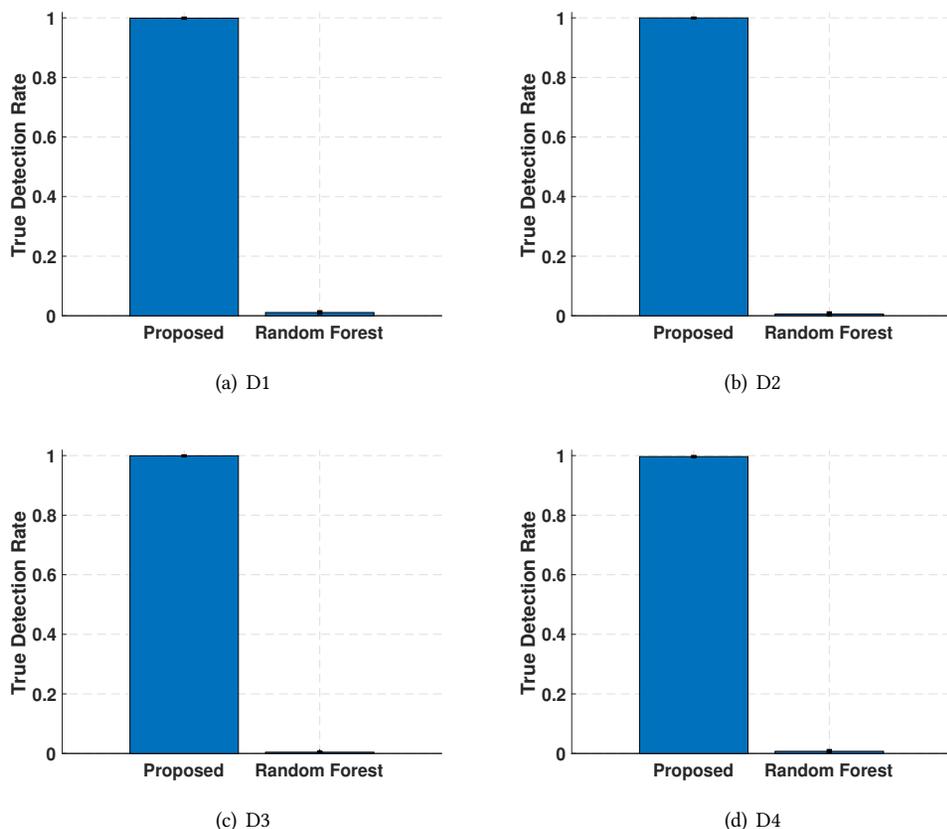


Fig. 13. The effect of deep learning

5.2.2 Architecture of the DNN model for Activity Dataset. Table 3 shows the architecture of the DNN model we use for AD1 and AD2. This architecture still uses our proposed general framework but different settings with SignFi dataset. The reasons for the different settings are two folds. First, the inputs of the CSI data are different. There is only one transmit-receive antenna pair used in the configuration of the Activity dataset, but CSI from multiple antenna pairs are collected in SignFi datasets. More subcarriers from one channel can be accessed to provide CSI from WASP device (i.e. 52 subcarriers instead of 30). Secondly, the Activity datasets consider the RFI in one secondary dataset AD2. RFI significantly increases the difficulties of pattern analysis, which is required to include additional efforts in the DNN model design.

Table 3 shows the model used for the Activity datasets. In this model, we employ 7 convolutional layers, 3 parallel average pooling layers followed with a concatenate layer, and 2 fully connected layers each followed with a dropout layer. As discussed in the general framework, the first convolutional layer considers the complex-valued input with 2×1 kernel size and stride. More convolutional layers are added in this model to address the challenges of noisy data and compensate for the lack of available antenna pairs. 3 average pooling layers are attached with pool sizes 1×2 , 1×3 and 1×4 . We use less average pooling layers than that used on the SignFi datasets because the usable patterns in the neighbouring samples decrease with the impact of the RFI. Furthermore, the

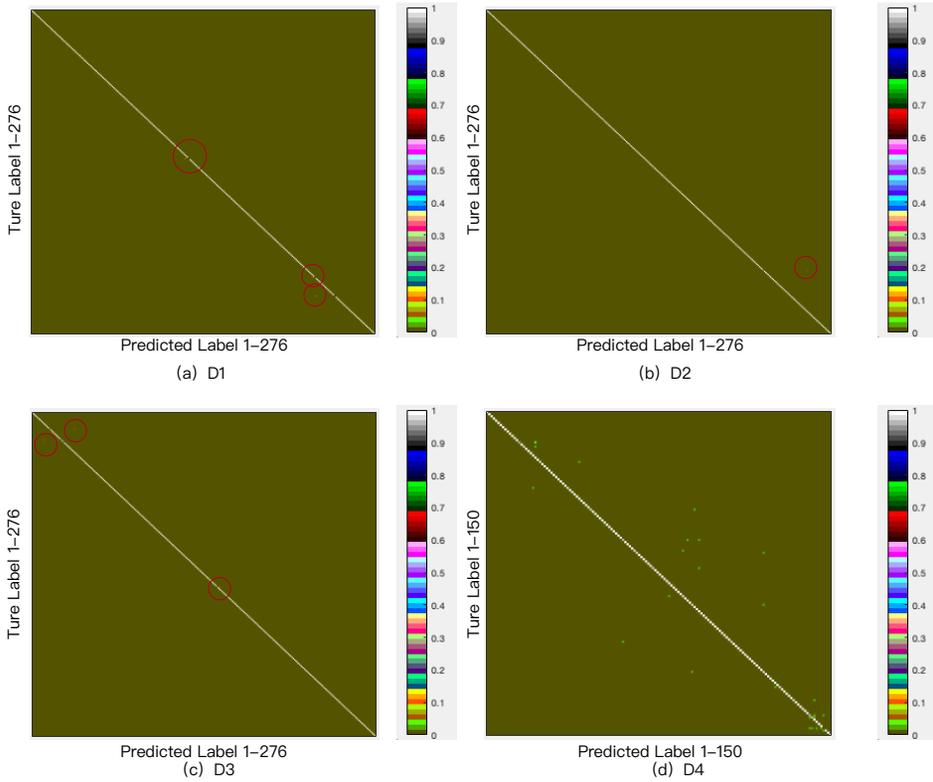


Fig. 14. Confusion matrix for D1, D2, D3, and D4. Due to the large number of classes and the small number of mismatch items, a high contrast colour bar is used to highlight mislabelled items. To further support the illustration items, the non-zeros elements not in the main diagonal of a matrix (meaning labelled mislabelled percentage) are circled for D1, D2, and D3 where the relevant points are still not very clearly visible due to low percentage of mismatched items.

consecutive samples in the AD2 with RFI may not be correlated, so we set the first dimension of the pool sizes of these three average pooling layers as 1. In this case, we only consider one sample in each average pooling layer. To further improve the performance, we use 1 more fully connected layer. The softmax layer is used as the final layer for training and inference.

5.2.3 Comparing with the benchmark. In this section, we compare the performance using our proposed general framework with the method in the previous research work [32]. The research work [32] introduced one extra step to obtain complex-valued CSI and used it to improve the performance of device-free activity recognition within the radio frequency interference. Differently, we explore the raw CSI without any processing. Figure 15 shows the accuracy of our proposed architecture and the method in [32]⁸. The true detection rate increases from 93.48% to 97.40% in AD1 without RFI, which indicates the advantages of our proposed deep learning based architecture. The true detection rate using our proposed framework in AD2 is 85.08% the same as that in [32]. Please note our general framework does not need any signal processing, while the previous method requires the explicit complex-valued CSI processing method design [32]. The results confirm the

⁸The performances of SRC are from [32], so error bars are not drawn due to the missing of the relevant data in [32].

Table 3. Configuration of DNN used in Activity datasets

	Kernel Size	Stride
Conv_1	2 x 1	2 x 1
Conv_2	1 x 2	1 x 1
Conv_3	1 x 3	1 x 1
Conv_4	1 x 4	1 x 1
Conv_5	1x 8	1 x 1
Conv_6	1 x 12	1 x 1
Conv_7	1 x 16	1 x 1
Pool Size		
AP_1	1 x 2	
AP_2	1 x 3	
AP_3	1 x 4	
Number of Unit / Dropout Rate		
Fully connected	1000	
Dropout	0.8	
Fully connected	1000	
Dropout	0.8	

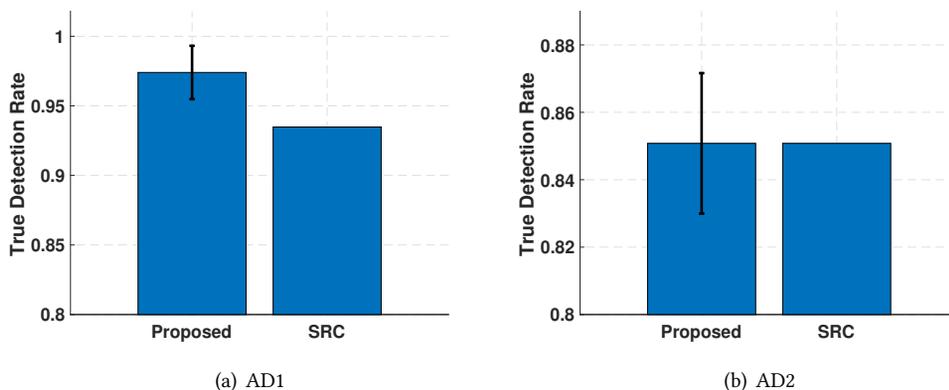


Fig. 15. The performance of our proposed architecture and the method in [31]

feasibility of the application of our proposed deep learning based architecture in challenging RFI environments.

5.2.4 Effect of the Convolutional Layer. In the section, we show the effect of convolutional layers in Activity datasets. Because Conv_1 layer is designed for the complex-valued CSI, we only explore the effect of the remaining 6 layers in this section. Figure 16 shows the performance of our proposed architecture with different numbers of convolutional layers. Figure 16(a) demonstrates that the number of convolutional layers does not impact the performance in AD1 without RFI. The similar results are also observed in SignFi D3 shown in Figure 10(c). The CSI data without RFI usually contain distinct patterns which can be easily explored for context awareness without considerable effort. However, when introducing RFI, radio-based context awareness becomes much

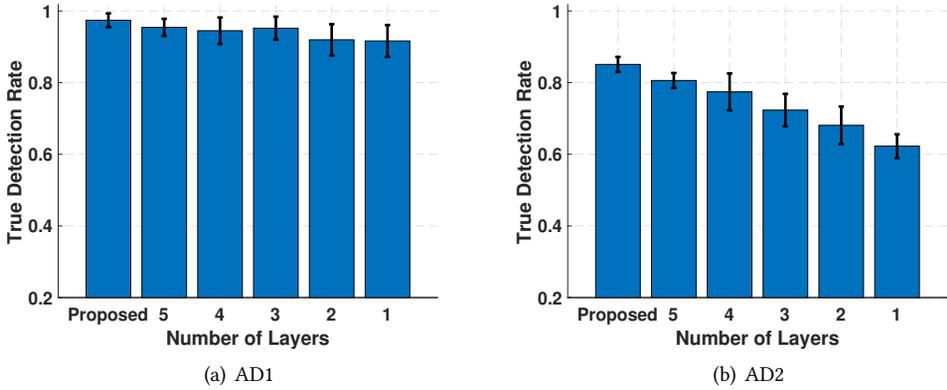


Fig. 16. The true detection rate of the proposed deep learning model with different number of Convolutional Layers

more challenging. There is an apparent trend of the accuracy decrease with the reduction of the number of convolutional layers, as shown in Figure 16(b). This confirms that the RFI introduces the extra challenges to the radio-based context awareness applications, and incorporating more convolutional layers can increase its performance of deep learning based context awareness model.

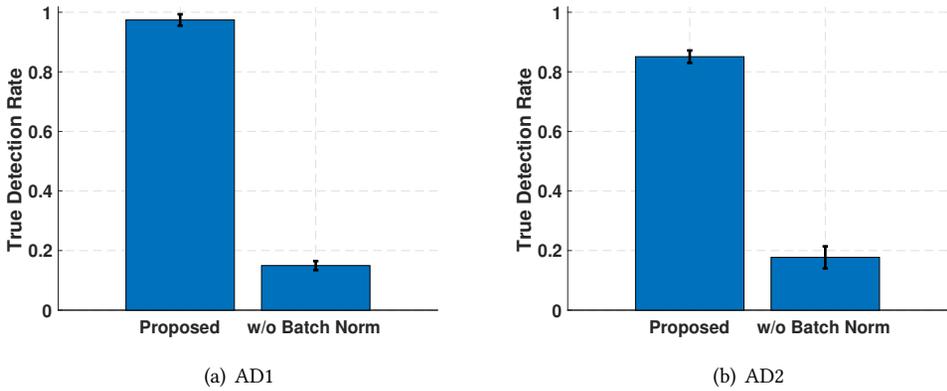


Fig. 17. The true detection rate of the proposed deep learning model with and without batch normalisation layers.

5.2.5 Effect of The Batch Normalisation Layer. It is observed from Section 5.1.5 that batch normalisation can significantly improve inference performance and accelerate the training process. This is further confirmed in Activity datasets that the batch normalisation layers can help analyse patterns in radio-based device-free context awareness applications. As shown in Figure 17, the accuracy is below 20% without using batch normalisation layers in both AD1 and AD2. When applying the batch normalisation layers along with convolutional layers, it is conspicuous that the accuracy increases more than 80% and 60% in AD1 and AD2, respectively.

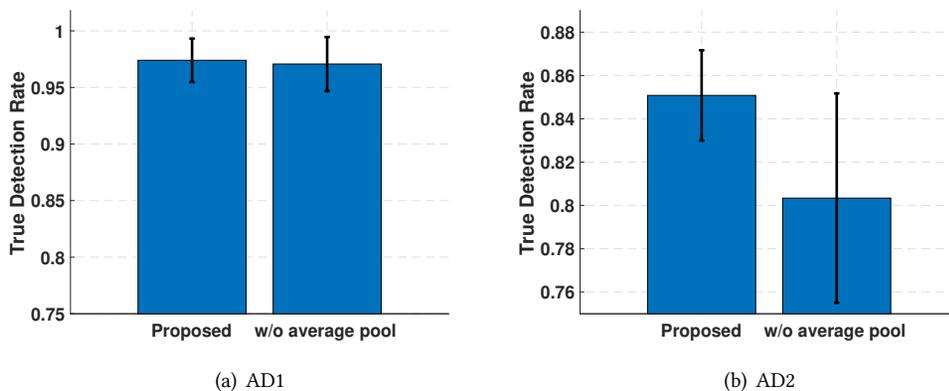


Fig. 18. The true detection rate of the proposed deep learning model with and without average pooling

5.2.6 Effect of The Average Pooling Layer. In the section, we will discuss the effect of average pooling layers. Figure 18 shows the true detection rates of activity recognition in AD1 and AD2 with and without average pooling layers. What can be clearly shown in Figure 18(b) is the dramatic decrease of the accuracy from 85.08% to 80.34% in AD2 without average pooling layers, while Figure 18(a) shows the accuracy remains steady in AD1 without RFI. This reveals that average pooling layers can improve the performance in the RFI environment.

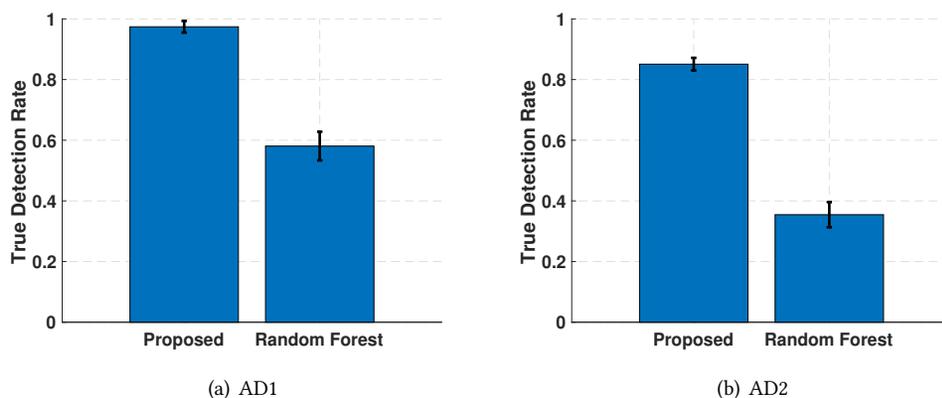


Fig. 19. The effect of deep learning

5.2.7 Effect of Deep Learning. In this section, we show the effect of deep learning in Activity dataset. The same as Section 5.1.7, we compare the proposed method with the random forest method. Figure 19 clearly shows the comparison between the proposed method and random forest. In AD1 and AD2, the random forest can only achieve 58.07% and 35.45% true detection rates, while the proposed method conspicuously implies its advantage with the accuracy of 97.40% and 85.08%, respectively.

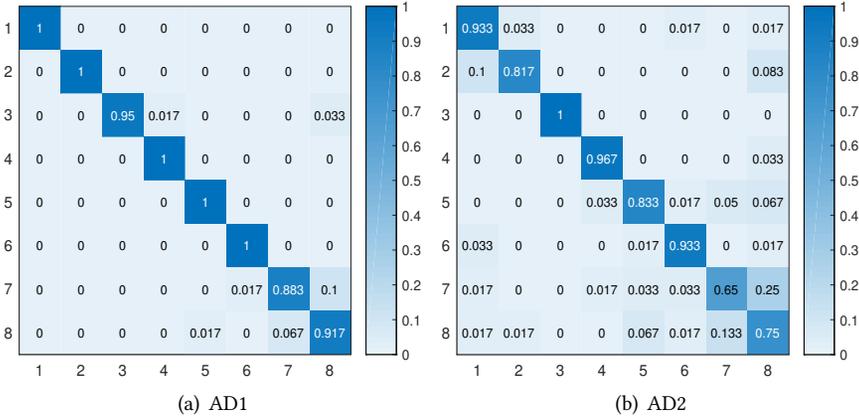


Fig. 20. Confusion matrix for AD1 and AD2

5.2.8 The performance in each class. As shown in Section 5.2.3, the proposed method can achieve equally good performance in AD2 and better performance in AD1. Figure 20 shows the confusion matrix using our proposed method on AD1 and AD2. It can be seen that static activities 1-6 (such as sitting, standing, etc) perform nearly perfect in AD1, while the performances of dynamic activities 7-8 (i.e. walking in two paths) are not as good as activities 1-6, but still have good true detection rates (i.e. 88.3% and 91.7%). When looking at the performance of AD2 in Figure 20(b), due to the radio interference, the performances of activities decrease, but activities 1-6 can still maintain good true detection rates with more than 80%. Activities 1, 3, 4, and 6 can achieve more than 90% true detection rates. We can see acceptable true detection rates (i.e. 65.0% and 75.0%) in dynamic activities 7-8 albeit a further reduction compared with AD1 due to radio interference.

5.3 Summary of Evaluation

The proposed framework shows better performance in both SignFi dataset and Activity dataset, which thus confirms its generalisation. The datasets for these application scenarios have various data dimension and signal quality. Therefore, different configurations are used for these two application scenarios. However, they both use the proposed general framework with different numbers of convolutional layers, dense layers, kernel sizes, etc. CSI data in Activity dataset are with radio frequency interference, which has a more complex situation. Therefore, more convolutional layers, dense layers, and larger kernel size are employed to improve the recognition performance.

The same as the majority of CSI based device-free context awareness methods, this paper explores a supervised machine learning method. This means the system can recognise context with the training set collected in various scenarios. In most radio-based device-free context awareness applications, one person is usually monitored[13, 25, 30–32, 40], and radio monitoring devices are used instead of carers or monitors. For example, in this paper, we have discussed two application scenarios, i.e. sign language recognition and activity recognition. In the first application scenario, sign language recognition, radio monitoring system is set up instead of a sign language interpreter. Activity recognition in the second application scenario is used to assist or replace carers. Therefore, we use the same mechanism as [13, 25, 30–32, 40] for designing experiments so that scenarios with multiple people are not considered when collecting training data. When considering a newly involved scenario, e.g. two users doing activities or gestures within the area of interest, the system

can still use the proposed framework for training and inference. The only adjustment is the collection of CSI data within that scenario with multiple people and re-tune the DNN model.

6 CONCLUSION

In this paper, we have proposed an innovative DNN based framework for radio-based device-free context awareness applications. We are the first to design a general framework of radio-based based applications, and we have shown the generalisation using extensive experiments. Without any data pre-processing, our proposed framework can still achieve equivalent or better performance. We have used public radio-based gesture recognition datasets and activity recognition datasets to evaluate our proposed model, which results in approximately 100% accuracy for recognising 276 gestures and 85% accuracy for activity recognition with RFI. We have also shown the capability of our proposed framework for radio-based applications in RFI environments. Finally, we have discussed the design of DNN structures using our proposed general framework and shown the effects of different layers in the DNN models in both datasets.

REFERENCES

- [1] Fadel Adib, Zach Kabelac, Dina Katabi, and Robert C. Miller. 2014. 3D Tracking via Body Radio Reflections. In *NSDI '14*. Seattle, WA.
- [2] Fadel Adib and Dina Katabi. 2013. See Through Walls with WiFi!. In *SIGCOMM '13* (Hong Kong, China). ACM, New York, NY, USA, 75–86.
- [3] Sourav Bhattacharya and Nicholas D Lane. 2016. From smart to deep: Robust activity recognition on smartwatches using deep learning. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. IEEE, 1–6.
- [4] Yu Gu, Jinhai Zhan, Yusheng Ji, Jie Li, Fuji Ren, and Shangbing Gao. 2017. MoSense: An RF-based motion detection system via off-the-shelf WiFi devices. *IEEE Internet of Things Journal* 4, 6 (2017), 2326–2341.
- [5] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. 2011. Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM Computer Communication Review* 41, 1 (2011), 53–53.
- [6] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi, and Hong-Jiang Zhang. 2005. Face recognition using laplacian-faces. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 3 (2005), 328–340.
- [7] Jawbone. 2014. UP. <https://jawbone.com/up>. [Online; accessed 28-August-2014].
- [8] Ossi Kaltiokallio, Maurizio Bocca, and Neal Patwari. 2012. Enhancing the accuracy of radio tomographic imaging using channel diversity. In *IEEE MASS 2012*. IEEE, 254–262.
- [9] Bryce Kellogg, Vamsi Talla, and Shyamnath Gollakota. 2014. Bringing Gesture Recognition to All Devices. In *NSDI 14*. USENIX, Seattle, WA.
- [10] Keun-Chang Kwak and Witold Pedrycz. 2005. Face recognition using a fuzzy fisherface classifier. *Pattern recognition* 38, 10 (2005), 1717–1732.
- [11] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436.
- [12] Jian Liu, Yingying Chen, Yan Wang, Xu Chen, Jerry Cheng, and Jie Yang. 2018. Monitoring vital signs and postures during sleep using WiFi signals. *IEEE Internet of Things Journal* 5, 3 (2018), 2071–2084.
- [13] Yongsan Ma, Gang Zhou, Shuangquan Wang, Hongyang Zhao, and Woosub Jung. 2018. SignFi: Sign Language Recognition Using WiFi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 23.
- [14] Pedro Melgarejo, Xinyu Zhang, Parameswaran Ramanathan, and David Chu. 2014. Leveraging directional antenna capabilities for fine-grained gesture recognition. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 541–551.
- [15] Microsoft. 2014. Kinect. <http://www.microsoft.com/en-us/kinectforwindows/>. [Online; accessed 28-August-2014].
- [16] Tom M. Mitchell. 1997. *Machine Learning*. McGraw-Hill, New York.
- [17] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al. 2015. Deep face recognition.. In *bmvc*, Vol. 1. 6.
- [18] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. 2013. Whole-home gesture recognition using wireless signals. In *MobiCom 2013*. 27–38.
- [19] Kun Qian, Chenshu Wu, Zimu Zhou, Yue Zheng, Zheng Yang, and Yunhao Liu. 2017. Inferring motion direction using commodity wi-fi for interactive exergames. In *CHI 2017*. ACM, 1961–1972.
- [20] Souvik Sen, Božidar Radunovic, Romit Roy Choudhury, and Tom Minka. 2012. You are facing the Mona Lisa: spot localization using PHY layer information. In *MobiSys 2012*. ACM, 183–196.

- [21] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- [22] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* 15, 1 (2014), 1929–1958.
- [23] Jiancheng Sun, Chongxun Zheng, Xiaohe Li, and Yatong Zhou. 2010. Analysis of the distance between two classes for tuning SVM hyperparameters. *IEEE transactions on neural networks* 21, 2 (2010), 305–318.
- [24] Fangxin Wang, Wei Gong, and Jiangchuan Liu. 2018. On Spatial Diversity in WiFi-based Human Activity Recognition: A Deep Learning based Approach. *IEEE Internet of Things Journal* (2018).
- [25] Guanhua Wang, Yongpan Zou, Zimu Zhou, Kaishun Wu, and Lionel M Ni. 2014. We can hear you with Wi-Fi!. In *MobiCom 2014*. ACM, 593–604.
- [26] Wei Wang, Alex X Liu, and Muhammad Shahzad. 2016. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 363–373.
- [27] Xuyu Wang, Lingjun Gao, Shiwen Mao, and Santosh Pandey. 2015. DeepFi: Deep learning for indoor fingerprinting using channel state information. In *2015 IEEE wireless communications and networking conference (WCNC)*. IEEE, 1666–1671.
- [28] Xuyu Wang, Lingjun Gao, Shiwen Mao, and Santosh Pandey. 2017. CSI-based fingerprinting for indoor localization: A deep learning approach. *IEEE Transactions on Vehicular Technology* 66, 1 (2017), 763–776.
- [29] Xuyu Wang, Chao Yang, and Shiwen Mao. 2017. PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 1230–1239.
- [30] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. 2014. E-eyes: device-free location-oriented activity identification using fine-grained WiFi signatures. In *MobiCom 2014*. ACM, 617–628.
- [31] Bo Wei, Wen Hu, Mingrui Yang, and Chun Tung Chou. 2015. Radio-based device-free activity recognition with radio frequency interference. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*. ACM, 154–165.
- [32] Bo Wei, Wen Hu, Mingrui Yang, and Chun Tung Chou. 2019. From Real to Complex: Enhancing Radio-based 22 Activity Recognition Using Complex-Valued CSI. *ACM Transactions on Sensor Networks* 15, 3 (2019).
- [33] Bo Wei, Ambuj Varshney, Neal Patwari, Wen Hu, Thiemo Voigt, Chou, and Chun Tung. 2015. dRTI: Directional Radio Tomography. In *IPSN '15 (Seattle, WA, USA)*. ACM, Seattle, WA, USA, 12.
- [34] J. Wilson and N. Patwari. 2010. Radio Tomographic Imaging with Wireless Networks. *IEEE TMC* 9, 5 (2010), 621–632. <https://doi.org/10.1109/TMC.2009.174>
- [35] Xiaomi. 2014. Mi Band. <http://www.mi.com/shouhuan>. [Online; accessed 28-August-2014].
- [36] Yaxiong Xie, Zhenjiang Li, and Mo Li. 2015. Precise Power Delay Profiling with Commodity WiFi. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (Paris, France) (MobiCom '15)*. ACM, New York, NY, USA, 53–64. <https://doi.org/10.1145/2789168.2790124>
- [37] Chenren Xu, Mingchen Gao, Bernhard Firner, Yanyong Zhang, Richard Howard, and Jun Li. 2012. Towards Robust Device-Free Passive Localization Through Automatic Camera-Assisted Recalibration. In *ACM SenSys*.
- [38] Yunze Zeng, Parth H Pathak, and Prasant Mohapatra. 2016. WiWho: wifi-based person identification in smart spaces. In *Proceedings of the 15th International Conference on Information Processing in Sensor Networks*. IEEE Press, 4.
- [39] Dongheng Zhang, Yang Hu, Yan Chen, and Bing Zeng. 2019. BreathTrack: Tracking Indoor Human Breath Status via Commodity WiFi. *IEEE Internet of Things Journal* (2019).
- [40] Jin Zhang, Bo Wei, Wen Hu, and Salil S Kanhere. 2016. Wifi-id: Human identification using wifi signal. In *Distributed Computing in Sensor Systems (DCOSS), 2016 International Conference on*. IEEE, 75–82.
- [41] Jin Zhang, Bo Wei, Fuxiang Wu, Limeng Dong, Wen Hu, Salil S Kanhere, Chengwen Luo, Shui Yu, and Jun Cheng. 2020. Gate-ID: WiFi-based Human Identification Irrespective of Walking Directions in Smart Home. *IEEE Internet of Things Journal* (2020).
- [42] Jin Zhang, Fuxiang Wu, Bo Wei, Qieshi Zhang, Hui Huang, Syed W Shah, and Jun Cheng. 2020. Data augmentation and dense-1stm for human activity recognition using wifi signal. *IEEE Internet of Things Journal* (2020).
- [43] Jun Zhang, Yong Yan, and Martin Lades. 1997. Face recognition: eigenface, elastic matching, and neural nets. *Proc. IEEE* 85, 9 (1997), 1423–1435.
- [44] Mingmin Zhao, Fadel Adib, and Dina Katabi. 2016. Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM, 95–108.
- [45] Yang Zhao and N. Patwari. 2011. Noise reduction for variance-based device-free localization and tracking. In *SECON 2011*. 179–187. <https://doi.org/10.1109/SAHCN.2011.5984895>
- [46] Yang Zhao, Neal Patwari, Jeff M. Phillips, and Suresh Venkatasubramanian. 2013. Radio Tomographic Imaging and Tracking of Stationary and Moving People via Kernel Distance. In *IPSN '13 (Philadelphia, Pennsylvania, USA)*. ACM, New York, NY, USA, 229–240. <https://doi.org/10.1145/2461381.2461410>

- [47] Zimu Zhou, Zheng Yang, Chenshu Wu, Longfei Shangguan, and Yunhao Liu. 2013. Omnidirectional Coverage for Device-free Passive Human Detection. *IEEE TPDS* (2013).

Received February 2007; revised March 2009; accepted June 2009