Multi-Adversarial In-Car Activity Recognition Using RFIDs

Fangxin Wang, *Student Member, IEEE*, Jiangchuan Liu, *Fellow, IEEE*, and Wei Gong, *Member, IEEE*

Abstract—In-car human activity recognition opens a new opportunity toward intelligent driving behavior detection and touchless human-car interaction. Among the many sensing technologies (e.g., using cameras and wearable sensors), radio frequency identification (RFID) exhibits unique advantages given its low cost, easy deployment, and less privacy concerns. Existing RFID-based solutions for activity recognition are mostly confined to working in stable indoor spaces. The inside space of a car however is much more compact and complex, not to mention the fast-changing driving conditions. All these introduce non-negligible noises that pollute the activity-related information, and the existence of various car models in the market further complicates the problem. In this article, we for the first time closely examine the distinct factors that affect the RFID-based in-car activity recognition. We present *RF-CAR*, a novel RFID-based tag-free solution that well adapts to different in-car environments. RF-CAR smartly filters the domain-specific features in RF signals and retains activity-related features to the maximum extent. It then integrates a deep learning architecture and an advanced multi-adversarial domain adaptation network for training and prediction. With only one-time pre-training, RF-CAR can adapt to new data domains such as new driving conditions, car models, and human subjects for robust activity recognition. We also demonstrate that it is readily deployable in cars with commercial off-the-shelf (COTS) RFID devices. Our extensive experiments suggest that RF-CAR achieves an overall recognition accuracy of around 95 percent, which significantly outperforms the state-of-the-art solutions.

Index Terms—Activity recognition, RFID based sensing, domain adversarial network, deep learning

1 Introduction

Human activity recognition in cars plays a key role towards safe driving [1] and human-car interaction [2]. It can effectively remind drivers of distracting behaviors, such as forgetting shoulder check; it is also an essential building block for in-car entertainment, especially in the emerging autonomous driving scenarios where people can enjoy convenient gesture-based control. Existing works on in-car activity recognition are mostly camera-based [3]. They not only require line-of-sight visual sensing but also have high risks of user privacy leakage. Other sensor-based approaches rely on wearable devices or smartphones, which is inconvenient for practical usage. Recently, there have been significant studies on activity recognition with wireless sensing, given the observation that different activities can affect the surrounding wireless signals in different ways [4], [5]. Radio Frequency Identification (RFID) [6], [7], [8], [9] is of particular interest for its low tag costs and the batteryless nature of the tags. Using multiple tags also facilitates the creation of multiple observation paths, so as to achieve higher reliability.

RFID-based activity recognition in the literature has concentrated on the indoor environment, with both tag-based [6], [8] and tag-free solutions [10], [11], [12]. In-car recognition however incurs a series of new challenges, as illustrated in Fig. 1. First, different from a stationary indoor space, the external driving conditions change fast. The reflected RF signals from external moving objects such as cars and pedestrians can cause unexpected signal fluctuations, affecting the recognition accuracy. Second, the facilities, items, and interior space of different car models are highly heterogeneous and can incur quite distinct multipath effects [13], [14] in RF signals, degrading the recognition accuracy differently. Third, people of different genders, heights, shapes may perform the same activity with different frequencies and amplitudes, which contributes to additional recognition error. As a result, a well-trained model in one domain-specific situation (e.g., one specific driving condition, car model and human subject) may not work effectively in another domain. This renders the traditional RFID-based recognition system infeasible for in-car scenarios given the innumerable domain-specific situations.

To address these challenges, we for the first time closely examine the impact of different driving conditions, car models, and human subjects on RFID-based in-car activity recognition. In this paper, we present *RF-CAR*, an RFID-based tag-free in-car activity recognition framework that can filter the domain-specific features in RF signals while retaining the activity-related features to the maximum extent. Our framework can apply to new driving conditions, car models, and human subjects after just one-time pre-training.

Given the complex raw data from RFID, we first carefully extract the time and frequency features from the original signals with a preprocessing step. Such features are next fed

- *F. Wang and J. Liu are with the School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. E-mail: {fangxinw, jcliu}@sfu.ca.*
- *W. Gong is with the School of Computer Science and Technology, University of Science and Technology of China, Hefei 230052, China. E-mail: weigong@ustc.edu.cn.*

Manuscript received 12 Apr. 2019; revised 17 Dec. 2019; accepted 18 Feb. 2020. Date of publication 3 Mar. 2020; date of current version 5 May 2021. (Corresponding author: Jiangchuan Liu.)
Digital Object Identifier no. 10.1109/TMC.2020.2977902

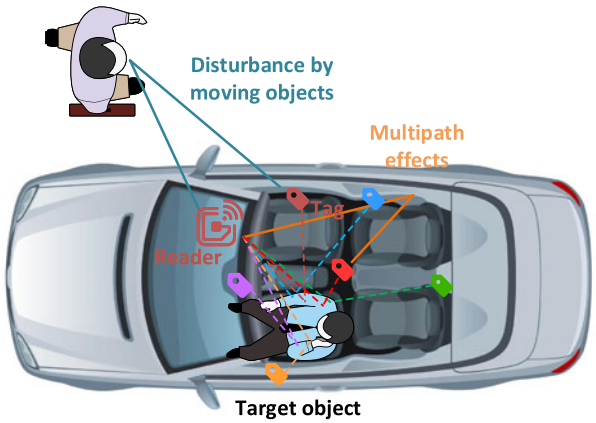


Fig. 1. The illustration of multiple impacts on RFID-based in-car activity recognition.

into a deep learning model, which mainly consists of three components: a *feature encoder*, an *activity classifier*, and a set of *domain discriminators*. The feature encoder employs a stacked convolutional neural network (CNN) to capture the dynamics in both frequency and time dimensions from multiple RF links. It plays a cooperative game with the activity classifier to achieve high activity recognition and simultaneously plays a minimax game to prevent the domain discriminators from classifying different domains. Pioneer works [15], [16] only consider single domain discriminator in indoor scenarios. Yet for the in-car scenario, the complex discriminative structures of different domains can be easily mixed up using such solutions [17], leading to false domain discrimination. RF-CAR tackles the problem through a multi-adversarial domain adaptation that is optimized for the car context.

We have prototyped RF-CAR using a commercial off-the-shelf (COTS) RFID reader with multiple tags in vehicles. We have conducted real-world experiments on 8 common in-car activities, involving 6 different driving conditions, 6 different car models, and 8 volunteers. The evaluation results report an average in-car activity recognition accuracy of around 95 percent for RF-CAR, which is much higher than that of state-of-the-art solutions with no adversarial network (67 percent) or with single-adversarial domain adaptation (86 percent).

The rest of this paper is organized as follows. We conduct experiments to examine the impact factors in Section 2. We present the system overview in Section 3. Preprocessing steps are introduced in Section 4. We describe the domain adversarial model in Section 5, followed by an evaluation in Section 6. We introduce the related work in Section 7. We discuss our work in Section 8 and conclude it in Section 9.

2 Motivation

A typical RFID sensing system consists of a UHF RFID reader and one or more tags, where tags are powered by the received RF signals and then backscatter the signals to the reader. Existing COTS readers, such as Impinj Speedway R420[1] and Thingmagic M6e,[2] can report such raw data as

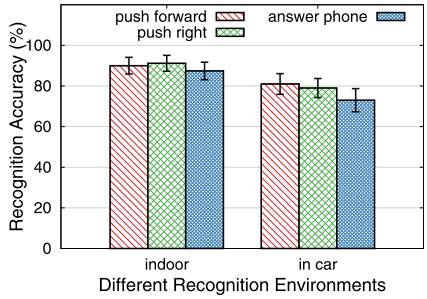


Fig. 2. Recognition performance within the indoor environment and the in-car environment using traditional RFree-ID approach.

received signal strength indicator (RSSI) and phase angle through their APIs. The RSSI has been explored earlier as an indicator to distinguish different activities [18], while the phase information has been more widely used in recent studies given its higher reliability and robustness [10], [12]. In the in-car activity recognition scenario, we fix multiple tags to the in-car surfaces (e.g., doors and seats) to construct multiple sensing links with the reader. The received phase readings of each link will be affected by the activities around, in particular, the in-car activities of the driver or passengers,[3] which can then be recognized through analyzing the changing signal metrics.

The captured phase patterns however can be extremely noisy in in-car environments compared to indoor environments since the in-car space is naturally much more compact, incurring serious multipath effects. We use a state-of-the-art activity recognition approach, RFree-ID [10], to examine the overall recognition accuracy in the in-car environment and indoor environment, respectively. Unless specified, we RFree-ID to study the performance of traditional RFID-based approaches. Fig. 2 illustrates the general recognition performance for three typical activities (e.g., pushing forward, pushing right, and answering phone) using Rfree-ID. We can find that the overall recognition accuracy is about 92 percent in the indoor environment, while it drops to below 80 percent in the in-car environment. This result indicates that existing solutions can render reasonably good performance in the indoor environment, while they are not so effective in the in-car environment.

Besides the impact of rich multipaths in in-car scenarios, there also exist other impact factors that can affect RF signals and hence the activity recognition performance. We mainly consider three most important impact factors, i.e., different driving conditions, human subjects and car models, each of which is referred to as a *class*. A certain combination of such classes (i.e., impact factors) forms a *domain* [19], where domain-specific information is included, affecting the activity recognition. We next conduct real-world experiments to study the impact of these impact factors.

We first consider the impact of the driving conditions, particularly the outside pedestrians and other cars, which are also reflectors of RFID signals. Although the outside impact is weaker than the inside objects, it is not negligible, especially in a congested environment. To demonstrate this, we select several typical driving conditions (different from

1. https://support.impinj.com/
2. https://www.jadaktech.com/

3. We mainly focus on the single driver activity recognition in this paper. The multi-person recognition scenario is discussed in Section 8.

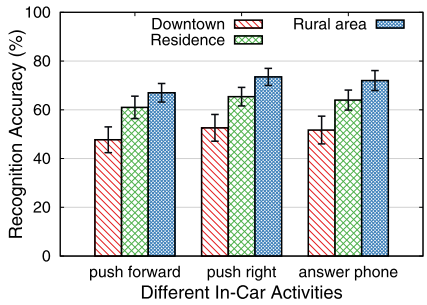


Fig. 3. Activity recognition accuracy in different driving conditions using traditional RFree-ID approach.

the training conditions) with different congestion levels to test the recognition performance on some common in-car activities using RFree-ID [10]. As illustrated in Fig. 3, we can see that the overall accuracy in the rural area is only 72 percent, while the accuracy in downtown further drops to below 45 percent. Intuitively, downtown is more crowded with people and cars, and hence with stronger interference from the environment, which overwhelms the useful data features for activity recognition. Such external interference unfortunately has not been well studied in the literature, which typically assumes a stationary indoor environment with internal interference only.

We next examine the impact of human subjects, which is probably the most important factor. The people in a car can vary in ages, heights, genders, and weights. As such, the same activity performed by different people can have different frequencies, durations, angles, and ranges, yielding diverse phase patterns and even opposite recognition results. We ask different volunteers to perform the same activity multiple times while keeping other configurations consistent and collect the RF signals. As in Fig. 4, the collected phase shift patterns are similar for samples of the same volunteer. However, those patterns are quite different for samples of different volunteers. Since the driver for a car is unknown in advance and the potential drivers are countless, simply using traditional phase and Doppler-based approaches (e.g., [8], [10]) may not work perfectly. Data gathering and training have to be re-done whenever there is a new driver, which is quite costly.

The car model is another key impact factor since different models can have quite diverse spaces, shapes, and interior decorations. An in-car space is much more compact than a typical indoor environment so that a slight difference can lead to dramatic pattern changes in the phase measurement, given the rich reflection, diffraction, and absorption [20]. We

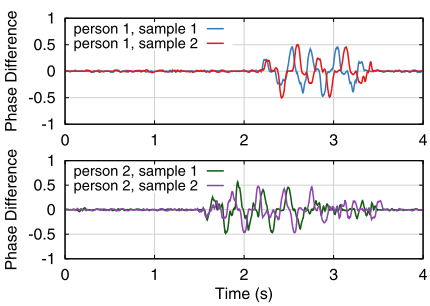


Fig. 4. The processed phase difference when different human subjects performs the same activity with other configurations consistent.

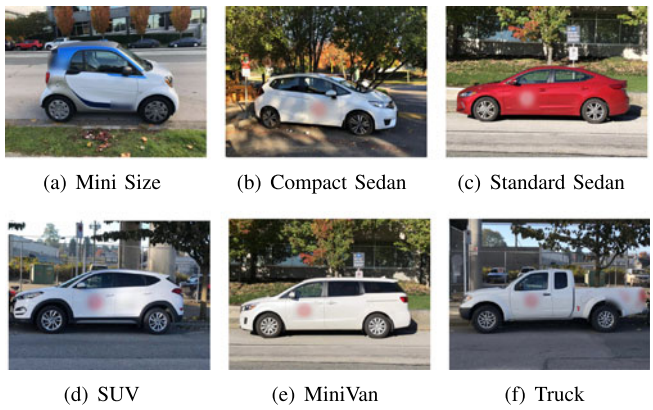


Fig. 5. Different car models we consider in our experiments.

examine such impacts using 6 different car models, including a mini size car (Benz Smart), a compact sedan (Honda Fit), a standard sedan (Hyundai Elantra), an SUV (Hyundai Tucson), a minivan (Kia Sedona), and a pickup truck (Nissan Frontier), as illustrated in Fig. 5. Similar to previous works for indoor environment [20], we use the dynamic time warping (DTW) distance to quantitatively describe the dissimilarities of the phase shift patterns when the same volunteer is performing an activity in two different cars and the same car, respectively. Intuitively, a smaller DTW distance indicates a higher similarity in activity features. For each car, we randomly select 20 samples and compare the DTW distance between them in Fig. 6. We also measure the DTW distance for the same car. Clearly, different cars have quite different measurements for the same activities, with the average DTW distance increased by 45 percent more than that of the same car.

In summary, our real-world experiment suggests that a practical RFID-based in-car activity recognition system must adapt to the highly varying driving conditions, human subjects and car models. Yet the existing solutions work only in well-trained domains; once the domain changes, costly retraining has to be invoked or the recognition performance will degrade dramatically. It is necessary to develop new flexible and robust solutions to deal with the extraneous implicit information in activity data that contaminates the activity-related features in different domains.

3 RF-CAR: High-Level System Overview

We present RF-CAR, a robust environment-independent system for RFID-based in-car activity recognition. With one-time

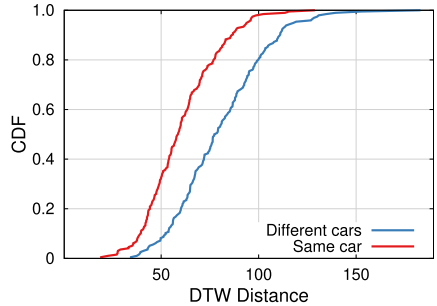


Fig. 6. The CDF plot of DTW distance when the same person performs the same activity in different cars.

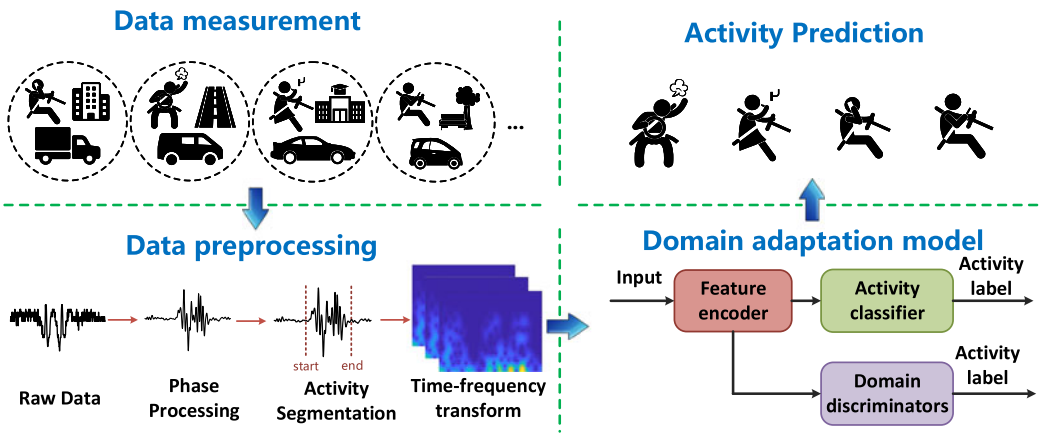


Fig. 7. The framework of RF-CAR.

training, it works for diverse driving conditions, human subjects, and car models. We integrate a customized RFID signal processing and a novel multi-adversarial domain adaptation network for training and prediction. RF-CAR consists of three cascaded components, namely, data measurement, data preprocessing, and domain adaptation model, as in Fig. 7.

Data Measurement. We use a UHF RFID reader with multiple antennas and commercial tags (four and six, respectively, in our current implementation) deployed in cars for activity sensing. The collected phase data of each sensing link through the system APIs are streamed to the backend module for further preprocessing and recognition.

Data Preprocessing. The raw phase data needs to be transformed into proper feature representations before training. We first conduct phase unwrapping and smoothing, where the Doppler frequency shift is derived thereafter. Principal component analysis (PCA) is then applied to extract the shared characteristics from multiple antennas. We leverage a KL-divergence-based algorithm to detect and segment each activity sample. At last, we use a short-time Fourier transform (STFT) to obtain the feature spectrogram on both time dimensions and frequency dimensions.

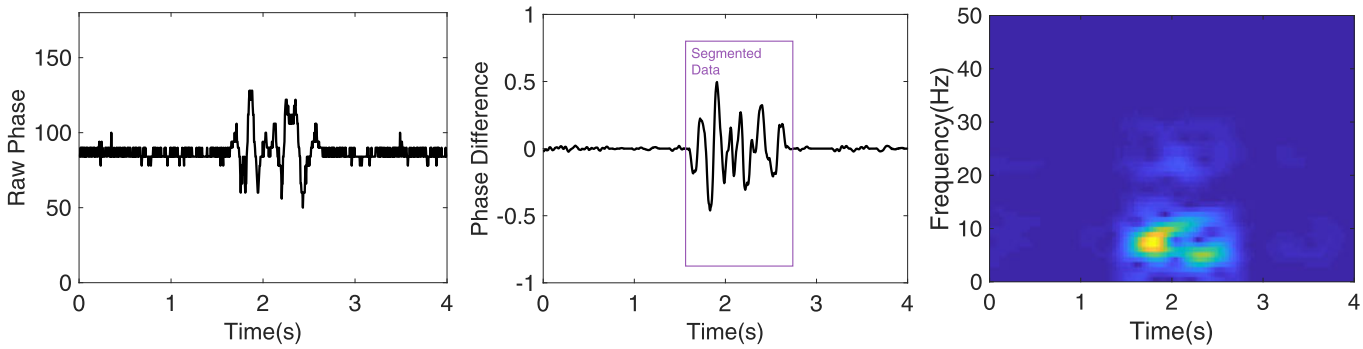
Domain Adaptation Model. As demonstrated earlier, in the in-car environment, the derived feature spectrogram can be complex with much extraneous information. We design a deep-learning-based domain adaptation model to effectively remove the domain-specific information while retaining the activity-related features. A CNN-based feature encoder cooperates with an activity classifier to maximize the activity recognition accuracy. To prevent the model from extracting domain-related information, a set of domain discriminators then work against the feature encoder. Note that our domain adaptation model is able to learn the transferred features even when the data of the target domain is unlabelled. Through such a domain adaptation approach, the trained model can be well applied to other untrained domains such as new cars and new drivers for activity recognition.

4 Data Preprocessing

In this section, we give a detailed demonstration of our preprocessing steps before the model training and inference, including phase processing, activity segmentation, and feature representation.

4.1 Phase Processing

In a practical wireless environment, the received RF signals contain multipath components (MPCs) due to the multipath effect. MPCs can arrive at a receiver through such various propagation mechanisms as reflection, diffraction, and scattering [20]. The multipath effect is strong in the compact in-car space. Fig. 8a shows the raw phase of a pushing activity reported by an RFID reader Thingmagic M6e in a car (Hyundai Tucson). When there is an activity nearby, such components can be classified into a superposition of static components $s_s(t)$ and dynamic components $s_d(t)$



(a) The phase pattern of raw signals. (b) The phase difference of processed signal. (c) Spectrogram of STFT.

Fig. 8. The result of data preprocessing.

$$s(t) = s_s(t) + s_d(t) = s_s(t) + \sum_{k=1}^{N} \alpha_k(t) e^{-j2\pi f \tau_k}, \quad (1)$$

where N is the number of propagation paths, $\alpha_k(t)$ includes the signal attenuation and the initial phase offset in the complex form, and $e^{-j2\pi f \tau_k}$ denotes the phase shift through the corresponding propagation path k with a delay τ_k on carrier frequency f. According to Euler formula, Eq. (1) can be transformed into a trigonometric form and we can get the angle

$$\angle s(t) = \arctan \frac{|s_s(t)| \sin \angle s_s(t) + |s_d(t)| \sin \angle s_d(t)}{|s_s(t)| \cos \angle s_s(t) + |s_d(t)| \cos \angle s_d(t)}, \quad (2)$$

where $|\cdot|$ and $\angle\cdot$ represent the amplitude and phase, respectively. From the above two equations we can see that any signal propagation path length change is highly correlated with the phase change [12], which we use as the key signal metrics for activity recognition.

The raw signal phases reported by the reader are confined to a periodic function ranging from 0 to 2π, which cannot be directly applied for processing. We therefore first need to unwrap the phase to correct values. We adopt the commonly used One-Dimensional Phase Unwrapping method [21] to restore the original phase values.

After the phase unwrapping process, we then smooth the phase values since the random access mechanism of RFID reading can lead to signal impulse and inconsistent time interval. We first apply the Hampel identifier to detect the outliers and filter them out from the processed phase. We then use a Savitzky-Golay filter [22] to remove the random noise and smooth the data. It is a widely used weighted moving average filter, which is able to effectively increase the signal-to-noise ratio with limited distortion to the signal.

In a practical RFID deployment, a reader can have multiple antennas closely placed as an array (e.g., 4 in our deployment), where the backscatter links between the antenna array and one tag usually share similar wave patterns. Given the correlations therein, we apply principal component analysis to the processed phase data of these links to extract the common characteristics. For each tag, we use PCA to obtain multiple feature components. We have randomly selected more than 100 samples and find that the percentage of the top two components account for more than 90.5 percent of the total captured variance. And from the third PCA component, the retained features with low resolution of different frequency component values are not sufficient enough for a good classification. We therefore use the top two PCA components for analysis. The average of them are referred to as the representative phase stream, denoted as p-stream. The individual phase value actually does not show specific meaning after unwrapping, yet the phase difference of two consecutive sample points indicates how fast the surrounding activity is performed. We next calculate the phase difference of each p-stream as shown in Fig. 8b, denoted as dp-stream for further processing.

4.2 Activity Segmentation

Given the phase difference, the next step is to detect whether there exists an in-car activity and segment the activity related samples from the whole data sequences. From Fig. 8b we know that the phase value keeps relatively stable in absence of human activities, while the variance increases dramatically with activities. Since different activities can have diverse time durations and gaps, a dynamic activity segmentation approach is required for adaptive and real-time activity sensing. Motivated by such observations, we use the Kullback-Leibler divergence (KL-divergence) [23] to detect and segment the effective human activities from the dp-stream. The KL-divergence is a measure of how one probability distribution is different from another distribution. The key insight here is that the KL value of two stationary samples are very small, while the KL value of two activity samples or one stationary sample and one activity sample will be significantly higher.

In our activity segmentation method, we first segment the dp-stream data into frames of every 0.5 second with 50 percent overlap, which balances both the granularity and computational efficiency. By calculating the KL-divergence value of each consecutive frames, we can easily ascertain whether a frame is belonging to samples in absence of activities. Naturally, the rest frames belong to samples in presence of activities.

4.3 Spectrogram Generation

The segmented frame contains the wave patterns of the collected phase data. Yet it is still not a good feature representation for the subsequent deep learning since it only shows the phase difference without exhibiting the phase changing frequency explicitly. To this end, we take advantage of time-frequency transform tools to reveal the frequency features. RF-CAR uses short-time Fourier transform on every segmented sample to extract the frequency characteristics and converts the time domain wave patterns to time-frequency spectrograms. Instead of using the discrete wavelet transform (DWT) as most existing systems [24], STFT enables more fine-grained and homogeneous resolution on the frequency domain, which is particularly suitable for in-car activity recognition because spectrograms with higher resolution can provide more details for the learning model. Fig. 8c illustrates the spectrogram of a pushing activity. We can see that the corresponding part has an obvious frequency increase when a person is performing the motion, and the frequency component drops when the motion ends.

Note that our RFID-based sensing system can easily support multiple tags, where the sensing links between the antenna array and each tag generate a representation spectrogram. With multiple tags deployed in a car, these sensing links can observe activities from multiple angles, and the spatial diversity therein can greatly improve the robustness and accuracy. We then concatenate the generated spectrogram data from multiple links together and feed such input to the learning model for further processing.

5 Multi-Adversarial Domain Adaptation

We describe the design of our domain adaptation model for activity recognition in this section. As illustrated in Fig. 9, the learning model mainly includes three components, a *feature encoder*, an *activity classifier* and a set of *domain discriminators*. This learning model is designed to only capture the activity-specific features while filtering out those domain-specific

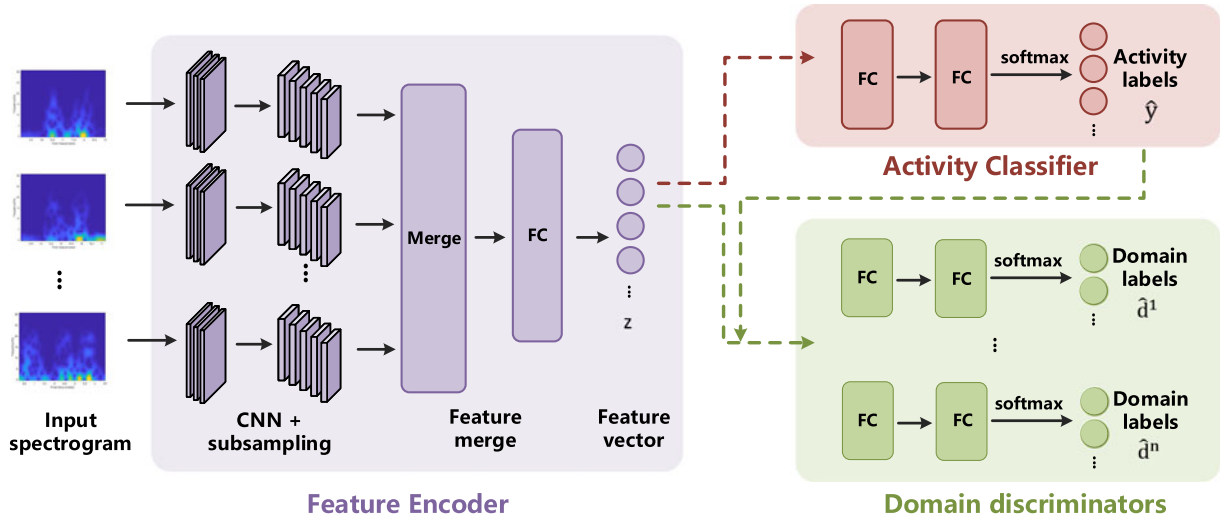


Fig. 9. The main components of the learning model.

features, so as to achieve domain-adaptive activity recognition. To do so, the activity classifier aims to boost the performance of the activity prediction, and the domain discriminators aim to discriminate the target domains from the source domains. The feature encoder however is designed to cooperate with the activity classifier to minimize the loss of activity prediction and play against the domain discriminators to maximize the loss of the domain prediction. Through this way, we are training the feature encoder to be domain adaptive. After enough training, the feature encoder and activity classifier can be used to predict labels from target domains.

5.1 Feature Encoder

We denote $x_i \in X$ as an input data spectrogram associated with tag i and X is the whole input space. Each input data spectrogram has an activity label $y_i \in Y$ indicating the current activity, where Y is the label space. Besides the activity label, each x_i also has a domain label from the domain space D, which consists of the source domain space D_s with labels and the target domain space D_t without labels.

The goal of the feature encoder G_f is to take the spectrograms as input and transform the original complex high dimensional features to low dimensional feature representations $\mathbf{z}$. Given that CNN has exhibited powerful abilities in extracting the spatial relationships from images, we adopt a CNN-based architecture to extract the features from the image-like structures in our input spectrogram. Note that the CNN model requires a uniform input feature representation. Yet the spectrogram transformed from each segmented data samples can have diverse time lengths with different activity durations. In our experiment, the window size is empirically set as 5 seconds, which is enough to include most activity features. For those continuous activities such as making phone call and texting, we mainly detect the start of the motion, which is also less than 5 seconds. We fill shorter activities with padding zeros to guarantee the fixed time length.

In the feature encoder, we stack two CNN layers to process each input spectrograms, each followed with a rectified linear unit (ReLU) layer as the activation function. Max pooling layers are also applied to reduce the feature dimensions. Since the extracted features from these multiple input spectrograms are of the same dimension, we can easily merge the results with fully connected networks and obtain the feature vector with a multi-view observation. The output is represented as

$$\mathbf{z} = G_f(\mathbf{x}; \theta_f), \tag{3}$$

where θ_f represents all the parameters of the feature encoding layers.

5.2 Activity Classifier

The output feature z is fed into the activity classifier G_y as input. We use two fully connected layers with activation layers to learn the discriminative activity-related features. We then put a softmax layer at last to map the features to a latent space with the same size as the activity label space. In this way, we can represent the predicted activity label distribution probabilities for input $\mathbf{x}$ as

$$\hat{\mathbf{y}} = G_y(G_f(\mathbf{x}; \theta_f); \theta_y), \tag{4}$$

where θ_y denotes all the parameters in the activity classifier.

The integrated loss function of G_f and G_y can be obtained by calculating the cross-entropy function between the actual labels and the predicted labels as

$$\begin{aligned} \mathcal{L}_y(G_f, G_y) &= \mathbb{E}_{\mathbf{x},y}[-log\ G_y(G_f(\mathbf{x}; \theta_f); \theta_y)] \\ &= -\frac{1}{|D_s|} \sum_{x_i \in D_s} \sum_{j=1}^{M} y_{ij}\ log(G_y(G_f(x_i; \theta_f); \theta_y)), \end{aligned} \tag{5}$$

where M is the number of activities labels and $|D_s|$ is the number of samples in source domains. In the training process, the feature encoder G_f cooperates with the activity classifier G_y to minimize the label prediction loss $\mathcal{L}_y$, so as to maximize the recognition accuracy.

5.3 Domain Discriminators

Domain adversarial network emerges as a key technique in transfer learning [25] given its strong ability to learn transferable features between source domains and target domains. In our in-car activity recognition scenario, a

generic model for activity recognition is necessary since there always exist new domains (e.g., new driving conditions and drivers). We consider using unsupervised domain adaptation [19], which is able to remove those domain-specific characteristics even when the target domains are fully unlabeled.

State-of-the-art indoor activity recognition models [15], [16] mainly consider a single domain discriminator for domain adaptation, which however is not sufficient for the in-car scenario. As shown earlier, in-car activities can be affected by multiple impact factors (or classes) such as driving conditions, car models and human subjects. The implicit features within each particular class usually exhibit specific structures, indicating either the boundaries of different classes in supervised learning or the boundaries of different clusters in unsupervised learning. Without exploiting the multimode structures, single domain adaptation is prone to either under transfer or negative transfer [17], which easily leads to false alignment of discriminative structures. To address this challenge, in RF-CAR, we for the first time incorporate multi-adversarial domain adaptation for RFID-based in-car activity recognition.

We use a set of domain discriminators $G_d = \{G_d^1, G_d^2, \ldots, G_d^K\}$, where K is the number of classes. Each domain discriminator includes two fully connected layers as well as ReLU activation function. A softmax layer is also employed as last to generate the domain distributions for each class. Correspondingly, the domain space can be divided into K classes as $D = \{D^1, D^2, \ldots, D^K\}$. We also assign each input sample a domain label d_i^k for every class D^k. Each domain discriminator takes as input the concatenation of the feature representations $\mathbf{z}$ from the feature encoder and the label distributions $\hat{\mathbf{y}}$ from the activity classifier, and predicts the domain labels distributions $\hat{\mathbf{d}^k}$ of class k as

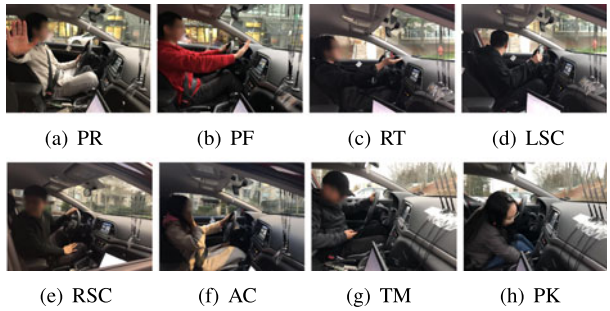
$$\hat{\mathbf{d}^k} = G_d^k(\mathbf{z}, \hat{\mathbf{y}}; \theta_d^k) = G_d^k(G_f(\mathbf{x}; \theta_f), G_y(G_f(\mathbf{x}; \theta_f); \theta_y); \theta_d^k), \quad (6)$$

where θ_d^k denotes the parameters of the kth domain discriminator. As mentioned, the target of the feature encoder G_f and the domain discriminators G_d is to play a minimax game to remove the domain-specific characteristics of the input data. To do this, we first calculate the integrated loss function of these two components as

$$\begin{aligned}\mathcal{L}_d^k(G_f, G_d) &= \mathbb{E}_{\mathbf{x},d}[-log\ G_d^k(\mathbf{z}, \hat{\mathbf{y}}; \theta_d^k)] \\ &= -\frac{1}{|D|}\sum_{x_i \in D}\sum_{j=1}^{|D^k|} d_{ij}^k\ log(G_f(x_i; \theta_f), G_y(G_f(x_i; \theta_f); \theta_y); \theta_d^k),\end{aligned} \quad (7)$$

where $|D|$ is the number of samples belonging to the whole sample space, $|D^k|$ is the number of labels for class k and d_{ij}^k is the corresponding domain label. Combining the loss of all the K discriminators together, we get the total loss for discriminators G_d as

$$\mathcal{L}_d(G_f, G_d) = \sum_{k=1}^{K} \mathcal{L}_d^k(G_f, G_d). \quad (8)$$



(a) PR (b) PF (c) RT (d) LSC

(e) RSC (f) AC (g) TM (h) PK

Fig. 10. Collecting activity samples in different domains.

5.4 Optimization Objective

The objective of our learning model is to minimize the label prediction loss $\mathcal{L}_y(G_f, G_y)$ and maximize the domain discrimination loss $\mathcal{L}_d(G_f, G_d)$, so as to achieve domain independent activity recognition. However, these two sub-objectives are mutually exclusive with each other and cannot be applied together directly. We introduce the gradient reversal layer proposed in [19] for training, so as to make the sub-objectives consistent. Based on Eqs. (5), (7) and (8), we integrate these two sub-objectives together and have the final joint loss function as follows:

$$\mathcal{L}(G_f, G_y, G_d) = \mathcal{L}_y(G_f, G_y) - \lambda \mathcal{L}_d(G_f, G_d), \quad (9)$$

where λ is a hyper-parameter to trade-off the two objectives in the model optimization. During the training process, our learning model aims to minimize this joint loss function, so as to achieve domain independent in-car activity recognition.

6 Evaluation

In this section, we implement a prototype to evaluate the performance of RF-CAR. We conduct trace-driven experiments to compare our RF-CAR framework with the state-of-the-art RFID-based solutions and evaluate the impacts of different settings.

6.1 Implementation and Evaluation Setup

Implementation. We build our RF-CAR prototype using the commercial-off-the-shelf devices without any hardware or software modification. The frontend module includes a Thingmagic M6e UHF reader connected to a common laptop (Thinkpad T430 in our prototype) and several passive RFID tags attached at different places (e.g., mirrors, doors, and seats) in a car. The frequency range of our RFID reader is our region is between 902 MHz and 928 MHz, and each channel has a bandwidth of 250 kHz. The sampling frequency is about 300 Hz. The Thingmagic reader provides APIs to customize the hopping table. We thus fix the channel on a common frequency 908.25 MHz to avoid the phase offset caused by channel hopping. The backend module on a server keeps fetching tag readings from the frontend module and perform the subsequent training and inference process.

Activity Collection. In our evaluation, we consider 8 common activities in cars for recognition, including pushing right (PR), pushing forward (PF), raise hand twice (RT), left shoulder check (LSC), right shoulder check (RSC), answering cellphone (AC), texting message (TM), and picking up things (PK), as illustrated in Fig. 10. For each activity, we

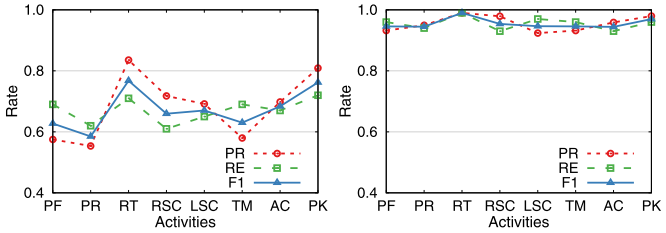
Actual Label	Predicted Label							
	PR	PF	RT	LSC	RSC	AC	TM	PK
PR	0.69	0.15	0.03	0	0.05	0.03	0.02	0.03
PF	0.16	0.62	0.02	0	0	0.11	0.07	0.02
RT	0.06	0.05	0.71	0	0	0.11	0.07	0
LSC	0.02	0.08	0	0.61	0.13	0.11	0	0.05
RSC	0.15	0	0	0.15	0.65	0.03	0	0.02
AC	0.07	0.08	0.05	0	0.11	0.69	0	0
TM	0	0.11	0.04	0.05	0	0.08	0.67	0.05
PK	0.05	0.03	0	0.04	0	0.03	0.13	0.72

(a) RFree-ID (without domain adaptation).

Actual Label	Predicted Label							
	PR	PF	RT	LSC	RSC	AC	TM	PK
PR	0.96	0.02	0	0	0.02	0	0	0
PF	0.03	0.94	0.01	0	0	0	0.02	0
RT	0	0.01	0.99	0	0	0	0	0
LSC	0.02	0	0	0.93	0.02	0.03	0	0
RSC	0	0	0	0.01	0.97	0.01	0	0.01
AC	0	0	0	0.01	0.02	0.96	0.01	0
TM	0.01	0.02	0	0	0	0.03	0.93	0.01
PK	0.01	0	0	0	0.02	0	0.01	0.96

(b) RF-CAR.

Fig. 11. The prediction confusion matrix of all activities when using RFree-ID approach and our RF-CAR approach.



(a) RFree-ID. (b) RF-CAR.

Fig. 12. Precision, recall and F1 score of all activities when using our RF-CAR approach and using RFree-ID approach.

have collected about a total of 2,500 samples in different settings, including 4 different driving conditions (i.e., downtown, school, rural, and residential area), 4 types of cars (i.e., compact sedan, standard sedan, SUV, and Minivan), and 4 volunteers differing in gender, weight, height, and habit. For each collection, the antenna array and the tags are placed at the relatively same location to guarantee the consistency in our measurement.

Learning Setup. We implement the RF-CAR learning model based on tensorflow [26] and train the learning model on a desktop equipped with a GTX 1,080 Ti GPU card, dual Intel I7 3.6 GHz CPU cards and 32 GB memory. The default filters of convolutional layers are 5×5 with the stride of 1. And the default filters of max-pooling layers are 2×2 with the stride of 2. We set the default neuron numbers in the two fully connected layers in both the activity predictor and domain discriminators to 256. In the testing stage, we collect new data samples from untrained domains for model evaluation. Specifically, we ask another 4 volunteers to perform activities in different driving conditions and cars, collecting about 2,000 activity samples for testing. These are then used to verify the generality of the domain adaption model in new environments.

Baseline Methods. We compare our approach with several existing learning models, including Random Forest (RF), RFree-ID [10] and EI [16]. RFree-ID is a state-of-the-art RFID-based system that uses traditional approaches without domain adaptation network to recognizes human gait patterns for identification. We incorporate its basic idea for the in-car activity recognition in our context. EI is the latest indoor activity recognition framework that uses single domain discriminator. RF is one of the most widely used classification methods, which can effectively inhibit the overfitting effect comparing with other classification methods. We extract the same features as used in EI. To guarantee the fairness of comparison and better evaluate the learning capability of our domain adversarial model, all the baseline models are fed with the same feature space as RF-CAR.

6.2 General Activity Recognition Performance

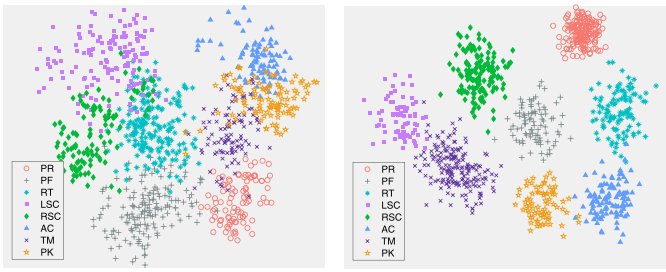
We first present the general in-car activity recognition performance of RF-CAR. Before diving into the evaluation, we introduce our used metrics in the classification problem: 1) *Precision* (*PR*) is defined as $\frac{TP}{TP+FP}$, where TP is the ratio of correctly labeled activities and FP is the ratio of falsely labeled activities as another activity; 2) *Recall* (*RE*) is $\frac{TP}{TP+FN}$, where FN is the ratio of mislabeled true activities; 3) *F1-score* (*F1*) is a combined metric for precision and recall, defined as $\frac{2*PR*RE}{PR+RE}$; 4) *Accuracy* (*ACC*) is defined as $\frac{TP+TN}{TP+TN+FP+FN}$, representing the correctly classified sample ratio. Besides the four metrics specific to each individual class, we also denote the *overall accuracy* (*O-ACC*) as $\frac{\sum TP_i}{all\ samples}$ to evaluate the overall performance of the classification model.

Fig. 11 shows the confusion matrix of the recognition results on 8 activities, comparing RF-CAR with the state-of-the-art RFree-ID approach (without domain adaptation). Note that the number of samples for each activity is the same, i.e., the data for each activity is balanced. Each row represents the actual activity label and each column represents the predicted label. Given the above metric definition, we can see that RF-CAR can achieve an overall recognition accuracy of 95.5 percent with a standard deviation of 2.1 percent, while RFree-ID only has an overall recognition accuracy of 67 percent with a standard deviation of 4 percent. Specifically, for activities with very special feature patterns (such as RT), our approach can achieve nearly 100 percent recognition performance, while the approach without domain adaptation seems unable to well extract and capture such patterns for recognition. This result indicates that RF-CAR can effectively extract activity related features and remove domain specific information. With one-time training, RF-CAR can effectively recognize those target activities even under new domains.

We next examine more metrics from the perspective of statistics to comprehensively understand the classification result. Fig. 12 compares these metrics of the state-of-the-art RFree-ID and RF-CAR. We can find that RFree-ID only achieves about 70 percent for both precision and recall. In contrast, both the precision and recall of RF-CAR are over 90 percent. The recognition of several activities even achieves near 100 percent precision and recall. These results indicate that our approach can not only accurately but also comprehensively recognize in-car activities.

6.3 Evaluation on Feature Encoder

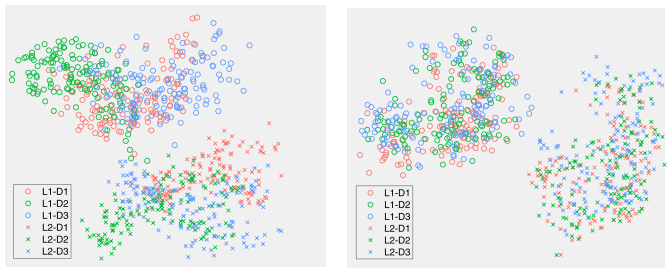
The feature encoder is a key role in our learning model, which cooperates with the activity classifier and plays against the domain discriminators to extract domain invariant feature representations. To intuitively evaluate the performance of the feature encoder, we use the t-SNE [27] tools to visualize the response of the CNN-based feature encoder. The t-SNE embedding tool maps the high-dimensional features into a 2D or 3D space, where the relative positions of data samples can be used to characterize the difference among them. As shown in Fig. 13, we randomly select about 100 data samples for each activity in our dataset for visualization in a 2D space and compare the different performance of simple neural



(a) The feature representations of NN layers.

(b) The feature representations of C-NN layers.

Fig. 13. Comparison of feature representations between simple NN and CNN with t-SNE visualization.



(a) EI using single-adversarial domain adaptation.

(b) RF-CAR using multiple-adversarial domain adaptation.

Fig. 14. Comparison of feature representations using single-adversarial domain adaptation and multi-adversarial domain adaptation.

network (NN) based representation and CNN-based representation used in RF-CAR. We can find that with NN-based feature encoder, although samples belonging to different activities are mostly concentrated in their respective regions, the boundaries are quite unclear and many data samples are mixed together, leading to inaccurate activity recognition. As a contrast, our CNN-based feature encoder has much better feature representation visually. As presented in Fig. 13b, samples belonging to one activity are grouped with clear boundaries, with only a few samples misclassified into other labels. This result indicates that our CNN-based feature encoder can effectively capture the feature differences of multiple activities even under untrained domains.

We next consider the impact of different neural network settings on the overall recognition accuracy, especially the number of CNN layers. As shown in Table 1, each column represents the number of CNN layers we used in the learning model and each row represents the corresponding performance metric. The default CNN filter size is set as 5×5 for layer number from 1 to 4. We can find that when we do not use CNN architecture (instead we use two NN layers), the general performance is relatively poor, with both recall and F1-score below 50 percent. This result shows that simple NN architecture may not be able to extract the useful features for recognition effectively. For CNN architecture, our experiment demonstrates that the performance is best when using two CNN layers, with 95.6 percent precision, 95.5 percent recall and 168 seconds training time. Yet when we keep increasing the CNN layer number, these metric values begin to decrease. This is because too many layers will compress spatial features too much and further cause information loss and too few layers are not able to fully capture the spatial correlations in the input data spectrogram.

6.4 Evaluation on Domain Adversarial Network

We next consider the domain discriminators and evaluate their performance on in-car activity recognition. For comparison, we consider our multi-adversarial domain adaptation approach, RF-CAR, and EI, the state-of-the-art single-adversarial domain adaptation approach. We select two activities and each activity contains data samples from three domains, where each domain is one particular combination of the three classes in our in-car activity recognition context. As described in Fig. 14, we also use t-SNE tool to compare the feature representations, where $L1$ and $L2$ denote the two activity labels, and $D1$ to $D3$ denote the three domain labels. We can find that the feature representations of RF-CAR are clearly separated into two parts according to their activity labels, yet the domain labels within each activity group are mixed together. This observation indicates that our model achieves good activity recognition with extracted features and successfully removes those domain-specific features. As a comparison, we can see the feature representations of the single-discriminator-based approach in Fig. 14a. Even the samples belonging to different activities are relatively separated, the domain labels are not completely mixed. For example, most samples of $L2-D1$ and $L2-D3$ appear in relatively different regions. On one hand, our experiment shows that different domains do have an obvious impact on activity recognition. On the other hand, the extracted latent features of different activities are clearly separated while the features of the same activities in different domains almost have the same distribution. Note that the testing data collected is from domains that are different from the training data. This experiment therefore confirms that domain adversarial network, especially the multi-adversarial domain adaptation network, is capable of removing domain-specific features and retaining activity-related features.

TABLE 1
Activity Recognition Performance With Different CNN Layers

CNN layers	0	1	2	3	4
Precision	0.508	0.829	0.956	0.917	0.893
Recall	0.364	0.808	0.955	0.891	0.864
F1-score	0.424	0.818	0.955	0.903	0.878
Training time	65s	159s	168s	175s	182s

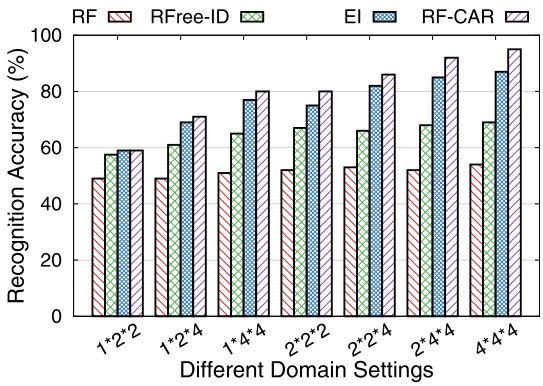


Fig. 15. Overall recognition accuracy under different training domain settings.

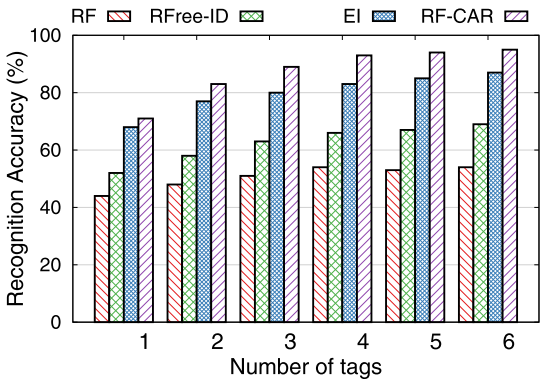


Fig. 16. Overall recognition accuracy with different numbers of tags.

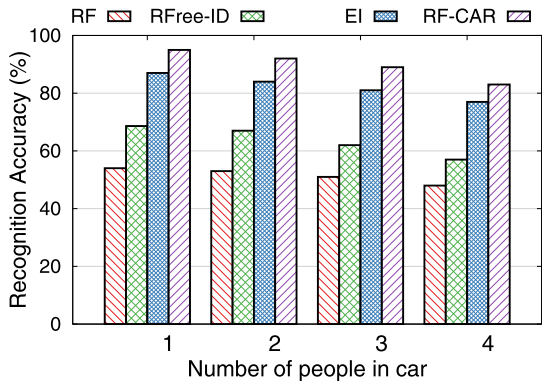


Fig. 18. Recognition accuracy given multiple people's presence.

Fig. 15 compares the overall recognition accuracy of RF-CAR and the baseline approaches varying different domain settings for training, where the x-axis k_i indicates the number of domains used in the ith class. We can observe that our approach outperforms all other baseline approaches for the in-car scenario with different training settings. Specifically, RFree-ID only has about 70 percent overall recognition accuracy even trained with enough domains. This is because traditional methods consider all captured information during activity sensing, while the extraneous domain-specific features can distort the activity related features, leading to low recognition accuracy. For domain adversarial approaches, EI has 87 percent accuracy with enough domains while RF-CAR can achieve about 95 percent overall accuracy. This comparative result shows that RF-CAR is more capable of distinguishing the inherent multimode structures of different domains and can remove such domain-specific impacts more effectively.

6.5 Impact of the Number of Tags and Antennas

The number of tags we used in the activity sensing is a key factor for recognition performance. We conduct experiments to evaluate such impacts. As described in Section 4, the backscatter signals between the RFID reader and each tag form a communication link that is used for activity sensing. As these tags are deployed at different positions, multiple tags together with the reader can generate multiple sensing links from different angles, providing more sensing information with spatial diversity in the in-car space. Fig. 16 presents the overall recognition accuracy when using different numbers of tags. We can see that with more tags, both RF-CAR and the baseline approaches could achieve better recognition performance. Comparing to using only one tag, using 6 tags can increase the overall recognition accuracy from 71.2 to 95.5 percent for RF-CAR, which is a great performance improvement with little extra overhead since the tag cost is negligible. This experiment indicates that the multi-tag sensing and the collaborative analysis in the learning stage can extract the activity features from multiple orientations, which greatly improves the overall recognition accuracy and robustness of RF-CAR.

In addition to the number of tags, the number of antennas is also an influential parameter given that commercial RFID readers usually support multiple antennas. Since the antennas are usually placed close as an array, the phase pattern information between one tag and multiple antennas are correlated with each other and we use PCA to extract the common characteristics. Fig. 17 illustrates the overall recognition accuracy of using different numbers of sensing antennas. From this experiment, we can know that different from adding tags, increasing the number of antennas only results in quite marginal performance improvement, especially when there are already multiple antennas. For example, RF-CAR only has 10 percent overall accuracy improvement when adding antenna number from 1 to 4. This is because in our preprocessing methods, the antenna array actually offers high reliability for sensing from one orientation compared to using only one antenna, while more tags provide extra sensing information from more orientations.

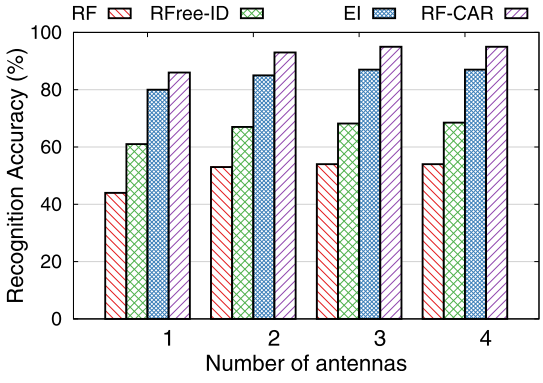


Fig. 17. Overall recognition accuracy with different numbers of antennas.

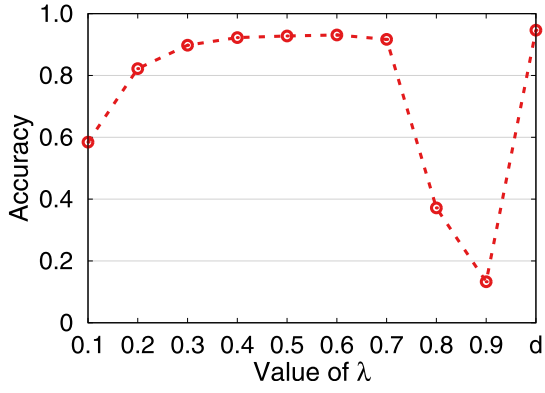


Fig. 19. Overall recognition accuracy with different setting of λ.

TABLE 2
Cross Validation and Statistical Test

Models	Accuracy	P-value
Random Forest	0.513 ± 0.032	2.8e-10
RFree-ID	0.679 ± 0.039	7.7e-9
EI	0.855 ± 0.026	1.3e-5
RF-CAR	0.949 ± 0.023	N/A

6.6 Evaluation on Hyperparameters

We next consider the impact of different settings for the hyperparameters on the recognition performance. As described in Eq. (9), λ controls the tradeoff between activity prediction and domain adaptation. Similar to [19], we dynamically set the value of λ based on the following rules

$$\lambda = \frac{2}{1 + exp(-10 \cdot p)} - 1, \tag{10}$$

where p is the percentage of current training steps to the maximum steps. Using such a dynamic setting, we can suppress noisy signal from the domain classifier at the early stages of the training procedure and encourage the domain adaptation at the late stage. Fig. 19 shows the recognition performance with different settings of λ, where d in the x-axis means the dynamic setting. We can find that either a too small or a too large value will cause a poor domain adaptation. And in general, the dynamic setting outperforms all the static settings.

6.7 Cross Validation and Statistical Test

In our experiment, given that the test domain should not appear in the training process, we cannot directly apply the conventional cross-validation method to verify the effectiveness of our model (because an arbitrary division may not well separate different domains). Note that our collected dataset contains 4 driving conditions, 4 persons and 4 types of cars. We therefore design a cross validation: each time we select one combination for testing, and the remaining $3 \times 3 \times 3$ domains served as the training set. We randomly conducted 10 groups of experiments for cross validation. The recognition accuracy of our model and the baseline models are presented in Table 2, which indicates that our model achieves good and stable recognition performance.

To further assess whether the performance of RF-CAR is significantly different from other baseline models, we conduct paired t-test over each experiment group between RF-CAR and other baseline models. The null hypothesis is rejected with respective p-values of $2.8 * 10^{-10}$, $7.7 * 10^{-9}$, and $1.3 * 10^{-5}$, for RF, RFree-ID, and EI, all of which are well below 0.05. This indicates that their performances are significantly different from a statistical perspective.

6.8 Evaluation on Multiple People's Presence

In many situations, there will be passengers inside a car other than the driver. Thus, we next evaluate the recognition performance when there are multiple people in a car. Fig. 18 presents the overall recognition accuracy for different approaches when the people in a car is ranging from 1 to 4. We can find that when the number of people increases, the overall recognition accuracy gradually decreases for all approaches. RF-CAR however still outperforms other baseline models, achieving 83.2 percent even when there are 4 people inside the car. This is because passengers will to some extent block the RF signals between the reader and the tags, making the multi-view observing incomplete. Besides, the small movements of these passengers will affect the received RF signals, making the recognition inaccurate. Note that in this paper we mainly focus on activity recognition for one target person (the driver in particular). Hence, the solution will work as long as the passengers do not move close to the operational space and perform these activities simultaneously. The scenario of activity recognition for multiple people is discussed in Section 8.

TABLE 3
Activity Recognition for Sequential Activities

1st Activity	2nd Activity	
	N	Y
N	0.01	0.03
Y	0.065	0.895

6.9 Evaluation on Sequential Activity Recognition

In practical driving scenarios, a person can perform a series of (sequential) activities, e.g., a left shoulder check followed with a right shoulder check when the driver tries to turn left and then turn right. We next examine the overall recognition accuracy of two sequential activities in different environments as previous settings. The recognition ratio results of the first and the second activity are presented in Table 3, where N indicates failure and Y indicates success. We can see that the recognition ratio for both two activities is still good enough, which achieves 89.5 percent. Comparing the situation of recognizing only one activity, the ratio of recognizing the first activity is almost twice that of the second activity. This is probably because the interval is quite ambiguous for some sequential activities, which makes the start detection of the second activity not so accurate.

6.10 Evaluation on Activity Detection

The preliminary step for activity recognition is accurately detecting and segmenting the activity samples. We also evaluate the activity detection ratio for RF-CAR. As presented in Table 4, our approach can detect and segment the target activities with an average ratio of 96 percent, and only TM and LSC fall below 95 percent detection ratio. This is because compared to other activities, these two activities have relatively small movement. Even so, the results show that our approach can effectively detect and segment target activities, which is necessary for in-car activity recognition.

TABLE 4
Activity Detection Ratio

Activity	PR	PF	RT	LSC
Ratio	0.983	0.965	0.984	0.931
Activity	RSC	AC	TM	PK
Ratio	0.962	0.956	0.925	0.978

7 Related Work

7.1 RFID-Based Activity Recognition

Radio Frequency Identification is rising as a promising sensing technology in recent years given its low cost, tiny size, low dependency, and the batteryless feature, making it widely used in a range of applications. Many pioneer researches have explored the sensing ability of RFID in activity recognition, which can be classified as tag-based approaches and tag-free approaches. Tag-based approaches usually attach tags to the surface of moving objects or people. The different motions of tags will produce different patterns for RSSI, phase, and Doppler shifts, which can be leveraged to recognize human breath [6], free-weight exercise [8], indoor activities [28], etc. Tag-free approaches usually attach tags to stationary objects such as walls and furniture instead of moving objects, and the patterns of reflected signals are used for recognition. For example, Zhang *et al.* [10] constructed a tag array attached on a wall to analyze people's gait characteristics, so as to identify different people. Fan *et al.* [11] analyzed the rich multipath effects caused by different motions in an indoor environment and further utilized the diverse phase patterns therein to recognize activities. Zou *et al.* [12] mainly profiled the phase changes when performing different gestures nearby and captured the spatial features to achieve gesture recognition. Besides, Yao *et al.* [29] used semi-supervised learning to analyze both tag-based and tag-free dataset for human activity recognition.

Previous activity recognition systems are usually confined to working in trained environments and are not well designed for fast-changing scenarios, such as in-car activity recognition. In contrast, RF-CAR adopts multi-adversarial domain adaptation to remove activity-irrelevant information while only keeping activity-related features for training, which achieves domain independent activity recognition.

7.2 Other RF-Based Activity Recognition

Besides RFID, there are many other radio frequency-based technologies that can be used for human activity recognition. Among them, WiFi is one of the most widely explored sensing technology given its ubiquity, low overhead and easy management. Many researchers relied on the coarse-grained received signal strength indicator [30], [31], [32] for sensing. For example, Sigg *et al.* [30] focused on the detection of static and dynamic activities of single individuals by leveraging active and passive, non-adhoc device-free systems to recognize four activities. WiGest [31] leveraged the change of patterns in RSSI to sense 8 in-air gestures even not in line-of-sight scenarios. As a relatively more fine-grained metric, channel state information (CSI) [33], [34], [35], [36] was also widely used to recognize more various activities. For example, WiDraw [37] was able to harness the Angle-of-Arrival values of incoming wireless signals at the mobile device to track the user's hand trajectory. To achieve more subtle activity recognition, Ali *et al.* [35] proposed WiKey that was able to recognize keyboard typing based on the variance of collected CSI patterns. Furthermore, mode advanced RF hardware such as USRP was employed to generate frequency-modulated carrier wave (FMCW) sweeping across a certain band to achieve more fine-grained sensing, such as constructing images of objects [38] and tracking 3D motions [39].

TABLE 5
The Electric Field Strength Measurement of In-Car RFID and Other Common Devices as well as the Reference Values for Public Exposure

Scenarios/Standards	E-Field (V/m)	Distance (m)
in-car RFID (Thingmagic M6e)	1.82	0.5
Smart Phone (iPhone 11)	2.17	0.05
Smart Watch (Apple Watch 4)	2.58	0.05
WiFi Router (ASUS AC1900)	3.23	1
ICNIRP Stardard [42]	41.8	N/A
Canada Stardard [43]	32.4	N/A

7.3 Domain Adversarial Learning

The learning model used in RF-CAR is related to adversarial networks for domain adaptation. Adversarial networks are mainly introduced to effectively train the generative model, where the most representative one is the generative adversarial network (GAN) [40]. In GAN, a generator is trained to fool the discriminator, so as to generate near-realistic data. Yet for domain adversarial network [17], [19], [41], the final objective is to train a feature encoder so that the extracted features are discriminative for the main task and keep stable when the domain changes. This unique function enables domain adversarial network to be used in activity recognition to eliminate the environmental impact. Pioneer researchers have explored it in WiFi-based activity recognition. Zhao *et al.* [15] improved domain adversarial network to remove the individual and condition-specific information during sleeping and used the extract features for accurate sleep stage prediction. Jiang *et al.* [16] proposed an adversarial learning framework to remove environment and subject related features in indoor activity recognition. Existing activity recognition systems all use single domain discriminator, where the multimode structures are easily mixed up, affecting the recognition performance. Different from these approaches, RF-CAR fully explores the in-car recognition and integrates multi-adversarial domain adaptation, which achieves a much higher recognition accuracy.

8 Discussion

This paper mainly introduces the fundamental design of RF-CAR, yet it can be further enhanced from many aspects before practical deployment in cars. We discuss the most important ones as follows.

Directional Activity Sensing. For traditional indoor activity recognition, different locations and orientations of activities can lead to quite diverse recognition performance. However, for in-car recognition scenario, people are usually restricted by seats and safety belts so that the location and orientation of activities are usually fixed. For this reason, RF-CAR focuses more on dealing with the extraneous domain information rather than considering the location and orientation. Even so, our model can be easily extended to recognize directional activities if necessary.

Scalability for Multiple People. In the current stage, RF-CAR mainly focuses on recognizing single person's activities in a car. Given that the positions of the RFID reader, tags and people's seats are fixed, activities should also yield some specific patterns in collected metrics including RSSI and phase.

Advanced learning tools such as deep learning can be further used for feature extraction and activity recognition. We leave the multi-activity recognition as our future work.

Practical Implementation. Our RFID-based sensing system requires a frontend module including a reader and multiple tags to collect the reflected backscatter signals. In the practical deployment, the reader can be integrated at the far end of the central control system in a vehicle. Tags can be simply fixed at many places such as doors and car body. Compared with other devices in a vehicle, deploying such sensing system does not cost much for car manufacturers considering its functionality.

Health Concern. The deployment of RFID-based sensing with enough distance actually does not introduce harm to human health. We have measured the electric field strength of different devices and compare it with several international RF explosure standards [42], [42] as in Table 5. The results indicate that even when RFID is deployed, the electric strength is within 1/10 of the reference level of these RF exposure standards, which is even lower than that of our daily used device. It is also known that RFID-based sensing has been used safely in Hongkong Airport for bagging handling for a long time [44]. Sensitive people with extra concerns such as pregnant women and infants can choose to close this function freely although FCC does not forbid the usage in this scenario.

Data Collecting and Model Training. Service providers can recruit users to drive various vehicles in different driving conditions and collect the sensing data accordingly for the subsequent training. Our model requires only one-time data collecting as well as training. The well-trained model can be delivered to other vehicles with limited overhead. Furthermore, the training data can be updated in a crowdsourced manner, i.e., users can contribute their raw data to the service provider with an incentive reward. We believe richer data collected in various conditions will help improve the overall recognition accuracy as well as robustness.

Model Generality. In this paper, we mainly focus on the in-car activity recognition scenario. The learning model seeks to remove extraneous domain information while retaining task-related features in training. Therefore, our learning model has the potential to accommodate other activity sensing technologies (e.g., WiFi, Ultrasound, and visible light) and application scenarios (e.g., indoor and outdoor environment). A future direction is to extend our model to other sensing methods and verify the effectiveness therein.

9 Conclusion

In this paper, we have proposed RF-CAR, an RFID-based in-car activity recognition framework. RF-CAR uses COTS RFID devices to collect the backscatter signal metrics. A preprocessing scheme is then applied to convert raw phase data into representative features. We have employed a multi-adversarial domain adaptation learning model to remove domain-specific extraneous information and keep activity-related features. Through the domain adversarial training, RF-CAR achieves domain independent in-car activity recognition. Extensive experiments have further demonstrated the superiority of RF-CAR compared to state-of-the-art recognition systems.

Acknowledgments

This work was supported by a Canada Technology Demonstration Program and a Canada NSERC Discovery Grant. Wei Gong's work was supported in part by NSFC Grant Nos. 61932017 and 61971390.

References

[1] T. Liu, Y. Yang, G.-B. Huang, Y. K. Yeo, and Z. Lin, "Driver distraction detection using semi-supervised machine learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 1108–1120, Apr. 2016.

[2] R. Zheng, K. Nakano, H. Ishiko, K. Hagita, M. Kihira, and T. Yokozeki, "Eye-gaze tracking analysis of driver behavior while interacting with navigation systems in an urban area," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 4, pp. 546–556, Aug. 2016.

[3] Y. Xing, C. Lv, H. Wang, D. Cao, E. Velenis, and F.-Y. Wang, "Driver activity recognition for intelligent vehicles: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 5379–5390, Jun. 2019.

[4] M. Raja, V. Ghaderi, and S. Sigg, "WiBot! in-vehicle behaviour and gesture recognition using wireless network edge," in *Proc. IEEE 38th Int. Conf. Distrib. Comput. Syst.*, 2018, pp. 376–387.

[5] S. Duan, T. Yu, and J. He, "WiDriver: Driver activity recognition system based on WiFi CSI," *Int. J. Wireless Inf. Netw.*, vol. 25, no. 2, pp. 146–156, 2018.

[6] Y. Hou, Y. Wang, and Y. Zheng, "TagBreathe: Monitor breathing with commodity RFID systems," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst.*, 2017, pp. 404–413.

[7] J. Yu, W. Gong, J. Liu, L. Chen, and K. Wang, "On efficient tree-based tag search in large-scale RFID systems," *IEEE/ACM Trans. Netw.*, vol. 27, no. 1, pp. 42–55, Feb. 2019.

[8] H. Ding *et al.*, "FEMO: A platform for free-weight exercise monitoring with RFIDs," in *Proc. 13th ACM Conf. Embedded Netw. Sensor Syst.*, 2015, pp. 141–154.

[9] X. Fan, F. Wang, F. Wang, W. Gong, and J. Liu, "When RFID meets deep learning: Exploring cognitive intelligence for activity identification," *IEEE Wireless Commun.*, vol. 26, no. 3, pp. 19–25, Jun. 2019.

[10] Q. Zhang, D. Li, R. Zhao, D. Wang, Y. Deng, and B. Chen, "RFree-ID: An unobtrusive human identification system irrespective of walking cofactors using COTS RFID," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, 2018, pp. 1–10.

[11] X. Fan, W. Gong, and J. Liu, "TagFree activity identification with RFIDs," *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 1, 2018, Art. no. 7.

[12] Y. Zou, J. Xiao, J. Han, K. Wu, Y. Li, and L. M. Ni, "GRfid: A device-free RFID-based gesture recognition system," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 381–393, Feb. 2017.

[13] D. Wu, D. Zhang, C. Xu, Y. Wang, and H. Wang, "WiDir: Walking direction estimation using wireless signals," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 351–362.

[14] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, "Widar: Decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi," in *Proc. 18th ACM Int. Symp. Mobile Ad Hoc Network. Comput.*, 2017, Art. no. 6.

[15] M. Zhao, S. Yue, D. Katabi, T. S. Jaakkola, and M. T. Bianchi, "Learning sleep stages from radio signals: A conditional adversarial architecture," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 4100–4109.

[16] W. Jiang *et al.*, "Towards environment independent device free human activity recognition," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, 2018, pp. 289–304.

[17] Z. Pei, Z. Cao, M. Long, and J. Wang, "Multi-adversarial domain adaptation," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 3934–3941.

[18] L. Yao *et al.*, "Compressive representation for device-free activity recognition with passive RFID signal strength," *IEEE Trans. Mobile Comput.*, vol. 17, no. 2, pp. 293–306, Feb. 2018.

[19] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 1180–1189.

[20] J. Wang *et al.*, "LiFS: Low human-effort, device-free localization with fine-grained subcarrier information," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 243–256.

[21] K. Itoh, "Analysis of the phase unwrapping algorithm," *Appl. Opt.*, vol. 21, no. 14, pp. 2470–2470, 1982.

[22] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures.," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
[23] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.
[24] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using WiFi," in *Proc. 15th Annu. Int. Conf. Mobile Syst. Appl. Services*, 2017, pp. 252–264.
[25] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko, "Simultaneous deep transfer across domains and tasks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 4068–4076.
[26] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Conf. Operating Syst. Des. Implementation*, 2016, vol. 16, pp. 265–283.
[27] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. Nov, pp. 2579–2605, 2008.
[28] L. Wang, T. Gu, X. Tao, and J. Lu, "Toward a wearable RFID system for real-time activity recognition using radio patterns," *IEEE Trans. Mobile Comput.*, vol. 16, no. 1, pp. 228–242, Jan. 2017.
[29] L. Yao, F. Nie, Q. Z. Sheng, T. Gu, X. Li, and S. Wang, "Learning from less for better: Semi-supervised activity recognition via shared structure discovery," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 13–24.
[30] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl, "RF-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals," *IEEE Trans. Mobile Comput.*, vol. 13, no. 4, pp. 907–920, Apr. 2014.
[31] H. Abdelnasser, M. Youssef, and K. A. Harras, "WiGest: A ubiquitous WiFi-based gesture recognition system," in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 1472–1480.
[32] J. Xiong and K. Jamieson, "ArrayTrack: A fine-grained indoor location system," in *Proc. 10th USENIX Conf. Netw. Syst. Des. Implementation*, 2013, pp. 71–84.
[33] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 65–76.
[34] F. Wang, W. Gong, J. Liu, and K. Wu, "Channel selective activity recognition with WiFi: A deep learning approach exploring wideband information," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 181–192, Jan. 2020.
[35] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using WiFi signals," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 90–102.
[36] F. Wang, W. Gong, and J. Liu, "On spatial diversity in WiFi-based human activity recognition: A deep learning-based approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2035–2047, Apr. 2019.
[37] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "WiDraw: Enabling hands-free drawing in the air on commodity WiFi devices," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 77–89.
[38] D. Huang, R. Nandakumar, and S. Gollakota, "Feasibility and limits of Wi-Fi imaging," in *Proc. 12th ACM Conf. Embedded Netw. Sensor Syst.*, 2014, pp. 266–279.
[39] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," in *Proc. 11th USENIX Conf. Netw. Syst. Des. Implementation*, 2014, vol. 14, pp. 317–329.
[40] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[41] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *The J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030, 2016.
[42] Guideline,ICNIRP, "Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz)," *Health Phys.*, vol. 74, no. 4, pp. 494–522, 1998.
[43] Consumer and Clinical Radiation Protection Bureau. "Limits of Human Exposure to Radiofrequency Electromagnetic Energy in the Frequency Range From 3 kHz to 300 GHz," Ontario, Canada: Health Canada, 2015. (Safety Code 6) (Cat.: H129-48/ 2015E-PDF).
[44] Y. Wong, P. W. Wu, D. M. H. Wong, D. Y. K. Chan, L. C. Fung, and S. W. Leung, "RFI assessment on human safety of RFID system at hong kong international airport," in *Proc. 17th Int. Zurich Symp. Electromagn. Compat.*, 2006, pp. 108–111.

Fangxin Wang (Student Member, IEEE) received the BS degree from the Department of Computer Science of Technology, Beijing University of Post and Telecommunication, Beijing, China, in 2013 and the MS degree from the Department of Computer Science and Technology, Beijing, China, in 2016. He is currently working toward the PhD degree in the School of Computing Science, Simon Fraser University, Burnaby, BC, Canada. His research interests include Internet-of-Things, wireless networks, big data analytics, and machine learning.

Jiangchuan Liu (Fellow, IEEE) received the BEng (Cum Laude) degree from Tsinghua University, Beijing, China, in 1999, and the PhD degree from The Hong Kong University of Science and Technology, Hong Kong, in 2003, both in computer science. He is currently a full professor (with university professorship) with the School of Computing Science, Simon Fraser University, British Columbia, Canada. He is a fellow of Canadian Academy of Engineering, and an NSERC E.W.R. Steacie Memorial fellow. He is a Steering Committee member of the *IEEE Transactions on Mobile Computing*, and associate editor of the *IEEE/ACM Transactions on Networking*, etc. He is a corecipient of the Test of Time Paper Award of IEEE INFOCOM (2015), ACM TOMCCAP Nicolas D. Georganas Best Paper Award (2013), and ACM Multimedia Best Paper Award (2012).

Wei Gong (Member, IEEE) receiving the BE degree in computer science from the Huazhong University of Science and Technology, Wuhan, China, and the ME degree in software engineering and PhD degree in computer science from Tsinghua University, Beijing, China. He is currently a professor of the School of Computer Science and Technology, University of Science and Technology of China, Hefei, China. His research focuses on wireless networks, Internet-of-Things, and distributed computing. He had also conducted research in Simon Fraser University, Burnaby, Canada and the University of Ottawa, Ottawa, Canada.