2024 21st Annual IEEE International Conference on Sensing, Communication, and Networking (SECON) | 979-8-3315-1918-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/SECON64284.2024.10934903

LoDiHAR: A Low-Cost Distributed Human Activity Recognition System Based on RFID

Xiaoqi Sun[†‡], Yanwen Wang[†‡], Chenwei Zhang[†], Zheng Wang[†], Xiaokang Shi[†], Yuanqing Zheng[§]
[†]Hunan University, China [‡]Shenzhen Research Institute, Hunan University, China
[§]The Hong Kong Polytechnic University, Hong Kong, China
{sunxiaoqi, wangyw, zhangcw, wangzheng, shixiaokang}@hnu.edu.cn, csyqzheng@comp.polyu.edu.hk

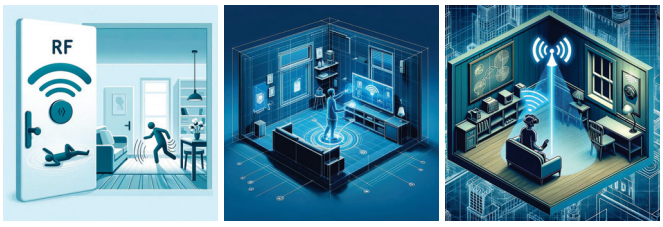
***Abstract*—Human Activity Recognition has been extensively applied to fulfill tasks such as fall detection, human-computer interaction, virtual reality, etc. Existing radio frequency-based HAR methods, although overcoming limitations of wearable-, visual-, and acoustic-based sensing technology, still suffer from high costs and low efficiency, which limits their pervasive use. In this paper, we propose LoDiHAR, a low-cost, distributed HAR system leveraging Radio Frequency Identification technology. LoDiHAR employs low-cost and fully programmable commercial wireless components, providing full access to the PHY samples of the backscattered signals, in which signal phases can be extracted to infer different activities. Different from COTS RFID systems that adopt a polling interrogation scheme, LoDiHAR supports a distributed sensing scheme, which profiles human activities more efficiently. LoDiHAR addresses a series of technical challenges such as accurate phase extraction from backscattered signals, asynchronous distributed RF data fusion and insufficient training data. A Conditional Generative Adversarial Network framework combined with a Transformer model is designed for accurate time-series activity classification. LoDiHAR demonstrates proficiency in recognizing eight types of human activities across diverse environments, achieving an accuracy of up to 94.9% while only costing 10% of the mainstream COTS RFID systems.**

***Index Terms*—RFID, low-cost, distributed sensing, data augmentation.**

I. Introduction

Motivation. Recent years have witnessed a huge proliferation of Internet of Things devices and the application of artificial intelligence technology in human activity recognition (HAR). As shown in Fig. 1, HAR plays a crucial role across various domains, including fall detection, virtual reality (VR) and human-computer interaction (HCI). In the era where aging is becoming increasingly severe, HAR technology facilitates real-time monitoring of elderly individuals, enabling prompt detection of accidents and ensuring their safety and health. In education, HAR enhances interaction with instructional multimedia, thereby improving the efficiency of both learning and teaching experience. In the entertainment industry, such as gaming and movies, HAR technology offers immersive experiences for users. In security, HAR enables behavioral analysis and helps prevent abnormal intrusion. So many potential application scenarios arouse extensive interest in research for HAR, facilitating the exploration of more pervasive use, low-cost and efficient technologies.

Prior works and limitations. Recent efforts have explored the potential of various commercially available devices to develop HAR systems. Existing HAR solutions primarily employ wearable sensors [14], [28], cameras [13], [21], acoustic sensors [3], [25], [29], [33], or RF signal-based recognition



(a) Fall detection. (b) VR. (c) HCI.

Fig. 1: HAR application scenarios (this figure is generated by ChatGPT 4.0).

[4], [31], [34], [35]. Wearable sensor-based HAR requires the user to constantly wear the devices and maintain persistent network connectivity, which incurs extra burden for users [14], [28]. Camera-based HAR struggles under poor light conditions and is prone to raise significant privacy concerns [7], [15]. Acoustic-based solutions offer contact-free HAR and are unaffected by lighting, yet still suffer from narrow bandwidth, rapid signal attenuation and a limited sensing range [25], [27]. Frequency Modulated Continuous Wave (FMCW) radar supports wider bandwidth, enabling fine-grained HAR. However, the FMCW-based method entails specialized high-end devices, which are expensive and hard for pervasive use [17], [20]. Currently, WiFi-based HAR has been extensively investigated due to its wide deployment in people's daily lives [9], [12], [31], [34]. Nevertheless, extracting channel state information (CSI) on limited hardware as well as its relatively low recognition resolution restrain its public use [31], [34].

RFID-based HAR is popularized for its passive, flexible and low-cost tags [18], [24], [30], [35], [36]. RFID tags harvest energy from the signals sent by RFID readers, eliminating the need for built-in batteries and intricate circuit design, which are small in size, cost-effective and flexible to deploy. Unfortunately, despite the low cost of RFID tags, mainstream commercial RFID readers tend to be expensive due to their complex hardware design, advanced signal processing techniques and patent barriers of the companies [30], [32]. More importantly, although COTS RFID readers can employ multiple antennas to enlarge the sensing coverage, the current standard EPC Gen2 protocol employs a polling communication scheme to prevent communication collision among readers [8], which sacrifices the read rate, rendering it inefficient for HAR applications.

In this study, we propose LoDiHAR, a low-cost and dis-

979-8-3315-1918-6/24/$31.00 ©2024 IEEE

tributed HAR system based on RFID technology, which is robust across different environments and can accurately identify eight types of human activities. The hardware design of LoDiHAR is based on a developed low-cost distributed sensing system yet only costs 10% of mainstream COTS readers [32], which poses great potential for economic and efficient HAR applications. Specifically, the tag backscattered signals can be simultaneously received by multiple receivers, which captures the human body more efficiently and provides more comprehensive information about human activities than that of a polling scheme. LoDiHAR enables full access to the PHY samples of RFID communication, in which signal phase can be extracted to infer different activities.

Challenges. However, we face many practical challenges when fulfilling HAR sensing tasks with such a low-cost sensing system. The first challenge lies in how to accurately distinguish between the PHY samples from absorb and reflect states of the backscattered signal to measure the signal phase. Existing approaches simply differentiate these two states based on their distinct amplitude levels in PHY samples due to the On-Off Keying (OOK) encoding scheme. However, our extensive experiments reveal that human activity imposes significant and non-negligible ambiguity on amplitudes of PHY samples for two states in tag backscattered signal, primarily due to significant multipath effects. As such, PHY samples for both two states exhibit the same amplitude level, which are unable to reliably separate, thereby significantly complicating the phase extraction process.

The second challenge involves how to effectively capture phase changes induced by the human body. LoDiHAR fully leverages the distributed infrastructure by simultaneously receiving tag backscattered signals through multiple receivers. However, when a human body moves within the detection range, the impacted multipath to different receivers are distinct due to different propagation distances. As a result, the extracted phase from each antenna is out of sync in both time and space, even for the same activity. A key challenge lies in how to effectively fuse the extracted signal phase from all antennas, ensuring that the fused phase encompasses most of the features captured by each individual antenna.

The third challenge stems from the inherent diversity of human activities. In HAR, variations in human bodies and individual habits result in phase diversity for the same activity, significantly degrading recognition accuracy. Specifically, differences in body shapes, heights and weights as well as different distances and speeds of the same activity remarkably impact the backscattered signals, struggling with fluctuations of the signal phase values for the same activity across different individuals. An intuitive approach is to manually gather a sufficient amount of data, which covers as much diversity as possible. However, such a burdensome data collection is hard, sometimes impractical, to implement in real-world settings.

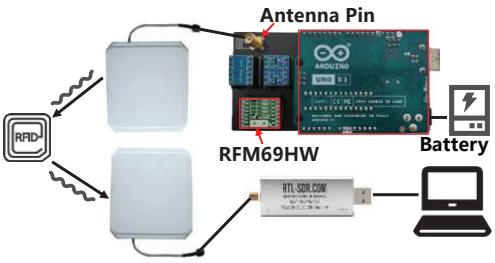
Solutions. To effectively distinguish between the absorb and reflect states in the tag backscattered signals, we exploit signal phase instead of amplitude. Our intuition is that the phases of backscattered signals corresponding to the two states are less likely to be identical during human activities, which creates two distinct clusters in the IQ plot. To further accurately map the clusters to the corresponding state, we conduct a deep investigation into the standard EPC Gen2 protocol and select typical PHY samples merely from CW (e.g., T1 defined in EPC Gen2 protocol [8]) as a reference signal, as it shares similar properties to the absorb state with regard to amplitude and phase. By doing so, the cluster with the closest distance to the T1 cluster in the IQ plot can be assigned to absorb state, while the cluster with the furthest distance is identified as the reflect state. By accurately separating two states, phase information can be extracted for HAR.

To fully leverage the information obtained from multiple receivers, we synthetically design a two-step data fusion scheme to address the time and space asynchronization inherent in multi-receiver system. In the first step, we align the phase waveform extracted from multiple receivers by matching their start and end points in the phase sequence of the activity and then perform linear interpolation across all waveforms ensuring that the length of the interpolated phase waveform for all distributed receivers is unified, achieving temporally synchronization among extracted phase sequences. In the second step, we employ a weighted sum strategy to integrate the phase information from all distributed receivers. The weights are determined based on the phase variations at each receiver. A phase waveform exhibiting more significant phase variations is assigned a higher weight, thereby emphasizing its signal characteristics, while a phase waveform with less pronounced variations is assigned a lower weight. This strategy effectively addresses the spatial asynchronization issues caused by the different deployments of receivers, ensuring that the measured phase sequence effectively retains the unique phase characteristics captured by each receiver.

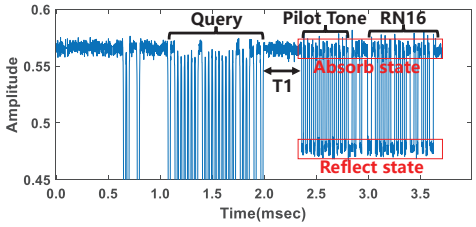
To release the burden of large-scale data collection, inspired by the Time Series Data Augmentation (TSDA) scheme [1], [2], we design a Conditional Generative Adversarial Network (CGAN) based on Long Short-Term Memory (LSTM) network. The CGAN is capable of generating high-quality samples, while the LSTM effectively extracts temporal features from the phase sequences. This is particularly important because the phase values associated with human activities are temporally related and contain rich information about the sequential relationships inherent in typical activities. Such a CGAN-based data augmentation scheme enables the automatic generation of signal phase sequences covering human activity diversity, significantly reducing the need for heavy manual data collection and enhancing the accuracy of activity recognition.

We summarize our contributions as follows:

- We accurately distinguish between reflect and absorb states using a low-cost RFID sensing platform, effectively resolving amplitude ambiguity of PHY samples in tag backscattered signals, and enabling precise phase extraction of the backscattered signal.
- We design a data fusion method for distributed parallel sensing for HAR, effectively fusing data that are both



(a) Low-cost RF system.



(b) The received physical samples.

Fig. 2: Low-cost RF system and PHY samples.

temporally and spatially asynchronous from multiple receivers, thereby providing reliable data for HAR.

- We design a CGAN deep learning framework suitable for time-series data augmentation, providing an effective approach for generating large-scale datasets, remarkably alleviating the burden from manual data collection.

II. Background

The employed low-cost distributed RF sensing system entirely consists of low-cost and general-purpose RF modules, which decouple the functionality of full-duplex communication of COTS RFID readers into Tx-only and Rx-only modules [32]. The transmitter employs an RFM69HW chip, which only costs less than $5USD$ and supports the OOK encoding scheme. The RFM69HW chip is controlled by being embedded into an Arduino UNO R3 board ($< 20USD$), as shown in Fig. 2a. We kindly refer the readers to find how to connect the IC pins of the RFM69HW chip to an Arduino board in [6]. On the receiver side, an RTL-SDR dongle equipped with RTL2832 ADC chip ($< 25USD$) is applied to serve as a sniffer to capture the tag backscattered signals. The circularly polarized antenna ($< 35USD$) with 9dBi gain is employed for both the transmitting and receiving antenna.

To successfully communicate with RFID tags, the employed low-cost system strictly follows the design of the interrogation step in the EPC Gen2 standard protocol [8]. In specific, the transmitter is programmed to send Continuous Wave (CW) and protocol-compatible commands to RFID tags, in which the *Query* command is deeply investigated and re-encoded to be compatible with the data format in RFM69HW chip. As a result, the emitted *Query* commands in the air are exactly the same as those generated by a COTS RFID reader in the view of tags. After receiving the *Query* command, RFID tags backscatter a 16-bit random number named RN16 prepended with a pilot tone, in which the signal phase can be measured. Fig. 2b shows an emitted *Query* command and tag replied RN16 in a single inventory round.

Note that this decoupling scheme significantly reduces the complexity of hardware circuit design, and more importantly supports distributed sensing with multiple receivers. Said differently, tag backscattered signals can be simultaneously captured by multiple receivers, which is not supported by COTS RFID systems that apply a polling interrogation scheme. A series of practical challenges have been addressed to extract accurate phase information from the backscattered signals, including carrier frequency offset (CFO), self-interference and cross-tech communication incurred by inevitable hardware heterogeneity due to applied low-cost RF modules for transmission and reception. We kindly refer the reader to [32] for more details. In this study, we focus on addressing particular challenges arising from employing this low-cost and distributed sensing system to implement distributed and efficient HAR.

III. System Design

A. System Overview

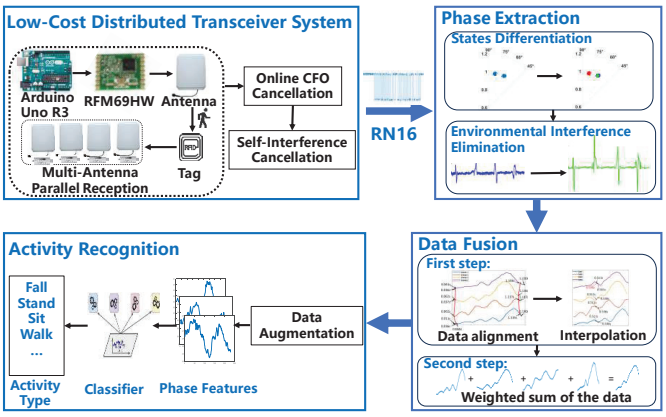
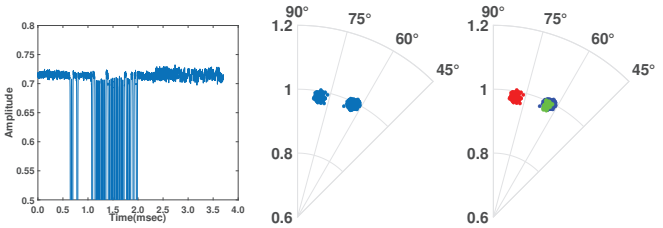


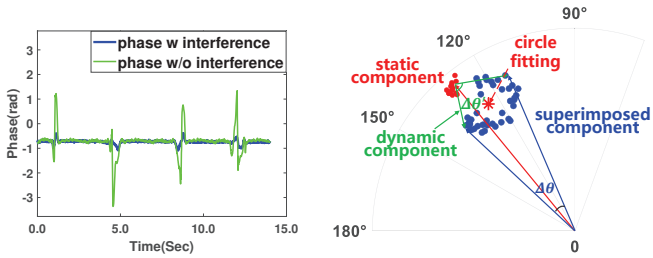
Fig. 3: System overview.

Fig. 3 depicts the system overview of LoDiHAR. LoDiHAR emits an excitation signal to activate the RFID tag and then transmits *Query* command that are compatible with the EPC Gen2 protocol to communicate with the tag. Once activated, the tag replies RN16 in response to the *Query* command. This backscattered signal, after undergoing multipath propagation and reflection from the human body, is simultaneously received by multiple distributed low-cost receivers. In the Phase Extraction module, LoDiHAR first differentiates between two states from the backscattered signals using signal phase. Next, the environmental interference is effectively eliminated from the extracted phase, obtaining phase information that is exclusively influenced by human movements. Third, the extracted phase values for each receiver are fused both temporally and spatially using our dedicatedly designed fusion algorithm in the Data Fusion module. The fused phase effectively captures the variations in multipath propagation caused by human activity. Finally, to cover the diversity of human activities as much as possible, we employ a CGAN model to augment the fused phase sequence, followed by a Transformer model to extract features and recognize the activities.



(a) RN16 with similar amplitude levels for both states. (b) Two indeterminate states. (c) Import reference state.

Fig. 4: Differentiate two states with reference state.



(a) Activity phase. (b) IQ plot of activity.

Fig. 5: Environmental interference removal.

B. Differentiate between Absorb and Reflect States

An RFID tag backscatters its signal using the OOK modulation scheme, which generates two distinct states, i.e., reflect state and absorb state. The phase of the backscattered signals can be measured between these two states if they can be clearly distinguished, which is a prerequisite for correct phase measurement [32]. Intuitively, one can simply differentiate these two states on the basis of their amplitudes on PHY samples due to their different electrical levels between CW and tag modulated signals induced by OOK scheme, as shown in Fig. 2b. Sensing tasks that impose distinct phase variations in the backscattered signals, such as object localization and human breath monitoring, can be fulfilled by configuring an appropriate threshold to differentiate between these two states, thereby enabling accurate extraction of phase information.

However, our extensive experiments manifest that the amplitude of the backscattered signal can be significantly impacted by human activities due to their multipath effects on signal energy, resulting in similar amplitude levels for both states. Fig. 4a illustrates the received RN16 PHY samples in one of the inventory rounds when a user performs activities within the sensing range of the low-cost system. The amplitude of PHY samples in RN16 exhibits a flat pattern, indicating that both absorb and reflect states share the same amplitude level. Such an ambiguity of amplitudes poses difficulties in differentiating between these two states of the backscattered signal, hindering reliable phase extraction.

To effectively separate the two states, we first project the PHY samples of RN16 pilot tone into an IQ plot, as shown in Fig. 4b. Two states share the same length of radius due to their similar amplitude, while can still be clearly separated by different phases (i.e., angles). However, it remains essential to accurately assign the states to each cluster for phase extraction. To this end, we introduce a reference state, which is derived from PHY samples collected in T1 duration in the same inventory round, as shown in Fig. 2b. This is because PHY samples in T1 characterize CW patterns that are similar to the absorbing pattern in RN16, which can be directly obtained in each inventory round. Luckily, in the EPC Gen2 protocol, the minimal length of T1 is defined as $268\mu s$ [8], which yields a sufficient number of samples (i.e., approximately 268 at a $1MHz$ sampling rate) and is easy to segment from an inventory round.

Therefore, we simultaneously project these T1 PHY samples and those from two states in the same inventory into the same IQ plot, as depicted in Fig. 4c. Due to the similar pattern in T1 and absorb state in RN16, the reference cluster (i.e., the green cluster) closely overlaps with the RN16 absorb state. More importantly, the reference cluster is clearly isolated from the reflect state due to different phases. Thus, the cluster center closer in Euclidean distance to the reference state can be identified as the absorb state, while the other cluster indicates the reflect state. By doing so, two states of tag can be distinctly differentiated for phase extraction, even when they share the same amplitude in PHY samples.

C. Elimination of Environmental Interference

After differentiating two states, we can now continuously measure the phase inventory by inventory, calculating the relative angle between two clusters to infer human activities. Fig. 5a shows the measured phase when a user repeats standing up for a while followed by sitting down for a while twice. The phase waveform slightly fluctuates when the user performs activities while remaining stable when the user keeps static, as depicted by the blue line. Such a small fluctuation in phase sometimes becomes even undetectable due to strong static reflection from the environment (e.g., the red line in Fig. 5b), resulting in an extremely small phase (i.e., $\Delta\theta$ in Fig. 5b).

Current studies eliminate the static component and extract dynamic component by estimating the center of the superimposed backscattered signals using circle fitting methods. The vector between the origin and the estimated center is regarded as the static component. This is because commercial RFID readers only output the phase of superimposed backscattered signal, without providing access to the PHY data. However, this method is effective only under the assumption that the superimposed signal forms an arc of a standard circle, which, however, is difficult to guarantee in practice. We plot the estimated static component derived from circle fitting and the actual static component extracted from PHY samples without human activity in Fig. 5b. The estimated static component significantly deviates from the one that is actually measured in both amplitude and phase, indicating that using the estimated static component may yield inaccurate phase information.

In our work, LoDiHAR employs a low-cost and distributed RF sensing system, which supports full access to the PHY samples of the backscattered signals. In another word, we

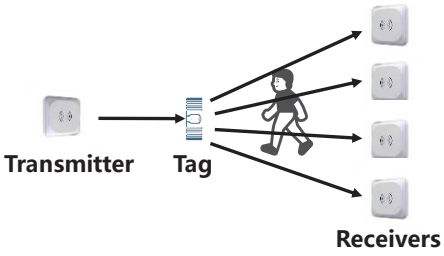
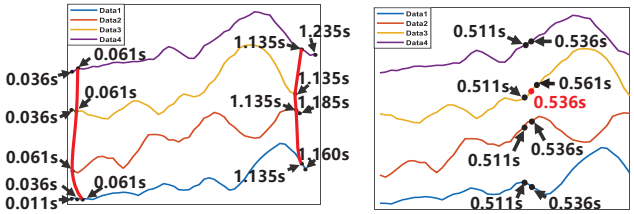


Fig. 6: Distributed sensing infrastructure.



(a) Data alignment. (b) Data interpolation.

Fig. 7: Temporal synchronization in phase fusion.

can directly obtain the raw PHY samples of the entire communication, involving the static component. In specific, we measure the static component by averaging a certain period of PHY samples in the absence of human activity, then perform vector subtraction to remove it from the PHY samples when performing activity. By doing so, the static reflection from the environment can be successfully canceled out, only remaining the phase induced by human activities.

The green curve in Fig. 5a illustrates the signal phase after applying our interference elimination method. The phase fluctuation significantly increases from 0.5 radians to 3.1 radians when only dynamic component is extracted, demonstrating a successful environmental interference mitigation. In addition, considering the environmental variations, we ensure timely updates of the static component by acquiring CW signals close to the tag backscattered signal in the same inventory round.

D. Distributed Phase Fusion

Implementing distributed sensing for HAR offers the potential to profile human activities with more reliable information. However, COTS RFID systems apply a polling interrogation scheme for multiple receivers such that only a single receiver is permitted to receive the backscattered signal at any given time slot. While this polling scheme effectively avoids communication collision, it results in inefficient data collection, particularly for tasks involving time series data, such as HAR.

To effectively capture reliable phase features incurred by human activities, we fully exploit the distributed characteristics of the employed low-cost sensing system. Specifically, we can deploy one transmitter to communicate with the tag and multiple receivers to concurrently receive the tag backscattered signal thanks to the decoupled functionalities of a duplex communication scheme, as shown in Fig. 6. As such, the tag backscattered signals can be received by each receiver in a parallel manner, which potentially provides richer information than polling communication scheme.

An intuitive data fusion method involving summing up the measured phase from all receivers when performing activities. However, simply adding multiple phases from different receivers is inherently difficult in distributed systems for two primary reasons. First, the propagation distance of the backscattered signal received by each receiver is distinct, particularly in reflections from the human body, resulting in temporal asynchronization of phase sequences extracted from each receiver. Second, variations in propagation distance incur different phase information for each receiver, leading to spatially asynchronous phase values for the same activity.

To address temporal asynchronization, we begin by connecting all receivers to a central hub, allowing us to configure the receiver such that all receivers start to sniff the communication channel at exactly the same time. We record the time stamp of the received backscattered signal for each receiver at a sampling rate of 1M/s and segment the phase sequence based on the timestamp when performing activities. Typically, we measure a single phase value for each inventory round, where the first point of RN16 serves as the timestamp of this phase value. Next, we identify the latest timestamp at the beginning of all phase sequences and the earliest timestamp at the end of these sequences to establish the start and end points of the recorded activity. Fig. 7a shows the aligned phase sequence for a fall activity collected by four receivers in parallel. For this particular fall activity, timestamps of 0.061s and 1.135s are selected as the start and end points, respectively.

However, phase sequences from all receivers are still asynchronous due to the multipath of the backscattered signals. To further align the phase values corresponding to each timestamp, we employ linear interpolation to supplement the phase sequences across all receivers to the same length. Specifically, in Fig. 7b, when a phase value for a given antenna is missing at a particular timestamp (i.e., the phase value at 0.536s for Data2), we supplement this missing phase value through linear interpolation between the two adjacent phase values (i.e., 0.511s and 0.561s). By iteratively applying this method across all sequences, we can achieve strict synchronization of the phase sequences for all receivers, ensuring that they are aligned across all timestamps.

To tackle spatial asynchronization induced by different phase variations among receivers, we fuse the time-synchronized phase sequences from all receivers by performing a weighted sum strategy. Our intuition is that the signal phase is linearly correlated with the propagation distance of the backscattered signals. As such, a phase waveform exhibiting more pronounced variations can be assigned to a higher weight, thereby highlighting its characteristics of human activity. In contrast, a lower weight is assigned to the waveform characterized by lower amplitude fluctuations. The assigned weight can be calculated as $W_i = A_i / \sum_{i=1}^{n} A_i$, where W_i represents the weight assigned to each receiver, A_i denotes the difference between max and min phase for receiver i, and n is the total number of receivers. Therefore, the weighted phase sequence can be expressed as $A = \sum_{i=1}^{n} W_i \times \theta_i$, where θ_i is the phase for receiver i.

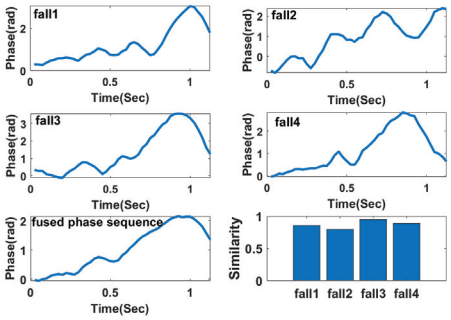


Fig. 8: Spatial synchronization in phase fusion.

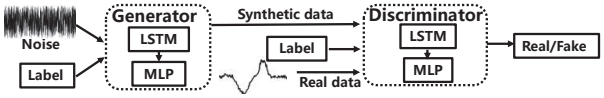


Fig. 9: CGAN model.

Fig. 8 depicts the results of our data fusion for a particular fall activity. We measure the correlation between the fused phase sequence and the one from each receiver. The average similarity exceeds 0.89, proving that the fused phase sequence retains the information from each receiver. By employing the weighted sum strategy, we can effectively integrate the phase features across all receivers, allowing for a more significant representation of the phase sequence and providing more reliable data for HAR.

E. Data Augmentation

In HAR tasks, a straightforward method to acquire training data involves manually collecting a sufficient amount of activity data, which covers as much diversity of activities as possible. However, such a burdensome data collection process is hard, or even impossible, to conduct in practice. Inspired by the effectiveness of the Time Series Data Augmentation approach [10], we design a CGAN model based on LSTM network to automatically generate activity-related phase sequences, as shown in Fig. 9. In our HAR scenario, the Generator and Discriminator fulfill complementary roles, where the Generator responses for producing synthetic time series phase sequences that aim to be indistinguishable from real-world measurements, while the Discriminator evaluates the authenticity of the presented phase sequences. This adversarial process facilitates the Generator to iteratively enhance its outputs until the Discriminator is unable to reliably identify the source of the phase sequences.

The Generator in the CGAN for LoDiHAR incorporates an LSTM unit with 64 hidden units, coupled with an MLP consisting of 6 layers. Furthermore, the discriminator is designed with a bidirectional LSTM that adeptly captures the impact of both preceding and subsequent states on the current state within time-series phase sequences, complemented by an MLP consists of 5 layers. Finally, we employ a transformer model, which manifests superior performance in time series classification, as our classifier to effectively distinguish among eight types of activities.

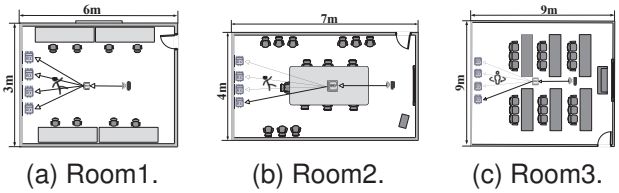


Fig. 10: Three different rooms.

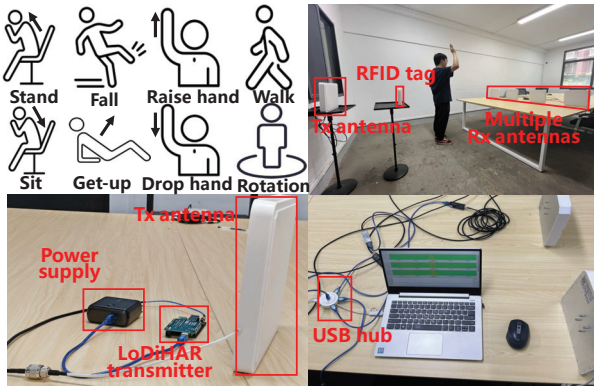


Fig. 11: The experiment setup.

IV. EXPERIMENTS AND EVALUATION

A. Experiment Setup.

1) Hardware: The experiment setup is illustrated in Fig. 11. LoDiHAR applies an Arduino Uno R3 board to control the RFM69HW chip, enabling the emission of CW and $Query$ command at a carrier frequency of 915MHz to illuminate and communicate with RFID tag. Unlike the system implementation presented in [32], we enhance the low-cost RF sensing system by integrating all components onto a single PCB board for higher system stability. Four identical commodity RTL-SDR dongles equipped with an RTL2832 ADC chip are employed as the receiver. All receivers are connected to a central hub for synchronization. Signal reception is configured using GnuRadio, which captures backscattered signals from the tags at a sampling rate of 1M/s, outputting complex values of the PHY samples.

2) Data collection and Model Training: We invite eight volunteers (five males and three females) to radomly perform eight types of activities across three rooms with varying sizes and layouts. The dimensions of these rooms are 6 m × 3 m × 3 m, 7 m × 4 m × 3 m and 9 m × 9 m × 3 m, respectively, as depicted in the Fig. 10. Each activity is performed 300 times with different speeds and distances, resulting in a total of 57600 phase sequences. To augment the collected dataset, we apply a data augmentation factor of 25× to enhance the dataset using our designed CGAN model.

We use Matlab to extract phase information from the tag backscattered signals. The training and testing of both CGAN and the HAR classifier are conducted using Pytorch on a PC server equipped with 32 GB of RAM, an Intel Core i7-13700K CPU from the 13th Generation lineup, and an NVIDIA GeForce RTX 4070 GPU. The dataset is carefully partitioned into three subsets for model evaluation: 80% for

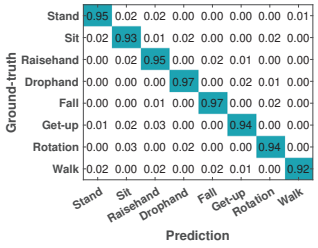


Fig. 12: Overall performance.

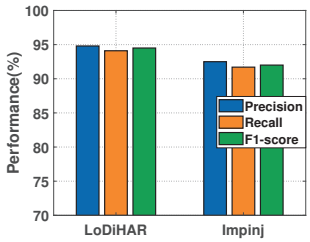


Fig. 13: Comparison with RFID reader.

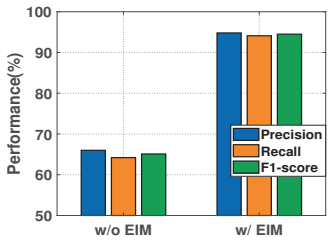


Fig. 14: Performance on EIM.

training to capture underlying data patterns, 10% for testing to assess model accuracy on new data, and 10% for validation to fine-tune hyper-parameters and prevent overfitting.

B. Evaluation

1) Overall system performance: Fig. 12 illustrates the confusion matrix that demonstrates the overall performance of LoDiHAR across all eight activities and various environments. For this evaluation, we utilize a hybrid dataset including both real collected and augmented data during training and testing. LoDiHar achieves an average recognition accuracy of 94.9%, with the accuracy for each individual activity exceeding 92.0% across diverse experiment settings. Our holistic design of LoDiHAR not only ensures cost-effectiveness and robustness but also delivers high accuracy in recognizing human activities, offering a significant potential to broaden the applicability of RFID-based HAR systems.

2) Comparison with COTS RFID reader: We evaluate the performance of LoDiHAR in comparison to the mainstream Impinj Speedway R420 reader, connecting four antennas to the Impinj reader and performing the same data fusion method and data augmentation strategy used in our study. The experiments are conducted in a single room, ensuring other experiment settings remain the same. The experiment result are shown in Fig. 13. LoDiHAR achieves an average precision of 94.9% in identifying various human activities, which outperforms that of the Impinj R420 reader. Such a superior performance owns to the distributed system design and our comprehensive data fusion approach, effectively maintaining a high read rate with an increased number of antennas while ensuring reliable phase information. In contrast, Impinj R420 operates on a polling communication scheme, which inherently limits the amount of information for HAR.

3) Performance on environmental interference mitigation: To validate the effectiveness of our environmental interference mitigation (EIM) method, we compared the performance of LoDiHAR before and after applying the interference elimination algorithm. Throughout the experiments, all other experimental settings are kept constant. Receivers are deployed in three different rooms at a distance of over 2 meters from the tag and the volunteers are allowed to perform activities in random speeds and distances. Fig. 14 depicts that the system performance manifests a substantial enhancement of 28% when employing our environmental interference mitigation algorithm, rising from 66% to 94%, thereby demonstrating the efficacy of our EIM approach.

4) Performance on different numbers of receivers: To evaluate the effectiveness of LoDiHAR in recognizing human activities using multiple receivers, we conduct an experiment with an increasing number of receivers. During this experiment, we maintained a constant experiment settings across different rooms, while only varying the number of receivers. Fig. 15 presents the results of the experiment. As the number of receivers increases from 1 to 4, the performance of LoDiHAR correspondingly improves from 90% to 94.9%. This experiment result validates that the use of multiple receivers enhances the system's ability to capture rich information about human activities, thereby improving the effectiveness of HAR. Importantly, LoDiHAR can flexibly add more receivers thanks to its distributed architecture, which provides a potential for HAR in more challenging environments.

5) Performance on Data Augmentation: In this evaluation, we test the impact of varying data augmentation factors on LoDiHAR to verify the effectiveness of our CGAN model. We employ the F1 score as our evaluation metric, as it is a comprehensive indicator that takes into account both precision and recall. We vary the augmentation factor from 1× to 35× applied to our collected phase sequences while keeping all other experiment settings the same. As shown in Fig. 16, the performance of LoDiHAR significantly improves in accordance with the elevated augmentation factors. However, when the augmentation factor exceeds 25×, the system performance slightly degrades since the augmented data at higher factors is prone to cause overfitting problem. Therefore, we adopt augmentation factors of 25× for model training.

6) Performance on Different Environments: In this evaluation, we conduct experiments across three different rooms with different sizes and layouts. For each room, we collect and augment the phase sequences to train a model, while applying this model to test the phase sequences collected from the other two rooms. As the experiment results depicted in Fig. 17, LoDiHAR still achieves an average precision of 90% for HAR, even when the model had not been trained in the unseen environments. The model trained in the room1 achieves the best performance due to its relatively simple layout, resulting in less multipath and more reliable collected phase information.

7) Performance on Different Distances: We evaluate LoDiHAR for identifying all eight activities performed at varying distances from the receivers. In this evaluation, we maintain a constant distance between the transmitter and tag to guarantee

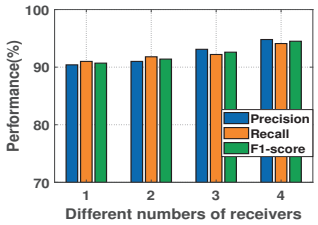


Fig. 15: Performance on different receivers.

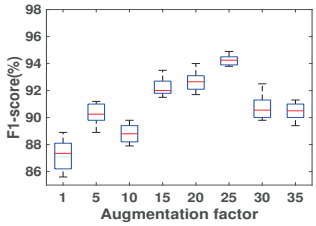


Fig. 16: Performance on different augmentation factors.

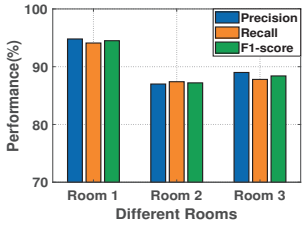


Fig. 17: Performance on different Environments.

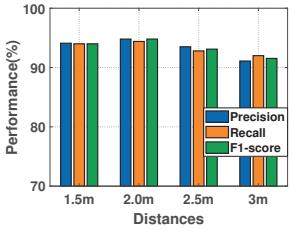


Fig. 18: Performance on different distances.

the activation of the tag while varying the distance between receivers and tag. Each activity is repeated 300 times in room3. At each distance, a data augmentation factor of $25\times$ is applied to the collected phase sequences. As shown in Fig. 18, the performance of LoDiHAR slightly decreases from 94% to 92% across all distances. The precision at a distance of 3m still achieves 92%, demonstrating the significant robustness of our LoDiHAR system. Note that the distance could be further enhanced by adding an RF amplifier for the receive antennas.

V. Related Work

A. Wearable Device Based HAR

Wearable sensors have been widely applied in HAR due to their capability to monitor the physiological parameters of the human body [14], [16], [23], [28]. A Fall Detection System is explored using the frequency of the inertial sensor across various datasets, demonstrating the effectiveness of deep neural networks for HAR [23]. Salient features from sensor data are extracted via Gaussian kernel-based principal component analysis and Z-score normalization, followed by training a deep CNN for HAR [28]. However, wearable device based methods encounter limitations in real-time transmission of raw inertial signals from wearable devices to servers due to high sampling rates and unstable communication networks. Crucial fall features in the inertial signal are used to achieve high fall detection accuracy with reduced data, thereby improving real-time performance [16]. However, the requirement for continuous wearing of the devices and persistent network connectivity pose burdensome for users [14].

B. Vision Based HAR

The combination of cameras with advanced deep learning technologies has significantly enhanced HAR in recent decades, enabling contact-free HAR [11], [13], [19], [21], [26]. ActivityNet provided an initial overview of vision-based HAR, focusing on conventional image processing methods [13]. Poppe [19] highlights the shift from traditional feature extraction to using Convolutional Neural Networks for complex activity recognition. UESTC-MMEA-CL [11] combines LSTM networks with CNN to capture temporal information of activities in videos. OFF [26] tackles real-time HAR, proposing a dual-stream CNN architecture that efficiently processes spatial and temporal data, enhancing real-time activity recognition. However, vision-based approaches are restricted by their reliance on good lighting conditions. Furthermore, they are prone to raise privacy concerns [21].

C. Acoustic Based HAR

The pervasive application of speakers and microphones embedded in smart devices has greatly facilitated acoustic-based HAR [3], [22], [25], [27], [29], [33]. RobuCIR enables identifying 15 gestures by extracting the channel impulse response of the acoustic signals [33], effectively recognizing gestures with varying duration, speed, and range. Vskin [25] achieves a more fine-grained hand motion recognition on the back surface of mobile devices when holding the devices by measuring multiple fingers' movement. Those works, however, can merely identify human motions in near-field scenarios (i.e., $\leq 1m$) by intentionally discarding the far-field interference. RemoteGesture [27] extends the acoustic sensing range by correlating the length of the transmitted signal with the sensing distances. However, acoustic-based solutions are constrained by the narrow bandwidth and rapid signal attenuation for far-field HAR.

D. RF Based HAR

RF sensing technology has become increasingly popular in human activity recognition due to the fact that human motion influences the propagation of RF signals, affecting both signal strength and phase [4], [5], [17], [31], [34], [35]. WiFall [34] successfully implemented HAR using WiFi by utilizing the amplitude and phase of channel state information (CSI). RTFall [31] demonstrated the capability of applying WiFi routers to detect human falls. However, extracting CSI on limited hardware and relatively low recognition resolution in WiFi signals limit its widespread application. A novel dynamic range-Doppler trajectory (DRDT) method based on the FMCW radar system is proposed to achieve fine-grained HAR [17], as FMCW radar operates with a wider frequency bandwidth. Nevertheless, FMCW-based method entails specialized high-end devices, which are expensive and hard for pervasive use. TACT [35] achieves accurate HAR recognition of eight activities through COTS RFID system. Yet, the relatively high cost of RFID readers and the polling mechanism in multi-antenna modes hinder its large-scale deployment. In contrast to these approaches, LoDiHAR is superior in achieving cost-effective and distributed RF sensing for HAR.

VI. Conclusion

This study proposes LoDiHAR, a cost-effective and distributed sensing system for HAR based on RFID technology. LoDiHAR entirely consists of low-cost general-purpose RF modules and provides full access to the PHY samples of

the backscattered signals. LoDiHAR first extracts phase information by accurately differentiating the reflect and absorb states in the backscattered signals. Then a comprehensive data fusion method is designed to integrate the extracted phase from distributed receivers, yielding a reliable phase sequence for HAR. Finally, LoDiHAR automatically generates a sufficient amount of training data from a limited number of collected data by employing a deep learning framework, effectively capturing the human activity diversity across different users. Extensive experiments demonstrate that LoDiHAR achieves an overall accuracy of 94.9% across various settings and environments.

ACKNOWLEDGMENT

This work was supported by the National Nature Science Foundation of China under Grant 62102139, the Nature Science Foundation of Hunan Province of China under Grant 2023JJ20015, the Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515011687, and the Science and Technology Innovation Program of Hunan Province under Grant 2024RC3105 (Yanwen Wang is the corresponding author).

REFERENCES

[1] L. Alawneh, T. Alsarhan, M. Al-Zinati, M. Al-Ayyoub, Y. Jararweh, and H. Lu. Enhancing human activity recognition using deep learning and time series augmented data. *Journal of Ambient Intelligence and Humanized Computing*, 12:10565 – 10580, 2021.

[2] D. Cheng, L. Zhang, C. Bu, H. Wu, and A. Song. Learning hierarchical time series data augmentation invariances via contrastive supervision for human activity recognition. *Know.-Based Syst.*, 276(C), sep 2023.

[3] H. Cheng and W. Lou. Push the limit of device-free acoustic sensing on commercial mobile devices. In *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, pages 1–10, 2021.

[4] C. Ding, H. Hong, Y. Zou, H. Chu, X. Zhu, F. Fioranelli, J. Le Kernec, and C. Li. Continuous human motion recognition with a dynamic range-doppler trajectory method based on fmcw radar. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6821–6831, 2019.

[5] C. Ding, L. Zhang, H. Chen, H. Hong, X. Zhu, and F. Fioranelli. Sparsity-based human activity recognition with pointnet using a portable fmcw radar. *IEEE Internet of Things Journal*, 10(11):10024–10037, 2023.

[6] H. Electronic. Rfm69hw ism transceiver module v1.3. http://www.hoperf.com/upload/rf/RFM69HW-V1.3.pdf, 2018.

[7] G. Garcia-Hernando, S. Yuan, S. Baek, and T.-K. Kim. First-person hand action benchmark with rgb-d videos and 3d hand pose annotations. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 409–419, 2018.

[8] GS1 EPCglobal, Inc. EPC UHF Gen2 Air Interface Protocol. https://www.gs1.org/standards/rfid/uhf-air-interface-protocol, 2024.

[9] B. Han, L. Wang, X. Lu, J. Meng, and Z. Zhou. *Cross-modal meta-learning for WiFi-based human activity recognition*. Association for Computing Machinery, New York, NY, USA, 2023.

[10] M. A. Hasan, F. Li, A. Piet, P. Gouverneur, M. T. Irshad, and M. Grzegorzek. Exploring the benefits of time series data augmentation for wearable human activity recognition. In *Proceedings of the 8th International Workshop on Sensor-Based Activity Recognition and Artificial Intelligence*, iWOAR '23, New York, NY, USA, 2023. Association for Computing Machinery.

[11] C. He, S. Cheng, Z. Qiu, L. Xu, F. Meng, Q. Wu, and H. Li. Continual Egocentric Activity Recognition via Foreseeable-Generalized Multi-Modal Representations. 8 2023.

[12] Y. He, J. Liu, M. Li, G. Yu, J. Han, and K. Ren. *SenCom: Integrated Sensing and Communication with Practical WiFi*. Association for Computing Machinery, New York, NY, USA, 2023.

[13] F. C. Heilbron, V. Escorcia, B. Ghanem, and J. C. Niebles. Activitynet: A large-scale video benchmark for human activity understanding. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 961–970, 2015.

[14] K. Karagiannaki, A. Panousopoulou, and P. Tsakalides. A benchmark study on feature selection for human activity recognition. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16, page 105–108, New York, NY, USA, 2016. Association for Computing Machinery.

[15] S. Khamis, V. I. Morariu, and L. S. Davis. A flow model for joint action recognition and identity maintenance. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1218–1225, 2012.

[16] J.-K. Kim, D.-S. Oh, K. Lee, and S. G. Hong. Fall detection based on interpretation of important features with wrist-wearable sensors. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, MobiCom '22, page 823–825, New York, NY, USA, 2022. Association for Computing Machinery.

[17] B. Langen, G. Lober, and W. Herzig. Reflection and transmission behaviour of building materials at 60 ghz. In *5th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Wireless Networks - Catching the Mobile Future.*, volume 2, pages 505–509 vol.2, 1994.

[18] X. Liu, X. Xie, K. Li, B. Xiao, J. Wu, H. Qi, and D. Lu. Fast tracking the population of key tags in large-scale anonymous rfid systems. *IEEE/ACM Transactions on Networking*, 25(1):278–291, 2017.

[19] R. Poppe. A survey on vision-based human action recognition. *Image Vis. Comput.*, 28:976–990, 2010.

[20] Y. Qifan, T. Hao, Z. Xuebing, L. Yin, and Z. Sanfeng. Dolphin: Ultrasonic-based gesture recognition on smartphone platform. In *2014 IEEE 17th International Conference on Computational Science and Engineering*, pages 1461–1468, 2014.

[21] J. Rajasegaran, G. Pavlakos, A. Kanazawa, C. Feichtenhofer, and J. Malik. On the benefits of 3d pose and tracking for human action recognition. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 640–649, 2023.

[22] W. Ruan, Q. Z. Sheng, L. Yang, T. Gu, P. Xu, and L. Shangguan. Audiogest: enabling fine-grained hand gesture detection by decoding echo signal. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, page 474–485, New York, NY, USA, 2016. Association for Computing Machinery.

[23] J. A. Santoyo-Ramón, E. Casilari, and J. M. Cano-García. A study of the influence of the sensor sampling frequency on the performance of wearable fall detectors. *Measurement*, 193:110945, 2022.

[24] S. Shen, M. Yang, X. Hou, L. Yang, S. Chen, W. Dong, B. Yu, and Q. Wang. Stmultiple: Sparse transformer based on rfid for multi-object activity recognition. *International Journal of Software Engineering and Knowledge Engineering*, 33(11n12):1813–1833, 2023.

[25] K. Sun, T. Zhao, W. Wang, and L. Xie. Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, MobiCom '18, page 591–605, New York, NY, USA, 2018. Association for Computing Machinery.

[26] S. Sun, Z. Kuang, L. Sheng, W. Ouyang, and W. Zhang. Optical flow guided feature: A fast and robust motion representation for video action recognition. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1390–1399, Los Alamitos, CA, USA, jun 2018. IEEE Computer Society.

[27] M. Tian, Y. Wang, Z. Wang, J. Situ, X. Sun, X. Shi, C. Zhang, and J. Shen. Remotegesture: Room-scale acoustic gesture recognition for multiple users. In *2023 20th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pages 231–239, 2023.

[28] M. Z. Uddin and M. M. Hassan. Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network. *IEEE Sensors Journal*, 19(19):8413–8419, 2019.

[29] A. R. Valiente, A. Trinidad, J. R. G. Berrocal, C. Górriz, and R. R. Camacho. Extended high-frequency (9–20 khz) audiometry reference thresholds in 645 healthy subjects. *International Journal of Audiology*, 53(8):531–545, 2014. PMID: 24749665.

[30] F. Wang, J. Liu, and W. Gong. Multi-adversarial in-car activity recognition using rfids. *IEEE Transactions on Mobile Computing*, 20(6):2224–2237, 2021.

[31] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li. Rt-fall: A real-time and contactless fall detection system with commodity wifi devices. *IEEE Transactions on Mobile Computing*, 16(2):511–526, 2017.

[32] Y. Wang, J. Cao, and Y. Zheng. Toward a low-cost software-defined uhf rfid system for distributed parallel sensing. *IEEE Internet of Things Journal*, 8(17):13664–13676, 2021.

[33] Y. Wang, J. Shen, and Y. Zheng. Push the limit of acoustic gesture recognition. In *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, pages 566–575, July 2020.

[34] Y. Wang, K. Wu, and L. M. Ni. Wifall: Device-free fall detection by wireless networks. *IEEE Transactions on Mobile Computing*, 16(2):581–594, 2017.

[35] Y. Wang and Y. Zheng. Modeling rfid signal reflection for contact-free activity recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(4), dec 2018.

[36] Y. Wang and Y. Zheng. Tagbreathe: Monitor breathing with commodity rfid systems. *IEEE Transactions on Mobile Computing*, 19(4):969–981, apr 2020.