TARF: Technology-agnostic RF Sensing for Human Activity Recognition

Chao Yang, Xuyu Wang, *Member, IEEE*, Shiwen Mao, *Fellow, IEEE*

***Abstract*— With the rapid development towards smart internet of things (IoT), detection of human activity has become essential in a variety of applications. Various radio-frequency (RF) sensing technologies, such as WiFi, Radio-Frequency Identification (RFID), and Frequency-Modulated Continuous Wave (FMCW) radar, have been utilized for non-invasive human activity recognition (HAR). It will be highly desirable to develop a HAR solution that can work with different types of RF technologies, such that the cost and the barrier of wide deployment can both be greatly reduced, and more robust performance can be achieved by utilizing the complementary RF sensory data. In this paper, we propose a technology-agnostic approach for RF-based HAR, termed TARF, which works with several different RF sensing technologies. A novel data generalization technique is proposed to mitigate the disparity in measured data from different RF devices. A domain adversarial neural network is proposed to combat the interference from various RF sensing technologies. The performance of the proposed system is evaluated with experiments using four different RF sensing technologies. TARF is shown to outperform the state-of-the-art Convolutional Neural Network (CNN)-based solution with considerable gains.**

***Index Terms*— Human Activity Recognition, Technology-agnostic RF sensing, Internet of Things (IoT), Domain Adversarial Learning.**

I. Introduction

Human activity recognition (HAR) has been recognize as one of the most important technology for many Internet-of-Things (IoT) applications, such as smart homes, safety surveillance, and health-care monitoring [1]. Video cameras and wearable sensors, such as smart watches and gyroscopes embedded in smartphones, are mostly used in traditional HAR solutions [2]. However, vision based HAR is usually constrained by the lighting condition and interference from the background, and may raise privacy concerns, while wearable sensors are uncomfortable for prolonged usage. As a result, several RF-based HAR solutions, such as WiFi [3], [4], Radio Frequency Identification (RFID) [5], and various types of radars [6], have been developed to overcome such constraints. By incorporating deep learning algorithms, these RF-based HAR approaches were shown effective to distinguish various types of human activities.

However, the existing solutions are each closely designed and tailored for a specific RF technology or platform. In the growing trend towards smart IoT systems, various RF sensing technologies are emerging. The limitation of being tied up with a specific technology or platform will hinder the development of large-scale, and easy-to-deploy HAR systems. It will be highly desirable to develop a HAR solution that can work with different types of RF technologies. First, such a *technology-agnostic solution* will greatly reduce the cost and overcome the barrier of wide deployment of HAR systems. For example, an existing RFID-based solution, e.g., [5], does not work with WiFi or radar. However, a user that does not have access to RFID can still make use of a technology-agnostic system with whatever RF sensing platform that is available, e.g., WiFi, without needing to acquire an RFID system. In addition, a technology-agnostic system can be used both in the lab, where both radar and RFID are available, and in the home, where there is only WiFi. Such a technology-agnostic solution will be of great value to users in such scenarios. Second, due to the different frequency band, wireless communication protocols, and hardware design, various RF platforms have their unique strengths and weaknesses in specific deployment environments. WiFi-based techniques, for example, can cover a large area, but are also susceptible to interference from the surrounding environment. RFID, on the other hand, is more resistant to interference from the environment, but is restricted by the shorter interrogation range, and the collisions and the mutual coupling effect induced by crowded tags. With a generalized signal processing module and a suitable movement feature extractor, the common activity-related features could be effectively extracted from different types of RF data, so that HAR could be performed using various RF technologies.

Obviously, this is a highly challenging problem given the variety of frequency bands, protocols, and hardware used in different RF sensing systems. The same propagation environment will become very different wireless channels and the same human activity will be transformed into very diverse RF representations. Not only the metrics used to describe such measurements are very different, but also the characteristics as indicated by the measurements for the same human activity exhibit a high degree of diversity. It is a great challenge and open problem to develop a technology-agnostic system to detect the

The manuscript was submitted on Dec. 14, 2021 for review. This work is supported in part by the NSF under Grants ECCS-1923163 and CNS-2107190, and through the Wireless Engineering Research and Education Center at Auburn University.

C. Yang and S. Mao are with the Department of Electrical and Engineering, Auburn University, Auburn, AL 36849-5201 USA (e-mail: czy0017@auburn.edu, smao@ieee.org). Xuyu Wang is with the Department of Computer Science, California State University, Sacramento, CA 95819-6021 USA (e-mail: xuyu.wang@csus.edu).

DOI: 10.1109/JBHI.2022.3175912

original human activity from such diverse representations.

In this paper, we propose a novel generalized, **t**echnology-**a**gnostic **RF** sensing system, termed TARF, for flexible and accurate human activity recognition utilizing a wide range of different RF sensing technologies. We first investigate the causes of the barriers between various RF technologies and find that the diversity is mostly caused by three factors: metric diversity, measurement sensitivity, and distinct translation of human activity to RF features. To address these problems, we first calibrate the RF data from different RF sensing technologies to represent them in a unified format. We then propose a signal preprocessing module that uses the Short Time Fourier Transform (STFT) to generate a generalized RF feature tensor, which can limit the interference of metric diversity and sensitivity diversity of different RF sensing technologies. In addition, we propose a Domain Adversarial Neural Network (DANN) to compensate for the discrepancy in RF signal translation. The domain discriminator of the DANN is to optimize the training variables in the feature extractor, thus allowing the network to concentrate on learning the generalized motion features, and ignore the technology-specific features for HAR.

The main contributions of this paper include the following.

- To the best of our knowledge, the TARF system is the first technology-agnostic human activity identification system capable of performing generalized and accurate HAR using various RF sensing platforms.
- We investigate the challenges in technology-agnostic HAR and show that they are caused by three main factors: metric disparities, heterogeneous sensitivity distributions, and diverse motion feature translations.
- A universal RF data preprocessing module is proposed to reduce the disparity between different RF sensing technologies. The sensitivity diversity is addressed by mapping the signal strength measurements, and generalized tensor data is constructed using STFT. The DANN is utilized to categorize different types of human activities, which further mitigates the interference from diverse RF domains.
- We develop a prototype of TARF to demonstrate the robustness of human activity recognition when data collection and testing are conducted using four different RF sensing technologies, including FMCW radar, WiFi in 2.4GHz and 5GHz bands, and RFID. The proposed system is compared with the traditional Convolutional Neural Network (CNN)-based technique, and the results validate that the proposed TARF system is resilient to technology-agnostic human activity recognition.

The remainder of this paper is organized as follows. We first review the related work on RF-based HAR and adversarial domain adaptation in Section II, We then introduce the preliminaries and the problem statement in Section III. Section IV provides an overview of the proposed TARF system, and Section V presents the detailed design of the key TARF components. Section VI presents the experimental evaluation of the TARF system and Section VII concludes this paper.

II. Related Work

The prior works on human activity recognition can be roughly categorized as camera-based, sensor-based, and wireless-based techniques [7]. In this paper, we mainly focus on RF-based human activity recognition, including Radar-based, WiFi channel state information (CSI)-based, and RFID-based methods. We will review such related work, as well as the recent works on adversarial domain adaptation for wireless human activity recognition in this section.

A. RF-based Human Activity Recognition

Several radar systems have been utilized for human activity recognition, such as the Frequency-Modulated Continuous Wave (FMCW) radar, Doppler radar, and Ultra Wide-band (UWB) radar [8]. FMCW radar was first employed for human activity monitoring, e.g., through-wall monitoring [9], 3D passive human tracking [10], and vital sign monitoring [11], by measuring the distance and velocity of body movement. However, these works require special hardware (e.g., Universal Software Radio Peripheral (USRP)) to implement the RF sensing system (usually operating at 5.467.25 GHz), thus incurring a higher cost. Commodity mmWave radars (e.g., IWR1443BOOST from Texas Instruments) operating at 77 GHz have also been utilized for various RF sensing tasks, such as human activity recognition [6], user authentication [12], and vital sign monitoring [13]. In [14], vision data captured by the Vicon motion capture system was used for supervised training of the deep learning model, which constructs 3D human meshes from sparse point clouds. Doppler radar can detect the velocity and direction of the subjects, and has also been utilized for human activity recognition [8]. Low-cost UWB devices have been shown useful for vital sign monitoring [15] and human activity recognition [16], where meta-learning was used to adapt to different deployment scenarios.

As a dominant wireless communications technology, there has been great interest in utilizing WiFi for human activity recognition. Several open-source tools have been developed to extract channel state information (CSI) from the orthogonal frequency division multiplexing (OFDM) channel, such as for Intel 5300 cards [17], the ESP32 WiFi microcontroller [18], the nexmon CSI Extractor for Broadcom and Cypress WiFi chips [19], the Atheros CSI tool [20], and the openwifi tool [21]. CSI amplitude and phase difference data have been used in applications for activity recognition, vital sign monitoring, and gesture recognition [3], [4]. In addition, deep learning techniques have great potential for achieving high recognition accuracy [22]. For example, long short-term memory (LSTM), a recurrent neural network (RNN) architecture, outperformed the traditional model-based method on human activity recognition using WiFi CSI amplitude data [23]. Generative adversarial networks (GAN) have been utilized to augment training data for human activity classification [24]. As in [14], high precision 3D skeletons captured by the Vicon motion capture system were used to supervise the training of the deep learning model that works with WiFi CSI data [25].

RFID is a near-field communication system originally developed for identifying tags attached to objects. Low-cost

and lightweight RFID tags can be attached to the human body as wearable sensors for activity monitoring. Commodity RFID readers (e.g., the Impinj R420 reader) can extract RF phase angle, Doppler frequency, and Peak RSSI from received signals, which can be used to estimate the range between the tag and the reader antenna. RFID-based sensing is usually more resilient to environmental interference than other RF sensing methods (e.g., radar and WiFi) due to the short range. RFID sensing techniques have been developed vital sign monitoring [26], driver fatigue detection [27], activity recognition [5], and 3D human pose estimation [28].

B. Adversarial Domain Adaptation for RF Sensing

Although deep learning has a great potential for RF sensing applications, it still faces great challenges for real-world applications. This is because different deployment environments, different users, or different wireless devices will lead to different data distributions, i.e., domain (or distributional) shift will occur between the source domain and target domain. A well trained deep learning model may fail when applied to unseen data. To address this challenge, generative adversarial network (GAN) [24], meta-learning [16], [29], and adversarial domain adaptation [30] have been proposed to adapt a trained deep learning model in the source domain to the new RF data in the target domain. In the following, we review several related works on adversarial domain adaptation methods for human activity monitoring.

Adversarial domain adaptation comprises feature learning, classifier learning, and domain adaptation, where adversarial training is leveraged to address the domain adaptation problem. The goal is to obtain an effective feature representation, which is discriminating for the learning tasks but invariant for the domain classifier. For example, the conditional domain adaptation architecture was used for radar-based sleep stage classification in different indoor scenarios, which was focused on supervised tasks [31]. In [30], unlabeled data was used in adversarial training for human activity classification, where four wireless devices were adopted to remove the environment and subject specific information. Furthermore, multi-view deep learning was introduced to improve the classification accuracy in different environments by fusing different wireless data [32], while multi-adversarial domain adaptation was proposed for WiFi based in-Car activity recognition [33]. All the above related works were focused on environment and user adaptation using adversarial domain adaptation. Unlike the related works, in this paper, we develop a novel technology-agnostic human activity recognition framework utilizing different wireless techniques such as radar, WiFi, and RFID.

III. Technology-Agnostic Generalization

A. Preliminaries of the Wireless Technologies

To develop the generalized technology-agnostic approach for RF-based human activity sensing, we first present the preliminaries of RF sensing with different RF technologies.

1) FMCW Radar: Frequency-Modulated Continuous Wave (FMCW) radar is a useful technology to provide both distance and velocity measurements. With the FMCW radar, the transmitted signal is modulated in the form of chirps [34], whose frequency keeps on increasing periodically. In each period, the signal frequency f_M is modulated as $f_M(t) = f_0 + \frac{Bt}{T_c}, \quad 0 \leq t \leq T_c$, where f_0 is the starting frequency, B is the bandwidth of the channel, and T_c is the duration of each period. The reflected chirp signal is received by the radar and fused with the transmitted signal. The fused signal $S_{FMCW}(t)$ at time t is given by [35]: $S_{FMCW}(t) = A\exp\left\{-j2\pi\left(f_0\tau + \frac{B\tau t}{T_c} - \frac{B\tau^2}{2T_c}\right)\right\}$, where A is the gain of the signal and τ represents the propagation delay of the backscattered signal. The expression of $S_{FMCW}(t)$ indicates that the frequency of the fused signal, $B\tau/T_c$, is determined by the propagation delay of the signal τ. By multiplying the speed of light, τ will be translated to the distance between the reflecting object and radar. Taking Fourier Transform on the fused signal, the power spectrum can therefore reveal the reflected signal strengths from various distances, which is referred to as the range profile [35]. When a person moves inside the radar detecting range, the gathered range profile will change with the body movements. Thus the received range profile can be leveraged to distinguish between different movement types.

2) Commodity 2.4GHz and 5GHz WiFi: The WiFi technology has also been explored as a promising solution to non-intrusive RF sensing of human activity. Compared to FMCW radar, WiFi is quite accessible due to the wide deployment of the infrastructure and the low-cost commodity devices. Recent WiFi sensing techniques mostly utilize the Channel State Information (CSI) extracted from the device driver, which provides a fine-grained representation of the orthogonal frequency-division multiplexing (OFDM) channel.

Considering the multipath effect of signal propagation, the CSI of a channel c can be written as [20]: $S_{WiFi}(c) = \sum_{n=1}^{N_c(t)} A_n \exp\{-j2\pi(f_c\tau_n + \phi_c)\}$, where $N_c(t)$ is the total number of propagation paths, f_c is the central frequency of channel c, ϕ_c is the phase offset of channel c, and A_n and τ_n are the gain and the propagation delay of the nth path, respectively. It can be seen that the channel offset largely determines the received CSI for all propagation paths. Human activity is captured in the CSI because, as a part of the propagation environment (or, the WiFi channel), moving human body parts can create considerable variations in most propagation paths, such as the gain A_n, the propagation delay τ_n, and even the total number of paths $N_c(t)$. Both CSI amplitude and phase can be used for learning human activity.

3) RFID: RFID devices have also been utilized in recent years for human activity monitoring. As wearable sensors, RFID tags are more resistant to environmental interference than broadband systems such as WiFi. Furthermore, RFID systems' low power consumption makes them a suitable RF sensing technology for the Internet of Things (IoT). The line-of-sight (LOS) path usually contributes to the dominant component in the received signal, and hence the received signal on a channel c can be written as: $S_{RFID} =$

$A_c \exp\{-j2\pi(f_c\tau + \phi_c)\}$, where A_c, f_c, and ϕ_c are the gain, frequency, and phase on channel c, respectively, and τ is the propagation delay. With the Low Level Reader Protocol (LLRP) [36] used in RFID systems, the phase value of signal S_{RFID} can be extracted for sensing of human activities. By attaching tags to the human body, the propagation delay τ of each tag changes along with the movements of body parts. Thus, human activities can be captured by the variations in phase values of the attached tags.

B. Problem Statement

Developing the technology-agnostic approach for human activity recognition is highly challenging. So we need to formulate the problem and figure out the mechanism to integrate the RF based techniques. Following the Fourier expansion, we can write the source signal $S(t)$ of body movement as a combination of different periodical components, as $S(t) = \sum_{n=1}^{M} K_n \cos(W_n t + \phi_n)$, where W_n is the frequency of the sinusoidal signal, K_n denotes the coefficient, ϕ_n represents the initial phase of each component, and M is the total number of periodic components. The set of parameters, i.e., $\{W_n, K_n, \phi_n, M\}$, represent the unique features of the corresponding human activity. The received signal reflected from the human body is dynamically distorted by the moving human body, and the distortion is mainly due to the prorogation path changed by the activity $S(t)$. Therefore, the reflected signal can be written as $R(t) = A_T \exp\left\{-j\left(2\pi\frac{D+S(t)}{\lambda} + \phi_T\right)\right\}$, where D is the average distance of the propagation path, λ denotes the wavelength of the transmitted signal, and A_T and ϕ_T represent the amplitude and the initial phase of the signal, respectively. The expression of $R(t)$ indicates that the human activity introduces an offset on the phase component $\angle R(t)$. The relationship between $\angle R(t)$ and the source signal $S(t)$ can be further investigated in the frequency domain. Taking Fourier transform on $\angle R(t)$, we have

$$\Gamma(\omega) = \int_{-\infty}^{\infty} \left\{\left(2\pi\frac{D+S(t)}{\lambda} + \phi_T\right)\right\} e^{-j\omega t}\, dt \qquad (1)$$

$$= \phi_c\delta(0) + \frac{\pi}{\lambda}\sum_{n=1}^{M} K_n[e^{j\phi_n}\delta(\omega - W_n) + e^{-j\phi_n}\delta(\omega + W_n)],$$

where $\delta(\omega)$ is the Dirac function, and ϕ_c is a constant given by $\phi_c = 2\pi D/\lambda + \phi_T$. The expression of $\Gamma(\omega)$ illustrates the mapping from the source signal $S(t)$ to the phase of the reflected signal $\angle R(t)$. The phase is determined by the unique features of the human activity, i.e., $\{W_n, K_n, \phi_n, M\}$.

The challenge in many RF sensing applications is, accurate phase angle of the reflected signal is usually hard to obtain due to the multipath effect. The mapping from the human activity signal to the received phase sample becomes highly complex. Traditional RF HAR based systems employ various signal preprocessing techniques to combat the interference caused by the complex mapping, to extract useful features for motion classification. Unfortunately, a specific preprocessing method developed for one RF technology is usually not applicable to other RF technologies, due to their different frequency bands, different communication protocols, and different types of hardware. To address this challenge, the primary objective of our technology-agnostic approach TARF is to learn the generalized features of the human activity signal $S(t)$ from various RF technologies, which will facilitate the accurate classification of different human activities.

IV. System Overview

A. Main Challenges

Each existing RF sensing based activity recognition system is closely tailored for the specific technology and has its unique advantages and certain limitations. Such a system designed for one RF technology usually does not work for a different technology (e.g., FMCW radar vs. RFID). Given the availability of various RF technologies in our daily lives, a technology-agnostic approach would be highly desirable to achieve better adaption to different sensing scenarios, as well as greatly reduce the barrier to deploying the system.

However, pursuing a generalized approach that works with very different RF technologies is a great challenge due to two main reasons. First, different RF technologies are established on different frequency bands. For example, the Ultra High Frequency (UHF) RFID systems operate on the 900 MHz band, WiFi works on the 2.4GHz or 5GHz bands, and the FMCW radar used in our experiments is on the $76 \sim 81$ GHz millimeter wave (mmWave) band. Even deployed in the same environment, different RF technologies see different propagation channels and different signal characteristics. Second, due to their different physical layer and medium access control layer protocols, as well as different device drivers that are available, the RF data collected by different RF devices are highly diverse. It is extremely challenging to develop a generalized approach to accurately detect human activities from such diverse RF data.

B. System Architecture

To address the above challenges, we design a novel system TARF that is generalizable to diverse RF data measured with different RF sensing technologies to perform technology-agnostic human activity recognition. Fig. 1 provides an overview of the system architecture of TARF, which is composed of three main components, including (i) RF signal collection, (ii) generalized RF signal preprocessing, and (iii) domain adversarial deep neural network based activity recognition. In the RF signal collection module, raw RF signals are sampled by several different RF sensing platforms. According to (1), we are interested in the phase angle $\angle R(t)$ of the collected signals. We use the phase signal from the RFID system and the phase difference signal from 2.4GHz and 5GHz WiFi systems. For FMCW radar, phase is not a good indicator of human activity because of the modulated frequency. Instead, we leverage the range profile from the FMCW radar as an input signal, which is indicative the propagation distance of the signal and is readily available from the device.

In the proposed technology-agnostic TARF system, the signal will be treated using the same generalized signal preprocessing module, no matter which RF technology is used for sensing. Generalization to multiple RF technologies

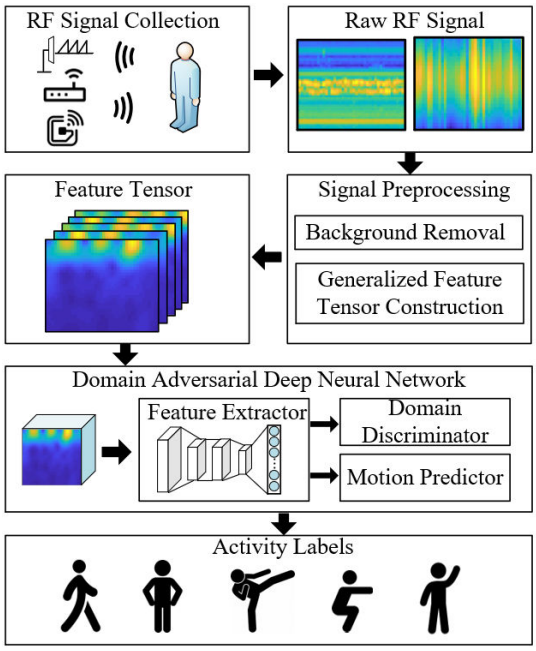


Fig. 1. Architecture of the proposed technology-agnostic RF sensing system TARF.

should begin with a standardized RF data format. We propose a generalized RF data preprocessing module for different RF sensing systems, where the input signal is treated as a group of different observations of the same source signal $S(t)$ and is converted to a generalized input data matrix. Then the general background removal is implemented with Hampel filters, where the interference from the common static background is removed. Finally, the observations are reordered according to their signal strengths to mitigate the sensitivity diversity across different RF devices.

Then, generalized feature tensors will be constructed and fed into a domain adversarial deep neural network for activity classification. In comparison to a traditional convolutional neural network, the domain adversarial neural network is able to acquire more generalized features of diverse human activities by combating characteristics gained from other domains. The details of the signal preprocessing and the domain adversarial deep neural network structure will be elaborated in Section V.

V. Design of the Technology-Agnostic System

In this section, we present the detailed design of the proposed TARF system. We will examine the challenges in diverse RF data, generalized feature mapping, and activity recognition with domain adversarial neural networks, and then present our proposed solutions to address these challenges.

A. Metric Generalization

1) Challenge: Diversity in Measured Data: The first challenge of generalization arises from the use of very different types of channel data. Fig. 2 illustrates the raw data collected for the same human activity sampled by three different RF sensing platforms over a 4-second period. For convenience, we have normalized the data from each platform. In these "temperature" plots, a lighter color, e.g., yellow, indicates larger values or higher strengths, while a darker color, e.g., dark blue, represents smaller values. Each plot in Fig. 2 presents a different type of sampled data. Specifically, Fig. 2(a) shows the raw range profile sampled by an FMCW radar, where the 2-dimensional data (or, matrix) consists of the signal strengths sampled over time for different ranges from 1 m to 5 m. The RFID data, plotted in Fig. 2(b), comprises the phase values sampled from 12 RFID tags. The WiFi data in Fig. 2(c) consists of phase differences, i.e., the difference of phases from a pair of WiFi antennas, sampled from 30 subcarriers.

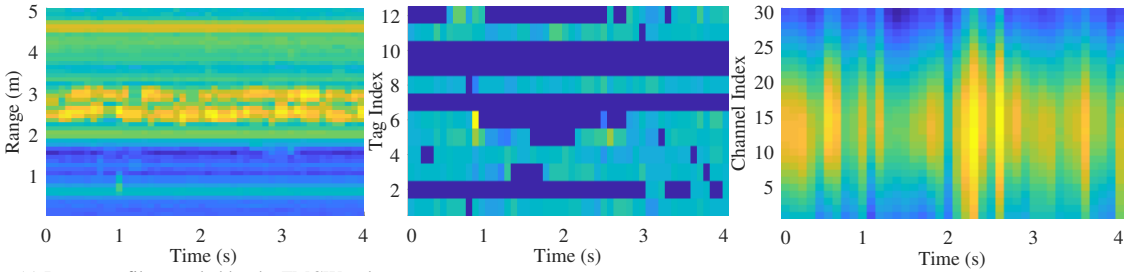
It is obvious that such RF data are very different, each with their unique features, making it extremely hard to handle with a generalized model. In particular, for the three RF technologies, the range profile of FMCW radar is measured in decibel (dB) typically ranging from 20 dB to 120 dB. The phase values sampled from the RFID tags, as well as the phase differences from the 30 subcarriers of the WiFi channel, vary from 0 to 2π rad. The different metrics incur considerable diversity in the scale of data measurement. Moreover, such RF data with different metrics cannot be directly generalized into a normalized format.

2) Proposed Solution: The first step in data preprocessing is to remove the disparity in metrics of different RF hardware platforms. We begin by defining the generalized data matrix in order to gather raw data from different kinds of RF platforms. The generalized data matrix has the following format:

$$\mathbf{S}_G = \begin{bmatrix} F_1^1 & F_2^1 & \dots & F_{N_t}^1 \\ F_1^2 & F_2^2 & \dots & F_{N_t}^2 \\ \vdots & \vdots & \ddots & \vdots \\ F_1^{N_F} & F_2^{N_F} & \dots & F_{N_t}^{N_F} \end{bmatrix}, \tag{2}$$

where F represents an RF data measurement for human activities and N_F denotes the total number of measurements. The integer N_t denotes the number of time frames captured by the RF device. The sampling frequency of all RF platforms is set to 10 Hz in (2), therefore the length of the x-axis N_t is given by $N_t = 10 \times t$. Meanwhile, the amount of measurements taken by different RF platforms determines the size of the y-axis dimension. Each measurement is regarded as an observation of the source signal $S(t)$, which is generated by the human activity.

With FMCW radar, the subject performs activities within its range and the range profile in the form of power spectrum is obtained by 256-point fast Fourier transform (FFT). Therefore we have $N_F = 256$ for the FMCW platform. With WiFi, a transmitter sends packets to a receiver, with the subject in the middle. The receiver has three antennas and each can extract phase data from 30 subcarriers, which results in 90 RF measurements for each received packet. For RFID based sensing of human activities, we attach 12 tags to the joints of the subject, including neck, left shoulder, right shoulder, left elbow, right elbow, left wrist, right wrist, pelvis, left hip, right hip, left knee, and right knee. Three RFID readers are used to interrogate the tags, while phase data is collected from received responses. Unfortunately, because of the Slotted-Aloha-like collision avoidance protocol, only one phase measurement can be collected by the reader at



(a) Range profile sampled by the FMCW radar. (b) Phase values from different RFID tags. (c) Phase difference from 30 WiFi subcarriers.

Fig. 2. Raw data sampled by different RF technologies for the same human activity over a 4-second period.

a time. We employ an effective tensor completion based data interpolation method [28] to augment the sparse RFID phase data. After tensor completion, each time frame has 36 phase samples (i.e., from 3 antenna and 12 tags). With the generalized data matrix employed in the system, we can remove the diversity in metrics of the diverse RF platforms.

B. Generalized Feature Mapping

1) Challenge: Diversity in Sensitivity: No matter which RF technique is employed, a number of measurements are gathered at the same time. However, the sensitivity of these measurements to human activity may be highly different. For example, in Fig. 2(a), the signal strength around 2.5 m is more sensitive to the human movements, because this is the average distance between the subject and the FMCW radar. As a result, measurements taken at a distance of 2.5 m should contribute more to the correct extraction of motion features. When it comes to the RFID technology, however, the situation is completely different. The sensitivity, as shown in the measurements, is strongly dependent on the limb where the RFID tags are attached, since the received phase value is determined by the movements of the RFID tags. The sensitive data should be more emphasized for accurate activity detection. Such diverse sensitivity also poses a challenge to developing the technology-agnostic approach.

2) Proposed Solution: To deal with the diverse sensitivity in measured RF data, we map the RF measurements from different wireless technologies into a generalized order, so that the same human action introduces a comparable distribution of measurements. Since the wireless propagation environment has a great impact on the measured signal, the environmental influence should be firstly removed before the mapping process. To do this, we measure the component corresponding to the impact of the static background and eliminate it from the sampled signal. Such a background removal operation is performed on each row of matrix $\mathbf{S}_G$ using two separate Hampel filters. The first filter uses a window size of 4 for thermal noise reduction, while the second uses a window size of 15 for extracting the background component. To extract the component corresponding to human activity, we subtract the signal filtered by the Hampel filter with the larger window size from the signal filtered by the Hampel filter with the smaller window size. After removing the background component, we reorder all the rows in $\mathbf{S}_G$ according to the signal strength.

The signal strength for row i, denoted by P_i, is computed by the variance of the time sequence as $P_i = \frac{1}{N_t}\sum_{t=1}^{N_t}(F_t^i - \mu_i)^2$, where μ_i represents the mean value of each row. Since the background component has been removed, the variance of the signal indicates its strength. The new matrix is sorted in descending order of signal strength. We always choose the first N_P rows of data for human activity recognition. The reordered matrix $\mathbf{S}_R$ is given by:

$$\mathbf{S}_R = \begin{bmatrix} F_1^1 & F_2^1 & \dots & F_{N_t}^1 \\ F_1^2 & F_2^2 & \dots & F_{N_t}^2 \\ \vdots & \vdots & \ddots & \vdots \\ F_1^{N_P} & F_2^{N_P} & \dots & F_{N_t}^{N_P} \end{bmatrix}, \tag{3}$$

where N_P is the number of the most powerful signals chosen for activity recognition.

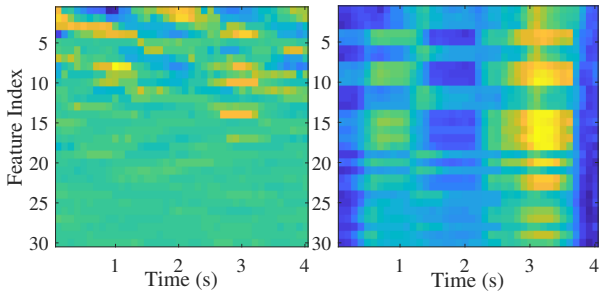


Fig. 3. Examples of the calibrated generalized feature matrix $\mathbf{S}_R$ measured from the kicking activity. Left: sampled with FMCW radar; Right: sampled with 5GHz WiFi.

In Fig. 3, we present the examples of the reordered matrix $\mathbf{S}_R$ of the data collected by FMCW radar and 5GHz WiFi devices, where $N_P = 30$ for a period of four seconds. It can be seen from the figures that the dimensions of the two signals are equivalent for the same sampling period. Furthermore, since the background component has been eliminated and the samples are reordered according to their strength, the overall sensitivity distributions of the two different RF technologies are now similar to each other. With the metric generalization and the generalized feature mapping process, regardless of the physical meaning of the raw sampled data, all RF data measurements are converted to a generalized data tensor for human activity recognition. Thus, the TARF system implements the same signal preprocessing framework for different wireless technologies, which means the system is also applicable to RF technologies other than the four used in this paper.

C. Activity Recognition with DANN

1) Challenge: Diversity in Motion Feature Translation: Since different RF technologies utilize different protocols and fre-

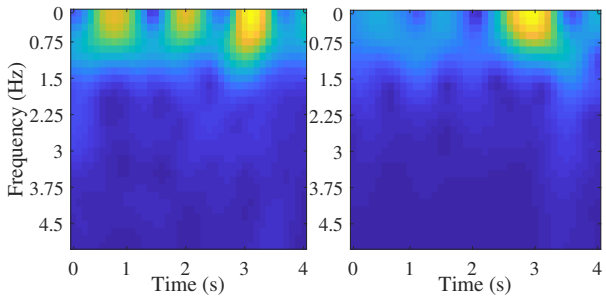


Fig. 4. Examples of one slice of the generalized feature tensor for the kicking activity. Left: sampled with FMCW radar; Right: sampled with 5GHz WiFi.

quency bands, the translation from received RF measurement to the target activity is highly diverse. Although the same source signal is generated by the same activity, it is transformed into very different RF data by the different protocols, frequency bands, and hardware. For example, with the RFID system, the human activity directly changes the positions of the RFID tags attached to the body. The change of tag position will introduce significant variation in the propagation delay τ. However, with the WiFi system, the human activity only affects part of the propagation paths of the OFDM channel. Furthermore, the different channel frequencies used in RFID and WiFi also generate large diversity in measured RF data. The considerable frequency diversity (i.e., 900 MHz in RFID and 2.4 GHz and 5 GHz in WiFi) causes large variation on the motion feature transformation in the raw sampled signals. Since such translation diversity is complicated and nonlinear, we propose a two-step solution to deal it i.e., (i) time-frequency (TF) domain transformation and tensorization, and (ii) a domain adversarial neural network model.

2) Proposed Solution Step 1: TF Domain Transformation and Tensorization: Given that human activity can be seen as a mixture of distinct periodic components [37], the characteristics recorded in the frequency domain are more universal than that in the time domain. To extract the generalized features from the reordered matrix $\mathbf{S}_R$, we perform Short Time Fourier Transform (STFT) on the reordered matrix to convert each row to a Time-Frequency (TF) domain matrix, and then construct a TF tensor with N_P slices. The TF domain data incorporates both frequency domain properties and variations over time. Fig. 4 presents one slice of the generalized TF tensor data, when the window size used in STFT is set to 16. Although sampled by two different RF technologies, i.e., FMCW radar and 5GHz WiFi, the TF domain data is well generalized with the proposed approach. Thus the same human activity will produce similar features as shown in the generalized feature tensor. Deep learning models can then be applied to classify different activities using such generalized TF tensor data.

3) Proposed Solution Step 2: Domain Adversarial Neural Network: We propose to use a domain adversarial deep neural network [38] to recognize human activities using the generalized feature tensors. Compared with the traditional CNN models, the domain adversarial neural network can further optimize the feature extractor with the domain discriminator. The network structure used in TARF is shown in Fig. 5, which is composed of the CNN based feature extractor, the activity predictor, and the domain discriminator.

a) Feature Extraction with CNN: The feature extractor used in the deep neural network is based on CNN. As a classic neural network structure, CNN can effectively extract features from all the slices in the generalized tensor. As Fig. 5 shows, the CNN feature extractor consists of two convolution layers, where all the convolutional kernels used for feature extraction have a size of 5×5. Each convolution layer is connected to a 2×2 max pooling layer to downsample the extracted feature. The final feature is formalized as a one-dimensional vector, which is used for the following activity predictor and domain discriminator. The generalized tensor used as the input is sampled every five seconds and transformed by 64-dot STFT. We only use data in the positive frequency domain, including 0 Hz, so the dimension of each slice is 33×50. The slice number is determined by N_p, which is equal to 30. We find the CNN-based feature extraction for all data slices may generate too many training variables, making the training time-consuming. Thus, we downsample the data tensor from 30 slices to 5 slices to reduce the complexity of network training. After the two feature extraction layers, the final feature is reshaped into a vector of 5616 elements.

b) Motion Identifier with Domain Discriminator: After extracting features using the CNN, the activity predictor and domain discriminator are applied, which consist of two fully connected layers. The final classification probability is calculated by the Softmax function. The loss function of the activity label predictor is calculated by the cross entropy between the Softmax output and the activity label as: $L_\alpha = \frac{1}{N_b} \sum_{b=1}^{N_b} \sum_{k=1}^{N_a} \hat{y}_k^b \log\left(y_k^b\right)$, where N_b is the number of training data in a batch, N_a is the number of classes of human activities, $\hat{y}_k^b$ denotes the estimated probability for class k with data sample b, and y_k^b is the class label which is either 0 or 1. L_α represents the accuracy of prediction, and the deep neural network is trained by minimizing L_α using the gradient descent algorithm.

In addition to the activity predictor, the domain adversarial neural network also employs a domain discriminator to combat the diversity between different domains, i.e., different RF technologies. The loss function of the domain discriminator, denoted by L_β, is calculated similarly as L_α: $L_\beta = \frac{1}{N_b} \sum_{b=1}^{N_b} \sum_{q=1}^{N_d} \hat{y}_q^b \log\left(y_q^b\right)$, where N_d indicates the number of RF technologies for data sampling and $\hat{y}_q^b$ denotes the estimation probability for the qth RF technology in the bth sample in the batch. Unlike the normal gradient descent learning algorithm used for maximizing L_β, the domain adversarial neural network performs a reversal gradient update for minimizing L_β, and the training variables of the network are updated as [38]: $\hat{X_\gamma} = X_\gamma - \xi\left(\frac{\partial L_\alpha}{\partial X_\gamma} - C_r \frac{\partial L_\beta}{\partial X_\gamma}\right)$, $\hat{X_\alpha} = X_\alpha - \xi \frac{\partial L_\alpha}{\partial X_\alpha}$, $\hat{X_\beta} = X_\beta - \xi C_r \frac{\partial L_\alpha}{\partial X_\beta}$, where X_γ denotes the training variables in the feature extractor; X_α and X_β represents the training variables for the label predictor and the domain discriminator, receptively; ξ denotes the learning rate; and C_r is the combating rate. The training goal for the feature extractor is to maximize L_β and minimize L_α, hence the feature extractor will be trained to ignore the

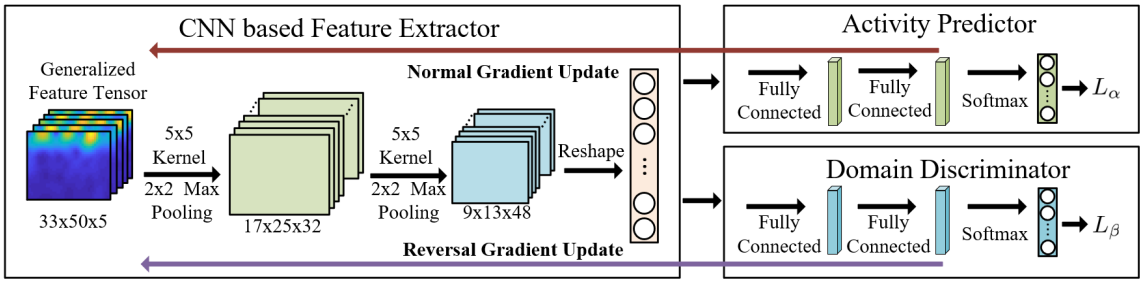


Fig. 5. Structure of the domain adversarial deep neural network used in the TARF system.

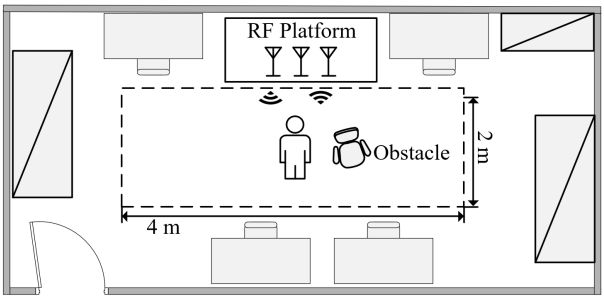


Fig. 6. The environment where human activity data is sampled.

domain-related features. Accordingly, the network will learn the generalized human activity related features and abandon the features associated with different RF technologies.

VI. Implementation and Evaluation

A. Experiments Setup

1) Hardware Platforms: To evaluate the proposed technology-agnostic HAR system, we develop a prototype using several RF technologies, including FMCW radar, 2.4GHz WiFi, 5GHz WiFi, and the UHF RFID system. The FMCW radar employed in the system, as shown in the figure, is an IWR1843 BOOST single-chip FMCW mmWave sensor that operates at 76 $\sim$ 81 GHz. The WiFi devices are integrated with a standard Intel 5300 network interface card (NIC), which operates at either 2.4 GHz or 5 GHz. The RFID platform consists of three S9028PCR polarized antennas, one Impinj R420 reader, and ALN-9634 (HIGG-3) passive RFID tags. An MSI laptop with an NVIDIA GTX 1080 GPU and an Intel Core i7-6820HK CPU are used for signal processing, model training, and inference.

2) Dataset Collection: RF data has been collected by sampling activities performed by a subject in front of the RF sensing platforms. The individual conducts seven types of different activities, including standing still, walking, running, squatting, body twisting, kicking, and hand waving. The data is sampled when the subject continuously repeats the activities. During the data acquisition using WiFi devices, the WiFi transmitter is set to the *injection* mode while the receiver is set to the *monitor* mode [20]. Two industrial, scientific and medical (ISM) bands, 2.472 GHz and 5.3 GHz, are used for the WiFi system, which allows us to examine the impact of different bands with the same WiFi protocol. RFID-based sampling is carried out with 12 passive RFID tags attached to the 12 joints on the subject's body, including neck, left shoulder, right shoulder, left elbow, right elbow, left wrist, right wrist, pelvis, left hip, right hip, left knee, and right knee. Three polarized antennae are used to interrogate these tags to ensure that each RFID tag is covered by at least one antenna. The FMCW radar employed in the investigations is a well-developed commodity mmWave sensor that produces range profiles for the scanned area. Each of these four RF technologies can independently sample human activities and the data is processed by the proposed TARF framework.

The detailed configuration of the experiment configuration is illustrated in Fig. 6. As the figure shows, the experiments are conducted in a lab. The subject performs different activities in the 2m$\times$4m scanning area as shown in Fig. 6. In the experiments, we emulate three different deploy scenarios, which includes an LOS scenario, an NLOS scenario, and a dynamic environment. The LOS scenario is to deploy the system in a clean scanning area, while the NLOS scenario is to adding chairs between the subject and the RF platforms. Although the LOS propagation is not entirely eliminated in this case, the obstacle, i.e., the chairs, effectively attenuates the strength of the LOS signal. Further, the dynamic environment is introduced by having another subject moving around the tested subject when RF data is being collected. We sample one-hour of data for each activity with each RF technology. That is, with seven activities and four RF technologies, we sampled 7 hours of data for each RF technology and 28 hours of data in total. 90% of the sampled data is used for model training, and the remaining 10% is used for testing.

B. Performance with Different RF Technologies

To analyze the experimental results, we define the number of correctly classified data samples as the true positive number (TP), and the number of mistakenly recognized results as the false negative number (FN). The true positive rate (TPR) and false negative rate (FNR) are calculated as: $TPR = \frac{TP}{TP+FN}$ and $FPR = \frac{FN}{TP+FN}$. The overall evaluation result is presented in the confusion matrix format, which is composed of the TPRs and FPRs for all the seven types of activities. The overall accuracy η is calculated by: $\eta = \frac{\sum_{i=1}^{7} TP_i}{\sum_{i=1}^{7}(TP_i+FN_i)}$, where TP_i and FN_i denotes the true positive number and false negative number for target activity i, respectively. For convenience, we label different activities with the following acronyms: standing still–*ST*, walking–*WA*, running-*RU*, squatting–*SQ*, body twisting–*BT*, kicking–*KI*, and hand waving–*HW*.

To demonstrate the performance of the TARF system, we evaluated it using different combinations of the RF plat-

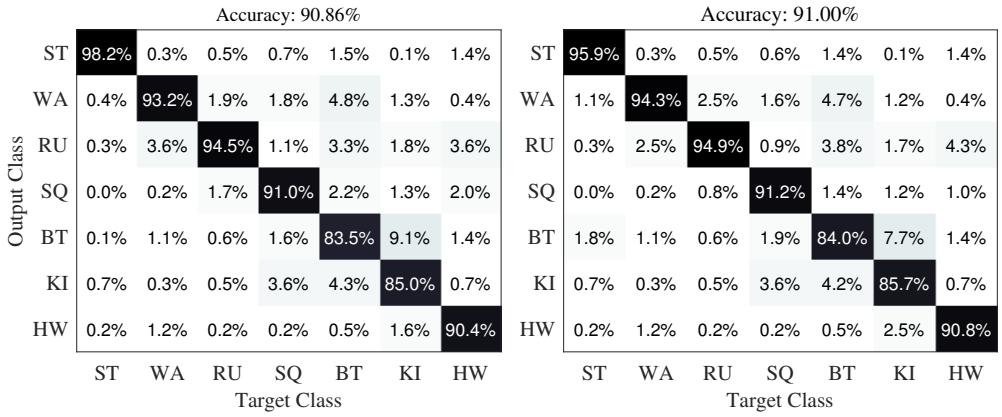


Fig. 7. Confusion matrix of human activity recognition with a single RF technology (FMCW radar). Left: CNN baseline scheme; Right: TARF.

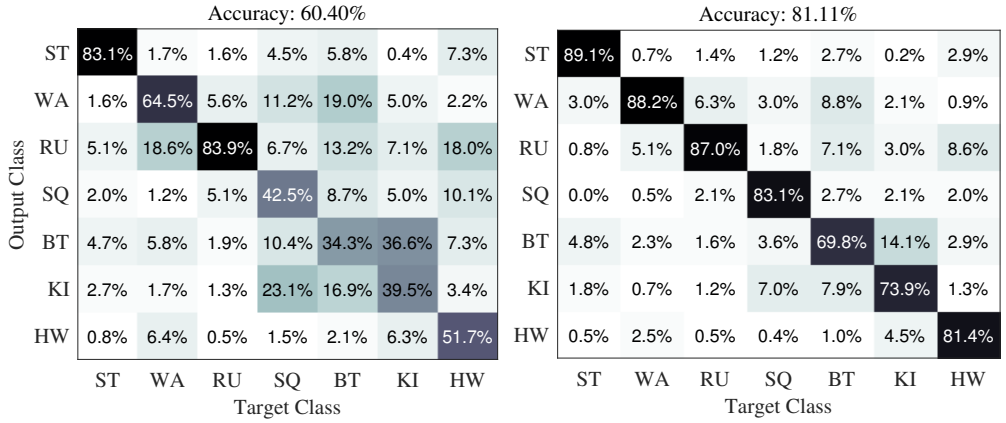


Fig. 8. Confusion matrix of human activity recognition obtained using TARF.

forms, ranging from using a single RF technology to using all four RF technologies. The baseline for comparison is the traditional CNN based classification network [39]. The CNN network is composed of the same feature extractor and activity predictor as in TARF, but without the domain discriminator or reversal gradient update. Fig. 7 presents the confusion matrices obtained with a single RF technology (i.e., the FMCW radar). The left confusion matrix is the result obtained by the traditional CNN based approach, and the right confusion matrix is generated by the proposed technology-agnostic TARF system. The overall accuracy is 90.86% for the baseline CNN method and 91.00% for the proposed TARF approach. These results demonstrate that both CNN and TARF can effectively distinguish the seven types of human activity. This is because, when there is only one data domain, the influence of the domain discriminator could be ignored. Therefore, the performance of domain adversarial deep neural network is equivalent to that of the traditional CNN model.

We next examine the case when all the four RF sensing technologies are used for data acquisition. Fig. 8 presents the confusion matrices when all four technologies are utilized for human activity recognition. The confusion matrix on the left is obtained with the CNN baseline method, whose accuracy is significantly reduced from 90.86% in Fig. 7 to 60.40% here. The *TPR* of identifying body twisting and kicking is mostly affected, which drop to 34.3% and 39.5%, respectively. The confusion matrix shows that, with the data sampled with four

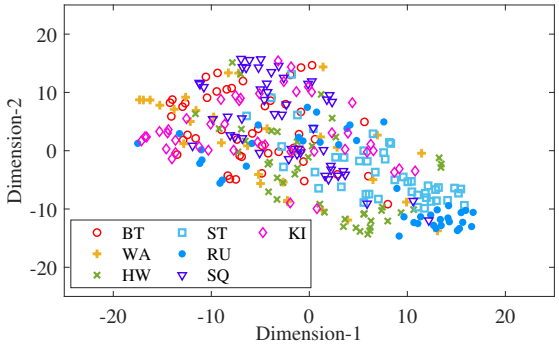


Fig. 9. T-NSE illustration of human activity recognition using four RF technologies obtained with the CNN baseline scheme.

different RF technologies, the CNN method fails to effectively learn the generalized features of different human activities.

The confusion matrix obtained with the proposed TARF approach is presented on the right side of Fig. 8. In contrast, TARF still achieves an overall accuracy of identification of 81.11%, although still affected by using the more diverse RF data collected from four different platforms. Such robustness to diverse RF data is achieved by the domain discriminator used in TARF. The domain discriminator can prevent the network from learning the domain-related features, and thus the technology-agnostic learning approach is quite effective to adapt to different RF technologies.

We also perform T-distributed Stochastic Neighbor Embed-

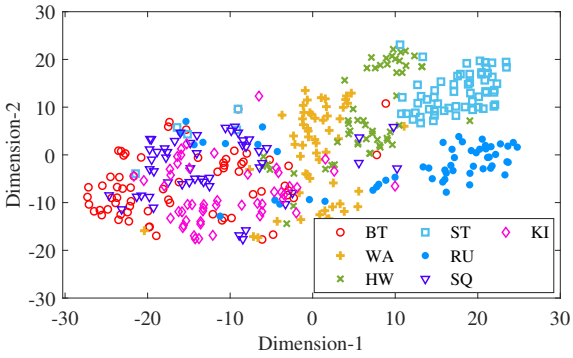


Fig. 10. T-NSE illustration of human activity recognition using four RF technologies obtained with TARF.

TABLE I
ACCURACY WITH DIFFERENT COMBINATION OF RF TECHNOLOGIES (1: RFID ONLY; 2: RFID AND 2.4GHZ WIFI; 3: RFID, 2.4GHZ WIFI, AND 5GHZ WIFI; 4: RFID, 2.4GHZ AND 5GHZ WIFI, AND FMCW RADAR)

No. of RF Technologies	One	Two	Three	Four
CNN Baseline	90.86%	93.73%	75.65%	60.4%
TARF	91.00%	88.21%	85.41%	81.11%

ding (T-SNE) on tested data for the CNN baseline scheme and TARF. T-SNE is an effective approach for visualizing high dimensional data introduced by the feature extractor. As shown in Fig. 9 for the CNN baseline scheme, the data collected from the same activity has not been classified effectively. Except for a few activities such as standing still and running, the features of other human activities are not satisfactorily extracted due to the interference from various RF technologies. As shown in Fig. 10 for TARF, the data from different human activities are better grouped in the 2D map. Although there are still some overlap between data groups of body-twisting, squatting, and kicking, the RF data sampled from different activities are more effectively grouped. The visualization results of T-SNE intuitively demonstrate that the domain discriminator effectively optimizes the activity feature extractor by mitigating the interference from the diverse RF technologies.

The superiority of the TARF system is further demonstrated in Table I, which shows the variations in accuracy as the number of RF technologies is increased from one to four. The figure shows that both CNN based scheme have similar classification accuracy when single RF technology is involved in the system, which are 90.86% and 91.00%, respectively. The figure also shows that using more RF technologies introduces increased diversity in the acquired data, which affects the classification performance. However, compared with the CNN baseline, the proposed TARF system is effective in combating such diversity and adapting to different data domains.

C. Evaluation Under Different Scenarios

We also evaluate the proposed TARF system in different scenarios and compare it with the HAR system trained by RF data from a single RF technology. In these experiments, we intend to investigate system performance with three different deployment scenarios, including an LOS testing scenario, an NLOS testing scenario, and a dynamic RF environment testing

TABLE II
ACCURACY IN THE LOS TESTING SCENARIO

RF Technology	WiFi 2.4G	WiFi 5G	FMCW	RFID
Single RF Technology	91.86%	89.37%	91.22%	90.73%
CNN Baseline	59.71%	63.41%	62.12%	60.89%
TARF	81.86%	80.34%	81.22%	82.73%

TABLE III
ACCURACY IN THE NLOS TESTING SCENARIO

RF Technology	WiFi 2.4G	WiFi 5G	FMCW	RFID
Single RF Technology	90.76%	88.71%	81.77%	71.22%
CNN Baseline	58.11%	56.35%	61.29%	51.79%
TARF	81.24%	80.61%	76.88%	71.84%

scenario. The NLOS environment is emulated by adding obstacles between the subject and the RF platforms, and the dynamic RF environments are emulated by introducing another subject moving around. We trained the network with RF data from a single RF platform to generate the corresponding technology-specific system baseline, while the technology-agnostic schemes are trained with data from all the four RF technologies. Table II illustrates the activity recognition accuracy of different systems in the LOS testing scenario. The figure shows that all technology-specific schemes can achieve a satisfactory activity recognition accuracy in the LOS testing scenario, which are all over 89.37%. The performance of technology-generalized systems, such as the CNN baseline and TARF, is worse than the technology-specific system. However, as a generalized system, the accuracy of TARF is still much higher than the CNN baseline. Although TARF does not outperform the single RF technology scheme, it has the unique advantage of being technology-agnostic. The lowest recognition accuracy of TARF, i.e., 80.34%, is achieved when tested with WiFi 5G, and its highest accuracy of 82.73% is achieved with RFID.

However, the LOS testing scenario may be too ideal for practical applications. So we also evaluate the system in the NLOS scenario and the dynamic, noisy environment. Tables III and IV present the accuracy results when the test data is sampled from the NLOS and dynamic RF environments, respectively. From the figures, we find that the impacts of the NLOS environment on different RF technologies are different. Among the four RF technologies, WiFi-based schemes, such as 2.4GHz and 5GHz, can still achieve high accuracy, but the accuracy of FMCW and RFID-based systems drops to 81.77% and 71.22%, respectively. This is because the WiFi signal scatters better with a larger range, so its NLOS component can still convey informative features of human activities. In contrast, due to the limited range of FMCW radar and RFID reader, the LOS component is dominant but blocked. Especially for the RFID system, most tags cannot be effectively interrogated as blocked by the obstacle.

When it comes to dynamic RF environments, the performance of each single RF system becomes very different from before. As Table IV shows, the accuracy of the two WiFi schemes is significantly degraded by the interference caused by the moving person. The accuracy of the WiFi-

TABLE IV
ACCURACY IN THE DYNAMIC RF ENVIRONMENT TESTING SCENARIO

RF Technology	WiFi 2.4G	WiFi 5G	FMCW	RFID
Single RF Technology	75.05%	71.44%	79.29%	89.38%
CNN Baseline	43.54%	42.85%	55.02%	58.54%
TARF	71.85%	70.06%	77.76%	80.18%

TABLE VI
PERFORMANCE COMPARISON OF THE GENERALIZED MATRIX-BASED APPROACH AND THE PROPOSED STFT TENSOR-BASED APPROACH

No. of RF Technologies	One	Two	Three	Four
Matrix-based	90.86%	84.23%	78.12%	71.32%
STFT Tensor-based	91.00%	88.21%	85.41%	81.11%

TABLE VII
ACTIVITY RECOGNITION ACCURACY WHEN DIFFERENT NUMBERS OF MEASUREMENTS ARE USED

	10	15	20	25	30	35
Single	85.86%	87.37%	89.44%	90.86%	91.62%	90.93%
Four	75.28%	77.42%	78.44%	80.21%	81.11%	80.97%

specific schemes decreases to 75.05% and 71.44% for 2.4GHz WiFi and 5GHz WiFi, respectively. However, the accuracy of the RFID system remains relatively high at 89.38%. This is because the RFID tags attached to the subject's clothes can convey reliable human movement features, which are more robust to the dynamic environment than WiFi-based schemes.

The results in Tables III and IV verify that a single RF technology does not adapt well to different kinds of testing environments. In contrast, an effective technology-agnostic system can leverage all the accessible RF technologies that are complementary to each other. Table V presents the accuracy results of the four single RF technology schemes, the CNN baseline, and TARF. The table shows that, when all the accessible RF technologies are used in the generalized system, TARF achieves 81.24% and 80.18% activity recognition accuracy for the NLOS and the dynamic environments, which is comparable to that in the ideal LOS testing scenario.

D. Impact of the Generalized Feature Tensor

We also conduct experiments to examine the benefit of utilizing the extended STFT feature tensor, and to establish the most appropriate tensor-related parameters. The accuracy performance of human activity recognition is presented in Table VI, where the blue bars are the results obtained by just utilizing the generalized matrix $\mathbf{S}_R$ as input to the deep neural network. It can be seen that using the generalized matrix can achieve a 90.86% recognition accuracy in a single-technology situation, but the accuracy drops dramatically to 71.32% when all the four technologies are used. The proposed STFT tensor-based technique results are represented by the green bars, which degrades from 91.00% to 81.11% instead, and is more resilient to the influence of various sensing technologies. The robustness demonstrated by these results validates that the proposed STFT feature tensor can successfully extract the general characteristics of human behavior from diverse RF data collected by different RF technologies.

We also conduct experiments to explore appropriate parameter setting for the proposed TARF system. Tables VII and VIII show the impacts of the measurement number N_P and size of the sliding window, respectively. As Table VII shows, the accuracy increases when more measurements are used for feature extraction. In the single-technology scenario, the accuracy is over 90% when $N_P \geq 25$, and the highest accuracy 91.62% is achieved when $N_P = 30$. Similarly, In the four-technology scenario, the highest accuracy is achieved when 30 measurements are used for tensor generation. Table VIII shows the accuracy when different sliding window sizes are employed for STFT. The figure shows that in both scenarios, high accuracy is achieved when the sliding windows is 6 seconds. However, we notice that when the window size is 5 seconds, the accuracy in the two scenarios are 91.00% and 81.11%, respectively, which is sufficiently high for human activity recognition. To reduce the training complexity, we choose the smallest sliding window size, which still achieves an acceptable system accuracy for STFT. As a result, we set N_P to 30 and the window size to 5 seconds. Although the STFT tensor requires 5 seconds of RF data, the sliding window structure allows the system to perform activity recognition for each newly sampled RF data. Furthermore, since the classification of the trained DANN is executed very fast, TARF is suitable for realtime tracking of human activities.

VII. CONCLUSION

In this paper, we proposed a generalized approach to human activity recognition, termed TARF, to mitigate the impact of technology-agnostic data acquisition. A novel signal preprocessing solution was proposed to combat the diversity caused by different RF sensing platforms. The generalized tensor construction method was proposed to break the barriers of RF data collected using different RF technologies and extract the generalized features related to human activities. We then utilize a domain adversarial neural network to address the diversity issue of motion feature translation in different RF platforms. The experiments results demonstrate that TARF can be effectively implemented with various RF devices so that different RF technologies can complement each other. The technology-agnostic scheme can achieve robust HAR performance in different scenarios by incorporating all accessible RF technologies.

REFERENCES

[1] A. Subasi, M. Radhwan, R. Kurdi, and K. Khateeb, "IoT based mobile healthcare system for human activity recognition," in *Proc. 15th Learning and Technol. Conf.*, Jeddah, Saudi Arabia, Feb. 2018, pp. 29–34.
[2] R. Liu, A. A. Ramli, H. Zhang, E. Datta, E. Henricson, and X. Liu, "An overview of human activity recognition using wearable sensors: Healthcare and artificial intelligence," *arXiv preprint arXiv:2103.15990*, Aug. 2021. [Online]. Available: https://arxiv.org/abs/2103.15990
[3] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–36, June 2019.
[4] X. Wang, C. Yang, and S. Mao, "Tensorbeat: Tensor decomposition for monitoring multi-person breathing beats with commodity WiFi," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 1, pp. 8:1–8:27, Sept. 2017.
[5] F. Wang, J. Liu, and W. Gong, "Multi-adversarial in-car activity recognition using RFIDs," *IEEE Trans. Mobile Comput.*, 2020, in press. DOI: 10.1109/TMC.2020.2977902.

TABLE V
ACCURACY COMPARISON WITH DIFFERENT TESTING SCENARIOS

Testing Environment	*WiFi 2.4GHz*	*WiFi 2.4GHz*	*FMCW*	*RFID*	*CNN Baseline*	*TARF*
LOS	91.86%	89.37%	91.22%	90.73%	63.41%	82.73%
NLOS	90.76%	88.71%	81.77%	74.22%	61.29%	*81.24%*
Dynamic Environment	75.05%	71.44%	79.29%	89.38%	62.54%	*80.18%*

TABLE VIII
ACTIVITY RECOGNITION ACCURACY WHEN DIFFERENT SLIDING WINDOW SIZES ARE USED

Sliding Window Size	2	3	4	5	6	7	8
Single RF Technology	81.86%	86.73%	86.65%	91.00%	91.65%	89.63%	90.59%
Four RF Technologies	70.27%	74.55%	78.32%	81.11%	83.23%	80.17%	81.98%

[6] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*, Los Cabos, Mexico, Oct. 2019, pp. 51–56.

[7] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1629–1645, Aug. 2019.

[8] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *MDPI Remote Sensing*, vol. 11, no. 9, p. 1068, Apr. 2019.

[9] F. Adib and D. Katabi, "See through walls with WiFi!" in *Proc. ACM SIGCOMM 2013*, Hong Kong, China, Aug. 2013, pp. 75–86.

[10] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," in *Proc. USENIX NSDI'14*, Seattle, WA, Apr. 2014, pp. 317–329.

[11] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Proc. ACM CHI 2015*, Seoul, Republic of Korea, Apr. 2015, pp. 837–846.

[12] X. Yang, J. Liu, Y. Chen, X. Guo, and Y. Xie, "MU-ID: Multi-user identification through gaits using millimeter wave radios," in *Proc. IEEE INFOCOM'20*, Toronto, Canada, Aug. 2020, pp. 2589–2598.

[13] U. Ha, S. Assana, and F. Adib, "Contactless seismocardiography via deep learning radars," in *Proc. ACM MobiCom 2020*, London, UK, Sept. 2020, pp. 1–14.

[14] H. Xue, Y. Ju, C. Miao, Y. Wang, S. Wang, A. Zhang, and L. Su, "mmMesh: towards 3D real-time dynamic human mesh construction using millimeter-wave," in *Proc. ACM MobiSys 2021*, Virtual Conf., June 2021, pp. 269–282.

[15] Z. Chen, T. Zheng, C. Cai, and J. Luo, "MoVi-Fi: motion-robust vital signs waveform recovery via deep interpreted RF sensing," in *Proc. ACM MobiCom 2021*, New Orleans, LA, Oct. 2021, pp. 392–405.

[16] S. Ding, Z. Chen, T. Zheng, and J. Luo, "RF-net: A unified meta-learning framework for RF-enabled one-shot human activity recognition," in *Proc. ACM SenSys'20*, Virtual Conf., Nov. 2020, pp. 517–530.

[17] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 159–170, Aug. 2010.

[18] S. M. Hernandez and E. Bulut, "Performing WiFi sensing with off-the-shelf smartphones," in *Proc. IEEE PerCom'20 Workshops*, Austin, TX, Mar. 2020, pp. 1–3.

[19] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free your CSI: A channel state information extraction platform for modern Wi-Fi chipsets," in *Proc. ACM WiNTECH'20*, Los Cabos, Mexico, Oct. 2019, pp. 21–28.

[20] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity WiFi," in *Proc. ACM Mobicom'15*, Paris, France, Sept. 2015, pp. 53–64.

[21] X. Jiao, M. Mehari, W. Liu, M. Aslam, and I. Moerman, "Openwifi CSI fuzzer for authorized sensing and covert channels," *arXiv preprint arXiv:2105.07428*, May 2021. [Online]. Available: https://arxiv.org/abs/2105.07428

[22] X. Wang, X. Wang, and S. Mao, "RF sensing for Internet of Things: A general deep learning framework," *IEEE Communications Mag.*, vol. 56, no. 9, pp. 62–69, Sept. 2018.

[23] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using WiFi channel state information," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, June 2017.

[24] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, "Multimodal CSI-based human activity recognition using GANs," *IEEE Internet of Things Journal*, vol. 8, no. 24, pp. 17 345–17 355, Dec. 2021.

[25] W. Jiang, H. Xue, C. Miao, S. Wang, S. Lin, C. Tian, S. Murali, H. Hu, Z. Sun, and L. Su, "Towards 3D human pose construction using WiFi," in *Proc. ACM MobiCom'20*, London, UK, Sept. 2020, pp. 1–14.

[26] C. Yang, X. Wang, and S. Mao, "Respiration monitoring with RFID in driving environments," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 2, pp. 500–512, Feb. 2021.

[27] C. Yang, X. Wang, and S. Mao, "Unsupervised drowsy driving detection with RFID," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8151–8163, Aug. 2020.

[28] C. Yang, X. Wang, and S. Mao, "RFID-Pose: Vision-aided 3D human pose estimation with RFID," *IEEE Trans. Reliability*, vol. 70, no. 3, pp. 1218–1231, Sept. 2021.

[29] C. Yang, L. Wang, X. Wang, and S. Mao, "Meta-Pose: Environment-adaptive human skele-ton tracking with RFID," in *Proc. IEEE GLOBECOM 2021*, Madrid, Spain, Dec. 2021, pp. 1–6.

[30] W. Jiang *et al.*, "Towards environment independent device free human activity recognition," in *Proc. ACM MobiCom 2018*, New Delhi, India, Sept. 2018, pp. 289–304.

[31] 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. ICML'17*, Sydney, Australia, Aug. 2017, pp. 4100–4109.

[32] H. Xue, W. Jiang, C. Miao, F. Ma, S. Wang, Y. Yuan, S. Yao, A. Zhang, and L. Su, "DeepMV: Multi-view deep learning for device-free human activity recognition," *Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technol.*, vol. 4, no. 1, pp. 1–26, Mar. 2020.

[33] F. Wang, J. Liu, and W. Gong, "WiCAR: WiFi-based in-car activity recognition with multi-adversarial domain adaptation," in *Proc. IWQoS'19*, Phoenix, AZ, June 2019, pp. 1–10.

[34] C. Iovescu and S. Rao, "The fundamentals of millimeter wave sensors," whitepaper, Texas Instruments, July 2017. [Online]. Available: https://www.ti.com/lit/wp/spyy005a/spyy005a.pdf

[35] V. Winkler, "Range doppler detection for automotive fmcw radars," in *Proc. 2007 European Radar Conf.*, Munich, Germany, Oct. 2007, pp. 166–169.

[36] M. Lenehan, "Application note - Low level user data support," Feb. 2019. [Online]. Available: https://support.impinj.com/hc/en-us/articles/202755318-Application-Note-Low-Level-User-Data-Support

[37] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. ACM Mobicom'15*, Paris, France, Sept. 2015, pp. 65–76.

[38] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *J. Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, Apr. 2016.

[39] F. M. Noori, M. Riegler, M. Z. Uddin, and J. Torresen, "Human activity recognition from multiple sensors data using multi-fusion representations and CNNs," *ACM Trans. Multimedia Comput., Commun., and Appl.*, vol. 16, no. 2, pp. 1–19, May 2020.