A Simulation-Based Framework for the Design of Human Activity Recognition Systems Using Radar Sensors

Sahil Waqar⬚ and Matthias Pätzold⬚

Abstract—**Modern human activity recognition (HAR) systems are designed using large amounts of experimental data. So far, real-data-driven or experimental-based HAR systems using Wi-Fi or radar systems have shown considerable results. However, the acquisition of large, clean, and labeled training data sets remains a crucial impediment to the progress of experimental-based HAR systems. Therefore, in this article, a paradigm shift from the experimental to a fully simulation-based design of HAR systems is proposed in the context of radar sensing. An end-to-end simulation framework is proposed as a proof-of-concept that can simulate realistic millimeter-wave radar signatures for synthesized human motion. We designed a human motion synthesis tool that emulates different types of human activities and generates the spatial trajectories accordingly. These trajectories are processed by a geometric model with respect to user-defined antenna configurations. Considering the long- and short-time stationarity of wireless channels, we synthesize the raw in-phase and quadrature data and process the data to simulate the radar signatures for emulated human activities. Finally, a simulated and a real HAR data set were used to train and test a simulation-based HAR system, respectively, which gave an average (maximum) classification accuracy of 94% (98.4%). The main advantage of the proposed simulation framework is that the training effort for radar-based classifiers, e.g., gesture recognition systems, can be minimized drastically.**

Index Terms—**Data augmentation, data generation, deep learning, human activity recognition (HAR), micro-Doppler analysis, Mixamo animation, motion synthesis, multiclass classification, radar simulation.**

I. INTRODUCTION

A. Background

WITHIN the domain of radio frequency (RF) sensing, an important and continuously evolving research area is human activity recognition (HAR), where the classification performance greatly depends on the quality, impartiality, and comprehensiveness of experimental data. Such merits of empirical data are hard to come by, especially when dealing with real humans as subjects. Over the years, researchers have endeavoured to classify different types of human activities using several sensing modalities, such as vision [1], [2], wearable [3], [4], [5], and RF sensors [6], [7], [8], [9], [10].

Various sensor types have been employed in HAR systems, each with distinct advantages and limitations [11]. Vision sensors, driven by advanced computer vision methods, have shown significant success in HAR [1]. However, they are vulnerable to lighting conditions and privacy concerns, unlike RF sensors. Wearable sensors [4], though effective, face criticism due to their fragility, intrusiveness, and reliance on user care. The need for continuous wear renders them impractical, particularly for elderly and ill individuals. Hence, RF sensors, particularly millimeter wave (mm-wave) radars, have garnered growing interest despite the challenges and complexities they entail [12].

In this research, we primarily focus on developing a human activity classification system using mm-wave radar technology. The collection of radar micro-Doppler signatures corresponding to real human subjects is a time-consuming, expensive, and laborious task. The recorded radar data set usually has a narrow scope because of its validity for a particular scenario and fixed radar parameters. To create diverse training data sets for radar-based HAR systems, a simulation-based approach becomes a compelling and viable alternative.

We design a fully simulation-based HAR system that exclusively relies on simulated radar data for training and validation. Unlike conventional methods, we avoid the use of real radar micro-Doppler signatures during these stages. Instead, experimental measurements from a real mm-wave radar system are solely employed for testing, showcasing our simulation-based HAR system's real-world performance. To ensure accurate radar system modeling and realistic radar micro-Doppler signature simulation, we adopt scatterer-level signal modeling (see Section VI). This proof-of-concept approach facilitates the generation of diverse simulated radar micro-Doppler signatures, thereby providing essential training data for simulation-based HAR systems.

B. Our Approach

In this article, we present an end-to-end simulation framework for HAR using frequency-modulated continuous wave (FMCW) radar systems. First, we design a human motion synthesis tool using the Unity software [13] from Unity Software Inc. that emulates different types of human activities and accordingly generates the 3-D trajectories for the virtual

markers placed on a humanoid character. The 3-D marker trajectories are processed by a geometric model (see Section V) with respect to a user-defined antenna configuration. Taking into account the long- and short-time stationarity properties of wireless channels and using our radar signal synthesizer, we simulate the raw in-phase and quadrature (IQ) components. Finally, the radar micro-Doppler signatures or, equivalently, the time-variant (TV) radial velocity distributions are generated for several types of emulated human activities.

Our proposed simulation-based framework offers several advantages over experimental-based designs, such as flexibility, to simulate radar data sets with specific distributions or target motion characteristics, ability to augment training data, cost-effectiveness, and mitigation of legal and privacy issues. With the proposed simulation framework, we can augment human motion data at a motion-synthesis layer, e.g., by varying an avatar's size and speed. The proposed simulation framework gives control over several radar parameters as well, thus it enables us to generate different types of training data sets corresponding to different radar-operating conditions and different applications. Above all, the proposed simulation framework drastically minimize the overall training effort of radar-based HAR systems.

Note that our simulation-based framework, basically designed for HAR, has versatile applications across various domains, including gesture recognition [14], sports [15], autonomous vehicles [16], social robotics [17], and smart homes [18]. In this research, our validation process involves real experiments covering five human activities, highlighting the effectiveness of our proof-of-concept. The core strength of this simulation-based framework, however, lies in its innovative capability to translate motion capture (MoCap) data into radar data (see Sections IV and VI), making it adaptable to a wide array of real-world scenarios. The availability of extensive online MoCap data repositories like Mixamo [19], covering domains, such as sports, multimedia, healthcare, and more, further enhances the framework's applicability. With our proposed framework, these repositories can be used to simulate radar signatures for a multitude of real-world scenarios. For instance, in healthcare, we demonstrate the framework's capability for fall detection, providing a tangible example of its real-world utility. In sports, our solution can be extended to simulate radar signatures for activities, such as running, swimming, and various exercises, thereby enhancing its practicality.

Changes in radar configurations in practical applications, driven by shifts in operational requirements, technological advancements, and emerging applications, necessitate the generation of new data sets. For emerging radar-based classifiers, the need to simulate new data sets is inevitable as it aligns with the dynamic nature of radar sensors. This constraint is common to all radar-based classifiers, whether realized using simulation or experimental data. Our simulation-based framework stands out for its efficient and rapid generation of diverse data sets for new or modified radar configurations, presenting a more resource-efficient alternative compared to the classifiers based on experimental data. The ability to

swiftly and easily adapt to varied radar configurations stands as a distinctive strength of our proposed framework.

C. Contributions

The multiple contributions of this research can be summarized as follows.

1) We propose a novel end-to-end simulation framework to avoid the need of real radar data for training. By using the proposed simulation framework, large quantities of realistic synthetic radar data are generated for human gross motor activities. It is worth noting that the proposed simulation framework is also useful for many other radar-based classifiers, for instance, gesture recognition systems.

2) We leverage a geometrical 3-D indoor channel model (see Section V) to simulate TV radial distances from the spatial trajectories of an avatar with 21 nonstationary virtual markers. By employing the proposed approach, we emulate and diversify various human activities by varying parameters, such as location, speed, acceleration, deceleration, and avatar's height. Our unique simulation framework offers the flexibility to augment data at the motion-synthesis layer, enabling the generation of diverse and customizable data sets for training HAR systems.

3) We simulate high-fidelity radar signatures, namely, TV range distribution, TV radial velocity distribution, and TV mean velocity for the emulated human activities. By computing the dynamic time warping (DTW) distance metric [20], it is shown that the simulated radar signatures closely resemble the radar signatures measured in reality. This shows the effectiveness of our simulation framework, which can simulate realistic radar signatures for adults and children alike, and can even be extended to simulate realistic radar data for animals, vehicles and airplanes.

4) For the radar-signal synthesis, we expound the long- and short-time stationarity properties of the indoor wireless channel (see Section VI). The short-time stationarity assumption is quite advantageous because it significantly simplifies the synthesis of the radar signal.

5) Through our proposed approach, we establish a novel simulated HAR data set to train our simulation-based HAR system, which was developed by using a deep convolutional neural network (DCNN). This data set comprises simulated radar signatures computed from emulated human activities. To demonstrate the practical relevance of our simulation-based HAR system, we tested its performance on unseen data acquired with a real mm-wave FMCW radar system involving real persons. The mean (maximum) classification accuracy of the fully simulation-based HAR system was 94% (98.4%). The classification performance of the proposed simulation-based HAR system over the experimental data set demonstrates the utility and efficacy of our proposed end-to-end simulation framework.

D. Paper Organization

This article is structured as follows. Section II presents the related work, and Section III gives an overview of conventional and the simulation-based HAR systems. Our human motion synthesis module is elucidated in Section IV. Section V details the 3-D geometrical model. The synthesis of realistic radar data is explained in Section VI. Section VII describes the processing of the radar data. The design, training, and testing of the simulation-based HAR system is detailed in Section VIII. Finally, Section IX draws the conclusions.

II. RELATED WORK

Recently, the availability of commercial mm-wave sensors has led to the development of numerous human-centric research areas. For instance, many studies have been conducted on radar-based HAR systems [21], [22], [23], sign language [24] and gesture [25], [26] recognition systems. So far, most of the studies have focused on HAR systems that are realized by utilizing the scarcely available recorded radar data [27]. In [27], for instance, the HAR classifier was based on a long short-term memory (LSTM) neural network and was trained on manually labeled 3-D point cloud data. Zhao et al. [28] addressed the problem of HAR in multiangle scenarios by exploiting measured characteristics of a mm-wave radar, such as received power, range, Doppler frequencies, azimuth, and elevation. Another problem with experimental data collected with radar systems is the reusability of the data. Generally, the recorded data of the radar system is not reusable due to its fixed operating parameters and antenna configurations. When the operating conditions of the radar system are fixed, the few-shot learning scheme [29] is useful to enhance the capability of the already trained HAR system.

To address the lack of real radar data, some studies have suggested to use data augmentation techniques. For instance, Yang et al. [30] proposed a data augmentation technique based on a generative adversarial network (GAN) to create diverse micro-Doppler signatures for human activities. Apart from GANs, a self-supervised HAR approach has been recently proposed to tackle the issue of limited labeled data [31]. Li et al. [32] presented a technique called supervised few-shot adversarial domain adaptation for HAR. This approach addresses the challenge of having a limited amount of radar training data available for a particular scenario. Moreover, Yu et al. [33] proposed a rotational shift method to augment radar point cloud data. Recently, a two-stage domain adaptation scheme was presented in [34] to address the lack of training data for radar-based HAR systems. For data augmentation, they used a GAN-based domain-translation network to translate the simulated spectrograms into measurement-like spectrograms with the help of small measurement data sets. Even with this data augmentation technique, it is not possible to completely get rid of the tedious data collection process that requires real radar data sets and real human subjects.

The lack of publicly available real radar data sets, the limited reusability of radar data, and the resource-intensive data collection are the main factors driving us to pursue fully

simulation-based HAR system development. So far, only a handful of studies have been carried out in this direction. To model the intricate details of human motion, high-fidelity MoCap systems are preferred to eventually reanimate more realistic and complex human motion [35]. Chen et al. [36] formulated Doppler modulations and established equations for micro-Doppler effects caused by various micro motions, such as vibration, rotation, tumbling, and coning. They validated these formulations through simulation studies. A simulation tool has been developed recently called SimHumalator, which simulates the target echoes for passive Wi-Fi radar (PWR) scenarios [37]. Manfredi et al. [38] developed a simulation tool that characterizes the near-field radar cross section of a walking person in the K-band, but the approach is not suitable to model the finer details of human motion.

III. SYSTEM OVERVIEW

In the following, let us first describe the basic building blocks of a conventional HAR system, which is employed solely to evaluate the proposed simulation-based HAR system.

A. Conventional HAR System

The building blocks of a conventional (experimental-based) HAR system are depicted in Fig. 1(a). For each human activity, the mm-wave radar system produces real raw IQ data. The IQ data is subsequently processed by the radar signal processing module to generate the real micro-Doppler signature or, equivalently, the TV radial velocity distribution capturing the characteristics of a human activity (see Section VII). The TV radial velocity distributions of the recorded human activities are stored in files and represent the real radar data set, as shown in Fig. 1(a). Generally, the real radar data samples are used to train the experimental-based HAR classifier. Subsequently, a portion of the real radar data is used to test the performance of experimental-based HAR classifier, as shown in Fig. 2. In this research, we are mainly interested in devising a simulation-based HAR system that matches the performance of state-of-the-art HAR systems. Thus, we will only use the entire recorded data set from our conventional HAR system to test our proposed simulation-based HAR system, as shown in Fig. 2. An overview of the proposed simulation-based HAR system is presented in the following section.

B. Proposed Simulation-Based HAR System

Conventionally recorded training data sets may not be reusable as they are only valid for specific radar parameters and a specific antenna configurations. A revision or redesign of even a single radar parameter may render the training data set useless, e.g., the redesign of the radar system using a different pulse repetition interval (PRI). Therefore, a pragmatic alternative is proposed in this article to overcome the aforementioned issues associated with the acquisition of large training data sets. We propose a fully simulation-based approach, as shown in Fig. 1(b), to develop a real-world HAR system. For training the HAR classifier, our simulation-based approach enables easy generation of a large amount of
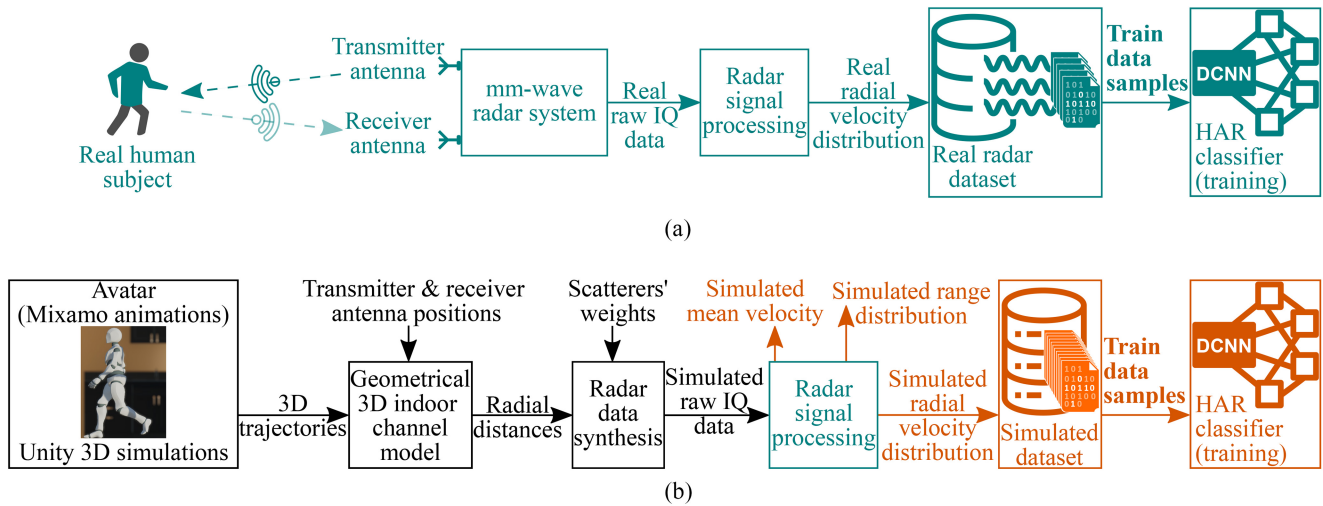
Fig. 1. (a) Design of conventional HAR systems that require real human subjects and a real radar system for training. (b) Design of the proposed simulation-based HAR system that only needs the simulated radar data set for training.
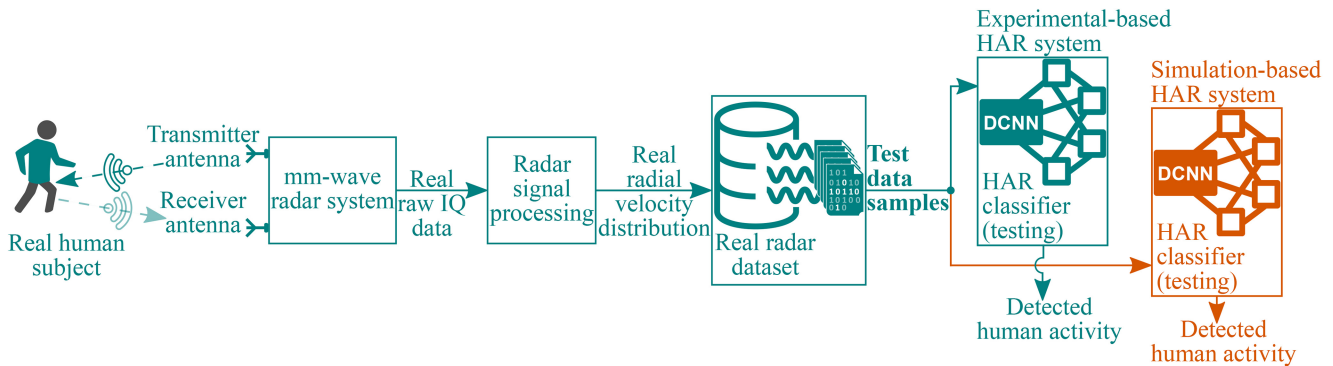

Fig. 2. Testing of the conventional (experimental-based) and the proposed simulation-based HAR systems on unseen real radar data samples.

training data without involving real human subjects and a real radar system, which makes the simulation-based approach very feasible and pragmatic.

The overall view of the proposed simulation-based approach is shown in Fig. 1(b). We start with six types of basic animations: 1) standing still; 2) falling; 3) walking in two steps; 4) standing up; 5) sitting down; and 6) picking up an object. Based on these six basic animations, we synthesize five different types of human activities: 1) falling on the floor; 2) walking forward with more than two steps; 3) standing up from a chair; 4) sitting down on a chair; and 5) picking up an object from the ground. The first 3-D simulations block in Fig. 1(b), implemented in the Unity software [13], synthesizes the human motion in the 3-D space and generates the corresponding TV 3-D trajectories of moving body segments, such as head, arms, legs, hands, and chest (see Section IV-C). Subsequently, the geometrical 3-D indoor channel model converts the TV 3-D trajectories into the TV radial distances with respect to the positions of the transmitter and receiver antennas of the radar system (see Section V). For the TV radial distances and a set of scatterers' weights (see Section VI), our radar data synthesizer in Fig. 1(b) simulates the raw IQ data in the slow- and fast-time domains. Note that the virtual markers in our simulation framework are analogous to the real scatterers on the human

body segments, which scatter the electromagnetic energy to the receive antenna of the radar system (see Section IV-C). Finally, the radar signal processing block of our simulation framework generates the simulated range distribution, the simulated radial velocity distribution, and the simulated mean velocity for a synthesized human activity.

In this study, we refrain from using simulated range distributions for HAR due to their limited intelligibility resulting from the radar systems' restricted range resolution. Additionally, we solely use the mean velocity for comparison, not for HAR, as it contains less information about scatterers' velocity compared to the radial velocity distribution. This is detailed in Section VII and evident from Figs. 6–8 as well. The simulated radial velocity distributions corresponding to the synthesized human activities are stored in a simulated radar data set. We have developed our simulation-based HAR system by training it using only the examples from our simulated radar data set as depicted in Fig. 1(b). The simulation-based HAR system was designed by using a DCNN approach. In order to demonstrate the practical significance of our simulation-based HAR system, we need to test its performance on unseen data collected by a real radar system and real human subjects. Therefore, we recorded real human activities (falling, walking, picking, standing and sitting) in front of a mm-wave radar system to

create a real radar data set, which is used to test our simulation-based HAR system, as shown in Fig. 2. It is noteworthy that the raw IQ data of the real and simulated radar are similar in structure. Therefore, we used the same radar signal processing block to process the real and simulated raw IQ data.

The proposed simulation framework emulates five distinct human activities and generates corresponding simulated radial velocity distributions for moving body segments. The simulated radar signatures (radial velocity distributions) are used to train the DCNN-based HAR classifier [see Fig. 1(b)]. Real mm-wave radar signatures are used to test the simulation-based HAR system as shown in Fig. 2. The details of the individual components of the proposed simulation-based HAR system are explained in the following sections.

IV. HUMAN MOTION SYNTHESIS

In this section, we elucidate the first component of our simulation framework, which is the synthesis of the human activities for our simulation-based HAR system.

A. Basic Humanoid Animations

To synthesize realistic human activities, we use a prerigged 3-D humanoid character and six types of basic humanoid animations from a well-known source called Mixamo [19]. It is a royalty-free library from Adobe Inc. offering countless realistic humanoid animations, which have been created with the help of professional actors and real-world MoCap systems [39]. We used the following animations from the Mixamo online library: idle, walking, falling, standing, picking, and sitting. In the idle animation, the avatar stands still in a natural upright posture, which causes a negligible in-place motion of all body segments. The walking animation consists of two steps in a forward direction on a flat floor. The falling animation portrays the avatar imitating a heart attack and collapsing abruptly to the ground. In the standing animation, the avatar gradually rises from a sitting position, while in the picking animation, it retrieves an object from the ground. In the sitting animation, the avatar is first in the idle upright position and then sits down on a chair.

We imported the six basic animations into the Unity software using the Filmbox (FBX) file format with a frame rate of 60 frames per second (fps). While importing an animation from Mixamo's online library, a keyframe reduction parameter must be configured to optimize the animation data. We have refrained from applying keyframe reduction to the animation data, as this could in some cases alter or degrade the animation itself. In fact, we used linear interpolation in the Unity software to upscale the frame rate of the animations from 60 fps to 2000 fps to emulate and match the radar's pulse repetition frequency (PRF).

B. Unity Animation System

Among other things, the Unity animation system estimates the spatial positions of the avatar's body segments between frames by performing an interpolation operation. While shape-preserving interpolation methods, such as spline interpolation, can offer more accurate representations of the motion data,

they often come with higher computational costs. In the context of our framework, where we aim to synthesize motion data at a high-frame rate of 2000 fps, computational efficiency is an important consideration. Linear interpolation provides a computationally efficient solution while still preserving the general shape and trajectory of the motion. Moreover, it is important to keep the animation frame rate f_r equal to the real radar's PRF, because the PRF samples the motion of an object and thereby dictates the maximum measurable radial velocity v_{max} according to $v_{max} = \text{PRF} \cdot \lambda/4$, where λ is the wavelength of the radar transmit signal. Analogously, in our simulation framework, the frame rate f_r dictates the maximum synthesizable radial velocity v'_{max} according to $v'_{max} = f_r\lambda/4$. Any motion of the avatar with a radial velocity component greater than the maximum synthesizable radial velocity v'_{max} reverts to a lower velocity, just as in a real radar system. The Unity animation system is, by and large, quite versatile in supporting a wide range of animation techniques, e.g., procedural, MoCap, and keyframe animations.

We use the Unity's animation state controller to enable the transition of the avatar between the six basic animation states. To synthesize a realistic human walking activity, we first need to switch between the basic idle and walking animations to merely start and end the overall walking activity. In addition, we need to gradually increase and decrease the walking speed during the transition periods of the emulated activity. These natural and smooth transitions with gradual acceleration and deceleration are provided by a special type of state in the Unity's animation state machine called the blend tree.

Note that we do not have any animation data for the transition periods. This gap is filled by blend trees. When emulating a human walking activity and transitioning between idle and walking animation, the blend tree state dynamically creates new animation data in the 3-D space in real time by aptly varying the avatar limbs to different degrees. With the help of blend trees, we can thus seamlessly transition from (to) idling animation to (from) walking animation with varying speeds while blending the two animations during the transition period. The human falling, standing, sitting, and picking activities are synthesized straightforwardly by combining the idle animation with the respective falling, standing, sitting, and picking animations.

C. 3-D Trajectories and Data Augmentation

We have synthesized five realistic human activities in the Unity software. In this section, we will explain how to capture the 3-D trajectories of the synthesized motion for the five types of human activities. First, we need to place several virtual markers on different body segments of the avatar, as shown in Fig. 3. These virtual markers are simulated point scatterers that resemble real scatterers on a human body.

In order to thoroughly capture the movements of the avatar, we placed a total of 21 virtual markers on different segments of the avatar body, which are represented by numbered stars in Fig. 3. The body segments associated with the virtual markers in Fig. 3 are listed in the ascending order: upper head, lower head, neck, right shoulder, left shoulder, right arm, left arm,
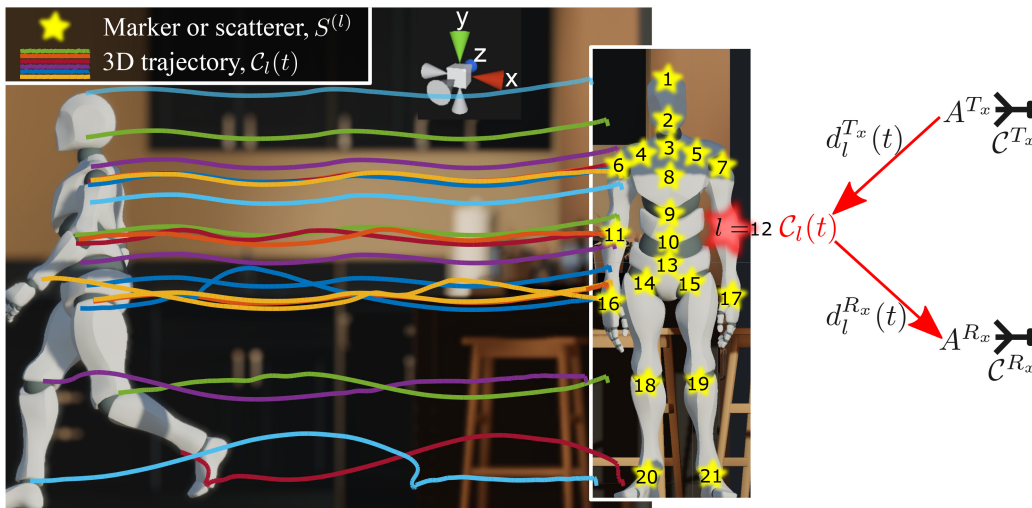
Fig. 3. Emulated propagation scenario composed of a radar system and a moving avatar with 21 nonstationary virtual markers.

upper spine, spine, lower spine, right forearm, left forearm, hips, right upper leg, left upper leg, right hand, left hand, right leg, left leg, right foot, and left foot. We need to spatially track the virtual markers and record the corresponding TV 3-D trajectories of the virtual markers for the synthesized human activities. For instance, for a walking activity consisting of three steps in a forward direction, we can visualize the progression of the TV 3-D trajectories associated with the 21 virtual markers, as represented by the colored curves in Fig. 3. With the ability to synthesize the human activities and the corresponding 3-D trajectories, we have created a data set of diverse human activities that is used to train the simulation-based HAR classifier.

The synthesized human activities can be augmented and diversified in the Unity software by varying the emulation parameters, such as the avatar's location, speed, acceleration, and deceleration. Thus, for each type of human activity, ten additional activity samples were generated by varying the above emulation parameters. For example, for the synthesized walking activities, random accelerations (decelerations) were assumed during the transition from the idling (walking) state to the walking (idling) state. For the five types of human activities, a total of fifty activity samples were generated in the Unity software. The TV 3-D trajectories were recorded for the synthesized human activity samples. Subsequently, the TV 3-D trajectories were exported to MATLAB for further data augmentation and processing. Using the geometrical 3-D indoor channel model, which is detailed in the following section, we simulated eight slightly different radar antenna positions by moving virtually the transmitter and receiver antennas laterally for data augmentation. We also scaled the weights of the scatterers [see Fig. 1(b) and Section VI-A] to vary the power levels of the simulated radar signatures for further data augmentation.

V. GEOMETRICAL 3-D INDOOR CHANNEL MODEL

In this section, we formulate a geometrical 3-D indoor channel model corresponding to an indoor propagation scenario

equipped with a radar system. Analogous to the real 3-D indoor propagation scenario, the emulated indoor environment is shown in Fig. 3. The emulated (real) propagation scenario is composed of a moving avatar (human) with $\mathcal{L}$ nonstationary virtual markers (scatterers), where the lth virtual marker (scatterer) is denoted by $S^{(l)}$ and $l = 1, 2, \ldots, \mathcal{L}$. In our simulation framework, the total number of virtual markers is $\mathcal{L} = 21$. The geometrical channel model is used to compute the TV radial distances between the $\mathcal{L}$ virtual markers (scatterers) and the radar transmit and receive antennas. In the simulation framework, the radar antennas can be placed freely as per the designer's requirements. Note that the antenna configuration greatly affects the simulated radar signatures. So, we can easily optimize the transmit and receive antenna positions with the help of the proposed simulation framework. For this research, the transmit and receive antennas of the radar system are placed in a monostatic configuration.

The radar transmit antenna A^{T_x} and receive antenna A^{R_x} are, respectively, placed at fixed positions $\mathcal{C}^{T_x} = [x^{T_x}, y^{T_x}, z^{T_x}]^\top$ and $\mathcal{C}^{R_x} = [x^{R_x}, y^{R_x}, z^{R_x}]^\top$, where $[\cdot]^\top$ represents the vector transpose operation. In Fig. 3, $\mathcal{C}_l(t) = [x_l(t), y_l(t), z_l(t)]^\top$ is the TV 3-D trajectory of the lth scatterer, $d_l^{T_x}(t)$ denotes the TV distance between the lth scatterer and the transmitter antenna A^{T_x}, and $d_l^{R_x}(t)$ denotes the TV distance between the lth scatterer and the receiver antenna A^{R_x}. Let $\|\cdot\|$ represents the Euclidean norm, then the TV distances $d_l^{T_x}(t)$ and $d_l^{R_x}(t)$ can be expressed as [40]

$$d_l^{T_x}(t) = \left\| \mathcal{C}_l(t) - \mathcal{C}^{T_x} \right\| \tag{1}$$

and

$$d_l^{R_x}(t) = \left\| \mathcal{C}_l(t) - \mathcal{C}^{R_x} \right\| \tag{2}$$

respectively. For the lth nonstationary virtual marker (scatterer), the TV radial distance $d_l(t)$ can be obtained as

$$d_l(t) = \frac{1}{2} \left[d_l^{T_x}(t) + d_l^{R_x}(t) \right]. \tag{3}$$

It is evident from (1)–(3) that the geometrical channel model maps the 3-D trajectory $\mathcal{C}_l(t)$ to the TV radial distance $d_l(t)$

for a particular antenna configuration $\{\mathcal{C}^{T_x}, \mathcal{C}^{R_x}\}$. For the monostatic configuration, we have $\mathcal{C}^{T_x} = \mathcal{C}^{R_x}$, and thus $d_l(t) = d_l^{T_x}(t) = d_l^{R_x}(t)$. In the context of radar sensing, the TV radial distance $d_l(t)$ characterizes the synthesized motion of the lth virtual marker (scatterer), which is used to simulate the radar raw IQ data in the fast- and slow-time domain as explained in the next section.

For all $\mathcal{L}$ virtual markers of the synthesized falling, walking, picking, standing and sitting activities, we used the geometrical channel model to simulate the TV radial distances $d_l(t)$, as shown in Fig. 4. As the $\mathcal{L}$ virtual markers are spatially distributed on the avatar's body segments, they have distinct TV radial distances, which are represented by the colored curves in Fig. 4. In the simulated falling activity of Fig. 4, some virtual markers exhibit more variations than others because the virtual markers on the lower body segments are less mobile than the virtual markers on the upper body segments. The simulated walking activity in Fig. 4 exhibits its periodic nature and it consists of three walking steps. Compared to the falling and walking activities, the radial distances of the virtual markers do not vary as much in the simulated standing, sitting, and picking activities.

VI. RADAR DATA SYNTHESIS

One of the main modules of our simulation framework is the radar data synthesis module, which simulates realistic raw IQ data by emulating an FMCW radar system. The simulated raw IQ data depends entirely on the TV radial distances $d_l(t)$ of all $\mathcal{L}$ virtual markers (scatterers) and their respective weights which are modeled in Section VI-A. In the following sections, we elucidate the synthesis of the radar baseband signal called beat signal and explore the relevant stationary and nonstationary aspects of the indoor wireless channel.

A. Beat Signal Synthesis

The FMCW radar system periodically emits RF pulses, where the intrapulse modulation is a linear chirp [41] waveform $c(t')$, where t' denotes the fast-time domain [42]. These RF pulses, also called the transmitted chirp signals $c(t')$, are reflected to the radar receiver by several scatterers in the environment. In this article, we only consider and model the nonstationary scatterers because the stationary scatterers do not cause any Doppler shift and can therefore easily be filtered out in the radar signal preprocessing unit [43]. Furthermore, we assume that the $\mathcal{L}$ scatterers are stationary in the fast time t' and nonstationary in the slow time t as explained in the following section. From the lth nonstationary scatterer, a copy of the transmitted chirp waveform $c(t')$ is received with the TV propagation delay $\tau^{(l)}(t)$, which is proportional to the TV range (radial distance) of the lth scatterer $d_l(t)$ according to $\tau^{(l)}(t) = 2d_l(t)/c_0$, where c_0 denotes the speed of light.

In FMCW radar systems, the quadrature mixture module downconverts the received passband signal and produces the complex baseband signal, also known as the composite beat signal $s_b(t', t)$ [42]. The raw IQ data from the FMCW radar system is the digitized version of the composite beat
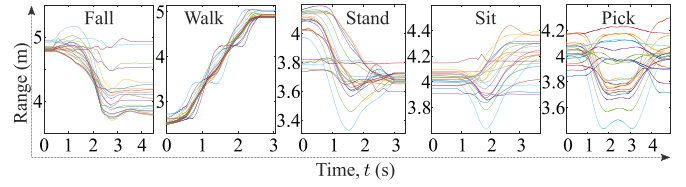

Fig. 4. Simulated TV radial distances $d_l(t)$ of 21 virtual markers for the five activities.

signal $s_b(t', t)$. In FMCW radar systems, the analog to digital converter (ADC) samples the composite beat signal $s_b(t', t)$ in fast time t' with the sampling interval T_s. For the radar's coherent processing interval (CPI), in which the phase of the scatterers is preserved [44], the discrete fast-time samples are arranged in the fast- and slow-time domain to form the raw IQ data matrix $\mathcal{D}$, i.e.,

$$
\mathcal{D} = \begin{bmatrix}
s_b(0, 0) & s_b(T_s, 0) & \cdots \\
s_b(0, T_{sw}) & s_b(T_s, T_{sw}) & \cdots \\
\vdots & \vdots & \vdots \\
s_b(0, (N_c-1)T_{sw}) & s_b(T_s, (N_c-1)T_{sw}) & \cdots \\
\end{bmatrix}
$$
$$
\begin{bmatrix}
s_b(T_{sw}-T_s, 0) \\
s_b(T_{sw}-T_s, T_{sw}) \\
\vdots \\
s_b(T_{sw}-T_s, (N_c-1)T_{sw})
\end{bmatrix}
\tag{4}
$$

where T_{sw} is the chirp duration and N_c is the number of chirps in the CPI of the radar system.

Now we want to model the composite beat signal $s_b(t', t)$, so that we can synthesize the raw IQ data of the FMCW radar system. Let $s_b^{(l)}(t', t)$ be the beat signal corresponding to the lth virtual marker, then the received composite beat signal $s_b(t', t)$ can be expressed as [42]

$$
s_b(t', t) = \sum_{l=1}^{\mathcal{L}} s_b^{(l)}(t', t).
\tag{5}
$$

Note that in (5), the composite beat signal $s_b(t', t)$ is composed of $\mathcal{L}$ distinct beat signals $s_b^{(l)}(t', t)$ corresponding to $\mathcal{L}$ virtual markers. In particular, for the lth virtual marker, the beat signal $s_b^{(l)}(t', t)$ is fully characterized by its TV path gain $a^{(l)}(t)$, beat frequency $f_b^{(l)}(t)$, phase $\phi^{(l)}(t)$, and propagation delay $\tau^{(l)}(t)$ according to

$$
s_b^{(l)}(t', t) = \sum_{n=0}^{\infty} a^{(l)}(t) \exp\left[j\left(2\pi f_b^{(l)}(t)t' + \phi^{(l)}(t)\right)\right]
$$
$$
\times \delta\left(t - \tau^{(l)}(t) - T_n\right)
\tag{6}
$$

where T_n is the discrete slow time that relates to the chirp duration T_{sw} by $T_n = nT_{sw}$ for $n = 0, 1, \ldots$. The function $\delta(\cdot)$ in (6) represents the Dirac delta function.

For the lth virtual marker, the syntheses of the TV beat frequency $f_b^{(l)}(t)$, phase $\phi^{(l)}(t)$, and propagation delay $\tau^{(l)}(t)$ in (6) are solely determined by the TV radial distance $d_l(t)$. Also, the lth TV path gain $a^{(l)}(t)$ is inversely proportional to the lth TV radial distance $d_l(t)$. The beat frequency $f_b^{(l)}(t)$

associated with the lth virtual marker can be modeled according to

$$f_b^{(l)}(t) = \frac{2d_l(t)\gamma}{c_0} \qquad (7)$$

where γ is the slope of the chirp waveform $c(t')$. The phase $\phi^{(l)}(t)$ of the lth virtual marker is related to the radial distance $d_l(t)$ according to

$$\phi^{(l)}(t) = \frac{4\pi d_l(t)}{\lambda}. \qquad (8)$$

Recall that the lth propagation delay component $\tau^{(l)}(t)$ in (6) can be obtained as $\tau^{(l)}(t) = 2d_l(t)/c_0$. Thus, the synthesis of the lth beat signal $s_b^{(l)}(t', t)$ in (6) is mainly determined by the lth TV radial distance $d_l(t)$.

We use the TV path gain $a^{(l)}(t)$ in (6) to model and simulate the amount of energy reflected back to the radar receiver from the lth scatterer (virtual marker). Thus, in the synthesis of the lth beat signal $s_b^{(l)}(t', t)$, the TV path gain $a^{(l)}(t)$ simulates the power or strength of the lth virtual marker. In this research, for the sake of simplicity, we have used $\mathcal{L} = 21$ time-invariant path gains, i.e., $a^{(l)}(t) = a^{(l)}$. For the five types of simulated human activities, we have accordingly used five sets of path gains to synthesize the composite beat signal $s_b(t', t)$ in (5). The body surface area [45] and the real TV radar signatures (see Section VII) helped us adjust the path gains of the $\mathcal{L}$ virtual markers corresponding to the five types of simulated human activities. Note that multiple sets of path gains can be used to simulate multiple radar signatures for a single simulated human activity.

B. Long- and Short-Time Stationarity of the Channel

In this section, we explain the long- and short-time stationarity of the indoor wireless channel. Since the transmitter and receiver antennas are spatially fixed, the nonstationarity of the wireless channel is due to the motion of a human subject. For fixed antennas, the wireless channel is non-stationary due to the motion of the scatterers. We assume that the channel is long-time nonstationary or, equivalently, nonstationary over the slow time t. But in the fast time t', we assume that the channel is stationary over the limited duration of a chirp waveform T_{sw}. This assumption simplifies the synthesis of the radar beat signal. In the following, we will see that the short-time stationarity of the channel basically comes down to the radar's range resolution denoted as d_{res}, which is related to the radar's bandwidth B according to $d_{res} = c_0/2B$.

In the radar signal processing module (see Section VII), we first apply the fast Fourier transform (FFT) to each row of the raw data matrix $\mathcal{D}$, called the range FFT. The frequency resolution f_{res} of the range FFT is equal to the inverse of the observation interval T_{sw}, i.e., $f_{res} = 1/T_{sw}$ [46]. For a row of the raw data matrix $\mathcal{D}$ and a slow-time instant t_0, the range FFT computes the spectrum containing the beat frequencies $f_b^{(l)}(t_0)$ corresponding to $d_l(t_0)$ for $l = 1, 2, \ldots, \mathcal{L}$. To resolve the spectral components corresponding to the $\mathcal{L}$ scatterers (virtual markers), the scatterers (virtual markers) must be at least f_{res}

apart in the spectrum or, equivalently, d_{res} apart in the range (see (7)).

Let Δd_l and $\Delta f_b^{(l)}$ denote the overall change in the lth radial distance $d_l(t)$ and beat frequency $f_b^{(l)}(t)$, respectively, over one chirp duration T_{sw}. Then, a small change in the lth radial distance Δd_l results in a small change in the lth beat frequency $\Delta f_b^{(l)}$ according to (7). In practice, these changes are insignificant over the chirp duration T_{sw} and are not discernible in the spectrum of (5), such that $\Delta d_l \ll d_{res}$ and $\Delta f_b^{(l)} \ll f_{res}$, especially for indoor channels. Thus, the lth beat frequency $f_b^{(l)}(t_0)$ is assumed to be constant at the slow-time instant t_0 and over the fast-time duration $t_0 < t' < t_0 + T_{sw}$. Therefore, the wireless channel is assumed to be short-time stationary, which makes the synthesis of the discrete beat signals $s_b^{(l)}(t', t_0 + T_n)$ fairly simple for $n = 0, 1, \ldots$ and $l = 1, 2, \ldots, \mathcal{L}$. For instance, for the lth radial distance $d_l(t_0)$, the real and imaginary components of the lth beat signal $s_b^{(l)}(t', t_0)$ in (6) can be digitally synthesized as simple tone signals with fixed frequency $f_b^{(l)}(t_0)$ and phase $\phi^{(l)}(t_0)$.

VII. RADAR SIGNAL PROCESSING

This section describes the radar signal processing module of Fig. 1, which can be used to process either the simulated or the real raw IQ data. First, the FFT operation is performed on the rows of the raw data matrix $\mathcal{D}$ (see (4)) to obtain the beat frequency function $S_b(f_b, t)$, which can be expressed as [47]

$$S_b(f_b, t) = \int_0^{T_{sw}} s_b(t', t) e^{-j2\pi f_b t'} dt' \qquad (9)$$

where f_b denotes the beat frequency. Subsequently, the short-time Fourier transform (STFT) of the beat frequency function $S_b(f_b, t)$ is carried out over the slow-time domain t to acquire the beat- and Doppler-frequency function $X(f_b, f, t)$ [42], which is given as

$$X(f_b, f, t) = \int_{-\infty}^{\infty} S_b(f_b, t'') W_r(t'' - t) e^{-j2\pi f t''} dt'' \qquad (10)$$

where f and t'' represent the Doppler frequency and running time, respectively. The function $W_r(\cdot)$ in (10) represents a rectangular window function spanning over the slow-time duration of $64 T_{sw}$.

Note that in (10), the beat- and Doppler-frequency function $X(f_b, f, t)$ can be integrated with respect to the Doppler frequency f (beat frequency f_b) to acquire the TV beat-frequency (micro-Doppler) signature. Thus, the expressions for the TV beat-frequency signature $S'(f_b, t)$ and the TV micro-Doppler signature $S(f, t)$ are given as

$$S'(f_b, t) = \left| \int_0^{PRF} X(f_b, f, t) df \right|^2 \qquad (11)$$

and

$$S(f, t) = \left| \int_0^{f_{b,max}} X(f_b, f, t) df_b \right|^2 \qquad (12)$$

respectively, where $f_{b,\max}$ is the maximum beat frequency. By using the TV beat-frequency signature $S'(f_b, t)$ in (11), we can express the TV range distribution $p'(r, t)$ as

$$p'(r, t) = \frac{S'\left(\frac{2\gamma}{c_0}r, t\right)}{\int_{-\infty}^{\infty} S'\left(\frac{2\gamma}{c_0}r, t\right)dr}. \tag{13}$$

The symbol r in (13) denotes the radar range, which is related to the beat frequency f_b by $r = c_0 f_b/(2\gamma)$.

Similarly, the TV radial velocity distribution $p(v, t)$ can be expressed as [42]

$$p(v, t) = \frac{S\left(\frac{2f_0}{c_0}v, t\right)}{\int_{-\infty}^{\infty} S\left(\frac{2f_0}{c_0}v, t\right)dv} \tag{14}$$

where v is the radial velocity and f_0 is the carrier frequency, which are related to the Doppler frequency f according to $v = c_0 f/(2f_0)$. Finally, the TV mean radial velocity $\bar{v}(t)$ can be obtained as

$$\bar{v}(t) = \int_{-\infty}^{\infty} v p(v, t)dv. \tag{15}$$

The TV mean radial velocity $\bar{v}(t)$ in (15) encapsulates the dominant characteristics of the TV radial velocity distribution $p(v, t)$ [42]. It provides a measure of the average radial velocity of all body segments at time t. Recall that the scatterers found on the human body segments reflect the electromagnetic energy back to the radar system. When a human body (avatar) moves, each scatterer (virtual marker) follows a spatially distinct trajectory and thus has a distinct TV radial velocity component with respect to the radar system. The TV radial velocity components corresponding to the scatterers (virtual markers) appear in the TV radial velocity distribution $p(v, t)$. Similarly, the TV range components of all the scatterers (virtual markers) appear in the TV range distribution $p'(r, t)$.

For the real radar data, the TV range distribution $p'(r, t)$ is not very intelligible due to the limited range resolution d_{res} of the radar system. Therefore, the real TV range distribution $p'(r, t)$ is usually not used for HAR. However, as an example, we show in Fig. 5 the simulated TV range distribution $p'(r, t)$ with a range resolution d_{res} of 75 mm for the simulated fall, walk, stand, sit and pick activities. We also show the simulated (real) TV radial velocity distributions $p(v, t)$ for these five types of simulated (real) activities in Fig. 6 (Fig. 7). We can clearly see the striking similarities between the simulated and the real TV radial velocity distributions $p(v, t)$ in Figs. 6 and 7, respectively. It is worth noting that we will use the images of the simulated (real) TV radial velocity distributions $p(v, t)$ to train (test) the proposed simulation-based HAR system (see Section VIII). Finally, for the five types of simulated and real activities, the TV mean radial velocities $\bar{v}(t)$ are shown in Fig. 8(a) and (b), respectively.

The similarities between the simulated and real radar results in Figs. 6, 7, and 8(a) and (b) demonstrate the quality of the data generated by the proposed simulation framework. Furthermore, to quantify the similarity between the
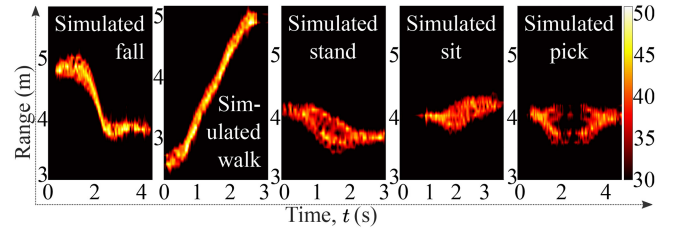

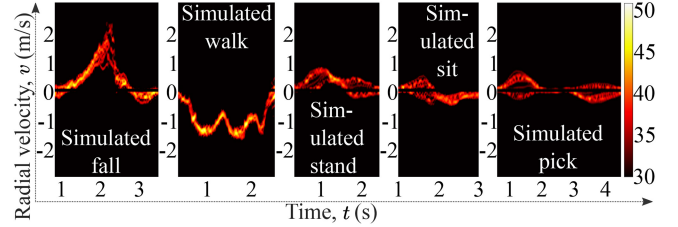Fig. 5. Simulated TV range distributions $p'(r, t)$ for the emulated human activities.

Fig. 6. Simulated TV radial velocity distributions $p(v, t)$ for the emulated human activities.
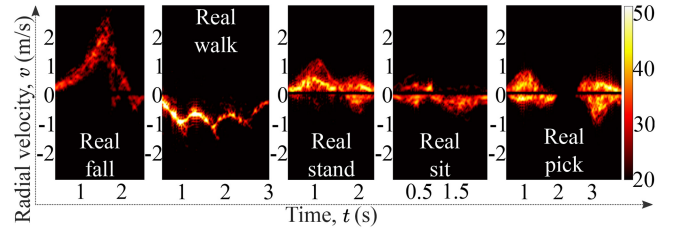

Fig. 7. Real TV radial velocity distributions $p(v, t)$ for the real human activities.
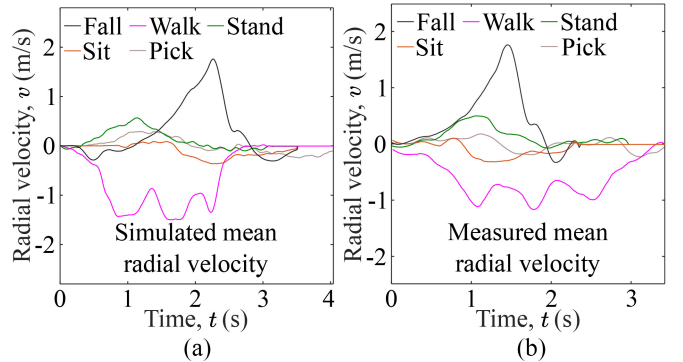

Fig. 8. TV mean radial velocities $\bar{v}(t)$ for (a) the emulated and (b) real human activities.

simulated and real radar signatures, we employ the DTW algorithm [20]. This DTW distance metric measures the resemblance between TV mean radial velocities $\bar{v}(t)$ of simulated and real human activities (see Fig. 8). The normalized DTW distances, presented in Table I, indicate the efficacy of our simulation-based approach in capturing the kinematic characteristics of various human activities. Notably, for all the activities, the DTW distance metric is minimized when comparing a given simulated activity to its corresponding real counterpart. For instance, for the walking activity, the DTW distance of 0.07 between the simulated and real TV

TABLE I
DTW DISTANCES BETWEEN THE SIMULATED AND REAL TV
MEAN RADIAL VELOCITIES $\bar{v}(t)$

Activity type		Real				
		fall	walk	stand	sit	pick
Simulated	fall	**0.03**	0.36	0.18	0.26	0.20
	walk	0.49	**0.07**	0.55	0.36	0.34
	stand	0.07	0.26	**0.01**	0.10	0.05
	sit	0.13	0.08	0.16	**0.05**	0.07
	pick	0.08	0.19	0.04	0.03	**0.01**

mean radial velocities $\bar{v}(t)$ highlights the precise simulation of the walking pattern. This trend persists across all simulated activities, affirming the fidelity of our simulation framework in accurately simulating real-world radar signatures. Note that the DTW distance between certain activities, such as sitting and picking, is smaller. This is due to their closely aligned patterns, thereby making them relatively challenging to classify.

VIII. SIMULATION-BASED HAR SYSTEM: REALIZATION AND TESTING

In this section, we first explain how we realized the proposed HAR system using a DCNN-based multiclass classifier and how we trained it using only the simulated HAR data set. A range of variations of the DCNN classifier is systematically analyzed through model ablations, facilitating the process of model analysis and selection. Subsequently, we demonstrate the performance of our trained simulation-based HAR classifier on the unseen real radar data using the best DCNN model.

A. Supervised Learning Using Simulated HAR Data Set

First and foremost, we need a simulated HAR data set for training purposes. To create the simulated HAR data set, we first synthesized the human motion using the Unity software as described in Section IV. The position, speed, acceleration, and deceleration parameters were randomly varied in the Unity software to synthesize ten unique activity samples for each of the five activity types: 1) falling; 2) walking; 3) standing; 4) sitting; and 5) picking. Subsequently, the spatial trajectories of these fifty activity samples were imported into MATLAB for further data augmentation. For each activity sample, eight slightly different radar positions $\{\mathcal{C}^{T_x}, \mathcal{C}^{R_x}\}$ and three different power levels were simulated in MATLAB. Low, medium, and high-power levels were simulated by scaling the $\mathcal{L}$ weights $a^{(l)}(t)$ of the scatterers. In conclusion, the simulated HAR data set consists of five types of human activities, ten different emulations of each activity type, eight radar positions, and three power levels. Thus, the total number of simulated TV radial velocity distributions $p(v, t)$ was 1200 in our simulated HAR data set, which was used to train the DCNN-based HAR classifier.

The simulated TV radial velocity distributions $p(v, t)$ [see Fig. 6] were transformed into images of dimension 224 × 224 × 3. Thus, for each image, the number of pixels in the horizontal and vertical dimensions are 224 and the number of color channels are 3 (red, green and blue). These 1200

images serve as input feature maps to the DCNN-based HAR classifier as shown in Fig. 9. The four convolutional layers of the DCNN classifier in Fig. 9 contain 32, 64, 128, and 256 filter channels, respectively, which extract features from the simulated TV radial velocity distribution $p(v, t)$. Each filter in a convolutional layer is a 2-D trainable kernel with the dimension k_d equal to 6 × 6 pixels. Note that, for the DCNN classifier shown in Fig. 9, the network complexity (depth of hidden layers), kernel dimension k_d, max-pool layers, learning rate l_r, and other hyperparameters were determined through systematic analysis of a range of model variations. Further details are provided in the subsequent subsections.

To train the weights of the kernels, the L2 regularization technique was adopted to overcome the potential issue of overfitting [48]. The stride parameter was set to 1 in the DCNN, so that the feature-extraction filters stepped by one pixel. We employed the rectified linear unit (ReLU) function to alleviate the problem of vanishing gradients [49]. In Fig. 9, the convolutional layers are followed by the max-pool layers of the order 2 × 2. The purpose of the max-pool layers is to reduce the redundancies by downsampling the output of the convolutional layers by a factor of 2. The features extracted by the multiple layers of the convolutional filters are flattened prior to the multilayer perceptron (MLP) layers (see Fig. 9).

In the DCNN, the feature vector of dimension 50176×1 is obtained from the input TV radial velocity distribution $p(v, t)$. Then, the feature vector undergoes three MLP layers of dimensions 256, 128 and 32 with a dropout rate of 30%, as shown in Fig. 9. The dropout layers mitigate the problems related to overfitting and generalizability of the network [50]. Finally, the Softmax layer of order 5×1 was employed to compute the probabilities corresponding to the five types of human activities. For the training and validation of our simulation-based HAR classifier, we used our simulation data set with a training–validation split ratio of 80:20. To optimize the weights and biases of our simulation-based HAR system in Fig. 9, we adopted the adaptive moment estimation (Adam) optimizer [51] and the human activity samples from our simulated data set. The decay factors β_1, β_2, and the parameter ϵ of the Adam optimizer were set to 0.9, 0.999, and 10^{-8}, respectively, and the batch size was set to 32.

B. Real Data Collection and Model Variations

To test our proposed simulation-based HAR system, we used real human activities recorded by Ancortek's mm-wave FMCW radar system. During the measurement campaign, the operating parameters, such as the carrier frequency f_c, bandwidth BW, chirp duration T_{sw}, and PRF of the mm-wave radar system, were set to 24.125 GHz, 250 MHz, 500 μs, and 2 kHz, respectively. Note that the same values were chosen in the radar simulation model. The antennas of the real and simulated radar systems were chosen to be placed in a monostatic configuration.

We conducted in-depth experiments with Ancortek's mm-wave radar system in an indoor propagation scenario to compose the real radar-based HAR data set consisting of five
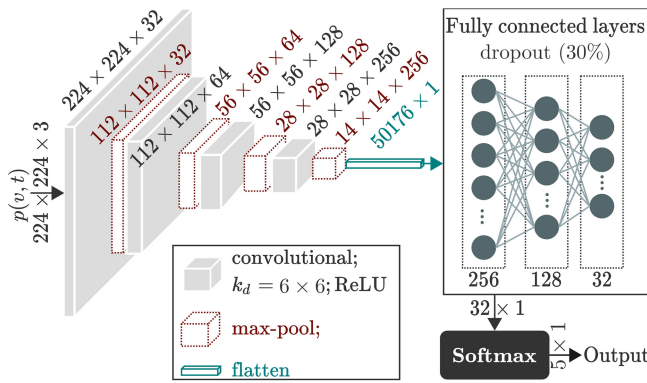
Fig. 9. Design of our DCNN-based HAR classifier that uses the simulated (real) HAR data set for its training (testing).

TABLE II
MEAN CLASSIFICATION ACCURACIES OF THE DCNN MODELS

Model	CNN layers	MLP layers	Trainable parameters	Mean accuracy $\pm$ SD (%)
0	[16, 32]	[32, 16]	3,232,117	80.5 ± 7.5
1	[16, 32, 48]	[48, 16]	1,882,805	81.9 ± 5.7
2	[16, 16, 32, 32, 48, 48]	[48, 32, 16]	658,485	77.4 ± 16.5
3	[32, 48, 64, 80]	[128, 64, 32]	2,371,557	85.3 ± 4.7
4	[32, 64, 72, 80]	[256, 128, 32]	4,502,205	87 ± 3.3
5	[32, 64, 96, 128]	[256, 128, 32]	7,201,029	86.2 ± 4.7
6	[32, 64, 128, 256]	[256, 128, 32]	14,434,725	86.7 ± 4.8
7	[48, 128, 256, 512]	[256, 128, 32]	31,853,109	47.3 ± 31.1
8	[48, 128, 512, 512]	[256, 128]	37,747,957	66.4 ± 30.8
9	[48, 128, 256, 256, 512, 512]	[256, 128, 64]	43,654,645	19.7 ± 5.5

(a)

(b)

Fig. 10. Model 6 performance analysis: the (a) mean accuracy curves and (b) mean accuracy heatmap for kernel dimensions k_d and learning rates l_r.

types of human activities, namely, falling, walking, picking, sitting, and picking. Five male adults and one female adult repeatedly performed the human activities in the presence of various indoor objects. The mm-wave radar's IQ data corresponding to the real human activities were processed by the radar signal processing module to generate the TV radial velocity distributions $p(v, t)$. Note that the human activities were recorded for more than 5 s, but the actual duration of the activities was mostly 3 s (see Fig. 7). The total number of radar signatures in the real radar data set is 306.
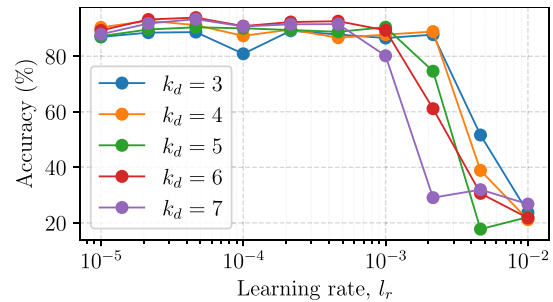
For our simulation-based HAR system, we systematically explored various DCNN network configurations, detailed in Table II. Utilizing simulated and real radar signatures, we, respectively, trained and tested the DCNN classifiers with varying depths and complexities of the convolutional neural network (CNN) and MLP layers. Models 4, 5, and 6 demonstrated mean accuracies exceeding 86% with standard deviations (SDs) of less than 5% (see Table II). It is important to highlight that other DCNN models with lower and higher complexities displayed suboptimal performance, as indicated by the mean test accuracies in Table II. Subsequently, we systematically determined optimal hyperparameters, including kernel dimension k_d and learning rate l_r, for Models 4–6. Among these, Model 6 emerged as the most promising classifier, achieving the average (maximum) accuracy of 94% (98.4%) with optimized hyperparameters. The average percentage accuracies of Model 6 across different kernel dimensions k_d and learning rates l_r are depicted through the curves in Fig. 10(a) and the heatmap in Fig. 10(b).

C. Testing of the Simulation-Based HAR System Employing Model 6
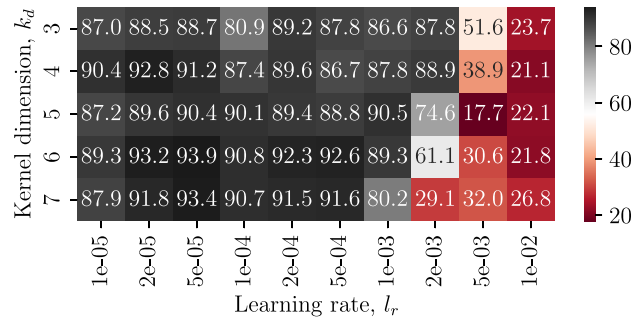
The train–test (or simulation–real) data split ratio was 80:20. From the real radar-based HAR data set, the 306 TV radial velocity distributions $p(v, t)$ corresponding to the real human subjects were used to test our trained simulation-based HAR system. The confusion matrix presented in Fig. 11 shows the performance of our simulation-based HAR system (see Fig. 9), specifically focusing on the trained model with the maximum performance. The x- and y-axes of the confusion matrix correspond to the predicted and true class of a human activity, respectively. Thus, the first five diagonal elements of the confusion matrix represent the number of correct classifications. The number of misclassifications is represented by the off-diagonal elements in the first five rows and columns of the confusion matrix. For example, a "walking" activity was misclassified as a "falling" activity, as shown in the second row of the first column. Fig. 11 also shows that four "sitting" activities were misclassified as "picking" activities as indicated by the fourth row of the fifth column. In the confusion matrix, the precision and recall quantities [52] are shown by the green color in the last row and last column, respectively. The worst precision and recall values are 95.9% and 92.3%, respectively. Most importantly, the overall classification accuracy of our simulation-based HAR system is 98.4% as shown by the white entry in Fig. 11.

Note that the classification accuracy of our simulation-based HAR system is similar to today's real or experimental-based HAR systems [21], [53], [54]. However, the proposed simulation-based approach is quite unique in that it effortlessly

TABLE III
STATE-OF-THE-ART RADAR-BASED HAR APPROACHES AND THEIR PERFORMANCE

HAR method	System description	Training dataset	Dataset simulation method	Accuracy
JDS-TL [8]	Semisupervised TL and domain adaptation using partially (10%) labeled radar data for HAR	Experimental	None	87.6%
Hybrid CNN-RNN [55]	Distributed radar network with spatial-temporal pattern extraction for HAR	Experimental	None	90.8%
Multi-view CNN-LSTM [56]	CAE-based unsupervised feature learning and multi-view data fusion via LSTM	Experimental	None	92%
FS-ADA [32]	Learning common feature space of pre-existing and newly acquired training datasets	Experimental	None	91.6%
MBGAN [57]	Physics-aware GAN-based systems that utilize limited real data to synthesize micro-Doppler signatures	Simulated and experimental	GAN-based data synthesis using limited real data	89.2%
WRGAN-GP [58]	Generation and refinement models for realistic spectrogram synthesis	Simulated and experimental	GAN-based data synthesis using limited real data	94.9%
Vid2Doppler [59]	Cross-domain translation: Creating synthetic Doppler signatures from videos of human activities	Simulated	Translation of video dataset to Doppler dataset	81.4%
Proposed simulation-based framework	Simulating RF signals and radar signatures by using simulated 3D trajectories of virtual point scatterers	Simulated	Translating MoCap data to radar data using sensor modeling	Mean: 94% Max: 98.4%

Fig. 11. Confusion matrix of the simulation-based HAR classifier with a classification accuracy of 98.4% on real data.

generates a large amount of high-quality simulation data for training purposes. In the context of radar-based HAR, it is difficult to claim the superiority of one method or system over another as these systems are designed to address different constraints and resolve distinct problems. Nevertheless, in Table III, we have reported the performance of various state-of-the-art HAR systems using classification accuracy as the base metric.

The joint domain and semantic transfer learning (JDS-TL) [8] approach employed semi-supervised transfer learning (TL) and domain adaptation on partially labeled radar data to achieve an accuracy of 87.6%, as shown in Table III. Utilizing a hybrid architecture of CNNs and recurrent neural networks (RNNs) for spatial-temporal pattern extraction, the hybrid CNN-RNN [55] approach achieved a classification accuracy of 90.8% in recognizing human activities. Through a combination of convolutional auto encoder (CAE)-based unsupervised feature learning and multiview data fusion, the CNN-LSTM method in [56] achieved an accuracy of 92%. The few-shot adversarial domain adaptation (FS-ADA) [32] method learned a common feature space from existing and new data sets, yielding a 91.6% accuracy in radar-based HAR despite limited training data. The aforementioned state-of-the-art HAR systems relied on experimental-based training data sets, as outlined in Table III. Now, let us turn our attention to HAR systems trained with either partially (GAN-based) simulated data sets or fully simulated data sets.

To tackle kinematic inconsistencies associated with GAN-based data synthesis, the multibranch generative adversarial network (MBGAN) system in [57] employed physics-aware GAN-based techniques to synthesize micro-Doppler signatures, achieving 89.2% classification accuracy. For data set augmentation, Qu et al. [58] employed a Wasserstein refined generative adversarial network with gradient penalty (WRGAN-GP) to generate synthetic micro-Doppler spectrograms. Vid2Doppler [59] employed cross-domain translation to generate synthetic Doppler signatures from videos, achieving an accuracy of 81.4% through entirely simulated training data. In contrast, our proposed simulation-based framework translated MoCap data into radar data via channel modeling, achieving a mean (maximum) accuracy of 94% (98.4%) using entirely simulated training data.

In this section, we have explained the design of the proposed simulation-based HAR system. It is worth noting that the proposed simulation framework of Fig. 1(b) can be easily extended to other mm-wave radar-based application areas, such as gesture classification. The only difference would be to animate different types of gestures in the Unity software, while the rest of the modules of Fig. 1(b) would remain the same.

IX. CONCLUSION

The development of the modern radar-based HAR systems is mostly hindered by the scarce, unbalanced and partial data sets, because the acquisition of real radar data is not an easy task, especially for real human subjects. Therefore, in this article, we alleviated the problems related to data scarcity for radar-based HAR classifiers. As a proof-of-concept, we presented an end-to-end simulation framework that synthesizes human motion and simulates the realistic mm-wave FMCW radar signatures. By generating large amounts of high-quality synthetic data, the proposed simulation framework significantly decreases the overall training effort of radar-based HAR systems. We used the synthetic and real data to train and test the HAR system, respectively. The proposed simulation-based HAR system demonstrated a classification accuracy of 98.4% on the unseen real radar data. Since the proposed end-to-end simulation framework reduces the involvement of real human

subjects, it is crucial to improve the capabilities of future radar-based HAR classifiers.

In addition, the proposed simulation framework provides control over numerous radar and target parameters, such as avatar speed, acceleration, deceleration, height, position, motion type, radar antenna configuration, frequency, PRF, and bandwidth. This allows us to generate different types of radar data sets corresponding to different radar-operating conditions and different applications. Additionally, the proposed framework enables us to augment the data at the motion synthesis layer. Thus, at the base motion synthesis layer, the target motion characteristics can be randomized to generate impartial, unbiased or balanced data sets that can be used to train radar-based classifiers.

In the proposed simulation framework, the scatterer-level modeling of the radar signal opens up new avenues of research for the radar-based classifiers. For instance, different optimization techniques can be explored to further improve the quality of the simulated radar signal and ultimately the simulated radar signatures. Furthermore, the work presented in this article can be extended to classify other types of everyday human activities. The proposed approach can also be used to actualize other mm-wave radar-based classifiers, such as gesture recognition. We anticipate that the proposed end-to-end simulation framework will empower future radar-based classifiers with enhanced capabilities. We plan to extend the proposed simulation framework to multiple-input multiple-output radar systems incorporating multidirectional HAR.

REFERENCES

[1] W. Qi, N. Wang, H. Su, and A. Aliverti, "DCNN based human activity recognition framework with depth vision guiding," *Neurocomputing*, vol. 486, pp. 261–271, May 2022.

[2] J. Roche, V. De-Silva, J. Hook, M. Moencks, and A. Kondoz, "A multimodal data processing system for LiDAR-based human activity recognition," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 10027–10040, Oct. 2022.

[3] Y. Tang, L. Zhang, F. Min, and J. He, "Multiscale deep feature learning for human activity recognition using wearable sensors," *IEEE Trans. Ind. Electron.*, vol. 70, no. 2, pp. 2106–2116, Feb. 2023.

[4] M. A. A. Al-Qaness, A. Dahou, M. A. Elaziz, and A. M. Helmi, "Multi-ResAtt: Multilevel residual network with attention for human activity recognition using wearable sensors," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 144–152, Jan. 2023.

[5] Y. Tang, L. Zhang, Q. Teng, F. Min, and A. Song, "Triple cross-domain attention on human activity recognition using wearable sensors," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 5, pp. 1167–1176, Oct. 2022.

[6] M. Muaaz, A. Chelli, M. W. Gerdes, and M. Pätzold, "Wi-sense: A passive human activity recognition system using Wi-Fi and convolutional neural network and its integration in health information systems," *Ann. Telecommun.*, vol. 77, nos. 3–4, pp. 163–175, 2022.

[7] Z. Chen, C. Cai, T. Zheng, J. Luo, J. Xiong, and X. Wang, "RF-based human activity recognition using signal adapted convolutional neural network," *IEEE Trans. Mobile Comput.*, vol. 22, no. 1, pp. 487–499, Jan. 2023.

[8] X. Li, Y. He, F. Fioranelli, and X. Jing, "Semisupervised human activity recognition with radar micro-doppler signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5103112.

[9] L. Zhang, W. Cui, B. Li, Z. Chen, M. Wu, and T. S. Gee, "Privacy-preserving cross-environment human activity recognition," *IEEE Trans. Cybern.*, vol. 53, no. 3, pp. 1765–1775, Mar. 2023.

[10] M. Muaaz, A. Chelli, and M. Pätzold, "WiHAR: From Wi-Fi channel state information to unobtrusive human activity recognition," in *Proc. IEEE Veh. Technol. Conf.*, May 2020, pp. 1–7.

[11] S. K. Yadav, K. Tiwari, H. M. Pandey, and S. A. Akbar, "A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions," *Knowl.-Based Syst.*, vol. 223, Jul. 2021, Art. no. 106970.

[12] S. Waqar, M. Muaaz, and M. Pätzold, "Direction-independent human activity recognition using a distributed MIMO radar system and deep learning," *IEEE Sensors J.*, vol. 23, no. 20, pp. 24916–24929, Oct. 2023.

[13] "Unity real-time development platform | 3D, 2D VR & AR engine." Unity. Accessed: Feb. 20, 2023. [Online]. Available: https://unity.com/

[14] S. Ahmed, W. Kim, J. Park, and S. H. Cho, "Radar-based air-writing gesture recognition using a novel multistream CNN approach," *IEEE Internet Things J.*, vol. 9, no. 23, pp. 23869–23880, Dec. 2022.

[15] T. Tuncer, F. Ertam, S. Dogan, and A. Subasi, "An automated daily sports activities and gender recognition method based on novel multikernel local diamond pattern using sensor signals," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9441–9448, Dec. 2020.

[16] M. Tammvee and G. Anbarjafari, "Human activity recognition-based path planning for autonomous vehicles," *Signal, Image Video Process.*, vol. 15, no. 4, pp. 809–816, 2021.

[17] F. J. R. Lera, F. M. Rico, A. M. G. Higueras, and V. M. Olivera, "A context-awareness model for activity recognition in robot-assisted scenarios," *Expert Syst.*, vol. 37, no. 2, 2020, Art. no. e12481.

[18] K. Kim, A. Jalal, and M. Mahmood, "Vision-based human activity recognition system using depth silhouettes: A smart home system for monitoring the residents," *J. Electr. Eng. Technol.*, vol. 14, no. 6, pp. 2567–2573, 2019.

[19] "Get animated." Mixamo. Accessed: Feb. 20, 2023. [Online]. Available: https://www.mixamo.com/#/

[20] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. AAAIWS*, 1994, pp. 359–370.

[21] W.-Y. Kim and D.-H. Seo, "Radar-based human activity recognition combining range–time–Doppler maps and range-distributed-convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022.

[22] G. Lee and J. Kim, "Improving human activity recognition for sparse radar point clouds: A graph neural network model with pre-trained 3D human-joint coordinates," *Appl. Sci.*, vol. 12, no. 4, p. 2168, 2022.

[23] A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli, "Continuous human activity classification from FMCW radar with bi-LSTM networks," *IEEE Sensors J.*, vol. 20, no. 22, pp. 13607–13619, Nov. 2020.

[24] M. M. Rahman, E. A. Malaia, A. C. Gurbuz, D. J. Griffin, C. Crawford, and S. Z. Gurbuz, "Effect of kinematics and fluency in adversarial synthetic data generation for ASL recognition with RF sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 4, pp. 2732–2745, Aug. 2022.

[25] Z. Xia and F. Xu, "Time-space dimension reduction of millimeter-wave radar point-clouds for smart-home hand-gesture recognition," *IEEE Sensors J.*, vol. 22, no. 5, pp. 4425–4437, Mar. 2022.

[26] H. Liu et al., "M-Gesture: Person-independent real-time in-air gesture recognition using commodity millimeter wave radar," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 3397–3415, Mar. 2022.

[27] Z. Yu et al., "A radar-based human activity recognition using a novel 3-D point cloud classifier," *IEEE Sensors J.*, vol. 22, no. 19, pp. 18218–18227, Oct. 2022.

[28] Y. Zhao, A. Yarovoy, and F. Fioranelli, "Angle-insensitive human motion and posture recognition based on 4D imaging radar and deep learning classifiers," *IEEE Sensors J.*, vol. 22, no. 12, pp. 12173–12182, Jun. 2022.

[29] Z. Liu, C. Wu, and W. Ye, "Category-extensible human activity recognition based on doppler radar by few-shot learning," *IEEE Sensors J.*, vol. 22, no. 22, pp. 21952–21960, Nov. 2022.

[30] Y. Yang, Y. Zhang, Y. Lang, B. Li, S. Guo, and Q. Tan, "GAN-based radar spectrogram augmentation via diversity injection strategy," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.

[31] Y. Zhou, C. Xie, S. Sun, X. Zhang, and Y. Wang, "A self-supervised human activity recognition approach via body sensor networks in smart city," *IEEE Sensors J.*, early access, Jun. 8, 2023, doi: 10.1109/JSEN.2023.3282601.

[32] X. Li, Y. He, J. A. Zhang, and X. Jing, "Supervised domain adaptation for few-shot radar-based human activity recognition," *IEEE Sensors J.*, vol. 21, no. 22, pp. 25880–25890, Nov. 2021.

[33] C. Yu, Z. Xu, K. Yan, Y.-R. Chien, S.-H. Fang, and H.-C. Wu, "Noninvasive human activity recognition using millimeter-wave radar," *IEEE Syst. J.*, vol. 16, no. 2, pp. 3036–3047, Jun. 2022.

[34] Y. Yang, Y. Zhang, H. Ji, B. Li, and C. Song, "Radar-based human activity recognition under the limited measurement data support using domain translation," *IEEE Signal Process. Lett.*, vol. 29, pp. 1993–1997, 2022.

[35] M. J. Bocus et al., "OPERAnet, a multimodal activity recognition dataset acquired from radio frequency and vision-based sensors," *Sci. Data*, vol. 9, no. 1, p. 474, 2022.

[36] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler, "Micro-doppler effect in radar: Phenomenon, model, and simulation study," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 2–21, Jan. 2006.

[37] S. Vishwakarma, W. Li, C. Tang, K. Woodbridge, R. Adve, and K. Chetty, "SimHumalator: An open-source end-to-end radar simulator for human activity recognition," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 37, no. 3, pp. 6–22, Mar. 2022.

[38] G. Manfredi, P. Russo, A. De Leo, and G. Cerri, "Efficient simulation tool to characterize the radar cross section of a pedestrian in near field," *Progr. Electromagn. Res. C*, vol. 100, pp. 145–159, Mar. 2020.

[39] Mixamo, *Mixamo Motion Capture (Behind the Scenes)*, (2012). [Online Video]. Available: https://www.youtube.com/watch?v=q7jWMd7ix98

[40] S. Waqar and M. Pätzold, "Interchannel interference and mitigation in distributed MIMO RF sensing," *Sensors*, vol. 21, no. 22, p. 7496, 2021.

[41] S. Sana, S. Waqar, H. Yusaf, M. Waqas, and F. A. Siddiqui, "Software defined digital beam forming processor," in *Proc. 13th Int. Bhurban Conf. Appl. Sci. Technol.*, 2016, pp. 671–676.

[42] S. Waqar, M. Muaaz, and M. Pätzold, "Human activity signatures captured under different directions using SISO and MIMO radar systems," *Appl. Sci.*, vol. 12, no. 4, p. 1825, 2022.

[43] S. Waqar, H. Yusaf, S. Sana, M. Waqas, and F. A. Siddiqui, "Reconfigurable monopulse radar tracking processor," in *Proc. 15th Int. Bhurban Conf. Appl. Sci. Technol.*, 2018, pp. 805–809.

[44] Z. Peng et al., "A portable FMCW interferometry radar with programmable low-IF architecture for localization, ISAR imaging, and vital sign tracking," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 4, pp. 1334–1344, Apr. 2017.

[45] M. Mance, M. Prutki, A. Dujmovic, M. Milošević, V. Vrbanovic-Mijatovic, and D. Mijatovic, "Changes in total body surface area and the distribution of skin surfaces in relation to body mass index," *Burns*, vol. 46, no. 4, pp. 868–875, 2020.

[46] J. W. Ting, D. Oloumi, and K. Rambabu, "FMCW SAR system for near-distance imaging applications—Practical considerations and calibrations," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 1, pp. 450–461, Jan. 2018.

[47] M. Muaaz, S. Waqar, and M. Pätzold, "Radar-based passive step counter and its comparison with a wrist-worn physical activity tracker," in *Proc. 4th Int. Conf. Intell. Technol. Appl. INTAP*, 2022, pp. 259–272.

[48] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. 7th Int. Conf. Learn. Represent. ICLR*, 2019, pp. 1–19.

[49] H. Ide and T. Kurita, "Improvement of learning for CNN with ReLU activation by sparse regularization," in *Proc. Int. Joint Conf. Neural Netw.*, 2017, pp. 2684–2691.

[50] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.

[51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[52] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 233–240.

[53] Y. Zhao, H. Zhou, S. Lu, Y. Liu, X. An, and Q. Liu, "Human activity recognition based on non-contact radar data and improved PCA method," *Appl. Sci.*, vol. 12, no. 14, p. 7124, 2022.

[54] M. Muaaz, S. Waqar, and M. Pätzold, "Orientation-independent human activity recognition using complementary radio frequency sensing," *Sensors*, vol. 23, no. 13, p. 5810, 2023.

[55] S. Zhu, R. G. Guendel, A. Yarovoy, and F. Fioranelli, "Continuous human activity recognition with distributed radar sensor networks and CNN–RNN architectures," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art no. 5115215.

[56] H.-U.-R. Khalid, A. Gorji, A. Bourdoux, S. Pollin, and H. Sahli, "Multi-view CNN-LSTM architecture for radar-based human activity recognition," *IEEE Access*, vol. 10, pp. 24509–24519, 2022.

[57] M. M. Rahman, S. Z. Gurbuz, and M. G. Amin, "Physics-aware generative adversarial networks for radar-based human activity recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 3, pp. 2994–3008, Jun. 2023.

[58] L. Qu, Y. Wang, T. Yang, and Y. Sun, "Human activity recognition based on WRGAN-GP-synthesized micro-doppler spectrograms," *IEEE Sensors J.*, vol. 22, no. 9, pp. 8960–8973, May 2022.

[59] K. Ahuja, Y. Jiang, M. Goel, and C. Harrison, "Vid2Doppler: Synthesizing doppler radar data from videos for training privacy-preserving activity recognition," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2021, pp. 1–10.

Sahil Waqar received the bachelor's degree (with Distinction) in electrical engineering from the University of Engineering & Technology, Taxila, Pakistan in 2010, and the master's degree in electrical engineering with a Gold Medal Award from the National University of Science and Technology, Islamabad, Pakistan, in 2013.

His master's research was related to probabilistic data association filters. From 2012 to 2020, he worked on the design and development of digital communication systems, signal processors, radar systems, and embedded systems. Since October 2020, he has been working on the CareWell Project as a Ph.D. Research Fellow with the University of Agder, Grimstad, Norway. His current research interests include machine/deep learning, sensors, MIMO radars, communications, signal processing, and embedded systems.

Matthias Pätzold received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from Ruhr University Bochum, Bochum, Germany, in 1985 and 1989, respectively, and the Habilitation degree in communications engineering from Hamburg University of Technology, Hamburg, Germany, in 1998.

From 1990 to 1992, he was with the ANT Nachrichtentechnik GmbH, Backnang, Germany, where he was involved in digital satellite communications. From 1992 to 2001, he was with the Department of Digital Networks, Hamburg University of Technology, Hamburg, Germany. Since 2001, he has been a Full Professor of Mobile Communications with the University of Agder, Grimstad, Norway. He has authored several books and numerous technical articles.

Dr. Pätzold's publications received 14 best paper awards. He has been actively participating in numerous conferences as a TPC member and a TPC chair.