

Review

# A Survey of Deep Learning-Based Human Activity Recognition in Radar

Xinyu Li <sup>†</sup>, Yuan He <sup>\*,†</sup> and Xiaojun Jing

School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China; lixinyu@bupt.edu.cn (X.L.); jxiaojun@bupt.edu.cn (X.J.)

\* Correspondence: yuanhe@bupt.edu.cn

† These authors contributed equally to this work.

Received: 3 April 2019; Accepted: 30 April 2019; Published: 6 May 2019



**Abstract:** Radar, as one of the sensors for human activity recognition (HAR), has unique characteristics such as privacy protection and contactless sensing. Radar-based HAR has been applied in many fields such as human–computer interaction, smart surveillance and health assessment. Conventional machine learning approaches rely on heuristic hand-crafted feature extraction, and their generalization capability is limited. Additionally, extracting features manually is time-consuming and inefficient. Deep learning acts as a hierarchical approach to learn high-level features automatically and has achieved superior performance for HAR. This paper surveys deep learning based HAR in radar from three aspects: deep learning techniques, radar systems, and deep learning for radar-based HAR. Especially, we elaborate deep learning approaches designed for activity recognition in radar according to the dimension of radar returns (i.e., 1D, 2D and 3D echoes). Due to the difference of echo forms, corresponding deep learning approaches are different to fully exploit motion information. Experimental results have demonstrated the feasibility of applying deep learning for radar-based HAR in 1D, 2D and 3D echoes. Finally, we address some current research considerations and future opportunities.

**Keywords:** human activity recognition; radar; deep learning; human backscattering echoes

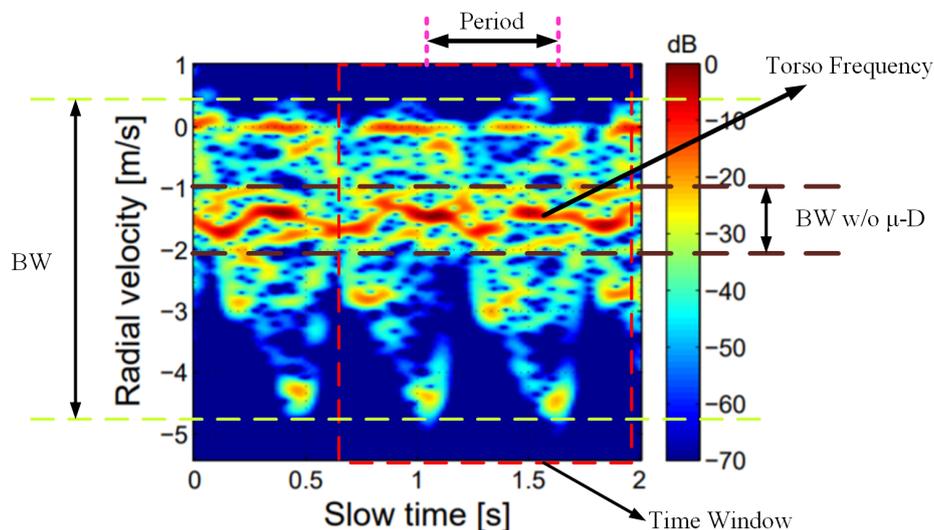
## 1. Introduction

Research on human activity recognition (HAR) has made significant progress over the past decade. Successful HAR applications include surveillance [1], smart home [2], video analytics [3], autopilot [4] and human–computer interaction [5]. The purpose of HAR is to identify a user’s behavior so as to allow computing systems to proactively provide assistance for the user [6]. There are two main categories of HAR [7]: vision-based and sensor-based. Taking advantage of the high resolution of optical sensors and the rapidly evolving computer vision (CV) techniques, vision-based HAR has yielded fruitful results [8–12]. Despite the superiority of vision-based HAR, there are still many open issues, such as illumination, occlusion, privacy leakage, etc. [13–15]. With the rapid development of sensor technology, sensor-based methods have set off a new wave in HAR [16–19]. Sensor-based activity recognition acquires data from accelerometer, gyroscope, radar, acoustic sensor and so on, and seeks the profound high-level information of human behaviors from multitudes of low-level sensor readings.

As one of the sensor-based methods, radar-based HAR [20–23] has drawn much attention due to the following reasons. Firstly, radar is robust to light and weather conditions so it is able to be applied in harsh environments. Secondly, radar could protect visual privacy. Instead of capturing the visual shape of the target, the returned signals modulated by the target carry abundant time-varying range and velocity information of activities [24]. Thirdly, radar is able to detect human through walls,

which makes radar-based HAR applicable to more scenarios. Lastly, radar systems do not need any tag attached to the human body, which makes it more user-friendly. Consequently, radar has been adopted more and more for recognizing human activities recently.

In the past, radar-based HAR systems often adopt conventional machine learning (ML) techniques [25–29]. These traditional algorithms are built on theoretical foundations so they are explicable and could be optimized theoretically. Compared with deep learning models, their complexity is often lower so the computation burden is lighter. Support vector machine (SVM) [30], dynamic time warping (DTW) [31], random forest classifier [27] are the most used conventional ML algorithms for radar-based HAR. Y. Kim et al. [30] used SVM to recognize human activities based on micro-Doppler signatures. Features were extracted manually from the time–Doppler spectrograms, as illustrated in Figure 1. By employing a decision-tree structure, the system that consists of six SVMs succeeded to classify 12 different human activities. In [27], a gesture recognition system using a 60 GHz mm-wave radar was built. A random forest classifier was employed in this system to perform real-time gesture recognition. In [31], an improved DTW algorithm was proposed for hand gesture recognition with a terahertz radar. Experiments showed that the improved DTW algorithm was capable of fully exploring the properties of range profiles and Doppler signatures.



**Figure 1.** Illustration of features extracted from a time–Doppler map. Adopted from [30].

Although widely used for radar-based activity recognition, traditional ML solutions have several drawbacks that hinder the further improvement of robustness and generalization. Firstly, features are extracted heuristically and manually, which highly relies on human experience and domain knowledge. Secondly, hand-crafted features often refer to some low-level statistical information including mean, variance, frequency and amplitude [32], which are task-specific. When a model trained with shallow hand-crafted features is applied to a new dataset, the performance is always not as good as in the original dataset. Thirdly, traditional ML methods mainly learn on small-scale static data. However, in the real world, activity data are coming in a stream and changeable. Conventional ML approaches are not competent to train a robust model in this circumstance. Deep learning (DL), a rapidly evolving technology, tends to break through these restrictions. As a new branch of ML, DL came into sight and has emerged as a powerful tool in the past few years. DL approaches extract high-level deep features automatically through hierarchical architectures. Artificial feature extraction using specialized knowledge is not acquired. Furthermore, with the advent of GPU, it is capable of fast computing with huge amounts of data. DL algorithms can take full advantage of parallel computing to achieve fast processing. With its excellent feature learning ability from massive data, DL has not only promoted

the development of visual object recognition, speech processing, natural language processing, etc. [33], but also made HAR more intelligent and versatile.

To the best of our knowledge, currently there is no survey that addresses DL-based HAR progress in radar, and this is the first article to present the recent advance of this area. We hope it provides a comprehensive summary and motivates more inspirations for relevant future research.

The remainder of the paper is organized as follows: In Section 2, we compactly overview DL techniques. Section 3 reviews examples of radar systems adopted to recognize human activities. Section 4 presents DL approaches for radar-based HAR in detail. We divide existing literature into three parts according to the dimension of radar echoes, and then DL techniques applied to the three parts are discussed respectively. Future research considerations and directions are discussed in Section 5. Finally, the paper is concluded in Section 6.

## 2. Deep Learning Techniques

With the emergence of deep learning algorithms, the development of many fields such as speech recognition, visual object recognition and even drug discovery has been accelerated [34]. A deep learning model has multiple processing layers to learn high-level representations automatically. Heavy feature engineering and domain knowledge are not required for DL. Furthermore, with so many deep transformations, very complex functions could be learned and difficult classification and recognition problems could be solved [35,36]. As a result, deep learning has advanced the development of many fields, including HAR. In this section, we investigate several deep learning models and analyze their unique advantages for HAR tasks. Table 1 describes all the models in brief.

**Table 1.** DL models and advantages for human activity recognition.

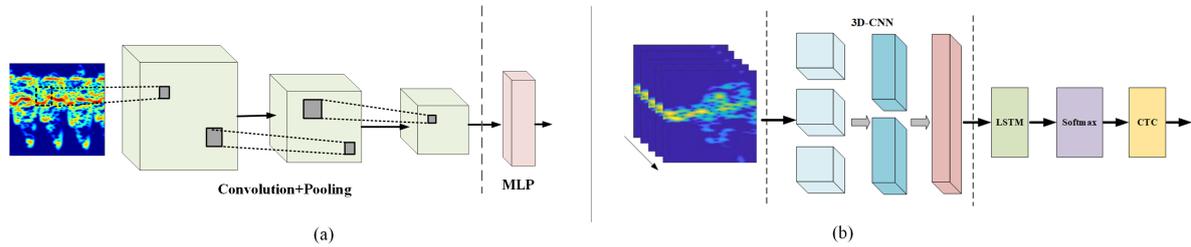
Models	Descriptions and Advantages
CNN	capturing the spatial relationship by multiple convolutional layers, often utilized as an excellent localized feature extractor
RNN	exploring the temporal relationship in data, variants are often utilized, such as LSTM
Auto-encoder	a feed-forward neural network that learns deep features in an unsupervised fashion
Hybrid deep models	the combination of some deep models, built on each model's own strength to obtain better performance

### 2.1. Convolutional Neural Network

The convolutional neural network is inspired by the visual cortex structure which is composed of simple cells and complex cells. It adopts four key ideas: local connections, parameter sharing, pooling and multi-layers. In CNN, convolution operation replaces the general matrix multiplication in general neural network. In this way, the complexity of the network is reduced due to the decreased number of weights. It should be noted that CNN is the first DL architecture with the hierarchical layers [37]. Multiple convolutional layers enable CNN to extract higher-level spatial features from the lower-level ones hierarchically, avoiding the manual feature extraction procedure in the conventional ML algorithms. After convolutional layers, pooling and fully connected layers are usually employed for classification or regression tasks. Additionally, thanks to the excellent deep feature learning ability, CNN is often employed as an automatic feature extractor for a variety of tasks [38,39].

When utilized for HAR, CNN has two main advantages [32]: taking nearby signals into consideration and scale-invariant for different paces or frequencies. The first advantage allows CNN to extract localized features from the positions which are space-related, rather than from a single position. The second advantage allows pace or frequency information to be retained in the extracted features. In [23,40–49], CNNs with different architectures and convolution kernels of various sizes

were employed as classifiers to recognize human activities with time–Doppler maps, as illustrated in Figure 2a. In [28,50–52], CNN acts as a spatial feature extractor and extracts high-level representations of human activities for further identification, as illustrated in Figure 2b.

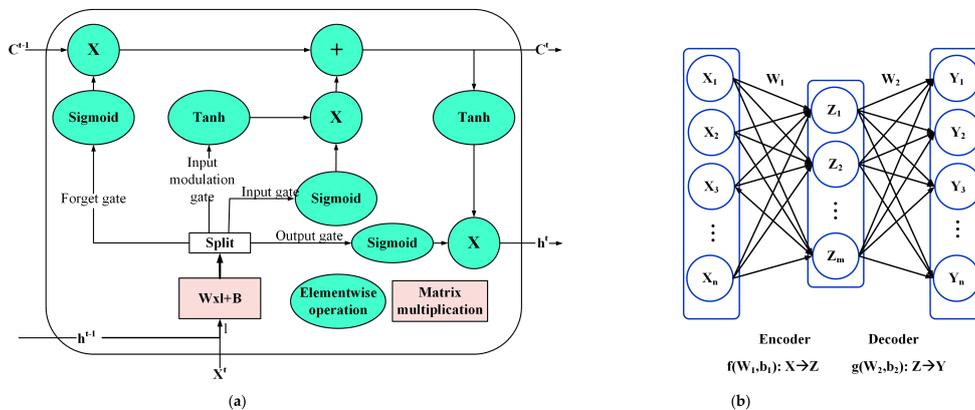


**Figure 2.** (a) CNN is performed as a classifier for radar-based HAR. Adopted from [41]. (b) CNN is adopted as a feature extractor to learn high-level features from the input time–range maps. Adopted from [28].

2.2. Recurrent Neural Network

With the successful application in NLP and speech recognition, recurrent neural network (RNN) has caught researchers’ attention in HAR. RNN has shone light on modeling temporal sequences because of the ability of mining temporal and semantic information. From the perspective of network structure, RNN remembers the previous information and uses it to influence the output of the following nodes. However, conventional RNN has its own limit: long-term dependencies. To overcome this shortcoming, long short term memory (LSTM) (see Figure 3a) came into being [53] and performs better in many tasks. LSTM owns three special gates: input gate, output gate and forget gate. By using these memory units especially the forget gate, LSTM is able to access a long-range context of sequential data. Compared with CNN, which can only process data with a predefined size, the prediction of RNN and its variants is assumed to increase in accuracy when more data is available. The prediction result is changing with time. Consequently, RNN is more sensitive to the change of the input data than CNN.

For HAR, RNN and its variants have the superiority of exploiting temporal correlations in an activity, which is a crucial issue for recognizing human activities. Ref. [28,50,51,54] all utilized RNN and its variants to model the temporal characteristics in human activities. In [50], the features that extracted from range–Doppler maps by a CNN are time-correlated, so LSTM was utilized for learning complex dependencies across time in those features. In this way, both spatial and temporal information was explored under the cooperation of CNN and LSTM.



**Figure 3.** A schematic overview of (a) LSTM cell and (b) auto-encoder.

### 2.3. Auto-Encoder

Auto-encoder (AE) (see Figure 3b) is a feed-forward neural network that aims to reconstruct the input under certain constraints [52]. It learns deep feature representations of unlabeled input via several rounds of encoding–decoding procedures in an unsupervised fashion. Especially when the input data are highly similar, AE is able to discover nuances in the data itself by the layer-wise unsupervised pre-training principle. Furthermore, unsupervised pre-training tends to function as a regularizer, which potentially prevents the network from overfitting [55].

The commonly used variants of AE in HAR are the following kinds: (a) stack auto-encoder (SAE) that stacks multiple sparse AEs together to acquire more compact feature coding. (b) convolutional auto-encoder (CAE) that essentially combines CNN and AE, and the encoding–decoding procedures are accomplished by convolution and deconvolution. (c) de-noising AE and contractive AE that make the models more generic by adding noise to the input or adding a well-chosen penalty term to the loss function.

In [52], CAE was employed for unsupervised pre-training, as illustrated in Figure 4. Then the decoding part of the network was removed, and fully connected layers, as well as a softmax classifier, were added to the encoder for classification. Taking advantages of unsupervised pre-training and localized feature learning, CAE outperformed CNN and plain AE for identifying human activities. In [56], stacked AEs were applied to obtain the most prominent features in radar echoes, and then softmax classifiers were utilized for recognizing human motions. A similar approach was also adopted in [57–59].

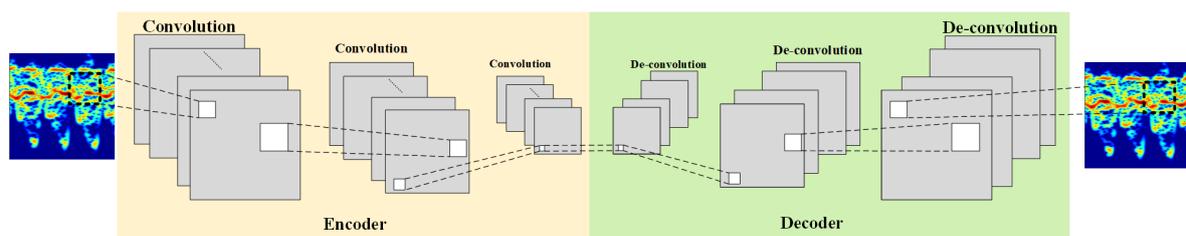


Figure 4. CAE architecture adopted from [52].

### 2.4. Hybrid Deep Model

Every model has its own disadvantages and is not competent to all tasks. Hybrid deep models integrate several networks together and take advantage of all these networks. Such cooperation is built on each model's own strength so as to obtain better performance. So far, in HAR, CNN and RNN are commonly combined as shown in Figure 2b, because they are good at abstracting different domain features: CNN captures spatial relationships while RNN captures temporal relationships [32]. Ref. [28,50,51] provided good examples for how to combine CNN and RNN. Those work has demonstrated that combining CNN and RNN tends to reinforce the power of recognizing the activities that vary in time and space. In addition, AE is often combined with CNN or RNN owing to its ability of unsupervised extracting high-dimensional features [52].

## 3. Radar System for Human Activity Recognition

Radar is an active sensing system that transmits radio wave and receives returned signals modulated by illuminated objects. It has been mostly used in remote sensing systems such as satellite remote sensing, air and terrestrial traffic control and geophysical monitoring in the past few decades [60]. Moreover, there has been a recent expansion of short-range radar for HAR tasks.

The radar-based HAR approaches are more robust than vision-based ones because of radar's insensitivity to light and weather condition. They can detect human presence and activities directly without any tag attached to the human body. When a person is moving, the speeds/Doppler frequencies of body parts are time-varying with respect to the person's movement. Subsequently,

the ranges of these parts are not linear with respect to time. The targets' range, speed and angle information that radar obtains could be utilized to recognize human activities [61].

Due to its intrinsic advantages such as simple architecture, easy system integration, relatively low cost, and penetration capability, radar is feasible as a kind of human motion measurement technology [62]. There are a series of radars used for HAR, such as continuous-wave radar, ultra-wide band radar and noise radar. While there are many advanced researches of noise radar HAR systems [63–68], they do not involve ML techniques, and thus are out of scope of this paper. Next, we introduce several kinds of radar designed for HAR purposes. Table 2 briefly outlines those radars and their basic characteristics.

**Table 2.** Radar system and basic characteristics.

<b>CW radar</b>	Doppler radar	sending out single-tone radio waves able to acquire the Doppler/radial velocity information of targets
	FMCW radar	providing range and speed information of targets simultaneously suitable for scenarios with the presence of multiple targets
	Interferometry radar	obtaining angular velocity of the target regardless of the targets' moving direction, with the output of two antennas cross-correlated
<b>UWB radar</b>		providing fine range resolution able to distinguish the major scattering centers of the target

### 3.1. Continuous-Wave (CW) Radar

CW radar transmits a known stable-frequency CW radio signal and receives the reflected signal that is modulated by objects on the radio signal path [60]. It is able to operate on either modulated mode or unmodulated mode. CW radar has a simple architecture with easy system integration and low power consumption, which makes CW radar attractive for mobile and portable applications. Various commercial CW radar chips and systems are available for HAR applications [61], such as 77 GHz AWR1642 and AWR1443 of Texas Instruments (Dallas, TX, USA), 77 GHz TEF8181EN and TEF8102EN of NXP (Eindhoven, The Netherlands) and 24 GHz BGT24MTR11 of Infineon (Neubiberg, Germany). Typical CW radar systems for HAR are Doppler radar, frequency-modulated CW Radar and interferometry radar.

#### A. Doppler Radar

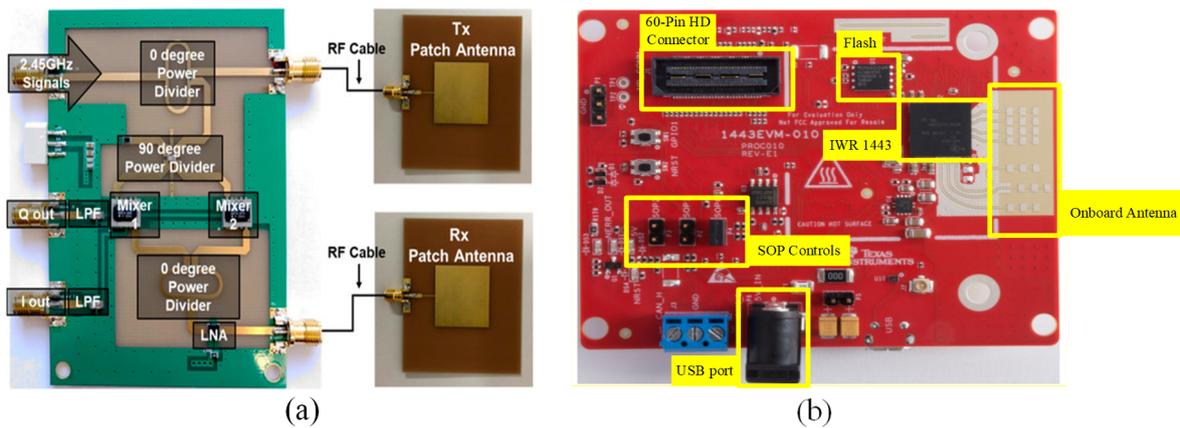
Doppler radar, as shown in Figure 5a, is one of the most popular radars in HAR [20,41–43,45,52,69]. A Doppler radar sends out single-tone radio waves and no modulation is involved. When the target is moving, the frequency of the received signals is shifted away from the transmitted ones because of the Doppler effect. The frequency  $f_r$  of backscattered signals is shown as follows [62],

$$f_r = f_t(1 + v/c)/(1 - v/c), \quad (1)$$

where  $f_t$  is the frequency of single tone radio signals sent by the Doppler radar,  $c$  is the speed of light,  $v$  is the radial speed of the target. The Doppler frequency shift  $f_d$  is thus

$$f_d = f_r - f_t = 2vf_t/(c - v) \quad (2)$$

Doppler radar is used to detect time-varying radial speeds of human motion due to its ability of capturing Doppler shifts. Owing to the relatively simple signal processing, Doppler radar is capable of acquiring appreciable performance in motion and displacement measurement [70,71].

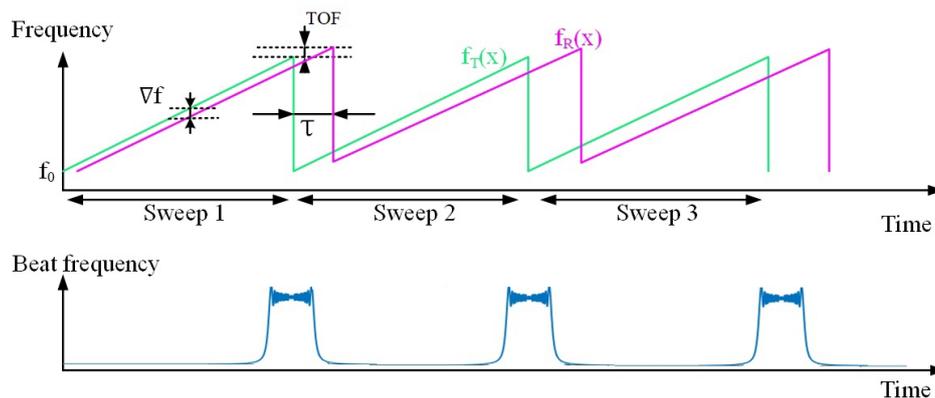


**Figure 5.** (a) Doppler radar adopted from [72]. (b) IWR1443, which is a 77 GHz FMCW radar produced by Texas Instruments.

### B. Frequency-Modulated Continuous-Wave Radar

Frequency-modulated continuous-wave (FMCW) radar, as shown in Figure 5b, is able to sense the range and Doppler properties of targets simultaneously. When more than one source of reflection reaches at radar antennas at the same time, both range and Doppler information are indispensable. Consequently, FMCW radar is widely employed in various short-range scenarios [28,50,51,57,73,74], especially scenarios with the presence of multiple targets [75,76].

In an FMCW radar system, a known stable frequency continuous wave that varies up and down in frequency over a fixed period of time is transmitted, such as a sine wave and sawtooth wave [77]. As illustrated in Figure 6, the backscattered echoes are mixed with transmitted signals to produce beat signals. By demodulating the beat signals and calculating the frequency delay of the received signals, range information could be extracted [73]. The range resolution of an FMCW radar that refers to the minimum separation in range of two objects of an equal cross section, is proportional to  $c/2B$ , where  $B$  is modulation bandwidth. So, the larger the bandwidth is, the higher the range resolution is. As for the Doppler information, it is obtained in the same way of unmodulated CW radar.



**Figure 6.** Illustration of how an FMCW radar acquires range and Doppler information, taking sawtooth wave as an example.  $f_d$  is Doppler shift while  $\tau$  is time delay. Adopted from [73].

### C. Interferometry Radar

One disadvantage of Doppler radar is that the frequency shift highly depends on radial velocity. As a result, it is hard to recognize noncooperative activities performed along the tangent direction. In this circumstance, interferometry radar is more helpful. It utilizes an interferometric receiver

composed of two antennas, and the output of the two antennas are cross-correlated [78]. When a target moves under the interferometric mode of radar, a signal whose frequency is proportional to the angular velocity of the target is produced. As a consequence, interferometry radar produces micro-Doppler signatures regardless of the moving direction of a person.

Interferometry radar has been applied in many fields, such as engineering metrology, remote sensing and small-displacement measurement [79]. In HAR, interferometry radar is also adopted thanks to its ability of acquiring tangential motion information [80,81]. In addition, the interferometry mode is often combined with FMCW mode for indoor precise positioning, versatile life activity monitoring and vital sign tracking. In [82], time–Doppler maps of a walking person acquired from an interferometric radar and a Doppler radar were compared. It is seen that the two time–Doppler maps look similar and both contain micro-Doppler features, which means that it is possible to apply the classification algorithms of Doppler radar to interferometric radar in a straight-forward manner.

### 3.2. Ultra-Wide Band Radar

Ultra-wideband (UWB) radar, whose fractional bandwidth of the transmitted signals is greater than 25%, is another type of radar that is often utilized for human detection and activity recognition [22,44,46,83–86]. Fractional bandwidth of UWB radar is defined as

$$FBW = \frac{2(f_H - f_L)}{f_H + f_L}, \quad (3)$$

where  $f_H$  refers to the upper bound of frequency and  $f_L$  refers to the lower bound frequency. UWB radar is to transmit pulses with very short durations in nanosecond range or even less. Due to the wideband, UWB radar has the capacity of anti-interference, penetrability, fine range resolution and short range detection. Thus, UWB radar is able to distinguish the major scattering centers of the target and identify short-range human activities [87,88]. Despite the contradiction between range resolution and Doppler resolution, UWB radar is able to acquire the Doppler information of each scattering center of the human body when compromising the range resolution and the Doppler resolution. Additionally, it has low power consumption, which makes it more applicable for portable HAR devices with limited computational capabilities. The transmitted pulse signal of UWB radar has a certain bandwidth, and theoretically has the ability for multi-target activity recognition. However, there is no related work at present, which is mainly due to the high complexity of the algorithms for separating targets and identifying individual activities.

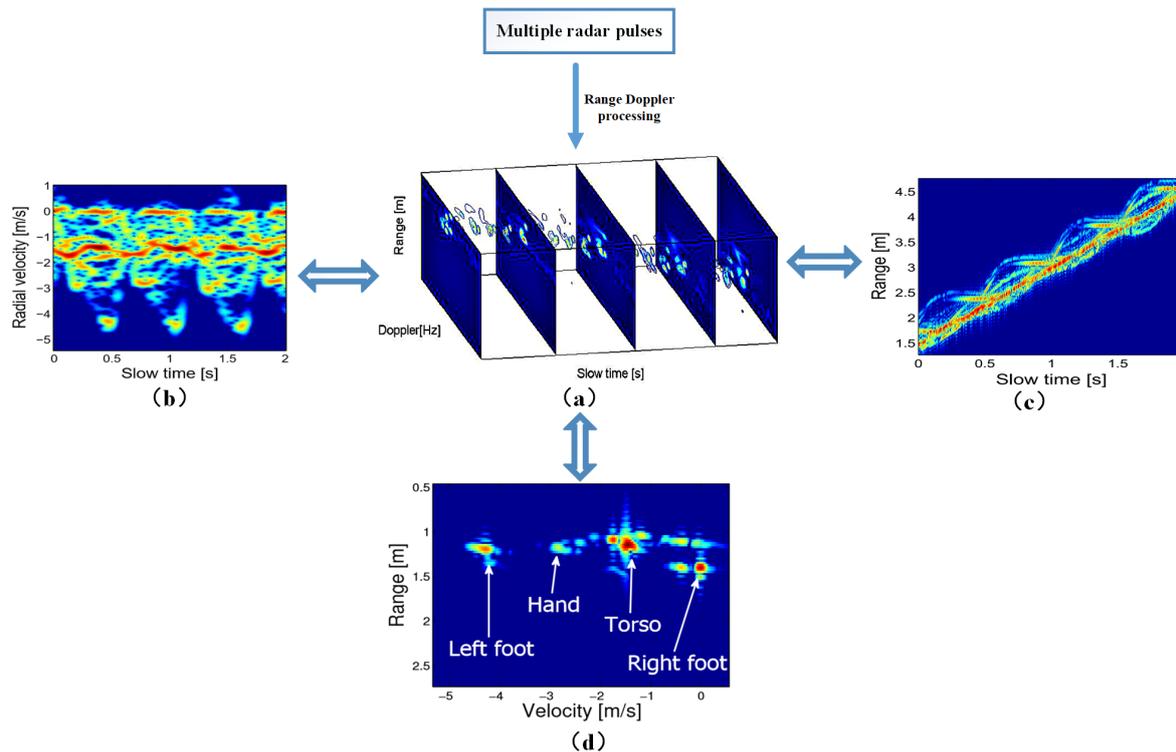
## 4. Deep Learning Approaches for Human Activity Recognition in Radar

In Section 2, we have discussed several common DL models and their advantages for HAR. As for radar, since radar echoes contain time, range and Doppler information, it is desirable that the DL algorithms are designed specifically for radar echoes. Motivated by this, in this section, we describe deep learning approaches for human activity recognition in radar according to the dimension of radar returns. Table 3 lists all the surveyed work in this section.

**Table 3.** Summation of existing works on DL based human activity recognition in radar.

	Echo Form	Literature	Radar Type	Central Frequency	Deep Model	
3D echoes	time–range–Doppler maps	[50,51]	FMCW radar	60 GHz	CNN + LSTM	
		[89]	CW radar	4 GHz	CNN	
		[45]	CW radar	8 GHz	CNN	
		[48]	CW radar	24 GHz	CNN	
		[54]	CW radar	8 GHz	LSTM	
		[56]	CW radar	6 GHz	SAE	
		[52]	CW radar	4 GHz	CAE	
		[41]	Doppler radar	2.4 GHz	CNN	
		[90]	Doppler radar	7.3GHz	CNN	
		time–Doppler maps	[47]	Doppler radar	24 GHz	CNN
			[49]	Doppler radar	5.8 GHz	CNN
			[69]	Doppler radar	25 GHz	LSTM
			[20]	Doppler radar	24 GHz	SAE
			[42]	pulse Doppler radar	5.8 GHz	CNN
[43]	pulse Doppler radar		5.8 GHz	CNN		
2D echoes		[84]	UWB radar	4 GHz	CNN	
		[85]	UWB radar	4.3 GHz	CNN	
		[83]	UWB radar	7.3GHz	CNN	
		[46]	UWB radar	4 GHz	CNN	
		[44]	UWB radar	4 GHz	CNN	
		[91]	FMCW radar	24 GHz	CNN	
		[21]	FMCW radar	5.8 GHz	CNN	
		[22]	UWB radar	3.9GHz	CNN	
		[28]	FMCW radar	24 GHz	3D CNN + LSTM	
		time–range maps	[74]	FMCW radar	24 GHz	3D CNN
		range–Doppler maps	[58]	FMCW radar	25 GHz	SAE
		time–Doppler maps and time–range maps	[57]	FMCW radar	24 GHz	SAE
		time–Doppler maps, time–range maps and range–Doppler maps				

Radar signals are transformed into 3D time–range–Doppler data cube by range–Doppler (RD) processing [92], which uses Doppler effect to determine the radial component of target’s velocity. In this way, multiple components of a target are resolved not only in range but also in Doppler. The 3D RD ‘video’ describes the slow-time evolution of the target’s activity, as shown in Figure 7a. Radar signals can also be represented in 2D, namely time–Doppler map (Figure 7b), time–range map (Figure 7c) and range–Doppler map (Figure 7d). In order to make full use of the information in echoes, deep learning methods should be designed more carefully for different forms of echoes.



**Figure 7.** 1D, 2D and 3D radar echoes: (a) 3D time–range–Doppler data cube, (b) 2D time–Doppler map, (c) 2D time–range map, (d) 2D range–Doppler map.

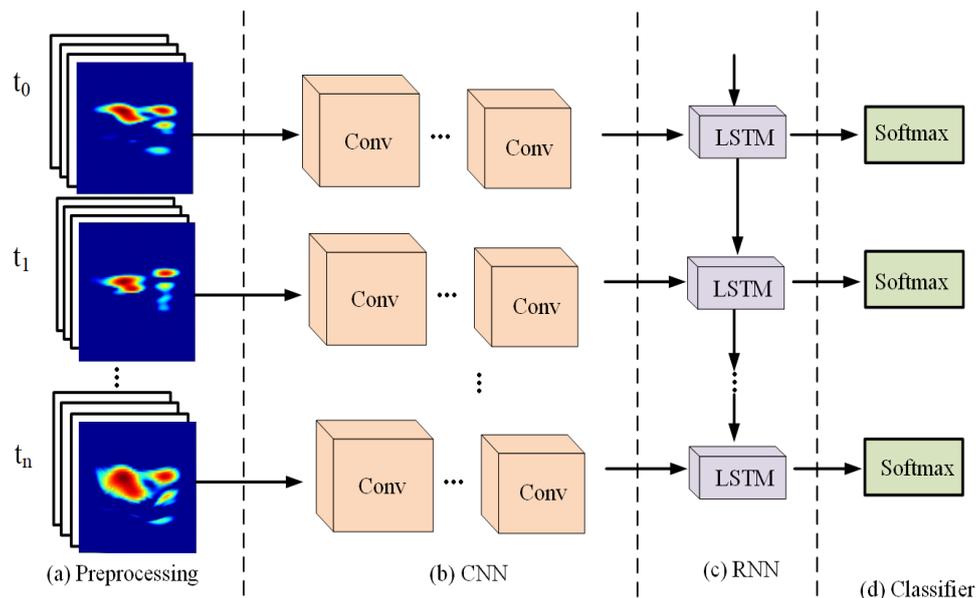
#### 4.1. Deep Learning Approaches in 3D Radar Echo

range–Doppler frames reveal moving properties, as well as micro-Doppler properties of targets [61]. Consisting of  $N$  time-sampled 2D range–Doppler frames, the 3D RD video sequence demonstrates both spatial and temporal characteristics. Range and Doppler information consists in every RD frame while time information exists between frames. Compared with 1D and 2D echoes, the joint time–range–Doppler echoes contain almost all the activity information that radar receives. Models that are able to extract both temporal and spatial information are required. Since it is difficult to design features manually from 3D echoes, DL methods are more feasible and preferable for 3D echo-based HAR, thanks to its capability of automatically extracting deep features. Furthermore, the advent of GPU makes it possible for DL models to process 3D data quickly and efficiently. Although there are few DL algorithms proposed for 3D radar echoes till now, DL approaches on 3D echoes are promising for HAR.

3D CNN is one of the most used models for processing 3D data recently [4,93,94]. It extends the spatial CNN into a spatio–temporal model, and spatial–temporal features are learned automatically. Z. Zhang et al. [28] proposed a recurrent 3-D CNN model for continuous dynamic gesture recognition using an FMCW radar. 3D CNN was used for extracting short temporal-spatial features in continuous time–range maps and then an LSTM was adopted for global temporal feature learning. Experiment showed that when 3D CNN was substituted with a traditional 2D-CNN, the recognition was reduced by around 5%, which demonstrated that compared with 2D CNN, 3D CNN was able to learn better representations of hand gestures. Though the input of 3D CNN is time–range maps, this approach is also suitable for a 3D data cube because the cube contains almost all the activity information in continuous time–range maps.

A representative example using 3D radar echoes for HAR is *GoogleSoli*, as shown in Figure 8. *GoogleSoli* is the first gesture recognition system capable of recognizing a rich set of dynamic gestures based on short-range FMCW radar [50,51]. It is based on an end-to-end trained combination of deep convolutional and recurrent neural networks, and the dataset is comprised of 3D radar echoes.

Combining CNN and LSTM could enhance the ability to recognize different activities that have varied time span and spatial distributions. It was shown that the approach with 3D range–Doppler videos was better than the frame-level classification approaches, and the end-to-end ‘CNN + LSTM’ method was able to explore the gesture information more fully than the single CNN or LSTM models. With the advent of *GoogleSoli*, other DL architectures have been proposed based on it [28,31,94].



**Figure 8.** Deep learning architecture of Google Soli, a hybrid model that consists of CNN and LSTM. Adopted from [50].

#### 4.2. Deep Learning Approaches in 2D Radar Echo

Containing plentiful information of human activity, 3D human backscattering echoes are still complicated to process. 2D radar echoes, which are mainly referred as time–Doppler map, time–range map and range–Doppler map, also carry sufficient human activity information. Generally, 2D echoes are treated as images, so along with the line of computer vision, CNN has become the most commonly utilized model for 2D echoes. Thus, 2D echo-based HAR is often transformed into an image classification task.

(1) **time–Doppler map** (also referred to as micro-Doppler signatures) includes sufficient time–varying Doppler information that is pivotal for radar-based HAR [95]. When a human target is moving, the main Doppler shift is caused by torso while micro-Doppler is produced by rotating or vibrating parts, such as legs, feet and hands. The range and velocities of every body parts are often different, as shown in Figure 9. When the target acts differently, the time–Doppler maps corresponding to these activities are various. time–Doppler maps are easy to obtain by transforming raw echoes with STFT [96] and other joint time–frequency analysis methods. A simple CW radar with one transmitter and one receiver could be employed for identifying basic human activities with time–Doppler maps. In addition, time–Doppler maps are intuitive and explicable. As a consequence, compared with other 2D radar echoes, the time–Doppler maps are most commonly used for radar-based HAR up to now [20,41–43,45,48,49,54,56,69].

R.P. Trommel et al. [45] applied a 14-layer deep CNN (DCNN) on time–Doppler maps to classify human gaits. The experimental result showed that the DCNN architecture was able to extract effective micro-Doppler features of human gaits even at lower frequencies or low SNR levels, which exceeded the performance of SVM and the artificial neural network. M.S. Seyfioglu et al. [52] employed a CAE architecture to discriminate 12 indoor human activities involving aided and unaided human motions, which often resulted in highly similar micro-Doppler signatures. The CAE model is composed of 3 convolutional layers and three deconvolutional layers, as illustrated in Figure 4. It is able to learn nuances in

the micro-Doppler signatures and obtains a good recognition performance of 94.2%. This HAR method shows the potential of radar-based health monitoring systems for assisted living. In [42], a DCNN-based hand gesture recognition system using time–Doppler maps was proposed. There were three convolutional layers and a fully connected layer in the model. In addition, how the DCNN effectively recognizing hand gestures in uncontrolled environments was investigated. Results showed that micro-Doppler signatures varied with aspect angle and distance to the radar, and recognition performance of the model under different scenarios. Ref. [47] proposed a DCNN architecture composed of cascaded convolutional network layers to classify human activities with time–Doppler maps, as shown in Figure 10. The Bayesian optimization with Gaussian prior process was utilized to optimize the network. Experimental results showed that the performance of this method was better than three existing feature-based methods.

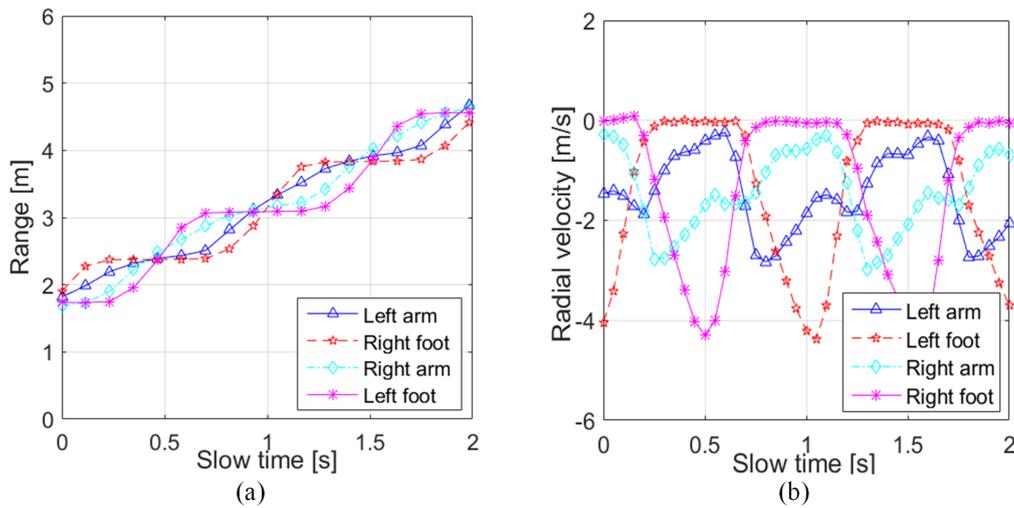


Figure 9. Moving trajectories of different body parts when a human target is walking: (a) Range of different parts. (b) Radial velocity of different parts. Adopted from [97].

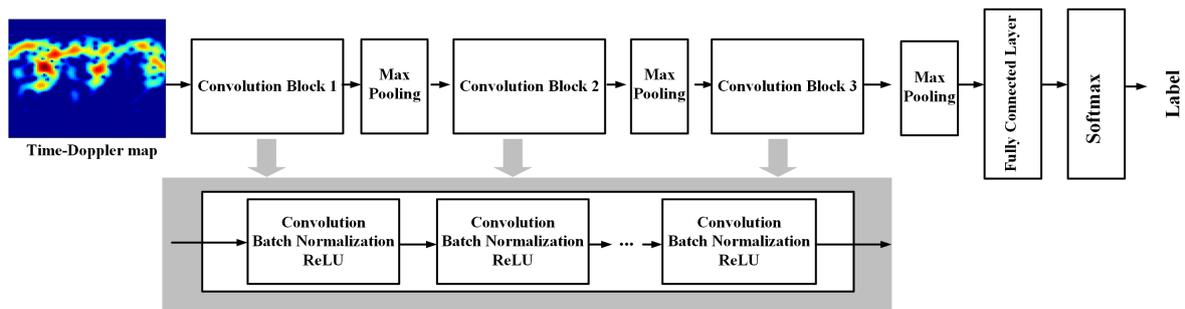


Figure 10. Cascaded DCNN optimized by Bayesian learning technique. Adopted from [47].

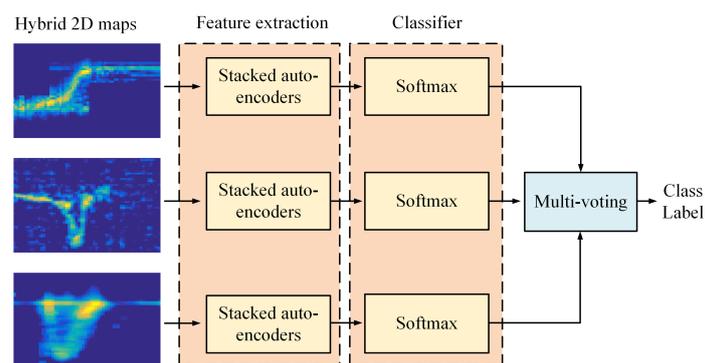
(2) **Time–range map** is composed of multiple pulses along time (see Figure 7c). It contains time-varying range information between the target and the radar. When a person is moving, different components of the human body have different relative distances from the radar, as illustrated in Figure 9a. As a result, although time–range maps neglect Doppler information, the time-varying range information of the human body is still able to be used for recognizing human activities [28].

In [98], time–range maps were utilized to detect falling in assisted living. By providing range information, the false alarms caused by fall-like activities such as sitting were reduced. In [22], Y. Shao et al. employed a three-layer DCNN to classify six human motions such as walking, running and boxing. It was shown that the time–range maps were more robust than the time–Doppler maps, especially when the radial velocity was low. Additionally, when increasing the incident angle,

the recognition accuracy was maintained at a stable value, because the range information did not change significantly with the signal to noise ratio.

(3) **range–Doppler map** (see Figure 7d) illustrates range and Doppler information of a moving target at a specific time. It has the ability to separate different components of the moving human body parts and locate the target accurately. In addition, range–Doppler maps are able to track multiple targets simultaneously, which is promising for multiple human activity recognition. P. Molchanov et al. [73] utilized a short-range monopulse FMCW radar with one Tx and three Rx to sense dynamic hand gestures. A 4D vector representing spatial coordinates and radial velocity of the hand was estimated with range–Doppler maps from three antennas. Similarly, in [74], a 4D vector obtained from three range–Doppler maps was combined with a mask from a depth image. Then a resulting velocity layer was fed into a 3D CNN to identify dynamic car-driver hand gestures. The 3D CNN is able to extract the spatial–temporal features, which is indispensable for recognizing dynamic hand gestures of short durations. In [57], two sparse AEs were stacked to learn sparse representation from range–Doppler maps gradually, and a Softmax layer was employed for classification. In [58], a stack AE was utilized to extract features from range–Doppler maps, and logistic regression was applied for identifying fall/non-fall. Ref. [57,58] gave examples of applying DL methods on range–Doppler maps for HAR.

(4) **Hybrid 2D maps** Up to now, most HAR systems based on 2D radar echoes only utilize one of the above three kinds of maps. However, sometimes it is observed that activities which could be easily distinguished with one map may not be correctly identified with another map. This motivates the use of multiple maps aiming at reducing false alarms. Ref. [99] utilized time–Doppler map, time–range map and range–Doppler map for falling detection. By extracting range and Doppler information from the three maps, the false alarm rate of fall detection was reduced. In [57], three stack AEs and three Softmax classifiers were employed to classify four human motions (falling, sitting, bending and walking), as described in Figure 11. In this method, time–Doppler maps, time–range maps and range–Doppler maps were all applied in order to fully explore the motion information that radar echoes contained. Then three classification results were combined to deliver the final result by voting strategy. Experiments showed that the performance was better than the one that only used one kind of maps. In [58], fall detection procedure was divided into two stages: using a stacked AE composed of two sparse AEs to distinguish fall/walk from sit/bend with time–range maps and using another stacked AE with the same structure to distinguish fall from walk with time–Doppler maps. Detection accuracy of 97.1% was achieved.

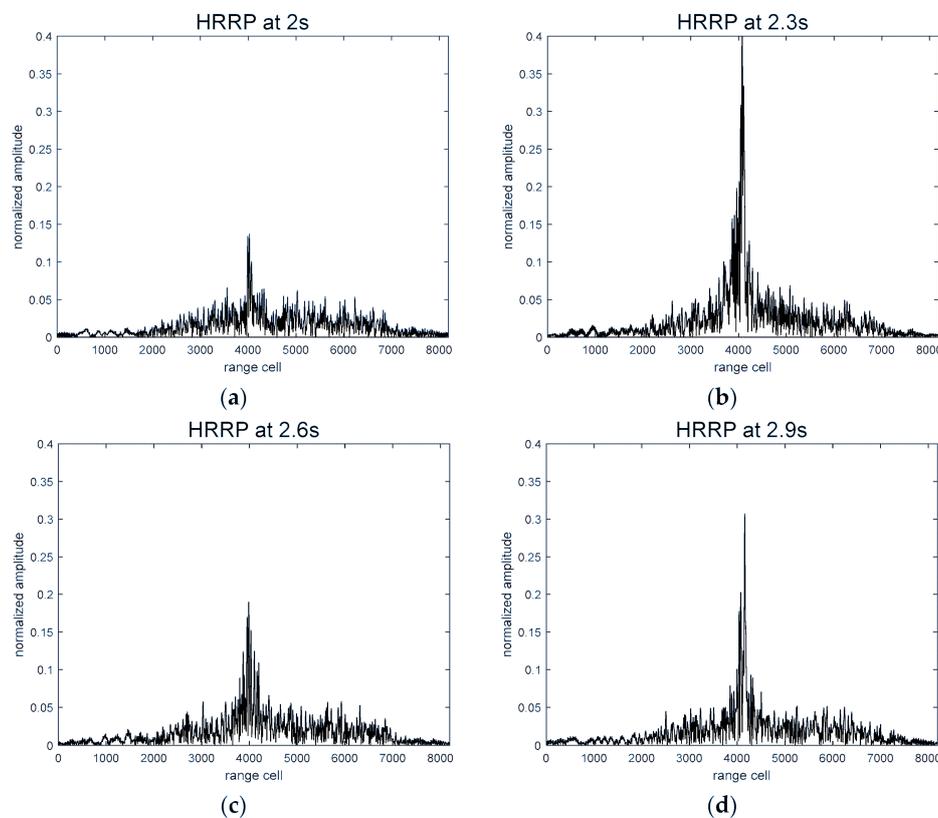


**Figure 11.** The scheme for hybrid 2D maps based recognition. Adopted from [57].

#### 4.3. Deep Learning Approaches in 1D Radar Echo

Projecting the time–range–Doppler data cube on range dimension results in 1D radar echoes, namely high resolution range profile (HRRP), as shown in Figure 12. Though HRRP is not as intuitive as 2D and 3D radar echoes, it carries enough information for identifying human activities likewise. Ref. [100] applied HRRP to analyze human target gaits with an ultra-wideband radar. Ref. [101] combined HRRP and micro-Doppler signatures to classify human gaits. Z. Zhou et al. adopted

multi-modal signals, including HRRPs and Doppler signatures acquired from a terahertz radar system to recognize dynamic gestures and the recognition rate reached more than 91% [31].



**Figure 12.** High resolution range profiles of a hand at a different time. Adopted from [31]. Each sub-figure illustrates the HRRP at a specific time.

1D radar echoes are essentially time-series, and similar to the data obtained from sensors like accelerometer and gyroscope. Thus, many approaches used for time series could be adopted to 1D echo-based HAR. RNN is often utilized for 1D data due to advantages of modeling sequential data. For instance, A. Graves et al. proposed a speech recognition architecture composed of LSTM and Connectionist Temporal Classification (CTC) algorithm that is suitable to label unsegmented sequence data [102]. This provides us insights on how to recognize continuous activities without annotating manually in advance. A. Hamid et al. [103] applied 1D CNN to hybrid NN-HMM model for speech recognition and proposed partial weight sharing for the first time. Although there are few DL-related studies for 1D radar echoes, DL approaches have the potential to extract sequential features and deliver good classification results for 1D radar echoes.

## 5. Future Directions

Despite radar-based HAR with DL algorithms has made noteworthy progress, there is a way to go before it matures. As a tool for feature extraction and activity identification, it is essential for the designed DL architectures to be capable of exploring activity information in radar echoes as much as possible. A few future research considerations are listed below.

### A. Complex human activity recognition.

Complex human activity, such as drinking coffee and cooking, is composed of several simple activities that are simultaneous. Compared with simple human activities such as walking, running and sitting, complex activities, which is more reflective of people's intentions, are worth studying. Due

to the complicated semantic and context information, complex activities, are harder to be recognized than the simple activities.

**(1) Hybrid deep model design.** In Table 3, most work adopts single and basic DL models, such as CNN and AE. However, as described in Section 2, each type of DL models has its own unique characteristics for HAR task. In order to fully take advantage of the semantic and context information for identification, it is far from enough to purely use a single DL model. Motivated by this, designing hybrid DL models for recognizing complex activities is imperative.

**(2) Multiple forms of echoes.** Table 3 shows that compared with 2D radar echoes, there is less work based on 1D and 3D echoes so far. It is mainly because that the current radar-based HAR tasks mostly focus on identifying simple activities. In this case, the information in 2D echoes is enough to obtain good recognition performance. In addition, 2D echoes are intuitive and explicable, which makes it more acceptable. Generally speaking, there is a loss of information during the radar signal transformation process, no matter the signals are converted into 1D, 2D or 3D. However, in order to make radar-based HAR systems more robust and generic for complex scenarios, more activity information in radar echoes should be utilized. To this end, different types of radar echoes could be employed for information extraction. Consequently, it is necessary to cooperate with multiple forms of echoes for HAR.

**(3) Aspect angle sensitivity.** In HAR task, Doppler shift is caused by the radial velocity of moving targets, and the radial velocity changes with the relative position between the target and radar. When the motion directions are different, radar backscattered signals produced by a subject differ a lot [104]. In this regard, the designed model should be robust to the aspect angle changes. Since there is a little research on this issue [22,42], it still needs to be investigated more.

## B. Radar-based human activity recognition in real-world scenarios.

So far, most of the recent radar-based HAR approaches are only applicable in the controllable environments, where a human target acts several discontinuous and assigned activities with little interference. In addition, the real-time processing capability of the model is not taken into account. However, in order to make radar-based HAR applied in real-world scenarios, several issues should be considered carefully.

**(1) Light-weight deep model design.** Training a DL model often requires lots of computing resources, which makes it often be executed off-line with a limited amount of data. However, in reality, activity data often come in a stream and require robust online and incremental learning. Though capable of processing and classifying data in real-time, huge feature engineering and hand-craft feature extraction hinder the use of traditional ML approaches for real-world HAR. Consequently, it is necessary to design light-weight DL models for radar-based HAR. There are two ideas available for investigation: combining hand-crafted features with deep features, and cooperating DL models with conventional ML algorithms.

**(2) Continuous activity segmentation and recognition.** In real-world scenarios, a person always acts continuously and freely, not merely performing assigned activities. Accurate segmentation and recognition of the interested activities is crucial. Recently, there is a trend of addressing segmentation and recognition jointly. For example, in [28], a Connectionist Temporal Classification (CTC) algorithm [105] was employed to recognize continuous dynamic hand gestures. CTC enables gesture recognition without explicit pre-segmentation and addresses segmentation and recognition simultaneously. In further research, more algorithms aiming at jointly segmenting and recognizing a series of activities are desired.

**(3) Multi-target activity recognition.** How to identify multiple targets' activities or separate the target from a group is worth studying. In [45,54], multi-target human gait recognition with DL approaches was studied. In [75], an FMCW radar was utilized to separate and recognize several assigned hand gestures in the presence of multiple targets. However, those solutions often work in less disturbing scenarios, such as the scenario where the human target is making gestures, and another

person is walking toward the radar meanwhile. When it comes to the circumstances where the radar echoes are modulated by multiple moving targets, applying DL models to learn high-level features are of great significance. More elaborate DL models should be designed for multi-target activity recognition.

### C. Unsupervised activity recognition in radar.

DL models require large-scale labeled data to prevent overfitting and obtain good generalization. In radar applications, however, acquiring a mass of measured labeled data is challenging due to constraints on manpower, cost, and other resources. As a result, unsupervised HAR in radar is urgent.

**(1) Deep Transfer learning.** Transfer learning generally refers to transferring the knowledge or models learned from a certain task to another related, but different task. Up to now, transfer learning for radar-based HAR mainly includes two perspectives: transferring the models trained with large-scale natural image datasets, such as ImageNet [84,89,90] and transferring the models trained with simulated radar image dataset [83,85]. How to elaborate a DL model that is capable of adequately learning the relatedness between source domain and target domain is an open issue in radar-based HAR area.

**(2) Cross-modal knowledge distillation.** For the activity recognition task, it is verified that the shared representations exist in different types of sensory data. In addition, the shared representations could be utilized as supervision for training a radar-based DL model. Only synchronized but unlabeled data are employed during the cross-modal knowledge distillation process. Ref. [106] demonstrates that training a model by cross-modal knowledge distillation not only reduces the amount of required labeled data but also speeds up the training process. For radar-based HAR, cross-modal knowledge distillation is also effective and large-scale labeled data are no longer necessary.

## 6. Conclusions

Human activity recognition is one of the interesting research topics in human–computer interaction and smart surveillance. As an active system for human activity recognition, radar has many unique advantages and has attracted the attention of researchers gradually. Deep learning is able to extract deep hierarchical features automatically and has achieved desirable classification performance. In this paper, we first survey several state-of-the-art deep learning models. Those models have different characteristics for identifying human activities and there is a trend of combining multiple models to better learn the features of human activities. Then, radar systems that are mostly employed for HAR are described. Doppler radar is able to obtain Doppler information for HAR while FMCW radar provides both range and Doppler information. UWB radar has a high range resolution and is capable of distinguishing the scattering centers of the human body. Interferometry radar provides Doppler information regardless of the directions of human movement. Furthermore, by classifying radar echoes into three different forms: 1D, 2D and 3D, we discuss the development of deep learning based HAR in radar. Various deep learning techniques designed specially for 1D/2D/3D radar echoes have been discussed, and the experiment results demonstrate the feasibility of such techniques. 2D radar echoes, especially time–Doppler maps, are more commonly used for radar-based HAR because they are more intuitive and contain sufficient activity information. 3D echoes contain more information, but they are also more difficult to process than 2D and 1D echoes. Because of the simple form of 1D echoes, the activity information contained in them is still waiting to be fully mined. Thanks to the ability of feature learning, DL techniques shows potential for radar-based HAR. Finally, several future research directions for radar-based HAR is presented. Though the adoption of radar for HAR is still lagging behind vision-based technologies, we should be optimistic about the potential of radar-based HAR techniques because of radar’s unique advantages such as environment-insensitivity and better privacy protection.

**Author Contributions:** X.L. was responsible for conducting research and writing the manuscript. Y.H. was responsible for formulating the research question, conducting research and writing the manuscript. X.J. was responsible for formulating the research question and revising the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors thank Francois Le Chevalier to give valuable advice to the manuscript. The authors would also be grateful to anonymous reviewers and academic editors for their constructive comments and helpful suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

HAR	Human activity recognition
FMCW	Frequency-modulated continuous-wave
CV	Computer vision
CNN	Convolutional neural network
TSN	Temporal segment network
ML	Machine learning
SVM	Support vector machine
DTW	Dynamic time warping
DL	Deep learning
NLP	Natural language processing
RNN	Recurrent neural network
LSTM	Long short term memory
RD	range–Doppler
HRRP	High resolution range profile
CTC	Connectionist temporal classification

## References

1. Cristani, M.; Raghavendra, R.; Bue, A.D.; Murino, V. Human behavior analysis in video surveillance: A Social Signal Processing perspective. *Neurocomputing* **2013**, *100*, 86–97. [[CrossRef](#)]
2. Qian, W.; Li, Y.; Li, C.; Pal, R. Gesture recognition for smart home applications using portable radar sensors. In Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 6414–6417.
3. Wang, J.L.; Singh, S. Video analysis of human dynamics—A survey. *Real-Time Imaging* **2003**, *9*, 321–346. [[CrossRef](#)]
4. Molchanov, P.; Gupta, S.; Kim, K.; Kautz, J. Hand gesture recognition with 3D convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–7.
5. Chen, L.; Zhou, M.; Wu, M.; She, J.; Liu, Z.; Dong, F.; Hirota, K. Three-layer Weighted Fuzzy Support Vector Regression for Emotional Intention Understanding in Human-Robot Interaction. *IEEE Trans. Fuzzy Syst.* **2018**, *26*, 2524–2538. [[CrossRef](#)]
6. Bulling, A.; Blanke, U.; Schiele, B. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv. (CSUR)* **2014**, *46*, 33. [[CrossRef](#)]
7. Huang, X.; Dai, M. Indoor Device-Free Activity Recognition Based on Radio Signal. *IEEE Trans. Veh. Technol.* **2017**, *66*, 5316–5329. [[CrossRef](#)]
8. Jalal, A.; Kim, Y.H.; Kim, Y.J.; Kamal, S.; Kim, D. Robust human activity recognition from depth video using spatiotemporal multi-fused features. *Pattern Recognit.* **2017**, *61*, 295–308. [[CrossRef](#)]
9. Jalal, A.; Kamal, S.; Kim, D. A Depth Video-based Human Detection and Activity Recognition using Multi-features and Embedded Hidden Markov Models for Health Care Monitoring Systems. *Int. J. Interact. Multimed. Artif. Intell.* **2017**, *4*, 54–62. [[CrossRef](#)]
10. Yang, X.; Tian, Y. Super normal vector for human activity recognition with depth cameras. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1028–1039. [[CrossRef](#)] [[PubMed](#)]

11. Simonyan, K.; Zisserman, A. Two-stream convolutional networks for action recognition in videos. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December, 2014; pp. 568–576.
12. Wang, L.; Xiong, Y.; Wang, Z.; Qiao, Y.; Lin, D.; Tang, X.; Van Gool, L. Temporal segment networks: Towards good practices for deep action recognition. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 20–36.
13. Ren, Y.; Zhu, C.; Xiao, S. Deformable Faster R-CNN with Aggregating Multi-Layer Features for Partially Occluded Object Detection in Optical Remote Sensing Images. *Remote Sens.* **2018**, *10*, 1470. [[CrossRef](#)]
14. Markman, A.; Shen, X.; Javidi, B. Three-dimensional object visualization and detection in low light illumination using integral imaging. *Opt. Lett.* **2017**, *42*, 3068–3071. [[CrossRef](#)] [[PubMed](#)]
15. Bouachir, W.; Gouiaa, R.; Li, B.; Noumeir, R. Intelligent video surveillance for real-time detection of suicide attempts. *Pattern Recognit. Lett.* **2018**, *110*, 1–7. [[CrossRef](#)]
16. Reyes-Ortiz, J.L.; Oneto, L.; Samà, A.; Parra, X.; Anguita, D. Transition-aware human activity recognition using smartphones. *Neurocomputing* **2016**, *171*, 754–767. [[CrossRef](#)]
17. Liu, Y.; Nie, L.; Liu, L.; Rosenblum, D.S. From action to activity: Sensor-based activity recognition. *Neurocomputing* **2016**, *181*, 108–115. [[CrossRef](#)]
18. Ronao, C.A.; Cho, S.B. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Syst. Appl.* **2016**, *59*, 235–244. [[CrossRef](#)]
19. Shoaib, M.; Bosch, S.; Incel, O.D.; Scholten, H.; Havinga, P.J. Complex human activity recognition using smartphone and wrist-worn motion sensors. *Sensors* **2016**, *16*, 426. [[CrossRef](#)] [[PubMed](#)]
20. Le, H.T.; Phung, S.L.; Bouzerdoum, A. Human Gait Recognition with Micro-Doppler Radar and Deep Autoencoder. In Proceedings of the IEEE 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 3347–3352.
21. Lin, Y.; Le Kernec, J.; Yang, S.; Fioranelli, F.; Romain, O.; Zhao, Z. Human Activity Classification With Radar: Optimization and Noise Robustness with Iterative Convolutional Neural Networks Followed with Random Forests. *IEEE Sens. J.* **2018**, *18*, 9669–9681. [[CrossRef](#)]
22. Shao, Y.; Guo, S.; Sun, L.; Chen, W. Human Motion Classification Based on Range Information with Deep Convolutional Neural Network. In Proceedings of the International Conference on Information Science and Control Engineering (ICISCE), Changsha, China, 21–23 July 2017; pp. 1519–1523.
23. Chen, Z.; Li, G.; Fioranelli, F.; Griffiths, H. Personnel Recognition and Gait Classification Based on Multistatic Micro-Doppler Signatures Using Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 669–673. [[CrossRef](#)]
24. Chen, V.C.; Li, F.; Ho, S.S.; Wechsler, H. Micro-Doppler effect in radar: Phenomenon, model, and simulation study. *IEEE Trans. Aerosp. Electron. Syst.* **2006**, *42*, 2–21. [[CrossRef](#)]
25. Zenaldin, M.; Narayanan, R.M. Radar micro-Doppler based human activity classification for indoor and outdoor environments. In Proceedings of the SPIE Conference on Radar Sensor Technology XX, Baltimore, MD, USA, 18–21 April 2016.
26. Qi, F.; Lv, H.; Liang, F.; Li, Z.; Yu, X.; Wang, J. MHHT-based method for analysis of micro-Doppler signatures for human finer-grained activity using through-wall SFCW radar. *Remote Sens.* **2017**, *9*, 260. [[CrossRef](#)]
27. Smith, K.; Csech, C.; Murdoch, D.; Shaker, G. Gesture Recognition Using mm-Wave Sensor for Human-Car Interface. *IEEE Sens. Lett.* **2018**, *2*, 1–4. doi:10.1109/LENS.2018.2810093. [[CrossRef](#)]
28. Zhang, Z.; Tian, Z.; Zhou, M. Latern: Dynamic Continuous Hand Gesture Recognition Using FMCW Radar Sensor. *IEEE Sens. J.* **2018**, *18*, 3278–3289. [[CrossRef](#)]
29. Kim, Y. Detection of eye blinking using Doppler sensor with principal component analysis. *IEEE Antennas Wirel. Propag. Lett.* **2015**, *14*, 123–126. [[CrossRef](#)]
30. Kim, Y.; Ling, H. Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 1328–1337.
31. Zhou, Z.; Cao, Z.; Pi, Y. Dynamic Gesture Recognition with a Terahertz Radar Based on Range Profile Sequences and Doppler Signatures. *Sensors* **2017**, *18*, 10. doi:10.3390/s18010010. [[CrossRef](#)]
32. Wang, J.; Chen, Y.; Hao, S.; Peng, X.; Hu, L. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognit. Lett.* **2018**, *119*, 3–11. [[CrossRef](#)]
33. Deng, L.; Yu, D. Deep learning: Methods and applications. *Found. Trends Signal Process.* **2014**, *7*, 197–387. [[CrossRef](#)]

34. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [[CrossRef](#)]
35. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the National Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 4278–4284.
36. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
37. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)] [[PubMed](#)]
38. Kang, M.; Ji, K.; Leng, X.; Lin, Z. Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection. *Remote Sens.* **2017**, *9*, 860. [[CrossRef](#)]
39. Lin, Z.; Ji, K.; Leng, X.; Kuang, G. Squeeze and Excitation Rank Faster R-CNN for Ship Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 751–755. [[CrossRef](#)]
40. Lin, Z.; Ji, K.; Kang, M.; Leng, X.; Zou, H. Deep convolutional highway unit network for sar target classification with limited labeled training data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1091–1095. [[CrossRef](#)]
41. Kim, Y.; Moon, T. Human Detection and Activity Classification Based on Micro-Doppler Signatures Using Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 8–12. [[CrossRef](#)]
42. Kim, Y.; Toomajian, B. Hand Gesture Recognition Using Micro-Doppler Signatures with Convolutional Neural Network. *IEEE Access* **2016**, *4*, 7125–7130. [[CrossRef](#)]
43. Kim, Y.; Toomajian, B. Application of Doppler radar for the recognition of hand gestures using optimized deep convolutional neural networks. In Proceedings of the European Conference on Antennas and Propagation, Paris, France, 19–24 March 2017; pp. 1258–1260.
44. Lang, Y.; Hou, C.; Yang, Y.; Huang, D.; He, Y. Convolutional neural network for human micro-Doppler classification. In Proceedings of the European Microwave Conference, Nuremberg, Germany, 8–13 October 2017.
45. Trommel, R.P.; Harmanny, R.I.A.; Cifola, L.; Driessen, J.N. Multi-target human gait classification using deep convolutional neural networks on micro-doppler spectrograms. In Proceedings of the European Radar Conference, London, UK, 5–7 October 2016; pp. 81–84.
46. Yang, Y.; Hou, C.; Lang, Y.; Guan, D.; Huang, D.; Xu, J. Open-set human activity recognition based on micro-Doppler signatures. *Pattern Recognit.* **2019**, *85*, 60–69. [[CrossRef](#)]
47. Le, H.T.; Phung, S.L.; Bouzardoum, A.; Tivive, F.H.C. Human Motion Classification with Micro-Doppler Radar and Bayesian-Optimized Convolutional Neural Networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 2961–2965.
48. Shao, Y.; Dai, Y.; Yuan, L.; Chen, W. Deep Learning Methods for Personnel Recognition based on Micro-Doppler Features. In Proceedings of the 9th International Conference on Signal Processing Systems, AUT, Auckland, New Zealand, 27–30 November 2017; pp. 94–98.
49. Zhang, J.; Tao, J.; Shi, Z. Doppler-Radar Based Hand Gesture Recognition System Using Convolutional Neural Networks. In Proceedings of the IEEE International Conference in Communications, Signal Processing, and Systems, Harbin, China, 14–16 July 2017; pp. 1096–1113.
50. Wang, S.; Song, J.; Lien, J.; Poupyrev, I.; Hilliges, O. Interacting with Soli: Exploring Fine-Grained Dynamic Gesture Recognition in the Radio-Frequency Spectrum. In Proceedings of the ACM Symposium on User Interface Software and Technology, Tokyo, Japan, 16–19 October 2016; pp. 851–860.
51. Lien, J.; Gillian, N.; Karagozler, M.E.; Amihhood, P.; Schwesig, C.; Olson, E.; Raja, H.; Poupyrev, I. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Trans. Graph.* **2016**, *35*, 142. [[CrossRef](#)]
52. Seyfioğlu, M.S.; Özbayğlı, A.M.; Gurbuz, S.Z. Deep Convolutional Autoencoder for Radar-Based Classification of Similar Aided and Unaided Human Activities. *IEEE Trans. Aerosp. Electron. Syst.* **2018**, *54*, 1709–1723. [[CrossRef](#)]
53. Graves, A. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780.
54. Klarenbeek, G.; Harmanny, R.I.A.; Cifola, L. Multi-target human gait classification using LSTM recurrent neural networks applied to micro-Doppler. In Proceedings of the European Radar Conference, Nuremberg, Germany, 11–13 October 2017; pp. 167–170.

55. Erhan, D.; Bengio, Y.; Courville, A.; Manzagol, P.A.; Vincent, P.; Bengio, S. Why does unsupervised pre-training help deep learning? *J. Mach. Learn. Res.* **2010**, *11*, 625–660.
56. Jokanovic, B.; Amin, M.; Ahmad, F. Radar fall motion detection using deep learning. In Proceedings of the IEEE Radar Conference (RadarConf), Philadelphia, PA, USA, 2–6 May 2016; pp. 1–6.
57. Jokanovic, B.; Amin, M.; Erol, B. Multiple joint-variable domains recognition of human motion. In Proceedings of the IEEE Radar Conference, Seattle, WA, USA, 8–12 May 2017; pp. 0948–0952.
58. Jokanović, B.; Amin, M. Fall detection using deep learning in range-Doppler radars. *IEEE Trans. Aerosp. Electron. Syst.* **2018**, *54*, 180–189. [[CrossRef](#)]
59. Kang, M.; Ji, K.; Leng, X.; Xing, X.; Zou, H. Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder. *Sensors* **2017**, *17*, 192. [[CrossRef](#)]
60. Li, C.; Peng, Z.; Huang, T.Y.; Fan, T.; Wang, F.K.; Horng, T.S.; Muñoz-Ferreras, J.M.; Gómez-García, R.; Ran, L.; Lin, J. A Review on Recent Progress of Portable Short-Range Noncontact Microwave Radar Systems. *IEEE Trans. Microw. Theory Tech.* **2017**, *65*, 1692–1706. [[CrossRef](#)]
61. Peng, Z.; Li, C. Portable Microwave Radar Systems for Short-Range Localization and Life Tracking: A Review. *Sensors* **2019**, *19*, 1136. [[CrossRef](#)]
62. Nanzer, J.A. A Review of Microwave Wireless Techniques for Human Presence Detection and Classification. *IEEE Trans. Microw. Theory Tech.* **2017**, *65*, 1780–1794. [[CrossRef](#)]
63. Lukin, K.; Konovalov, V. Through wall detection and recognition of human beings using noise radar sensors. In Proceedings of the NATO RTO SET Symposium on Target Identification and Recognition Using RF Systems, Oslo, Norway, 11–13 October 2004; pp. 15-1–15-11.
64. Lai, C.P.; Ruan, Q.; Narayanan, R.M. Hilbert-Huang transform (HHT) analysis of human activities using through-wall noise radar. In Proceedings of the International Symposium on Signals, Systems and Electronics, Montreal, QC, Canada, 30 July–2 August 2007.
65. Narayanan, R.M. Through-wall radar imaging using UWB noise waveforms. *J. Frankl. Inst.* **2008**, *345*, 659–678. [[CrossRef](#)]
66. Lai, C.P.; Narayanan, R.; Ruan, Q.; Davydov, A. Hilbert–Huang transform analysis of human activities using through-wall noise and noise-like radar. *IET Radar Sonar Navig.* **2008**, *2*, 244–255. [[CrossRef](#)]
67. Lai, C.P.; Narayanan, R.M. Ultrawideband random noise radar design for through-wall surveillance. *IEEE Trans. Aerosp. Electron. Syst.* **2010**, *46*, 1716–1730. [[CrossRef](#)]
68. Susek, W.; Stec, B. Through-the-wall detection of human activities using a noise radar with microwave quadrature correlator. *IEEE Trans. Aerosp. Electron. Syst.* **2015**, *51*, 759–764. [[CrossRef](#)]
69. Wang, M.; Zhang, Y.D.; Cui, G. Human motion recognition exploiting radar with stacked recurrent neural network. *Digit. Signal Process.* **2019**, *87*, 125–131. [[CrossRef](#)]
70. Mercuri, M.; Liu, Y.; Lorato, I.; Torfs, T.; Wieringa, F.P.; Bourdoux, A.; Van Hoof, C. A Direct Phase-Tracking Doppler Radar Using Wavelet Independent Component Analysis for Non-Contact Respiratory and Heart Rate Monitoring. *IEEE Trans. Biomed. Circuits Syst.* **2018**, *12*, 632–643. [[CrossRef](#)] [[PubMed](#)]
71. Rahman, A.; Lubecke, V.; Boriclubecke, O.; Prins, J.; Sakamoto, T. Doppler Radar Techniques for Accurate Respiration Characterization and Subject Identification. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2018**, *8*, 350–359. [[CrossRef](#)]
72. Kim, J.Y.; Park, J.H.; Jang, S.Y.; Yang, J.R. Peak Detection Algorithm for Vital Sign Detection Using Doppler Radar Sensors. *Sensors* **2019**, *19*, 1575. [[CrossRef](#)] [[PubMed](#)]
73. Molchanov, P.; Gupta, S.; Kim, K.; Pulli, K. Short-range FMCW monopulse radar for hand-gesture sensing. In Proceedings of the IEEE Radar Conference, Arlington, VA, USA, 10–15 May 2015; pp. 1491–1496.
74. Molchanov, P.; Gupta, S.; Kim, K.; Pulli, K. Multi-sensor system for driver’s hand-gesture recognition. In Proceedings of the 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; Volume 1, pp. 1–8.
75. Peng, Z.; Li, C.; Muñoz-Ferreras, J.M.; Gómez-García, R. An FMCW radar sensor for human gesture recognition in the presence of multiple targets. In Proceedings of the 2017 First IEEE MTT-S International Microwave Bio Conference (IMBIOC), Gothenburg, Sweden, 15–17 May 2017; pp. 1–3.
76. Zhou, H.; Cao, P.; Chen, S. A novel waveform design for multi-target detection in automotive FMCW radar. In Proceedings of the IEEE Radar Conference (RadarConf), Philadelphia, PA, USA, 2–6 May 2016; pp. 1–5.
77. Kim, B.S.; Jin, Y.; Kim, S.; Lee, J. A Low-Complexity FMCW Surveillance Radar Algorithm Using Two Random Beat Signals. *Sensors* **2019**, *19*, 608. [[CrossRef](#)]

78. Nanzer, J.A. Millimeter-Wave Interferometric Angular Velocity Detection. *IEEE Trans. Microw. Theory Tech.* **2010**, *58*, 4128–4136. [[CrossRef](#)]
79. Hariharan, P.; Creath, K. Basics of Interferometry. *Phys. Today* **1993**, *46*, 75. [[CrossRef](#)]
80. Peng, Z.; Muñoz-Ferrer, J.M.; Tang, Y.; Liu, C.; Gómez-García, R.; Ran, L.; Li, C. A portable FMCW interferometry radar with programmable low-IF architecture for localization, ISAR imaging, and vital sign tracking. *IEEE Trans. Microw. Theory Tech.* **2017**, *65*, 1334–1344. [[CrossRef](#)]
81. Wang, G.; Gu, C.; Inoue, T.; Li, C. A Hybrid FMCW-Interferometry Radar for Indoor Precise Positioning and Versatile Life Activity Monitoring. *IEEE Trans. Microw. Theory Tech.* **2014**, *62*, 2812–2822. [[CrossRef](#)]
82. Nanzer, J.A. Micro-motion signatures in radar angular velocity measurements. In Proceedings of the IEEE Radar Conference, Philadelphia, PA, USA, 2–6 May 2016; pp. 1–4.
83. Kim, Y.; Park, J.; Moon, T. Classification of micro-Doppler signatures of human aquatic activity through simulation and measurement using transferred learning. In Proceedings of the Radar Sensor Technology XXI. International Society for Optics and Photonics, Anaheim, CA, USA, 10–12 April 2017; Volume 10188, p. 101880V.
84. Du, H.; He, Y.; Jin, T. Transfer Learning for Human Activities Classification Using Micro-Doppler Spectrograms. In Proceedings of the IEEE International Conference on Computational Electromagnetics (ICCEM), Chengdu, China, 26–28 March 2018; pp. 1–3.
85. Lang, Y.; Wang, Q.; Yang, Y.; Hou, C.; Huang, D.; Xiang, W. Unsupervised Domain Adaptation for Micro-Doppler Human Motion Classification via Feature Fusion. *IEEE Geosci. Remote Sens. Lett.* **2018**, *6*, 392–396. [[CrossRef](#)]
86. Yarovoy, A.; Lighthart, L.; Matuzas, J.; Levitas, B. UWB radar for human being detection. *IEEE Aerosp. Electron. Syst. Mag.* **2006**, *21*, 10–14. [[CrossRef](#)]
87. Bryan, J.; Kim, Y. Classification of human activities on UWB radar using a support vector machine. In Proceedings of the IEEE Antennas and Propagation Society International Symposium, Toronto, ON, Canada, 11–17 July 2010; pp. 1–4.
88. Bryan, J.; Kwon, J.; Lee, N.; Kim, Y. Application of ultra-wide band radar for classification of human activities. *IET Radar Sonar Navig.* **2012**, *6*, 172–179. [[CrossRef](#)]
89. Seyfioğlu, M.S.; Gürbüz, S.Z. Deep neural network initialization methods for micro-Doppler classification with low training sample support. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2462–2466. [[CrossRef](#)]
90. Park, J.; Javier, R.J.; Moon, T.; Kim, Y. Micro-Doppler based classification of human aquatic activities via transfer learning of convolutional neural networks. *Sensors* **2016**, *16*, 1990. [[CrossRef](#)]
91. Cao, P.; Xia, W.; Ye, M.; Zhang, J.; Zhou, J. Radar-ID: Human identification based on radar micro-Doppler signatures using deep convolutional neural networks. *IET Radar Sonar Navig.* **2018**, *12*, 729–734. [[CrossRef](#)]
92. He, Y.; Le Chevalier, F.; Yarovoy, A.G. Range-Doppler processing for indoor human tracking by multistatic ultra-wideband radar. In Proceedings of the 13th International Radar Symposium (IRS), Warsaw, Poland, 23–25 May 2012; pp. 250–253.
93. Ji, S.; Xu, W.; Yang, M.; Yu, K. 3D convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 221–231. [[CrossRef](#)] [[PubMed](#)]
94. Sang, Y.; Shi, L.; Liu, Y. Micro Hand Gesture Recognition System Using Ultrasonic Active Sensing. *IEEE Access* **2017**, *6*, 49339–49347. [[CrossRef](#)]
95. Tahmoush, D. Review of micro-Doppler signatures. *IET Radar Sonar Navig.* **2015**, *9*, 1140–1146. [[CrossRef](#)]
96. Chen, V.C.; Qian, S. Joint time-frequency transform for radar range-Doppler imaging. *IEEE Trans. Aerosp. Electron. Syst.* **1998**, *34*, 486–499. [[CrossRef](#)]
97. He, Y.; Molchanov, P.; Sakamoto, T.; Aubry, P.; Chevalier, F.L.; Yarovoy, A. Range-Doppler surface: A tool to analyse human target in ultra-wideband radar. *IET Radar Sonar Navig.* **2015**, *9*, 1240–1250. [[CrossRef](#)]
98. Erol, B.; Amin, M.; Zhou, Z.; Zhang, J. Range information for reducing fall false alarms in assisted living. In Proceedings of the IEEE Radar Conference, Philadelphia, PA, USA, 2–6 May 2016; pp. 1–6.
99. Erol, B.; Amin, M.G. Fall motion detection using combined range and Doppler features. In Proceedings of the 24th European Signal Processing Conference (EUSIPCO), Budapest, Hungary, 29 August–2 September 2016; pp. 2075–2080.
100. Wang, Y.; Fathy, A.E. UWB micro-doppler radar for human gait analysis using joint range-time-frequency representation. *Proc. SPIE* **2013**, *8734*, 873404.

101. Cammenga, Z.A.; Smith, G.E.; Baker, C.J. Combined high range resolution and micro-Doppler analysis of human gait. In Proceedings of the IEEE Radar Conference, Arlington, VA, USA, 10–15 May 2015; pp. 1038–1043.
102. Graves, A.; Mohamed, A.R.; Hinton, G. Speech recognition with deep recurrent neural networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC, Canada, 26–31 May 2013; pp. 6645–6649.
103. Abdel-Hamid, O.; Mohamed, A.R.; Jiang, H.; Penn, G. Applying Convolutional Neural Networks concepts to hybrid NN-HMM model for speech recognition. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Kyoto, Japan, 25–30 March 2012; pp. 4277–4280.
104. Vignaud, L.; Ghaleb, A.; Kernec, J.L.; Nicolas, J.M. Radar high resolution range & micro-Doppler analysis of human motions. In Proceedings of the International Radar Conference—Surveillance for a Safer World, Bordeaux, France, 12–16 October 2010; pp. 1–6.
105. Graves, A.; Gomez, F. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In Proceedings of the International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 369–376.
106. Xing, T.; Sandha, S.S.; Balaji, B.; Chakraborty, S.; Srivastava, M. Enabling Edge Devices that Learn from Each Other: Cross Modal Training for Activity Recognition. In Proceedings of the 1st International Workshop on Edge Systems, Analytics and Networking, Munich, Germany, 10–15 June 2018; ACM: New York, NY, USA, 2018; pp. 37–42.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).