

# Human Activity Detection and Coarse Localization Outdoors Using Micro-Doppler Signatures

Fei Luo, Stefan Poslad, and Eliane Bodanese

**Abstract**—Human activity detection outdoors is emerging as a very important research field due to its potential application in surveillance, assisted living, search and rescue, and military applications. For such applications it is important to have detailed information about the human target, for example, whether the detected target is a single person or a group of people, what activity a target is performing, and the rough location of the target. In this paper, we propose novel usage of machine learning techniques to perform subject classification, human activity classification, people counting, and coarse localization by classifying micro-Doppler signatures obtained from a low-cost and low-power radar system. Our experiments were performed outdoors. For feature extraction of micro-Doppler signatures, we applied a two-directional two-dimensional principle component analysis (2D2PCA). Our results show that by applying 2D2PCA, the accuracy results of Support Vector Machine (SVM) and  $k$ -nearest neighbors ( $k$ NN) classifiers were greatly improved. We also designed and implemented a Convolutional Neural Network (CNN) for the target classifications in terms of type, number, activity and coarse localization. Our CNN model obtained very high classification accuracies (97% to 100%), which are superior to the best results obtained by SVM and  $k$ NN. Finally, we investigated the effects of the frame length of the sliding window, the angle of the direction of movement, and the number of radars used on the classification performance, providing valuable guidelines for machine learning modeling and experimental setup of micro-Doppler based research and applications.

**Index Terms**—Micro-Doppler, Machine Learning, Convolutional Neural Network (CNN), two-directional two-dimensional principle component analysis (2D2PCA), Support Vector Machine (SVM),  $k$ -nearest neighbors ( $k$ NN), human activity classification, coarse localization, people counting.

## I. INTRODUCTION

THE detection, recognition, and classification of human targets and human activities are an increasingly important topic in many applications, such as surveillance, search and rescue, and person centered monitoring. Numerous sensors, such as cameras, LIDAR [1], GPS [2] and radars, have been employed to achieve contactless measurement of humans and human kinetic characteristics. Although such a wide variety of sensors exist, Doppler radar is emerging as an increasingly popular device that it is especially useful for motion analysis [3]. Unlike these other sensor technologies, it is all-weather, contactless, it works day-and-night, and it operates in non-line of sight situations such as through building walls, clothes, and foliage [4]. It is non-intrusive, and it does not generate privacy concerns because the identity or personal identifiable features of a target cannot be obtained with radar detection. Furthermore, low-cost and low-power radar components are becoming more available and accessible, which makes Doppler

radars more suitable to be deployed outdoors and on a larger scale.

Human activity classification can be obtained by comparing the differences in the radar micro-Doppler signatures of different targets and activities. A moving target relative to a radar sensor induces a frequency shift of the echo as a result of the well-known Doppler Effect. Additional movements of smaller parts of the target, called micro-motions, will result in additional modulation of the main Doppler frequency shift, known as the *micro-Doppler effect* [5], [6]. The distinctive characteristics of the observed micro-Doppler effect of an object or a process are called *micro-Doppler signatures* [7]. For walking and running, a unique micro-Doppler signature is the periodic motion of arms and legs that produces sidebands to the main Doppler frequency [8]. Micro-Doppler signatures are typically represented using joint time-frequency analysis such as Short-Time Fourier Transform (STFT) [8], [9].

In recent years, there has been a great research interest in human activity classification using micro-Doppler signatures [10], [11], [12], [13]. In [14], a low-power pulse-Doppler radar that operates at 5.8 GHz was used to collect the micro-Doppler signatures of three different activities (walking, running, and crawling) performed by four subjects on a treadmill. Kim and Lin [15] used a 2.4 GHz Doppler radar to classify seven human activities, including running, walking, boxing, etc. I. Bilik et al. [16] employed a Pulse-Doppler radar operating at 9 GHz to perform automatic target recognition on multiple people, wheeled vehicles, tracked vehicles, and animals. Fairchild et al. [17] built a bistatic radar system operating at 4 GHz to differentiate three human motions, such as no activity, arm swinging, and picking up an object.

In the published work above, various methods, such as Principle Component Analysis (PCA) [18], Empirical Mode Decomposition (EMD) [19], and Mel-frequency Cepstral Coefficients (MFCC) [20] were used to extract the micro-Doppler features. These are computer algorithms that extract features automatically, like other possible techniques such as Linear Discriminate Analysis (LDA) and Gabor wavelet filter. They are more efficient and informative compared to handcrafted feature extraction, where features are extracted manually by human visual judgement such as in [14], [15], [21], [22].

After being extracted, the micro-Doppler features are fed into classifiers. The most used classifiers in micro-Doppler based human activity classification are Support Vector Machine (SVM) [15], [21], [23],  $k$ -Nearest-Neighbors ( $k$ NN) [14] and Naïve Bayes [24]. Deep learning is increasingly being used to classify micro-Doppler signatures [25], [26]. The hierarchical structure of deep learning enables automatic learning of the features from raw data without relying on

The author sequence of this paper follows the first-last-author-emphasis norm (FLAE).

feature extraction methods.

So far, most related work was implemented in confined spaces [14], [27], [20]. Micro-Doppler based human activity detection outdoors has not been extensively investigated. In [21], the authors investigated human motion in four different environments including free space, through-the-wall, leaf tree foliage, needle tree foliage. Their results indicate that movement of leaves and trees outdoors can cause interference to the micro-Doppler signatures and may contribute to making the environment less stable. Additionally, animals may be confused with humans, creating false positives. Hence, before performing human activity detection, it is also important to differentiate humans from such confusers. Furthermore, the radars used in the related work tend to be high power radars and need constant mains power to function, which is an important shortcoming for lasting continuous detection outdoors, especially in areas where the electricity supply may be scarce or absent. Our proposed research addresses this problem by using low-power pulsed Doppler radars operating at 5.8 GHz that allows long-lasting battery powered operation.

Outdoors usually comprises a large area, and it is important not only to know what the target is doing but also where the target is. It is not feasible to perform localization by using range-Doppler analysis, because the low-power Doppler radars used here cannot provide information of the azimuth and the distance to the target. To the best of our knowledge, there is no relevant literature research applying micro-Doppler signature classification for coarse-grained location estimation.

The main contributions of this paper can be summarized as follows: Firstly, we obtained micro-Doppler signatures from a simple, very low-power pulsed radar system and proposed novel usage of machine learning that achieved high accuracy classification results for subject classification, human activity classification, people counting, and coarse localization in foliage cluttered outdoors; our designed CNN (RadarNet) achieved the best results with 96.69% to 100% average class accuracy for all four classification tasks; we proposed the use of the 2D2PCA to extract the micro-Doppler features to greatly improve the classification accuracy of SVM and  $k$ NN. Secondly, we developed a novel approach to perform coarse-grained estimation of the distance between the radar and the target by solely using micro-Doppler signature classification; Finally, our investigations on the effects of frame length of the sliding window, angles of the movement, and the number of radars on the classification performance provide valuable guidelines for micro-Doppler based research and applications.

The remainder of this paper is organized as follows. Section II introduces the fundamental concepts of micro-Doppler, the characteristics of the Bumblebee radar, and describes the composition of the Doppler radar system in detail. Section III describes the three experiments performed in this research; it presents and analyzes the pattern of spectrograms that were generated by the different subjects, angles, and ranges. Section IV shows how the radar signals were collected and stored in a database and describes the methods used for data processing and the composition of the samples after data processing. Section V describes the mathematical fundamentals of 2D2PCA and how it can be applied to micro-Doppler feature extraction.

Section VI presents our proposed structure of the CNN for classification of the gathered micro-Doppler signatures, and models the SVM and  $k$ NN classifiers by optimizing and listing their hyper-parameters. Section VII presents, analyzes, and compares the classification results of the five classifiers for human activity, people counting, human versus animal, and coarse localization. Section VIII investigates the effects of the frame length of the sliding window, angles of the movement, and the number of radars. Section IX summarizes the contributions and discusses the future prospects of the research.

## II. MICRO-DOPPLER AND THE LOW-POWER RADAR SYSTEM

### A. Micro-Doppler

The concept of micro-Doppler was first proposed by Chen in 2000 [9]. Micro-Doppler signatures reflect the periodic kinetic characteristics of a moving object. Modulations of the radar resulted from the arms, the legs and even the body sway have been investigated by researchers [4], [28], [29].

Given an electromagnetic wave transmitted by a radio frequency (RF) radar, the frequency of the received signals due to a moving target with a constant radial velocity  $v$  with respect to the radar is:

$$f = f_0(1 + 2v/c), \quad (1)$$

where  $f_0$  is the carrier frequency of the radar and  $c$  is the speed of light. The Doppler frequency shift due to the moving target is:

$$f_D = f_0(2v/c), \quad (2)$$

which is proportional to the velocity of the target relative to the radar.

In the case of an articulated body such as a walking person, the torso, each arm and each leg has its own velocity, and even when the torso's velocity is constant, the velocity of the limbs changes over time [29]. The Doppler signature  $f_{Dsig}$  for such a complex object has multiple time-dependent frequency shifted components and it is defined as:

$$f_{Dsig}(t) = f_0 \sum_{i=1}^N 2v_i(t)/c, \quad (3)$$

where  $N$  is the number of parts of the moving target,  $v_i(t)$  is the velocity of each part as a function of the time.

The transmitted signal is generally expressed in polar format as

$$\hat{S}_T(t) = e^{j2\pi f_0 t}, \quad (4)$$

The analytic signal of the returned echo from such a target is given by:

$$\hat{S}_R(t) = e^{j2\pi f_0 t} e^{j2\pi f_{Dsig}(t)t}, \quad (5)$$

The combination of the received signal  $\hat{S}_R(t)$  with the transmitted signal  $\hat{S}_T(t)$  is given as follows:

$$\hat{S}_R(t)\hat{S}_T(t)^* = e^{j2\pi f_{Dsig}t}, \quad (6)$$

This allows the extraction of the Doppler signature from the data. This is the component of the signal that contains

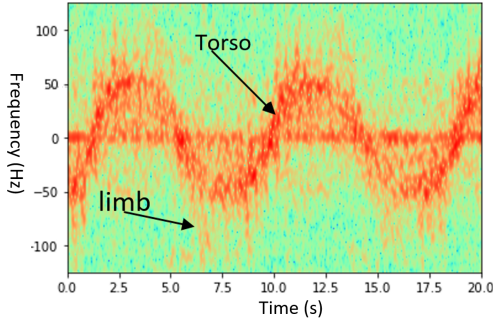


Figure 1. A time-frequency spectrogram of indoor human walking

the micro-Doppler information of the target and it can be used for target or activity recognition and classification. The bandwidth of this resulting signal is normally much smaller than the carrier frequency, because the radial speed of the target is small compared to the speed of light, producing lower Doppler frequencies [5]. The micro-Doppler signature can be represented in a two-dimensional time-frequency space using an STFT [5], [30]:

$$STFT(i, K) = \sum_{n=0}^{N-1} x_i(n) e^{-j2\pi(nK/N)}, \quad (7)$$

$$K = 0, \dots, N - 1$$

where  $x_i(n)$  is the sliding window with a given length  $N$ . The  $i$ th window is defined as:

$$x_i(n) = \hat{S}_R(n + i(N/2))w(n), \quad (8)$$

where  $w(n)$  is a weighting function.

The frequency resolution can be approximated as the inverse of the duration of the window:

$$T_w = N/f_s, \quad (9)$$

where  $f_s$  is the sampling rate, and therefore only Doppler shifts that are greater than  $1/T_w$ , corresponding to velocities

$$v > c/2f_o T_w, \quad (10)$$

will be clearly visible [5]. From Eq. (10), it can be shown that radars that work in higher frequencies have the additional advantage to induce a wider micro-Doppler bandwidth where small movements are more easily detected for a given frequency resolution, because the carrier frequency is higher [5].

In this paper, an STFT has been used to generate time-frequency spectrograms. Fig. 1 shows a spectrogram of the radar signals of a human walking indoors that we collected. The spectrogram shows the evident periodic characteristics of a human activity. The fluctuation resulted by the limbs is attached to the main Doppler frequency resulted by the torso. The line around the zero frequency is mainly the DC component of the signal.

### B. The Low-power Radar System

The Doppler radar system built in this research consists of two BumbleBee radars from Samraksh [31]. The BumbleBee radar is a low-power Pulsed Doppler Radar that is designed

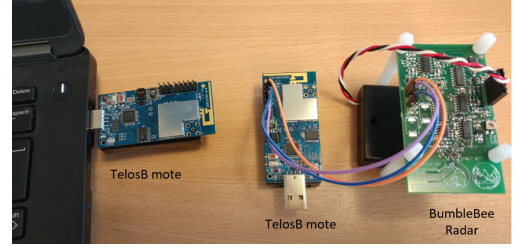


Figure 2. BumbleBee radar and TelosB mote

for a variety of Wireless Sensor Network (WSN) applications. Its center frequency is 5.8 GHz; and its detection range is up to 8 meters outdoors. Unlike traditional radars, the BumbleBee is designed to work with small, battery powered nodes. It only consumes about 12 mA, so when using typical 1.5v AA alkaline batteries with a capacity of 2400 mA, it can run at 100% duty cycle for about 8 days. Each BumbleBee radar outputs data on two channels providing the in-phase ( $I$ ) and quadrature-phase ( $Q$ ) signal components, which are used to form the complex signal  $C = I + jQ$ . The  $I/Q$  output data of the BumbleBee radar represents the peak of the matched filtered data acquired from each pulse. Thus, the time interval between each data packet corresponds to the pulse repetition interval (PRI) of the radar.

Each BumbleBee radar was connected to one TelosB mote [32] and another TelosB mote was used as a base station connected to a PC (see Fig. 2). The TelosB mote provides radio communication at low-power consumption (IEEE 802.15.4). It has a long battery life and it is able to wake up fast from a sleep state. It is fully compatible with TinyOS, an open-source operating system that supports large-scale, self-assembling sensor networks.

We used a multi-radar system composed of two radars. A multi-radar system collects signals of human activity from multiple angles, this provides more information for the classifiers. However, it also means more data is required to be transferred concurrently, consequently packet loss in data transmission increases in detriment to the gains in information gathered. We found that a system consisting of one base station and two radars provided an increased gain in information gathering without a loss in data transmission, but for a higher number of radars the transmission data loss was noticeable.

The Doppler radar system was implemented outdoors. As shown in Fig. 3, two radars (BumbleBee 1, BumbleBee 2) were placed in a straight line, eight meters apart and opposite to each other. One radar was called the *primary node* (BumbleBee 1), the other was called the *secondary node* (BumbleBee 2). There was a *base station* (Telosb 03) connected to a laptop. The base station received data from the two Telosb motes (Telosb 01 and Telosb 02). Each mote was connected to one BumbleBee radar. The shadowed areas were the detection ranges of BumbleBee 1 and BumbleBee 2. The cross width of the detection ranges was from 10m to 12m. Only targets inside the detection ranges could be observed. It can be noticed that the union of two radars' detection ranges was split into three non-overlapping ranges, which were 1-3m, 3-5m, and

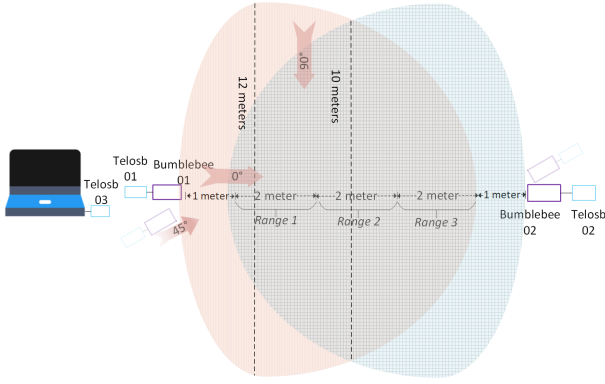


Figure 3. Doppler radar system implemented in outdoors

5-7m relative to the primary node. The experiments were made in these three different ranges in order to tag the range labels to the micro-Doppler signatures, which were used for coarse localization. The arrows represent the target movement direction relative to the radar beam. In order to investigate the effects of the angle between the direction of movement and the radar, three different directions ( $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ ) were considered. When the experiments were performed with the radar at  $45^\circ$  in relation to the target, the target moved on the ground in the same way as it did for the  $90^\circ$ , however we rotated each radar to a position where their beams were at  $45^\circ$  in relation to the target movement.

### III. EXPERIMENTS

In order to investigate the effectiveness of the human activity classification outdoors using our system, four different tasks were processed in the experiments. They were subject classification, human activity classification, people counting, and coarse localization. As illustrated in Fig. 4, the output metric for subject classification is whether the target is human or animal, or there is no target. If the target is human, then the system further estimates whether the target is running or walking, how many people the target represents, and the rough range that the target is located in.

Three different experiments were performed in three different outdoor locations respectively. The experimental locations were populated with trees and shrubs; they were realistic wild areas. These experiments are described in detail below:

*Case 1: Single Person Activity Classification:* Three individuals participated in this experiment. The activities of walking and running were performed from three different angles ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ) relative to the radar beam. The same set of experiments was performed by each participant one at a time. At  $45^\circ$  and  $90^\circ$ , activities were performed in three ranges (1-3m, 3-5m, and 5-7m) relative to the primary node.

Fig. 5 illustrates the micro-Doppler signatures of an individual participant walking at  $0^\circ$ ,  $45^\circ$  and  $90^\circ$ . The spectrograms were generated by a STFT with a sliding window size of 20s, which contains 5000 signal samples (250Hz sample rate). Comparing Fig. 5(a) and Fig. 5(b), the motion cycle for running is composed of three cycles and for walking only two cycles; this confirms the physical fact that running takes

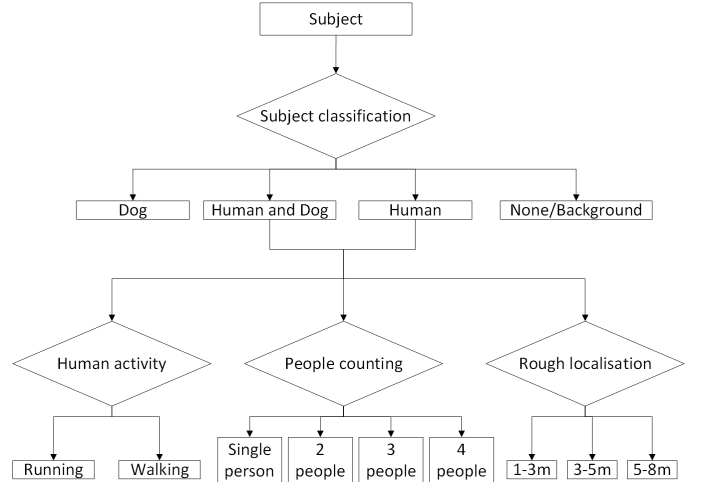


Figure 4. Workflow of four tasks, including subject classification, human activity classification, people counting, and coarse localization

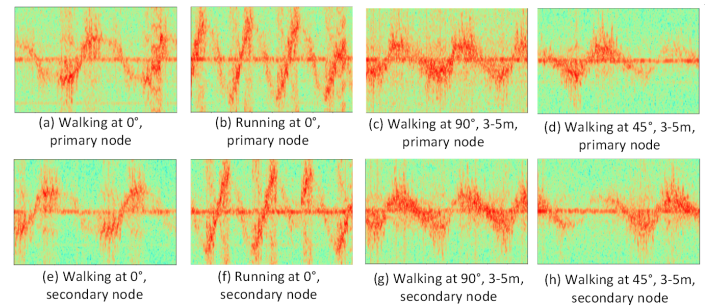


Figure 5. Micro-Doppler signatures of an individual at different angles

less time to finish a motion cycle than walking. Further, it is interesting to note that the wave directions of Fig. 5(a) and Fig. 5(b) are opposite to the wave directions of Fig. 5(e) and Fig. 5(f), this is because the directions of the primary node and the secondary node are opposite to each other. When a person moves at  $0^\circ$ , he/she is moving towards one radar and moving away from the other radar, consequently the frequencies of the echoed signals of the two radars fluctuate in opposite directions. When a person moves at  $90^\circ$  and  $45^\circ$ , he/she gets close to or further away from both primary and second radars, so the frequencies of the echoed signals of the two radars fluctuate in the same direction as shown in Fig. 5(c) and Fig. 5(g), and Fig. 5(d) and Fig. 5(h) respectively. Fig. 5(d) and Fig. 5(h) present more faded cycle segments, but in different positions. This is because when the participant is walking at  $45^\circ$ , he/she is at the edge of or outside the detection range of each radar for a small period of time. In the spectrograms, the change in color intensity results from the changes in the radar cross-sections (RCS). The RCS is the measure of a target's ability to reflect the radar signals in the direction of the radar's receiver, i.e. It is a measure of the ratio between the backscatter density in the direction of the radar (from the target) and the power density that is intercepted by the target. Larger RCS indicates a greater energy is reflected by the target, and it produces more intensive color in the spectrograms.

The spectrograms in Fig. 6 represent the micro-Doppler



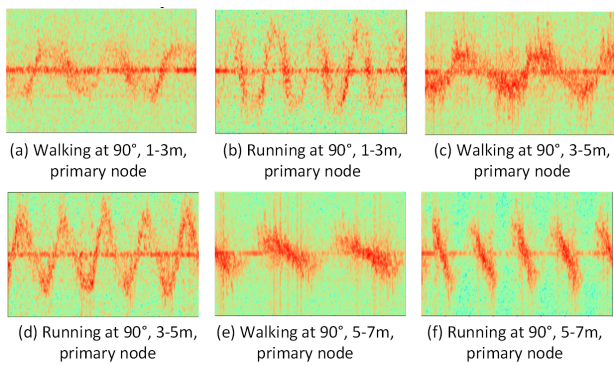


Figure 6. Micro-Doppler signatures of an individual walking and running in different ranges at  $90^\circ$

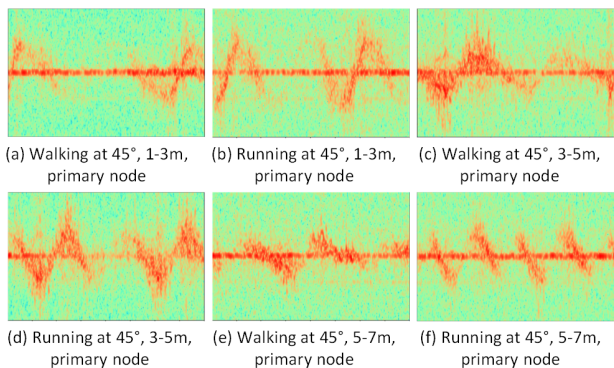


Figure 7. Micro-Doppler signatures of individual walking and running at  $45^\circ$  with three different ranges

signatures of an individual walking and running at  $90^\circ$  in each one of the three distance ranges. The patterns of the signal are different depending on the target's distances to the primary node, indicating the distance indeed affects the spectrogram pattern. The spectrograms of the secondary node are not shown, because the patterns collected by the secondary node are similar to the primary node in each correspondent range. The patterns of the spectrograms of an individual walking and running at  $45^\circ$  in the three different ranges are illustrated in Fig. 7. They are similar to those in Fig. 6, except that in a section of each spectrogram the wave form is less clear because the target is outside the detection range.

*Case 2: Activity classification and people counting in a group of people:* Nine individuals participated in this experiment. Fig. 8 shows snapshots of the experiments performed. Participants were arranged into three groups. The first group had two participants, the second group had three participants, and the third group consisted of four participants. Each group walked and ran in three ranges from three different angles ( $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ ) relative to the radar beam. Participants in the first or second group moved abreast. Participants in the third group were divided into two rows with two people in each row and they ran or walked simultaneously inside the detection range. The main difference between *Case 1* and *Case 2* is the number of people that composed the target. Fig. 9 illustrates the spectrograms of the target with different numbers of people. It is difficult to find visual differences



Figure 8. Snapshots of case 2 and case 3 experiments

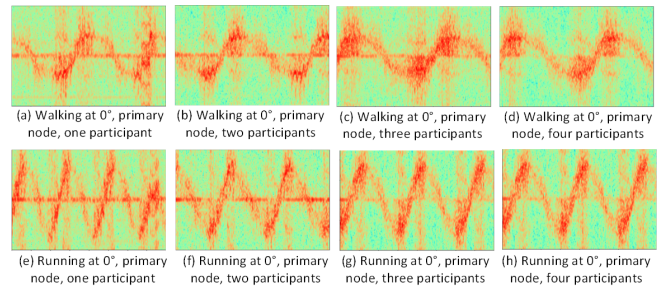


Figure 9. Micro-Doppler signatures of different numbers of people

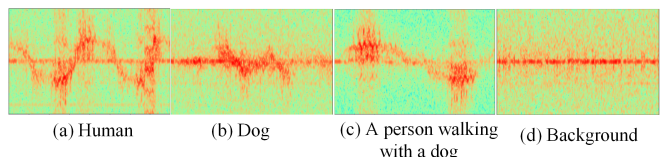


Figure 10. Micro-Doppler signatures of different subjects

among the spectrograms, however as we will present later, we obtained very good results for the micro-Doppler signatures classification used for people counting.

*Case 3, differentiation between humans and dogs:* One human volunteer and two dogs participated in this experiment. As it is hard to control a dog's speed (i.e. walk and run) inside the detection range, the experiments did not differentiate between walking and running, and only two angles ( $0^\circ$  and  $90^\circ$ ) were investigated. In the first experiment, separately, each dog was encouraged by their owners to move alone back and forth inside the detection range. In the second experiment, the volunteer walked a dog back and forth, as shown in Fig. 8. Unavoidably, some noisy data was produced during the experiments, but the amount was small.

Combined with the data collected in *Case 1* and *Case 2*, all samples could be classified into 4 categories according to the different types of the targets, including human, dog, a person and a dog, and the background. Examples of spectrograms generated by these different targets are illustrated in Fig. 10.

It is important to note that we selected the clearest spectrograms to illustrate the different micro-Doppler signatures. However, as shown in Fig. 11, many spectrograms are very noisy or vague. The uncontrolled movement from trees, small plants and possible small animals in the radar's detection range creates noises that increase the difficulty of micro-Doppler signature classification outdoors when compared to a more stable environment.

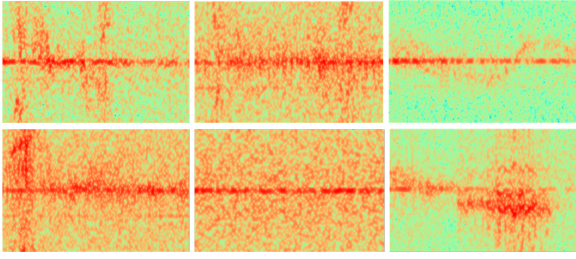


Figure 11. Spectrograms generated from outdoor environments

Table I  
FIELDS OF THE SIGNAL COLLECTION

Field	Information
ID	The identity number of a Doppler radar signal
Q	The quadrature power value of a signal
I	The in-phase power value of a signal
Time	The time a signal was collected (accurate to a millisecond)
SubjectType	The type of a target or targets. '0' is human, '1' is dog, '2' is human and dog, '3' is no target.
SubjectID	The ID of the targets.
SubjectNum	The number of participants as a target.
Activity	The type of human activity. '0' is walking, '1' is running, '2' is no human activity
Angle	The angle between the direction of human movement and the radar beam.
Range	The range of the target located in. '0' is 1-3m, '1' is 3-5m, '2' is 5-8m, '3' is not in any range.
RadarNum	The number of radars used in the system set up: 2.
NodeID	The ID of primary node that consists of BumbleBee 1 and Telsob 1.
SecNodeID	The ID of secondary node that consists of BumbleBee 2 and Telsob 2.

#### IV. DATA COLLECTION AND PREPROCESSING

The data collected in the experiments requires further processing before feature extraction and classification are made.

##### A. Data collection

With the Doppler radar system described in Section II, signals collected by the BumbleBee radars were transferred to the TelosB nodes and radio transferred to the TelosB base station. The TelosB base station was connected to a computer which received the signals through a serial port. Finally, all data was stored into a MongoDB database, which is an open source NoSQL database and it is usually utilized to store a large amount of time series data.

The final data stored into the database contains not only radar signals, but also labels that informed the time, the types of the target and the types of human activities, etc. Table I shows the information of each field of a radar signal recorded in our MongoDB system. The sample frequency of the radars is 250 Hz. The data collected in the experiments were used to train the classifiers. In prediction, only the ID of the signals, I and Q values, and the time were used.

##### B. Data preprocessing

The original signal is a complex value ( $I + jQ$ ). Micro-Doppler signatures are represented in a time-frequency domain. It is required to transform the original radar signals from the time-amplitude domain into the time-frequency domain

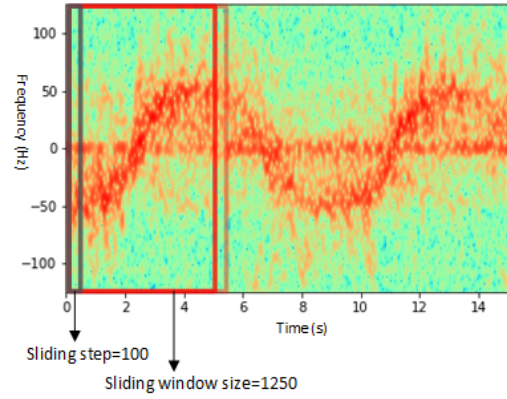


Figure 12. The use of a sliding window for STFT

Table II  
THE COMPOSITION OF SAMPLES

Subject	Angle	Walking	Running	Total
Single person	0	1150	1200	15400
	45	2700	3150	
	90	3500	3700	
Two people	0	550	550	7000
	45	1600	1550	
	90	1150	1600	
Three people	0	600	550	7200
	45	1250	1700	
	90	1300	1800	
Four people	0	500	550	6950
	45	1200	1650	
	90	1150	1900	
Dog	90		1350	1350
Human and Dog	0		1200	3150
	90		1950	
Background				5850

using STFT. A windowed STFT processing technique with a window length of 64 samples and a Hanning weighting, transforms a sequence of 1250 signal samples into a frequency spectrogram with the size of  $2048 \times 148 \times 1$  (2048 is the height, 148 is the width, 1 is the depth). The frequency spectrogram can be taken as an image with one channel. Fig. 12 presents a frequency spectrogram that was created from a sliding-window with the window size of 1250 frames (1250 samples within the sliding window) and the sliding step with the length of 100 frames. The sampling frequency of the micro-Doppler signals is 250Hz. So by moving the sliding-window with continuous sliding steps, a spectrogram can be extracted at every 0.4 seconds, which is calculated by  $(Sliding\ step\ length)/(sample\ frequency)$ , i.e.  $100/250$ . The effect of the window size is investigated in Section VIII.

Through the above processing by using an STFT sliding window of 1250 frames (time length of 5s), the described experiments generated 46900 spectrograms in total. The composition of the samples/spectrograms is shown in Table II.

For the classification, the spectrogram samples were separated into two groups, 80% of the samples were used for training and validation, and 20% for testing. Note in Table II that the number of individual samples observed for each class differs. In order to overcome this problem, different weights were assigned to each class in the training process; and

the weight ratio was inversely proportional to the proportion of spectrograms of the various classes. However, for subject classification, because the numbers of spectrograms of the human and the background are far greater than the number of spectrograms of the dog (including human and dog), we randomly selected 2000 spectrograms of human and 2000 spectrograms of the background for the training.

## V. FEATURE EXTRACTION USING TWO-DIRECTIONAL TWO-DIMENSIONAL PRINCIPAL COMPONENT ANALYSIS

Feature extraction in micro-Doppler analysis is used to reduce the number of features of a spectrogram; the aim is to identify features of the signal, which are required for recognizing an activity, and to disregard other parts as background noise. A good feature extraction technique is an important component in a system before classification; because it will dictate the quality of the classification and the time that the classifiers will take to sufficiently train the model.

Two-directional Two-dimensional Principal Component Analysis was firstly developed by [33] and used for face representation and recognition. 2D2PCA can be regarded as a two-dimensional version of the PCA. PCA is well-known as a classic feature extraction and dimensionality reduction technique. A spectrogram is an image composed of a two-dimensional structure of pixels. 2D2PCA performs feature extraction on the rows and columns of an image simultaneously and it is more efficient in computing the covariance matrices, the eigenvalues and the eigenvectors than PCA alone.

A spectrogram generated by STFT can be denoted by a matrix  $A \in R^{m \times n}$ , where  $R$  indicates that the elements of  $A$  consist entirely of real numbers. Let  $X \in R^{n \times d}$ ,  $n \geq d$  be a projection matrix with orthonormal components.  $Y \in R^{m \times d}$  is generated by projecting  $A$  onto  $X$ , which can be written as  $Y = AX$ . The reduction of the dimensionality is achieved by the selection of a suitable value for  $d$ , which decides how many features will be kept in each row, also it is the final number of columns after the projection.

Matrix  $Y$  must preserve relevant information of the specific activity contained in the spectrogram. An ideal projection matrix should ensure that the result after the projection is very distinct from others of different activities. This makes the samples of different activities more independent to each other so that they do not cluster together. Therefore, it is beneficial to the classification if the most relevant information is kept after projection. In order to determine a good projection matrix  $X$ , the following criterion [33] is adopted:

$$\begin{aligned} J(X) &= \text{trace}\{E[(Y - EY)(Y - EY)^T]\} \\ &= \text{trace}\{E[(AX - E(AX))(AX - E(AX))^T]\} \quad (11) \\ &= \text{trace}\{X^T E[(A - EA)^T(A - EA)]X\}, \end{aligned}$$

where  $J(\cdot)$  is an objective function.  $J$  must be maximized in order to find the optimal projection matrix  $X$ .  $E$  is the expectation operator that when applied to a matrix produces a new matrix containing the expected values of the elements of the original matrix. The last term in Eq. (11) follows the commutative property of matrices where  $\text{trace}(QP) = \text{trace}(PQ)$ ,  $Q$  and  $P$  represent any two matrices.

The covariance matrix [33] of  $A$  is defined as:

$$G = E[(A - EA)^T(A - EA)], \quad (12)$$

Suppose that the training set consists of  $M$  spectrograms  $\{A_1, A_k, \dots, A_M\}$ , the covariance matrix  $G$  can be computed as:

$$G = (1/M) \sum_{k=1}^M (A_k - \bar{A})^T (A_k - \bar{A}), \quad (13)$$

where  $\bar{A}$  is the average spectrograms as  $\bar{A} = (1/M) \sum_{k=1}^M A_k$ . Eq. (11) can be simplified as:

$$J(X) = X^T G X, \quad (14)$$

Let  $A_k = [(A_k^{(1)})^T (A_k^{(2)})^T \dots (A_k^{(m)})^T]^T$  and  $\bar{A} = [(\bar{A}^{(1)})^T (\bar{A}^{(2)})^T \dots (\bar{A}^{(m)})^T]^T$ , where  $A_k^{(i)}$  and  $\bar{A}^{(i)}$  denote the  $i_{th}$  row of  $A_k$  and the  $i_{th}$  row of  $\bar{A}$ , respectively. The covariance matrix  $G_r$  can be rewritten as:

$$G_r = (1/M) \sum_{k=1}^M \sum_{i=1}^m (A_k^{(i)} - \bar{A}^{(i)})^T (A_k^{(i)} - \bar{A}^{(i)}), \quad (15)$$

The covariance matrix  $G_r$  is essentially the same as  $G$ , the only difference is that matrix  $A$  is taken as a set of row vectors. Through diagonalization of covariance matrix  $G_r$ , the ideal projection matrix  $X$  can be obtained.

In the same way, matrix  $A$  also could be taken as a set of column vectors. Let  $A_k = [(A_k^{(1)}) (A_k^{(2)}) \dots (A_k^{(n)})]$  and  $\bar{A} = [(\bar{A}^{(1)}) (\bar{A}^{(2)}) \dots (\bar{A}^{(n)})]$ , where  $A_k^{(j)}$  and  $\bar{A}^{(j)}$  denote the  $j_{th}$  column of  $A_k$  and the  $j_{th}$  column of  $\bar{A}$ , respectively. The covariance matrix  $G_c$  could be rewritten as:

$$G_c = (1/M) \sum_{k=1}^M \sum_{j=1}^n (A_k^{(j)} - \bar{A}^{(j)}) (A_k^{(j)} - \bar{A}^{(j)})^T, \quad (16)$$

Let  $Z \in R^{m \times q}$  be a matrix with orthonormal columns. Projecting matrix  $A$  onto  $Z$  yields matrix  $B = Z^T A$ ,  $B \in R^{q \times n}$ . The reduction of the dimensionality is achieved by the selection of a suitable value for  $q$ , which determines how many features will be kept in each column, it also corresponds to the number of rows of  $B$ .  $Z$  can be obtained through diagonalization of  $G_c$ .

After the projection matrices  $X$  and  $Z$  are obtained,  $A$  must be projected onto  $X$  and  $Z$  simultaneously in order to yield a matrix  $C$ :

$$C = Z^T A X, \quad (17)$$

the matrix  $C$  is called the *coefficient matrix*. It can be taken as the input features that are fed into SVM and  $k$ NN classifiers.

Fig. 13(a) is a spectrogram whose size is  $2048 \times 616$  pixels. It is the result of an echoed radar signal of a human walking indoors. The method 2D2PCA was applied on this original spectrogram for different row and column dimension reductions. The results are shown in Fig. 13(b) and Fig. 13(c), where they still retain detailed time-frequency characteristics (the periodic waves), but the image dimensions are well reduced, visually they are almost the same as the original spectrogram. Even when reducing the dimensionality to 8 rows and 8 columns the remaining characteristics can still reflect the periodic trend of a human walking as it can be seen in Fig.



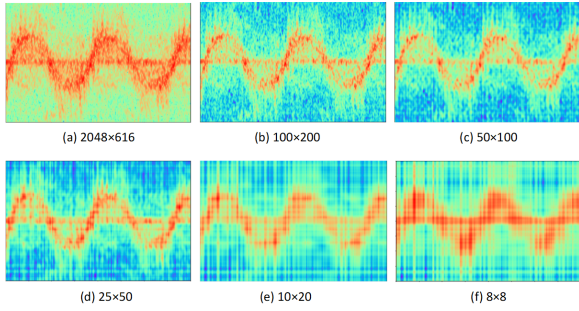


Figure 13. Dimensionality reduction with 2D2PCA

13(f). 2D2PCA is able to keep most of the distinctive features (pixels) while greatly reducing the dimensions of the micro-Doppler signatures.

## VI. MACHINE LEARNING MODELS

Three classifiers, including CNN, SVM, and  $k$ NN were modeled in this work. All the three classifiers were modeled to classify the samples for different classification outputs. This section details how the classifiers were built and what hyper-parameters were used and optimized.

### A. Convolutional neural network (CNN)

Deep learning has been applied to many research areas and has achieved quite remarkable results. CNNs are widely used in image recognition and classification due to their ability of automatic learning hierarchical representations directly from the raw data input. In this research, each spectrogram can be considered as an image with one channel. So a CNN is suitable to be applied for subject and activity classification using spectrograms.

Generally, a CNN contains three typical layers: *Convolutional Layers*, *Pooling Layers*, and *Fully-Connected Layers*. A convolutional layer allows the network to detect spatial patterns over different parts of the input, and a pooling layer to learn the translational invariance of the input. Fully-connected layers are placed after several convolutional and pooling layers to perform high-level reasoning in the network. The unique architectural configuration of a CNN is defined by its *hyperparameters*, whose values are set before the network is trained. An instance of a convolutional network is defined by its parameters whose values are learned during the network training [34].

An *activation Function* is not a typical function in CNNs, but it is a nonlinearity operation commonly used in neural networks. It increases the nonlinear properties of CNNs. As the result of a series of linear operations (like convolutions) can be a single-linear operation, an elementwise nonlinear function can be applied between convolutions in order to introduce nonlinearity [34]. This makes CNNs capable to learn more complex mappings of input to output. The activation function used in our work is the ReLU (Rectified Linear Units) [35], which is formulated as:

$$f(x) = \text{Max}(0, x), \quad (18)$$

ReLU is often preferred over other functions because it trains the neural network several times faster without a significant penalty to the generalization accuracy.

In the training process, a CNN model is trained to minimize an objective function in terms of the parameters of the network. For a given classification, let  $C$  be the number of labeled classes, the following cross-entropy loss function [36] is often used:

$$E_y(y') = - \sum_{i=1}^N y_i \cdot \log(y'_i) \quad (19)$$

where  $E$  is the loss function evaluated over  $N$  samples,  $y_i$  is the original label of the  $i_{th}$  sample and  $y'_i$  is the class score maps of a sample  $i$  calculated using a *softmax* activation function [37]:

$$y_j = \exp(x_j) / \left( \sum_{c=1}^C \exp(x_c) \right) \quad (20)$$

where  $y$  is the softmax score and  $x$  is the output layer containing unnormalized class scores.

Fundamental details of CNN explaining concepts of convolutional layers, pooling layers, fully connected layers, activation function and regularization can be found in [38].

*RadarNet – Our Proposed CNN*: Our radar system consists of two BumbleBee radars. In order to fuse the signals from the radars, the spectrograms are firstly downsampled into the size of  $50 \times 50 \times 1$ , then overlapped together. The downsampling of the spectrograms is beneficial to accelerate the training phrase by reducing the number of parameters since the size of  $2048 \times 148 \times 1$  is too large. A resulting overlapped spectrogram can be considered as an image with two channels. As shown in Fig. 14, the overlapped spectrogram with the shape of  $50 \times 50 \times 2$  is input into our RadarNet.

After investigating several different CNN designs for RadarNet, the complete CNN architecture that we propose, which provides the best performance results is described below and shown in Fig. 14.

RadarNet contains three convolutional layers (C1, C2, and C3), two Max Pooling layers (M1, M2) and two fully connected layers. All three convolutional layers use  $3 \times 3$  kernels to do the convolutionalization. Each feature map is generated by one filter. The C1 layer contains 16 feature maps, and the size of each feature map is  $48 \times 48 \times 1$ . A  $2 \times 2$  filter is used to perform the *Max Pooling* on C1, and to generate the M1 layer. The C2 layer contains 32 feature maps, and the size of each feature map is  $22 \times 22 \times 1$ . The resulting M2 layer by Max-Pool the C2 layer has the size of  $11 \times 11 \times 1$ . The C3 layer contains 48 feature maps, each feature map is a  $9 \times 9 \times 1$  matrix. The F1 layer contains 356 hidden units, and the F2 contains 160 hidden units.

In RadarNet, dropout has been used to control overfitting with an initial dropout rate of 0.4. The term ‘dropout’ refers to dropping out a part of the units in a neural network. By avoiding training all nodes on all training data, dropout decreases overfitting. Batch normalization [39] has been applied on M1, M2, F1, and F2 as a regularizer to accelerate convergence. Batch normalization normalizes the output of a

Table III  
OPTIMAL HYPERPARAMETERS FOR DIFFERENT CLASSIFICATION TASKS

Task	Model	Gamma	C
Subject classification	SVM	1.0E -3.57	7
	SVM+2D2PCA	1.0E -5.84	3.1
Human activity classification	SVM	1.0E -3.35	7.34
	SVM+2D2PCA	1.0E -5.67	5.37
People counting	SVM	1.0E -3.26	6
	SVM+2D2PCA	1.0E -6	6.5
Coarse localization	SVM	1.0E -3.57	7.9
	SVM+2D2PCA	1.0E -6.07	2.9

previous activation layer by subtracting the batch mean and dividing by the batch standard deviation. The use of dropout and batch normalization reduces overfitting and accelerates the training. The optimization function applied is *Adadelta*, whose initial learning rate is 0.1. *Adadelta* is an optimization function that can dynamically adapt over time using only first order information and it has a minimal computational overhead beyond standard stochastic gradient descent, which is one of the most popular methods used to perform optimization. *Adadelta* requires no manual tuning of a learning rate and appears robust to noisy gradient information, different model architecture choices, various data modalities, and selection of hyperparameters [40].

### B. Support vector machine (SVM)

SVM is one of the most used classifiers for micro-Doppler based human activity classification. SVMs are discrete algorithms that can be used to find the optimal separating hyperplane that maximizes the margin of the training data. A hyperplane is a decision boundary to separate data of different classes. As a dataset is usually non-linear separable in practice, the SVM algorithm is implemented using a kernel, which is a way of transforming the input data into a high-dimensional space. In the transformed feature space, it is possible to separate the data with a linear hyperplane. An SVM with RBF (Radial Basis Function) kernel has been applied because a Linear SVM cannot separate the micro-Doppler data successfully.

Differently from the works in [3], [14], [15], [21] where the features fed into SVMs were handcrafted, the input features used in our work were extracted automatically in two different ways. One way was to calculate the mean value of each column in each spectrogram and define it as a feature. Another way was applying 2D2PCA on the samples. The results of these two feature extraction methods are compared later.

Two hyperparameters ( $C$ ,  $\gamma$ ) in RBF SVM were specified manually. Cross-validation was used to tune the hyperparameters. Given a hyperparameter space  $C : [1, 30]$ ,  $\gamma : [0.1, 1.0E - 5]$ , a different pair of parameters was selected from the hyperparameter space by the cross-validation in each training and validation iteration in order to build the SVM model. The parameters that make the SVM to perform the best are the optimal parameters. Table III presents the optimal parameters for different classification tasks.

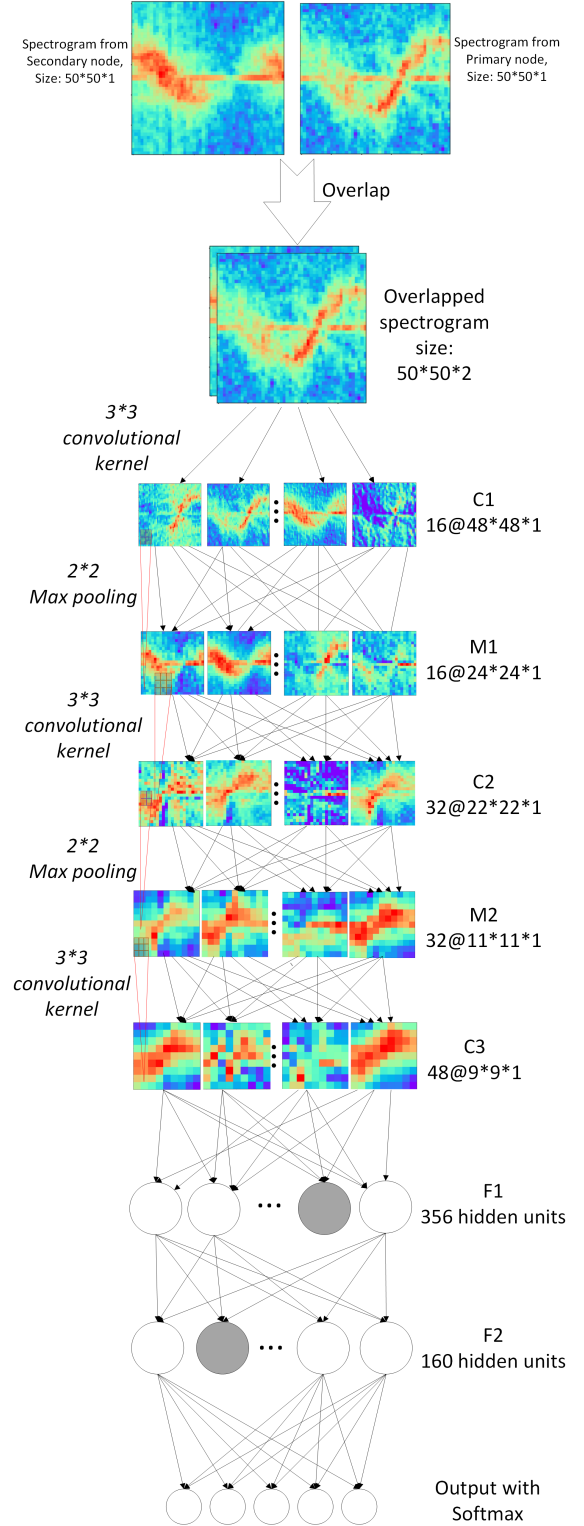


Figure 14. RadarNet structure for human activity classification

### C. $k$ -Nearest-Neighbors ( $k$ NN)

The  $k$ -Nearest-Neighbors is one of the most fundamental and simple classification methods and should be one of the first choices for a classification study when there is little or no prior knowledge about the distribution of the data [41]. It also has been widely used in micro-Doppler based human



Table IV  
THE VALUE  $k$  OF  $k$ NN IN DIFFERENT TASK

TASK	MODEL	K
Subject classification	kNN	1
	kNN+2D2PCA	1
Human activity classification	KNN	9
	kNN+2D2PCA	5
People counting	KNN	4
	kNN+2D2PCA	1
Coarse localisation	KNN	7
	kNN+2D2PCA	1

activity classification and detection [14], [42], [43].

In human activity classification, a set of features extracted from frequency spectrograms of a micro-Doppler radar can be represented by  $\{f_j\}_M^N$ , ( $1 \leq j \leq M$ ). Where  $N$  is the number of labeled samples (frequency spectrograms),  $M$  is the number of features in each sample,  $f_j$  is the  $j$ th feature, and an unlabeled sample can be represented  $S_i = \{f_j^i\}_M$ , ( $1 \leq j \leq M$ ). In order to find  $k$  closest samples to  $S_i$ , it is necessary to calculate the distance between each labeled sample  $S_c$ , ( $1 \leq c \leq N$ ) and  $S_i$ . The Euclidean distance is one option for such calculation:

$$\begin{aligned} Dist(S_c, S_i) &= Dist((f_1^c, \dots, f_M^c), (f_1^i, \dots, f_M^i)) \\ &= \sqrt{\sum_{p=1}^M (f_p^c - f_p^i)^2} \end{aligned} \quad (21)$$

Cross-validation is used to select a value of  $k$  that minimizes the overall distance between the  $k$  nearest labeled samples and the unlabeled sample. Finally, the unlabeled sample will be classified to the class label based upon a majority vote from the  $k$  nearest labeled samples. For each classification task, the chosen value of  $k$  is presented in Table IV.

## VII. RESULTS ANALYSIS

The models built in Section VI were used to classify the spectrograms. We compared the performance of the classifiers using different assessment methods: 1) overall classification accuracy (OA); 2) average class accuracy (AA); 3) Recall; and 4) F1 score. Details on these performance assessment methods can be found in [44].

### A. Results for subject classification

The aim of the subject classification is to differentiate human from other targets based on their different micro-Doppler signatures. Animals are the most common confusers. In this paper, a dog was used to represent four-legged animals. According to the type of the targets, the samples were divided into four categories (e.g. human, dog, human and dog, and background/none). The performances of CNN, SVM, and  $k$ NN are shown in Table V. RadarNet achieved the highest performance in all different metrics, with 97.5% in OA, 96.69% in AA, 97.23% in Recall, and 97.22% in F1. It outperforms SVM+2D2PCA and  $k$ NN+2D2PCA with 1%-7% in OA and AA, 0.8%-6.2% in Recall and F1. SVM and  $k$ NN had the poorest performance of around 72%-75%.

Table V  
PERFORMANCE FOR HUMAN RECOGNITION

MODEL	OA	AA	RECALL	F1
SVM	74.4%	72.35%	75.52%	74.76%
SVM+2D2PCA	96.5%	95.69%	96.4%	96.4%
KNN	66.2%	64.60%	66.23%	65.54%
KNN+2D2PCA	91.1%	89.69%	91.41%	91.24%
CNN (RadarNet)	97.5%	96.69%	97.23%	97.22%

Table VI  
PERFORMANCE FOR HUMAN ACTIVITY CLASSIFICATION

MODEL	OA	AA	RECALL	F1
SVM	86.6%	87.6%	85.8%	88.13%
SVM+2D2PCA	99.4%	99.33%	99.3%	99.38%
KNN	83.3%	83.16%	83.16%	84.3%
KNN+2D2PCA	98.35%	98.12%	97.2%	98.2%
CNN (RadarNet)	99.89%	99.89%	99.89%	99.89%

### B. Results for human activity classification

Two types of human activities (walking and running) were performed in the experiments. The ‘human’ samples can be divided into two further classes. The performance of the classifiers is shown in Table VI. The best performance of 99.89% in OA, AA, Recall, and F1 was achieved by our RadarNet, and SVM+2D2PCA followed it closely around 99.4%.  $k$ NN+2D2PCA also achieved a good result from 97.2% to 98.35%. The above three classifiers outperformed SVM and  $k$ NN by a wide margin again.

### C. Results for people counting

In *Case 1*, the micro-Doppler signatures of a single person were collected. In *Case 2*, samples for groups of two, three, and four people were collected by our radar system. For people counting, the ‘human’ samples can be further divided into four classes according to the number of people. The performance of the classifiers in people counting is shown in Table VII. RadarNet still achieved the best performance results around 98.85%, followed by SVM+2D2PCA around 95.8%.  $k$ NN+2D2PCA had 12.5% lower performance than SVM+2D2PCA. SVM and  $k$ NN had the poorest performance.

### D. Results for coarse localization of human targets

The detection area was split into three non-overlapping ranges, which were 1-3m, 3-5m, and 5-7m relative to the primary node. The coarse localization estimates which range the location of the human target belongs to. With these three ranges, the samples of the human target were divided into three categories. The performance of the classifiers for coarse localization is shown in Table VIII. The overall performance

Table VII  
PERFORMANCE FOR PEOPLE COUNTING

MODEL	OA	AA	RECALL	F1
SVM	65.3%	58%	65.3%	63%
SVM+2D2PCA	95.9%	95.7%	95.9%	95.9%
KNN	60.46%	52.6%	61%	58.4%
KNN+2D2PCA	83.3%	83.4%	81.88%	83.4%
CNN (RadarNet)	98.85%	98%	98.85%	98.7%

Table VIII  
PERFORMANCE FOR COARSE LOCALIZATION

MODEL	OA	AA	RECALL	F1
SVM	88.2%	87.25%	88%	88%
SVM+2D2PCA	99.9%	99.87%	99.87%	99.87%
KNN	81.8%	81.42%	81.08%	81%
KNN+2D2PCA	99.19%	99.24%	99.19%	99.19%
CNN (RadarNet)	100%	100%	100%	100%

achieved in coarse localization is higher than other three classification tasks. The CNN perfectly estimated the range where the target was located in. SVM+2D2PCA presented slight inferior results, with 99.9% in OA, and 99.87% in AA, Recall, and F1.  $k$ NN+2D2PCA also performed very well with the lowest accuracy metric achieving 99.19%. SVM underperformed SVM+2D2PCA by around 12% in the different accuracy metrics. Finally,  $k$ NN presented the lowest scores, achieving a percentage of around 81%.

Important to note that even though other types of radars may give more accurate range resolutions, we demonstrated that it is possible to obtain a coarse estimate of location with a very simple pulse radar system solely by using micro-Doppler signature classification. This is novel and needs to be further explored.

#### E. Analysis

Based on the results above, Fig. 15 presents a comparative performance graph for the different classification tasks. In all four tasks, the highest performance (OA) of the classifiers was given by RadarNet, followed by the SVM+2D2PCA, then  $k$ NN+2D2PCA. The classifiers with the lowest accuracy results were SVM and  $k$ NN. As one of the most popular deep learning algorithms, CNN has proved to be very suitable for micro-Doppler signature-based classification. RadarNet achieved the best OA scores, which were 97.5% in subject classification, 99.89% in human activity classification, 98.85% in people counting, and 100% in coarse localization. SVM+2D2PCA followed closely and  $k$ NN+2D2PCA was slightly inferior to SVM+2D2PCA, but both exceeded SVM and  $k$ NN by a wide margin (11%-18%). This shows that the performance of SVM and  $k$ NN was improved greatly by applying 2D2PCA and makes SVM+2D2PCA a very cost-effective and time-efficient classifier option for micro-Doppler signatures.

From the aspect of classification tasks, different classification tasks had different levels of difficulty, which indicated whether samples in a classification task were easier or more difficult to be classified. In order of difficulty we had: *People Counting* > *Subject Classification* > *Human Activity Classification* > *Rough localization*. This can be seen in Fig. 15, the average performance of all classifiers for coarse localization was the highest compared to the average performance of all others; this was followed by human activity classification and subject classification, while the average results for people counting were the lowest. Fig. 16 shows the *confusion matrices* of RadarNet for the four classification tasks. All correct predictions are located in the diagonal of the table, which allows the visualization

of the performance of a classifier. In Fig. 16(a), the most misclassified samples were from the ‘Human and Dog’, and they were classified into ‘Human’. This implies that it is relatively difficult to differentiate the micro-Doppler signatures of these two classes. In Fig. 16(b), six samples of ‘Walking’ were classified into ‘Running’, this could be probably because different participants have different walking speeds, and fast-walking people and slow-running people generated more similar micro-Doppler signatures that is harder to differentiate. Fig. 16(c) presents 12 out of 13 misclassified samples of ‘1 person’ that were classified into ‘2 people’; 13 out of 14 samples of ‘2 people’ were misclassified as ‘1 person’, and 17 samples of ‘3 people’ were misclassified as ‘4 people’. This indicates that adjacent categories are more likely to be misclassified between each other for the case of people counting. This is because there are more similarities in the micro-Doppler signatures between adjacent categories of people counting.

#### F. Comparison with the related work

There is some other similar work that is worth comparing with our study. In human and animal classification, the authors in [45] investigated animal confusers, they aimed to differentiate the categories of human walking and horse walking in an outdoor scenario. They used a Doppler radar operating at 9.2 GHz, which works in a higher frequency than the Bumblebee radar and consumes more energy. They achieved an OA of 92.7% in classification between humans and horses using SVM. Although our confuser was a different animal, we achieved a better result. This indicates that the performance of the human detection using micro-Doppler signatures depends on both the classifier and the radars. Although the radar used in [45] has better frequency and distance resolutions than the Bumblebee radar, a well modeled CNN like our RadarNet is able to compensate for these limitations.

In human activity classification, the authors in [21] investigated three motions (crawling, walking, and jogging) in four different environments, including (a) free space, (b) through-the-wall, (c) leaf tree foliage, and (4) needle tree foliage. They made their experiments using a continuous-wave Doppler radar operating at 6.5 GHz. They implemented an SVM classifier. The best classification results (OA of 91%) were obtained from the experiments in free space, followed by the experiments in leaf tree foliage and in through-the-wall. The lowest classification rate was from the experiments in needle tree foliage of around 71%. In our work, we only considered the classification of two activities (walking and running), but we could argue that we performed the experiments in an environment comparable to the leaf tree foliage and the needle tree foliage with considerably better results. In [21], the authors used a BumbleBee radar to measure micro-Doppler signatures of three motions (walking, running, and crawling) in an indoor scenario. The classification was implemented using the  $k$ NN method. Their classifier correctly classified the activity of walking 90% of the cases, 88% for running, and 93% for crawling. Although the radar used in [21] is the same as ours, they made the experiments in less noisy conditions (indoor environment), but their results are still not comparable to the 99.8% OA that we achieved in our work.

For people counting, the authors in [28] applied a Ka-band Doppler radar to measure the micro-Doppler signatures of people outdoors. The stride rate over the peak period was extracted from spectrograms as an important feature to classify whether the target was an individual or a small group. With a  $k$ NN classification approach, they achieved an overall classification rate of 80%. The authors in [25] measured the simulated walking of subjects using a simulated CW radar. They varied the subjects from no subject to a group of five people. The classifications were performed using a DCNN on the simulated data and achieved a high accuracy of 96.1% overall. However, real environments are far more complex than simulations, therefore our 98.85% overall accuracy for people counting is a good performance.

For micro-Doppler based localization, to the best of our knowledge, there is still no relevant comparable work.

For the CNN model, it is difficult to make a direct comparison because there is no benchmark dataset publicly available in radar activity detection. The accuracy performance of the CNN model greatly depends on the collected data. Doppler radar samples differ depending on the radar's operating frequency, the radar's power, and the sampling frequency of the radar signals. All these dependencies prohibit a direct comparison of the CNNs on the data collected from different Doppler radars. Although a direct comparison is not possible, we describe some of the differences between our work and [25], [26] apart from the DCNN designs. In [25], it is clearly shown that the CNN results improve when the radar operating frequency and SNR increase on the synthetic data. The best results were obtained with the highest frequency tested (10 GHz). For the measured data, they used a radar with a better hardware specification and higher operating frequency than ours to measure micro-Doppler signatures of different numbers of people (from 0 to 3 people). The accuracy achieved by their DCNN was 86.9%. We performed people counting (from 1 to 4 people) and achieved an accuracy of 98.85%. In [26], the author used a CW radar operating at 7.25 GHz to perform human recognition outdoors at  $0^\circ$  and a CW radar operating at 2.4 GHz to perform human activity detection indoors at  $0^\circ$  both in a clear line of sight environment. They investigated subjects and activities different from ours. Their CNN obtained 97.6% overall accuracy in human target classification and we achieved 97.5% using a simpler radar system in a more cluttered outdoor environment.

Therefore, it is plausible to infer that the methods including the signal processing and the classifiers, which we investigated and tested in our research, are solid implementation tools to be used in micro-Doppler signature based human activity classification outdoors. The comparative performance is even better for human activity classification and people counting. Our work promotes the micro-Doppler based application in long-term continuous human activity detection outdoors.

### VIII. FACTORS IN MICRO-DOPPLER BASED CLASSIFICATION

In this section, three factors, including the frame length of the sliding window, the angle of the movement, and the

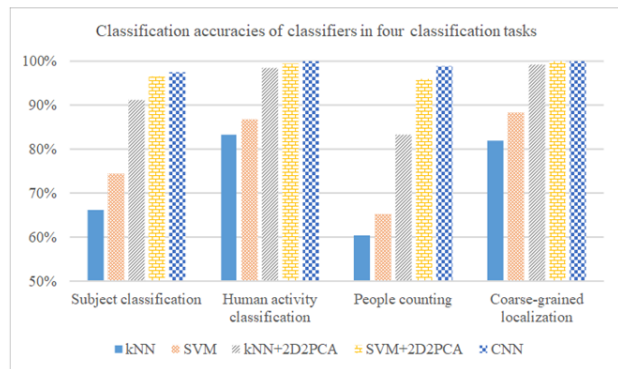


Figure 15. The performance of the five classifiers for four classification tasks

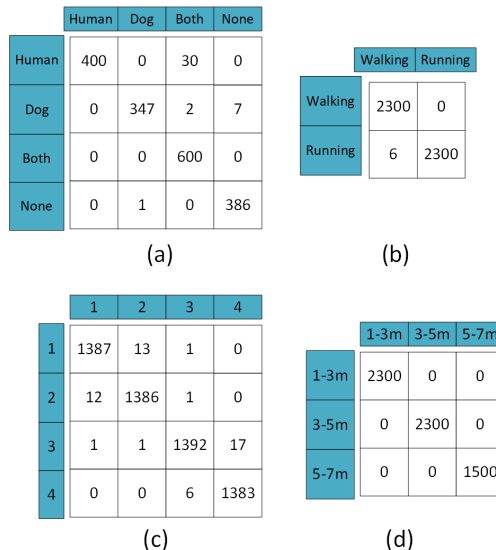


Figure 16. Confusion matrices of CNNs in (a). Subject classification, (b). Activity classification, (c). People counting, (d). Coarse localization

number of radars were investigated in micro-Doppler based classification. It is helpful to assess the influence of these factors on micro-Doppler data processing and model optimization to make its analysis more practical.

#### A. The frame length of the sliding window

As mentioned in Section IV, the radar signals are processed with STFT, which uses a sliding window with a given frame length to generate spectrograms. It is worth studying how the different frame lengths affect the classification.

We investigated six frame lengths, including 750, 1000, 1250, 1500, 1750, and 2000 samples. As the sample frequency is 250 Hz, the frame length also can be measured in the time domain. Then the six frame lengths also can be measured as 3s, 4s, 5s, 6s, 7s, and 8s. The five classifiers were used to perform the classification with the generated spectrogram samples. As shown in Fig. 17, the performance (OA used) of almost all classifiers increase with the increasing frame length for all four classification tasks, although the amount of the increase declines for greater frame lengths. The superiority of the classifiers performance remained the same as stated

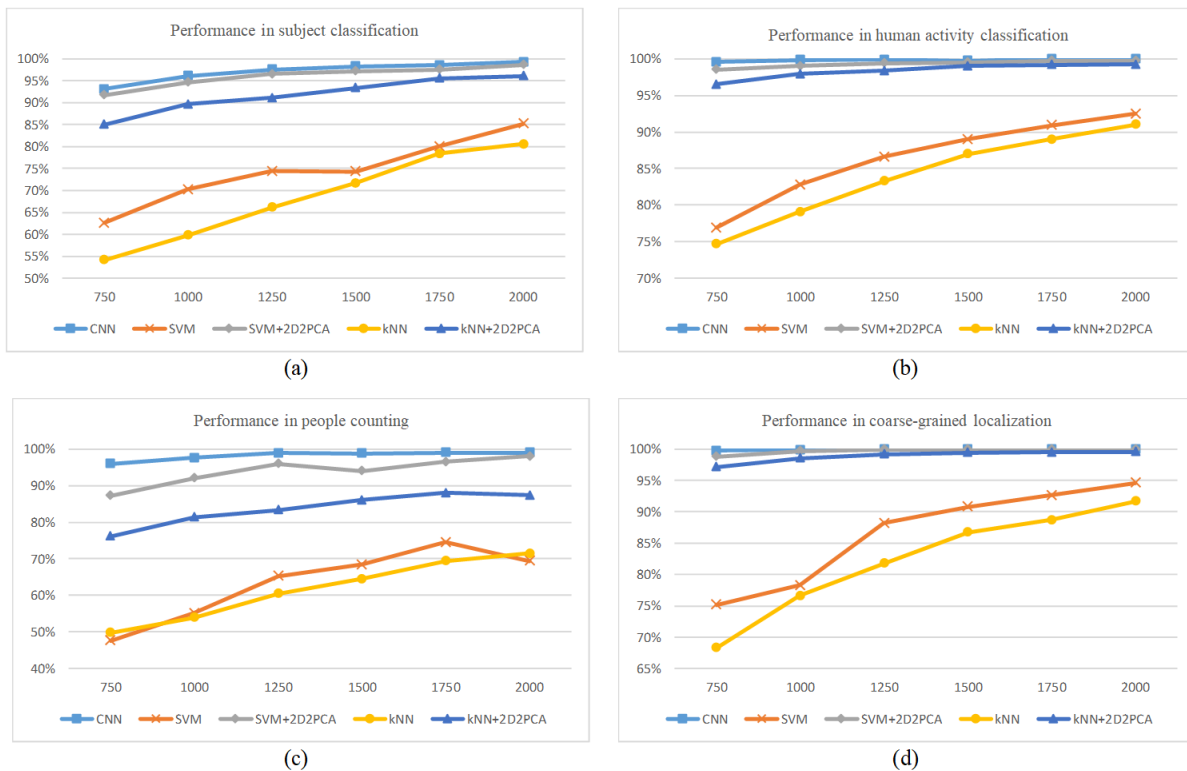


Figure 17. Performance (OA) of the classifiers with the changing frame lengths

in Section VII. Therefore, it is plausible to conclude that increasing the frame length of the sliding window can increase the classification accuracy. A longer frame length means the sliding window contains more information that makes the classification easier. In reality, it is not possible to increase the frame length endlessly, because each activity has a time period. Also, a longer frame length means a longer sampling time interval that results in higher latency. So, the frame length is determined based on the tradeoff between the classification accuracy and the latency. We used a 5s frame length that led to good accuracy and low latency. Also, the time of 5s was a suitable period to measure the movement in the detection area, because the walking or the running along the detection area usually took 4-7s.

### B. Angles of the movement

In Section III, we stated that the experiments were made from three different angles ( $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ ). The coarse localization made in *Case 3* was investigated from two angles ( $45^\circ$  and  $90^\circ$ ). As shown in Fig. 18, the classifications for human activity and people counting performed the best with spectrograms from  $0^\circ$ , reaching overall accuracies of 100% and 99.46% respectively. Samples from  $45^\circ$  produced the worst results (99.7% in human activity classification and people counting). For the coarse localization, the same OA (100%) was achieved at both  $45^\circ$  and  $90^\circ$ . This means the angle of the movement has no effect on the classification accuracy for localization. In conclusion, the direction of the movement to the radar beam can affect the classification in human activity and people counting, and  $0^\circ$  provides the best

accuracy, followed by  $90^\circ$  and  $45^\circ$ . This is probably because the RCS is largest when people move at  $0^\circ$ . While for the coarse localization, the angle of movement had no effect.

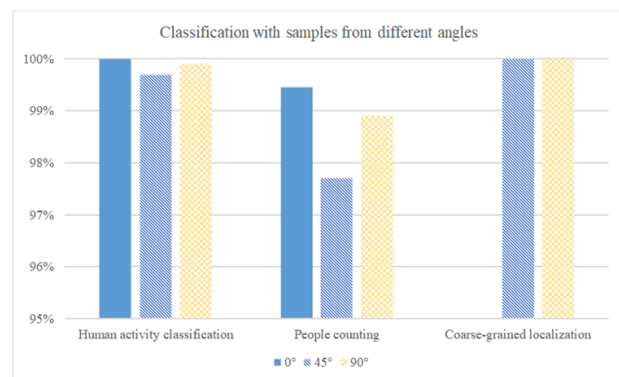


Figure 18. Performance of the classifiers for the tasks with the radar at different angles

### C. Number of Doppler Radars

The radar system built in this research consisted of two radars. In the data processing, the signals from both radars were fused together. In this section, we investigated how the number of radars affects classification. For this purpose, a new CNN model 'RadarNet<sub>one</sub>' was trained using only the data from the primary radar node.

The main difference between RadarNet<sub>one</sub> and RadarNet is the input size, which is  $50 \times 50 \times 1$  (one spectrogram) for RadarNet<sub>one</sub>, and  $50 \times 50 \times 2$  (two overlapped spectrograms)

Table IX  
STRUCTURE OF RADARNET AND RADARNET<sub>ONE</sub>

RadarNet	RadarNet <sub>one</sub>
X (50 × 50 × 2)	X (50 × 50 × 1)
Conv1-16@48 × 48 × 1	Conv1-12@48 × 48 × 1
maxpool	Maxpool
BN	BN
Conv2-32@22 × 22 × 1	Conv2-24@22 × 22 × 1
maxpool	Maxpool
BN	BN
Dropout (0.3)	Dropout (0.3)
Conv3-48@9 × 9 × 1	Conv3-24@9 × 9 × 1
Flatten	Flatten
BN	BN
Dropout (0.4)	Dropout (0.4)
FCL-350	FCL-256
BN	BN
Dropout (0.3)	Dropout (0.3)
FCL-160	FCL-76
FCL (Softmax)	

for RadarNet. As seen in Table IX, the structure of both CNNs are very similar, both contain 3 convolutional layers, 2 max-pooling layers, and 2 fully connected layers. However, RadarNet<sub>one</sub> presents less feature maps and hidden units. The comparison of their performance is shown in Fig. 19. The performance obtained from two radars was higher than the performance obtained by using only one radar in all four classification tasks. For human activity classification and coarse localization, the results presented by RadarNet<sub>one</sub> were slightly inferior to those of RadarNet. While for subject classification, and especially for people counting, the overall accuracy scores obtained by two radars exceeded those from one radar by a large margin. In people counting, the overall accuracy score with RadarNet<sub>one</sub> was 91.57%, while for RadarNet was 98.85%. In subject classification, the overall accuracy with one radar was 95.70%, while with two radars reached 97.5%. It is plausible to conclude that by increasing the number of radars, the accuracy of micro-Doppler based classification also increases. However, this increase is small for human activity classification and coarse localization, but very obvious for subject classification and people counting.

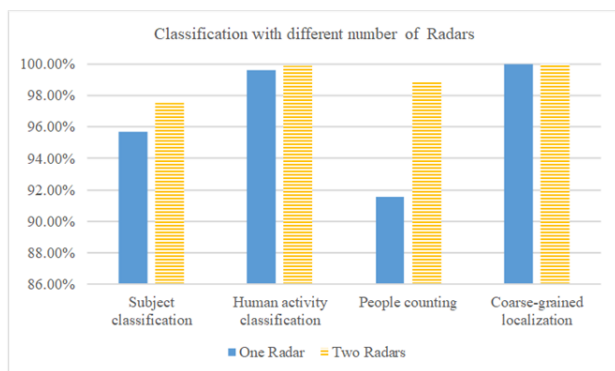


Figure 19. CNN classification with samples of different number of radars

## IX. CONCLUSION

This paper applied micro-Doppler signatures to perform four classification tasks, which are subject classification, hu-

man activity classification, people counting, and coarse localization. A radar system that consists of two pulsed Doppler radars operating at 5.8 GHz was used. With the collected radar signals and processing them with STFT, the patterns of the spectrograms of different subjects, activities, location ranges, and the number of people were presented and analyzed. Five classifiers, including CNN (RadarNet), SVM, SVM+2D2PCA, *k*NN, *k*NN+2D2PCA were implemented. It was found that our RadarNet performs the best for all four classification tasks. Also, 2D2PCA proved to be a very good feature extraction method for micro-Doppler analysis and improved the performance of SVM and *k*NN significantly. At last, three factors, including the frame length of the sliding window, the angle of the movement, and the number of radars were investigated for micro-Doppler signature applications. Our investigation provided a valuable guideline for model optimization and experiment setup of micro-Doppler based research and applications.

In conclusion, this research shows that low-power low-cost radars have a great potential for human activity detection, even in outdoor environments surrounded by trees and foliage. For future work, the method used in this research will be extrapolated into a Doppler radar with a greater detection range and extended to more application scenarios.

## ACKNOWLEDGMENT

This research was funded in part by a China Scholarship Council (CSC) and QMUL PhD Grant. The authors would also like to thank: Mahadi Hasan for assisting the first set up of the radar devices; and all participants in the experiments.

## REFERENCES

- [1] Ma Z., Bigham J., Poslad S., Wu B., Zhang X., and Bodanese E., "Device-free daily life (adl) recognition for smart home healthcare using a low-cost (2d) lidar."
- [2] Z. Zhang and S. Poslad, "Improved use of foot force sensors and mobile phone gps for mobility activity recognition," *IEEE Sensors Journal*, vol. 14, no. 12, pp. 4340–4347, 2014.
- [3] R. M. Narayanan and M. Zenaldin, "Radar micro-doppler signatures of various human activities," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1205–1215, 2015.
- [4] S. S. Ram and H. Ling, "Simulation of human microdopplers using computer animation data," in *Radar Conference, 2008. RADAR'08. IEEE*. IEEE, 2008, pp. 1–6.
- [5] A. Balleri, K. Chetty, and K. Woodbridge, "Classification of personnel targets by acoustic micro-doppler signatures," *IET radar, sonar & navigation*, vol. 5, no. 9, pp. 943–951, 2011.
- [6] S. Dura-Bernal, G. Garreau, C. Andreou, A. Andreou, J. Georgiou, T. Wennekers, and S. Denham, "Human action categorization using ultrasound micro-doppler signatures," in *International Workshop on Human Behavior Understanding*. Springer, 2011, pp. 18–28.
- [7] V. Chen, D. Tahmouh, and W. Miceli, "Micro-doppler signatures-review, challenges, and perspectives," *Radar-Micro Doppler Signatures: Processing and Applications*, pp. 1–25, 2014.
- [8] F. H. C. Tivive, S. L. Phung, and A. Bouzerdoum, "An image-based approach for classification of human micro-doppler radar signatures," in *Active and Passive Signatures IV*, vol. 8734. International Society for Optics and Photonics, 2013, p. 873406.
- [9] V. C. Chen and R. D. Lipps, "Time frequency signatures of micro-doppler phenomenon for feature extraction," in *Wavelet Applications VII*, vol. 4056. International Society for Optics and Photonics, 2000, pp. 220–227.
- [10] J. M. Garcia-Rubia, O. Kilic, V. Dang, Q. M. Nguyen, and N. Tran, "Analysis of moving human micro-doppler signature in forest environments," *Progress In Electromagnetics Research*, vol. 148, pp. 1–14, 2014.



- [11] Y. Kim, S. Ha, and J. Kwon, "Human detection using doppler radar based on physical characteristics of targets," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 2, pp. 289–293, 2015.
- [12] B. Jokanović and M. Amin, "Fall detection using deep learning in range-doppler radars," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 1, pp. 180–189, 2018.
- [13] Y. Yang, C. Hou, Y. Lang, D. Guan, D. Huang, and J. Xu, "Open-set human activity recognition based on micro-doppler signatures," *Pattern Recognition*, vol. 85, pp. 60–69, 2019.
- [14] B. Çağlıyan and S. Z. Gürbüz, "Micro-doppler-based human activity classification using the mote-scale bumblebee radar," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 10, pp. 2135–2139, 2015.
- [15] Y. Kim and H. Ling, "Human activity classification based on micro-doppler signatures using a support vector machine," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1328–1337, 2009.
- [16] I. Bilik, J. Tabrikian, and A. Cohen, "Gmm-based target classification for ground surveillance doppler radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 267–278, 2006.
- [17] D. P. Fairchild and R. M. Narayanan, "Multistatic micro-doppler radar for determining target orientation and activity classification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 1, pp. 512–521, 2016.
- [18] B. G. Mobasser and M. G. Amin, "A time-frequency classifier for human gait recognition," in *Optics and Photonics in Global Homeland Security V and Biometric Technology for Human Identification VI*, vol. 7306. International Society for Optics and Photonics, 2009, p. 730628.
- [19] D. P. Fairchild and R. M. Narayanan, "Classification of human motions using empirical mode decomposition of human micro-doppler signatures," *IET Radar, Sonar & Navigation*, vol. 8, no. 5, pp. 425–434, 2014.
- [20] L. Liu, M. Popescu, M. Skubic, M. Rantz, T. Yardibi, and P. Cuddihy, "Automatic fall detection based on doppler radar motion signature," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on*. Citeseer, 2011, pp. 222–225.
- [21] M. Zenaladin and R. M. Narayanan, "Radar micro-doppler based human activity classification for indoor and outdoor environments," in *Radar Sensor Technology XX*, vol. 9829. International Society for Optics and Photonics, 2016, p. 98291B.
- [22] S. Björklund, H. Petersson, A. Nezirovic, M. B. Guldogan, and F. Gustafsson, "Millimeter-wave radar micro-doppler signatures of human motion," in *Proceedings International Radar Symposium (IRS)*. IEEE, 2011, pp. 167–174.
- [23] J. Zabalza, C. Clemente, G. Di Caterina, J. Ren, J. J. Soraghan, and S. Marshall, "Robust pca micro-doppler classification using svm on embedded systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 3, pp. 2304–2310, 2014.
- [24] J. A. Nanzer and R. L. Rogers, "Bayesian classification of humans and vehicles using micro-doppler signals from a scanning-beam radar," *IEEE Microwave and Wireless Components Letters*, vol. 19, no. 5, pp. 338–340, 2009.
- [25] R. Trommel, R. Harmanny, L. Cifola, and J. Driessen, "Multi-target human gait classification using deep convolutional neural networks on micro-doppler spectrograms," in *Radar Conference (EuRAD), 2016 European*. IEEE, 2016, pp. 81–84.
- [26] Y. Kim and T. Moon, "Human detection and activity classification based on micro-doppler signatures using deep convolutional neural networks," *IEEE geoscience and remote sensing letters*, vol. 13, no. 1, pp. 8–12, 2016.
- [27] K. Saho, T. Sakamoto, T. Sato, K. Inoue, and T. Fukuda, "Pedestrian classification based on radial velocity features of uwb doppler radar images," in *2012 International Symposium on Antennas and Propagation (ISAP)*. IEEE, 2012, pp. 90–93.
- [28] D. Tahmouh and J. Silvious, "Radar micro-doppler for long range front-view gait recognition," in *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*. IEEE, 2009, pp. 1–6.
- [29] Z. Zhang, P. Pouliquen, A. Waxman, and A. G. Andreou, "Acoustic micro-doppler gait signatures of humans and animals," in *Information Sciences and Systems, 2007. CISS'07. 41st Annual Conference on*. IEEE, 2007, pp. 627–630.
- [30] X. Shi, F. Zhou, L. Liu, B. Zhao, and Z. Zhang, "Textural feature extraction based on time–frequency spectrograms of humans and vehicles," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1251–1259, 2015.
- [31] Samraksh Company, "Bumblebee user's manual," <https://samraksh.com/files/products/Bumblebee%20radar/BumbleBee-UM-v103.pdf>, 2008, [Online; accessed 19-July-2018].
- [32] J. Polastre, R. Szwedczyk, and D. Culler, "Telos: enabling ultra-low power wireless research," in *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*. IEEE, 2005, pp. 364–369.
- [33] D. Zhang and Z.-H. Zhou, "(2d) 2pca: Two-directional two-dimensional pca for efficient face representation and recognition," *Neurocomputing*, vol. 69, no. 1-3, pp. 224–231, 2005.
- [34] J. R. Bergado, C. Persello, and A. Stein, "Recurrent multiresolution convolutional networks for vhr image classification," *arXiv preprint arXiv:1806.05793*, 2018.
- [35] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.
- [36] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Annals of operations research*, vol. 134, no. 1, pp. 19–67, 2005.
- [37] R. A. Dunne and N. A. Campbell, "On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function," in *Proc. 8th Aust. Conf. on the Neural Networks, Melbourne*, vol. 181. Citeseer, 1997, p. 185.
- [38] S. Haykin, *Neural networks*. Prentice hall New York, 1994, vol. 2.
- [39] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
- [40] M. D. Zeiler, "Adadelata: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.
- [41] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.
- [42] B. Erol and S. Z. Gurbuz, "A kinect-based human micro-doppler simulator," *IEEE Aerospace and Electronic Systems Magazine*, vol. 30, no. 5, pp. 6–17, 2015.
- [43] S. Z. Gürbüz, B. Erol, B. Çağlıyan, and B. Tekeli, "Operational assessment and adaptive selection of micro-doppler features," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1196–1204, 2015.
- [44] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," 2011.
- [45] A. Miller, C. Clemente, A. Robinson, D. Greig, A. Kinghorn, and J. Soraghan, "Micro-doppler based target classification using multi-feature integration," 2013.



**Fei Luo** received the B.Sc. degree in Geographic Information System for Jiangxi University of Science and Technology, Jiangxi, China, in 2008, and the M.Sc. degree in Surveying and Mapping from Wuhan University, Wuhan, China, in 2016. He is currently pursuing the Ph.D. degree in the School of Electronic Engineering and Computer Science. His research interests include Geographic Information Systems, Human Activity Detection, and Machine Learning.



**Stefan Poslad** received the B.Sc. degree in Physics from the University of Southampton, the M.Sc. degree in Medical Physics from the University of Aberdeen, and the Ph.D. degree from Newcastle University, all in the UK. Since 2000, he has been a Principal Investigator of 12 funded collaborative research projects, and has received funding from the U.K. Research Council, Europe, Norway, the U.S., and from the industry.



**Eliane Bodanese** MSc, Ph.D., MIEEE. She is a senior lecturer in the School of Electronic Engineering and Computer Science and a member of the Communication Systems Research Group. Her research interests include human activity detection and recognition through minimal non-intrusive sensing, indoor and outdoor localization, and communication support for emergencies.