Human activity classification using micro-Doppler signatures and ranging techniques

by

Fei Luo

A thesis submitted to the Queen Mary University of London for the degree of Doctor of Philosophy

School of Electronic Engineering and Computer Science
Queen Mary, University of London
United Kingdom

January 2020

Statement of Originality

I, Fei Luo, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged and my contribution indicated. Previously published material is also acknowledged.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature: Fei Luo

Date: 11/11/2019

Abstract

Human activity recognition is emerging as a very import research area due to its potential applications in surveillance, assisted living, and military operations. Various sensors including accelerometers, RFID, and cameras, have been applied to achieve automatic human activity recognition. Wearable sensor-based techniques have been well explored. However, some studies have shown that many users are more disinclined to use wearable sensors and also may forget to carry them. Consequently, research in this area started to apply contactless sensing techniques to achieve human activity recognition unobtrusively. In this research, two methods were investigated for human activity recognition, one method is radar-based and the other is using LiDAR (Light Detection and Ranging). Compared to other techniques, Doppler radar and LiDAR have several advantages including all-weather and all-day capabilities, non-contact and nonintrusive features. Doppler radar also has strong penetration to walls, clothes, trees, etc. LiDAR can capture accurate (centimetre-level) locations of targets in real-time. These characteristics make methods based on Doppler radar and LiDAR superior to other techniques.

Firstly, this research measured micro-Doppler signatures of different human activities indoors and outdoors using Doppler radars. Micro-Doppler signatures are presented in the frequency domain to reflect different frequency shifts resulted from different components of a moving target. One of the major differences of this research in relation to other relevant research is that a simple pulsed radar system of very low-power was used. The outdoor experiments were performed in places of heavy clutter (grass, trees, uneven terrains), and confusers including animals and drones, were also considered in the experiments. Novel usages of machine learning techniques were implemented to perform subject classification, human activity classification, people counting, and coarse-grained localisation by classifying the micro-Doppler signatures. For the feature extraction of

the micro-Doppler signatures, this research proposed the use of a two-directional two-dimensional principal component analysis (2D2PCA). The results show that by applying 2D2PCA, the accuracy results of Support Vector Machine (SVM) and k-Nearest Neighbour (kNN) classifiers were greatly improved. A Convolutional Neural Network (CNN) was built for the target classifications of type, number, activity, and coarse localisation. The CNN model obtained very high classification accuracies (97% to 100%) for the outdoor experiments, which were superior to the results obtained by SVM and kNN. The indoor experiments measured several daily activities with the focus on dietary activities (eating and drinking). An overall classification rate of 92.8% was obtained in activity recognition in a kitchen scenario using the CNN. Most importantly, in nearly real-time, the proposed approach successfully recognized human activities in more than 89% of the time. This research also investigated the effects on the classification performance of the frame length of the sliding window, the angle of the direction of movement, and the number of radars used; providing valuable guidelines for machine learning modeling and experimental setup of micro-Doppler based research and applications.

Secondly, this research used a two dimensional (2D) LiDAR to perform human activity detection indoors. LiDAR is a popular surveying method that has been widely used in localisation, navigation, and mapping. This research proposed the use of a 2D LiDAR to perform multiple people activity recognition by classifying their trajectories. Points collected by the LiDAR were clustered and classified into human and non-human classes. For the human class, the Kalman filter was used to track their trajectories, and the trajectories were further segmented and labelled with their corresponding activities. Spatial transformation was used for trajectory augmentation in order to overcome the problem of unbalanced classes and boost the performance of human activity recognition. Finally, a Long Short-term Memory (LSTM) network and a (Temporal Convolutional Network) TCN was built to classify the trajectory samples into fifteen activity classes. The TCN achieved the best result of 99.49% overall accuracy. In comparison, the proposed TCN slightly outperforms the LSTM. Both of them outperform hidden Markov Model (HMM), dynamic time warping (DTW), and SVM with a wide margin.

Acknowledgments

First of all, I would like to express my sincere gratitude to my supervisor Dr. Eliane Bodanese for her useful suggestions, comments and encouragement. She not only provides me advice and guidance on research throughout the whole three years but also motivates me to achieve a variety of research approaches. I also wish to thank Dr. Stefan Poslad for his constructive comments on my reports, papers, and this thesis. I am deeply grateful for their help in the completion of this thesis.

Secondly, I would also like to thank my parents and my brother. They have been my strongest backup when I was frustrated. They are my motivations and strengths on the road of life that full of thorns.

Last but not least, I would like to thank my country for providing the CSC scholarship during the period of my PhD.

Contents

List of Figures

List of Tables

List of Abbreviations

2D2PCA	Two-directional two-dimensional principle component analysis
AA	Average class accuracy
ADL	Activities of Daily Living
AOL	Airborne Oceanographic Lidar
CEC	Constant error carousel
CNN	Convolutional Neural Network
CW	Continuous wave
DBSCAN	Density-based spatial clustering algorithm
DCNN	Deep convolutional Neural Network
DIAL	Differential absorption Lidar
DTW	Dynamic time warping
EMD	Empirical Mode Decomposition
FFT	Fast Fourier transform
FMCW	Frequency Modulated Continuous wave
GPS	Global Positioning System
HAR	Human activity recognition
HF	High frequency
HMM	Hidden Markov Model
ICA	Independent Component Analysis
IMF	Intrinsic mode function

IMU	Inertial measurement unit
kNN	k-Nearest Neighbour
LDA	Linear Discriminative Analysis
LiDAR	Light Detection and Ranging
LPRF	Low pulse repetition frequency
LSTM	Long Short-term Memory
MFC	Mel-frequency cepstrum
MFCC	Mel-frequency Cepstral Coefficients
MPRF	Medium pulse repetition frequency
OA	Overall classification accuracy
PCA	Principle Component Analysis
PDR	Pulsed Doppler Radar
PRF	Pulse repetition frequency
PRI	Pulse repetition interval
RBF	Radial Basis Function
ReLU	Rectified Linear Units
RF	Radio frequency
RFID	Radio-frequency identification
RNN	Recurrent Neural network
SFP	Spectrogram frequency profile
SNR	Signal-noise-ratio
STFT	Short time Fourier transform
SVD	Singular value decomposition
SVM	Support Vector Machine
TCN	Temporal Convolutional Network
UAV	Unmanned aerial vehicle
UHF	Ultra-high frequency
UWB	Ultra-wideband radar
VHF	Very high frequency

WSN Wireless Sensor Network

Chapter 1

Introduction

Human activity recognition is a wide field of study concerned with identifying the specific movement or action of a person or a group of people based on sensor data. Automatic human activity recognition is essential in many ambient intelligence applications such as surveillance, smart home, health monitoring, intelligent control, etc. For example, users can use gesture recognition [10, 11] to contactlessly interact with electronic devices such as TV, computers, virtual reality (VR)/augmented reality (AR) glasses, etc. Online abnormal activity recognition can perform foul detection in sports [12], fall detection [13], theft detection [14], illegal parking [15], etc. Daily activity monitoring can provide health assistance for the elderly and patients [16].

Wide research in human activity recognition has been done by using various sensors and techniques. Video-based systems have been well developed for HAR. It has several advantages. Firstly, videos and images are visual forms that can easily be perceived by people's eyes. Secondly, cameras are widely deployed and used. It is easy to obtain video data. Thirdly, Videos and images can be labelled after data collection. However, wearable-sensor based and radio-based systems need to perform data labelling while data collection. Fourthly, although different cameras have differences in resolution, size, brightness, etc., these differences can be easily addressed by using data

augmentation (crop, scale, blend, etc.). The disadvantages of video-based systems are also worthy to be mentioned. Video-based systems suffer from insufficient illumination. Video-based systems may cause privacy concerns because people's faces are exposed in front of cameras. The data volume of video-based systems is huge. Video data is 4-dimensional $(width, height, channel, time)$. Data of radars and wearable sensors is usually 2-dimensional $(channel, time)$. Video-based systems are very computation-intensive because of their huge data volume.

Wearable sensors including accelerometer, gyroscope, magnetometer, Bluetooth and RFID receivers have increasingly been applied in human activity recognition as they are ubiquitous in smart phones, smart watches, and sport bracelets. Wearable sensor based systems are inconvenient and intrusive to users in many situations as they require people to keep wearing them. Concerning privacy, convenience, power consumption, etc., more and more researchers are exploring other possible techniques including WiFi, GPS (Global Positioning System), and radar for human activity recognition. This thesis focuses on contactless human activity recognition by using two different techniques including micro-Doppler and LiDAR.

1.1 Motivations

The aim of human activity recognition system is to recognize actions and activities of people by using sensors. There are rapidly increasing demands for human activity recognition systems. In healthcare, it can record the daily changes of patients in dietary habits, medicine taking, and sleep conditions. With the coming of the aging society, activity recognition systems also can provide healthcare assistance for the elderly. For example, they can provide early warning for fall detection and report their daily activities to remote doctors for health evaluation. In surveillance, activity recognition systems can help to detect stealing and robbery, or even prevent potentially dangerous situations. Thus, human activity recognition can increase the security level and protect people from acts of violence, terrorism, vandalism, and stealing. In human and computer interaction,

more and more VR/AR devices are emerging and they have no keyboard or mouse. Activity recognition can provide users an immersive way to interact with them. In electronic games industry, more games try to interpret body motions to level up users' game experience, especially in sports games. In robotics, activity recognition can help autonomous vehicles to assess and anticipate pedestrian intention to cross a road, which is helpful to avoid potential car accidents. Partner robot systems require the ability of activity recognition to assist in healthcare and ordinary housework.

Video-based and wearable-based systems have been widely explored and well developed. However, video-based systems are privacy intrusive and wearable sensors are inconvenient for users to keep wearing them. For instance, a smart band has been used to monitoring sleep conditions of users, but some users feel uncomfortable to wear a smart band at sleeping time. Ambient sensing techniques such as radar, LiDAR, and WiFi are increasingly applied in human activity recognition because they can measure human activities remotely. These techniques do not require users to carry any receivers or transmitters and users' identities are not exposed. It is important to develop a user-friendly activity recognition system by using these ambient sensing techniques. This thesis focuses on two ambient sensing techniques: micro-Doppler and LiDAR.

1.2 Micro-Doppler based human activity recognition

Human activity recognition can be accomplished by measuring differences in radar micro-Doppler signatures of different 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-motion, will result in additional modulation of the main Doppler frequency shift, known as the micro-Doppler effect [17, 18]. In 2000, Chen first labelled the intratarget radar frequency shifts as the micro-Doppler effect in [19]. Later on, he further developed this concept to micro-Doppler signature. As he described in [20], 'The micro-Doppler signature is a distinctive characteristic of the observed micro-Doppler effect in an object. The "signature" is com-

monly used to refer to the characteristic expression of an object or a process. When examining the micro-Doppler effect, the distinctive micro-Doppler characteristic, i.e., the micro-Doppler signature of an object, allows the recognition of an object's identity through its movements.

Human activity is complex but very unique. A human body is an articulated object, comprising a number of rigid parts connected by joints. When a person moves, different parts of his body (torso, arms, legs) have a particular motion that produces characteristic micro-Doppler signatures. Usually, human activities can be decomposed into periodic motions of different parts of the human body. As shown in Figure 1.1, when a person is walking or running, his arms and legs are swinging back and forth periodically relative to his torso. The torso, arms, and legs have their own speeds, and each individual part will result in different Doppler frequency returns. In Figure 1.2, the spectrogram of a person walking with both arms swinging is generated by time-frequency analysis. It shows that the spine of the spectrogram is induced by the torso, and the main sidelobes represent the micro-Doppler signatures of the leg swing and the arm swing.

Micro-Doppler signatures are obtained by Doppler radars. Unlike other sensor technologies, Doppler radar has all-weather and day-and-night capabilities [21], while smoke, dust, and fog only slightly attenuate the signal. Radar signals are also able to penetrate building wall material, cloth, and dense foliage [22]. It does not intrude on people's privacy, their identities and faces will not be disclosed because of radar detection. So it is very suitable to be used in human activity recognition.

1.3 LiDAR based human activity recognition

Trajectory-based human activity recognition is gradually attracting attention from industry and academia as human activity is usually accompanied by the location changes of people. Location acquisition systems have been widely applied in people's life and several techniques have been used, such as GPS, radar, LiDAR, etc. These systems generate a large amount of location data, and in the past years, they have been applied

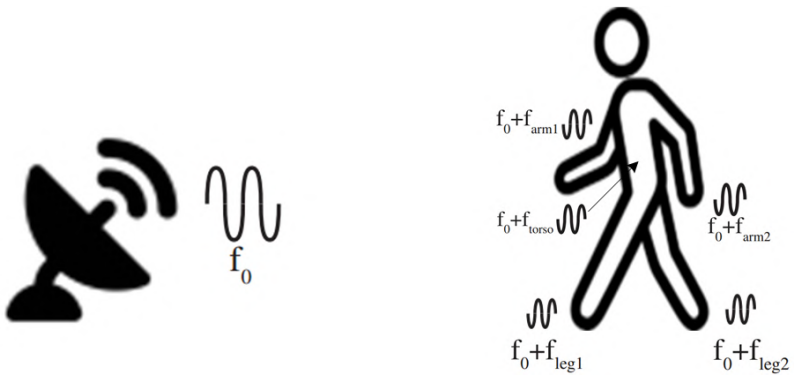


Figure 1.1: A person moving towards a radar

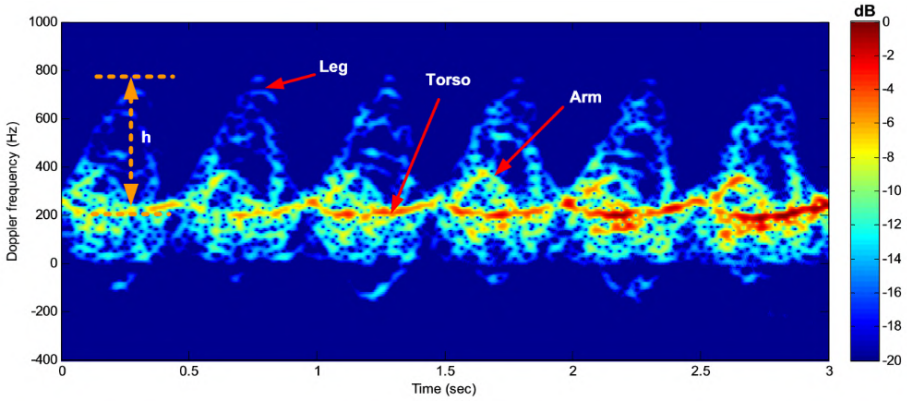


Figure 1.2: Spectrogram of a walking person [1]

in navigation, route planning, and mapping. Recently, more and more studies attempt to infer semantic information from the trajectories formed by the locations of people, in order to provide a higher level of services, for example, for personalized recommendations, smarter home or more intelligent human-like robot interaction. Although there are many localisation techniques, none of these can be applied ubiquitously for any purpose. GPS hardly works indoors as its signals cannot penetrate walls. Radars may suffer interference from the ground surrounding and multi-path effects. Bluetooth and RFID (Radio-Frequency Identification) can perform indoor localisation, however, due to their short-range sensing capability, they generally require larger deployments that may become difficult to maintain.

As a surveying and mapping technique, LiDAR has been increasingly used in self-driving and robots. LiDAR is a surveying method that measures the distance to a target

by illuminating the target with pulsed laser light and measuring the reflected pulses with a laser [23]. The data collected by a LiDAR is a set of points. These points show the locations of different objects. The trajectory of a moving person can be obtained by correlating its locations at different time steps. Trajectories reflect the activities of people from temporal and spatial dimensions. LiDAR has high localisation accuracy, high real-time performance, and it is easy to deploy. It also can be applied both indoors and outdoors. These merits make LiDAR an ideal technique in trajectory-based human activity recognition. However, research in trajectory-based human activity recognition using LiDAR is few in number.

1.4 Challenges

Human activity is a very challenging research topic. It requires multi-discipline knowledge including sensors, communication, signal processing, machine learning, etc. As human activity is a dynamic changing process, this process is affected by many factors such as occlusions, noise, and confusers. Even the same activity can be performed from different perspectives at different speeds by different people. The limitations of sensors also bring challenges to human activity recognition. For long-time continuous monitoring, it requires activity recognition systems to be power-efficient. For emergency, it requires activity recognition systems to support nearly real-time response. This section summarizes the main challenges in human activity recognition using micro-Doppler signatures and LiDAR.

Complexity of human activity

Human activity has high complexity. It is a dynamic process that contains the movements of multiple components of the human body. Human activity has many types, such as running, walking, sitting, etc. Some activities are highly dependent on scenarios, for instance, walking to get in, walking to get out, walking upstairs, walking downstairs. Activities of daily living are switching in different scenarios, which creates many types of activities. Different activities have different time cycles. Consequently, it is impossible

to capture these activities by using a fixed duration. Some activities usually happen simultaneously. For example, people can take a phone call while walking. Besides, some activities have high similarities, such as eating and drinking. It is very hard to discriminate such activities due to their similar movement patterns.

Activity recognition of multiple people

Currently, most research work focuses on single-person activity recognition. However, activity recognition of multiple people or a group of people is also important. For instance, the activity recognition of a team in sports can support the coach to do tactic analysis. Activity recognition of the crowd in subway stations is helpful to security and evacuation in an emergency. The changing spatial distribution of people and the interactions between people in a group increase the difficulty of multi-person activity recognition.

Noises and obstacles

It is very possible that non-human objects produce noises and interference in human activity recognition system. For micro-Doppler based human activity recognition, the noises resulted from other small movements and moving objects can affect the performance. Especially, outdoors, swaying leaves and trees generate additional frequency changes that interfere with the signal frequency of a target, and animals and birds are potential confusers to humans. For LiDAR, it is necessary to discriminate humans for non-human objects such as walls, tables, cars, etc., in a space.

Real-time response

Many applications require a real-time response in human activity recognition. For example, it needs to report abnormal behaviours (falling, abnormal heartbeat) in order to provide immediate assistance for the elderly and patients. Generally, human activity recognition requires a multi-stage process including data collection, data fusion from multiple sensors, data processing, classification. It is challenging to finish all these pro-

cesses in nearly real-time with limited computation resources. It requires to divide sensor signals to small segments with little compromise in recognition performance.

Spatio-temporal context association

In many human activity recognition tasks, it is not sufficient to recognize specific human activities. It is also important to know 'when' and 'where' these activities are happening or have happened. It requires to associate the human activity with time and location. It is not difficult to get the occurrence of the time because sensor signals are time-series. However, many sensors have no localisation capabilities. This requires to integrate human activity techniques and localisation techniques.

1.5 Main contributions

As mentioned above, the work in this thesis make use of Doppler radar and LiDAR to perform human activity recognition. The main contributions of this thesis are listed as follows:

1. **Usage of power-efficient Doppler radars**: this thesis uses low-cost and low-power Doppler radars to support long-time and continuous human activity recognition both indoors and outdoors. Most radars used in the literature are very power consuming. In some outdoor areas, there is even no power supply to support these radars to perform continuous human activity recognition. They are not suitable to perform continuous and long-time monitoring. The usage of energy-efficient radars is also able to reduce the cost of power and the effort of maintenance . The radar system built in this research consists of two Bumblebee radars. The center frequency of each Bumblebee radar is 5.8 GHz. Its detection range is up to 10m. Especially, it only consumes about 12 mA and can be power by 3 AA batteries for about 8 days. More details about Bumblebee radars are described in Section 3.3.1. By using such a radar system, this research still achieved the results that are comparable or even better than the related work. The results are 98.58% overall

accuracy (OA) in human recognition, 99.89% OA in activity recognition, 98.85% OA in people counting, and 100% in rough localisation.

2. **Consideration of the clutter from environments**: In reality, lots of objects can create interference to human activity recognition. Outdoor environments are relatively unstable, and the trees, leaves, and grass keep swaying because of airflow. They will introduce interference/noise to radar signals. Other moving objects such as animals, cars, drones may be confused with humans, creating false positives. It is necessary to discriminate humans from non-human objects. In this research, the radar signals were collected from three different scenarios that are surrounded by trees and grasses. Samples created from these signals are able to improve the robustness of machine learning models to noises. This research did not just collect the signals of humans, but also dogs, drones, and backgrounds. This can eliminate the false positives that may be resulted by dogs, drones, and backgrounds. For the LiDAR-based approach, the LiDAR points do not just contain humans, but also tables, chairs, and walls. In order to remove the clutters of tables, chairs, and walls, geometric features were extracted to classify them as human and non-human classes after clustering.

3. **Nearly real-time human activity recognition**: In the related work, many approaches are only able to perform offline human activity recognition after collecting all data. Although some of them can do online activity recognition, they use a long window to produce signal segments, which results in a significant delay. This thesis uses very short sliding windows to perform data segmentation in order to achieve nearly real-time human activity recognition. For radar-based activity recognition outdoors, a 5s sliding window was used to segment radar signals, which results in a delay of less than 5s. For radar-based human activity recognition indoors, a 2.5s sliding window was applied in signal segmentation, which results in a delay of less than 2.5s. For LiDAR-based human activity recognition indoors, a 2.5s sliding window was used to segment the trajectories of fifteen activities, which also results

in a delay of less than 2.5s.

4. **Activity recognition of multiple people**: This thesis does not just investigate the activity recognition of a single person. This research also investigated micro-Doppler signatures of a group of people and successfully estimated the number of people and the activity of groups. The overall accuracy achieved for micro-Doppler based people number estimation reaches 98.85%. For activity recognition of a group of people, the micro-Doppler based approach reaches 99.89%. LiDAR successfully recognise the activities of each person among a group of people, and the overall accuracy for the recognition of fifteen activities reaches 99.49%.

5. **Association of human activity and the corresponding location**: It is important to know where the human activity is happening or happened. This research associates human activities with their corresponding locations. For the micro-Doppler technique, an approach is proposed to perform coarse-grained estimation of the distance between the radar and the target according to the changing intensity of the micro-Doppler signatures due to the atmospheric attenuation, and reflections of the target and the environment. In this thesis, the overall accuracy achieved for micro-Doppler based coarse-grained localisation reaches 100%. As LiDAR itself is a localisation technique, it easy to get very accurate locations (centimeter-level) of humans from the point cloud of LiDAR.

6. **Datasets for human activity recognition**: In this research, three datasets were collected. They are radar signals for human activity recognition outdoors, radar signals for human activity recognition indoors, and 2D LiDAR data for human activity recognition indoors. The first dataset was collected from three angles by using the radar sensor network outdoors for human recognition, activity recognition, people counting, and rough localisation. The second dataset was collected by using the radar sensor network in a kitchen for 15 activities. The third dataset was collected by using a 2D LiDAR in the kitchen for 15 activities. The datasets are public now in order to encourage people who interested

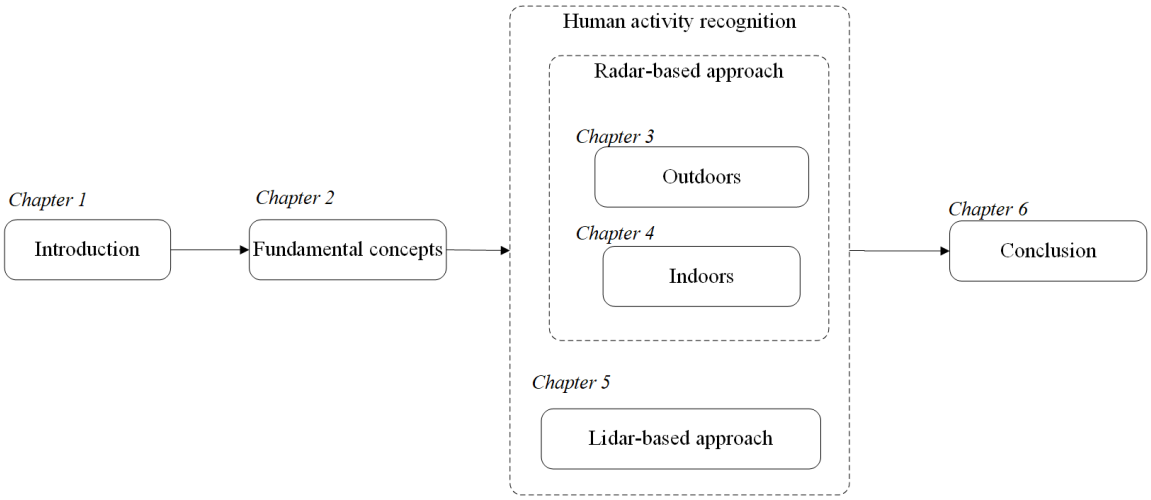


Figure 1.3: Thesis progression

in human activity recognition to develop better algorithms upon it. The link is https://drive.google.com/open?id=1pLwiKg8a5s-pAf58QRWOPGn4brzlPsd5.

1.6 Thesis outline

This thesis has been organized into 6 chapters, as shown in Figure 1.3. A brief description of the consecutive chapters is as follows:

- **Chapter 2 – Fundamental concepts of radar, LiDAR, and machine learning**

  This chapter introduces the history and fundamental concepts of both radar and LiDAR. The mathematical deduction of the micro-Doppler analysis is also presented. It describes machine learning algorithms including CNNs and LSTM used in human activity classifications in detail.

- **Chapter 3 – Human activity recognition, people counting, and coarse-grained localisation outdoors using micro-Doppler signatures**

  This chapter presents novel usages of machine learning techniques to perform human detection, human activity recognition, people counting, and coarse-grained

localisation by classifying micro-Doppler signatures obtained from a low-cost and low-power radar system. It describes the experiments performed outdoors in detail. It presents the proposed CNN that achieved much better results than SVM, kNN, and the other approaches in the related work for human activity recognition. It demonstrates the investigation of three factors in micro-Doppler analysis including the frame length of the sliding window, the angle of the direction of movement, and the number of radars.

- **Chapter 4 – Human activity recognition indoors using micro-Doppler signatures**

  This chapter describes the implementation of the human activity recognition in a kitchen scenario using the low-power radar system. Fifteen different activities of different people were investigated. It describes the proposed CNN for human activity recognition. It also presents a real-time human activity system that was developed in this research and it successfully recognized human activities in more than 89% of the time.

- **Chapter 5 – Multi-person activity recognition using a 2D LiDAR**

  This chapter describes the use of a 2D LiDAR in multiple people activity recognition. Raw LiDAR data was divided into different clusters. The clusters were further classified into human and non-human classes. For the clusters of humans, it presents the implementation of the Kalman Filter for tracking their trajectories which were further segmented and labeled with corresponding activities. It describes the use of spatial transformation for trajectory augmentation in order to overcome the problem of unbalanced classes and boost the performance of human activity recognition. It illustrates the proposed LSTM network and the TCN for classifying trajectory samples of 15 activities collected in a kitchen.

- **Chapter 6 – Analysis, conclusion, and future works**

This chapter concludes this dissertation, summarizes the main advantages and disadvantages of all the proposed approaches. This chapter makes a comparison between micro-Doppler based techniques and LiDAR based techniques, and presents possible future directions, extensions, and perspectives for future research.

1.7 Publications

Journal papers

1. **F. Luo**, S. Poslad, and E. Bodanese, "Human Activity Detection and Coarse Localization Outdoors Using Micro-Doppler Signatures," *IEEE Sensors Journal*. (Published)

2. **F. Luo**, S. Poslad, and E. Bodanese, "LSTM-based multi-person activity recognition using a 2D LIDAR," *IEEE Internet of Things Journal*. (Accepted and will be published soon)

Conference papers

1. **F. Luo**, S. Poslad, and E. Bodanese, "Kitchen Activity Detection for Healthcare using a Low-Power Radar-Enabled Sensor Network," *IEEE International Conference on Communications* , Shanghai, China, May, 2019. (Published)

Chapter 2

Fundamental concepts of Doppler radar, LiDAR, and machine learning

Human activity recognition generally involves sensor data collection, data processing, feature extraction, and classification. The sensor data in this research is collected by using a Doppler radar system and a LiDAR sensor. This chapter first introduces the history, types, and structures of Doppler radar and LiDAR respectively. As human activity is reflected and represented by micro-Doppler signatures while using Doppler radars, the mathematical principles of micro-Doppler are introduced. Activity classification is usually performed by using machine learning. Since deep learning has achieved state-of-the-art in many fields, this thesis implements deep learning including CNNs and LSTM to perform activity classification. The fundamental of CNNs and LSTM are described in detail in this chapter.

2.1 Doppler radar

This section first introduces the basic concepts and an abridged history related to the Doppler radar. The background of Doppler radar is put in a historical context to show what attempts have been made to extend the capabilities of radar, how Doppler radar was invented, the components of radar, and what area the Doppler radar has been used for. Then it presents the fundamental concepts of micro-Doppler and the frequency spectrum analysis used for the representation of micro-Doppler signatures.

2.1.1 The history of radar

The acronym RADAR (for Radio Detection And Ranging) was coined by the U.S. Navy in 1940, and the term "radar" became widely used. Before the technology was first demonstrated in the United States during December 1934, it has been explored by many scientists. In 1870s, Scottish physicist James Clerk Maxwell had formulated the general equations of the electromagnetic field, determining that both light and radio waves are examples of electromagnetic waves governed by the same fundamental laws but having widely different frequencies. During the late 1880s, German physicist Heinrich Hertz verified the earlier theoretical work of Maxwell, demonstrated that radio waves could be reflected by metallic objects. In 1900, Nicola Tesla suggested that the reflection of electromagnetic waves could be used for detecting of moving metallic objects. In 1917, he stated the principle of radar using standing electromagnetic waves along with pulsed reflected surface waves to determine the relative position, speed, and course of a moving object [24]. As these principles were becoming widely available in the early 20th century, German inventor Christian Hülsmeyer is the first to use radio waves to detect a ship in dense fog on the sea. The system was not able to provide range information, only warning of a nearby object. In 1922, the American electrical engineers Albert H. Taylor and Leo C. Young of the Naval Research Laboratory were inspired by Marconi's suggestion to use a continuous waveform interference radar with 5 m wavelength, with separate transmitter and receiver, to locate a wooden ship for the first time; and discovered that a ship passing

through the beam path caused the received signal to fade in and out. Eight years later, Lawrence A. Hyland at the Naval Research Laboratory observed similar fading effects from a passing aircraft; this led to a patent application as well as a proposal for serious work at the Naval Research Laboratory on radio-echo signals from moving targets.

Before the Second World War, researchers in different countries developed technologies that led to the modern version of radar independently and secretly. In France in 1934, the research branch of the Compagnie Générale de Télégraphie Sans Fil (CSF), headed by Maurice Ponte began developing an obstacle-locating radio apparatus, a part of which was installed on the liner Normandie in 1935. During the same time, the Soviet military engineer P. K. Oshchepkov, in collaboration with Leningrad Electrophysical Institute, produced an experimental apparatus, RAPID, capable of detecting an aircraft within 3 km of a receiver [25]. However, the French and Soviet systems had continuous-wave operation that could not provide the full performance ultimately equivalent to modern radar. Full radar evolved as a pulsed system, and the first such elementary apparatus was demonstrated in December 1934 by the American Robert M. Page, working at the Naval Research Laboratory [26]. The following year, the German physicist Rudolf Kühnhold invented a pulsed system. The Chain Home radar system, a five-station system, was developed and by 1940 stretched across the entire UK including Northern Ireland. In 1939, a small but powerful radar was developed by John Turton Randall and Henry Albert Howard Boot, this radar was deployed in the B–17 airplanes, which aimed to spot submarines from the air. During the Second World War, the radar technology underwent a strong development boost, which was driven by general war events and the development of the Air Force to major branch of service. After the war, progress in radar technology slowed considerably. The last half of the 1940s was devoted to developments initiated during the war. Two of these were the monopulse tracking radar and the moving-target indication (MTI) radar. Another notable development was the klystron amplifier, which provided a source of stable high power for very-long-range radars. Synthetic aperture radar first appeared in the early 1950s, but it took almost 30 more years to reach a high state of development, with the introduction of digital

processing and other advances. The airborne pulse Doppler radar also was introduced in the late 1950s in the Bomarc air-to-air missile. Also in 1950s, the publication of important theoretical concepts helped to put radar design on a more quantitative basis. These included the statistical theory of detection of signals in noise, called matched filter theory, which showed how to configure a radar receiver to maximize detection of weak signal; the Woodward ambiguity diagram, which made clear the trade-offs in waveform design for good range and radial velocity measurement and resolution; and the basic methods for Doppler filtering in MTI radars, which later became important when digital technology allowed the theoretical concepts to become a practical reality.

The Doppler frequency shift and its utility for radar were known before the Second World War, an Austrian physicist Christian Andreas Doppler first described how the observed frequency of light and sound waves was affected by the relative motion of the source and the detector in 1842. This phenomenon became known as the Doppler effect. But it took years of development to achieve the technology necessary for wide-scale adoption. Serious application of the Doppler principle to radar began in the 1950s, and today the principle has become vital in the operation of many radar systems. As previously explained, the Doppler frequency shift of the reflected signal results from the relative motion between the target and the radar. The use of the Doppler frequency is indispensable in continuous wave, MTI, and pulse Doppler radars, which must detect moving targets in the presence of large clutter echoes. The Doppler frequency shift is the basis for police radar guns. Synthetic-aperture radars (SAR) and Inverse synthetic-aperture radars (ISAR) make use of Doppler frequency to generate high-resolution images of terrain and targets. The Doppler frequency shift also has been used in Doppler-navigation radar to measure the velocity of the aircraft carrying the radar system. The extraction of the Doppler shift in weather radars, moreover, allows the identification of severe storms and dangerous wind shear not possible by other techniques.

A Doppler radar is a specialized radar that uses the Doppler effect to produce velocity data about objects at a distance. It allows accurate measurements of the radial compo-

nent of a target's velocity relative to the radar. There are four types of Doppler radar: Coherent Pulsed, Continuous Wave (CW), Frequency Modulated, and Pulse-Doppler radar. Early Doppler radars were CW Doppler radars, which only provide a velocity output, as the received signal from the target is compared in frequency with the original signal. But it quickly led to the development of the Frequency Modulated (FM-CW) radar, which sweeps the transmitter frequency to encode and determine range. CW, or FM radar was developed during World War II for United States Navy aircraft, to support night combat operation. Both of the CW and FM-CW radars can only process one target, which limits their applications. With the advent of the digital techniques, Pulse-Doppler (PD) radars were introduced, and Doppler processors for coherent pulse radars were developed at the same time. The advantage of combining Doppler processing to pulse radars is to provide accurate velocity information. During 1970s, digital technology underwent a tremendous advance, digital fast Fourier transform (FFT) filtering became practical. This was immediately connected to coherent pulsed radars, where velocity information was extracted. Its usability was proved in both weather and air traffic control radars. Pulse-Doppler radars are widely used in airborne early warning, navigation, missile guidance, satellite tracking, battlefield reconnaissance, range measurement, weapon fire control and weather detection, and has become an important military equipment. An early warning aircraft equipped with a pulsed Doppler radar has become an effective military weapon against low-level bombers and cruise missiles. For example, airborne fire control systems mainly use Pulse Doppler radars, the APG-68 radar-equipped in American aircraft represents the advanced level of airborne pulse Doppler fire control radar. In addition, Pulse-Doppler radars are also used for meteorological observation. In addition to all the functions of the conventional weather radars, Doppler weather radars can also provide signals of atmospheric wind. The velocity distribution of atmospheric turbulence in different heights of the atmosphere can be obtained by Doppler velocity resolution of meteorological echoes.

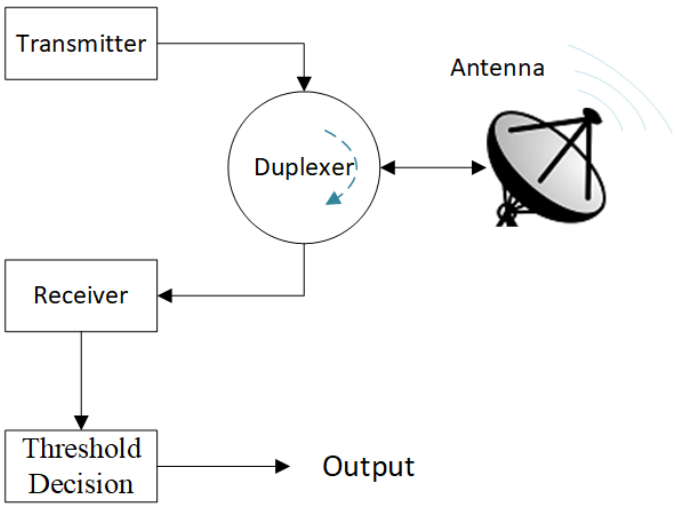


Figure 2.1: Structure of a monostatic radar system

2.1.2 Basics of radar

RADAR is the abbreviation for Radio Detection and Ranging System. It is basically an electromagnetic system used to detect the location and distance of an object from the point where the RADAR is placed. It works by radiating energy into space and monitoring echo or reflected signals from the objects. It operates in the ultra-high frequency (UHF) and microwave range.

2.1.2.1 The components of a radar

A radar system generally consists of a transmitter that produces an electromagnetic signal that is radiated into space by an antenna. When this signal strikes any object, it gets reflected or reradiated in many directions. This reflected echo or signal is received by the radar antenna that delivers it to the receiver, where it is processed to determine the geographical statistics of the object. The range is determined by calculating the time taken by the signal to travel from the radar to the target and back. The target's location is measured in angle, from the direction of maximum amplitude echo signal that the antenna points to. Doppler Effect is used to measure range and location of moving targets.

The major components of a radar system (shown in Figure 2.1) are listed below:

1. A transmitter: It can be a power amplifier like a Klystron, traveling wave Tube or a power oscillator like a magnetron. The signal is first generated by using a waveform generator and then amplified in the power amplifier.

2. Feed system: It refers to all components connecting the antenna to the transmitter or receiver. In a transmitting antenna, it includes all of the components involved conveying the RF electrical current into the radiating part of the antenna, where the current is converted to radiation; in a receiving antenna, it is the system that converts the electric currents already collected from incoming radio waves into a specific voltage to current ratio (impedance) needed at the receiver. In a radar system, a feed system might consist of a feed horn, orthomode transducer, polarizer, frequency diplexer, waveguide, waveguide switches, rotary joint, etc. [27].

3. Antenna: A common form of a radar antenna is a parabolic dish antenna fed from a feeding antenna at its focus. Phased-array antennas have also been used for radar. In a phased-array antenna, the antennas beam is scanned electronically by introducing phases to the phase shifters connected to elements. The functions of the antenna are to concentrate the transmitting signal into a narrow beam in a single preferred direction, intercept the target echo signal from the same direction [28].

4. Duplexer: A duplexer allows the antenna to be used as a transmitter or a receiver. It can be a gaseous device that would produce a short circuit at the input to the receiver when the transmitter is working.

5. Receiver: The function of the receiver is to detect wanted echo signals in the presence of noise, clutter, and interference [29]. It can be a superheterodyne type consisting of a low-noise RF amplifier, a mixer, an intermediate frequency (IF) amplifier, etc. Returned signals are enhanced by the low noise RF amplifier. The mixer converts the RF signal into an IF signal. The IF is the difference between

the target return frequency and the local oscillator frequency.

6. Threshold Decision: The output of the receiver is compared with a threshold to detect the presence of an object. If the output is below any threshold, the presence of noise is assumed.

2.1.2.2 Radar types

As shown in Figure 2.2, types of radar can be classified based on frequency, waveform, pulse repetition frequency (PRF), and application [30].

Radars can operate on different frequency bands: (i) high frequency (HF) band (3-30 MHz) radars; (ii) very high frequency (VHF) band (30-300 MHz) and ultra-high frequency (UHF) band (300-1000 MHz) radars; (iii) L-Band (1-2 GHz) radars; (iv) S-Band (2-4 GHz) radars; (v) C-Band (4-8 GHz) radars; (vi) X-Band (8-12 GHz) radars; (vii) Ku (12-18 GHz), K (18-26.5 GHz), Ka (26.5-40 GHz) band radars; (viii) infrared, visible light band (40-75 GHz) radars; (IX) W-Band (75-110 GHz) radars; (X) THz band

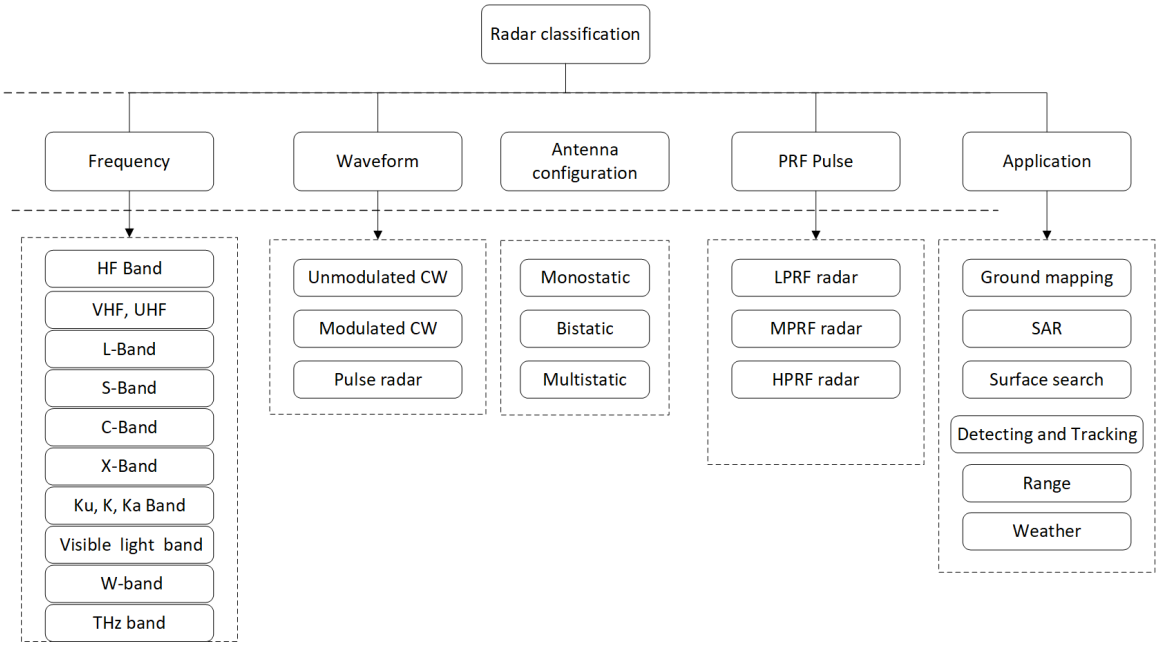


Figure 2.2: Radar types

(above 110 GHz).

Base on the waveform, radars can be classified as follows: (i) unmodulated CW radars; this kind of radar transmits a constant frequency with constant amplitude. The received echo signal has either exactly this frequency, or the echo signal is reflected on a moving target with a radial velocity of the reflector and it is then shifted by the amount of Doppler frequency. It only detects moving targets, as stationary targets (along with the line of sight) will not cause a Doppler shift. (ii) modulated CW radars, this kind of radar continuously transmits radio waves with varied frequencies at a known rate and the frequency of reflected signals is compared with the frequency of the transmitted signal. The frequency difference between the received signal and the transmit signal increases with delay, and hence with distance. (iii) A pulse radar is a radar device that emits short and powerful pulses and in the silent period receives the echo signals. In contrast to the continuous wave radar, the transmitter is turned off before the measurement is finished.

Based on the PRF, radars have three types. (i) Low PRF (LPRF), its parameters are designed so that range measurement is unambiguous. (ii) High PRF (HPRF) radars are ambiguous in range but can measure Doppler shifts without ambiguities. (iii) Medium PRF (MPRF) radars represent the intermediate case between low and high PRF radars. They are ambiguous in both range and Doppler. However, using MPRF is a good compromise for target detection because of the higher information update rate in comparison to LPRF. Based on antenna configurations, radars can be classified into three types. (i) Monostatic radar, the radar's transmitter and receiver share a common antenna. (ii) Bistatic radar, the radar receive antenna is located far from the radar transmit antenna. (iii) Multistatic radar, a multistatic radar system contains multiple spatially diverse monostatic radar or bistatic radar components with a shared area of coverage. A multistatic radar is an integration of multiple radars that provide multiple looks at a subject from different viewpoints [31].

Radars have lots of applications in many areas. Based on applications, radars can be classified into many types. In the military, the radar has three major applications.

In air defense, the radar is used for target detection, target recognition and weapon control (directing the weapon to the tracked targets). In a missile system, it can guide the weapon. It identifies the enemy location in a map. In air traffic control, the radar is used to control air traffic near airports, guide the aircraft to land in bad weather, and scan the airport surface for aircraft and ground vehicle positions. In remote sensing, radar can be used for observing weather or observing planetary positions and monitoring sea ice to ensure a smooth route for ships. In-ground traffic control, radar can also be used by traffic police to determine speed and range of the vehicle, controlling the movement of vehicles by giving warnings about the presence of other vehicles or any other obstacles behind them. In space, radar can be used to guide the space vehicle for a safe landing on the moon, observe the planetary systems, detect and track satellites, and monitor the meteors. With the development of radar, the application of radar is continuing extending.

2.1.3 Micro-Doppler

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 [32–34].

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:

$$f = f_0(1 + 2v/c), \tag{2.1}$$

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

$$f_D = f_0(2v/c), \tag{2.2}$$

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 [34]. The Doppler signature f_{Dsig} for such a complex object has multiple time-dependent frequency shifted components and it is defined as:

$$f_{Dsig}(t) = f_0 \sum_{i=1}^{N} 2v_i(t)/c, \tag{2.3}$$

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

$$\hat{S}_R(t) = e^{j2\pi f_0 t} e^{j2\pi f_{Dsig}(t)t}, \tag{2.4}$$

Mixing the received signal $\hat{S}_R(t)$ with the transmitted signal $\hat{S}_T(t)$ as follows [35]:

$$\hat{S}_R(t)\hat{S}_T(t)^* = e^{j2\pi f_{Dsig}t}. \tag{2.5}$$

This allows the extraction of the Doppler signature from the data. This is the component of the signal that contains the micro-Doppler information of the target and it can be used for target or activity recognition and classification. The bandwidth of the resulting signal is normally much smaller than the carrier frequency, because the radial speed of the target is very small compared to the speed of the light, producing lower Doppler frequencies [17]. It can be seen from Equation 2.2, only when $v > c/2$, the f_D can be higher than f_0. Any common observable moving target would present a much smaller speed than half of the speed of light, consequently the Doppler frequency would be normally much smaller than the carrier frequency, i.e., the bandwidth of the resulting signal is much smaller than the carrier frequency.

The micro-Doppler signature can be represented in a two-dimensional time-frequency

space using a Short Time Fourier Transform (STFT):

$$STFT(i,K) = \sum_{n=0}^{N-1} x_i(n) e^{-j2\pi(nK/N)}, \quad K = 0, \ldots, N-1 \tag{2.6}$$

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

$$x_i(n) = \hat{S}_R(n + i(N/2))w(n), \tag{2.7}$$

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

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

$$T_w = N/f_s, \tag{2.8}$$

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

$$v > c/2f_oT_w, \tag{2.9}$$

will be clearly visible [17]. From Eq. (2.9), 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 [17].

In this thesis, STFT has been used to generate time-frequency spectrograms for micro-Doppler signatures representation. Micro-Doppler signature will be used in Chapter 3 and Chapter 4 for human activity recognition.

2.2 LiDAR

This section introduces the history, basic concepts and the composition of LiDAR sensors. The types of LiDARs and their corresponding applications are presented.

2.2.1 The history of LiDAR

In 1930s, the first attempt was made to measure the composition of the atmosphere using sweeping searchlight beams. In 1938, pulses of light were used for the first time to measure cloud based heights. The acronym LiDAR for this kind of measurement technique was first introduced by Middleton and Spilhaus in 1953. Theodore Maiman built and demonstrated the first practical laser in 1960. Shortly after the invention of the laser, the Hughes Aircraft Company proposed the first LiDAR-like system in 1961. This system combined laser-focused image to calculate the distance by measuring the time for a signal to return in order to perform satellite tracking. In 1963, a large rifle-like laser rangefinder was produced for military targeting. It had a detection range of 7 miles and an accuracy of 15 feet. LiDAR was firstly applied in meteorology to measure clouds and pollution by the National Center for Atmospheric Research. In 1971, the general public realized the accuracy and usefulness of LiDAR systems due to that astronauts used a laser altimeter to map the surface of the moon during the Apollo 15 mission. In 1975, NASA launched the Airborne Oceanographic LiDAR (AOL) program. The French Starlette satellite was firstly launched for satellite laser ranging. The first commercial LiDAR system appear in 1980s. By the mid-1990s, manufactures of laser scanners were producing LiDAR sensors that were widely used for geographical mapping. Nowadays, with the development of robots and self-driving, LiDAR is increasingly being used in robots drones, self-driving cars for localisation, navigation and obstacle avoidance.

2.2.2 Basics of LiDAR

Light detection and ranging (LiDAR) is a surveying method that measures distance to a target by illuminating the target with laser light and measuring the reflected light with

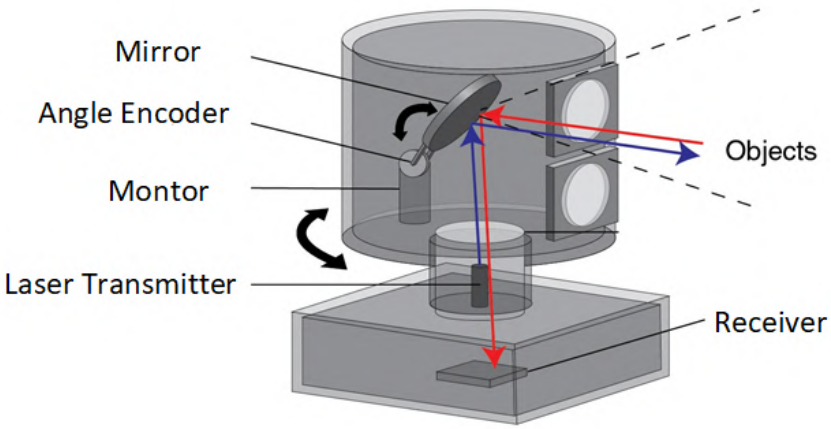


Figure 2.3: Basic structure of LiDAR [2]

a sensor.

2.2.2.1 The structure of LiDAR

The basic structure of a LiDAR system is shown in Figure 2.3. It consists of a transmitter, a receiver, control electronics, a mirror, an angle encoder, and a motor. It is a non-contact optical surveying system based on the principle of time-of-flight measurement. It calculates the distance of objects by the time it takes from emitting infrared lasers to receiving the reflected lasers. The maximum power of the laser transmitter is limited in order to make it eye-safe for the people around it. Wavelengths used in LiDAR depend on the application and extend from about $250nm$ to $11\mu m$. The Laser transmitters are categorized by their wavelength. The receiver reads and records the backscattered signal to the system. The mirror is rapidly rotated by the motor to perform scanning. The angle encoder converts the angular position or motion of a shaft or axle to analog or digital output signals.

2.2.2.2 The types of LiDAR

LiDAR systems can be classified based on platform, physical process, and scattering process, respectively.

Based on platform, LiDAR can be divided into ground-based LiDAR and airborne LiDAR. Ground-based LiDARs are placed on the Earth's surface. Ground-based LiDARs have many uses, including surveying and mapping, engineering and construction, archaeology, and self-driving cars. They acquire 3D point clouds that can be matched with digital images to create 3D models. Ground-based LiDARs can be either stationary or mobile. Mobile LiDARs are attached to a moving vehicle to collect data. For self-driving cars, gound-based LiDARs provide 3D point cloud of roads, pedestrians, and surrounding cars to help them perform navigation. Airbone LiDAR is attached to an aircraft. An Airborne LiDAR system consists of a laser scanner, an IMU (Inertial Measurement Unit), and a GPS. It creates a 3D model for a large landscape. Airbone LiDAR is a more accurate method of creating digital elevation models. One major advantage is that airbone LiDAR is able to filter out reflections from vegetation from the point cloud data.

Based on physical process, LiDAR can be divided into rangefinder LiDAR, differential absorption LiDAR (DIAL), and Doppler LiDAR. A rangefinder LiDAR uses a laser beam to determine the distance to an object based on time-of-flight principle by sending a laser pulse towards the object and measuring the time taken by the pulse to be reflected off the target and returned to the LiDAR. The operating principle of Doppler LiDAR is the same as Doppler radars. Doppler LiDAR measures the Doppler frequency shift of the backscatter signal. It has been used to monitor the changes of air flow such as wind speed, wind direction, etc.

DIAL is a laser remote-sensing technique that is used for range-resolved (profile) measurements of atmospheric gas concentrations [36]. Light is absorbed by many of the atmospheric components. Based on this, DIAL is used to measure the atmospheric concentrations of gases such as water vapor and ozone. The wavelength of the atmospheric backscatter echoes is sequentially alternated between two wavelengths (λ_{on} and λ_{off}), λ_{on} coinciding with an absorption line of the gas of internet, λ_{off} in the close vicinity but off the absorption line [37]. Assuming that the atmosphere and the detection system

have the same properties at the two wavelengths apart from the absorption due to the gas of interest, the range-resolved gas absorption profile can be measure from the ratio of the atmospheric echoes at two wavelenghts [37]:

$$\frac{P(\lambda_{on}, z)}{P(\lambda_{off}, z)} = K \exp\left\{-2 \int_0^z N(z^{'} \left[\sigma(\lambda_{on}) - \sigma(\lambda_{off})\right] dz^{'}\right\}, \tag{2.10}$$

where z is the distance, $P(\lambda_{on}, z)$ and $P(\lambda_{off}, z)$ are the backscattering signals at λ_{on} and λ_{off} at the distance z, $N(z)$ is the gas concentration at distance z, $\sigma(\lambda)$ is the gas absorption cross section, and K is the ratio of the transmitted laser power at on- and off- wavelengths. Thus, with known absorption cross sections at λ_{on} and λ_{off}, the gas distribution along the path of the laser beam can be retrieved through the DIAL equation:

$$N(z) = \frac{1}{2[\sigma(\lambda_{on}) - \sigma(\lambda_{off})]} \frac{d}{dz} \ln\left[\frac{P(\lambda_{off}, z)}{P(\lambda_{on}, z)}\right]. \tag{2.11}$$

These types of LiDARs have been applied in a wide variety of applications. LiDARs are widely applied in surveying and mapping to get 3D map and make digital elevation models. They also have been used in localisation and navigation on robots and drones. In autonomous vehicles, LiDARs are used for obstacle detection and avoidance to navigate safely through environments. In architecture, LiDAR can be used to capture buildings' structure. In the Atmospheric Physics, LiDAR is used to measure the concentration of oxygen, nitrogen, potassium, sodium and other gas particles in the middle and upper atmosphere. DIAL can be used to trace amount of gases above the hydrocarbon region, which helps to find the oil and gas deposits. More applications of LiDAR can be found in [38].

In this thesis, a 2D rangefinder LiDAR is implemented to perform localisation and tracking of people for collecting their trajectories related to different activities.

2.3 Machine learning

Generally, human activity recognition is achieved by performing classification upon sensor data. Activity classification is usually achieved by using machine learning. In machine learning and statistics, classification is used to identify a new observation belongs to which category based on a training set of data containing observations whose category membership is known [34]. In human activity recognition, the training dataset is collected by sensors, a machine learning algorithm is trained upon the sensor data to classify different activities. In recent years, deep learning has shown strong learning ability and achieved state-of-the-art in the fields of image recognition, speech recognition, game competition, etc. The hierarchical structure of deep learning enables automatic learning of the features from raw data without relying on feature extraction methods. This section reviews the machine learning algorithms implemented in the related work, then introduces the fundamental concepts of traditional machine learning algorithms including SVM and kNN, and deep learning including CNN and LSTM, which are applied in this thesis. SVM and kNN are compared with deep learning for human activity recognition in this thesis.

2.3.1 Machine learning in human activity recognition

A number of machine learning algorithms have been applied in human activity classification. Traditional classification methods such as decision tree, kNN, and SVM are trained to identify different activities using handcrafted features, which are extracted by a hand process and human visual judgement, not automatically by a computer system. For instance, in [39], the authors developed a kNN algorithm to classify the activities of walking, running, sitting and standing measurerd by mobile sensors. In [40], an SVM was built to classify handcrafted micro-Doppler features of seven different human activities. In [41], the authors used a decision tree to classify five activities collected by the accelerometer embedded in smartphones, such as staying still, walking, running, and going upstairs and downstairs. Some related work applied ensemble learning that combines

different algorithms to perform human activity classification [42, 43]. Deep learning also has been increasingly implemented in human activity classification and achieved very good results. Kim et al. [44] proposed the use of deep convolutional neural networks (DCNNs) for human detection and activity classification. A DCNN was implemented upon the sample spectrograms of four types of targets (human, dog, horse, car) and seven human activities that measured by a Doppler radar operating at 7.25 GHz, and it achieved 97.6% and 90.9% accuracy respectively. In [45], the authors built a bidirectional LSTM network to classify accelerometer and gyroscope data obtained from a smartphone, and an overall accuracy of 92.67% was achieved upon six human activities including sitting, standing, laying, walking, walking upstairs, and walking downstairs.

2.3.2 Support vector machine (SVM)

SVMs are a set of supervised learning algorithm which can be used for both classification and regression analysis. A support vector machine finds a hyperplane or a set of hyperplanes in a high- or infinite-dimensional space and selects the hyperplanes that maximize the distance between the nearest training samples and the hyperplanes.

Given a training dataset of N data points $\{x_i, y_i\}_i^N$, where $x_i \in R^n$ is the i_{th} input and $y_i \in \{-1, 1\}$ is the i_{th} output. The aim is to find a separating hyperplane, which divides the group of points x_i for which $y_i = 1$ from the group of points for which $y_i = -1$ with the maximum distance between the hyperplane and the nearest point x_i from either group.

Any hyperplane can be written as the set of points X satisfying

$$W^T \cdot X - b = 0, \tag{2.12}$$

where W is the normal vector to the hyperplane. The parameter $\frac{b}{\|W\|}$ determines the offset of the hyperplane from the origin along the vector W.

As shown in Figure 2.4, if the training data is linearly separable, there are two parallel

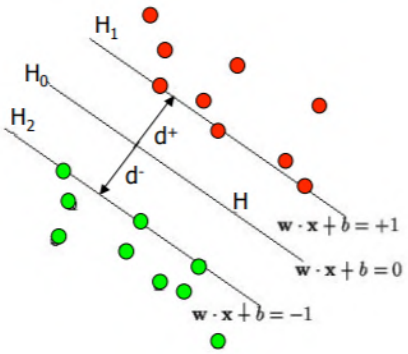


Figure 2.4: Maximum-margin hyperplane and margins for an SVM trained with samples from two classes [3]

hyperplanes that separate two classes of data, so that the distance between them is as large as possible. The region bounded by these two hyperplanes is called the margin, and the maximum-margin hyperplane is the hyperplane that lies halfway between them [46]. These hyperplanes can be described by the equations:

$$W^T \cdot X - b = 1, \tag{2.13}$$

and

$$W^T \cdot X - b = -1. \tag{2.14}$$

To maximize the distance between these two hyperplanes is equally to minimize $\|W\|$. To prevent data points from falling into the margin, the following constraint is added: for each i

$$\begin{aligned} W^T \cdot x_i - b \geq 1, \ if \ y_i = 1, \\ W^T \cdot x_i - b \leq -1, \ if \ y_i = -1, \end{aligned} \tag{2.15}$$

which is equivalent to

$$y_i(w^T \cdot x_i - b) \geq 1, \ for \ all \ 1 \leq i \leq N. \tag{2.16}$$

So the optimization problem can be described as minimizing $\|W\|$ subject to $y_i(w^T \cdot$

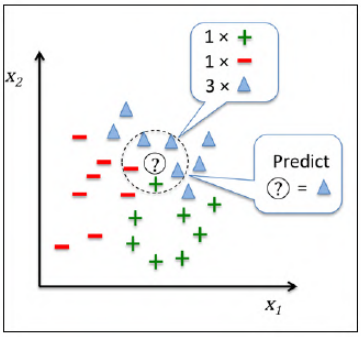


Figure 2.5: kNN (k=5) classification [4]

$x_i - b) \geq 1$, for $i = 1, \cdots, n$.

To extend SVM to cases in which the data are not linearly separable, the hinge loss function is introduced here [46],

$$max(0, 1 - y_i(\vec{w} \cdot \vec{x_i} - b)). \tag{2.17}$$

This function is zero if the constraint in Equation 2.16 is satisfied, in other words if $\vec{x_i}$ lies on the correct side of the margin. For data on the wrong side of the margin, the function's values if proportional to the distance from the margin. It is expected to minimize

$$\left[\frac{1}{n}\sum_{i=1}^{n} max(0, 1 - y_i(\vec{w} \cdot \vec{x_i} - b))\right] + \lambda \left\|\vec{w}\right\|^2, \tag{2.18}$$

where the parameter λ determines the tradeoff between increasing the margin-size and ensuring that $\vec{x_i}$ lie on the correct side of the margin.

2.3.3 K-nearest neighbors (kNN)

KNN is one of the 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.

The basic idea as shown in Figure 2.5 which presents a 5-nearest neighbour classifier

on a three-class problem in a two-dimensional feature space. In this example, the decision for ⓘ is simple. It has three neighbours of one class Δ, one neighbour of class $-$, and one neighbour of class $+$. This can be resolved by simple majority voting or by distance weighted voting. Because class Δ is the majority of its neighbours, so the class of ⓘ is decided as Δ.

So kNN classification has three stages, which can be summarized as follows:

1. Choose the number of k and a distance metric, which is commonly based on the Euclidean distance.

2. Find the k nearest neighbours of the sample that needs to be classified.

3. Assign a class label by majority vote.

In this thesis, both SVM and kNN have been trained on micro-Doppler features for human activity recognition. SVM is also applied in trajectory-based activity classification. The results obtained by them are compared with that obtained by deep learning.

2.3.4 Deep learning

This section describes the fundamental of artificial neural networks, then introduces CNNs that used in micro-Doppler signatures classification and LSTM that used in LiDAR data classification in detail.

2.3.4.1 Artificial neural network (ANN)

An artificial neural network is a biologically inspired computational model formed from hundreds of single units, artificial neurons, connected with coefficients (weights) which constitute the neural structure [47]. An artificial neuron in a neural network is called a perceptron. Perceptrons were developed in the 1950s and 1960s by the scientist Frank Rosenblatt [48]. The perceptron was designed for a specific type of machine Learning problems: binary classification problems. A perceptron takes several binary inputs, $x_1, x_2, ...$, and produce a single binary output.

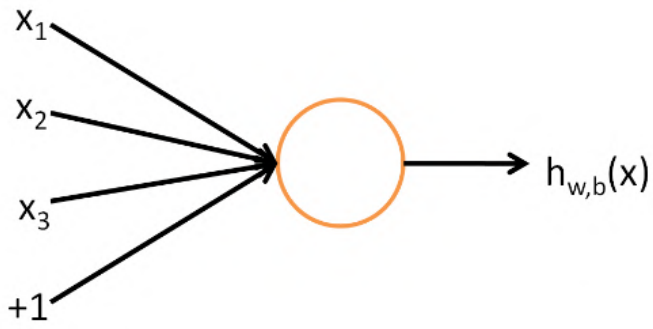


Figure 2.6: Perceptron

Figure 2.6 presents a perceptron with three inputs x_1, x_2, x_3, and '+1' is a bias unit. In general, it could have more or fewer inputs. Weights, $w_1, w_2, ...$, are introduced in order to express the importance of the respective inputs to outputs. An output can be formularized as:

$$h_{w,b}(x) = f(w^T x) = f(\sum_{i=1}^{3} w_i x_i + b), \tag{2.19}$$

where $f : \Re \mapsto \Re$ is called an activation function, b is the bias associated with the feature x. As shown in the subsequent equations, there are several available activation functions, such as tanh, sigmoid, and Rectified Linear Units (ReLU), in neural networks. Equation 2.20 shows the equations of three activation functions and Figure 2.7 presents the plots of tanh, sigmoid, and ReLU functions.

$$\begin{aligned} &sigmoid : f(z) = \frac{1}{1+e^{-z}}, \\ &tanh : f(z) = tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}, \\ &relu : f(z) = max(0, z). \end{aligned} \tag{2.20}$$

The $tanh(z)$ function is a rescaled version of the sigmoid, and its output range is $[-1, 1]$ instead of $[0, 1]$. The rectified linear function is a piece-wise linear and it saturates at exact point 0 whenever the input z is less than 0.

A neural network is created by hooking together many of simple perceptrons with a hierarchical structure, the output of the perceptrons in one layer can be the input of

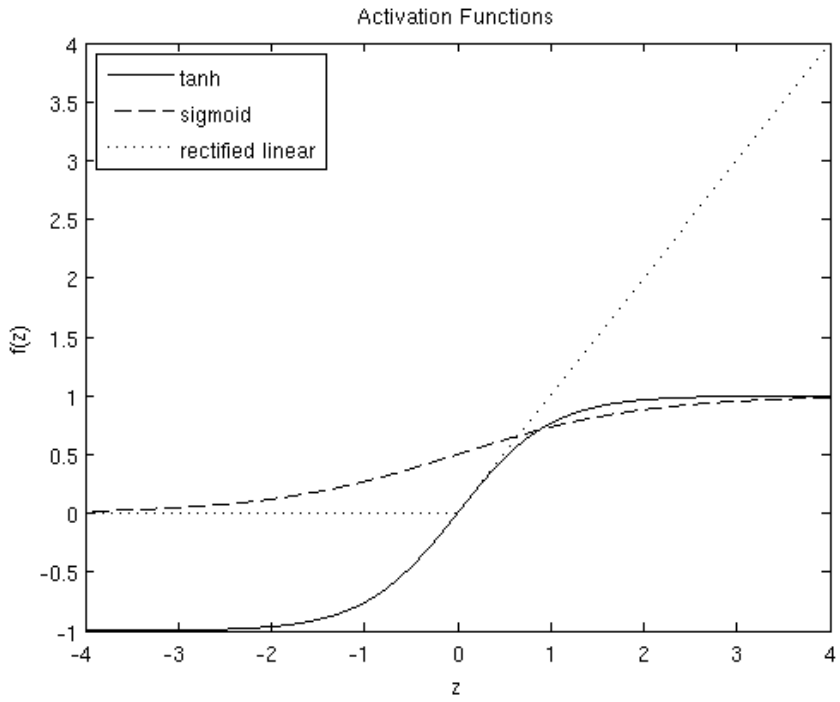


Figure 2.7: Activation functions

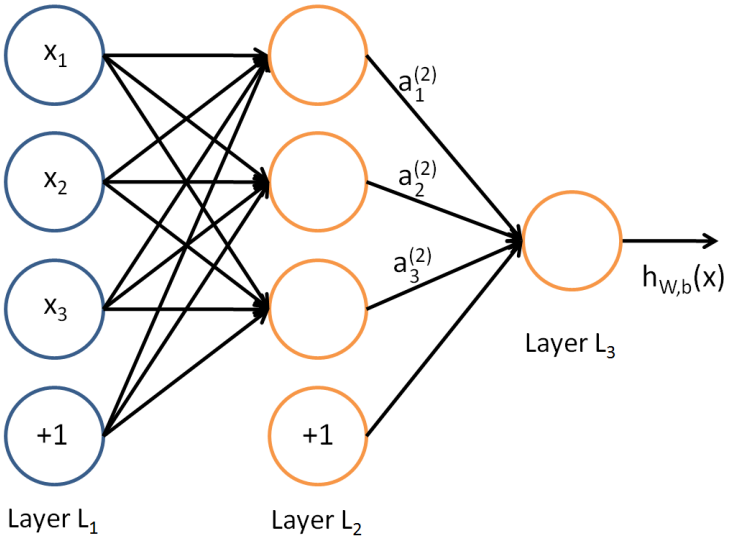


Figure 2.8: Neural network model

the perceptrons in another layer. Figure 2.8 shows a small neural network. The leftmost layer of the network is called the input layer, and the rightmost layer the output layer (which, in this example, has only one node). The middle layer of nodes is called the hidden layer, because its values are not observable in the training set. $a_i^{(l)}$ denotes the

activation (output value) of unit i in layer l. For $l = 1$, $a_i^{(1)} = x_i$. Given a fixed setting of the parameters W,b, a hypothesis $h_{W,b}(x)$ can be defined as a real number output. Specifically, the computation that this neural network represents is given by:

$$\begin{aligned} a_1^{(2)} &= f(W_{11}^{(1)}x_1 + W_{12}^{(1)}x_2 + W_{13}^{(1)}x_3 + b_1^{(1)}), \\ a_2^{(2)} &= f(W_{21}^{(1)}x_1 + W_{22}^{(1)}x_2 + W_{23}^{(1)}x_3 + b_2^{(1)}), \\ a_3^{(2)} &= f(W_{31}^{(1)}x_1 + W_{32}^{(1)}x_2 + W_{33}^{(1)}x_3 + b_3^{(1)}), \\ h_{W,b}(x) = a_1^{(3)} &= f(W_{11}^{(2)}a_1^{(2)} + W_{12}^{(2)}a_2^{(2)} + W_{13}^{(2)}a_3^{(2)} + b_1^{(2)}). \end{aligned} \tag{2.21}$$

Let $z_i^{(l)}$ denote the total weighted sum of inputs to unit i in layer l, the equations above can be written more compactly as:

$$\begin{aligned} z^{(2)} &= W^{(1)}x + b^{(1)}, \\ a^{(2)} &= f(z^{(2)}), \\ z^{(3)} &= W^{(2)}a^{(2)} + b^{(2)}, \\ h_{W,b}(x) &= a^{(3)} = f(z^{(3)}). \end{aligned}$$

Similarly, the number of hidden layers in a neural network can be extended to more than one.

2.3.4.1.1 The backpropagation algorithm

The backpropagation algorithm is one of the most popular methods of training artificial neural networks, it propagates error values backwards from the output, and uses these error values to calculate the gradient of loss function with weights in the network, then feeds this gradient to an optimization method, which in turn uses it to update the weights, in an attempt to minimize the loss function.

For a single training example (x, y), a cost function can be defined as:

$$J(W, b; x, y) = \frac{1}{2} \left\| h_{W,b}(x) - y \right\|^2, \tag{2.22}$$

it is a squared-error cost function. Given a training set of m examples, $\left\{(x^{(1)}, y^{(1)}), \cdots, (x^{(m)}, y^{(m)})\right\}$, then the overall cost function is defined as:

$$\begin{aligned} J(W, b) &= [\frac{1}{m} \sum_{i=1}^{m} J(W, b; x^{(i)}, y^{(i)})] + \frac{\lambda}{2} \sum_{l=1}^{n_l - 1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \\ &= [\frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} \left\| h_{W,b}(x) - y \right\|^2] + \frac{\lambda}{2} \sum_{l=1}^{n_l - 1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2, \end{aligned} \tag{2.23}$$

The first term in the definition of $J(W, b)$ is an average sum-of-squares error term. The second term is a regularization term (also called a weight decay term) that tends to decrease the magnitude of the weights, and helps prevent overfitting. The weight decay parameter controls the relative importance of the two terms.

The goal is to minimize $J(W, b)$ as a function of W and b. One iteration of gradient descent updates the parameters W, b as follows:

$$\begin{aligned} W_{ij}^{(l)} &= W_{ij}^{(l)} - \alpha \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b), \\ b_i^{(l)} &= b_i^{(l)} - \alpha \frac{\partial}{\partial b_i^{(l)}} J(W, b), \end{aligned} \tag{2.24}$$

where α is the learning rate, its value range is $(0, 1)$. The key step is computing the partial derivatives above. The derivative of the overall cost function $J(W, b)$ can be computed as:

$$\begin{aligned} \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b) &= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b; x^{(i)}, y^{(i)}) + \lambda W_{ij}^{(l)}, \\ \frac{\partial}{\partial b_i^{(l)}} J(W, b) &= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial b_i^{(l)}} J(W, b; x^{(i)}, y^{(i)}). \end{aligned} \tag{2.25}$$

For each node i in layer l, define $\delta_i^{(l)}$ that measures how much that node was "responsible" for any errors in the output. For an output node, the difference between the network's activation and the true target value can be defined as $\delta_i^{(n_l)}$. For each output unit i in layer n_l (the output layer), set

$$\delta_i^{(n_l)} = \frac{\partial}{\partial z_i^{n_l}} \frac{1}{2} \left\| y - h_{W,b}(x) \right\|^2 = -(y_i - a_i^{(n_l)}) \cdot f'(z_i^{(n_l)}), \tag{2.26}$$

then, each node i in layer l, where $l \in [n_l - 1, n_l - 2, n_l - 3, \cdots, 2]$, set

$$\delta_i^{(l)} = (\sum_{j=1}^{s_{l+1}} W_{ji}^l \delta_j^{(l+1)}) f'(z_i^{(l)}). \tag{2.27}$$

Compute the desired partial derivatives, which are given as:

$$\begin{aligned} \nabla_{W^{(l)}} J(W, b; x, y) &= \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b; x, y) = a_j^{(l)} \delta_i^{(l+1)}, \\ \nabla_{b^{(l)}} J(W, b; x, y) &= \frac{\partial}{\partial b_i^{(l)}} J(W, b; x, y) = \delta_i^{(l+1)}. \end{aligned} \tag{2.28}$$

Finally, one iteration of batch gradient descent can be implemented as follows:

1. Set $\Delta W^{(l)} := 0, \Delta b^{(l)} := 0$ (matrix/vector of zeros) for all l.
2. For $i = 0$ to m,
   (a) Use the backpropagation method to compute $\nabla_{W^{(l)}} J(W, b; x, y)$ and $\nabla_{b^{(l)}} J(W, b; x, y)$.
   (b) Set $\Delta W^{(l)} := \Delta W^{(l)} + \nabla_{W^{(l)}} J(W, b; x, y)$.
   (c) Set $\Delta b^{(l)} := \Delta b^{(l)} + \nabla_{b^{(l)}} J(W, b; x, y)$.
3. Update the parameters:

$$\begin{aligned} W^{(l)} &= W^{(l)} - \alpha[(\frac{1}{m} \Delta W^{(l)}) + \lambda W^{(l)}], \\ b^{(l)} &= b^{(l)} - \alpha[(\frac{1}{m} \Delta b^{(l)})]. \end{aligned} \tag{2.29}$$

To train a neural network, the above steps of gradient descent can be repeatedly implemented to reduce the cost function J(W,b).

2.3.4.2 Convolutional neural networks (CNNs)

Convolutional Neural Network (CNN) is one of the variants of artificial neural networks widely used in the field of computer vision. In this thesis, a CNN is built to classify frequency spectrograms generated from radar signals. CNNs have an associated terminology and a set of concepts that are unique to them, and that set them apart from other types of neural network architectures. The main ones are detailed in this section.

2.3.4.2.1 Input/Output Volumes

CNNs are usually applied to image data. Every image is a matrix of pixel values. The range of values that can be encoded in each pixel depends upon its bit size. The possible range of values a single pixel can represent is $[0, 255]$. However, with coloured images, particularly RGB (Red, Green, Blue)-based images, the presence of separate colour channels (3 in the case of RGB images) introduces an additional 'depth' field to the data, making the input 3-dimensional. As shown in Figure 2.9, for a given RGB image with $4 \times 4(Width \times Height)$ pixels, it has 3 matrices , one for each of the colour channels. Thus the image in its entirety, constitutes a 3-dimensional structure called the Input Volume $(4 \times 4 \times 3)$.

2.3.4.2.2 Filters (Convolution Kernels)

A filter (or kernel) is an integral component of the layered architecture. Generally, it refers to an operator applied to the entirety of the image such that it transforms the information encoded in the pixels. In practice, however, a kernel is a smaller-sized matrix in comparison to the input dimensions of the image, that consists of real valued entries.

The kernels are convolved with the input volume to obtain so-called activation maps or feature maps. Activation maps indicate 'activated' regions, i.e. regions where features specific to the kernel have been detected in the input. The real values of the kernel matrix

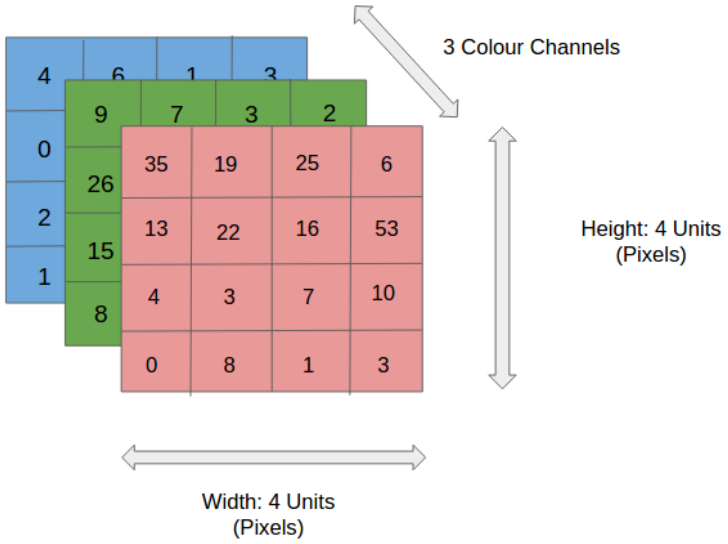


Figure 2.9: The cross-section of an input volume of size: $4 \times 4 \times 3$.

change with each learning iteration over the training set, indicating that the network is learning to identify which regions are of significance for extracting features from the data.

2.3.4.2.3 Receptive Field

It is impractical to connect all neurons with all possible regions of the input volume. It would lead to too many weights to train, and produce a too high computational complexity. Thus, instead of connecting each neuron to all possible pixels, a 2-dimensional sub-region called receptive field extending to the entire depth of the input, within which the encompassed pixels are fully connected to the following convolutional layer. It's over these small regions that the network layer cross-sections (each consisting of several neurons (called depth columns)) operate and produce the activation map. Figure 2.10 presents a receptive field 2×2 field.

2.3.4.2.4 Architecture of convolutional neural networks

Generally, a CNN contains three typical layers: *Convolutional Layers*, *Pooling Layers*, and *Fully-Connected Layers* [49]. As illustrated in Figure 2.11, these layers are stacked to form a full CNN architecture. These three types of layers are detailed further below:

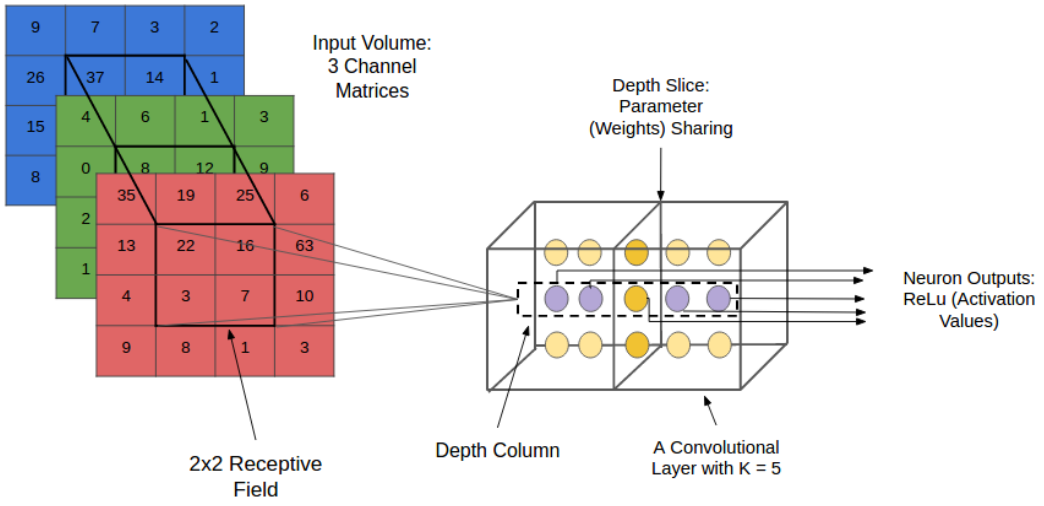


Figure 2.10: Concept of Receptive Field

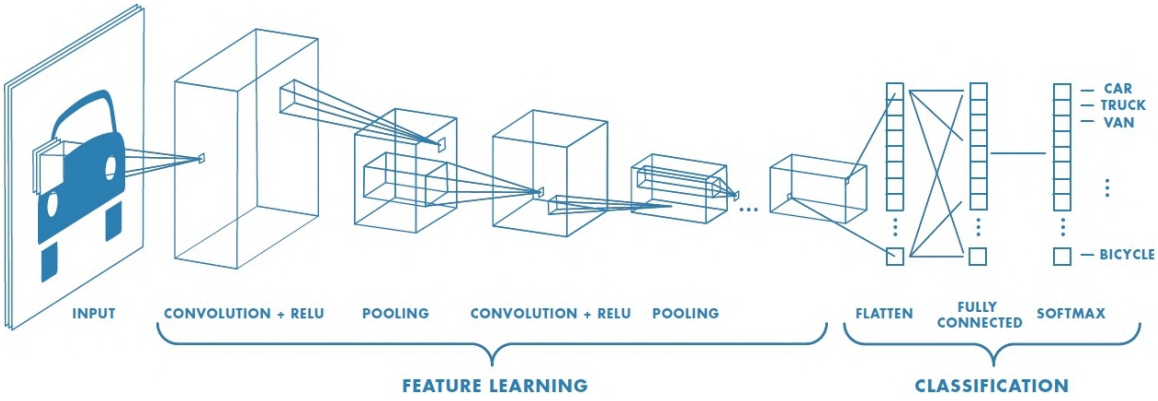


Figure 2.11: Architecture of a CNN

The convolutional layer is the main fundamental layer of a CNN. Each convolutional layer performs an aggregational operation aimed to learn the feature representation of the input and reduce the number of learnable parameters; it operates on the input/feature maps by applying a filter (e.g. kernel function) over the input. A convolution applies a linear operation on the input/feature maps using a set of filters F. The filter size is a matrix $G \times G$ of learnable parameters, on an input feature map x, it produces an output feature map $x^{'}$ as

$$x^{'} = \sum_{i=1}^{k} \sum_{r=1}^{G} \sum_{c=1}^{G} x_{rc} \cdot F_i + b^{'}, \tag{2.30}$$

where k is the number of filters F, $b^{'}$ is the bias parameter associated with the feature map $x^{'}$.

By convolving the input with the filter, a set of feature maps is produced. A convolutional layer is parameterized by a ***depth***, a ***kernel field size***, a ***stride***, and a ***zero-padding***. The depth determines the channel size of the output (i.e. the desired number of feature maps). The kernel field size, i.e. the filter size, covers a small region of pixels of the input image at each convolution step (see the bold square in Figure 2.12). Each element in a feature map is generated by the filter convolving with the covered region of the input (e.g. the generated elements in the feature map in Figure 2.12). The stride determines the number of pixels by the filter which is moved during each filtering step over the image. As can be seen in Figure 2.12 by the squares (representing the filter) over the input image. The zero-padding is used to pad the input with zeros on its border, as seen in Figure 2.12. The zero-padding is useful to determine a desired output size of the feature map. For example, given an input image of height and width $H \times W$, a kernel field size of $G \times G$, a stride S, and an amount of zero padding P, the spatial size ($H^{'} \times W^{'}$) of the feature map generated can be computed as

$$\begin{aligned} H^{'} &= \left(\frac{H - G + 2P}{s} + 1\right), \\ W^{'} &= \left(\frac{W - G + 2P}{s} + 1\right). \end{aligned} \tag{2.31}$$

Figure 2.12 provides a numerical example of the convolution process: assuming an image with the size of 4×4 pixels that is zero-padded with value 1, then convolved by a kernel of 3×3 with a stride of 2, a resulting feature map of 2×2 neurons is produced.

An image usually contains three channels (R, G, B), which also means the depth of an image is 3. For convolving with an image, the kernel is also initialize with 3 channels that is the same as the depth of the image . The convolution process with a 3×3 filter/kernel on an image is shown as Figure 2.13. The channel 1, 2, 3 of the filter convolves with the channel 1, 2, 3 of the image respectively. The final output value is generated by

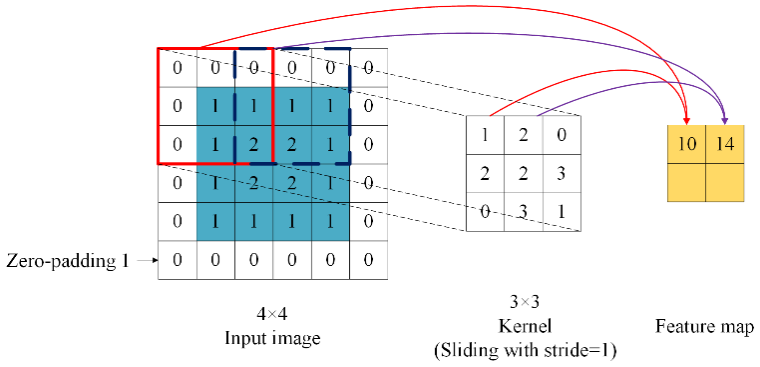


Figure 2.12: convolution process

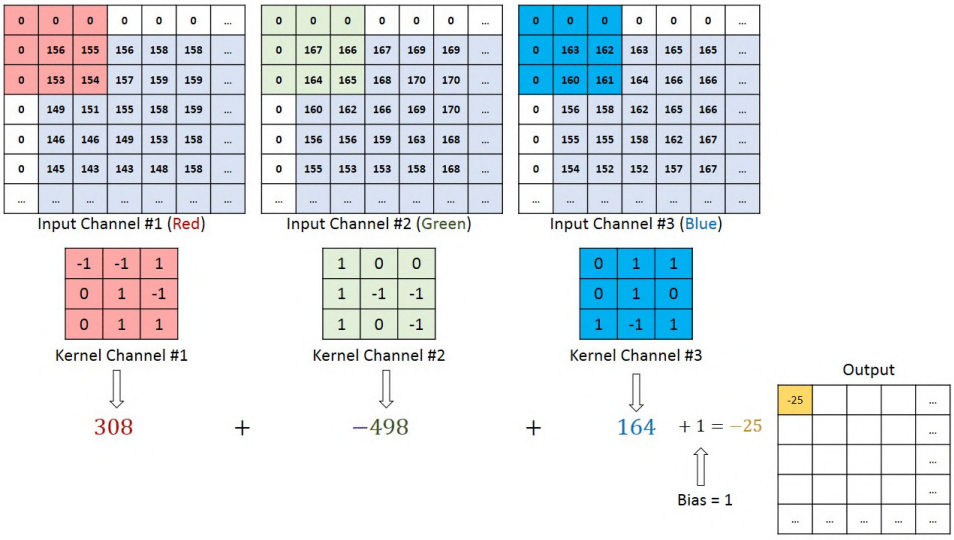


Figure 2.13: Multi-channel convolution in CNNs

summing the output on each channel and the bias. For an input with a different depth, the convolution process is similar, which the depth of the kernel is initialized as the same depth of the input.

The pooling layer is a form of non-linear down-sampling. It is designed to reduce the spatial size (dimensionality) of the input, in order to reduce the number of parameters (e.g. neurons and their connectivity) in the CNN. It aggregates the values of a local region of the input (window), commonly by applying an average or a maximum function. Max pooling is the most common function to implement pooling by using the maximum function. It partitions the input image into a set of non-overlapping windows. For

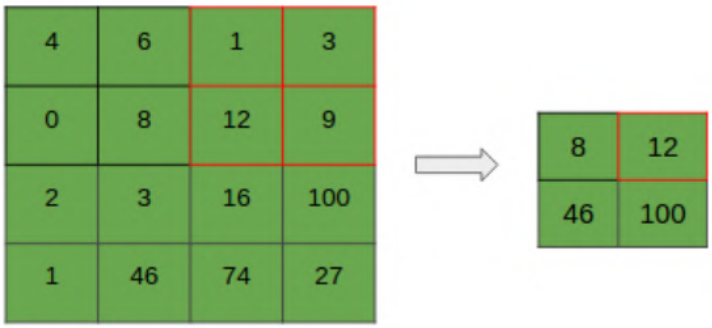


Figure 2.14: The Max-Pooling operation

each such window, it takes the maximum neuron value of that window and places it as an output. It is common to periodically insert a pooling layer between successive convolutional layers in a CNN architecture.

Figure 2.14 presents the max-pooling implemented on an image with the volume of size $4 \times 4 \times 1$, the sliding window is of size 2×2, it slides forward and down with a stride value of $[2, 2]$, the values in the sliding window are down-sampled to the maximum value.

The fully connected layer follows after several convolutional and max pooling layers to perform high-level reasoning in the neutral network. Perceptrons in a fully connected layer have full connections to all the input in the previous layer, as seen in regular Neural Networks.

CNNs model animal visual perception, has been widely applied to image recognition or classification tasks. In this research, frequency spectrograms generated from micro-Doppler signals can be taken as images with one channel, CNNs are very suitable to classify these frequency spectrograms related to different human activities.

2.3.4.3 Recurrent neural networks (RNNs)

Recurrent neural networks are a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence [50]. Figure 2.15 illustrates the basic structure of RNNs. Just the same as an ANN, a RNN consists of an input layer, one or more hidden layers, and an output layer. As shown in the unfolded

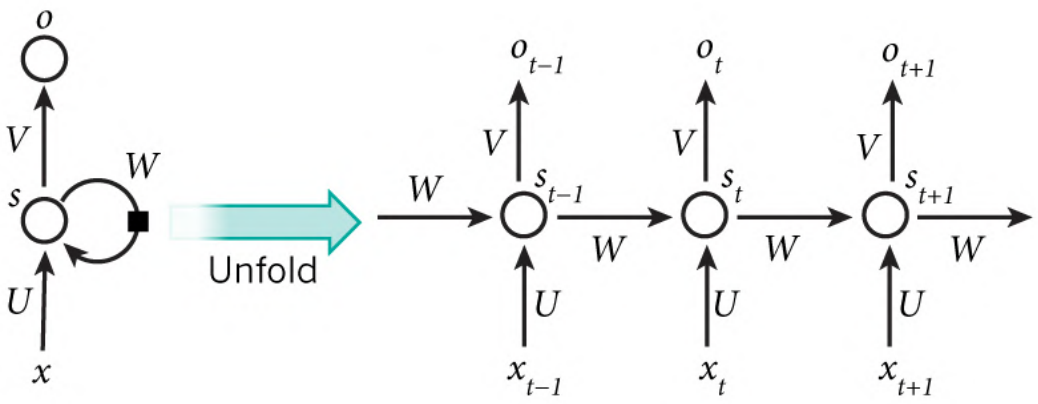


Figure 2.15: Unfolded basic recurrent neural network [5]

structure (the left side of Figure 2.15), each node in a given layer is connected with a directed connection to every other node in the next successive layer.

Given an input sequence $x = (x_0, ..., x_{T-1})$, the hidden states of a recurrent layer $s = (s_0, ..., s_{T-1})$ and the output of a single hidden layer RNN $o = (o_0, ..., o_{T-1})$ is formulated as follows:

$$\begin{aligned} s^t &= f(U_{xs}X^t + W_{ss}s^{t-1} + b_s), \\ o^t &= g(V_{so}s^t + b_o), \end{aligned} \tag{2.32}$$

where U_{xs}, W_{ss}, V_{so} denote the connection weights from the input layer x to hidden layer state s, the hidden layer s to itself and the hidden layer s to the output layer o, respectively. b_s and b_o are bias vectors. $f(\cdot)$ and $g(\cdot)$ are the activation functions in the hidden layer and the output layer.

2.3.4.4 Long short-term memory (LSTM)

Long short-term memory is a variant of RNNs. It is invented to address the exploding and vanishing gradients problem of RNNs [51]. This problem arises when back-propagating through RNNs during training, especially for networks with deep layers. The gradients have to go through continuous matrix multiplications during the back-propagation process due to the chain rule, causing the gradient to either shrink exponentially (vanish)

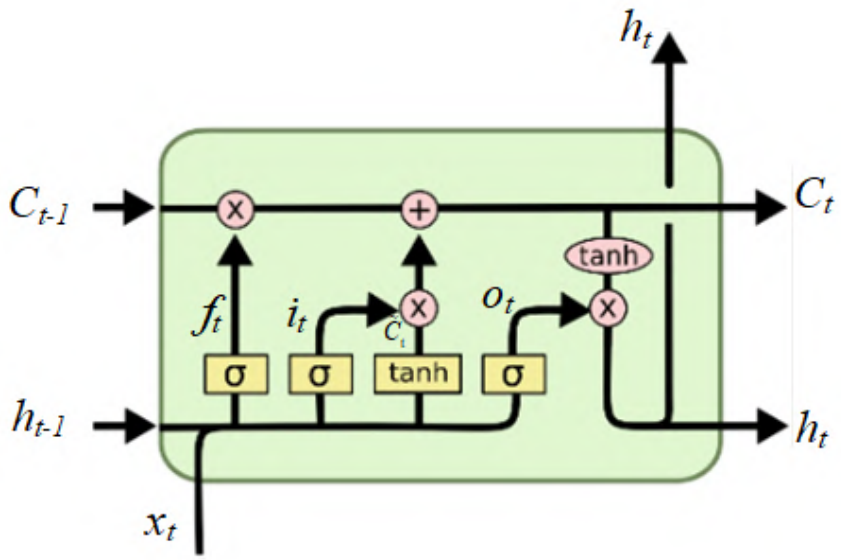


Figure 2.16: The structure of an LSTM unit [6]

or erupt exponentially (explode) [52]. To overcome the problem, LSTM uses a gating mechanism within each LSTM cell. In a normal RNN cell, the input at a time-step and the hidden state from the previous time step is passed through a *tanh* activation function to obtain a new hidden state or output. However, LSTM uses a few gates to control the passing of the information along the sequence and thus can learn long-range dependencies.

As shown in Figure 2.16, a typical LSTM unit contains an input gate i_t, a forget gate f_t, a cell c_t, an output gate o_t and an output response h_t. The input gate controls the extent to which a new value flows into the cell, the forget gate controls the extent to which a value remains in the cell, and the output gate controls how much information

form the cell is passed to the output h_t. The recursive computation of an LSTM unit is

$$\begin{aligned} i_t &= \delta(W_{xi}x_{x_t} + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \\ f_t &= \delta(W_{xf}x_{x_t} + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f), \\ c_t &= f_t \odot c_{t-1} + i_t \odot tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \\ o_t &= \delta(W_{xo}x_{x_t} + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o), \\ h_t &= \delta \odot tanh(c_t), \end{aligned} \tag{2.33}$$

where $\odot$ denotes element-wise product, $\delta(x)$ is the sigmoid function defined as $\delta(x) = 1/(1 + e^{-x})$, $W_{\alpha\beta}$ is the weight matrix between α and β (e.g., W_{xi} is the weight matrix from the input x_t to the input gates i_t), and b_β denotes the bias term of β with $\beta \in i, f, c, o$.

Notably, there exist several variations on the basic LSTM architecture. A common one is creating peephole connections that allow the gates to not only depend on the previous hidden state h_{t-1}, but also on the previous internal state c_{t-1} by adding an additional term in the gate equations. Figure 2.17 presents an LSTM unit with peephole connections. Peephole connections allow the gates to access the constant error carousel (CEC), whose activation is the cell state. The computation of a peephole LSTM unit is

$$\begin{aligned} i_t &= \delta(W_{xi}x_{x_t} + W_{hi}c_{t-1} + W_{ci}c_{t-1} + b_i), \\ f_t &= \delta(W_{xf}x_{x_t} + W_{hf}c_{t-1} + W_{cf}c_{t-1} + b_f), \\ c_t &= f_t \odot c_{t-1} + i_t \odot tanh(W_{xc}x_t + b_c), \\ o_t &= \delta(W_{xo}x_{x_t} + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o), \\ h_t &= tanh(\delta \odot c_t). \end{aligned} \tag{2.34}$$

LSTM has been applied to classify human trajectories collected by a LiDAR in this research in order to recognize human activities.

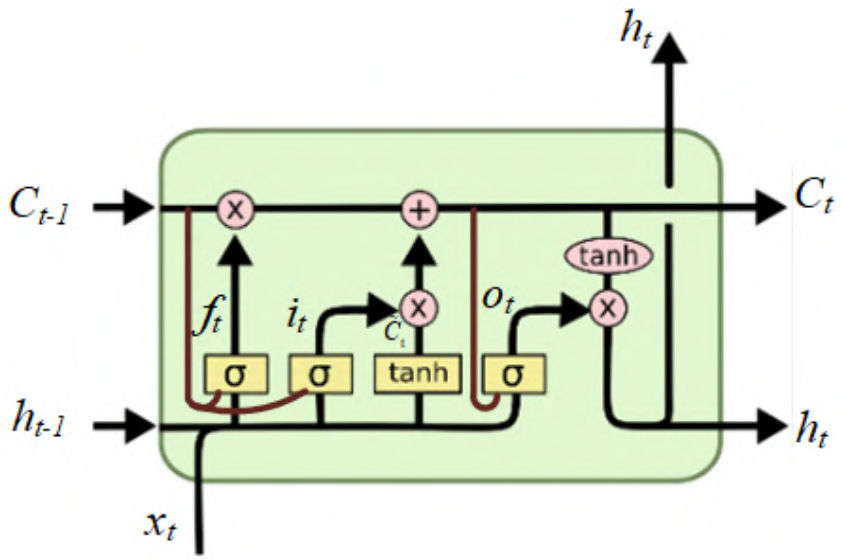
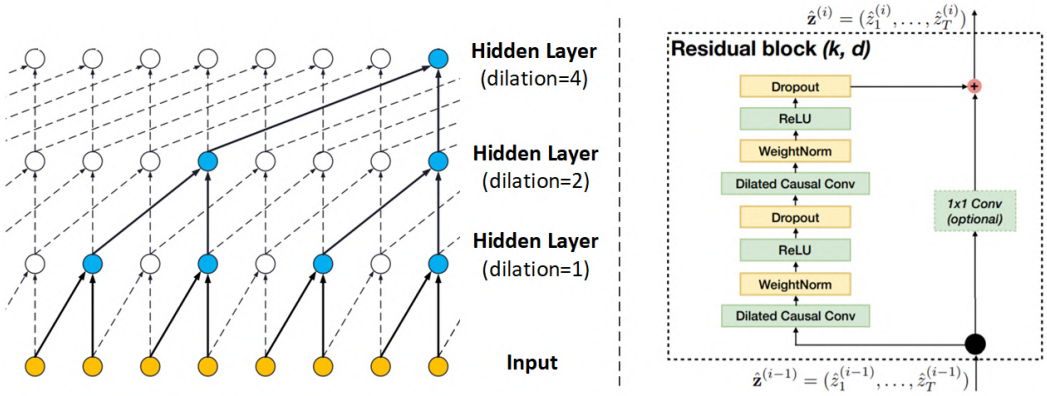


Figure 2.17: A peephole LSTM unit



(a) (b)

Figure 2.18: Temporal convolutional network, (a) dilate casual convolution, (b) residual block

2.3.4.5 Temporal convolutional network (TCN)

TCN is a class of time-series neural network models that capture long-range patterns using a hierarchy of temporal convolutional filters [53]. TCNs integrate dilated casual convolutions and the residual block structure to expand the receptive field and increase the depth. In dilated casual convolutions, an output at time t is convolved only with the elements from time t and earlier in the previous layer by using dilated convolution to

enable an exponentially large receptive field [54]. For a 1-D sequence input $x \in \mathbb{R}^n$ and a filter $f : 0, f : \{0, \cdots, k-1\} \rightarrow \mathbb{R}$, the dilated convolution operation F on element s of the sequence is defined as [54]

$$F(s) = (x *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i}, \tag{2.35}$$

where d is the dilation factor,k is the filter size, and $s - d \cdot i$ accounts for the direction of the past. With dilated casual convolutions, it is able to increase the receptive field of the TCN by using larger filter sizes k and increasing the dilation factor d. Figure 2.18(a) presents a dilated causal convolution with dilation factors $d = 1, 2, 4$ and filter size $k = 3$.

A TCN is consisted of several residual blocks. Each residual block contains a series of transformations τ. As shown in Figure 2.18(b), a residual block consists of two sets of layers including a dilated casual convolution layer, a weight normalization layer, a ReLU activation layer, and a dropout layer. The output of each residual block is the sum of the output of these transformations and the input x:

$$o = Activation(x + \tau(x)). \tag{2.36}$$

This is also called skip connection. Residual blocks avoid the vanishing gradient problem by carrying gradient throughout the extent of very deep network.

2.3.4.6 Regularization techniques in neural networks

One of the most common problem in machine learning is to aviod overfitting. Overfitting is a machine learning model fits too well to the training set but does not perform as well on the test set. Deep learning is very prone to overfitting since the model is able to adapt freely to large amount of complex data. Regularization is a technique which is used to solve the overfitting problem in machine learning. The most common methods of regularization are weight penalty (L1 and L2), data augmentation, dropout, and early

stopping.

2.3.4.6.1 Weight penalty

Weight penalty reduces the complexity of a model by penalizing larger individual weights, since large weights will result in a higher loss. A large weight has the effect of making the output of the network sensitive to small changes in the feature that it corresponds to because changes in this feature are amplified by being multiplied by a large weight [55]. Small changes in the value of this input feature can lead to large changes in the output. The methods of weight penalty mainly are L1 regularization and L2 regularization. In L1, the penalty is the sum of the absolute values of weights. The loss function with L1 regularization is formalized as:

$$Loss = Min(\sum_{i=1}^{n}(y_i - w_i x_i)^2 + p\sum_{i=1}^{n}|w_i|), \tag{2.37}$$

p is the tuning parameter which decides the extent in penalization.

In L2, the penalty is the the sum of the squared values of weights. The loss function with L2 regularization is formalized as:

$$Loss = Min(\sum_{i=1}^{n}(y_i - w_i x_i)^2 + p\sum_{i=1}^{n}(w_i)^2). \tag{2.38}$$

2.3.4.6.2 Data augmentation

Data augmentation is a strategy that is able to significantly increase the diversity of data available for training models, without actually collecting new data. It can be used as a regularization technique that is used to avoid overfitting when training dataset is small. For images, augmentation techniques such as cropping, padding, and horizontal flipping are commonly used to enrich training samples. For text, the authors of [56] applied synonym replacement, random insertion, random swap, and random deletion to create more text data for boosting performance on text classification tasks. Other

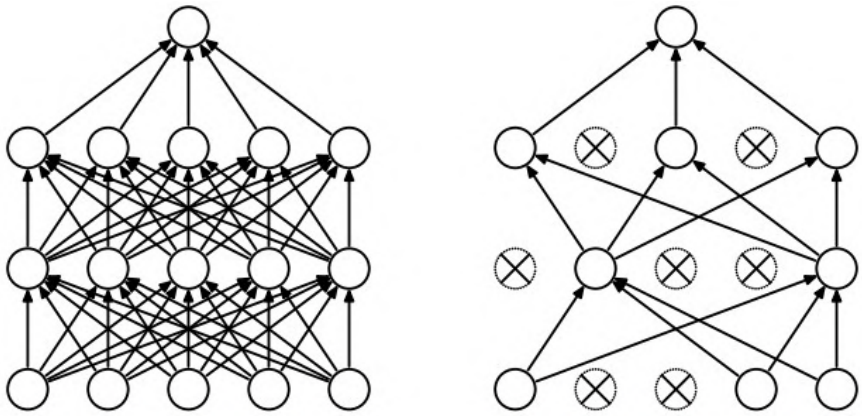


Figure 2.19: Dropout neural network model. (a) a standar neural netwrok with 2 hidden layers, (b) a thinned network produced by applying dropout to (a).

augmentation techniques such as adding random noise, is also very effective.

2.3.4.6.3 Dropout

Dropout prevents overfitting and provides a method of approximately combining exponential many different neural network architectures efficiently [57]. The term 'dropout' refers to randomly dropping out units (along with their connections) from the neural network during training. Dropping a unit out, means temporarily removing it from a network, along with all its incoming and outgoing connections, as shown in Figure 2.19. The choice of which units to drop is random. Dropout is initialized with a rate p. At each training stage, individual nodes are individual nodes are either "dropped out" of the net with probability $1 - p$ or kept with probability p. Applying dropout to a neural network equates to sampling a 'thinned' network from it. Training a neural network with dropout can be seen as training a collection of thinned networks extensive weight sharing.

2.3.4.6.4 Early stopping

Early stopping provides guidance as to how many iterations can be run before the learner begins to overfit. Early stopping rules work by splitting the original training set into a

new training set and a validation set. The error on the validation set is used as a proxy for the generalization error in determining when overfitting has begun. These methods are most commonly employed in the training of neural networks. A naive implementation of holdout-based early stopping is as follows [58]:

1. Split the training data into a training set and a validation set.

2. Train only on the training set and evaluate the per-example error on the validation set once in a while, e.g. after every fifth epoch.

3. Stop training as soon as the error on the validation set is higher than it was the last time it was checked.

4. Use the weights the network had in that previous step as the result of the training run.

2.4 Summary

This chapter introduced the basic concepts and abridged histories of radar and LiDAR, respectively. The types and applications of LiDAR and radar were presented. The fundamental concepts of deep learning including CNN, RNN, and LSTM are reviewed. In follow chapters, radar and LiDAR sensors are combined with deep learning respectively, to perform human activity recognition both indoors and outdoors

Chapter 3

Human activity recognition, people counting, and coarse-grained localisation outdoors using micro-Doppler signatures

Human activity recognition and classification can be achieved by comparing the differences in the radar micro-Doppler signatures of different targets and activities. Currently, most related work focuses on recognizing different human activities in different indoor scenarios. However, human activity recognition outdoors using micro-Doppler signatures is more challenging because outdoor environments are more complex than indoor environments. There are complex terrains, many kinds of animals, and trees in outdoor areas. Additionally, the power supply is also a problem in some outdoor areas. This chapter builds a low-power radar system to perform human activity recognition, people counting, coarse-grained localisation in outdoor areas with consideration of the

surrounding trees and grasses, and confusers such as animals and drones.

3.1 Introduction

The detection, recognition, and classification of human targets and human activities outdoors is an important topic in many applications, such as surveillance, search and rescue, and military. Since Doppler radar is contactless, privacy-preserving, and able to penetrate walls, clothes, etc., it has been increasingly used in human activity recognition. In recent years, there has been great research interest in human activity classification using micro-Doppler signatures [21, 44, 59, 60]. In [7], 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 [40] used a 2.4 GHz Doppler radar to classify seven human activities, including running, walking, boxing, etc. I. Bilik et al. [61] 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. [62] 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) [63], Empirical Mode Decomposition (EMD) [64], and Mel-frequency Cepstral Coefficients (MFCC) [65] 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 judgment such as in [7, 40, 66, 67].

After being extracted, the micro-Doppler features are fed into classifiers. The most used classifiers in micro-Doppler based human activity classification are SVM [40, 66, 68], kNN [7] and Naïve Bayes [69]. Deep learning is increasingly being used to classify micro-Doppler signatures [70, 71]. The hierarchical structure of deep learning enables automatic

learning of the features from raw data without relying on feature extraction methods.

So far, most related work was implemented in confined spaces [7, 65, 72]. Micro-Doppler based human activity detection outdoors has not been extensively investigated. With the constant development and improvement of radar manufacturing, the Doppler radar is becoming cheaper and able to operate with low power consumption, this makes Doppler radar more suitable to be deployed outdoors and on a large scale. In [66], 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 the movement of leaves and trees outdoors can cause interference to the micro-Doppler signatures and may contribute to making the environment less stable. It is more challenging to perform human activity detection in outdoors. These challenges can be listed as follows:

1. Complex background, trees, and foliage will introduce noise resulting in complex multipath effects and signal attenuation.

   Most objects indoors are simple and static. With the isolation of walls, indoor environments are stable. However, there are many plants (grass, trees) in outdoor environments. Because of the airflow, these plants keep producing tiny movements. This can result in low SNR (signal-noise-ratio). If there is wind, the movement would affect the quality of micro-Doppler signatures seriously. Multipath is the propagation phenomenon that results in radio signals reaching the receiving antenna by two or more paths. Uneven terrains and foliage can result in complex multipath propagation and signal attenuation.

2. Animals and drones can be strong confusers to humans

   Many moving subjects can appear in the detection range of radar systems. Besides humans, there are animals, drones, etc. They also can produce strong and unique micro-Doppler signatures. This creates a considerable challenge to human recognition. It is necessary to take these confusers into consideration in the experiments

in order to differentiate humans from them.

3. Lack of power supply

   Most radar systems are power-consuming. However, many monitoring areas may be far away from residential areas, where has no power station to support long-time monitoring with a radar system. However, a low-power radar system usually compromises its performance. So it is challenging to use a low-power radar system and achieve the results that are comparable to other radar systems.

Additionally, outdoor areas are usually large. It is important not only to know what the target is doing but also where the target is, and how many people are in the detection area. To the best of the author's knowledge, there is no relevant literature research applying micro-Doppler signature classification for coarse-grained location estimation. This chapter implements subject classification, human activity recognition, people counting, and coarse-grained localisation by using a low-power and low-cost Doppler radar system.

3.2 Related work

Generally, human activity recognition involves data collection, feature extraction, and classification with machine learning. So this section reviews related work from three aspects: radar sensors used in data collection, feature extraction in micro-Doppler analysis, and machine learning in radar-based human activity classification.

3.2.1 Radars used in human activity recognition

Radar sensors have a great impact on micro-Doppler based human activity recognition. As micro-Doppler signatures are extracted in frequency scale, a radar that operates at a higher frequency can provide more detailed micro-Doppler signatures [73]. In addition to the frequency, a high range resolution can offer additional information about the target, which contains the size and the number of moving parts. Range resolution is the minimum separation (in range) of two targets or equal cross-section that can be

resolved as separate targets [74]. It is determined by the bandwidth of transmitted signals. With the range resolution, the received signal returns are separated into a set of discrete elements, called range bins. When a moving target is observed using a wideband radar system, the individual scatterers of the target can be expected to migrate between range bins, each scatterer follows a different trajectory based upon target's macro-motion (like the torso's movement of human walking) and micro-motion (like the arm and leg swings of human walking), the different trajectories result in varying radar-target ranges for each scatterer, the composition of all individual scatter range paths is called range-Doppler signatures [75]. In [76], the authors investigated range-Doppler signatures of pedestrian walking/running with an S-band (2.9-3.4 GHz) radar system, which had a wide frequency bandwidth of 800 MHz; as shown in Figure 3.1, two spectrograms generated from range-Doppler analysis show the range-Doppler signatures of walking and running respectively, the moving parts in human walking are clearly separated in Figure 3.1(a). In [77], two radar systems, which are an acoustic system with a 70 KHz and an RF pulse Doppler with 92.5 GHz central frequency and 1 GHz bandwidth were used to investigate the range-Doppler signatures of individual walking, the generated spectrograms from both radars are similar, it shows that at positions corresponding to when the limbs are near the torso the number of discrete frequency components in the observed range bins increases. Fogle et al. [78] employed a Doppler radar that operates at 10 GHz which has a bandwidth of 4 GHz to measure range-Doppler signatures of human walking, and proposed an expectation-maximization (EM) to decompose range-Doppler signatures into the responses of constituent body parts. With a wideband (a large bandwidth) radar, it is very promising to enhance the accuracy of human activity classification by combing range-Doppler signatures and micro-Doppler signatures.

Polarimetric radars have been employed in some research work. Different from common radars that transmit and receive radio waves with a single horizontal polarization, polarimetric radars transmit and receive both horizontal and vertical polarizations. The transmit and receive polarizations usually denoted by a pair of symbols, which are listed

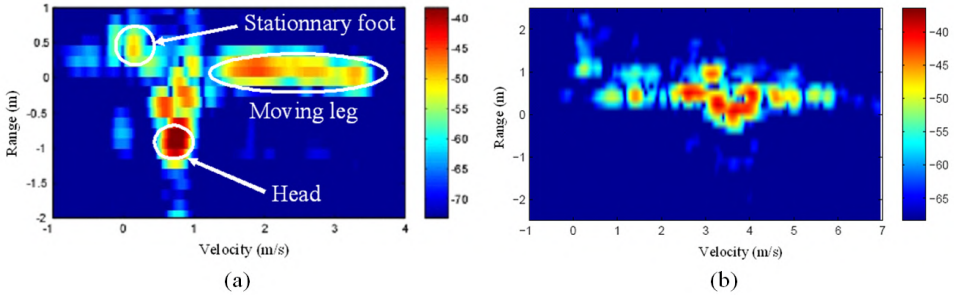


Figure 3.1: Experimental range-Doppler images of (a) walking and (b) running

as follows: (a) HH, for horizontal transmit and horizontal receive, (b) VV, for vertical transmit and vertical receive, (c) HV, for horizontal transmit and vertical receive, (d) VH, for vertical transmit and horizontal receive. HH and VV are referred to as co-polarized (co-pol), because the transmit and receive polarizations are the same; HV and VH are referred to as cross-polarized (cross-pol) because the transmit and receive polarizations are orthogonal to one another. Fully polarimetric radar data can potentially improve the separability of different parts of human motion [79]. In [79], a polarimetric radar operating at Ka band was used to measure human walking with or without carrying a load, with analysis of the collected data, it showed the arm was often bent at the elbow, providing a surface similar to a dihedral which could create a double bounce, this phenomenon is easier to be observed when the subject is not carrying a load. In [80], *J. Park et al.* measured polarimetric micro-Doppler signatures of simulated human walking, the results showed only minor difference between VV and HH polarized returns, but large difference between co- and cross-polarizations; the motions of the torso and head produces strong co-pol but weaker cross-pol scattering, so it is expected that cross-polarized observations will emphasize the movements of legs and arms.

Monostatic Doppler radar can only observe the radial component of the micro-Doppler signatures. If the target moves perpendicularly to the radar beam, the radar cannot detect the movement. For overcoming the limitations of monostatic radar systems, wide research has been done with bistatic or multistatic radars. A bistatic radar is a radar that the transmitter and the receiver are separated in different locations. Multi-

static radar is an integration of multiple radars that provides multiple looks at a subject from different viewpoints [31]. In [81], a multistatic radar system NetRAD, which is a 2.4 GHz, 45 MHz bandwidth, pulsed coherent radar system developed at University College London, was used to gather the polarimetric micro-Doppler signatures for determining whether a person is armed or not, the classification accuracies provided by Naive Bayes and LDA were up to 0.9-0.95, the results showed classification performance with co-pol (HH, VV) data has lower classification error in comparison to cross-pol (HV) data, and the combining co-pol and cross-pol data has shown positive improvement in the classification accuracy. In [62], a bistatic radar system operating at 4 GHz was used to differentiate three motion classes (no activity, swinging arm, picking up an object) from three aspect angles (0°, 45°, 90°), it was compared with a monostatic radar in classification with an SVM classifier, the results showed it can achieve a higher accuracy when the motion is perpendicular to the radar beam. In [82], the NetRAD radar system was used to detect individual walking, and the data of each sensor (a bistatic channel) was compared with the data of the monostatic Doppler radar operating at 2.4 GHz, the low level of correlation between the monostatic and bistatic channels indicates they carry independent information about the target, which provides additional information to improve the detection performance, with this knowledge, it can be used to reduce the redundancy of the data gathered by a multistatic radar system.

Micro-Doppler signatures also can be observed in ultrasound radar. In [34], an acoustic sensing system which operates at 40kHz was used to measure the micro-Doppler signature of human, dog, and horse. The spectrograms generated from these targets showed a big difference between them. *Balleri et al.* [17] used an acoustic radar operating at 80KHz to perform indoor human behaviour classification, micro-Doppler signatures were gathered for three different personnel targets undertaking various actions. PCA, Cepstrum, and Mel-Cepstrum were employed and compared in feature extraction, kNN (k=5) and Naive Bayesian were trained to distinguish walking gaits of different targets or different actions (walking with/without a heavy weight on the shoulders) undertaken by the same target; in the targets differentiation, the accuracy of kNN combining with

PCA was above 80%, the accuracy of Naive Bayesian combining with PCA was 92.57%; in the activity classification, the accuracy of kNN combining with PCA was 75%, the accuracy of Naïve Bayesian combining with PCA was 82.05%. In [83], a continuous-wave sonar at 40 KHz was applied to identify individual subjects and gender recognition, with the k-means clustering algorithm, the classification rate was 87.1% for individual recognition, 92.4% for gender classification. Compared with electromagnetic radars, acoustic radar systems offer the advantages of lower cost, easier signal processing and immunity from electromagnetic interference. But the speed of sound propagation is likely affected by temperature, atmospheric pressure, and humidity.

Table 3-A summarizes some radar systems used in related work.

Table 3-A: Radars used in micro-Doppler analysis

Radar	**Frequency**	**Description**	**Citation**
CW radar	24 GHz	Measuring a walking person with or without carrying an object	[1]
	77 GHz	Investigating different human motions (walking, jogging, crawling, running, and creeping)	[67]
	6.5 GHz	Measuring human motions in free-space and through-the-wall environments.	[84]
Pulse radar	5.8 GHz	Measuring three human behaviours (walking, running, crawling).	[7]
Acoustic radar	40 KHz	Measuring human and horse	[85]
	80 KHz	Measuring human behaviours	[17]
	70 KHz	Investigating range-Doppler signatures of individual walking	[77]
Polarimetric radar	Ka-band	Measuring human walking with or without a load	[79]

Radar	**Frequency**	**Description**	**Citation**
	2.25 GHz	Measuring polarimetric micro-Doppler signatures of simulated human walking	[80]
A multistatic radar system (NetRAD)	2.4 GHz	NetRAD is used to distinguish between unarmed and armed personnel moving.	[81]
	4 GHz	Distinguishing three motion classes (no activity, swinging arm, picking up an object) from three aspect angles (0, 45, 90)	[62]
A bistatic radar system	5.8 GHz	Investigating walking back and forth, sitting down and standing up on an office chair, bending down to pick up an object from the floor and coming up to stand, and waving with one hand left and right	[86]

3.2.2 Feature extraction in micro-Doppler analysis

In machine learning and statistics, feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations [87]. In micro-Doppler analysis, the initial set of measured data is radar signals collected from the target or its movements, the derived features are micro-Doppler signatures extracted from the signals, these features are distinctive enough to differentiate the target or its movements from others.

Different from the data of optical sensors, radar micro-Doppler signatures are less intuitive for humans to interpret [78]. Time-frequency processing techniques have been

used to represent micro-Doppler signatures from the received signals [17]. Bjorklund et al. [67] studied micro-Doppler signatures of different human motions, such as walking, jogging, crawling, running, and creeping with different aspect angles using an FMCW (Frequency Modulation Continuous Wave) radar with a carrier frequency of 77GHz. It was illustrated that micro-Doppler properties, such as the envelope of the spectrogram, mean torso frequency, the bandwidth of the Doppler signal, the ratio of the strength of the torso echoes to the other echoes in the envelope, and the distribution of different strength of echoes can be exploited for classification. Ram M. Narayanan [84] gathered micro-Doppler signatures of six human subjects in free-space and through-wall environments using a 6.5 GHz C-band coherent radar. The collection of micro-Doppler signatures was divided into three categories: stationary, forward-moving, and multi-target; six features, including (i) average Doppler frequency, (ii) total Doppler bandwidth, (iii) Doppler offset, (iv) bandwidth without micro-Dopplers, (v) maximum cadence frequency, and (vi) arm swing rate, were investigated for their significance to human activity classification in two scenarios (free-space, through-wall). In [7], a low-power pulse-Doppler radar that operates at 5.8 GHz was used to collect the micro-Doppler signatures of four subjects' three different activities (walking, running, and crawling) performed on a treadmill; ten different features, such as the mean torso Doppler, the mean value of the lower envelope, the mean value of the upper envelope, the bandwidth of the torso return, the peak-to-peak (total) bandwidth, etc., were extracted and fed into kNN, with classification rate over 88% were achieved. Similar features also were used in [40] for the classification of seven human activities detected by a 2.4-GHz Doppler radar, the classification rate of above 90% was achieved with SVM.

However, the features used above are handcrafted. This way is inefficient and unable to process a large amount of Doppler signals, not to mention to achieve real-time human activity recognition and detection. Many functions have been applied in non-manual micro-Doppler feature extraction. In [64, 88] the researchers employed empirical mode decomposition (EMD) as a method for feature extraction. EMD is an adaptive technique that decomposes a signal into time-frequency components called intrinsic mode

functions (IMFs). EMD used in [64] extracted features from frequency spectrograms of four human motions (breathing, swinging arms, and picking up an object, standing up from a crouching position) in a through-wall indoor environment. Y Li et al. [88] applied EMD directly on radar signals to extract micro-Doppler features for classifying wheeled and tracked vehicles, the classification rate obtained by SVM reached to 84.61%. In [89, 90], EMD was utilized to suppress the interference and make the main target (aircrafts or vehicles) stand out. Singular value decomposition (SVD) also has been widely applied in micro-Doppler signatures extraction [91–94]. SVD can reduce the feature space dimensionality; information about target velocity, spectrum periodicity, and spectrum width, can be inferred from the SVD vectors. Francesco Fioranelli et al. [91] compared the armed/unarmed personnel classification using SVD-based features and handcrafted features, which were extracted from radar signals collected by a three-node radar system that operates at 2.4 GHz, a higher accuracy of 97.22% was achieved by using SVD-based features. In [92, 93], micro-Doppler signatures extracted by SVD were used to perform the classification between birds and unmanned aerial vehicles (UAVs). Jing Li et al. [95] successfully identified and located buried humans with an Ultra-Wideband radar (UWB), an SVD method was used to reduce the noise in weak life signals, and extract micro-Doppler signatures of breathing and heartbeat. PCA, the most popular dimensionality reduction and feature extraction tool, has been widely applied into micro-Doppler signatures extraction. In [63], the feature vector extracted by PCA from the spectrogram generated by STFT was utilized to classify three human behaviours (one-arm swing, two-arm swing, non-arm swing) that detected using a Doppler radar at 2.5 GHz. An SFP-CVDFP-PCA algorithm that combines spectrogram frequency profile (SFP), PCA, and cadence velocity diagram frequency profile (CVDFP) algorithms was proposed in [96]. It was applied to micro-Doppler signatures extraction of behaviours, including human walking, a horse with rider walking, and horse and human walking separately, the classification of 94.9% was achieved using the SFP-CVDFP-PCA and an SVM classifier. There are still many other feature extraction methods, such as LDA [97, 98], mel-frequency cepstrum (MFC), Gabor wavelet

filter, etc., that have been proposed and applied to human activity micro-Doppler features extraction. Some researchers [1, 17] have compared the effects of different feature extraction methods to human activity classification. Tivive et al. [1] compared the performance of two-directional two-dimensional principal component analysis ((2D)2-PCA), Gabor wavelet filter, Cadence frequency based method, and EMD on human activities (walking with no-arm swing, one-arm swing, two-arm swing) classification, the result showed (2D)2-PCA achieved the highest accuracy of 90.1%, while the EMD performed the worst. In [17], micro-Doppler signatures were gathered for three different personnel targets undertaking various actions using an acoustic radar; feature extraction methods, including PCA, Cepstrum, and MFC were compared, and the features extracted with PCA always provided the highest accuracy. Yang et al. [99] studied the performance of Independent Component Analysis (ICA) and PCA on motions, such as rotation, coning and tumbling, and PCA performed better for RCS (Radar Cross Section) model.

Table 3-B summarizes feature extraction methods that have been frequently used in related work.

Table 3-B: Feature extraction methods in micro-Doppler analysis

Method	**Description**	**Summary**	**Citation**
Handcraft	A hand process and human visual judgement, not extracting features automatically by a computer system.	Features, including average Doppler frequency, total Doppler bandwidth, Doppler offset, bandwidth, swing rate, etc. are usually extracted.	[7, 40, 66, 67, 84]

Method	**Description**	**Summary**	**Citation**
Empirical mode decomposition (EMD)	EMD is an adaptive technique that decomposes a signal into time-frequency components called intrinsic mode functions (IMFs).	One advantage of EMD is that it does not require a priori knowledge of the signal and is not dependent on the selection of a kernel function like the STFT, wavelet transform, Wigner-Vile distribution etc.	[64, 88–90]
Singular value decomposition (SVD)	Singular-value decomposition (SVD) is a factorization of a real or complex matrix. It is the generalization of the eigendecomposition of a positive semidefinite normal matrix to any mn matrix via an extension of the polar decomposition.	SVD allows the reduction of the feature space dimensionality, and it is shown that parameters with physical meaning, such as target velocity, spectrum periodicity, and spectrum width, can be inferred from the SVD vectors [91].	[91, 92, 94]

Method	**Description**	**Summary**	**Citation**
Principle component analysis (PCA)	Principal component analysis is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables	PCA is the best-known feature extraction and dimensionality reduction method.	[63, 68, 96]
Independent component analysis (ICA)	ICA is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals.	ICA can be seen as an extension to PCA and factor analysis. ICA is a much more powerful technique, however, capable of finding the underlying factors or sources when these classic methods fail completely.	[99, 100]

Method	**Description**	**Summary**	**Citation**
Mel-frequency cepstrum coefficients (MFCCs)	Mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound. MFCCs are coefficients that collectively make up an MFC.	MFCCs are usually used in sound processing. While radar signals are very similar to sound, both of them are time-series data. So some researchers applied MFCCs into micro-Doppler analysis.	[65, 101, 102]
2-Directional 2-Dimensional PCA (2D2PCA)	2D2PCA can be regarded as a two dimensional version of PCA with two projection matrices to reduce the number of rows and columns of an image simultaneously.	2D2PCA has been widely used in image classification. The spectrogram generated from radar signals can be seen as an image with one channel.	[1, 103]

3.2.3 Classification using machine learning

In micro-Doppler based human activity recognition, classification is used to determine what kind of activity is detected based on micro-Doppler features extracted from radar signals using a trained classifier, which are machine learning algorithms or statistic functions. Micro-Doppler features extracted from radar signals are fed into classifiers. In micro-Doppler based human activity recognition, the most commonly used classifiers are SVM, kNN, and Naive Bayes. Kim et al. [40] measured seven different human activities using a 2.4-GHz Doppler radar, an SVM was employed to classify them upon handcrafted micro-Doppler features. In [68], the authors used an SVM on an embedded system to

achieve real-time classification of different behaviours from a single person and a group of people. In [66], an SVM was employed to classify crawling, walking and jogging in four different environments, such as free space, through-the-wall, leaf tree foliage, and needle tree foliage. In [44], a continuous-wave Doppler radar that operates at 7.25 GHz was developed to measure human, dog, and bicycle. The physical characteristics of the target (target stride) was used to overcome the uncertainty of the micro-Doppler signatures, the SVM used in the research could detect a human subject with an accuracy of over 96%. Gurbuz et al. [104] investigated aided and unaided human gaits using four different radar and sonar systems including a 40 kHz sonar, a 5.8 GHz wireless pulse-Doppler radar mote, a 10 GHz CW radar, and a 24 GHz CW radar. The classification result achieved by an SVM showed the sonar not only had the distinct advantages of higher resolution and no interference in RF communication bands, but also exhibited the highest performance, with a correct classification rate of 86%. JA Nanzer et al. [69] built a Bayesian formulation for the classication of humans and vehicles using micro-Doppler signatures obtained from a 36 GHz scanning-beam continuous-wave radar, and achieved the accuracy of above 0.9. In [7], a kNN (k=2) was implemented to classify micro-Doppler signatures collected by a 5.8 GHz pulse-Doppler radar. Four human activities, including human walking, running, crawling were measured from three different aspect angles. Deep learning algorithms also have been applied into micro-Doppler signatures based human activity classification, and it achieved good results. Jokanovic et al. [105] applied Deep Neural Network for radar-based falling detection. Four motions, including walking, falling, bending/straightening, and sitting were investigated using a 6 GHz CW radar in the experiment, the accuracy of 87% was achieved, which was higher than 83% for SVM. Kim et al. [71] proposed the use of DCNNs for human detection and activity classification, a DCNN was applied on the sample spectrograms of four types of targets (human, dog, horse, car) and seven human activities that measured by a Doppler radar operating at 7.25 GHz, and it achieved 97.6% and 90.9% accuracy respectively. The same radar also used in [106] to investigate micro-Doppler signatures of humans on water, including a swimming person, a swimming person pulling a floating object, and a

rowing person in a small boat; three swimming styles, including free stroke, backstroke, and breaststroke were measured, after training with DCNN, the classification accuracy reached to 87.8%. The performance of different classifiers is compared in many research works. Liu et al. [65] compared the performance of SVM and kNN in fall detection based on two Doppler radar sensors, kNN (k=3) produced the best results that were close to 0.96 for both sensors. Trommel et al. [70] presented the use of DCNN in distinguishing multi-target human gait measured by an X-band CW radar; the DCNN outperformed SVM and kNN by a wide margin. In [85], the authors explored the utility of ultrasonic sensors to distinguish between people and animals (horse) walking and provided an SVM and a Bayesian classifier to classify them upon high SNR signals and low SNR signals respectively. SNR is a measure that compares the level of the desired signal to the level of background noise, low SNR means a high level of background noise. The accuracies achieved by both classifiers were above 80% with a small gap. GE Smith et al. [107] measured micro-Doppler signatures of a person, wheeled vehicles, tracked vehicles from different aspect angles using a Doppler radar operating at Ku-band; the performance of naive Bayesian classifier was compared with LDA, kNN and DTW. The classification rates achieved by them were 0.95, 0.90, 0.84, and 0.6, respectively.

Table 3-C: Machine learning algorithms in micro-Doppler based classification

Algorithms	**Class**	**Accuracy**	**Citation**
Support Vector Machine (SVM)	Walking while holding an object with both hands, walking while carrying an object with one hand, and walking with arms swinging	90%	[1]
	Running, walking, walking while holding a stick, crawling, boxing while moving forward, and boxing while standing	Around 92%	[40]

Algorithms	**Class**	**Accuracy**	**Citation**
	Crawling, walking, and jogging	98.11%	[66]
	Walking of horse and human, respectively	92.70%	[96]
	Walking, running of a person and a group of people respectively	94.62%	[68]
	Pedestrians swinging [both arms only one arm, and no arms], on crutches, pushing wheelchairs, and seated in wheelchairs	95%	[108]
K-Nearest-Neighbour (kNN)	Walking, running, and crawling	90.30%	[7]
	15 hand gestures	96%	[109]
	Walking, running, boxing, leaping, and random motion.	87%	[110]
Naïve Bayes	Personnel, wheeled vehicles, tracked vehicles	94%	[107]
	Armed versus unarmed personnel	95%	[81]
	Aided and unaided walking	86%	[104]
	Walking and running	97.08%	[111]
Artificial neural network (ANN)	Walking, falling, bending/straightening, and sitting	87%	[105]
	Running, walking, walking without moving arms, crawling, boxing, boxing while moving forward, and sitting still	87.80%	[40]

Algorithms	**Class**	**Accuracy**	**Citation**
Convolutional neural network (CNN)	Running, walking, walking while holding a stick, crawling, boxing while moving forward, boxing while standing in place, and sitting still	90.90%	[44]
	A swimming person, a swimming person pulling a floating object, and a rowing person in a small boat	87.80%	[106]
	Swiping from left to right, swiping from right to left, rotating clockwise, rotating counterclockwise, pushing, double pushing, holding, and double holding	85.60%	[112]
	Walking alone or side by side in groups of two or three people	86.90%	[70]
Dynamic time warping (DTW)	Personnel, tracked vehicle, and wheeled vehicle	80%	[113]
	Pushing arms and pulling back, Crossing arms and opening, Crossing arms, Rolling arms, Stop sign, Pushing arms and opening.	99%	[114]
Gaussian mixture model (GMM)	Human, vehicles, animals	93.30%	[115]
	One person, two persons, three persons, vehicle, animal, clutter	96%	[61]
	Humans, truck, vehicle and clutter	99.08%	[102]
Linear Discriminant Analysis (LDA)	Walking of three individuals	98.90%	[94]
	Walking, running	95.83%	[111]

After a wide exploration of related work, several conclusions can be made. Firstly, most radars used in related work are very expensive and power-consuming. This prevents the large-scale application of radar-based systems. It is necessary to apply a low-power low-cost radar system in human activity recognition. Secondly, handcrafted ways for feature extraction are time and labour-consuming, and not able to achieve real-time human activity recognition. Feature extraction algorithms are automatic and efficient, but the performance of human activity recognition still heavily depends on the quality of the extracted features. Thirdly, deep learning shows its superiority in human activity recognition without additional feature extraction. Based on these conclusions, this section implements a low-power low-cost radar system to measure human activity indoors and outdoors. Deep learning is also used for human activity classification in this section and it is compared with the way of the integration of feature extraction methods and traditional machine learning.

3.3 The low-power radar system

The Doppler radar system built in this chapter consists of two BumbleBee radars from Samraksh [116]. This section describes the BumbleBee radar and the low-power radar system in detail.

3.3.1 The BumbleBee radar

BumbleBee radar is a low-power Pulsed Doppler Radar (PDR) that is designed for a variety of Wireless Sensor Network (WSN) applications. Its center frequency is 5.8 GHz, which corresponds to a wavelength of about 5.2 cm. An important reason for using Bumblebee radar in this research is its low-power-consumption characteristic. It only consumes about 12 mA, so when using typical 1.5v AA alkaline batteries with a capacity of 2400 mA hours, it can run at 100% duty cycle for about 8 days [116]. This makes it suitable for long-term monitoring. BumbleBee radar is also low-cost. It is cheaper than several radar systems that are claimed as low-cost radar systems in related work. For example, the X-band CW radar designed in [115] costs $1,245. The K-band radar system

Table 3-D: Key features of BumbleBee radar

Center frequency	5.8 GHz
Detection range	Up to 10m
Coherent output	I & Q channels
Range gate sharpness	0.2m
Antenna	Onboard antenna
Responds to radial velocity	2.6 cm/s to 2.6 m/s
Total power draw	About 12 mA
Coverage pattern	60 degree conical coverage pattern

used in [117] costs about $200. The short-range FMCW radar operating at 24GHz used in [118] cost about $105. However, the BumbleBee radar only cost $100, including a TelosB mote. The BumbleBee radar enables a number of low-power, wireless sensor network applications including simple motion detection, robust intrusion detection using velocity estimation, velocity-based motion tracking and vibration monitoring. Key features of BumbleBee radar are shown in Table 3-D.

In particular, the BumbleBee is optimized for WSNs based on motes. BumbleBee radar and the TelosB Tmote usually work together (as Figure 3.2) with TinyOS 2.x operating system to provide a data downlink to a remotely located host computer.

TelosB mote (TPR2400) is an open-source platform designed by UC Berkeley, to enable state of the art experimentation for the research community. The TelosB bundles all the essentials for lab studies into a single platform including USB programming capability, an IEEE 802.15.4 radio with integrated antenna, a low-power MCU with extended memory and an optional sensor suite (TPR2420) [119]. TelosB mote offers many unique features. Amongst the main features offered are Zigbee acquiescent RF transceiver, incorporated onboard antenna, low current usage microcontroller of 8Mhz TIU MSP430 with 10kb RAM as well as programming and information accumulation through USB, optional sensor suite including incorporated temperature, light and humidity sensor, 1MB outside flash for information logging. TelosB delivers low power consumption allowing for long battery life as well as fast wake up from a sleep state. It is fully compatible with the open-source TinyOS. TinyOS is an energy-efficient software operating system

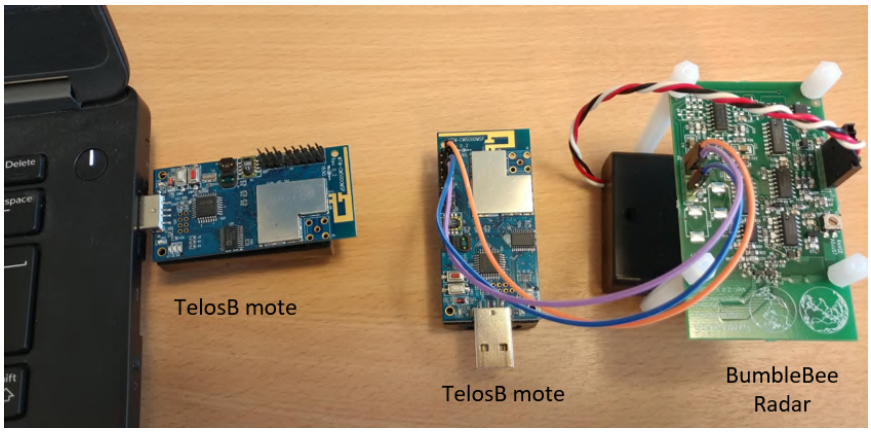


Figure 3.2: BumbleBee radar and TelosB mote

developed by UC Berkeley which supports large-scale, self-assembling sensor networks.

The sensor outputs data on two channels providing the in-phase (I) and quadrature-phase (Q) output signal components, which are used to form the complex signal $C = I+jQ$. The I/Q data output by the BumbleBee radar represents the peak of the matched filtered data acquired for each pulse. Thus, the time interval between each data packet corresponds to the pulse repetition interval (PRI) of the radar. As a result, this sensor is only able to supply users with the Doppler and velocity information pertaining to the target [116].

The time-domain and frequency-domain measurements [7] of the transmitted signals are shown in Figure 3.3 (a) and (b), respectively. Using the pulse analysis mode of the signal analyzer, the pulse duration was measured to be 40ns, whereas the pulse repetition frequency was found to be 2 MHz. In the frequency domain, the envelope of the received signal is that of a *sinc* function and corresponds to the Fourier Transform of the time-domain pulsed Doppler waveform. The bandwidth of the transmitted signal was measured to be 240 MHz.

3.3.2 The low-power radar system

A multi-radar system composed of two BumbleBee radars was built. The system collected signals of human activity from multiple angles, this provided more information for the

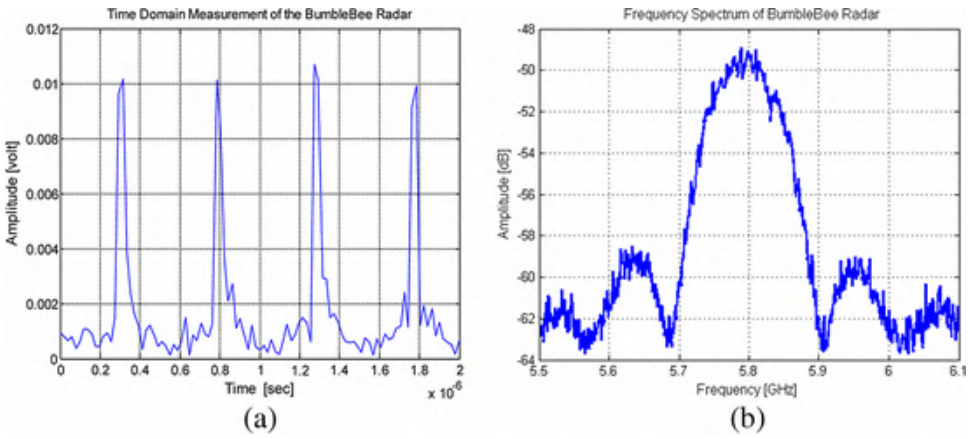


Figure 3.3: (a) Time-domain and (b) frequency-domain measurements [7]

classifiers. However, it also means more data was required to be transferred concurrently, consequently packet loss in data transmission increased in detriment to the gains in information gathered. It was found that a system consisting of one base station and two BumbleBee 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.

For validating the setup of the BumbleBee radars, a corner reflector was used as a target. The corner reflector is a passive device used to reflect radio waves directly back toward the radar. The corner reflector consists of three conducting sheet metal surfaces intersecting perpendicular to each other. The waves that hit the corner reflector are bounced by each surface to result in reversed direction waves sending back towards the radar. Therefore, the corner reflector is a useful device for radar system calibration. The corner reflector was tied to a string and hanged from the ceiling of a laboratory and pushed slightly to swing back and forth. This experiment followed the same set up as the corner reflector experiment presented in [8] that also used BumbleBee radars. The corner reflector's swinging was measured by the Bumblebee radars, and the collected signals were used to generate a frequency spectrogram to show the fluctuations of the reflected micro-Doppler signals. Figure 3.4(a) presents the frequency spectrogram shown in [8] and Figure 3.4(b) shows the spectrogram obtained from the radars used in this research.

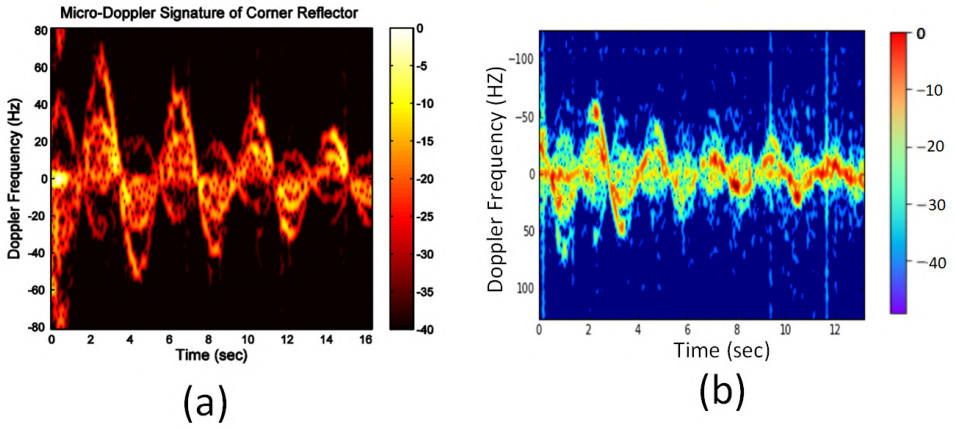


Figure 3.4: Micro-Doppler signatures of a swinging corner reflector, (a) the spectrogram in [8], (b) spectrogram of validation in this work

Both spectrograms show similar periodical fluctuations caused by the swinging of the corner reflector. In both, when the range of the swing angle declines, the amplitude of fluctuations decreases. Therefore, the setup of Bumblebee radars used in the work of this thesis was successfully validated.

The Doppler radar system was implemented outdoors. As shown in Figure 3.5, 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 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 localisation. 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

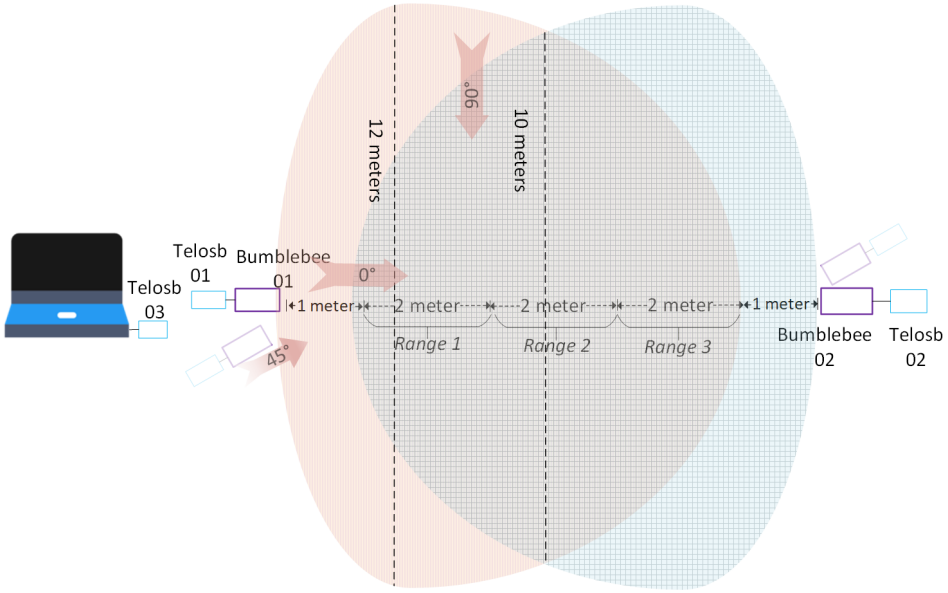


Figure 3.5: Doppler radar system implemented in outdoors

movement and the radar, three different directions (0°, 45°, and 90°) were considered. When the experiments were performed with the radar at 45° in relation to the target, the target moved on the ground in the same way as it did for the 90°, however, each radar was rotated to a position where their beams were at 45° in relation to the target movement.

3.4 Experiments

In order to investigate the effectiveness of human activity recognition outdoors using the radar system, four different tasks were processed in the experiments. They were human recognition, human activity classification, people counting, and coarse-grained localisation. As illustrated in Figure 3.6, the output metric for human recognition is whether the target is human, animal, drone, 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. Although the maximum number of people counted in a group was 4 and the ranges for rough localisation are relatively narrow (only 2 meters apart because of the short

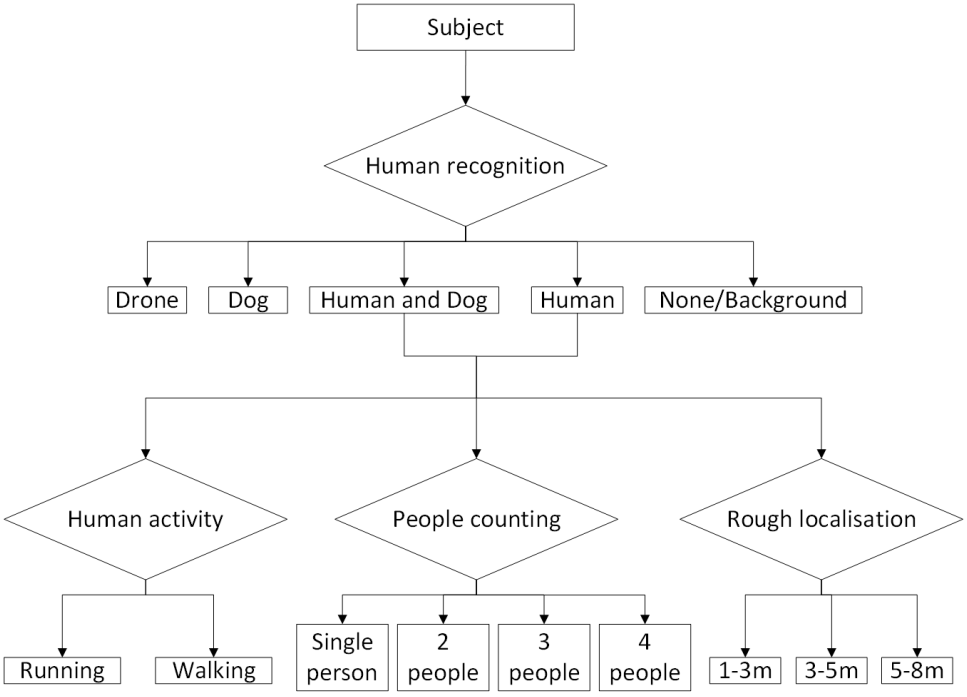


Figure 3.6: Workflow of four tasks, including human recognition, human activity detection, people counting, and coarse-grained localisation

detection range of the Bumblebee radar), it is foreseen that the methods used here could also be applied to a greater number of people and to radars that have greater detection ranges.

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°, 45°, 90°) relative to the radar beam. The same set of experiments was performed by each participant one at a time. At 45° and 90°, activities were performed in three ranges (1-3m, 3-5m, and 5-7m) relative to the primary node.

Figure 3.7 illustrates the micro-Doppler signatures of an individual participant walking at 0°, 45° and 90°. The spectrograms were generated by a STFT with a sliding

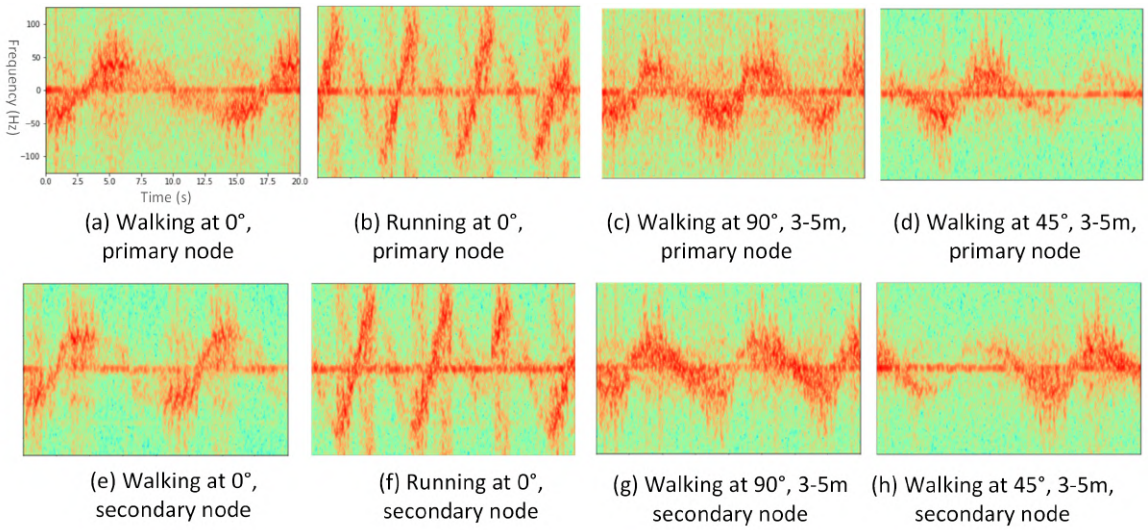


Figure 3.7: Micro-Doppler signatures of an individual at different angles

window size of 20s, which is also the length of the horizontal axis (time). The range of the vertical axis of the spectrograms (frequency) is from -100 Hz to 100 Hz. All spectrograms in this section are presented in the same axis range as shown in Figure 3.7(a). Comparing Figure 3.7(a) and Figure 3.7(b), the motion cycle of running is shorter than that of walking. There are more than three cycles in the spectrogram for running, but only two cycles in the spectrogram for walking. This confirms the physical fact that running takes less time to finish a motion cycle than walking. Interesting to note that the wave directions of Figure 3.7(a) and Figure 3.7(b) are opposite to the wave directions of Figure 3.7(e) and Figure 3.7(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°, 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° and 45°, 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 Figure 3.7(c) and Figure 3.7(g) and Figure 3.7(d) and Figure 3.7(h) respectively. Figure 3.7(d) and Figure 3.7(h) present more faded cycle segments, but in different positions. This is because when the participant is walking at 45°, 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 Figure 3.8 represent the micro-Doppler signatures of an individual walking and running at 90° in each one of the three ranges. As it can be seen, the patterns of the signal are different depending on the targets distances to the primary node. The waveforms in Figure 3.8(a) and Figure 3.8(b) are smooth, the waveforms in Figure 3.8(c) and Figure 3.8(d) are more zigzagged, and the waveforms in Figure 3.8(e) and Figure 3.8(f) are almost discrete. This indicates the distance does affect the wave pattern of the spectrograms of human activity. The spectrograms of the secondary node are not shown, because the wave patterns collected by the secondary node are similar to the primary node in each correspondent range.

Figure 3.9 represents micro-Doppler signatures of an individual walking and running at 45° in three different ranges. The patterns of these spectrograms are very similar to those in Figure 3.8, except that one part of each spectrogram is less clear, almost disappearing. This is because the target was not always in the detection range when he/she walked or ran back and forth at 45°. Notice that the amplitude of the waves in Figure 3.9 is a little smaller than in Figure 3.8.

Case 2: Activity classification and people counting in a group of people: Nine individuals participated in this experiment. Participants were arranged into three groups, as shown in Figure 3.10. 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°, 45°, and 90°) relative to the radar beam. Participants in the first or second group moved abreast. Participants in the third

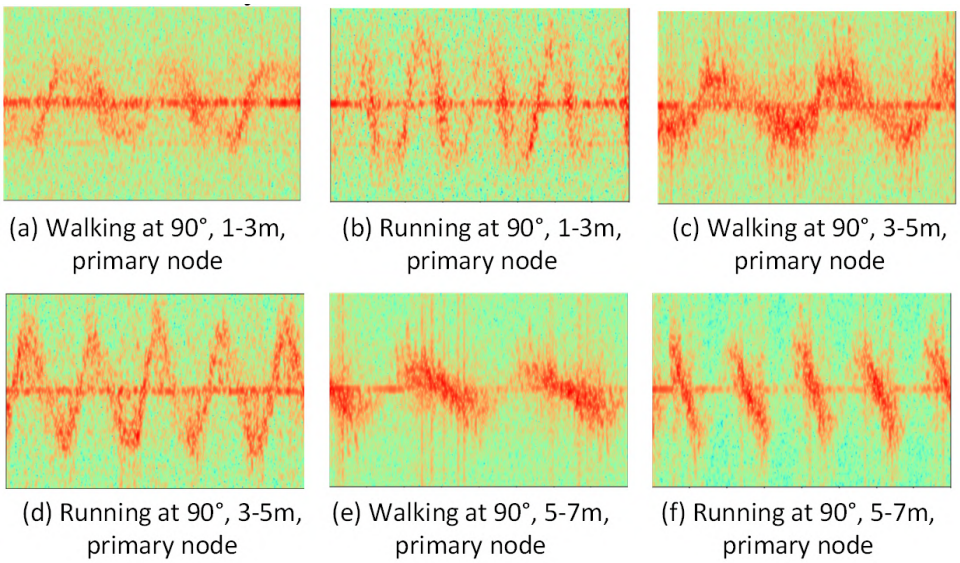


Figure 3.8: Micro-Doppler signatures of an individual walking and running in different ranges at 90°

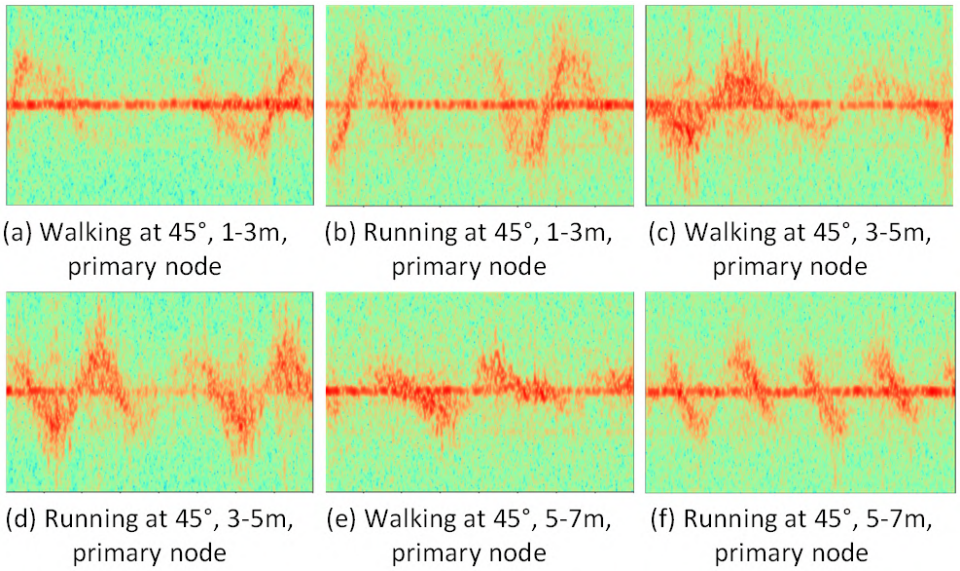


Figure 3.9: Micro-Doppler signatures of individual walking and running at 45° with three different ranges

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.

Figure 3.11 illustrates the spectrograms of the target with different numbers of people.

Figure 3.10: Group behaviour classification and people counting

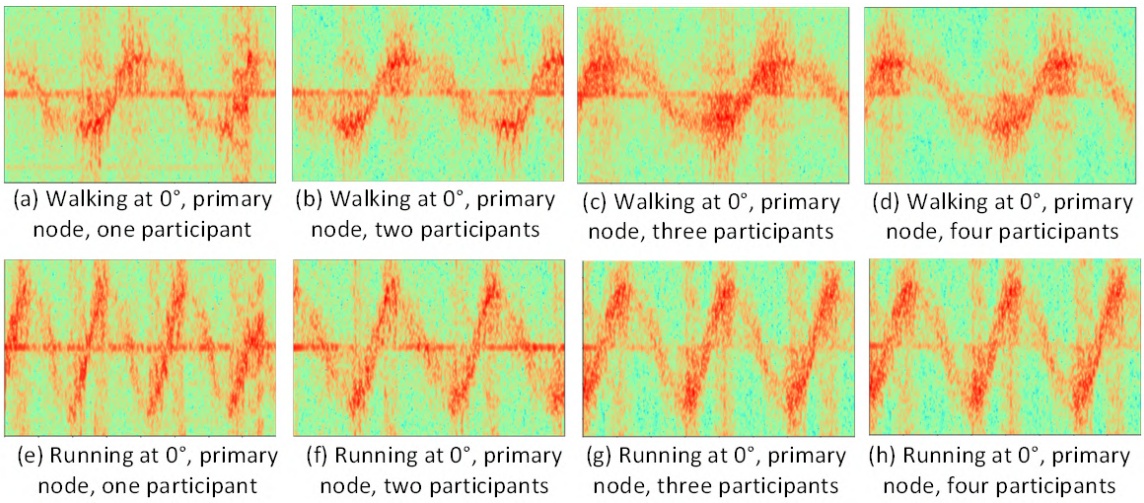


Figure 3.11: Micro-Doppler signatures of different number of people

of the target with different numbers of people. By the naked eye, it is difficult to find visual differences among the spectrograms. However, if we observe very carefully, it can be seen that the frequency waves in Figure 3.11(c) and Figure 3.11(d) are thicker than in Figure 3.11(a) and Figure 3.11(b). It could be understood that more people in a target would result in a more solid wave shape. However, this phenomenon cannot be observed in Figure 3.11(e), Figure 3.11(f), Figure 3.11(g), and Figure 3.11(h) possibly as a result of the speed of the movement.

Case 3: Differentiation among humans, dogs, and drones: One human volunteer, two dogs, and one drone participated in this experiment, as shown in Figure 3.12. In the first experiment, separately, each dog was encouraged by their owners to move back and forth inside the detection range, in order to obtain micro-Doppler signatures of a dog only. In the second experiment, the volunteer walked a dog back and forth. Because it is hard to

Figure 3.12: Differentiation among humans, dogs, and drones

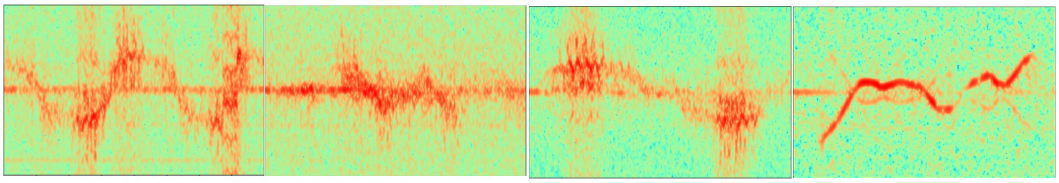


Figure 3.13: Micro-Doppler signatures of different subjects, (a) a person, (b) a dog, (c) a person walking with a dog, (d) a flying drone

control the dog's speed (i.e. walk and run) inside the detection range, the experiment did not differentiate between walking and running, and only two angles (0° and 90°) were investigated. Unavoidably, some noisy data was produced during the experiments, but the amount was small. In the third experiment, a drone was flying back and forward in the detection area, the flying height was from 0.5m to 3m.

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 a drone. As shown in Figure 3.13, the spectrograms generated by different targets are different. Figure 3.13(a) presents the periodic change of a human walking. Figure 3.13(b) shows the micro-Doppler signatures of a moving dog. Visually, the frequency wave in Figure 3.13(c) of a moving person and dog seems to have an overlapping shadow. Figure 3.13(d) shows the micro-Doppler signatures of a flying drone.

It is important to note that the clearest spectrograms were selected to illustrate the different micro-Doppler signatures. However, as shown in Figure 3.14, many spectro-

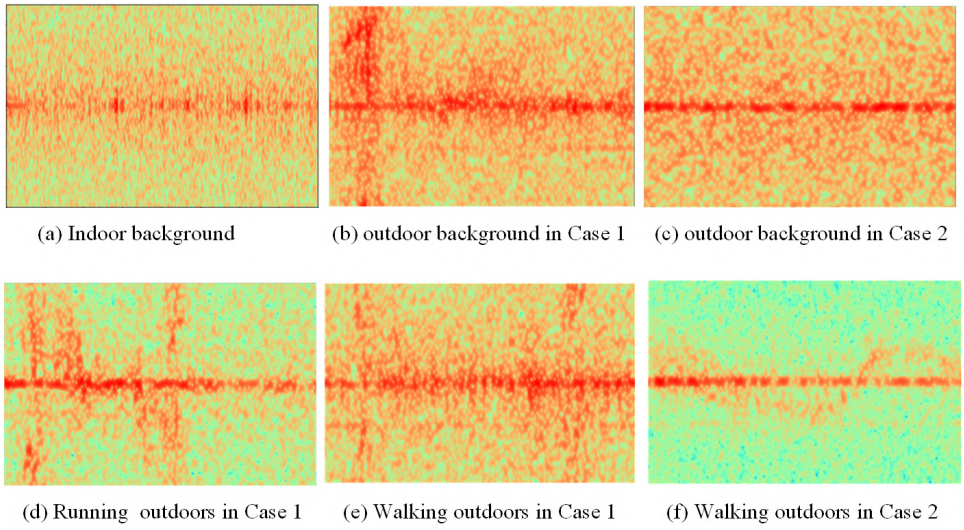


Figure 3.14: Noisy Spectrograms generated from outdoor environments

grams are very noisy or vague. Figure 3.14(a) shows the spectrogram of indoor background, Figure 3.14(b) and (c) are the spectrograms generated from the outdoor backgrounds in Case 1 and Case 2 experiments respectively. Compared to Figure 3.14(a), Figure 3.14(b) and (c) are more noisy and their lines around 0 Hz is more solid. This reflects that there is more clutter outdoors. The clutter is probably generated from uncontrolled movement from trees, small plants and small animals in the radar's detection range. The characteristics of human activities in some spectrograms become vague because of the clutter. As shown in Figure 3.14(d) and (e), the wave shapes of running and walking are not continuous. Figure 3.14(f) still shows the wave shape of walking, but it is very vague. This confirms that the outdoor environment increases the difficulty of micro-Doppler signature classification outdoors when compared to a more stable indoor environment.

3.5 Data collection and preprocessing

The data collected in the experiments requires further processing before feature extraction and classification are made. This is necessary to attenuate the background noise and better characterize the micro-Doppler signatures.

3.5.1 Data collection

With the Doppler radar system described in Section 3.3, radar signals were collected and stored in a database. As illustrated in Figure 3.15, signals collected by the BumbleeBee radars were transferred to the TelosB nodes and radio transferred to the TelosB base station. The TelsoB base station was connected to a computer which received the signals through a serial port. Finally, all data was stored into a MongoDB database system, which is an open-source NoSQL database and it is usually utilized to store a large amount of time-series data.

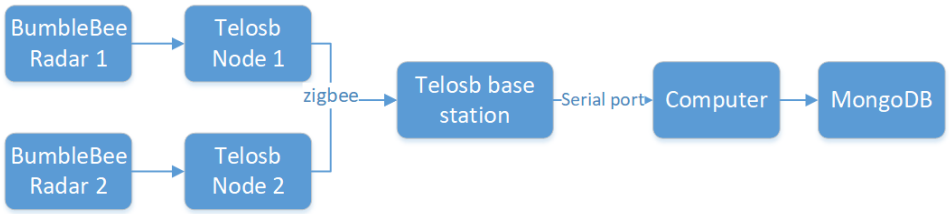


Figure 3.15: Data flow diagram of Doppler radar signals

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 3-E shows the information of each field of a radar signal recorded in the 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.

3.5.2 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 using STFT. Figure 3.16(b) shows a spectrogram generated by STFT. The bright zero line is mainly the direct-current (DC) component of the signal. In order to attenuate the DC component, a Butterworth high-pass filter has been applied to the raw signals. The Butterworth high-pass filter

Table 3-E: 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 here.
SecNodeID	The ID of secondary node that consists of BumbleBee 2 and Telsob 2 here.

can be represented by the following equation [120]:

$$|H(w)|^2 = 1/(1 + (w/w_c)^{2n}) = 1/(1 + \varepsilon^2(w/w_s)^{2n}), \tag{3.1}$$

where n is the order of the filter, w_c is the cutoff frequency (5 Hz here), w_s is the stopband boundary frequency.

After filtering, a new spectrogram is generated as can be seen in Figure 3.16(c). The fluctuations obtained from human movements are now more distinctive. Note the spine of the plot corresponds to the motion from the torso, the smaller fluctuations around the spine corresponds to the motions from arms and legs. Positive and negative Doppler frequencies correspond to the subject moving toward or away from the radar, respectively. For improving the robustness of the classifiers, the samples are required to be collected from environments with high-levels of clutter. By training the classifiers with those samples, they can learn how to reduce the interference of the noise in the

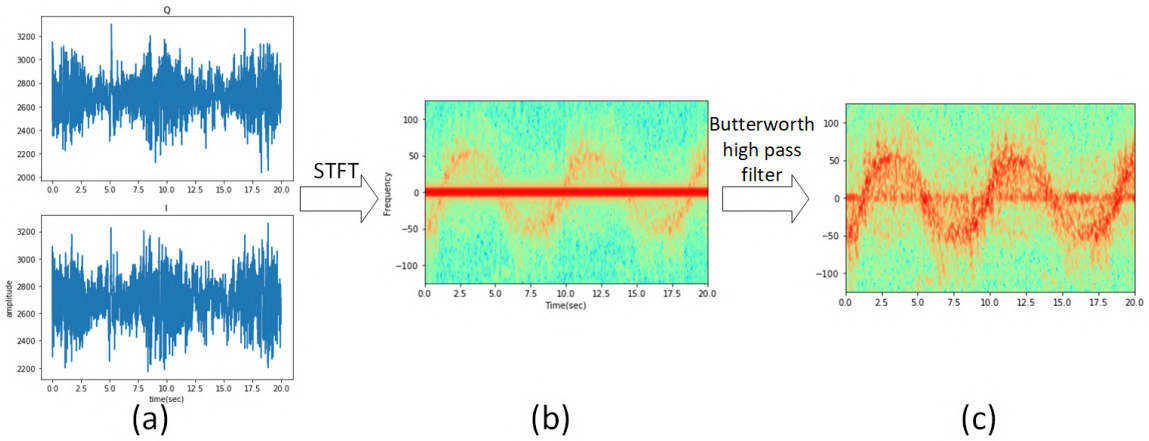


Figure 3.16: Data preprocessing, (a) original signals, (b) spectrogram after STFT, (c) spectrogram after high-pass filter

environment.

With the STFT technique, a different length of a sequence will result in a different width of the frequency spectrogram. 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. Figure 3.17 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 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 3.9.

Through the above processing methods and using an STFT sliding window of 1250 frames (time length of 5s), the described experiments generated 49720 spectrograms in total. The composition of the samples/spectrograms is shown in Table 3-F.

For the classification, the spectrogram samples were separated into two groups, 80%

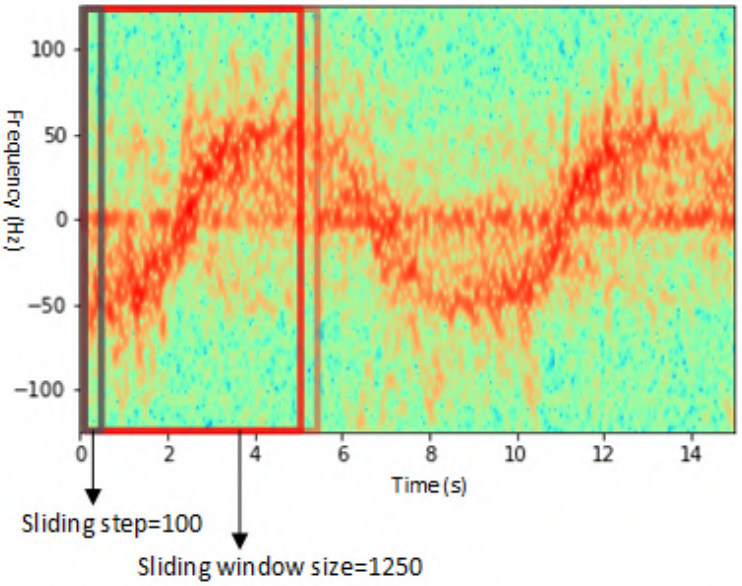


Figure 3.17: A sliding window of STFT

Table 3-F: 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		
Drone				2820
Background				5850

of the samples were used for training and validation, and 20% for testing. As can be seen from Table 3-F, the sample size for the classes is not balanced. 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 human recognition, because the numbers of spectrograms of the experiments with humans and the background are far greater than the number of spectrograms of the experiments with a dog (including human and dog), 2000 spectrograms of human and 2000 spectrograms of the background were randomly selected for the training.

3.6 Feature Extraction Using Two-Directional Two-Dimensional Principal Component Analysis

In this work, a time-frequency spectrogram presents the characteristics of human activity in the time-frequency domain. 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 recognition 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 [121] 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 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 [121] is adopted:

$$\begin{aligned} J(X) &= trace\{\mathbb{E}[(Y-\mathbb{E}(Y))(Y-\mathbb{E}(Y))^T]\} \\ &= trace\{\mathbb{E}[(AX-\mathbb{E}(AX))(AX-\mathbb{E}(AX))^T]\} \\ &= trace\{X^T\,\mathbb{E}[(A-\mathbb{E}(A))^T(A-\mathbb{E}(A))]X\}, \end{aligned} \tag{3.2}$$

where $J(\cdot)$ is an objective function. J must be maximized in order to find the optimal projection matrix X. $\mathbb{E}$ 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. (3.2) follows the commutative property of matrices where $trace(QP) = trace(PQ)$, Q and P represent any two matrices.

The covariance matrix [121] of A is defined as:

$$G = \mathbb{E}[(A-\mathbb{E}(A))^T(A-\mathbb{E}(A))], \tag{3.3}$$

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

$$G = (1/M)\sum_{k=1}^{M}(A_k-\bar{A})^T(A_k-\bar{A}), \tag{3.4}$$

where $\bar{A}$ is the average spectrograms as $\bar{A} = (1/M)\sum_{k=1}^{M} A_k$. Eq. (3.2) can be simplified

as:

$$J(X) = X^T G X, \tag{3.5}$$

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:

$$G_r = (1/M) \sum_{k=1}^{M} \sum_{i=1}^{m} (A_k^{(i)} - \bar{A}^{(i)})^T (A_k^{(i)} - \bar{A}^{(i)}), \tag{3.6}$$

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:

$$G_c = (1/M) \sum_{k=1}^{M} \sum_{j=1}^{n} (A_k^{(j)} - \bar{A}^{(j)})(A_k^{(j)} - \bar{A}^{(j)})^T, \tag{3.7}$$

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:

$$C = Z^T A X, \tag{3.8}$$

the matrix C is called the *coefficient matrix*. It can be taken as the input features that

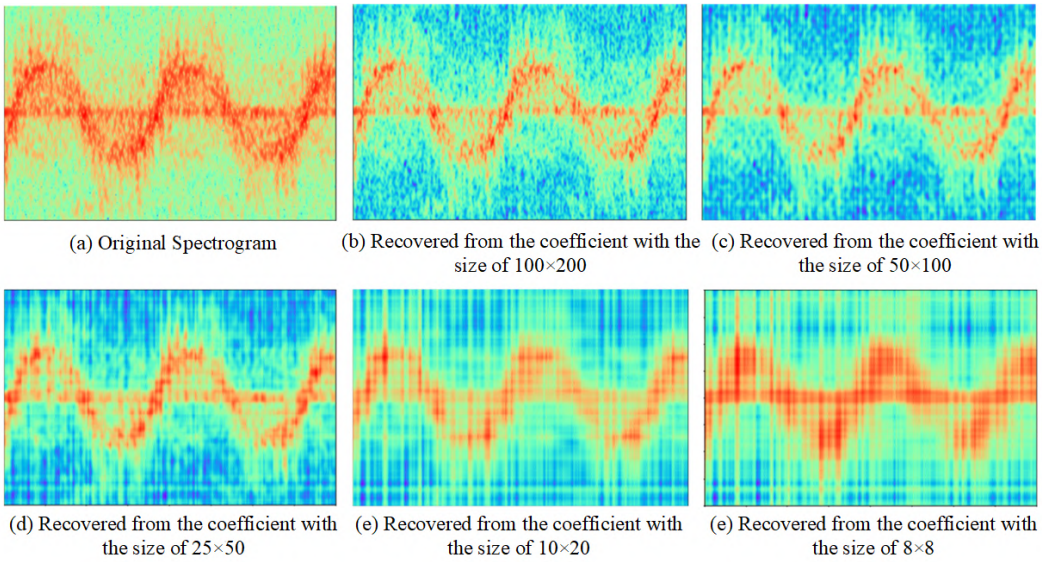


Figure 3.18: Dimension reduction with 2D2PCA

are fed into SVM and kNN classifiers.

Figure 3.18(a) is a spectrogram whose size is 2048×616 pixels, which is generated from a radar signal sequence with a length of 5000. It is the result of echoed radar signals of a human walking indoors. The method 2D2PCA was applied to this original spectrogram for different row and column dimension reductions. It is able to restore the spectrogram from the coefficient matrix C, which is obtained in Equation 3.8. The restore process can be formulated as:

$$A^{'} = ZCX^{T}, \tag{3.9}$$

where $A^{'}$ is the generated spectrogram after the restoration.

After restoration, it is able to visualize the effect of the dimension reduction by using 2D2PCA. Figure 3.18(b), (c), (d), (e), and (f) are restored from the coefficients with different number of row components and column components. Figure 3.18(b) and Figure 3.18(c) still retain detailed time-frequency characteristics (the periodic waves), visually they are almost the same as the original spectrogram. Even Restoring by using the coefficient matrix with the size of 8 rows and 8 columns, the remained characteristics

in Figure 3.18(f) still can reflect the periodic trend of human walking. 2D2PCA is able to keep most of the distinctive features (pixels) while greatly reducing the dimensions of the micro-Doppler signatures.

3.7 Machine learning models

Three classifiers, including CNN, SVM, and kNN 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.

3.7.1 Convolutional neural network

The radar system consisted of two BumbleBee radars. In order to fuse the signals from the radars, the spectrograms generated from them were 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 phase 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 Figure 3.19, the overlapped spectrograms with the shape of $50 \times 50 \times 2$ was input into a CNN built in this work, which was called RadarNet.

After investigating several different CNN designs for RadarNet, the complete CNN architecture that provided the best performance is described below and shown in Figure 3.19.

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 3×3 kernels to do the convolution. The C1 layer contains 16 feature maps, and the size of each feature map is $48 \times 48 \times 1$. A 2×2 filter is used to perform the *Max Pooling* on C1, and 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. Batch normalization [122] has been applied on $M1$, $M2$, $F1$, and $F2$ as a regularizer to accelerate convergence. Batch normalization normalizes the output of a previous activation layer by subtracting the batch mean and dividing by the batch standard deviation. The Batch normalization is presented in Algorithm 1. In the algorithm, ϵ is a constant added to the mini-batch variance for numerical stability [122].

Algorithm 1 Batch Normalization [122].

Input: Values of x over a mini-batch: $B = x_1 \ldots, x_m$; Parameters to be learned: γ , β
Output: $y_i = BN_{(\gamma,\beta)}(x_i)$

1: $\mu_B \leftarrow (1/m) \sum_{i=1}^{m} x_i$ // min-batch mean

2: $\sigma_B^2 \leftarrow (1/m) \sum_{i=1}^{m} (x_i - \mu_B)^2$ // min-batch variance

3: $\hat{x}_i \leftarrow (x_i - \mu_B)/\sqrt{\sigma_B^2 + \epsilon}$ // normalize

4: $y_i \leftarrow \gamma \hat{x}_i + \beta \equiv BN_{(\gamma,\beta)}(x_i)$ //scale and shift

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 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 [123].

3.7.2 Support vector machine

SVM is one of the most used classifiers in micro-Doppler based human activity detection. SVMs are discrete algorithms that can be used to find the optimal separating hyperplane

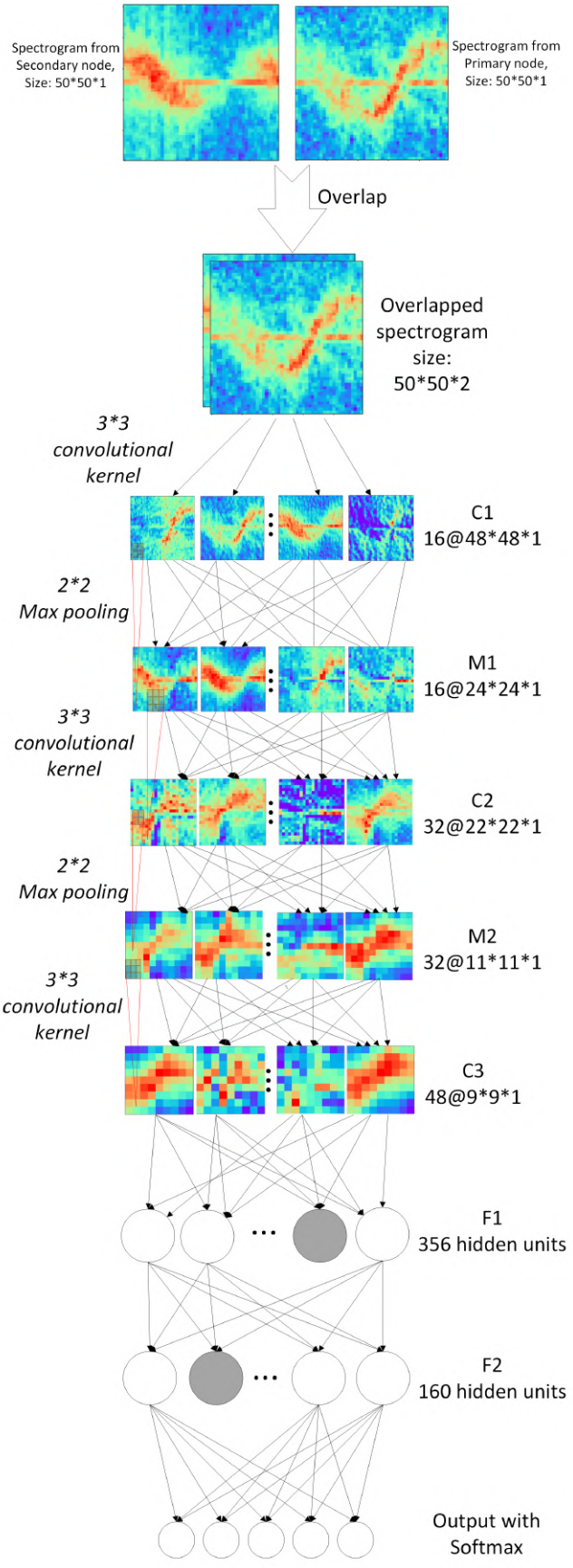


Figure 3.19: Structure of the CNN in human behaviour detection

Table 3-G: Optimal hyperparameters for different classification tasks

Task	**Model**	**Gamma**	**C**
Human recognition	SVM	$10^{-3.57}$	7
	SVM+2D2PCA	$10^{-5.84}$	3.1
Human activity classification	SVM	$10^{-3.35}$	7.34
	SVM+2D2PCA	$10^{-5.67}$	5.37
People counting	SVM	$10^{-3.26}$	6
	SVM+2D2PCA	10^{-6}	6.5
Coarse-grained localisation	SVM	$10^{-3.57}$	7.9
	SVM+2D2PCA	$10^{-6.07}$	2.9

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 [7, 40, 66, 84] where the features fed into SVMs were handcrafted, the input features used in this 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 3-G presents the optimal parameters for different classification tasks.

3.7.3 *k*–Nearest–Neighbor

The *k-Nearest-Neighbor* (kNN) classification 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 [124]. It also has been widely used in micro-Doppler based human activity detection [7, 110, 125]. The kNN classification algorithm itself is fairly straightforward and it can be summarized by the following steps:

1. Choose the number for k and a distance metric, which is commonly based on the Euclidean distance.

2. Find the k nearest neighbors of the sample that needs to be classified

3. Assign a class label by majority vote.

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:

$$\begin{aligned} Dist(S_c, S_i) &= Dist((f_1^c, \ldots, f_M^c), (f_1^i, \ldots, f_M^i)) \\ &= \sqrt{\sum_{p=1}^{M} (f_p^c - f_p^i)^2}. \end{aligned} \tag{3.10}$$

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 with a majority vote from the k nearest labeled samples. For each classification task, the chosen value of k is presented in Table 3-H.

Table 3-H: The value k of kNN in different Task

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

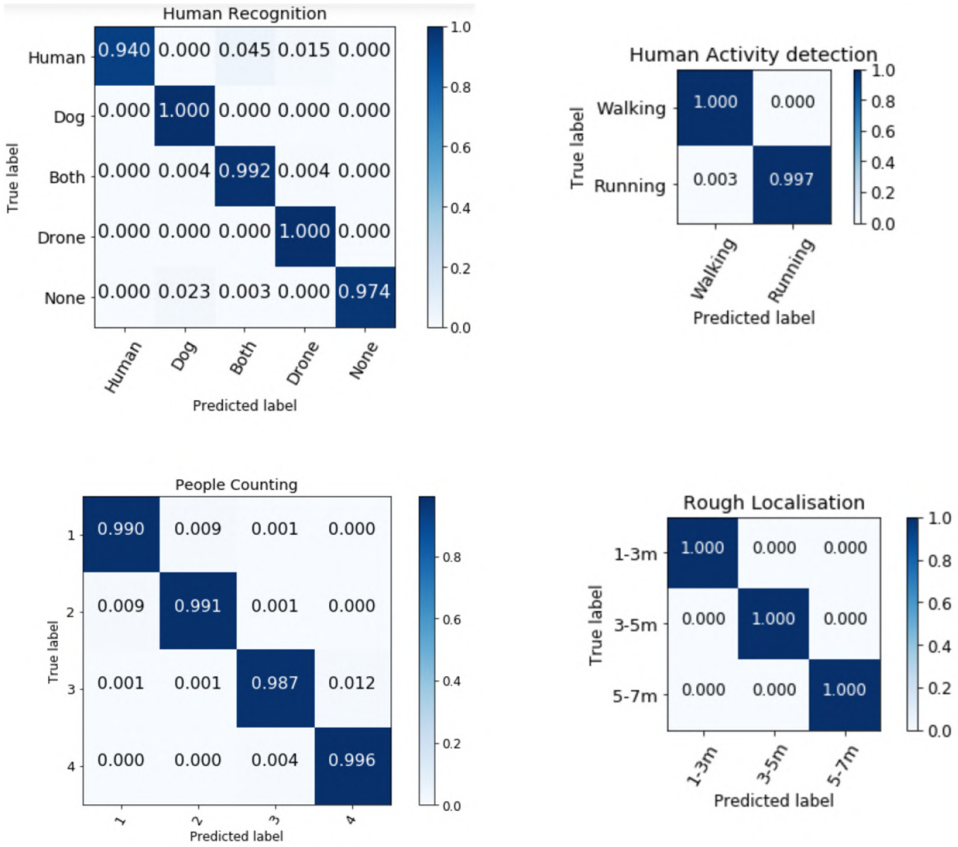


Figure 3.20: Normalized confusion matrices of CNNs for subject classification, activity classification, people counting, coarse localisation

3.8 Result analysis

The models built in Section 3.7 were used to classify the spectrograms. Different approaches were used to assess the accuracy obtained with these models for the different classification tasks.

3.8.1 Optimal performance

Before making a comparison between these five models, the optimal performance achieved by the RadarNet is presented by using confusion matrices, validation accuracy and validation loss in this section. Figure 3.20 shows the normalized *confusion matrices* of RadarNet for the four classification tasks. A confusion matrix is a table that is often used to describe the performance of a classification model (or 'classifier') on a set of test data for which the true values are known. All correct predictions are located in the diagonal of the table. A normalized confusion matrix is a confusion matrix with normalization. In Figure 3.20(a), the most misclassified samples were from the "Human', and most of them were classified into "Human and dog (Both)'. This implies that it is relatively difficult to differentiate the micro-Doppler signatures of these two classes and the micro-Doppler signatures of humans in 'human and dog' are more prominent. In Figure 3.20(b), 0.3% of 'running' were classified into 'walking', 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 are harder to differentiate. Figure 3.20(c) presents most of the misclassified samples from "1 person' that were classified into "2 people'; 0.9% of samples from "2 people' were misclassified as "1 person', and 1.2% of samples from "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.

Validation accuracy and validation loss are generated in validation of each epoch. By plotting the lines of validation accuracy and validation loss, it is able to present how the RadarNet converged to the best result visually. Figure 3.21 shows the lines of validation accuracy and validation loss for subject classification, activity classification, rough localisation, and people counting.

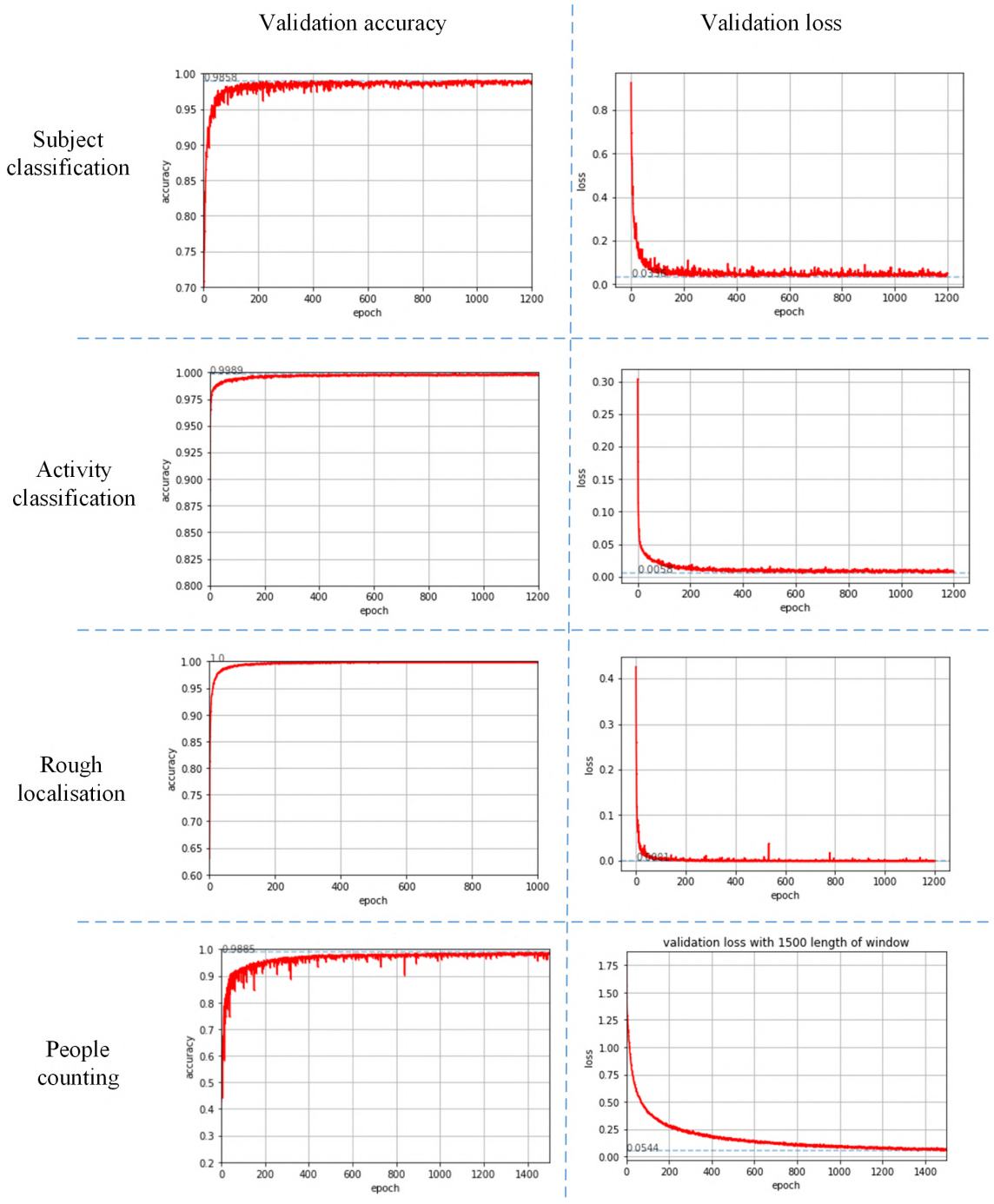


Figure 3.21: Lines of validation accuracy and validation loss of RadarNet for each classification task

3.8.2 Accuracy assessment

The performance of the classifiers is compared using different assessment methods: 1. overall classification accuracy (OA); 2. average class accuracy (AA); 3. Recall; 4. F1

score. OA is given as

$$OA = \frac{\sum_{i=1}^{C} n_{ii}}{n}, \tag{3.11}$$

where C is the number of existing classes, n_{ii} is the number of samples of class i that are classified rightly in the prediction and n is the total number of labeled samples used in the prediction. OA provides the rate of correctly classified samples. However, it is biased towards the classes that have a high frequency relative to other classes. Unlike OA, Recall, AA and F1 score provide an average of measures independent of class distribution. The Recall method can be calculated by

$$Recall = \frac{\sum_{i=1}^{C} n_{ii}}{\sum_{i=1}^{C} n_{i+}}, \tag{3.12}$$

where n_{i+} represents the number of reference samples.

The average class accuracy can be calculated by

$$AA = \frac{1}{C} \sum_{i=1}^{C} \frac{n_{ii}}{n_{i+}}. \tag{3.13}$$

The F1 score can be calculated by

$$F1 = 2\frac{Precision \times Recall}{Precision + Recall}, \tag{3.14}$$

where $Precision$ is given by

$$Precision = \frac{\sum_{i=1}^{C} n_{ii}}{\sum_{i=1}^{C} n_{+i}}, \tag{3.15}$$

where n_{+i} represents the number of samples that are predicted to be as class i.

Recall provides the rate of correctly classified samples of class $i, (1 \leq i \leq C)$ within all reference samples of class i, while AA computes the average rate of correctly classified spectrograms within each class, and F1 calculates the harmonic mean of the precision and recall.

Table 3-I: Performance in human recognition

MODEL	**OA**	**AA**	**RECALL**	**F1**
SVM	71.2%	67.1%	71.2%	70.2%
SVM+2D2PCA	97.57%	97.07%	97.57%	97.57%
KNN	57.5%	56.57%	57.5%	56.67%
KNN+2D2PCA	84.94%	85.1%	84.93%	85%
CNN (RadarNet)	98.58%	98%	98.58%	98.58%

3.8.3 Results for human recognition

The aim of human recognition is to differentiate humans from other targets based on their different micro-Doppler signatures. Animals are the most common confusers. In this thesis, a dog was used to represent four-legged animals. According to the type of the targets, the samples were divided into five categories (e.g. human, dog, human and dog, drone, and background/none). The performance of CNN, SVM, and kNN is shown in Table 3-I. RadarNet achieved the highest performance in all different metrics, with 98.58% in OA, 98% in AA, 98.58% in Recall, and 98.58% in F1. It outperforms SVM+2D2PCA with around 1% in OA, AA, Recall, and F1. The results of kNN+2D2PCA are around 85% in all four metrics. SVM and kNN had the poorest performance of around 57% and 71%, respectively.

3.8.4 Results for human activity detection

Two types of human activities (walking and running) were performed in the experiments. So the 'human' samples can be divided into two further classes. The performance of the classifiers are shown in Table 3-J. The best performance of 99.89% in OA, AA, Recall, and F1 was achieved by the RadarNet, and SVM+2D2PCA followed it closely around 99.4%. KNN+2D2PCA also achieved a good result from 97.2% to 98.35%. The above three classifiers outperformed SVM and kNN by a wide margin again.

3.8.5 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 the radar system.

Table 3-J: Performance in human activity detection

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%

Table 3-K: Performance in 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%

So 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 are shown in Table 3-K. RadarNet still achieved the best performance results around 98.85%, followed by SVM+2D2PCA around 95.8%. kNN+2D2PCA had 12.5% lower performance than SVM+2D2PCA. SVM and kNN had the poorest performance.

3.8.6 Results for coarse-grained localisation 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-grained localisation 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-grained localisation are shown in Table 3-L. The overall performance achieved in coarse-localisation 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. kNN+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, kNN presented the lowest scores, achieving a percentage of around 81%.

Table 3-L: Performance in coarse-grained localisation

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%

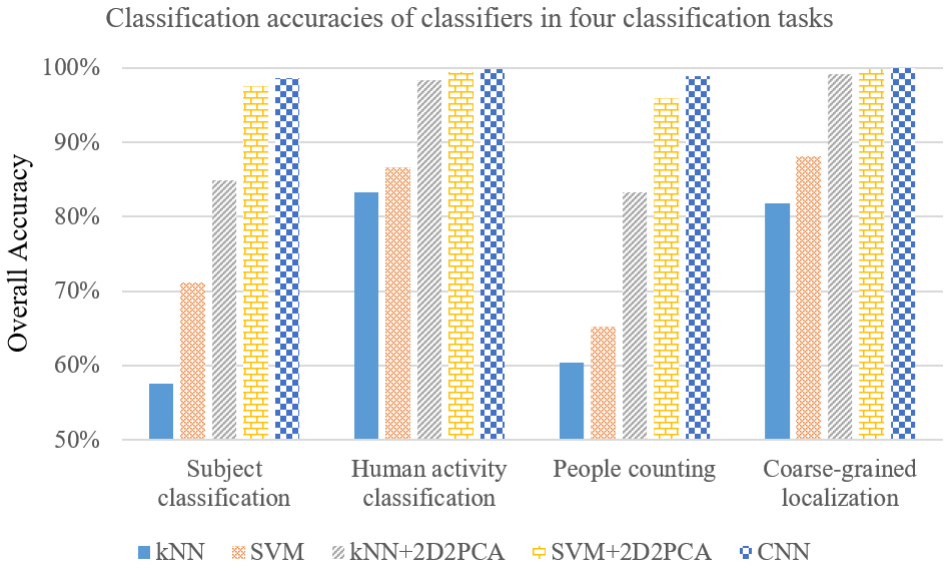


Figure 3.22: The performance of the five classifiers for four classification tasks

3.8.7 Analysis

Based on the results above, Figure 3.22 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 kNN+2D2PCA. The classifiers with the lowest accuracy results were SVM and kNN. 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 98.58% in subject classification, 99.89% in human activity classification, 98.85% in people counting, and 100% in coarse localisation. SVM+2D2PCA followed closely and kNN+2D2PCA was slightly inferior to SVM+2D2PCA, but both exceeded SVM and kNN by a wide margin (11%-18%). This shows that the performance of SVM and

kNN 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. People counting is the most difficult classification task, then followed by subject classification and activity recognition. Rough localisation is the easiest classification task. This can be seen in Figure 3.22, the average performance (OA) of all classifiers for coarse localisation 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.

3.8.8 Comparison with the related work

There is some other similar work that is worth comparing with this work. In human and animal classification, the authors in [96] 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 at 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 the confuser was a different animal in this thesis, a better result was achieved. 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 [96] has better frequency and distance resolutions than the Bumblebee radar, a well-modeled CNN like RadarNet is able to compensate for these limitations.

In human activity classification, the authors in [66] 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 this work, only two activities (walking and running) were considered, but the experiments were performed in an environment comparable to the leaf tree foliage and the needle tree foliage with considerably better results. In [66], 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 kNN method. Their classifier correctly classified the activity of walking 90% of the cases, 88% for running, and 93% for crawling. Although the same radar used in [66] , they made the experiments in less noisy conditions (indoor environment), but their results are still not comparable to the 99.8% OA achieved in this work.

For people counting, the authors in [33] 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 kNN classification approach, they achieved an overall classification rate of 80%. The authors in [70] 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 the 98.85% overall accuracy achieved in this thesis for people counting is a good performance.

For micro-Doppler based localisation, to the best of our knowledge, there is still no relevant comparable work. Because of the low-power consumption, some functions and the range of the radar system are compromised. According to the Radar Range Equation [29], if the operating frequency, the antenna gain, and the smallest power of a radar keep the same, the transmitting power of the radar transmitter needs to increase 16 times in order to double the detection range of the radar. But the transmitting power is not the total power of the radar, it still also depends on the manufacturing

level of radar. The methods proposed in this research can be transplanted to other radar systems with a larger detection range. The advantages of the outdoor system are its low-power consumption and high performance (accuracy) in human recognition, people counting, and activity classification. It can be implemented in the key places outdoors for surveillance, warning of dangers, etc.

For the CNN model, it is difficult to make a direct comparison because there is no benchmark dataset publicly available in radar-based activity detection. The accuracy 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, it is worth to describe some of the differences between this work and [70, 71] apart from the DCNN designs. In [70], 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%. However, people counting (from 1 to 4 people) performed in this thesis achieved an accuracy of 98.85%. In [71], the authors used a CW radar operating at 7.25 GHz to perform human recognition outdoors at 0° and a CW radar operating at 2.4 GHz to perform human activity recognition indoors at 0° both in a clear line of sight environment. They investigated subjects and activities different from the work presented in this thesis. Their CNN obtained 97.6% overall accuracy in human target classification, while 98.58% was achieved using a simpler radar system in a more cluttered outdoor environment in this research.

Therefore, it is plausible to infer that the methods including the signal processing and the classifiers that were investigated and tested in this research, are solid implemen-

tation 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. This work promotes the micro-Doppler based application in long-term continuous human activity detection outdoors.

3.9 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 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.

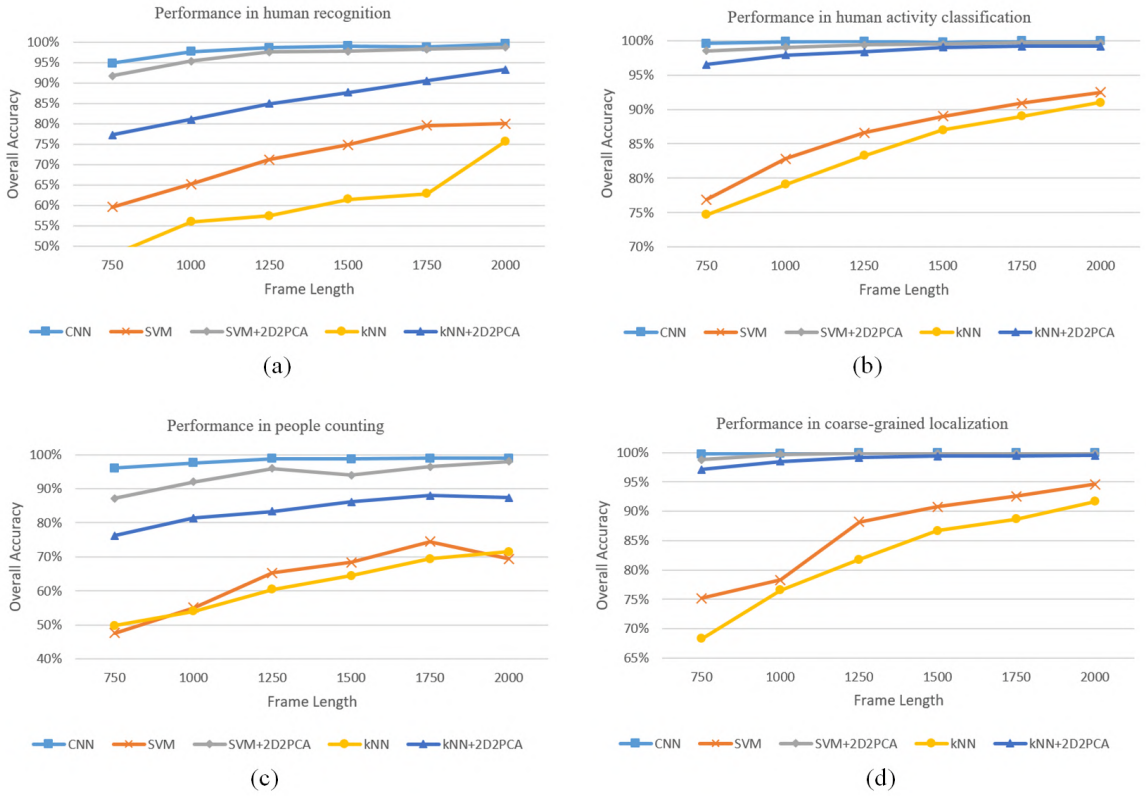


Figure 3.23: Performances (OA) of the classifiers with the changing frame lengths, (a) human recognition, (b) human activity classification, (c) people counting, (b) coarse-grained localisation

3.9.1 The frame length of the sliding window

As mentioned in Section 3.5, 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 different frame lengths affect the classification.

Six frame lengths, including 750, 1000, 1250, 1500, 1750, and 2000 samples were investigated. As the sampling 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, respectively. The five classifiers were used to perform the classification with the generated spectrogram samples. As shown in Figure 3.23, the performances (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 classifier's performance remained the same as stated in Section 3.8. 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. This thesis used a 5s frame length that led to good accuracies 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.

3.9.2 Angles of the movement

In Section 3.4, it describes that the experiments were made from three different angles (0°, 45°, and 90°). The coarse-grained localisation made in *Case 3* was investigated from two angles (45° and 90°). As described in Figure 3.5, the detection area was split into 3 ranges, the length of these ranges are from 10m to 12m in vertical aspect

(90°), but only 2m in horizontal aspect (0°). 2m is too short and very inconvenient for participants to perform walking and running. They would keep turning around in this short range. So this research only investigated coarsed-grained localisation from 45° and 90°. As shown in Figure 3.24, the classifications for human activity detection and people counting performed the best with spectrograms from 0°, reaching overall accuracies of 100% and 99.46% respectively. Samples from 45° produced the worst results (99.7% in human activity detection and people counting). For the coarse-grained localisation, the same OA (100%) was achieved at both 45° and 90°. This means the angle of the movements has no effect on the classification accuracy in localisation. In conclusion, the direction of the movement to the radar beam can affect the classification in human activity and people counting, and 0° can provide the best accuracy, followed by 90° and 45°. This is probably because the RCS is largest when people move at 0°. While for the coarse-grained localisation, the angle of movement had no effect.

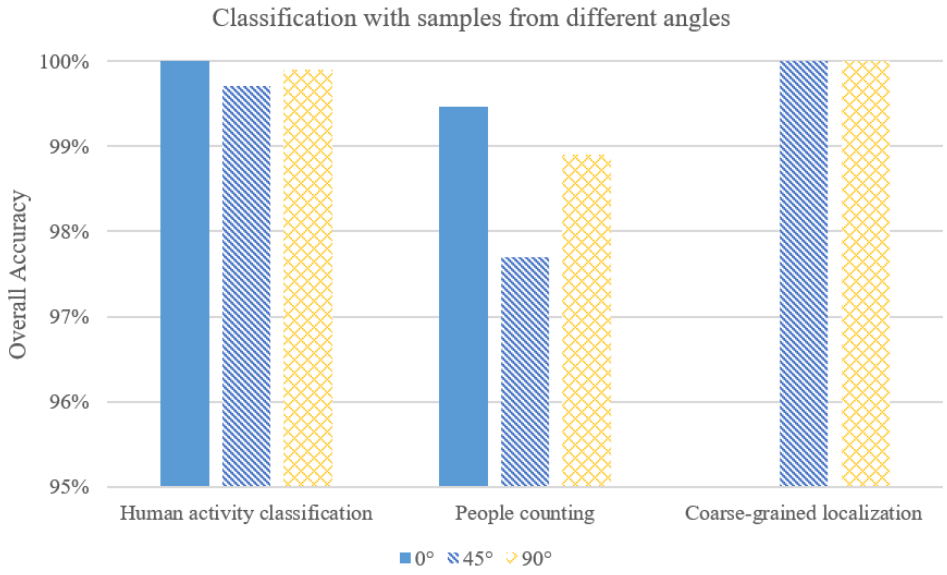


Figure 3.24: Performances of the classifiers in the tasks with different angles

3.9.3 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. This section investigates how the

Table 3-M: Structure of RadarNet and RadarNet$_{\text{one}}$

RadarNet	**RadarNet$_{\text{one}}$**
X ($50 \times 50 \times 2$)	X ($50 \times 50 \times 1$)
Conv1-16@$48 \times 48 \times 1$	Conv1-12@$48 \times 48 \times 1$
maxpool	Maxpool
BN	BN
Conv2-32@$22 \times 22 \times 1$	Conv2-24@$22 \times 22 \times 1$
maxpool	Maxpool
BN	BN
Dropout (0.3)	Dropout (0.3)
Conv3-48@$9 \times 9 \times 1$	Conv3-24@$9 \times 9 \times 1$
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)	

number of radars affects classification. For this purpose, a new CNN model 'RadarNet$_{\text{one}}$' was trained using only the data from the primary radar node.

The main difference between RadarNet$_{\text{one}}$ and RadarNet is the input size, which is $50 \times 50 \times 1$ (one spectrogram) for RadarNet$_{\text{one}}$, and $50 \times 50 \times 2$ (two overlapped spectrograms) for RadarNet. As seen in Table 3-M, the structure of both CNNs are very similar, both contain 3 convolutional layers, 2 max-pooling layers, and 2 fully connected layers. However, RadarNet$_{\text{one}}$ presents less feature maps and hidden units. The comparison of their performances is shown in Figure 3.25. The performance obtained from the two radars was higher than the performance obtained by using only one radar in all four classification tasks. For human activity detection and coarse-grained localisation, the results presented by RadarNet$_{\text{one}}$ were slightly inferior to those of RadarNet. While in human recognition, and especially in 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$_{\text{one}}$ was 91.57%, while for RadarNet was 98.85%. In human recognition, the overall accuracy with one radar was 95.70%, while with two

radars reached 98.58%. 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 to human activity detection and coarse-grained localisation, but very obvious to human recognition and people counting.

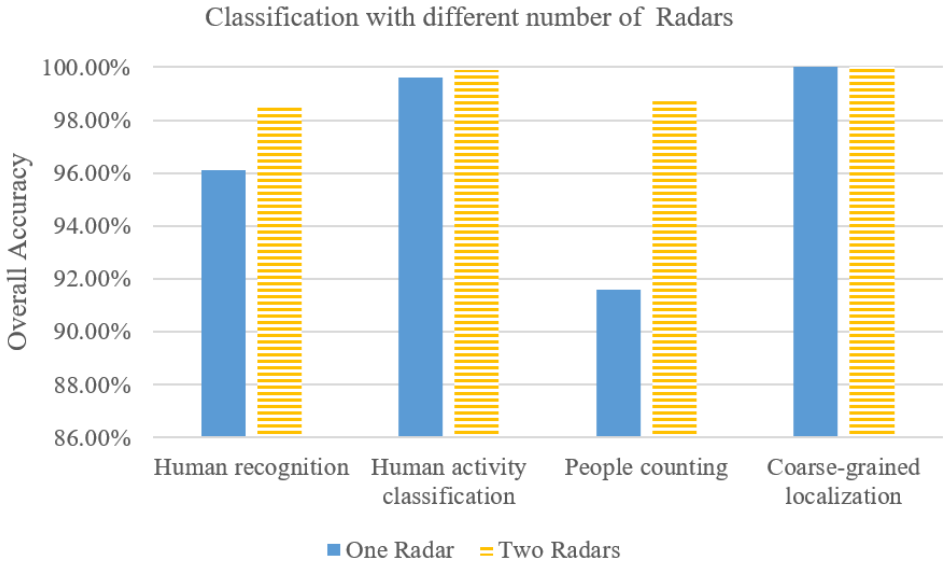


Figure 3.25: CNN classification with samples of different number of radars

3.10 Summary

This chapter applied micro-Doppler signatures to perform four classification tasks, which are subject classification, human activity classification, people counting, and coarse localisation. A radar system that consists of two pulsed Doppler radars operating at 5.8 GHz was built. 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, kNN, kNN+2D2PCA were implemented. It was found that 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 kNN 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. This investigation provided a valuable guideline for model optimization and experiment setup of micro-Doppler based research and applications.

In conclusion, this chapter shows that low-power low-cost radars have great potential for human activity detection, even in outdoor environments surrounded by trees and foliage.

Chapter 4

Human activity recognition indoors using micro-Doppler signatures

According to statistics presented in [126], humans spend more than 80% of their life indoors. Human activity recognition indoors is very important and has been widely researched. For micro-Doppler signatures based human activity recognition indoors, most researched activities are walking, running, crawling, falling, etc. These activities are human motions. It is difficult to infer our daily events from these motions. It is expected to extract higher level-activity information such as eating, drinking, watching TV, sleeping, etc., by human activity recognition. This chapter presents the work of human activity recognition in a kitchen scenario in order to provide an available way to monitor human dietary activities.

4.1 Introduction

Human activity recognition plays a crucial role in our daily life. Specifically, human activity recognition at home enables the development of important applications for assisted living for the elderly and people in need of rehabilitation and support due to mental health problems (e.g. dementia and depression). For example, statistics agree that in the next 30 years the number of elderly people will increase about a third, therefore, there is an increasing demand for remote healthcare systems for one-person households as it facilitates independent living in a smart home setting and does not require the presence of caregivers at all times. Routine activities such as eating, drinking, dressing, bathing, and toileting are commonly referred to as Activities of Daily Living (ADL), which people tend to do on a daily basis for normal self-care [127]. Nearly real-time ADL monitoring can allow timely intervention for medical emergencies, for treatment and rehabilitation support, obesity and diet control or other necessary behavioural health assessment linked to a physical/mental disease or injury. Activities in the kitchen are directly related to dietary health. For example, the indication that a person gets in or out of a kitchen can be used to estimate his/her dinning frequency, time and duration of dining; the activity detection of eating and drinking can provide the information of the intake of food and drink. Kitchen scene context-based activity detection can be a useful method for diet controlling and dietary treatment, also helpful for developing a smart kitchen as a part of a smart home.

Micro-Doppler signatures can be used to distinguish different moving objects and human activities. Researchers have applied micro-Doppler signatures in indoor activity recognition [86], differentiation between human and animal [128], heartbeat and respiration detection [129], etc. However, most radars used in these research works are power-consuming and high-cost, which preclude the use of radar in many scenarios. The use of low-power low-cost radars to perform micro-Doppler based human activity recognition are more accessible and suitable for long-time continuous human activity recognition. In [130], a radar-based network consisting of two short-range radars (25

GHz) and two long-range radars (5.8 GHz BumbleBee) were used to classify indoor activities. The highest accuracy was achieved when fusing data of the two short-range radars and using a random forest classifier. However, they used a much longer frame length and the short-range radars have higher frequency resolution than the BumbleBee radar. The approach proposed in this chapter achieves the same performance with only two BumbleBee radars in a more challenging scenario.

In this chapter, Bumblebee radars were used to detect human activities in a realistic kitchen scenario. The low-power radar-based sensor network built in this research is suitable for long-time continuous human activity recognition and is is easy to deploy. Combining the radar network with DCNNs, makes the system able to perform nearly real-time human activity recognition in a kitchen scenario. The contributions presented in this chapter can be summarized as follows:

1. A non-intrusive and device-free system was built by using a low-cost low-power radar-based wireless network. By using a very short time (2.5s) window to perform segmentation, it is able to achieve nearly real-time activity recognition .

2. Human activities are investigated in a very challenging scenario. Activities in the kitchen are difficult to be separately distinguished as they are short in duration, they occur in a confined space with multiple appliances and furniture, and the activities may be chained in many different sequences. The result on the recognition of several different activities in a single confined house room using a device/tag free system is a valuable source for future research and comparison on the field.

3. A Deep Convolutional Neural Network (CNN) was built for the target activities, it performs automatic feature extraction and classification on radar signals. The DCNN provided a high rate of classification success, and most importantly, it achieved very good classification results in near real-time.

4.2 Related work

People spend more than 80% of their lifetime indoors. Most human activities happen indoors, consequentely with modern living there is a need to perform indoor human activity recognition. Human activity recognition indoors is often used in the treatment and assessment of patients, continual in-home gait and life signs monitoring, and security in a smart home setting. As noncontact sensors, Doppler radars are very suitable to be applied in remote health monitoring, and smart home monitoring and control without being affected by light or intruding people's privacy. Doppler radars have been widely used in human activity recognition indoors. In [131], the authors built a hybrid radar system that incorporates a linear frequency-modulated continuous-wave (FMCW) mode and an interferometry mode for indoor human localisation and life activity monitoring applications. The FMCW mode is able to obtain the range information of the subject. The interferometry mode is utilized to measure micro-Doppler signatures of life activities. This hybrid radar system is used to perform localisation, respiration pattern measurement, and activity recognition. In [132], the authors applied a UWB radar in continuous in-home monitoring by detecting the vital signs (respiration and heart rates). With the classification by EMD, the overall accuracies reach up to 95% and 91% in detecting respiration and heart rates, respectively. In [133], the authors applied a deep neural network to classify aided and unaided human activity indoors. A CW radar operating at 4 GHz was used to measure 12 different activities. The performance achieved by the deep neural network was 89% in overall accuracy, while SVM only obtained 72%. In [134], the authors proposed behavior-based person identification with the use of an ultra-wideband impulse radar. A CNN was built and it achieved 95.21% accuracy on the identification of 15 people.

Currently, most approaches for human activity recognition are based on supervised machine learning. Although some activities such as heartbeat and breath have been investigating without training with machine learning algorithms, they are simple and they generate distinctive patterns (frequency peaks) in frequency. For example, some

research [135–137] detected heart-rate, and breath rate by extracting frequency peaks of the signals. In [135], the achieved accuracies for heart-rate detection in different distances between the subject and the radar are different, which are 96% at 1m, 81.5% at 2m, and 64.6% at 2.5m. However, most human movements and activities are too complex to be detected by just using signal processing. Most importantly, machine learning achieved results that are better than other methods. Again for heart-rate detection, the work in [138] achieved 98% accuracy for heart-rate detection at a distance of 1.3 between the subject and the radar and above 97% accuracy at the distance of 3m by using deep learning, which are better than the result achieved in [135]. Since deep learning has achieved state-of-the-art in many tasks, this work integrated deep CNN with micro-Doppler analysis for human activity recognition indoors.

Various activities have been investigated indoors for different purposes and applications by using micro-Doppler signatures. Table 4-A summarizes the activities investigated in the related work and their corresponding applications.

To the best of the author's knowledge, there is no relevant research applying micro-Doppler analysis to daily dietary activity recognition and monitoring. Dietary habits affect people's health in various ways. An imbalanced diet elevates health risks for many chronic diseases including obesity. Daily dietary behaviour monitoring can give health care providers an insight into the dietary habits of patients, like when, how much, and how frequently a patient go to the kitchen for eating and drinking. With Doppler radars, the long-term and noncontact monitoring of daily dietary behaviour can be achieved.

4.3 Implementation

This section details the setup of the Doppler radar network. The radar sensor network was deployed in a kitchen to recognize fifteen human activities. Also, this section presents the DCNN model built to classify the activities by using the micro-Doppler signatures.

Table 4-A: Activities and applications of human activity detection indoors using micro-Doppler signatures

Activities	**Sensors**	**Application**	**Citation**
Respiration, heartbeat, and body movement	A CW Doppler radar	Sleep monitoring and sleep quality evaluation	[139]
Falling, walking, walking with carrying an object, sitting down, standing up, picking up a phone, Bending, etc.	A FMCW radar operating at 5.8 GHz and RGB-D sensors	Falling detection for elderly people	[140]
Walking, running, jumping, turning, and standing up	A passive radar system	Activity of daily living monitoring	[141]
Falling, bending over, sitting, and standing.	A CW radar operating at 8 GHz	Falling detection	[142]
Respiration, heartbeat, walking, sitting while waving hands, and	FMCW radar operating at 5.8 GHz	Tracking individual location and monitoring life activity	[131]
Cooking, preparing meals, washing dishes, eating, sitting on the couch, resting in bed	A pulsed-Doppler radar operating at 4.3 GHz	Detecting early signs of illness and functional decline.	[132]
Walking, limping, falling, creeping, crawling, jogging and sitting	A CW radar operating at 4.4 GHz	Remote health monitoring	[133]
Walking	A pulse Doppler radar operating at 5.8 GHz	Estimating the speed of walking, and step recognition	[143]
Vital signs, minute motions, and major bodily motions	A short-range 60 GHz compact radar	People counting and occupancy sensing	[144]

4.3.1 Experimental setup

The low-power low-cost radar sensor network implemented in the experiment is almost the same as the radar system described in Chapter 3. It consists of two BumbleBee radars and three TelosB motes (see Figure 4.1). The sensor network contains two nodes ('Node 1' and 'Node 2') and one base station. Each node consists of a BumbleBee radar and

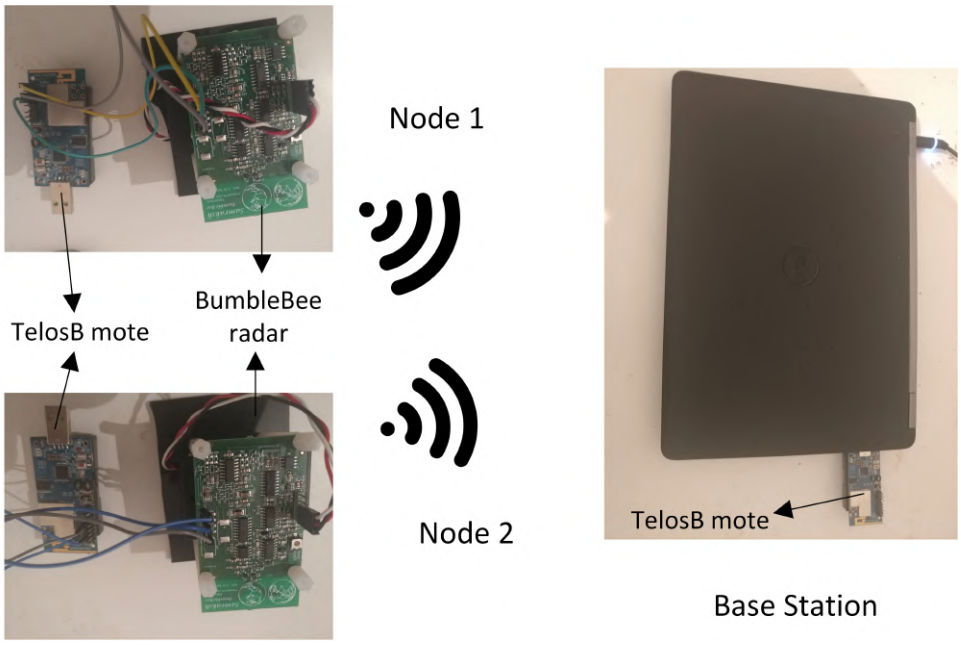


Figure 4.1: A wireless radar sensor network

a TelsoB mote. The TelosB mote is connected to the BumblebBee radar and transfers data packets that contain radar signals to the base station. The base station is built with a TelsoB connected to a PC, where the received data packets are processed.

This network was deployed in a kitchen. The plan of the kitchen is shown in Figure 4.2(a). In order to cover the whole kitchen space, two nodes (green rectangles in Figure 4.2(a)) were placed in two diagonally opposite corners of the kitchen, with an approximate angle of 30° in relation to their left wall. The height of both nodes was 1.2 meters. Figure 4.2(b) presents the bird's-eye view of 3D model of the kitchen. As it can be seen, the kitchen contains a long rectangular table, chairs, a cabinet, a microwave oven, a sink, etc.

4.3.2 Data collection

Fifteen activities of three people were investigated in the kitchen, including (a) Walking, (b) Eating, (c) Drinking, (d) Sitting, (e) Standing, (f) Washing, (g) Open door and get in, (h) Open door and get out, (i) Open cabinet, (j) Close cabinet, (k) Open oven, (l) Close oven, (m) Open freezer, (n) Close freezer, and (o) No activity. Each activity were

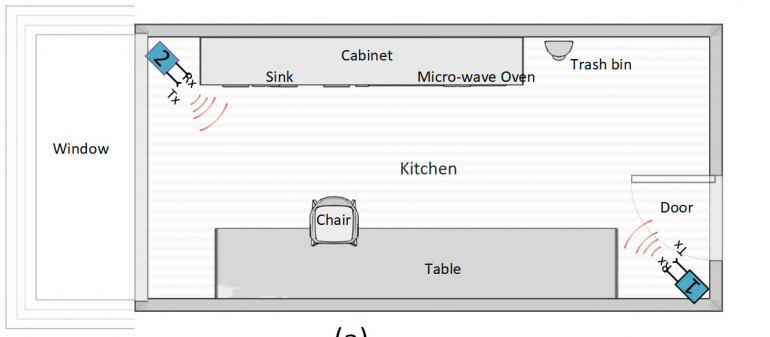


Figure 4.2: Kitchen scenario (a) Plan, (b) 3D bird's-eye view

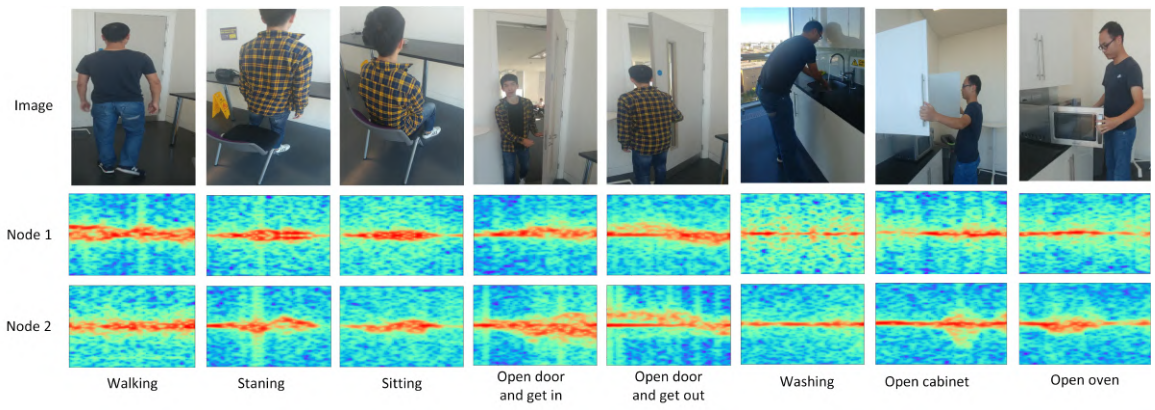


Figure 4.3: Spectrograms of human activities in the kitchen

performed hundreds of times by these three people. Radar signals for these activities were collected simultaneously by two BumbleBee radars in the network. Figure 4.3 shows the images and the spectrograms of several targeted activities measured by the radar sensor network.

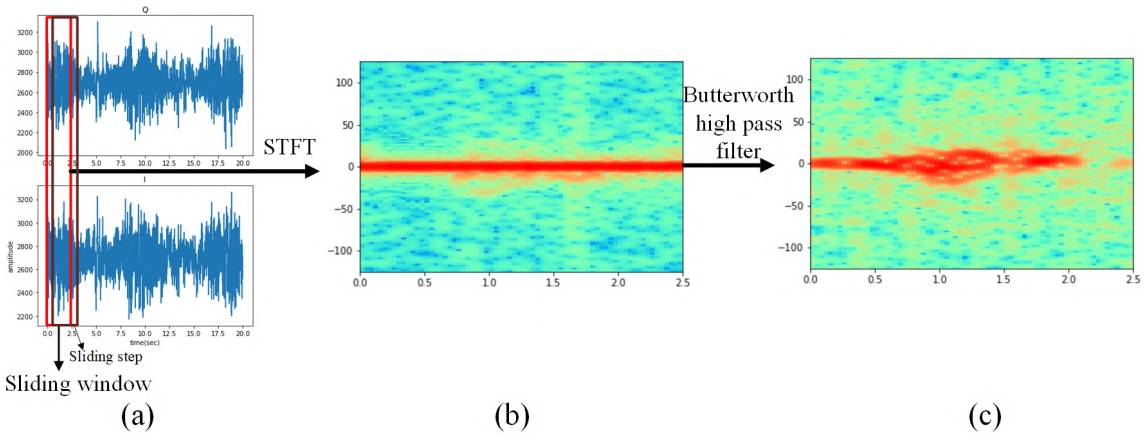


Figure 4.4: Data processing, (a) radar signals, (b) spectrogram after STFT, (c) spectrogram after high-pass filter

4.3.3 Data processing

Data processing implemented in this work is almost the same as the data processing described in Chapter 3. The only difference is the length of the sliding window used for signal segmentation. The durations of different activities are different. Walking, eating, and drinking take longer time than other activities. So, it is required to select a suitable length for the sliding window, in order to make sure the window covers at least one motion cycle of each activity. As shown in Figure 4.4(a), a sliding window of 2.5s was applied to create spectrograms with the interval of 0.5s. As the sampling rate of each node is 250 Hz, a sliding window contains 625 signals.

After signal segmentation, the segmented signal sequences were transformed from amplitude domain to frequency domain by using STFT. The generated spectrogram is shown as Figure 4.4(b). Because of the DC component, the micro-Doppler signatures of human activity is vague. For attenuating the DC component, the Butterworth high pass filter was implemented.

After radar signal processing, 15350 spectrograms were obtained in total in the experiment. The composition of the samples/spectrograms is shown in Fig 4.5.

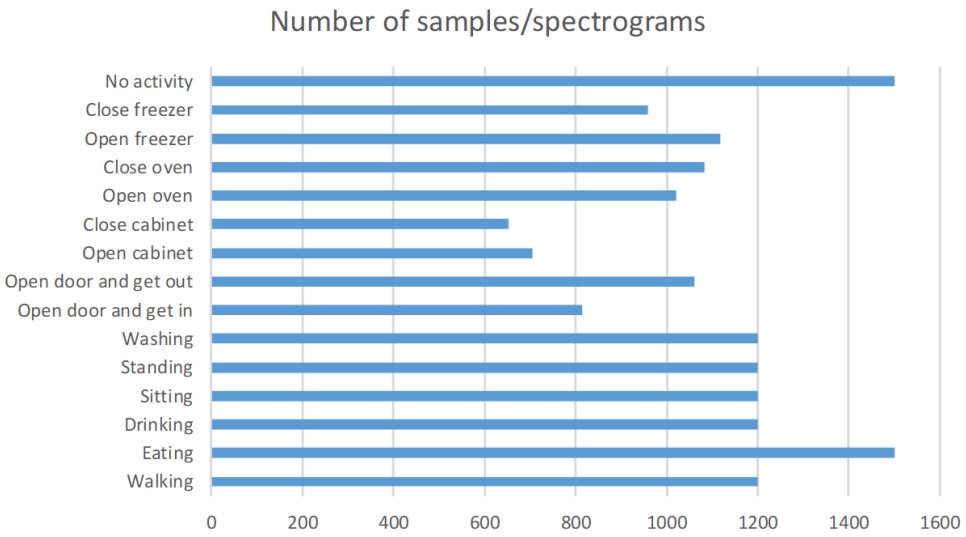


Figure 4.5: Number of samples of each type of activity

4.3.4 Micro-Doppler signature analysis

A large amount of spectrograms were collected from the experiments. Before classification of these spectrograms, it is worth to have a deep insight into the differences of different activities. Four typical spectrograms of four activities are described and analyzed as follows:

1. **Drinking**: a person picks up a cup on the table and has a drink, then puts it back. As it can be seen from the spectrogram in Figure 4.6(a), there are four lumps around the 0 frequency. The first two of them result from the subject picking up the cup. The front one reflects the movement of elbow, the latter one reflects the movement of the arm. The last two result from the subject putting the cup back. Conversely, the front one represents the movement of the arm, and the latter one represents the movement of the elbow. The frequency directions of the arm and the elbow are opposite, because when the subject picks up and puts down the cup, the elbow and the arm move in opposite directions.

2. **Eating**: a person scoops a spoon of food to eat, then puts down the spoon (Figure 4.6(b)). In the corresponding spectrogram, there are two lumps produced around

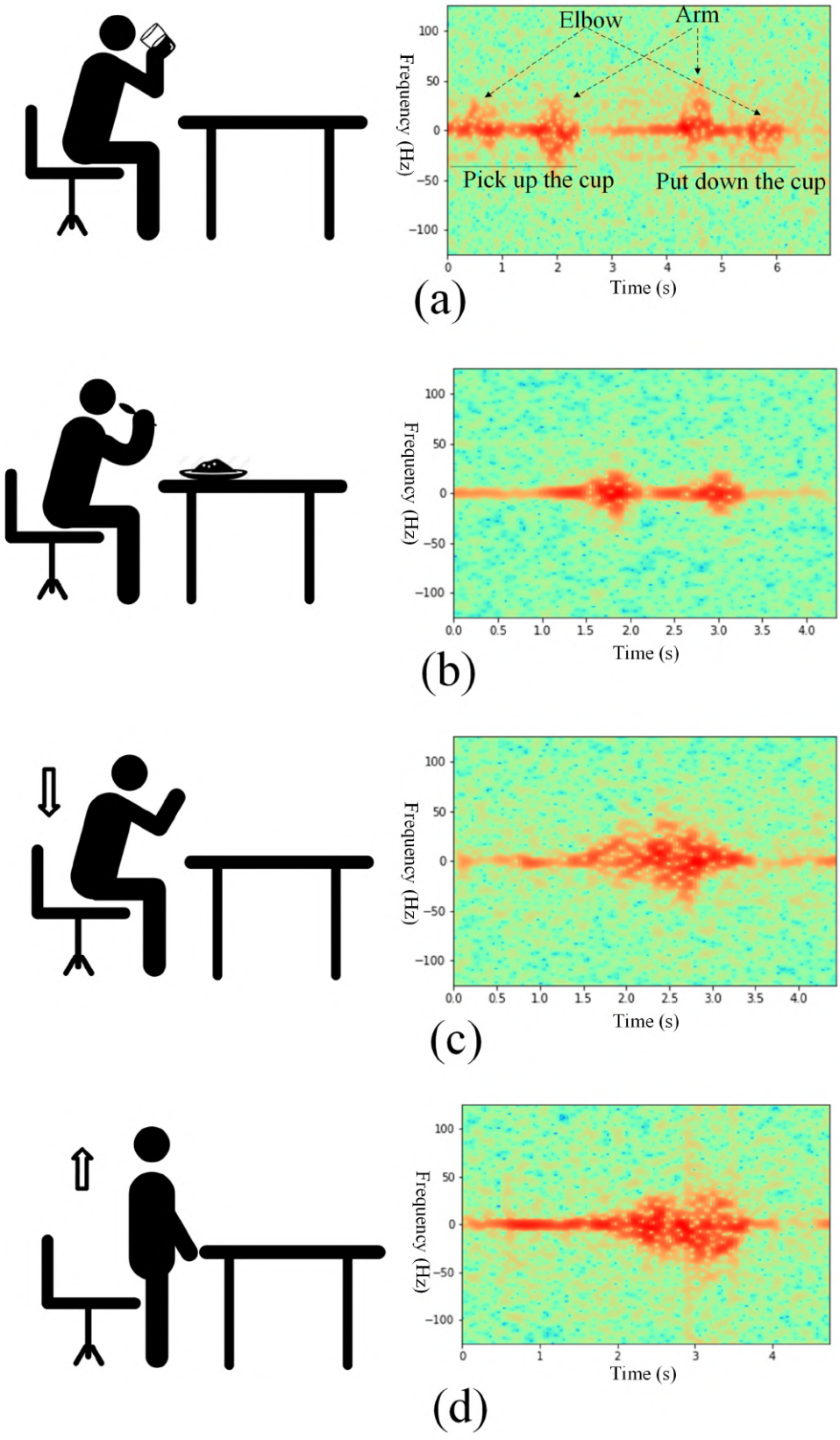


Figure 4.6: Four indoor activities and their spectrograms. (a) drinking, (b) eating, (c) sitting, (d) standing

zero. The front one is generated by picking up the spoon, and the latter results from putting the spoon down.

3. **Sitting**: a person sits down to a chair as shown in Figure 4.6(c). There is a big lump produced by this movement. The upper part of the lump is more prominent than its lower part.

4. **Standing**: a person stands up from a chair as shown in Figure 4.6(d). There also is a lump produced. Contrary to the spectrogram of sitting, the lower part is more prominent than its upper part.

In the experiments, not all spectrograms are the same as described above due to the variable factors such as the position of people and the multipath effect, but spectrograms of one activity have similar patterns. This section only presents the spectrograms of four activities because their patterns are relatively simple; other activities are more complex and scenario-dependent, but each activity has its own micro-Doppler signatures that can be used for recognition.

4.3.5 The DCNN Model

The radar-based network contains two radar nodes. Each node observes human activities from a different perspective. In order to fuse the signals from the radars, the spectrograms generated from them are firstly down sampled into the size of $80 \times 80 \times 1$, then overlapped together forming a spectrogram with the size of $80 \times 80 \times 2$ (it can be considered as an image with two channels). A DCNN model was built to take the fused spectrograms as input. The structure of the DCNN is shown in Figure 4.7. It contains three convolutional layers (C1, C2, and C3), two Max Pooling layers (M1, M2) and two fully connected layers (F1, F2). All three convolutional layers use 3 kernels to do the convolutionalization. The C1 layer contains 20 feature maps, and the size of each feature map is $78 \times 78 \times 1$. A 2×2 filter is used to perform the Max Pooling on C1, and generate the M1 layer. The C2 layer contains 36 feature maps, and the size of each feature map is $37 \times 37 \times 1$. The M2 layer is produced by Max Pooling the C2 layer. The C3 layer

contains 56 feature maps; each feature map is a $16 \times 16 \times 1$ matrix. The F1 layer contains 550 hidden units, and the F2 contains 240 hidden units. In the training process, dropout [145] has been used to control overfitting with an initial dropout rate of 0.4. Batch normalization [122] has been applied on M1 and M2 as a regularizer to accelerate the convergence. 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.

In the training process, the samples were randomly divided into two groups, 80% of the samples were used for training and 20% of them for testing. As it can be seen from Figure 4.5, the sample size of each class is different. For overcoming this imbalanced problem, different weights were assigned to each class and the weight ratio was inversely proportional to the proportion of samples of each class.

The DCNN was trained on a computer server containing two E5-2680 2.7GHz CPU, one K40 GPU and one K20 GPU, 128GB of RAM, and installed with Linux system. After training for around 500 epochs, the DCNN reached convergence.

4.4 Evaluation

In this section, the performance of the DCNN model is evaluated in three ways.

4.4.1 Evaluation on test dataset

The first evaluation was made by classifying a new dataset that has never been exposed to the DCNN model during the training. The achieved overall accuracy was 92.81%. As shown in the normalized confusion matrix of Figure 4.8, the accuracy achieved in 'Drinking' was 82%, which was the lowest. Most of the misclassified samples in 'Drinking' have been classified into 'Eating'. This is because there are a lot of similarities between these two activities. The classification rates of most activities were above 90%. Especially, 'Walking', 'Open door and get in' and 'Open door and get out' achieved the classification rate of 100%.

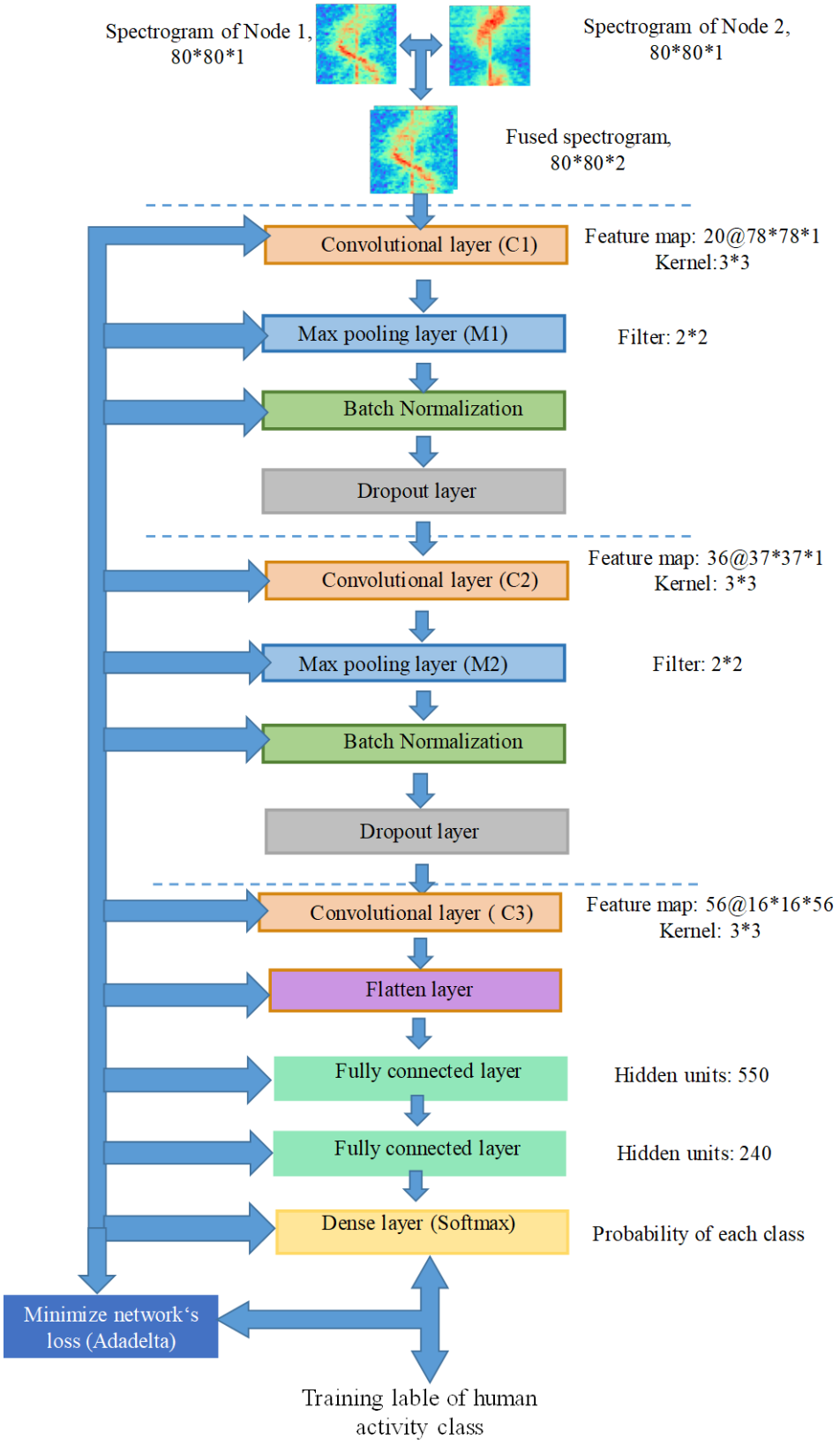


Figure 4.7: DCNN model

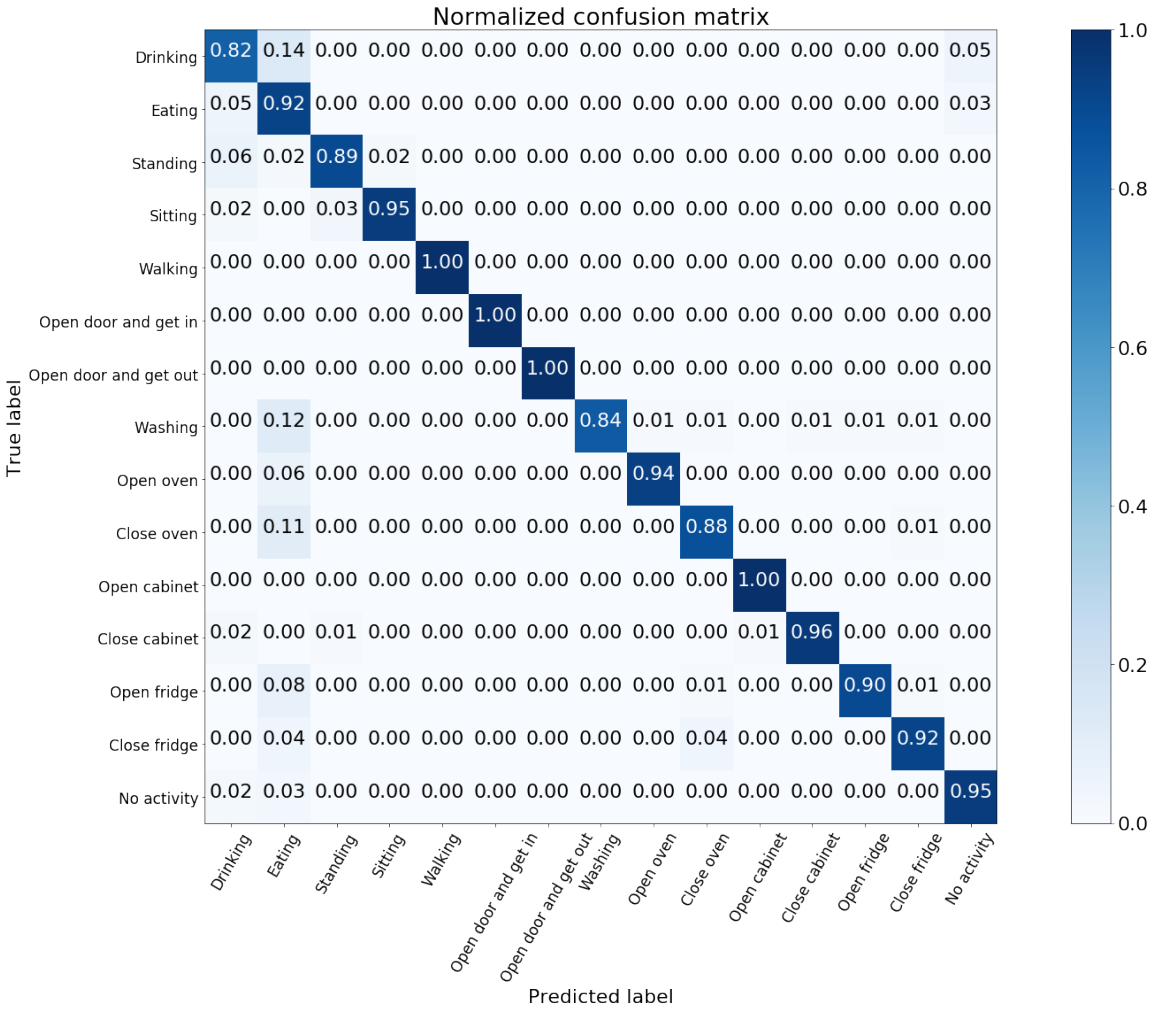


Figure 4.8: Normalized confusion matrix

4.4.2 Comparison with SVM

The second evaluation was a comparison between the DCNN model and an SVM model. SVM [146] is a popular traditional machine learning algorithm, which has been widely used in image classification. PCA was used [147] to extract features as the input of SVM. As shown in Table 4-B, the DCNN outperformed SVM+PCA by around two percentages in all three metrics, which were 92.81% in overall accuracy (OA), 93.14% in Recall, and 93.83% in F1.

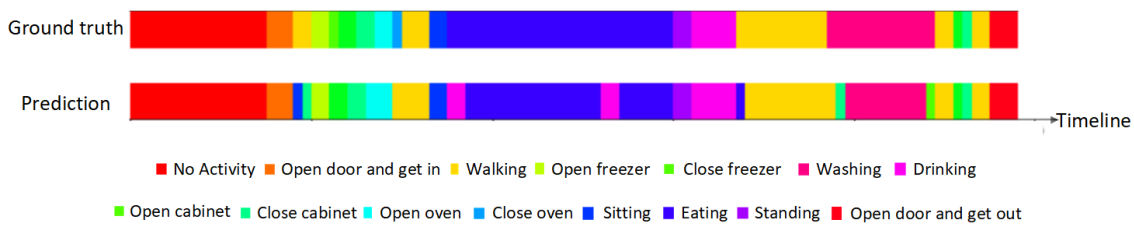


Figure 4.9: A timeline visualization of real-time human activity recognition results

Table 4-B: Comparison of the DCNN and the SVM+PCA

	OA	Recall	F1
DCNN	92.81%	93.14%	93.83%
SVM+PCA	90.9%	91.56%	91.71%

4.4.3 Real-time human activity recognition

The third evaluation was by developing a system combining the radar sensor network and the DCNN model. With this system, a three-minute real-time sequence of human activity was captured in the kitchen scenario. Also at the same time, a video camera recorded the same sequence for groundtruth comparison. In this period, all fifteen targeted activities were performed. Figure 4.9 summarized the real-time recognition results. The first row is the groundtruth and the second is the automatic recognition of the system. Different colors represent different activities. Except by 'Close oven', all other activities were recognized. 'Eating' took the longest time and 'Walking' was the most frequently performed activity. By comparing the groundtruth and the prediction, it was found that human activities were successfully recognized in more than 89% of the time, where the prediction exactly matched the groundtruth. It is possible to increase the recognition accuracy by using wider sliding windows. However, a wider sliding window also means more latency in the prediction. It is a tradeoff between the accuracy and the latency.

4.5 Analysis

The experiment and the result show that the integration of the radar sensor network and the DCNN achieved a good result in indoor human activity recognition. By using a very short sliding window for signal segmentation, it is able to achieve nearly real-time human activity recognition. It is possible to achieve a higher accuracy with better radars because the low-power capability of the radar sensor network built in this thesis may compromise some other characteristics, such as detection distance, sampling frequency, etc. A better design of the DCNN probably can be further explored for improving the accuracy of human activity classification.

Some activities investigated in this chapter are context-dependent, such as 'open door and get in', 'open door and get out', 'open/close oven', 'open/close cabinet', 'open/close fridge'. When a person is performing these activities, the frequency shifts are not only generated by his/her movements but also generated by the doors of the room, cabinet, fridge, and oven. When getting to a new scenario, the positions of doors and facilities are different. They will generate different frequency shifts and affect the results of activity recognition. So it necessary to collect the samples of these activities again in the new scenario and re-train the classifiers. However, for the activities including 'Drinking', 'Eating', 'Standing', 'Sitting', 'Walking' and 'Washing', it is unnecessary to collect their samples again. Because these activities are context-independent. The frequency shifts generated by these activities are all from the movements of the target person. The samples of these activities have already been collected from different positions, angles. So these activities don't need to be sampled again in a new scenario.

It is worthy to note that this work only investigated single-person activity in a kitchen. It is suitable for indoor scenario with one person. In many scenarios and at most of the time, multiple people can appear at the same time. However, the radars used in this thesis are not able to multi-person activity recognition. Because they are single-antenna radars that can not get the azimuth and the distance of targets, they are not able to separate

each target in the spatial dimension. For achieving multi-person activity recognition, a location-enabled technique is required. Radars with multiple antennas are possibly able to obtain the locations of targets and separate them. LiDAR also can be used to track multiple people. In next chapter, a 2D LiDAR is used to perform trajectory-based multi-person activity recognition.

4.6 Summary

This chapter presented the implementation of a radar sensor network whose good was to recognize 15 activities in a kitchen scenario. With the collected radar signals, STFT was used to generate spectrograms that contained micro-Doppler signatures of these human activities in the frequency domain. A DCNN was built to learn the micro-Doppler features automatically and classify the activities. The DCNN achieved 92.81% overall accuracy in the test, which exceeded the performance of SVM+PCA by around 2%. A system combing the radar network and the DCNN was implemented to achieve nearly real-time recognition. It successfully recognized human activities in more than 89% of the time. This chapter shows the great potential of a low-power low-cost radar-enable sensor network in a kitchen context-based human activity recognition, which is helpful in diet-controlling, patient monitoring and smart kitchen design. The proposed approach has the advantages of not requiring wearable sensors, it is power-saving and computationally efficient.

Chapter 5

Multi-person activity recognition using a 2D LiDAR

In many situations, when monitoring activities indoors, there are more than just one person in a room. In these cases, it is necessary to perform multiple people activity recognition simultaneously. Human activity is usually accompanied by location changes of humans. Consequently, it is required to track each person while he/she is moving when performing monitoring. LiDAR (Light Detection and Ranging) is a popular technique in localisation and surveying. It can achieve centimeter-level localisation that is able to separate multiple people spatially. It is also able to achieve real-time people tracking due to the fast response of the LiDAR. This chapter presents trajectory-based multiple people activity recognition in nearly real-time by using a 2D LiDAR to track and collect trajectories of each person.

5.1 Introduction

Location acquisition systems have been widely applied in many recent applications and they have achieved by various techniques, such as GPS, radar, LiDAR, etc. These systems generate a large amount of location data, and in the past years, they have been applied in navigation, route planning, and mapping. Recently, more and more studies attempt to infer semantic information from the trajectories formed by the locations of people, for example, for personalized recommendation, smarter home or more intelligent human-like robot interaction. Although there are many localisation techniques, none of those can be applied ubiquitously for any purpose. GPS hardly works indoors as its signals cannot penetrate walls. Radars may suffer interference from the ground surrounding and multi-path effects. Bluetooth and RFID (Radio-Frequency Identification) can perform indoor localisation, however, due to their short-range sensing capability, they generally require larger deployments that may become difficult to maintain. As a surveying and mapping technique, LiDAR has been increasingly used in self-driving vehicles and robots due to its high localisation accuracy, good real-time performance, and easy deployment. LiDAR also can be applied both indoors and outdoors. As 2D LiDARs are more affordable than 3D LiDARs, the work proposed in this chapter is to use a 2D LiDAR to perform multiple people activity recognition indoors.

A trajectory, e.g. a sequence of coordinates, can be collected by a LiDAR. A trajectory is a time series that contains both spatial and temporal information. RNN has been widely used to process time-sequential data due to the connections between its nodes form a directed graph along a temporal sequence. As a variant of RNN, LSTM further uses three gates to regulate the flow of information into and out of each node, which overcomes the exploding and vanishing gradient problems existing in conventional RNN models. LSTM has achieved state-of-art in many areas, such as natural language processing, weather forecasting, speech recognition, etc. In this chapter, an LSTM network and a temporal convolutional networks (TCN) is built to classify human trajectories into predefined activity and the performance of them are compared.

It is more challenging to perform human activity recognition using wireless sensing techniques than using wearable sensors. For wearable sensor-based HAR, each person wears sensor devices, the sensor data is collected from each specific person. However, wireless sensing techniques including LiDAR, Radar, and WiFi monitor a scenario that contains both humans and non-human objects such as walls, tables, animals, etc. It is necessary to differentiate humans from non-human objects before performing HAR. Another challenge is that multiple people can appear in the detection area of wireless sensors. For recognizing the activities of different people, it is necessary to track each person separately. LiDAR can achieve centimeter-level localisation accuracy, by using this characteristic of LiDARs, it is possible to separate multiple people in the spatial dimension. LiDARs also have a very fast response time that makes them suitable for real-time tracking. These reasons drove us to choose a 2D LiDAR to perform multi-person activity recognition.

5.2 Related work

Trajectory-based human activity recognition is gradually attracting the attention from industry and academia as human activity is usually accompanied by the location changes of people. A wide variety of sensors can be used to perform trajectory-based human activity recognition. Cameras are widely deployed for trajectory-based human activity recognition because videos and images are visual forms that can easily be perceived by our eyes. In [148], the authors proposed to detect abnormal behaviour through trajectory analysis and anomaly modeling upon a camera sensor network. Target trajectories are firstly reconstructed as symbol sequences. Then the sequences are fed into a Markov model to automatically analyse abnormal behaviour. Finally, the anomaly detection is decided through the majority voting of local results of individual camera nodes. In [149], the authors proposed to detect two types of motion anomalies based on people's trajectory. Firstly, the author segmented the surveillance scene as a Region Association Graph (RAG). The trajectory of individual motion activity is represented by the nodes and edges in the RAG. The authors further extracted statistical features from the

trajectories in order to train an SVM that used to detect anomalies. In [150], the authors presented a method based on hierarchical Dirichlet process hidden Markov models (HDP-HMM) to model human trajectories and recognize human activities. They applied this method to two real-world scenes, which include a shopping center and a university campus. For the shopping center, four activities including 'entering', 'leaving', 'passing', 'browsing' were investigated. For the university campus, seven activities including 'entering', 'leaving', 'crossing park up', 'crossing park down', 'passing through', 'walking along', and 'wandering' were considered. The overall accuracies achieved in these two scenarios were 96.76% and 96.62% respectively. In [151], the authors proposed an unsupervised method to perform human activity recognition by building a scene topology that defines a set of regions interacting with individuals. An activity is described by a trajectory cut over the topology. An activity recognition model is built by measuring the activity similarity. GPS as a popular positioning and navigation technique also has been used in trajectory-based activity recognition. In [152], the authors proposed an automatic activity detection method using POIs' (Point of Interests) spatial-temporal attractiveness to identify activity-locations as well as durations from raw GPS trajectory. In [153], the authors proposed a cost-sensitive approach for activity recognition from GPS-logs by measuring the importance of each activity from spatial and temporal perspectives. The cost function is combined with other activity recognition models to perform activity recognition. In the experiment, nine activities including 'Work', 'Stay home', 'Visit Friends', 'Shopping', 'Dining outside', 'Visit Scenic', 'Pick up', 'Entertainment', and 'Other' were investigated. They achieved near 70% in macro average accuracy. In [154], the authors extracted and labelled individual activities and significant places from traces of GPS data with the consideration of high-level context. They demonstrated a model of Relational Markov Networks (RMNs) that trained from a group of persons, and then applied successfully to a different person, achieving more than 85% accuracy in determining low-level activities.

LiDAR (Light Detection and Ranging) is becoming increasingly popular. It is an important component in robots and self-driving cars, which enables them to perform

localisation, navigation, and mapping. It has been widely applied in object localisation [155–157], tracking [158–160], and mapping [161, 162]. Currently, there is little research in trajectory-based human activity recognition using LiDAR. In [**?**], the authors used a multi-beam LiDAR sensor to perform gait analysis and activity recognition. Firstly, they performed target re-identification through silhouette print extraction. Then they proposed a Gait Energy Image-Based approach for feature extraction. Finally, CNNs were built and trained for activity recognition. In the experiment, nearly 80% overall accuracy was achieved in the classification of five activities including 'bend', 'watch', 'phone', 'wave', and 'wave2'. In [163], the authors used a 3D LiDAR to perform driving behaviours recognition. They proposed the concept of behaviour distance, which is DTW distance of the tracks of individual target and the car. Based on the behaviour distance, they successfully recognized 5 different car-related behaviours (Drive-GetOff, GetOn, Trunk, GetOn-Drive, DropOff) with 90% overall accuracy. In [164], the authors used a 2D laser to investigate interactions between persons and detect anomalies in three different environments: open layout laboratory, corridor, and an outdoor courtyard. Tracks of persons were split into segments. Then a Markov model was built to represent the activity segments. Kullback-Leibler distance was calculated to measure the similarity of the segments of two people for interaction detection. In [165], the authors used a 2D laser to perform activity recognition of an individual in a kitchen scenario. The trajectories of the target were annotated with activity labels. A seq2seq model was built to perform the classification. However, they did not differentiate human targets and obstacles, and the system recognized activities of only one person, which is fine for single people occupancy, but not for a family home.

It is worth to make a comparison between LiDAR sensors and other techniques in trajectory-based activity recognition. Compared to cameras, LiDAR is unaffected by light, and it is privacy-preserved due to targets' faces are unexposed. It is easy to integrate a LiDAR sensor's local coordinate system into a global coordinate system [166]. However, trajectories obtained from videos are hardly converted to a global coordinate system and the trajectory-based analysis in videos is heavily dependent on the azimuth

and inclination of the camera. The data of a LiDAR is a sequence of coordinates that indicate the directions and distances to the LiDAR. However, the data of a camera is a sequence of images consisting of RGB values. If the position of a LiDAR in a global coordinate system is known, it is able to convert the LiDAR points to this global coordinate system because the relative positions of these points to the LiDAR are known. However, with images collected by a camera, it is impossible to know the relative positions of the objects in the images. Trajectories extracted from images only reflect their pixel positions in the images, not the real position to the camera. Compare to GPS and other wearable sensors, LiDAR is device-free, which does not require people to carry any transmitters and receivers. In addition, GPS signals are difficult to be received indoors. All these advantages drive us to use a LiDAR to perform multiple people activity recognition.

Machine learning has been widely applied in trajectory-based human activity recognition to classify different activities. DTW is one of the most used algorithms in human activity detection [167, 168]. It measures the similarity between two temporal sequences to find the optimum distance between time series [169]. As almost all trajectories are time-series, DTW has innate advantages in processing them. In [170], the authors built a DTW model to cluster similar trajectories in a video surveillance system for suspicious behavior detection. HMM is also favored in HAR [171, 172]. It is a Markov chain with both hidden and observable stochastic processes. In [173], the authors built a hierarchical Dirichlet process HMM upon GPS trajectories to detect anomalies of human motion in dynamic traffic control. A RNN as a deep learning method has widely applied in sequential data classification. It also has been used in trajectory prediction and classification. In [174], the authors proposed an LSTM model to perform trajectory and activity prediction from videos. Ma et al. [165] implemented a RNN to classify location-driven sequences of a user to recognize human activity in a kitchen. This chapter implements DTW, HMM, and LSTM, and compares their performances in LiDAR-based human activity recognition.

Table 5-A summarizes the sensors, algorithms, and activities investigated in related work.

Table 5-A: Sensors, algorithms used in trajectory-based activity recognition

Sensors	**Algorithm**	**Activity**	**Accuracy**	**Citation**
Camera	Markov model	Abnormal behaviour (running, following) detection through trajectory analysis	93.7%	[148]
	SVM	Two distinct types of anomalies	86%	[149]
	Hierarchical Dirichlet process hidden Markov model	Four activities in shopping center and seven activities in a campus	96.76% and 96.62 respectively	[150]
	Channel-Separated Convolutional Network	400 activities in a benchmark dataset of Kinetics-400	53.3%	[175]
GPS	Spatial temporal POIs' Attractiveness based activity identification	Six activities: "Dining", "Shopping", "Entertainment", "Public facilities" and "Others"	70%	[152]
	Hidden Markov Model	Working, staying home, visiting friends, shopping, dining outside, visiting scenic, picking up, entertainment	About 70%	[153]
	Relational Markov Networks	Walk, sleep, leisure, visit, pickup, on/off car, other	85%	[154]

Sensors	**Algorithm**	**Activity**	**Accuracy**	**Citation**
LiDAR	Convolutional neural networks	Five activities including 'bend', 'watch', 'phone', 'wave', and 'wave2	80%	[176]
	Dynamic time warping	Five car-related behaviours (Drive-GetOff, GetOn, Trunk, GetOn-Drive, DropOff	90%	[163]
	Recurrent Neural Network	Seventeen activities in a kitchen	88%	[165]

5.3 Methodology

The main steps of LiDAR-based human activity recognition are demonstrated in Figure 5.1. Firstly, a 2D LiDAR is used to collect a set of points, which are reflected by walls, tables, humans, etc. Then this point set is grouped into different clusters that represent different objects in a scene. Human recognition is achieved upon these point clusters by using a random forest to classify geometric features extracted from them. The Kalman Filter is applied to perform continuous tracking of multiple people. Trajectories obtained from the tracking are further segmented and tagged with predefined activity labels. Trajectory augmentation is implemented to enrich the samples and overcome the problem of unbalanced classes. Finally, an LSTM network and a TCN were built to perform trajectory classification. After training, both two networks are able to perform human activity recognition online. The details of each step are described in the following subsections.

5.3.1 Point clustering

The raw data of a LiDAR is a sequence of polar coordinates and each coordinate is noted as (r, θ) where r is the distance to LiDAR and θ is the angle. For facilitating subsequent processing, these coordinates are transformed from the polar coordinate system to a

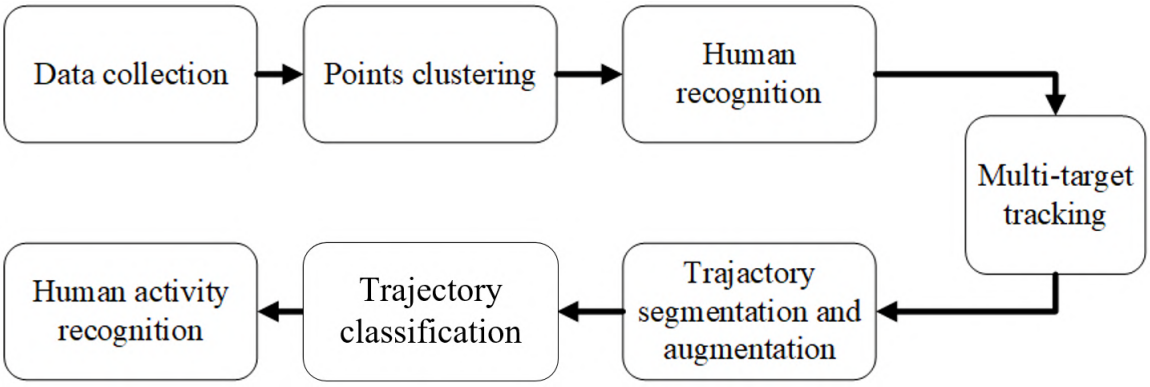


Figure 5.1: Workflow of LiDAR-based human activity recognition

plane coordinate system. Then the data at each timestep can be described as a sequence $S = \left\{p_1, p_2, ..., p_n\right\}$, where n is the total number of points.

It requires to perform clusters to group the points in S into different clusters for differentiating different objects. The DBSCAN is applied to perform clustering. DBSCAN is a density-based spatial clustering algorithm based on a threshold for the number of neighbors, $minPts$, within the radius ε [177]. A point with a neighbor count greater than or equal to $minPts$ within ε, is identified as a core point. A border point has the number of neighbors that less than $minPts$ but it belongs to the ε-neighborhood of other points. If a point is neither a core nor a border point, then it is called a noise point. For understanding the DBSCAN, three terms are defined:

Direct density reachable: a point A is directly density reachable from another point B if A is in the ε-the neighborhood of B and B is a core point.

Density reachable: a point A is density reachable from B if there are a set of core points leading from B to A.

Density connected: two points A and B are density connected if there are a core point C, such that both A and B are density reachable from C.

The DBSCAN can be abstracted into the following steps:

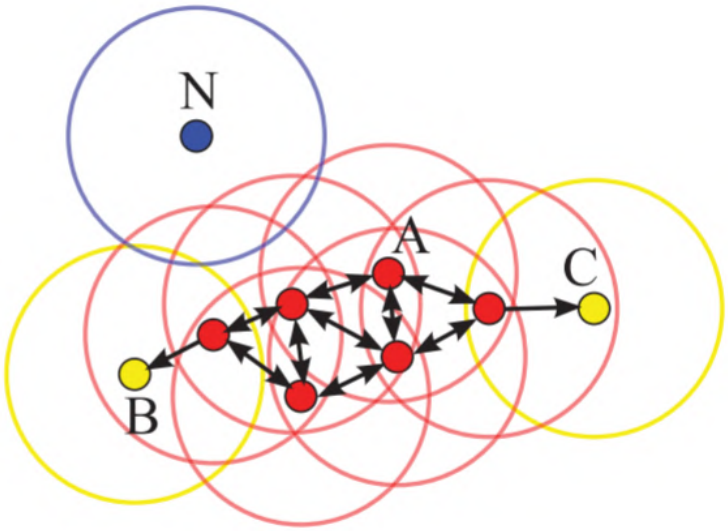


Figure 5.2: Illustration of the DBSCAN cluster algorithm

1. Find all neighbor points within ε of every point, and each point with more than $minPts$ neighbors within ε are marked as a core point.

2. For each core point if it is not already assigned to a cluster, create a new cluster. Find recursively all its density connected points and assign them to the same cluster as the core point.

3. Iterate through the remaining unvisited points in the dataset. Those points that do not belong to any cluster are marked as noise.

Figure 5.2 illustrates the concepts of DBSCAN. The $minPt$ is 4, and the ε is indicated by the size of the circles. N is a noise point, A is a core point, and points B and C are border points. Arrows indicate direct density reachability. Points B and C are density connected, because both are density reachable from A. N is not density reachable, and thus considered to be a noise point.

Compared to other clustering algorithms, DBSCAN does not need to specify the number of clusters and it is able to discover arbitrarily shaped clusters and robust to noise [178].

5.3.2 Human recognition

The clusters obtained after point clustering represent many different objects. It is necessary to recognize which cluster is human for getting its trajectory. The clusters show the differences in geometric characteristics between human and non-human. Geometric differences can be shown in linearity, circularity, angularity, etc. As Angus et al. did in [160], fifteen geometric features were extracted (shown in Figure 5.3) from each cluster. Base on these features, a random forest was built to classify human and non-human.

Random forest is an ensemble learning method for classification, regression by constructing a collection of decision trees. Decision tree is built by finding the the feature and decision threshold which splits the data as much as possible into its distinct classes. The splitting is repeated on each derived subset of data in a recursive manner until all data are fully split into distinct classes or a termination criterion is met. Algorithms for constructing decision trees usually work top-down, by choosing a threshold at each iteration that splits a subset of data. Several possible metrics are used to measure the optimality of the decision tree split. Gini impurity is the default metric in decision tree. It is a measure of how often a randomly chosen element from the set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset. To compute Gini impurity for a dataset with J classes, suppose $i \in \{1, 2, ..., J\}$, and let p_i be the fraction of items labeled with class i in the set.

$$\begin{aligned} I_G(p) = \sum_{i=1}^{J} p_i \sum_{k \neq i} p_k &= \sum_{i=1}^{J} p_i(1 - p_i) = \sum_{i=1}^{J} (p_i - {p_i}^2) \\ &= \sum_{i=1}^{J} p_i - \sum_{i=1}^{J} {p_i}^2 = 1 - \sum_{i=1}^{J} {p_i}^2. \end{aligned} \tag{5.1}$$

A random forest is trained by constructing a set of decision trees with the number of trees specified by a hyperparameter N. Thus, the random forest consists of N trees. For a new dataset, each item of the dataset is input to each of the N trees. The random forest chooses a class having the most out of N votes.

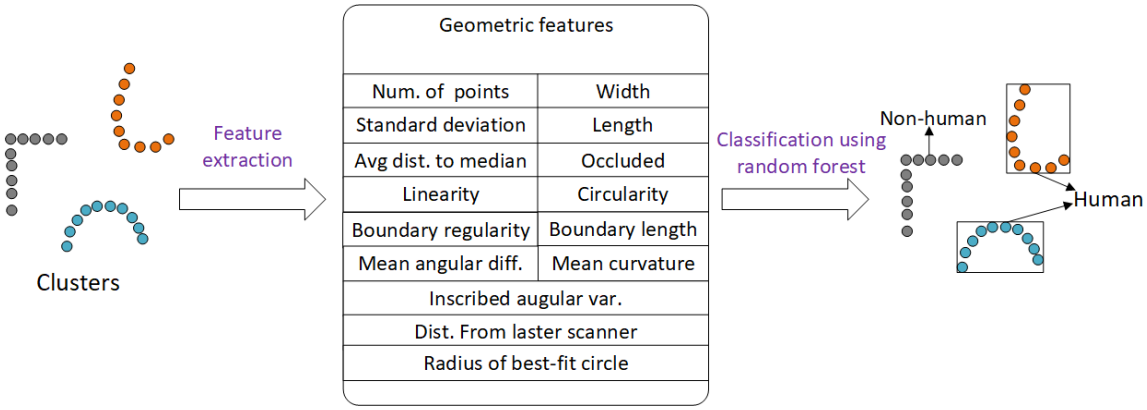


Figure 5.3: Human recognition

A random forest built with 25 decisions trees is used to classify human and non-human objects in this thesis. Positive (human) samples are collected from a clear and open space and negative (non-human) samples are collected from various indoor spaces without people. The total number of samples is 12000 that consists of 6000 positive samples and 6000 negative samples. The overall accuracy achieved in classification is 96%. The 2D geometric features may not contain enough information to fully discriminate humans from other objects, which probably compromises the results of human recognition. If using a 3D LiDAR or fusing with other sensors, the accuracy of human recognition is expected to be further improved.

One benefit of random forest is that it can measure the confidence level in the classification by aggregating the classification output of each individual tree. Thus, rather than just using the predicted labels, confidence level generated from random forest is considered to eliminate the noise and increase reliability.

5.3.3 Multiple people tracking

In order to obtain the trajectories of human clusters, it is necessary to perform continuous target tracking. Kalman filter is one of the most popular tracking algorithms due to its efficiency and accuracy. It is an optimal recursive data processing, which uses a series of measurements observed over time to produce an estimate of the desired variables and

finds the optimal state with the smallest possible variance error. It contains two steps: prediction and correction. In the prediction step, the state is predicted with the dynamic model, while in the correction step, the state is corrected with the observation model.

The process and measurement equations for the Kalman filter are given as follows:

$$\begin{aligned} x_k &= Ax_{k-1} + Bu_k + w_{k-1}, \\ z_k &= Hx_k + v_k, \end{aligned} \tag{5.2}$$

where k is the discrete time, x_k is the state vector, z_k is the observation vector, A and H are the transition matrix and observation matrix respectively. B_k is the control-input model which is applied to the control vector u_k. w_{k-1} and v_k are Gaussian random variable with zero mean, so their probability distributions are $p(w) \sim N(0, Q), p(v) \sim N(0, R)$ where the covariance matrix Q and R are referred to as transition noise covariance matrix and observation noise covariance matrix.

The prediction stage for the Kalman filter is as follows:

$$\begin{aligned} \hat{x}_k^- &= A\hat{x}_{k-1} + Bu_k, \\ P_k^- &= AP_{k-1}A^T + Q, \end{aligned} \tag{5.3}$$

a priori estimate of state $\hat{x}_k^-$ and covariance error P_k^- is obtained for the next time step k

The correction stage for the Kalman filter is as follows:

$$\begin{aligned} K_k &= P_k^- H^T (HP_k^- H^T + R)^{-1}, \\ \hat{x}_k &= \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-), \\ P_k &= (1 - K_k H)P_k^-, \end{aligned} \tag{5.4}$$

K_k is the Kalman gain which is computed by above equations. After that a posterior state estimate x_k^- and a posterior error estimate P_k is computed by the measurement z_k. The prediction and correction equations are calculated recursively with the previous

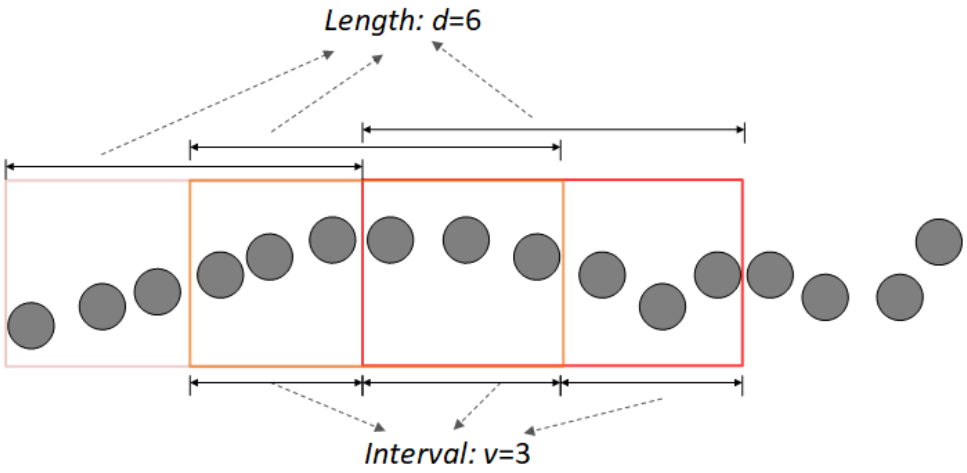


Figure 5.4: Trajectory segmentation

posterior estimates to predict new prior estimates.

In a tracking system, the state vector is $X = [x, y, v_x, v_y, a_x, a_y]^T$, where (x, y), (v_x, v_y) and (a_x, a_y) represent position, velocity and acceleration, respectively. The observation vector is $Z = (x', y')$.

5.3.4 Trajectory segmentation and augmentation

A trajectory can be the result of the location displacement when one or a sequence of activities is performed. For differentiating activities,, a trajectory is split into segments that each segment represents one activity. Several ways have been used to perform trajectory segmentation, such as distance-based segmentation, time-based segmentation, and the number of points-based segmentation. They split a trajectory into segments with the same distance, time, and the number of points, respectively. In this research, the number of points-based segmentation is implemented. A sliding window with the length of d is applied to perform segmentation at the interval of v. For instance, Figure 5.4 shows a sliding window with the length of 6 (location points) and the sliding interval of 3 (location points).

In our daily life, the frequency and duration of different activities are different. For example, the activity of 'having a meal' are mainly performed 3 times a day but the

activity like 'washing hands' can be performed many times, while the time consumed by 'having a meal' is much longer than 'washing hands'. This leads the problem of data sparsity that some activities have many samples but others have few. To overcome this problem, trajectory augmentation was implemented by using spatial transformation. In machine learning, data augmentation is used to create more training samples through different ways of processing or a combination of multiple processing upon the original samples. Data augmentation can increase the number of samples and boost the performance of deep learning. In this paper, spatial transformation is applied upon the raw trajectories to perform trajectory augmentation because the trajectories are spatial data. A spatial transformation is a mapping function that builds a spatial connection between all points in an image or geographic map and another set of points [179]. The methods implemented in this thesis are described as follows:

1. Translation: all points are translated to new positions by adding offsets T_x and T_y to x and y, respectively.

$$[x^{'} \quad y^{'} \quad 1] = [x \quad y \quad 1] \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ T_x & T_y & 1 \end{bmatrix} \tag{5.5}$$

2. Rotation: all points in the 2D plane are rotated around the origin through the counterclockwise angle θ.

$$[x^{'} \quad y^{'} \quad 1] = [x \quad y \quad 1] \cdot \begin{bmatrix} cos\theta & sin\theta & 0 \\ -sin\theta & cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.6}$$

3. Scale: all points are scaled by applying the scale factors S_x and S_y to the x and y

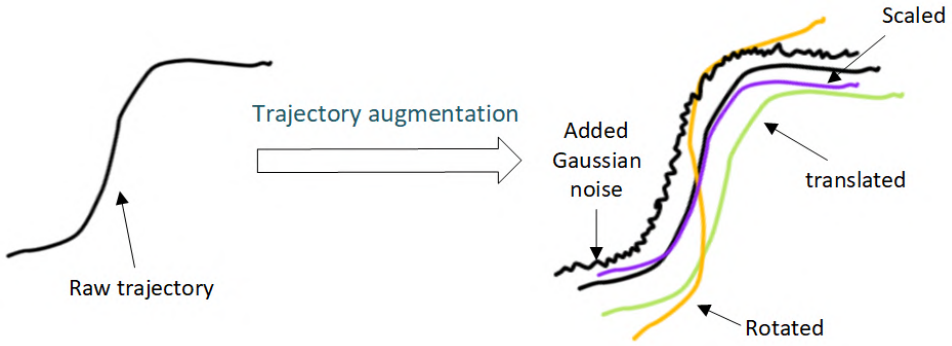


Figure 5.5: Trajectory augmentation

coordinates,respectively.

$$[x^{'} \quad y^{'} \quad 1] = [x \quad y \quad 1] \cdot \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_v & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.7}$$

Except for spatial transformation, Gaussian noise is also added to the raw trajectories. The process of trajectory augmentation is shown in Figure 5.5. It is worthy to note that it requires to control the range of spatial transformation to avoid to change the semantic information of trajectories.

5.4 Experimental setup

This section introduces the LiDAR and human activity recognition experiments.

The LiDAR used here is UST-10LX as shown in Figure 5.6(a). Its error in localisation accuracy is within 40mm. Figure 5.6(b) some characteristics of UST-10LX. It has a maximum detection distance of 30m and scan angle of 270°. Its angular resolution is 0.25°. One scan of it has 1081 measurement steps. More information about UST-10LX can be found in Table 5.4.

The experiments were made in the same kitchen as described in Chapter 4. An UST-10LX LiDAR was put in the height of 1.2m and close to the up-left corner of the

Figure 5.6: (a) UST-10LX LiDAR, (b) Laser scanning image

Table 5-B: Specifications of UST-10LX LiDAR [9]

Model	UST-10LX
Supply voltage	12VDC/24VDC (Operation range 10 to 30V ripple within 10%)
Supply current	150mA or less (When using DC24V) (during start up 450mA is necessary.)
Light source	Laser semiconductor (905nm) Laser class 1 (IEC60825-1:2007)
Detection range	0.06m to 10m (white Kent sheet) 0.06m to 4m (diffuse reflectance 10%) Max. detection distance : 30m
Accuracy	40mm
Repeated accuracy	$<$ 30mm
Scan angle	270°
Scan speed	25ms (Motor speed 2400rpm)
Angular resolution	0.25°
Start up time	Within 10 sec (start up time differs if malfunction is detected during start up)
Input	IP reset input, photo-coupler input (current 4mA at ON)
Output	Synchronous Output, photo coupler open collector output 30VDC 50mA MAX
Ambient temperature and humidity	−10°C to 50°C, below 85%RH (without dew, frost)
Storage temperature and humidity	−30°C to 70°C, below 85%RH (without dew, frost)
Weight	130g (Excluding cable)
Material	Front case: Polycarbonate, Rear case: Aluminum
Dimensions ($W \times D \times H$)	$50 \times 50 \times 70$mm (sensor only)

kitchen. The sampling frequency for data collection is 10 Hz (10 scans per second). The collected data was further processed using the methods described in Section 5.3

as shown in Figure 5.7. The raw LiDAR data (Figure 5.7(a)) consists of the points reflected from tables, walls, humans, etc. After clustering using DBSCAN, the points are clustered into different groups that shown with different colors in Figure 5.7(b). Then human recognition was achieved by classifying 15 geometric features extracted from these clusters using random forest with 25 trees. Finally, the Kalman filter was built for each human target in order to obtain their trajectories.

The kitchen was monitored using the LiDAR for a whole day and a density map of human trajectories was plotted to present the spatial distribution of their daily activities. As shown in Figure 5.8, human trajectories are mainly concentrated around the sink, the oven, the chair, and the door in the kitchen. These four areas can be seen as four stay areas that people usually stop for some amount of time. For example, people usually stay around the sink to wash their cooking utensils, stay by the oven for cooking, and sit on the chair for eating. Based on this, 15 activities were predefined, including: 'get in', 'get out', 'from sink to door', 'from chair to sink', 'from oven to chair', 'from sink to chair', 'washing', 'from door to sink', 'cooking', 'from oven to door', 'sitting', 'from oven to sink', 'from chair to oven', 'from door to oven', 'from sink to oven'. Most of the activities investigated are 'from ... to ...' can be taken as just walking. But walking cannot provide further more information about the activity that a person is engaging in and the object that the person is interacting with.

Trajectories of these activities (except 'get in' and 'get out') begin or end at these stay areas. Based on this, a tracker is defined for each person in the detection range. The properties of a tracker include 'id', 'current stay', 'last stay', 'stay time', and 'current trajectory' as shown in Figure 5.9. The trajectory collection process for each person is:

1. Initialize a tracker for each person in the detection area and set the tracker's 'current stay' and 'last stay' as none.

2. Update the tracker at each scanning frame of the LIDAR and judge whether the current location of the tracker is within a stay area. If it is within a certain stay

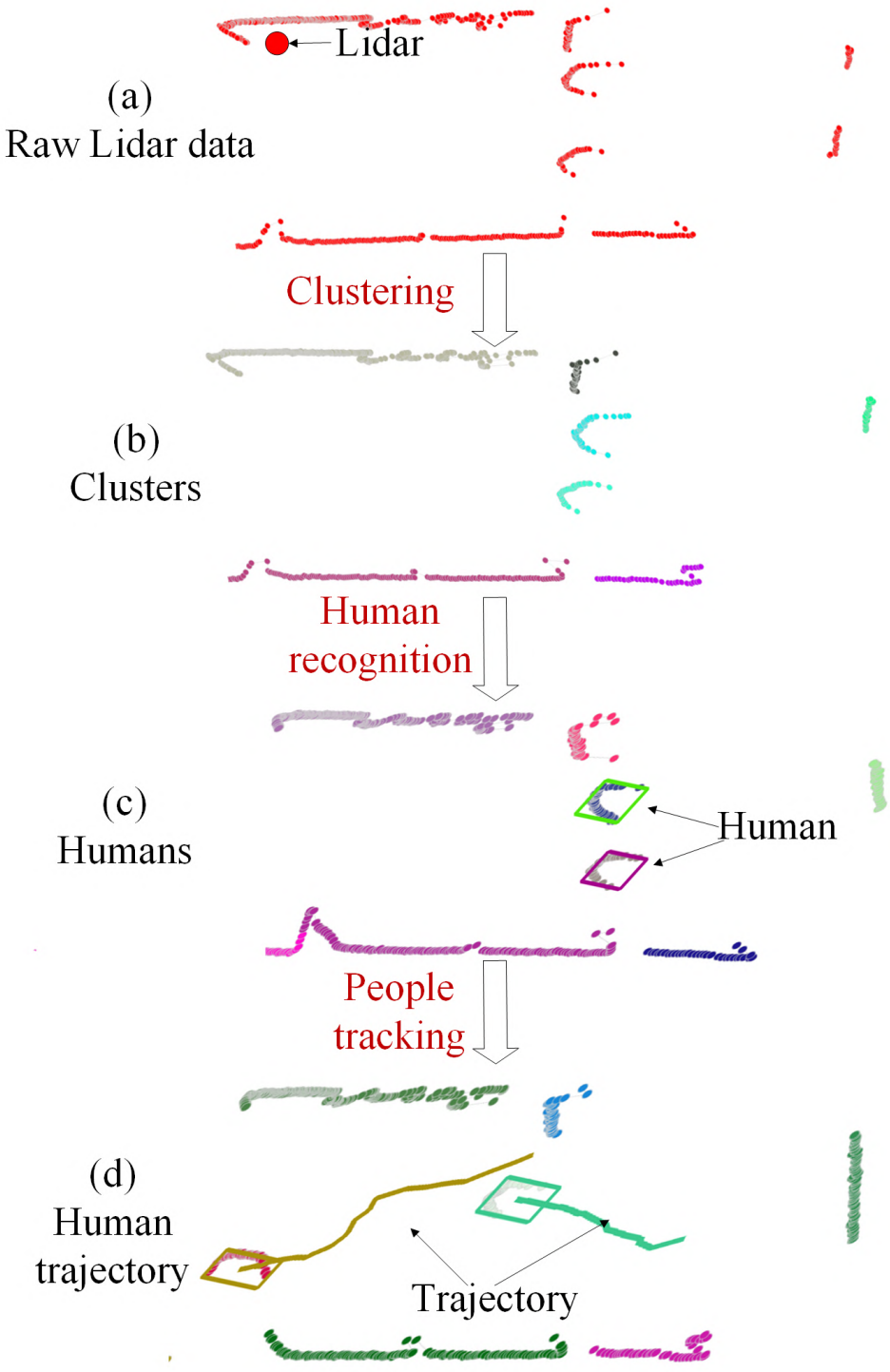


Figure 5.7: LiDAR data processing

area, then set its 'current stay' as the name of the stay area and initialize its 'current trajectory' as a point list to store the trajectory of the tracker.

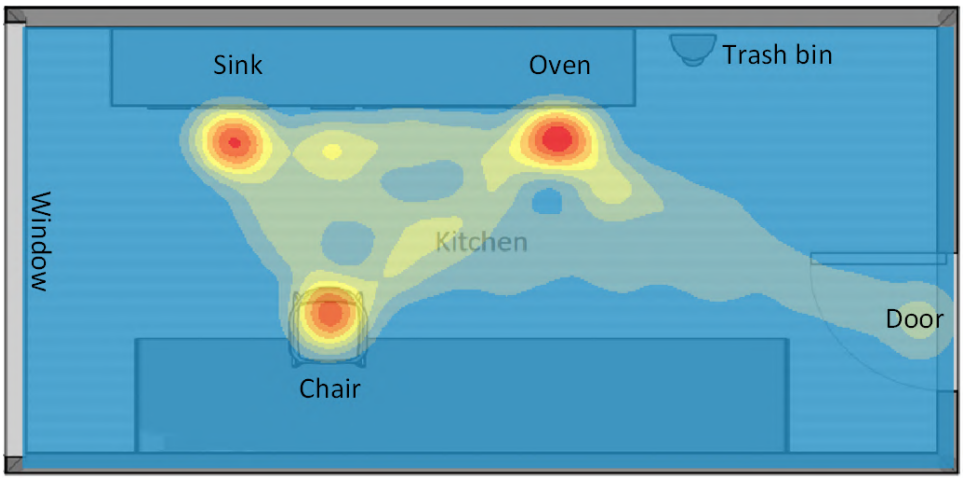


Figure 5.8: Density map of human trajectory in the kitchen

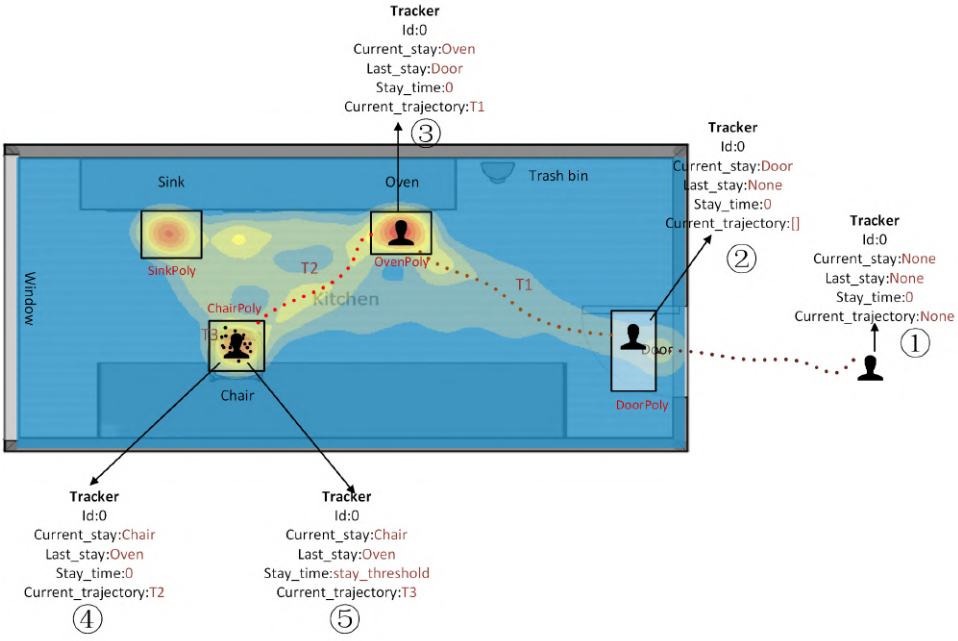


Figure 5.9: Process of trajectory collection

3. When the tracker reaches a new stay area, then set its 'last stay' as the value of 'current stay' and its 'current stay' as the name of this new stay area. Save its 'current trajectory' as a trajectory labelled with 'from 'last stay' to 'current stay'', then clear its 'current trajectory' to empty for storing the next trajectory of the tracker.

4. If the tracker does not get to a new stay area but stays in the same stay area for some amount of time and the stay time is beyond the threshold (80 frames used in this work), then save its 'current trajectory' and label it with 'cooking' if the stay area is 'OvenPoly', 'sitting' if the stay area is 'ChairPoly', and 'washing' if the stay area is 'SinkPoly' as shown in Figure 5.9. After this, set its 'current trajectory' as empty to store the next trajectory of the tracker.

5. Repeat step 3 and 4 until the tracker leaves the detection range of the LIDAR.

For the activities of 'get in' and 'get out', because they do not begin or end at any stay area. Trajectories of them were collected manually. It was started to record the trajectory of participants before they got in and stopped to record and save the trajectories and label them as 'get in' after the participants got in. For 'get out', the collection process is similar.

For ensuring the accuracy of the labels, because most activities were recorded and labelled automatically, the relationship between trajectories and the stay areas can make sure they were correctly recorded. For 'get in' and 'get out', they were recorded manually. Some trajectories of them may be mislabelled by the participants. But they can be filtered out by judging their directions. Because the directions of 'get in' and 'get out' are opposite. Besides, the collected trajectory can be visually checked.

The collected trajectories of these activities were segmented using a sliding window with the length of 25 and the interval of 10, so each trajectory has 25 points. As the sampling frequency is 10 Hz, the length of the sliding windows also can be measured by time and it is 2.5s. A longer trajectory may contain more than one activity and will lead to more latency, and a shorter trajectory does not have enough information related to human activity. A sliding window of 2.5s can avoid to contain more than one activity and has low latency in activity recognition in the kitchen. In the experiments, the number of people in the kitchen was varying and the maximum number of people is 4. The total time for monitoring is about 40 hours. Figure 5.10 demonstrates one trajectory sample

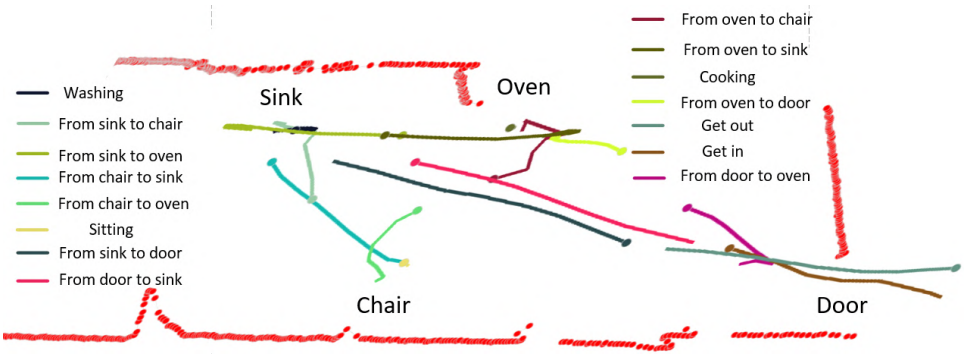


Figure 5.10: Trajectory samples of human activity

in the dataset for each activity with different colors. As it can be seen, the points of trajectories of 'washing', 'cooking', and 'sitting' tend to cluster together because people do not move much when they perform these activities.

As the shown in Figure 5.11, the quantity of each type of activity before augmentation is: 357 'from sink to door', 367 'from chair to sink', 279 'from sink to chair', 1000 'washing', 1000 'sitting', 350 'from door to oven', 353 'from oven to chair', 372 'from oven to door', 1000 'cooking', 267 'from door to sink', 270 'from chair to oven', 536 'from sink to oven', 656 'from oven to sink', 300 'get in', and 300 'get out'. Some activities (cooking, washing, sitting, etc.) have many samples while others (from sink to chair, from chair to oven, etc.) do not. We implemented trajectory augmentation to overcome this problem of unbalanced classes. After trajectory augmentation, we obtained 15000 total samples and the number of samples for each activity was 1000.

5.5 Trajectory classification using machine learning

This chapter implements four machine learning algorithms (DTW, HMM, LSTM, TCN) to perform trajectory classification, because they have been used in human activity recognition. SVM as a popular traditional machine learning method, is also applied in this work.

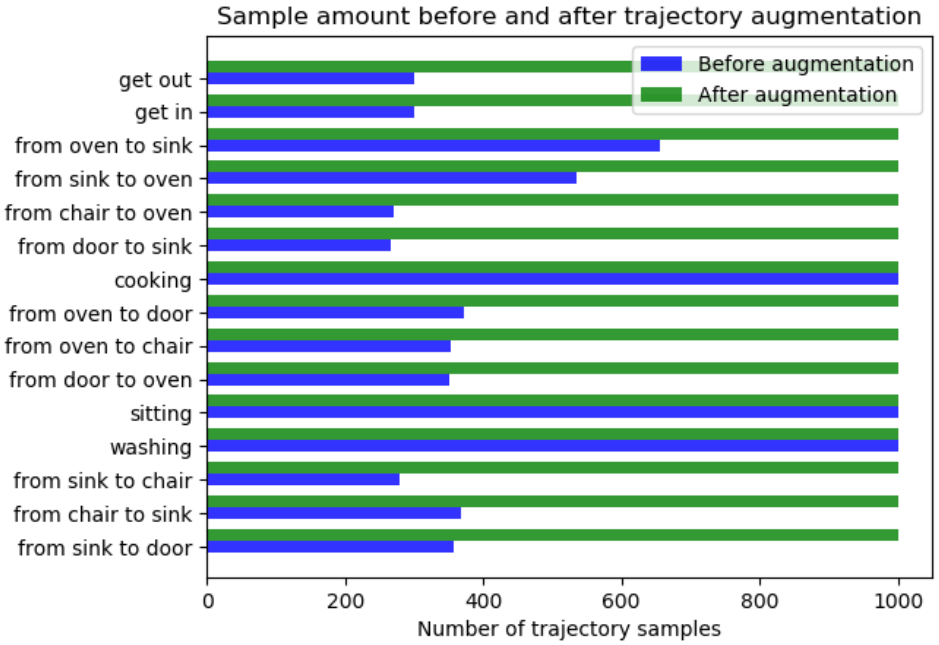


Figure 5.11: Sample amount before and after trajectory augmentation

5.5.1 Dynamic Time Warping (DTW)

Dynamic Time warping is usually used to measure similarity between two time series by minimizing the effects of shifting and distortion in time.

Given two time series $X = (x_1, x_2, \cdots, x_N), N \in \mathbb{N}$ and $Y = (y_1, y_2, \cdots, y_N), N \in \mathbb{M}$ represented by the sequences of values [180]. Firstly, DTW builds the distance matrix $C \in \mathbb{R}^{N \times M}$ representing all pairwise distances between X and Y. This distance matrix is called **local cost matrix** and used for the alignment of two sequences X and Y:

$$C_l \in \mathbb{R}^{N \times M} : c_{i,j} = \|x_i - y_j\|, i \in [1:N], j \in [1:M]. \tag{5.8}$$

Once the local cost matrix built, the algorithm finds alignment path which runs through the low-cost areas on the cost matrix. The alignment path built by DTW is a sequence of points $p = (p_1, p_2, \cdots, p_K)$ with $p_l = (p_i, p_j) \in [1:N] \times [1:M]$ for $l \in [1:K]$ which must satisfy to the following criteria:

1. Boundary condition: $p_1 = (1, 1)$ and $p_K = (N, M)$. The starting and ending points of the warping path must be the first and the last points of aligned sequences.

2. Monotonicity condition: $n_1 \leq n_2 \leq \cdots \leq n_K$ and $m_1 \leq m_2 \leq \cdots \leq m_K$. This condition preserves the time-ordering of points.

3. Step size condition: this criteria limits the warping path from long shifts in long while aligning sequences.

The cost function associated with a warping path computed with respect to the local cost matrix will be:

$$c_p(X, Y) = \sum_{L}^{l=1} c(x_{n_l}, y_{m_l}). \tag{5.9}$$

The warping path is called the optimal warping path that has a minimal cost associated with alignment. It is noted as P^*. DTW uses the Dynamic Programming-based algorithm to efficiently find the opimal warping path because it is computation-consuming to test every possible warping path between X and Y. The Dynamic Programming of DTW uses the DTW distance function [180]:

$$DTW(X, Y) = c_{p^*}(X, Y) = min\left\{c_p(X, Y), p \in P^{N \times M}\right\}, \tag{5.10}$$

where $P^{N \times M}$ is the set of all possible warping paths and builds the accumulated cost matric or global cost matrix D which defined as follows:

1. First row: $D(1, j) = \sum_{k=1}^{j} c(x_1, y_k), j \in [1, M]$.

2. First column: $D(i, 1) = \sum_{k=1}^{i} c(x_k, y_1), i \in [1, N]$.

3. All other elements: $D(i, j) = min\left\{D(i-1, j-1), D(i-1, j), D(i, j-1)\right\} + c(x_i, y_j), i \in [1, N], j \in [1, M]$.

where X and Y are the input time series and C is the local cost matrix representing all the pairwise distances between X and Y.

DTW usually works with kNN to perform classification. The implementation in this work combined kNN and DTW by replacing the Euclidean distance measurement with DTW distance measurement. The hyperparameter of kNN is the number of neighbours k and that of DTW is the *max warping window*. After tuning with cross validation, the optimal hyperparameters are $k : 1$, *max warping window*: 18.

5.5.2 Hidden Markov model (HMM)

HMM has been widely applied to deal with time series and can provide time-scale invariability in recognition. An HMM consists of a number of states each of which is assigned a probability of transition from one state to another state. As Markov models, states at any time step depend only on the sate at the preceding time. One symbol is yielded from one of the HMM states according to the probabilities assigned to the states [181]. HMM states are not directly observed symbols, and can be derived only through a sequence of observed symbols. An HMM is specified by the following notations:

$Q = \{q_1, q_2, \cdots, q_N\}$, a set of N states.
$A = \{a_{11}, \cdots, a_{ij}, \cdots, a_{NN}\}$, a transition probability matrix A, each a_{ij} representing the probability of moving from state i to state j, $s.t.$ $\sum_{N}^{j=1} a_{ij} = 1 \quad \forall i$.
$O = \{o_1, o_2, \cdots, o_T\}$, a sequence of T observations.
$B = b_i(o_t)$,a sequence of observation likelihoods, also called emission probabilities, each expressing the probability of an observation o_t being generated from a state i.
$\pi = \{\pi_1, \pi_2, \cdots, \pi_N\}$, an initial probability distribution over states. π_i is the probability that the Markov chain will start in state i. Some state j may have $\pi_j = 0$, meaning that they cannot be initial states. Also, $\sum_{i=1}^{n} \pi_i = 1$.

To recognize observed symbol sequences, a HMM is created for each category. For a classified of C categories, the model that best matches the observation from C HMMs $\lambda_i = \{A_i, B_i, \pi_i\}, i = 1, \cdots, C$ is choosen. For a sequence of uknown category, $P(\lambda_i \mid O)$

is calculated for each HMM λ_i and select λ_{c^*}, where

$$c^* = \underset{i}{argmax}(P(\lambda_i \mid O)). \tag{5.11}$$

Given the observation sequence $O = \{O_1, O_2, \cdots, O_T\}$ and the HMM λ_i, according to the Bayes rule, the problem is how to evaluate $P(\lambda_i \mid O)$, the probability that the sequence was generated by HMM λ_i. This probability is calculated by using the forward algorithm, which is formalized as follows:

$$a_t(i) \equiv P(O_1, O, \cdots, O_t, s_t = q_i \mid \lambda), \tag{5.12}$$

$a_t(i)$ is called the forward variable and can be calculated recursively as follows:

$$a_t(j) = \left\{ \sum_i a_{t-1}(i) a_{ij} \right\} b_j(O_t) a_1 = \pi_i b_i(O_i). \tag{5.13}$$

Then

$$P(O \mid \lambda) = \sum_{i \in S_P} a_T(i) \lambda_{c^*} c^* = \underset{i}{argmax}(P(\lambda_i \mid O)). \tag{5.14}$$

The likelihood of each HMM using the above equation and select the most likely HMM as the recognition result. The advantage of HMMs for time-sequential pattern recognition, which is robust to time scale variance and shift, results from this factor.

The hyperparameter of an HMM is the number of hidden components. We searched the number of hidden components from 0 to 20. It was found that 5 is the optimal number of hidden components.

5.5.3 Support vector machine (SVM)

A RBF SVM is also applied in trajectory classification for making a comparison with other models. Cross-validation was used to tune the hyperparameters. Given a hy-

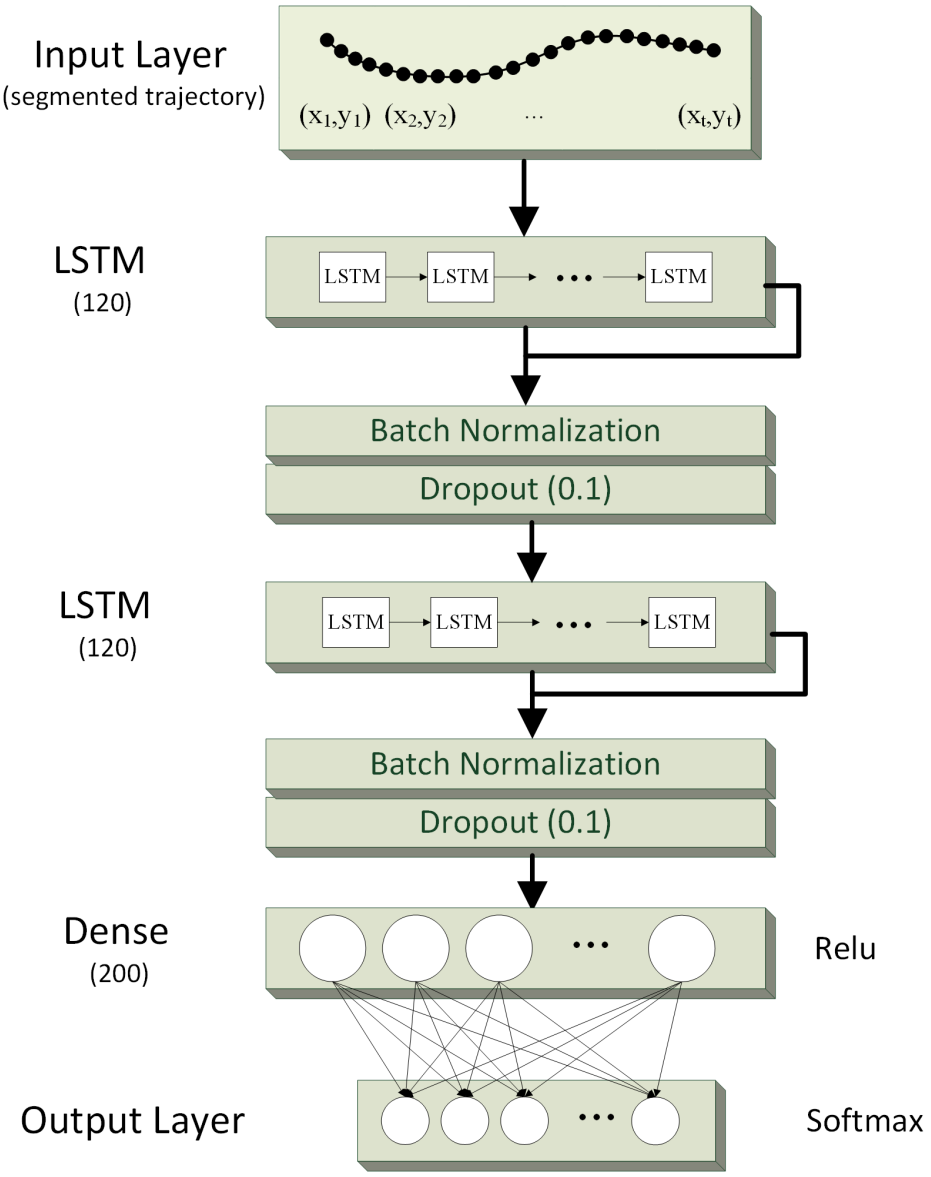


Figure 5.12: The proposed architecture of the LSTM network

perparameter 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 optimal parameters are $C = 4.8$, $gamma = 10^{-4.5}$.

5.5.4 LSTM

As shown in Figure 5.12, this thesis builds an LSTM network that consists of two LSTM layers and one fully-connected layer. Segmented trajectories are input into the network. As each point in a trajectory has two values (x, y), the input shape of the LSTM network is $(t, 2)$ where t is the length of a trajectory. Both two LSTM layers have 120 units and the fully-connected layer has 200 hidden units. The size of the output layer is determined by the number of activities that it is designed to recognize.

In training process, the LSTM network is trained to minimize an objective function in terms of the parameters of the network. For activity recognition, let C be the number of activities, the following cross-entropy loss function is often used:

$$E_y(y^{'}) = -\sum_{i=1}^{N} y_i \cdot log(y_i^{'}), \tag{5.15}$$

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 the sample i calculated using a *softmax* activation function:

$$y_j = exp(x_j)/(\sum_{c=1}^{C} exp(x_c)), \tag{5.16}$$

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

In training process, dropout and batch normalization have been applied after each LSTM layer to perform regularization that prevents overfitting. The dropout is initialized with the rate of 0.1. The optimizer used for the LSTM network is Adam whose learning rate was initialized as 0.0003.

5.5.5 Temporal Convolutional Network (TCN)

In this work, a TCN was built that consists of three residual blocks ('Block 1', 'Block 2', and 'Block 3') as shown in Figure 5.13. After hyperparameter tuning, the kernel sizes for these three blocks were set as 5, the dilation rates of them were 1, 2, and 4 respectively.

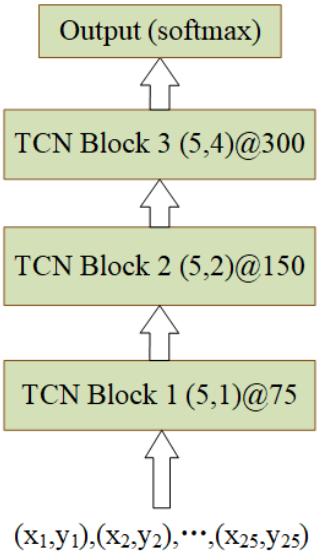


Figure 5.13: The proposed architectural based on TCN

'Block 1' has 75 filters, 'Block 2' has 150 filters, and 'Block 3' has 300 filters.

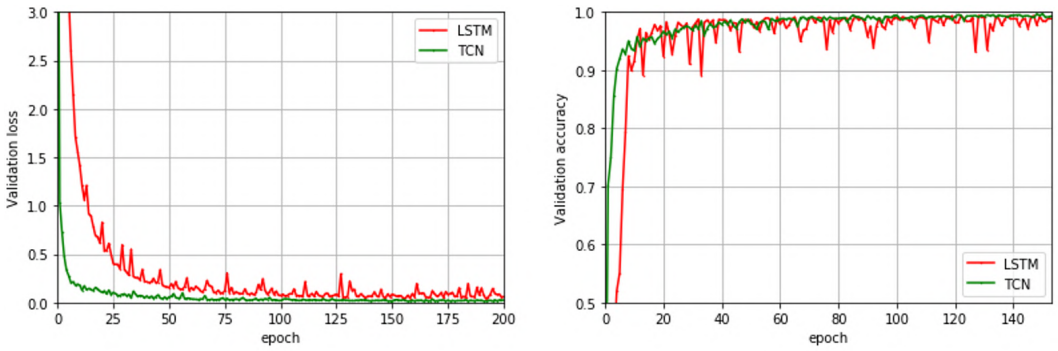
In training process, the dropout is initialized with the rate of 0.1. The optimizer used for the TCN network is Adam whose learning rate was initialized as 0.0003.

5.6 Evaluation

In this section, the proposed LSTM-based multiple people activity recognition is evaluated in three ways including evaluation on the test dataset, comparison with the baseline, and comparison with the related work.

5.6.1 Evaluation on the test dataset

The trajectory samples are grouped into two groups, 80% of the samples are used for training, and 20% for testing. In the training, a hold-out validation is performed with a 15% rate that randomly holds out 15% of the data for validation in training. The testing dataset was not exposed to the LSTM network and the TCN in training. The achieved overall accuracies by the LSTM network and the TCN in testing were 99.39% and 99.49% respectively. The TCN performed slightly better than the LSTM. More importantly, the



(a) Validation loss (b) Validation accuracy

Figure 5.14: Validation loss and accuracy of the proposed LSTM and TCN

convergence speed of the TCN is much faster than that of the LSTM. As shown in Figure 5.14, the lines of validation loss and accuracy of the TCN are smoother than those of the LSTM network. This is probably because TCN has a backpropagation path different from the temporal direction of the sequence that is beneficial to avoid the problem exploding/vanishing gradients [54].

As shown in the normalized confusion matrices of Figure 5.15 and Figure 5.16, which are generated by the LSTM and the TCN in test respectively. In both confusion matrices, most of the activities are correctly classified with 100% or 99%. In Figure 5.16, for 'from chair to sink', 1% samples have been misclassified into 'from chair to oven' and 1% samples have been misclassified into 'sitting'. It is because some trajectory samples from these classes begin from the chair in the kitchen and have some overlaps. The same situation also happened between 'from door to oven' and 'from door to sink'. 1% samples of 'from door to oven' have been misclassified into 'from door to sink'. Also, 1% samples of 'from sink to door' have been misclassified into 'from sink to chair'.

For validating the effect of trajectory augmentation, a comparison between the networks that trained on the samples after augmentation and those trained on the raw samples was made. As shown in Table 5-C, after applying the proposed trajectory augmentation, the performance of the LSTM network has been increased about 1.7% in OA, Recall, and F1; and the performance of the TCN has been increased about 1.5% in all

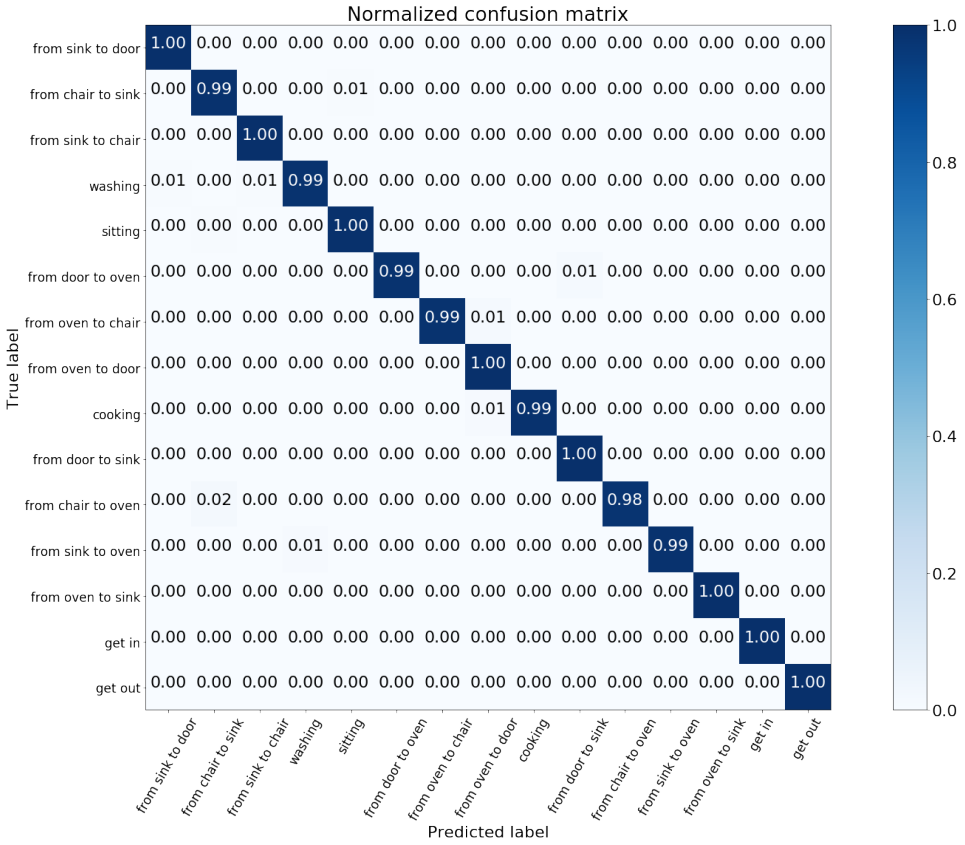


Figure 5.15: Normalized confusion matrix of the LSTM

three metrics.

Table 5-C: The effect of trajectory augmentation

	OA	Recall	F1
TCN with trajectory augmentation	99.49%	99.53%	99.51%
TCN without trajectory augmentation	97.96%	97.93%	97.96%
LSTM with trajectory augmentation	99.39%	99.41%	99.39%
LSTM without trajectory augmentation	97.68%	97.79%	97.65%

5.6.2 Comparison with the baseline

The DTW, HMM and SVM introduced in Section 5.5 are taken as the baseline to make a comparison with the proposed LSTM and TCN networks. As shown in Table 5-D, both the LSTM network and the TCN are superior to these three algorithms in all three

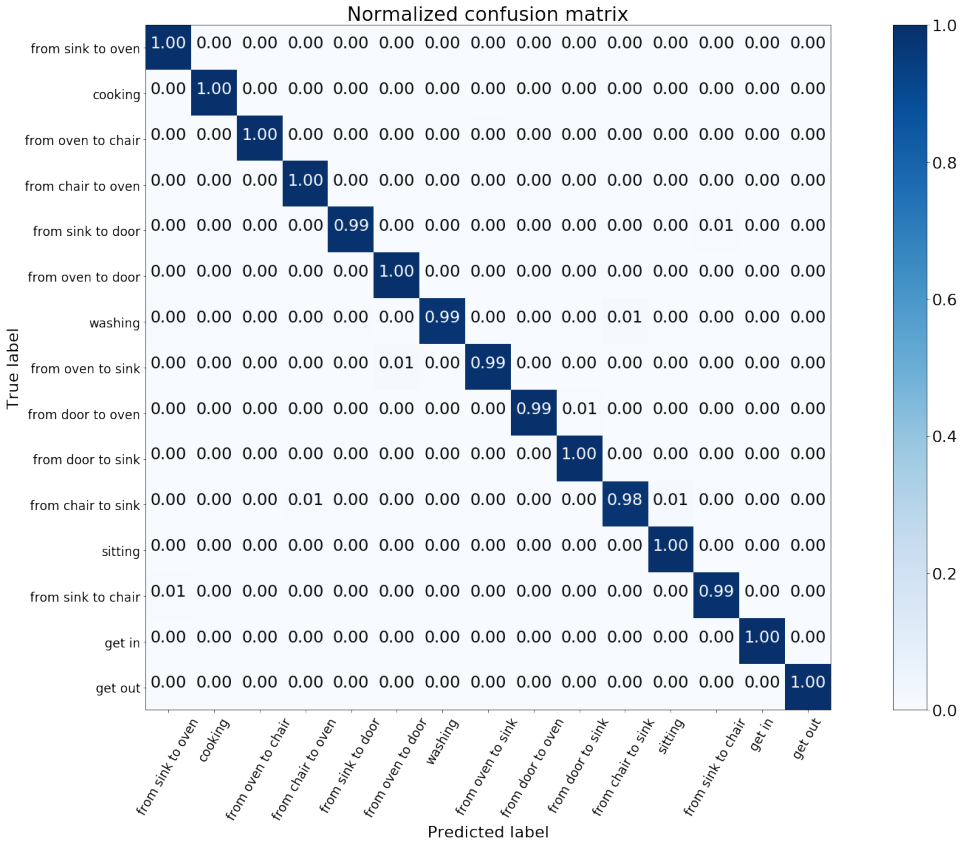


Figure 5.16: Normalized confusion matrix of the TCN

metrics, The TCN achieved the best result that is 99.49% in OA, 99.53% in Recall, and 99.51% in F1.

Table 5-D: Comparison of the classification models

	OA	Recall	F1
TCN	99.49%	99.53%	99.51%
LSTM	99.39%	99.41%	99.39%
SVM	95.88%	95.76%	96.69%
HMM	85%	86.7%	85.5%
DTW	90.9%	91.56%	91.71%

5.6.3 Comparison with the related work

In [165], the authors also used a 2D LIDAR to perform human activity recognition. They developed a seq2seq model to perform the classification of 17 activities and achieved 88% overall accuracy. It is impossible to make a completely direct comparison between the work in [165] and the approach proposed in this thesis. The layout of the kitchen and the LiDAR used in this work are different from those in [165]. It is unclear that the structure of the seq2seq model implemented in [165], the way of trajectory collection, and the quantity of their trajectory samples. However, the proposed approach in this thesis is able to perform multiple people tracking and it achieved very good accuracy in nearly real-time human activity recognition The work in [165] did not consider the interference from non-human objects and it can only be used to recognize the activities of a single person. The sep2sep model implemented in [165] probably is not a good method because it is difficult to know a trajectory is generated from how many activities. It is more suitable to represent each activity by using a small trajectory segmentation, which also can improve real-time capability in activity recognition.

5.6.4 Analysis

From above evaluations, it can be seen that deep learning has great performance in trajectory-based activity classifications in the comparison with traditional algorithms. Trajectory augmentation is an effective way to improve the classification accuracy when trajectories is few and imbalanced. The TCN performed better than the LSTM in trajectory classification. This is probably because the causal dilated convolutional layer is able to learn long-range temporal patterns. With a bigger dilation rate, a dilated convolutional layer can learn longer temporal patterns. The residual blocks enable the TCN have a deeper structure, which makes the deep layers in a TCN have a strong learning ability without losing information from the shallow layers. The structure of residual block was firstly applied in ResNet [182], which is a benchmark CNN model. However, the LSTM has great difficulty learning long-term dependencies because of the problem of vanishing gradients. If gradients vanish, there is little information propagated

back through back-propagation. As the length of the trajectory samples is only 25 (point), the capability of TCNs in learning long-range temporal patterns was probably not fully released. The high accuracy achieved in this work infers that indoor trajectory-based human activity recognition is applicable. Of course, it also has relationship with the layout of indoor scenarios. For a more complex indoor layout, the performance of activity recognition probably will decline.

5.7 Summary

In this chapter, a 2D Lidar was used to perform human activity recognition. The DBSCAN was applied to cluster the Lidar points and 15 geometric features were extracted from the clusters in order to perform human and non-human classification. The Kalman filter was used to perform multiple people tracking and obtain their trajectories. Then a sliding window was used to perform trajectory segmentation. For overcoming the problem of unbalanced classes, trajectory augmentation by using spatial transformation and adding Gaussian noise was implemented. Finally, an LSTM network and a TCN were built to perform human activity recognition upon the trajectory samples. The TCN achieved the best result of 99.49% overall accuracy, which is slightly superior to the LSTM network and much higher than the results achieved by SVM, HMM, and DTW. The proposed approach in this thesis outperforms the state-of-the-art approach.

Chapter 6

Analysis, conclusion, and future work

This thesis proposed two contactless sensing techniques for human activity recognition, which are a micro-Doppler signature based technique and a LiDAR-based technique. This chapter analyses the differences between these two techniques and compares their respective strengths and weaknesses. Then this chapter draws a set of definite conclusions from the novel research described in the previous chapters. Finally, this chapter discusses the possible avenues for further research in human activity recognition using micro-Doppler signatures and ranging techniques.

	Doppler radar	**Lidar**
Fundamental theory	Doppler effect	Propagation time
Data	Frequency spectrogram	Trajectory (location sequence)
Multi-person human recognition	Multi-antenna, Yes Single antenna, No	Yes
Application scenarios	Indoors and outdoors	Indoors and outdoors
Activity	Micro-activity	Macro-activity
Context-driven	No	Yes
Localisation ability	Multi-antenna, Yes Single antenna, No	Yes
Anti-interference ability	Weak	Strong
Contactless	Yes	Yes
Penetration ability	Yes (through wall and clothes)	No
Low-power	Depending on the operating frequency and output power	Depending on angle resolution, scan speed, etc.
Cost	Less expensive	Expensive
Range	From meters to kilometers	From meters to thousands of kilometers

Figure 6.1: Comparison between radar-based and LiDAR-based techniques in human activity recognition

6.1 Analysis

From the research in the previous chapters, it is known that both radar-based and LiDAR-based techniques are contactless and unintrusive to users' privacy. However, these two kinds of techniques are different in many aspects. Figure 6.1 compares radar-based techniques and LiDAR-based techniques from several aspects in detail. As it can be seen, Doppler radar achieves human activity recognition by using Doppler frequency shifts resulted by the Doppler effect; while LiDAR measures distances of objects based on the Time-of-Flight principle in order to obtain trajectories of users for human activity recognition. Radars with multiple antennas can achieve localisation by using range-Doppler information. However, radars with a single antenna cannot obtain rang-Doppler information. They are not able to perform localisation and separate multiple people in the space dimension. In Chapter 3, the group of people is taken as a whole target because Bumblebee radars are single-antenna and not able to discriminate each person in

a group. LiDAR can easily accomplish multiple people activity recognition by separating each person in the space dimension. Both radar and LiDAR can be used indoors and outdoors. However, the detection range of many radars is much larger than that of LiDAR sensors. So radar is more suitable for human activity recognition in large areas whose radius is hundreds or even thousands meters. Doppler radars are more suitable to recognise micro activities (motions) such as respiration, walking, jumping, etc. These motions will generate distinctive micro-Doppler signatures. LiDAR is more suitable to detect macro (context-dependent or interactive) activities such as 'from kitchen to live room', 'going upstairs', 'leaving office', etc. These activities are context-driven. For a different scenario, trajectory samples need to be reacquire by using LiDAR. LiDAR can only detect people in line-of-sight and can be affected by fog, dust, rain, etc., while radar can penetrate foliage, clothes, walls, etc. Both radar and LiDAR can be low-power consuming. For radars, power consumption depends on the operating frequency, output power, waveform, etc. For LiDAR, power consumption depends on angular resolution, scan rate, etc. The cost for both LiDAR and radar can be high and low, but on average, the cost of LiDAR is higer than the cost of radar. The range of radar can be from meters to thousands of kilometers. It depends on the operating frequency of a radar, the power of the transmitter, beamwidth (the more concentrated the beam, the greater is the detection range of the radar), receiver sensitivity, weather (clouds and rains), etc. The range of LiDAR can be from meters to kilometers. It depends on laser power, target surface diffuse reflectance, and the amount of ambient light coming from the target surface.

6.2 Conclusion

In this thesis, two contactless techniques (radar and LiDAR) for human activity recognition are investigated. For radar-based techniques, this thesis systematically investigated micro-Doppler based human activity recognition from fundamental concepts of micro-Doppler, radar signal storage and processing, feature extraction, and classification with machine learning. Radar-based techniques have been implemented both indoors and

outdoors. For LiDAR-based techniques, this thesis performed multiple people activity recognition indoors by implementing clustering, human recognition, tracking, and classification upon LiDAR data. From the research described in the previous chapters, it is possible to draw several conclusions.

Firstly, for achieving long-term human activity recognition, it is more suitable to use low-power techniques. This thesis built a low-power Doppler radar system by using BumbleBee radars which are Pulse-Doppler radars with a single antenna. Most radars used in the related work are more power-consuming and large in size. It prevents them to be implemented in areas without power supply. Admittedly, the range of BumbleBee radars are compromised due to the low-power consumption. According to the Radar Range Equation [29], if the operating frequency, the antenna gain, and the smallest power of a radar keep the same, the transmitting power of the radar transmitter needs to increase 16 times in order to double the detection range of the radar. The transmitting power is not the total power of the radar, it still also depends on the manufacturing level of radar. However, the performance of the radar system built in this thesis has not been compromised. The overall accuracies achieved in human activity recognition indoors and outdoors are not inferior to and even better than that in related work.

Secondly, it is necessary to consider the interference from the surroundings in micro-Doppler analysis of human activity. This thesis performed outdoor human activity recognition with the consideration of the noises generated by trees and foliage. Outdoor environments are relatively unstable compared to the indoors because of the unexpected movements generated by trees, small plants and leaves swaying due to airflow, and possible small animals (insects and small birds) moving in the detection range. Signal processing techniques such STFT and digital filtering are not able to completely remove all the noises. So it is necessary to collect the samples with the noises and input them to the classifiers or feature extraction techniques in order to make the classifiers and feature extraction techniques robust to noises and able to extract the important features for classification.

Thirdly, LiDAR-based human activity recognition is highly context-dependent. LiDAR-based techniques use the trajectories of users to perform human activity recognition. In Chapter 5, the investigated activities are mainly related to the stay points (chair, sink, and oven) in the kitchen. However, different rooms have different stay points and these stay points are located in different positions. This problem happens in all trajectory-based techniques. Whether radar-based techniques are context-dependent or not, it depends on the investigated activity. For example, activities such as walking, eating, sitting are context-independent, other activities such as 'open door and get out', 'open fridge' and 'open cabinet' are context-dependent. However, for LiDAR-based techniques, almost all the investigated activities are context-dependent.

Fourthly, deep learning shows its superiority in human activity recognition. Unlike traditional machine learning algorithms such as kNN, SVM, decision tree, etc., deep learning does not need to extract features before training. The hierarchical structure of deep learning enables automatic learning of the features from raw data without relying on feature extraction methods. And from the results achieved in the previous chapters, it can be seen that deep learning (CNN, LSTM, TCN) achieved better accuracies in human activity classification. However, it is worth to note that deep learning takes much longer time in training than traditional algorithms. Deep learning is computation intensive, it is possible to lead a higher latency in prediction because it takes longer time to perform prediction especially when the deep learning model becomes large (many layers and many feature maps).

Fifthly, radar-based techniques are affected by several factors such as the number of radars, the observation perspective of a radar system, and the size of the sliding window for segmentation, etc. More radars can provide more information from different observation perspectives. While it also means more data needs to be collected and processed. The observation perspective is related to the RCS of a target. RCS is a property of the target's reflectivity. RCS is greater when a radar observes a person from the front side. The size of sliding window for segmentation decides how much information

a sample contains. A large sliding window will lead a good accuracy in human activity classification. However, each activity has a duration. There is a limitation for the size of a sliding window. Also, a larger sliding window means a larger latency.

6.3 Future work

This section outlines four general directions that can be further explored in the same field.

6.3.1 Usage of multi-antenna radars and the Radar Data Cube technique

BumbleBee radars used in this research are single-antenna radars that are not able to obtain range-Doppler information. With a multi-antenna radar, it is possible to perform human activity recognition in both frequency and range domain. The Radar Data Cube is a convenient way to conceptually represent range-time-frequency processing of radar signals.

Assuming that preprocessing converts the RF signals received from multiple pulses across multiple array elements to complex-valued baseband samples, the radar data cube is constructed by arranging the complex-valued baseband samples in a three-dimensional array of size $K \times N \times L$. K defines the length of the first (fast-time) dimension. N defines the length of the second (spatial) dimension. L defines the length of the third (slow-time) dimension. Figure 6.2 shows the organization of the radar data cube. The fast time dimension is also referred to as the range dimension and the fast time sample intervals, when converted to distance using the signal propagation speed, are often referred to as range bins. The sampling interval between the L samples is the pulse repetition interval (PRI). Typical PRIs are much longer than the fast-time sampling interval. Because of the long sampling intervals, samples taken across multiple pulses are referred to as slow time. Processing data in the slow-time dimension allows users to estimate the Doppler spectrum at a given range bin. In this type of operation, the number of samples in the

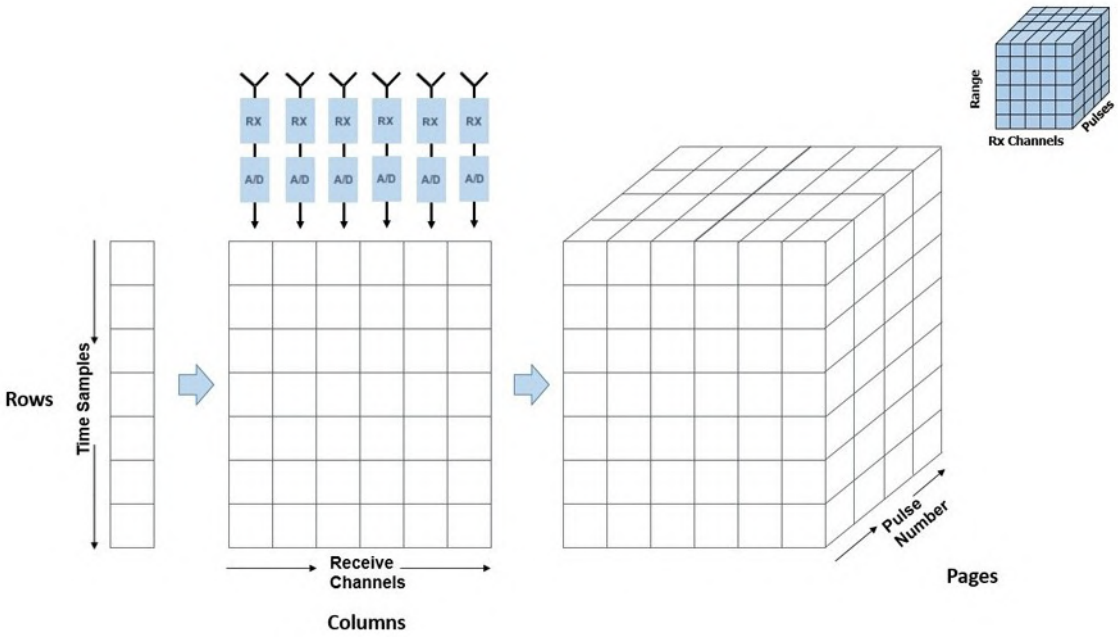


Figure 6.2: Radar Data Cube

third dimension of the data cube can change. The number of Doppler bins may not be the same as the number of pulses.

By using the Radar Data Cube technique, it is possible to achieve multiple people tracking with multi-antenna radars. In [183], the authors successfully separated and tracked two human targets by using the principles of Doppler processing and spatial beamforming on a continuous-wave Doppler radar with a multi-element receiver array. In [184], the authors tracked moving targets in the range domain, then used micro-Doppler signatures to perform target classification. The process is shown in 6.3. A threshold is applied to remove the noise and obstacles in range domain, then the clustering using DBSCAN is performed to detect the target, the Kalman-filter is further used to track the moving targets in range domain and micro-Doppler signatures of targets in frequency domain are extracted synchronously, finally target classification is performed using the extracted micro-Doppler signatures. So it is possible to perform multiple people activity recognition by using the Radar Data Cube technique that tracking multiple people in the range domain and recognizing human activity in the frequency domain. The combination

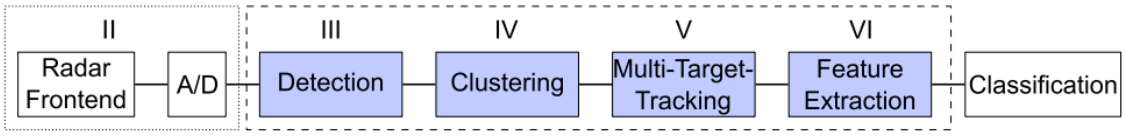


Figure 6.3: Process of target tracking and classification using Doppler Radar

of micro-Doppler signatures and range-Doppler signatures can improve the performance of human activity recognition. In [185], the author applied the Radar Date Cube to extract features from both the frequency domain and the range domain for activity recognition of an individual. From the work mentioned above, it is possible to achieve multiple people activity recognition by using a multi-antenna radar system and the Radar Data Cube technique.

6.3.2 Small-sample learning

In this thesis, the collection and labelling of samples took a large amount of time. Human activity recognition is a diverse field and many types of activities could be monitored. It is difficult to collect a large number of samples for each activity. In machine learning, small-sample learning is a class of machine learning tasks and techniques that also make use of a small number of labeled data. There are three possible ways to achieve small-sample learning. The first way is **data augmentation**. Data augmentation had been applied in LiDAR-based human activity recognition in this thesis. Data augmentation is more difficult to achieve for micro-Doppler signatures because many micro-Doppler signatures are almost unable to be observed by naked eyes. It is worth to explore data augmentation techniques for micro-Doppler signatures. The second way is **transfer learning**. Transfer learning is the improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned. For example, it is possible to train a model built for GPS trajectory classification upon a small amount of LiDAR trajectories or train a model built for micro-Doppler based target classification upon a small number of spectrograms of human activity. Compared to training a new model, transfer learning is able to obtain a higher accuracy with fewer samples. The third

way is **generative adversarial network (GAN)**. GAN composes of two networks, the generative network and the discriminative network. Given a training set, GAN learns to generate new data with the same statistics as the training set. The generative network generates new data while the discriminative network evaluates them. GAN can be used to generate fake spectrograms for human activity in order to increase the training samples.

6.3.3 A multi-LiDAR sensors network

LiDAR suffers from line-of-sight issues. For object detection and tracking, it requires to ensure there is no obstruction between the LiDAR and the targets. With a single LiDAR, the moving targets are easily blocked by obstacles or other moving targets that obstruct their LOS. For example, in this research, a person can block the LOS of another person frequently, this leads to a broken and uncontinuous trajectory. For overcoming this problem, it is necessary to use multiple LiDAR sensors to perform human recognition and tracking from multi-perspective. Even if one LiDAR sensor is obstructed, the other LiDAR sensors will still be able to detect the targets.

6.3.4 Data fusion with other sensors

Although radar and LiDAR show a great potential in human activity recognition, it does not mean they are good at recognize all human activities in all scenarios. LiDAR recognize human activity by trajectories. Some activities does not cause location changes or they have the same trajectory. For example, sitting and standing cannot be differentiated by their corresponding trajectories. In very noisy environments, it is difficult to extract micro-Doppler signatures from radar signals. In these situations, it is possible to improve human activity recognition by fusing several different sensors. In [186], the author achieved 100% accuracy in fall detection by combing Doppler radars and Kinect sensors. In [187], the authors applied magnetic and radar sensors to perform human activity recognition. The results showed that the fusion of magnetic and radar sensing greatly improved the performance of human activity recognition. It is worth to explore the fusion between radar or LiDAR with other sensors further in human activity

recognition.

6.4 Summary

This thesis presents two contactless techniques including radar-based and LiDAR-based techniques. By integrating with deep learning, both of them achieved very good results both indoors and outdoors. For the radar-based technique, a low-power low-cost radar system was proposed and applied in human activity recognition outdoors and indoors. Besides, micro-Doppler signatures measured by this radar system were used to perform human recognition, people counting, and coarse-grained localisation. For the LiDAR-based technique, a 2D LiDAR was used in human recognition and multi-person tracking in order to achieving trajectory-based human activity recognition.

This chapter makes a comparison between these two techniques in human activity recognition and summarizes the work that has been implemented in this thesis. It also provides four research directions in the same field.

References

[1] 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.

[2] Renishaw, "Optical encoders and LiDAR scanning," https://www.renishaw.com/en/optical-encoders-and-lidar-scanning--39244, accessed Jan 21, 2020.

[3] R. Berwick, "An idiot's guide to support vector machines (SVMs)," *Retrieved on October*, vol. 21, p. 2011, 2003.

[4] S. Raschka, *Python machine learning*. Packt Publishing Ltd, 2015.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[6] C. Olah, "Understanding LSTM networks," 2015.

[7] 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.

[8] B. Cagliyan, C. Karabacak, and S. Z. Gurbuz, "Human activity recognition using a low cost, COTS radar network," in *Radar Conference, 2014 IEEE*. IEEE, 2014, pp. 1223–1228.

[9] Hokuyo, "Scanning laser range finder Smart-URG mini UST-10LX (UUST003) specification," *Online available*, 2015.

[10] A. Rafii, S. B. Gokturk, C. Tomasi, and F. Sürücü, "Gesture recognition system using depth perceptive sensors," Mar. 26 2019, uS Patent App. 10/242,255.

[11] S. Wang, J. Song, J. Lien, I. Poupyrev, and O. Hilliges, "Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 2016, pp. 851–860.

[12] M. Jug, J. Perš, B. Dežman, and S. Kovačič, "Trajectory based assessment of coordinated human activity," in *International Conference on Computer Vision Systems*. Springer, 2003, pp. 534–543.

[13] N. Lu, Y. Wu, L. Feng, and J. Song, "Deep learning for fall detection: Three-dimensional CNN combined with LSTM on video kinematic data," *IEEE journal of biomedical and health informatics*, vol. 23, no. 1, pp. 314–323, 2018.

[14] D. Roy *et al.*, "Snatch theft detection in unconstrained surveillance videos using action attribute modelling," *Pattern Recognition Letters*, vol. 108, pp. 56–61, 2018.

[15] J. T. Lee, M. S. Ryoo, M. Riley, and J. Aggarwal, "Real-time illegal parking detection in outdoor environments using 1-D transformation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 7, pp. 1014–1024, 2009.

[16] H.-H. Hsu and C.-C. Chen, "Rfid-based human behavior modeling and anomaly detection for elderly care," *Mobile Information Systems*, vol. 6, no. 4, pp. 341–354, 2010.

[17] 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.

[18] 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.

[19] V. C. Chen and R. D. Lipps, "Timechen2014micro frequency signatures of micro-Doppler phenomenon for feature extraction," in *Wavelet Applications VII*, vol. 4056. International Society for Optics and Photonics, 2000, pp. 220–226.

[20] V. Chen, D. Tahmoush, and W. Miceli, "Micro-Doppler signatures-review chal-

lenges and perspectives," *Radar-Micro Doppler Signatures: Processing and Applications*, pp. 1–25, 2014.

[21] 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.

[22] T. Thayaparan, L. Stanković, and I. Djurović, "Micro-Doppler-based target detection and feature extraction in indoor and outdoor environments," *Journal of the Franklin Institute*, vol. 345, no. 6, pp. 700–722, 2008.

[23] R. O. Dubayah and J. B. Drake, "Lidar remote sensing for forestry," *Journal of Forestry*, vol. 98, no. 6, pp. 44–46, 2000.

[24] S. Salous, *Radio propagation measurement and channel modelling.* John Wiley & Sons, 2013.

[25] J. Erickson, "Radio-location and the air defence problem: The design and development of Soviet radar 1934-40," *Science Studies*, vol. 2, no. 3, pp. 241–263, 1972.

[26] R. M. Page, "The origin of radar," 1962.

[27] A. R. R. League, *The ARRL antenna book.* The League, 1949, no. 15.

[28] H. Rahman, *Fundamental Principles of Radar.* CRC Press, 2019.

[29] M. I. Skolnik, *RADAR systems.* McGraw-Hill, NY, 2001.

[30] E. Brookner, "Radar technology," *Dedham, Mass., Artech House, Inc., 1977. 444 p*, 1977.

[31] D. Tahmoush, "Review of micro-Doppler signatures," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1140–1146, 2015.

[32] S. S. Ram and H. Ling, "Simulation of human micro-Dopplers using computer animation data," in *Radar Conference, 2008. RADAR'08. IEEE.* IEEE, 2008, pp. 1–6.

[33] D. Tahmoush and J. Silvious, "Radar micro-Doppler for long range front-view gait recognition," in *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems.* IEEE, 2009, pp. 1–6.

[34] Z. Zhang, P. Pouliquen, A. Waxman, and A. G. Andreou, "Acoustic micro-Doppler gait signatures of humans and animals," in *2007 41st Annual Conference on In-*

formation Sciences and Systems. IEEE, 2007, pp. 627–630.

[35] R. M. Narayanan and M. Dawood, "Doppler estimation using a coherent ultrawideband random noise radar," *IEEE Transactions on Antennas and Propagation*, vol. 48, no. 6, pp. 868–878, 2000.

[36] G. J. Koch, B. W. Barnes, M. Petros, J. Y. Beyon, F. Amzajerdian, J. Yu, R. E. Davis, S. Ismail, S. Vay, M. J. Kavaya *et al.*, "Coherent differential absorption lidar measurements of CO_2," *Applied optics*, vol. 43, no. 26, pp. 5092–5099, 2004.

[37] L. Mei and M. Brydegaard, "Continuous-wave differential absorption Lidar," *Laser & Photonics Reviews*, vol. 9, no. 6, pp. 629–636, 2015.

[38] R. Harrap and M. Lato, "An overview of lidar: collection to application," *Norway*, 2010.

[39] P. Paul and T. George, "An effective approach for human activity recognition on smartphone," in *2015 Ieee International Conference on Engineering and Technology (Icetech)*. IEEE, 2015, pp. 1–3.

[40] 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.

[41] L. Fan, Z. Wang, and H. Wang, "Human activity recognition model based on decision tree," in *2013 International Conference on Advanced Cloud and Big Data*. IEEE, 2013, pp. 64–68.

[42] Z. Chen, C. Jiang, and L. Xie, "A novel ensemble ELM for human activity recognition using smartphone sensors," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2691–2699, 2018.

[43] C. Catal, S. Tufekci, E. Pirmit, and G. Kocabag, "On the use of ensemble of classifiers for accelerometer-based activity recognition," *Applied Soft Computing*, vol. 37, pp. 1018–1022, 2015.

[44] 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, 2015.

[45] F. Hernández, L. F. Suárez, J. Villamizar, and M. Altuve, "Human activity recog-

nition on smartphones using a bidirectional LSTM network," in *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*. IEEE, 2019, pp. 1–5.

[46] W. Karwowski and T. Ahram, *Intelligent Human Systems Integration: Proceedings of the 1st International Conference on Intelligent Human Systems Integration (IHSI 2018): Integrating People and Intelligent Systems, January 7-9, 2018, Dubai, United Arab Emirates*. Springer, 2017, vol. 722.

[47] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research," *Journal of pharmaceutical and biomedical analysis*, vol. 22, no. 5, pp. 717–727, 2000.

[48] M. Olazaran, "A sociological study of the official history of the perceptrons controversy," *Social Studies of Science*, vol. 26, no. 3, pp. 611–659, 1996.

[49] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE transactions on neural networks*, vol. 8, no. 1, pp. 98–113, 1997.

[50] T. Mikolov, M. Karafiát, L. Burget, J. Černockỳ, and S. Khudanpur, "Recurrent neural network based language model," in *Eleventh annual conference of the international speech communication association*, 2010.

[51] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[52] R. S. Bidgoli, "Higher order recurrent neural network for language modeling," 2016.

[53] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 156–165.

[54] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[55] L. Graesser, "Regularization for neural networks," 2016.

[56] J. W. Wei and K. Zou, "Eda: Easy data augmentation techniques for boosting

performance on text classification tasks," *arXiv preprint arXiv:1901.11196*, 2019.

[57] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[58] L. Prechelt, "Early stopping-but when?" in *Neural Networks: Tricks of the trade*. Springer, 1998, pp. 55–69.

[59] 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.

[60] 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.

[61] 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.

[62] 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.

[63] B. G. Mobasseri 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.

[64] 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.

[65] 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.

[66] M. Zenaldin 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.

[67] S. Björklund, H. Petersson, A. Nezirovic, M. B. Guldogan, and F. Gustafsson, "Millimeter-wave radar micro-Doppler signatures of human motion," in *Radar Symposium (IRS), 2011 Proceedings International*. IEEE, 2011, pp. 167–174.

[68] 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.

[69] 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.

[70] 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.

[71] 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.

[72] 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.

[73] S. Rahman and D. A. Robertson, "Radar micro-Doppler signatures of drones and birds at k-band and w-band," *Scientific reports*, vol. 8, no. 1, pp. 1–11, 2018.

[74] M. Jankiraman, *Design of multi-frequency CW radars*. SciTech Publishing, 2007, vol. 2.

[75] O. R. Fogle, "Human micro-range/micro-Doppler signature extraction, association, and statistical characterization for high-resolution radar," 2011.

[76] L. Vignaud, A. Ghaleb, J. Le Kernec, and J.-M. Nicolas, "Radar high resolution range & micro-Doppler analysis of human motions," in *2009 International Radar*

Conference" Surveillance for a Safer World"(RADAR 2009). IEEE, 2009, pp. 1–6.

[77] Z. Cammenga, G. Smith, and C. Baker, "Combined high range resolution and micro-Doppler analysis of human gait," in *2015 IEEE Radar Conference (RadarCon).* IEEE, 2015, pp. 1038–1043.

[78] O. R. Fogle and B. D. Rigling, "Micro-range/micro-Doppler decomposition of human radar signatures," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 4, pp. 3058–3072, 2012.

[79] D. Tahmoush and J. Silvious, "Radar measurement of human polarimetric micro-Doppler," *Journal of Electrical and Computer Engineering*, vol. 2013, p. 11, 2013.

[80] J. Park, J. T. Johnson, N. Majurec, M. Frankford, K. Stewart, G. E. Smith, and L. Westbrook, "Simulation and analysis of polarimetric radar signatures of human gaits," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 3, pp. 2164–2175, 2014.

[81] F. Fioranelli, M. Ritchie, and H. Griffiths, "Analysis of polarimetric multistatic human micro-Doppler classification of armed/unarmed personnel," in *2015 IEEE Radar Conference (RadarCon).* IEEE, 2015, pp. 0432–0437.

[82] ——, "Aspect angle dependence and multistatic data fusion for micro-Doppler classification of armed/unarmed personnel," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1231–1239, 2015.

[83] G. Garreau, C. M. Andreou, A. G. Andreou, J. Georgiou, S. Dura-Bernal, T. Wennekers, and S. Denham, "Gait-based person and gender recognition using micro-Doppler signatures," in *2011 IEEE Biomedical Circuits and Systems Conference (BioCAS).* IEEE, 2011, pp. 444–447.

[84] 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.

[85] T. Damarla, M. Bradley, A. Mehmood, and J. M. Sabatier, "Classification of animals and people ultrasonic signatures," *IEEE Sensors Journal*, vol. 13, no. 5, pp. 1464–1472, 2012.

[86] F. Fioranelli, M. Ritchie, and H. Griffiths, "Bistatic human micro-Doppler signatures for classification of indoor activities," in *Radar Conference (RadarConf), 2017 IEEE*. IEEE, 2017, pp. 0610–0615.

[87] P. Larranaga, B. Calvo, R. Santana, C. Bielza, J. Galdiano, I. Inza, J. A. Lozano, R. Armananzas, G. Santafé, A. Pérez *et al.*, "Machine learning in bioinformatics," *Briefings in bioinformatics*, vol. 7, no. 1, pp. 86–112, 2006.

[88] Y. Li, L. Du, and H. Liu, "Hierarchical classification of moving vehicles based on empirical mode decomposition of micro-Doppler signatures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 3001–3013, 2012.

[89] C. Cai, W. Liu, J. S. Fu, and Y. Lu, "A new approach for ground moving target indication in foliage environment," *Signal processing*, vol. 86, no. 1, pp. 84–97, 2006.

[90] X. Bai, M. Xing, F. Zhou, G. Lu, and Z. Bao, "Imaging of micromotion targets with rotating parts based on empirical-mode decomposition," *IEEE transactions on geoscience and remote sensing*, vol. 46, no. 11, pp. 3514–3523, 2008.

[91] F. Fioranelli, M. Ritchie, and H. Griffiths, "Classification of unarmed/armed personnel using the NetRAD multistatic radar for micro-Doppler and singular value decomposition features," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 9, pp. 1933–1937, 2015.

[92] J. De Wit, R. Harmanny, and P. Molchanov, "Radar micro-Doppler feature extraction using the singular value decomposition," in *2014 International Radar Conference*. IEEE, 2014, pp. 1–6.

[93] P. Molchanov, R. I. Harmanny, J. J. de Wit, K. Egiazarian, and J. Astola, "Classification of small uavs and birds by micro-Doppler signatures," *International Journal of Microwave and Wireless Technologies*, vol. 6, no. 3-4, pp. 435–444, 2014.

[94] F. Fioranelli, M. Ritchie, and H. Griffiths, "Performance analysis of centroid and SVD features for personnel recognition using multistatic micro-Doppler," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 5, pp. 725–729, 2016.

[95] J. Li, L. Liu, Z. Zeng, and F. Liu, "Advanced signal processing for vital sign extraction with applications in UWB radar detection of trapped victims in complex

environments," *IEEE journal of selected topics in applied earth observations and remote sensing*, vol. 7, no. 3, pp. 783–791, 2013.

[96] A. Miller, C. Clemente, A. Robinson, D. Greig, A. Kinghorn, and J. Soraghan, "Micro-Doppler based target classification using multi-feature integration," 2013.

[97] D. L. Hall, T. D. Ridder, and R. M. Narayanan, "Abnormal gait detection and classification using micro-Doppler radar signatures," in *Radar Sensor Technology XXIII*, vol. 11003. International Society for Optics and Photonics, 2019, p. 110030Q.

[98] M. Jahangir, K. Ponting, and J. O'loghlen, "Robust Doppler classification technique based on hidden Markov models," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 150, no. 1, pp. 33–36, 2003.

[99] Y. Yang, J. Lei, W. Zhang, and C. Lu, "Target classification and pattern recognition using micro-Doppler radar signatures," in *Seventh ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD'06)*. IEEE, 2006, pp. 213–217.

[100] V. C. Chen, "Spatial and temporal independent component analysis of micro-Doppler features," in *IEEE International Radar Conference, 2005.* IEEE, 2005, pp. 348–353.

[101] L. Liu, M. Popescu, M. Rantz, and M. Skubic, "Fall detection using doppler radar and classifier fusion," in *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*. IEEE, 2012, pp. 180–183.

[102] D. Yessad, A. Amrouche, M. Debyeche, and M. Djeddou, "Micro-Doppler classification for ground surveillance radar using speech recognition tools," in *Iberoamerican Congress on Pattern Recognition*. Springer, 2011, pp. 280–287.

[103] J. Li, S. L. Phung, F. H. C. Tivive, and A. Bouzerdoum, "Automatic classification of human motions using Doppler radar," in *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2012, pp. 1–6.

[104] S. Z. Gurbuz, C. Clemente, A. Balleri, and J. J. Soraghan, "Micro-Doppler-based in-home aided and unaided walking recognition with multiple radar and sonar systems," *IET Radar, Sonar & Navigation*, vol. 11, no. 1, pp. 107–115, 2016.

[105] B. Jokanovic, M. Amin, and F. Ahmad, "Radar fall motion detection using deep learning," in *2016 IEEE radar conference (RadarConf)*. IEEE, 2016, pp. 1–6.

[106] Y. Kim and T. Moon, "Classification of human activity on water through micro-Dopplers using deep convolutional neural networks," in *Radar Sensor Technology XX*, vol. 9829. International Society for Optics and Photonics, 2016, p. 982917.

[107] G. E. Smith, K. Woodbridge, and C. J. Baker, "Naïve bayesian radar micro-Doppler recognition," in *Radar, 2008 International Conference on*. IEEE, 2008, pp. 111–116.

[108] K. Saho, T. Sakamoto, T. Sato, K. Inoue, and T. Fukuda, "Accurate and real-time pedestrian classification based on UWB Doppler radar images and their radial velocity features," *IEICE transactions on communications*, vol. 96, no. 10, pp. 2563–2572, 2013.

[109] M. G. Amin, Z. Zeng, and T. Shan, "Hand gesture recognition based on radar micro-doppler signature envelopes," in *2019 IEEE Radar Conference (RadarConf)*. IEEE, 2019, pp. 1–6.

[110] 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.

[111] Y. Lin and J. Le Kernec, "Performance analysis of classification algorithms for activity recognition using micro-Doppler feature," in *2017 13th International Conference on Computational Intelligence and Security (CIS)*. IEEE, 2017, pp. 480–483.

[112] Y. Kim and B. Toomajian, "Hand gesture recognition using micro-doppler signatures with convolutional neural network," *IEEE Access*, vol. 4, pp. 7125–7130, 2016.

[113] G. E. Smith, K. Woodbridge, and C. J. Baker, "Radar micro-doppler signature classification using dynamic time warping," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 3, pp. 1078–1096, 2010.

[114] M. G. Amin, Z. Zeng, and T. Shan, "Arm motion classification using curve matching of maximum instantaneous doppler frequency signatures," *arXiv preprint arXiv:1911.03471*, 2019.

[115] M. G. Anderson, "Design of multiple frequency continuous wave radar hardware and micro-doppler based detection and classification algorithms," Ph.D. dissertation, 2008.

[116] Samraksh Company, "BumbleBee user's manual," https://samraksh.com/files/products/Bumblebee%20radar/BumbleBee-UM-v103.pdf, 2008, [Online; accessed 19-July-2018].

[117] J. Fang, H. Meng, H. Zhang, and X. Wang, "A low-cost vehicle detection and classification system based on unmodulated continuous-wave radar," in *2007 IEEE Intelligent Transportation Systems Conference*. IEEE, 2007, pp. 715–720.

[118] Y. G. Parra, B. Guzmán, P. González, J. Tomás *et al.*, "Development of a low-cost, short-range radar system to measure speed and distance," *Tecciencia*, vol. 12, no. 22, pp. 99–106, 2017.

[119] Willow Technologies, "Telosb datasheet," https://www.willow.co.uk/TelosB_Datasheet.pdf, 2008, [Online; accessed 19-July-2018].

[120] A. Dogra and P. Bhalla, "Image sharpening by gaussian and butterworth high pass filter," *Biomedical and Pharmacology Journal*, 2014.

[121] 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.

[122] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[123] M. D. Zeiler, "ADADELTA: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.

[124] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.

[125] 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.

[126] N. E. Klepeis, W. C. Nelson, W. R. Ott, J. P. Robinson, A. M. Tsang, P. Switzer, J. V. Behar, S. C. Hern, and W. H. Engelmann, "The national human activity pattern survey (NHAPS): a resource for assessing exposure to environmental pol-

lutants," *Journal of Exposure Science and Environmental Epidemiology*, vol. 11, no. 3, p. 231, 2001.

[127] M. E. Mlinac and M. C. Feng, "Assessment of activities of daily living, self-care, and independence," *Archives of Clinical Neuropsychology*, vol. 31, no. 6, pp. 506–516, 2016.

[128] W. Van Eeden, J. P. De Villiers, R. Berndt, W. A. Nel, and E. Blasch, "Micro-Doppler radar classification of humans and animals in an operational environment," *Expert Systems With Applications*, vol. 102, pp. 1–11, 2018.

[129] V. C. Chen, "Advances in applications of radar micro-Doppler signatures," in *Antenna Measurements & Applications (CAMA), 2014 IEEE Conference on*. IEEE, 2014, pp. 1–4.

[130] M. A. A. H. Khan, R. Kukkapalli, P. Waradpande, S. Kulandaivel, N. Banerjee, N. Roy, and R. Robucci, "RAM: Radar-based activity monitor," in *INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications, IEEE*. IEEE, 2016, pp. 1–9.

[131] G. Wang, C. Gu, T. Inoue, and C. Li, "A hybrid FMCW-interferometry radar for indoor precise positioning and versatile life activity monitoring," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 11, pp. 2812–2822, 2014.

[132] G. Diraco, A. Leone, and P. Siciliano, "A radar-based smart sensor for unobtrusive elderly monitoring in ambient assisted living applications," *Biosensors*, vol. 7, no. 4, p. 55, 2017.

[133] M. S. Seyfioğlu, S. Z. Gürbüz, A. M. Özbayoğlu, and M. Yüksel, "Deep learning of micro-Doppler features for aided and unaided gait recognition," in *2017 IEEE Radar Conference (RadarConf)*. IEEE, 2017, pp. 1125–1130.

[134] Y. Yang, C. Hou, Y. Lang, G. Yue, Y. He, and W. Xiang, "Person identification using micro-Doppler signatures of human motions and uwb radar," *IEEE Microwave and Wireless Components Letters*, vol. 29, no. 5, pp. 366–368, 2019.

[135] Y. Xiao, J. Lin, O. Boric-Lubecke, and V. M. Lubecke, "A ka-band low power doppler radar system for remote detection of cardiopulmonary motion," in *2005*

IEEE Engineering in Medicine and Biology 27th Annual Conference. IEEE, 2006, pp. 7151–7154.

[136] M. Mabrouk, S. Rajan, M. Bolic, M. Forouzanfar, H. R. Dajani, and I. Batkin, "Human breathing rate estimation from radar returns using harmonically related filters," *Journal of Sensors*, vol. 2016, 2016.

[137] L. Ren, Y. S. Koo, H. Wang, Y. Wang, Q. Liu, and A. E. Fathy, "Noncontact multiple heartbeats detection and subject localization using UWB impulse Doppler radar," *IEEE Microwave and Wireless Components Letters*, vol. 25, no. 10, pp. 690–692, 2015.

[138] H. Lee, B.-H. Kim, J.-K. Park, and J.-G. Yook, "A novel vital-sign sensing algorithm for multiple subjects based on 24-ghz fmcw doppler radar," *Remote Sensing*, vol. 11, no. 10, p. 1237, 2019.

[139] H. Hong, L. Zhang, C. Gu, Y. Li, G. Zhou, and X. Zhu, "Noncontact sleep stage estimation using a CW Doppler radar," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 8, no. 2, pp. 260–270, 2018.

[140] H. Li, A. Shrestha, F. Fioranelli, J. Le Kernec, H. Heidari, M. Pepa, E. Cippitelli, E. Gambi, and S. Spinsante, "Multisensor data fusion for human activities classification and fall detection," in *2017 IEEE SENSORS*. IEEE, 2017, pp. 1–3.

[141] W. Li, Y. Xu, B. Tan, and R. J. Piechocki, "Passive wireless sensing for unsupervised human activity recognition in healthcare," in *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*. IEEE, 2017, pp. 1528–1533.

[142] B. Jokanovic, M. G. Amin, Y. D. Zhang, and F. Ahmad, "Multi-window time–frequency signature reconstruction from undersampled continuous-wave radar measurements for fall detection," *IET Radar, Sonar & Navigation*, vol. 9, no. 2, pp. 173–183, 2014.

[143] F. Wang, M. Skubic, M. Rantz, and P. E. Cuddihy, "Quantitative gait measurement with pulse-Doppler radar for passive in-home gait assessment," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 9, pp. 2434–2443, 2014.

[144] A. Santra, R. V. Ulaganathan, and T. Finke, "Short-range millimetric-wave radar

system for occupancy sensing application," *IEEE sensors letters*, vol. 2, no. 3, pp. 1–4, 2018.

[145] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[146] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[147] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[148] Y. Wang, D. Wang, and F. Chen, "Abnormal behavior detection using trajectory analysis in camera sensor networks," *International Journal of Distributed Sensor Networks*, vol. 10, no. 1, p. 839045, 2013.

[149] M. Chebiyyam, R. D. Reddy, D. P. Dogra, H. Bhaskar, and L. Mihaylova, "Motion anomaly detection and trajectory analysis in visual surveillance," *Multimedia Tools and Applications*, vol. 77, no. 13, pp. 16 223–16 248, 2018.

[150] Q. Gao and S. Sun, "Trajectory-based human activity recognition with hierarchical Dirichlet process hidden Markov models," in *2013 IEEE China Summit and International Conference on Signal and Information Processing*. IEEE, 2013, pp. 456–460.

[151] G. Pusiol, F. Bremond, and M. Thonnat, "Trajectory based activity discovery," in *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, 2010, pp. 270–277.

[152] L. Huang, Q. Li, and Y. Yue, "Activity identification from GPS trajectories using spatial temporal POIs' attractiveness," in *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on location based social networks*. ACM, 2010, pp. 27–30.

[153] W. Huang, M. Li, W. Hu, G. Song, X. Xing, and K. Xie, "Cost sensitive GPS-based activity recognition," in *2013 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. IEEE, 2013, pp. 962–966.

[154] L. Liao, D. Fox, and H. Kautz, "Location-based activity recognition," in *Advances*

in Neural Information Processing Systems, 2006, pp. 787–794.

[155] J. Castorena and S. Agarwal, "Ground-edge-based LIDAR localization without a reflectivity calibration for autonomous driving," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 344–351, 2017.

[156] R. W. Wolcott and R. M. Eustice, "Robust LIDAR localization using multiresolution Gaussian mixture maps for autonomous driving," *The International Journal of Robotics Research*, vol. 36, no. 3, pp. 292–319, 2017.

[157] ——, "Fast LIDAR localization using multiresolution gaussian mixture maps," in *2015 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2015, pp. 2814–2821.

[158] J. Zhao, H. Xu, H. Liu, J. Wu, Y. Zheng, and D. Wu, "Detection and tracking of pedestrians and vehicles using roadside LiDAR sensors," *Transportation research part C: emerging technologies*, vol. 100, pp. 68–87, 2019.

[159] H. Wang, B. Wang, B. Liu, X. Meng, and G. Yang, "Pedestrian recognition and tracking using 3D LiDAR for autonomous vehicle," *Robotics and Autonomous Systems*, vol. 88, pp. 71–78, 2017.

[160] A. Leigh, J. Pineau, N. Olmedo, and H. Zhang, "Person tracking and following with 2d laser scanners," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 726–733.

[161] C. Park, S. Kim, P. Moghadam, C. Fookes, and S. Sridharan, "Probabilistic surfel fusion for dense LiDAR mapping," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2418–2426.

[162] J. Zhang and S. Singh, "Low-drift and real-time LiDAR odometry and mapping," *Autonomous Robots*, vol. 41, no. 2, pp. 401–416, 2017.

[163] K. Kim, M. Cao, S. Rao, J. Xu, S. Medasani, and Y. Owechko, "Multi-object detection and behavior recognition from motion 3D data," in *CVPR 2011 WORKSHOPS*. IEEE, 2011, pp. 37–42.

[164] A. Panangadan, M. Matarić, and G. S. Sukhatme, "Tracking and modeling of human activity using laser rangefinders," *International Journal of Social Robotics*, vol. 2, no. 1, pp. 95–107, 2010.

[165] Z. Ma, J. Bigham, S. Poslad, B. Wu, X. Zhang, and E. Bodanese, "Device-free, activity during daily life, recognition using a low-cost Lidar," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–6.

[166] X. Song, X. Shao, R. Shibasaki, H. Zhao, J. Cui, and H. Zha, "A novel laser-based system: Fully online detection of abnormal activity via an unsupervised method," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 1317–1322.

[167] W. Xi, D. Huang, K. Zhao, Y. Yan, Y. Cai, R. Ma, and D. Chen, "Device-free human activity recognition using CSI," in *Proceedings of the 1st Workshop on Context Sensing and Activity Recognition*. ACM, 2015, pp. 31–36.

[168] S. Sempena, N. U. Maulidevi, and P. R. Aryan, "Human action recognition using dynamic time warping," in *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*. IEEE, 2011, pp. 1–5.

[169] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowledge and information systems*, vol. 7, no. 3, pp. 358–386, 2005.

[170] K. L. Y. Siang and S. W. Khor, "Path clustering using dynamic time warping technique," in *2012 8th International Conference on Computing Technology and Information Management (NCM and ICNIT)*, vol. 1. IEEE, 2012, pp. 449–452.

[171] S. Kamal, A. Jalal, and D. Kim, "Depth images-based human detection, tracking and activity recognition using spatiotemporal features and modified hmm," *J. Electr. Eng. Technol*, vol. 11, no. 3, pp. 1921–1926, 2016.

[172] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of wifi signal based human activity recognition," in *Proceedings of the 21st annual international conference on mobile computing and networking*. ACM, 2015, pp. 65–76.

[173] T. Fuse and K. Kamiya, "Statistical anomaly detection in human dynamics monitoring using a hierarchical dirichlet process hidden markov model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 11, pp. 3083–3092, 2017.

[174] J. Liang, L. Jiang, J. C. Niebles, A. Hauptmann, and L. Fei-Fei, "Peeking into the

future: Predicting future person activities and locations in videos," *arXiv preprint arXiv:1902.03748*, 2019.

[175] D. Tran, H. Wang, L. Torresani, and M. Feiszli, "Video classification with channel-separated convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 5552–5561.

[176] C. Benedek, B. Gálai, B. Nagy, and Z. Jankó, "Lidar-based gait analysis and activity recognition in a 4d surveillance system," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 1, pp. 101–113, 2016.

[177] M. Parimala, D. Lopez, and N. Senthilkumar, "A survey on density based clustering algorithms for mining large spatial databases," *International Journal of Advanced Science and Technology*, vol. 31, no. 1, pp. 59–66, 2011.

[178] P. H. Ahmad and S. Dang, "Performance evaluation of clustering algorithm using different datasets," *International Journal of Advance Research in Computer Science and Management Studies*, vol. 3, pp. 167–173, 2015.

[179] J. Ashburner and K. J. Friston, "Spatial transformation of images," *Human brain function*, pp. 43–58, 1997.

[180] P. Senin, "Dynamic time warping algorithm review," *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, vol. 855, no. 1-23, p. 40, 2008.

[181] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden Markov model," in *Proceedings 1992 IEEE Computer Society conference on computer vision and pattern recognition.* IEEE, 1992, pp. 379–385.

[182] A. Veit, M. J. Wilber, and S. Belongie, "Residual networks behave like ensembles of relatively shallow networks," in *Advances in neural information processing systems*, 2016, pp. 550–558.

[183] S. S. Ram, Y. Li, A. Lin, and H. Ling, "Human tracking using Doppler processing and spatial beamforming," in *2007 IEEE Radar Conference.* IEEE, 2007, pp. 546–551.

[184] T. Wagner, R. Feger, and A. Stelzer, "Radar signal processing for jointly estimating tracks and micro-Doppler signatures," *IEEE Access*, vol. 5, pp. 1220–1238, 2017.

[185] B. Erol and M. G. Amin, "Radar data cube processing for human activity recognition using multi subspace learning," *IEEE Transactions on Aerospace and Electronic Systems*, 2019.

[186] S. Spinsante, M. Pepa, S. Pirani, E. Gambi, and F. Fioranelli, "Micro-Doppler radar and depth sensor fusion for human activity monitoring in aal," in *Convegno Nazionale Sensori*. Springer, 2018, pp. 519–528.

[187] H. Li, A. Shrestha, H. Heidari, J. Le Kernec, and F. Fioranelli, "Magnetic and radar sensing for multimodal remote health monitoring," *IEEE Sensors Journal*, 2018.