4D RADAR IMAGING FOR TARGET DETECTION AND CLASSIFICATION USING DEEP-LEARNING

BY

LIYAANA SHAHIRAH BINTI WAN ABD AZIZ

A thesis submitted in fulfilment of the requirement for the degree of Master of Science in Engineering

Kulliyyah of Engineering
International Islamic University Malaysia

JUNE 2025

ABSTRACT

The goal of this project is to develop an object detection and classification system for road crossing areas as part of a monitoring system using 4D radar imaging with a deep-learning neural network approach. In this work, we utilised deep neural networks powered by Keras and Tensorflow to detect and classify multiple pedestrians, cars, buses, and trucks. This paper presents Retina-4F, which is a multi-chip radar imaging system with high range resolution for object detection and localization. Retina-4F, which was developed by Smart Radar System, allows the system to provide 4D real-time information about the target. Retina-4F utilises a multi-chip cascade onboard with three transmitters and four receivers in each chip. We demonstrated two road-crossing scenes to collect data for creating a point cloud dataset with a target class label to be used for training and testing a deep learning model. There are two main sensors implemented in this work: Retina-4F as a 4D radar imaging and a mono-camera. The data from both sensors is pre-processed using DBSCAN and YOLOv7. Retina-4F operates at 77 GHz, and the test was conducted in two different road areas. After conducting data measurement at two road crossing areas, the collected data is passed for preprocessing and data fusion processing. This results in a complete point cloud dataset with approximately 10,000 frames of point cloud images that can be used for neural network training and testing. The model evaluation showed satisfying performance of the deep neural network in classifying multiple targets with 97 percent of overall accuracy. The approach of sensor data fusion for multiple target classification shows good results where it manages to distinguish different types of targets: cars, pedestrians, buses, and trucks. The proposed radar point cloud classification using sensor fusion can be applied to a wide variety of complex monitoring applications.

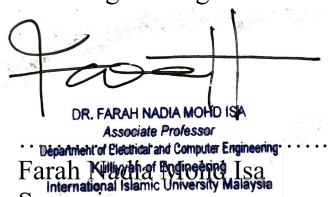
ملخص البحث

الهدف من هذا المشروع هو تطوير نظام كشف وتصنيف الأشياء لمناطق عبور الطرق كجزء من نظام المراقبة باستخدام التصوير الراداري رباعي الأبعاد مع نهج الشبكة العصبية للتعلم العميق. في هذا العمل، استخدمنا شبكات عصبية عميقة مدعومة من Keras وTensorflow لاكتشاف وتصنيف العديد من المشاة والسيارات والحافلات والشاحنات. تقدم هذه الورقة نظام Retina-4F، وهو نظام تصوير راداري متعدد الشرائح يتمتع بدقة عالية المدى لاكتشاف الأجسام وتحديد موقعها. ويتيح نظام Retina-4F، الذي تم تطويره بواسطة نظام راداري ذكي، يوفر هذا النظام معلومات فورية رباعية الأبعاد حول الهدف. يستخدم Retina-4F سلسلة متعددة الشرائح مدمجة مع ثلاثة أجهزة إرسال وأربعة أجهزة استقبال في كل شريحة. لقد أظهرنا مشهدين لعبور الطريق لجمع البيانات وذلك لإنشاء مجموعة بيانات سحابية نقطية مع تسمية الفئة المستهدفة لاستخدامها في التدريب واختبار نموذج التعلم العميق. هناك مستشعران رئيسيان تم تنفيذهما في هذا العمل: Retina-4F كجهاز تصوير راداري رباعي الأبعاد وكاميرا أحادية. تتم معالجة البيانات الواردة من كلا المستشعرين مسبقًا باستخدام DBSCAN وYOLOv7. يعمل Retina-4F بتردد 77 جيجاهرتز، وتم إجراء الاختبار في منطقتين مختلفتين على الطريق. بعد إجراء قياسات البيانات في منطقتين لعبور الطرق، تم تمرير البيانات المجمعة للمعالجة المسبقة ومعالجة دمج البيانات. وينتج عن ذلك مجموعة بيانات سحابية نقطية كاملة تحتوي على ما يقرب من 10000 إطار من الصور السحابية النقطية التي يمكن استخدامها للتدريب على الشبكات العصبية واختبارها. أظهر تقييم هذا النموذج أداءً مُرضيًا للشبكة العصبية العميقة في تصنيف أهداف متعددة بنسبة 97 % من الدقة الإجمالية. يظهر نهج دمج بيانات الاستشعار لتصنيف الأهداف المتعددة نتائج جيدة عندما يتمكن من التمييز بين أنواع مختلفة من الأهداف: السيارات والمشاة والحافلات والشاحنات. يمكن

تطبيق التصنيف المقترح لسحابة نقاط الرادار باستخدام دمج أجهزة الاستشعار على مجموعة واسعة من تطبيقات المراقبة المعقدة.

APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science in Engineering.



Farah Nadia Mohd Isa
Supervisor

……
Faridah Abd Rahman
Co-Supervisor

……
Huda Adibah Mohd Ramli
Co-Supervisor

……
Norun Farihah Abdul Malek
Co-Supervisor

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science in Engineering.

……
Teddy Surya Gunawan
Internal Examiner

……
Nur Emileen Abd Rashid
External Examiner

This thesis was submitted to the Department of Electrical and Computer Engineering and is accepted as a fulfilment of the requirement for the degree of Master of Science in Engineering.

…….
Othman O-Khalifa
Head, Department of Electrical and Computer Engineering

This thesis was submitted to the Kulliyyah of Engineering and is accepted as a fulfilment of the requirement for the degree of Master of Science in Engineering.

…….
Khairul Azami Sidek
Dean, Kulliyyah of Engineering

DECLARATION

I hereby declare that this thesis is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Liyaana Shahirah Binti Wan Abd Aziz

Signature ..... Date 18/6/2025

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

4D RADAR IMAGING FOR TARGET DETECTION AND CLASSIFICATION USING DEEP-LEARNING

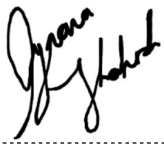
I declare that the copyright holder of this thesis are jointly owned by the student and IIUM.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below

1. Any material contained in or derived from this unpublished research may only be used by others in their writing with due acknowledgement.
2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purpose.
3. The IIUM library will have the right to make, store in a retrieval system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledge that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Liyaana Shahirah Binti Wan Abd Aziz



Signature

18/6/2025
Date

This thesis is dedicated to my parents and loved ones for laying the foundation of what I turned out to be in life.

ACKNOWLEDGEMENTS

All glory is due to Allah, the Almighty, whose Grace and Mercies have been with me throughout the duration of my programme. Although, it has been tasking, His Mercies and Blessings on me ease the herculean task of completing this thesis.

I am most indebted to by supervisor, Assoc. Prof. Dr. Farah Nadia Mohd Isa, whose enduring disposition, kindness, promptitude, thoroughness and friendship have facilitated the successful completion of my work. I put on record and appreciate her detailed comments, useful suggestions and inspiring queries which have considerably improved this thesis. Her brilliant grasp of the aim and content of this work led to her insightful comments, suggestions and queries which helped me a great deal. Despite her commitments, she took time to listen and attend to me whenever requested. The moral support she extended to me is in no doubt a boost that helped in building and writing the draft of this research work.

I am also grateful to my co-supervisor, Asst. Prof. Dr. Faridah Abd Rahmah, Assoc. Prof. Dr. Huda Adibah Mohd Ramli, and Assoc. Prof. Dr. Norun Farihah Abdul Malek whose support and cooperation contributed significantly to the outcome of this work. Your guidance, expertise, and dedication have been instrumental in shaping my research and helping me navigate the complexities of this academic journey.

I extend my heartfelt gratitude to my dedicated supervisor, Prof. George Shaker at the University of Waterloo, Canada for his unwavering support and guidance throughout my six-month attachment as International Visiting Graduate Student. His profound expertise and willingness to share his knowledge enriched my research journey and played an instrumental role in shaping the outcome of this thesis. I also want to express my sincere appreciation to my mentor, Ahmad Reza, for his invaluable assistance and mentorship during this period. Their insights, constructive feedback, and collaboration were integral to the success of my project and my growth as a researcher. I am honored to have had the privilege of working with such exceptional mentors during this pivotal phase of my academic journey.

I would like to express my sincere appreciation to Dr. Arvind Hari Narayanan for his generous sponsorship of the radar equipment used in my project. His support not only provided me with the necessary tools to conduct my research but also enabled me to explore new dimensions within my field of study. This invaluable contribution significantly enhanced the quality and depth of my project, and I am truly grateful for his investment in my academic pursuits.

Lastly, my gratitude goes to my beloved parents and lovely family; for their prayers, understanding, and endurance while away. Your unwavering support has been the cornerstone of my journey, providing me with the strength and motivation to overcome the challenges that came my way. Your love and encouragement have been my constant companions, and I am deeply thankful for the sacrifices you have made to see me succeed.

Once again, we glorify Allah for His endless mercy on us one of which is enabling us to successfully round off the efforts of writing this thesis. Alhamdulillah

TABLE OF CONTENTS

Abstract……ii
Abstract in Arabic……iii
Approval Page……...v
Declaration……...vii
Copyright……..viii
Dedication……ix
Acknowledgement……....x
List of Tables…… xiii
List of Figures……... xiv
List of Abbreviations……...xviii

CHAPTER ONE: INTRODUCTION……...1
1.1 Background of the Study……1
1.2 Problem Statement……3
1.3 Research Objectives……5
1.4 Research Methodology……5
1.5 Scope of Research……7
1.6 Research Significance……8
1.7 Thesis Organization……8

CHAPTER TWO: LITERATURE REVIEW……10
2.1 Introduction ……...10
2.2 Radar Imaging Technology……...10
2.2.1 Automotive Radar Imaging……14
2.3 Millimeter-Wave FMCW MIMO-Based 4d Radar Imaging……..17
2.3.1 FMCW MIMO Millimeter-Wave Radar ……17
2.3.2 FMCW Signal Model……22
2.3.3 Resolution: Range, Velocity and Angular……25
2.4 Radar Imaging vs Camera vs Lidar……27
2.5 Object Clustering and Classification using Deep Learning……30
2.5.1 Deep Learning……30
2.5.2 Clustering……...33
2.5.3 YOLO for Image Object Recognition……...41
2.5.4 Radar Imaging Classification ……47
2.6 Related Works……49
2.7 Summary……56

CHAPTER THREE: RESEARCH METHODOLOGY……57
3.1 Introduction ……57
3.1.1 Flow Chart……57
3.2 Retina-4F and Camera ……60
3.2.1 System Hardware Platform ……60
3.2.2 Radar System Parameter……62
3.2.3 System Software Platform……64
3.2.4 ROS Data Acquisition……65

3.3 Experimental Setup and Workflow ……...…66
3.3.1 Experimental Setup for Data Collection…….…….66
3.3.2 Experimental Workflow for Point Cloud Dataset……...…68
3.4 Data Processing Approaches…….……69
3.4.1 Clustering…….…….70
3.4.2 YOLOv7 Object Detection……...…….72
3.4.3 Point Cloud Projection onto 2D Image…….…….…….74
3.4.4 Point Cloud Classification…….…77
3.5 Summary……....81

CHAPTER FOUR: RESULTS AND DISCUSSIONS……83
4.1 Introduction…….…….83
4.2 Task One: Preliminary Test …….……..84
4.2.1 Experimental Setup…….84
4.2.2 Experimental Result……85
4.2.3 Discussion……...86
4.3 Task Two: Radar Performance Test……...87
4.3.1 Experimental Setup……88
4.3.2 Experimental Results…….…..…91
4.3.2.1 Maximum Detection Range of a Pedestrian……91
4.3.2.2 Multiple Pedestrians Detection …….……..….95
4.3.2.3 Motorcycle Detection…..……98
4.3.2.4 Car Detection……..……101
4.3.3 Discussion……..……...103
4.4 Task Three: Data Collection at Road Crossing Areas…….……105
4.4.1 Experimental Setup…….……..…105
4.4.2 Experimental Results……...109
4.4.2.1 Person…….……..…….109
4.4.2.2 Car……....……...….112
4.4.2.3 Bus……..…….……..…115
4.4.2.4 Truck…….…….……117
4.4.3 Discussion…….……..120
4.5 Task Four: Radar Point Cloud Classification …….…..……121
4.5.1 Target's Classification Performance…….123
4.5.2 Overall System Performance…….…..……128
4.5.3 Discussion…….……132
4.5.4 Challenges……139
4.6 Summary …….……...…….140

CHAPTER FIVE: RECOMMENDATIONS AND CONCLUSION…….……142
5.1 Conclusion…….…..……142
5.2 Recommendations for Future Research …….…….……144

REFERENCES……...…….146

LIST OF PUBLICATIONS……154

LIST OF TABLES

Table 2.1	Comparison Between Camera, Lidar and Radar	29
Table 2.2	Table of Comparison between YOLOv7 and YOLOv8	46
Table 2.3	Table of Summary of Related Works	52
Table 3.1	Radar Operating System Parameter	63
Table 4.1	Number of Data Points for each Class	122
Table 4.2	Model Evaluation Report (3 layers)	129
Table 4.3	Model Evaluation Report (4 layers)	130
Table 4.4	Model Evaluation Report (5 layers)	131
Table 4.5	Model Evaluation Report – Benchmark Paper	136
Table 4.6	Model Evaluation Report (4 layers)	137

LIST OF FIGURES

Figure 2.1	Synthetic Aperture Length	12
Figure 2.2	Roadmap of Driver Assistance Functions using Radar over Years	14
Figure 2.3	Millimetre Wave Signal	17
Figure 2.4	Angle-FFT plot. Two black circles describe two objects at ±10◦ individually	19
Figure 2.5	Antenna pattern of (a) Short Range Mode Radar (b) Long Range Mode Radar	21
Figure 2.6	FMCW Chirp Signal with Amplitude as Function of Time	23
Figure 2.7	Constant Frequency IF Signal from the Transmitted and Received Chirp Signals	25
Figure 2.8	Reflected Object Wave Receive at Two Adjacent Antennas	24
Figure 2.9	Supervised and Unsupervised Machine Learning	32
Figure 2.10	YOLO timeline	41
Figure 2.11	YOLO network architecture that made up with Convolutional Neural Network	43
Figure 3.1	Flow Chart	59
Figure 3.2	Retina-4F Imaging Radar	61
Figure 3.3	ELP USB Mono-Camera	62
Figure 3.4	RVIZ Visualization Tool Environment	64
Figure 3.5	Experimental Setup for Data Collection at Road Crossing Area	67
Figure 3.6	Radar and Camera Mounting Arrangement	68
Figure 3.7	Experimental Workflow	69

Figure 3.8	YOLOv7 Object Classification with Bounding Box	74
Figure 3.9	Radar to Pixel Coordinate System Transformation	76
Figure 3.10	Difference Between Camera and Radar Coordinate System. (a) Camera Coordinate System. (b) Radar Coordinate System.	77
Figure 3.11	Feedforward Neural Network	78
Figure 3.12	Proposed Feedforward Neural Network System Architecture	79
Figure 4.1	Indoor Testing of Retina-4F	84
Figure 4.2	Point Clouds of Object 1 in RVIZ	85
Figure 4.3	Radar Point Cloud Output Information	86
Figure 4.4	Doppler Velocity Colorbar Indicator	88
Figure 4.5	IIUM Marching Field as Observed in Google Maps	89
Figure 4.6	Experimental Setup of the Controlled Area	90
Figure 4.7	Radar Mounted at the Controlled Area	90
Figure 4.8	Pedestrian. (Camera View)	91
Figure 4.9	Mean Average of Maximum Detectable Distance for a Pedestrian	92
Figure 4.10	Doppler Velocity Distribution for a Pedestrian	93
Figure 4.11	Pedestrian. (X-Y-axis 2D Plot)	94
Figure 4.12	Radar Data for a Pedestrian at Range 32.9 meters	94
Figure 4.13	Multiple Pedestrians. (Camera View)	95
Figure 4.14	Doppler Velocity Distribution for Multiple Pedestrian	96
Figure 4.15	Multiple Pedestrians Clusters across the Range	97
Figure 4.16	Motorcycle. (Camera View)	98
Figure 4.17	Doppler Velocity Distribution for a Motorcycle	99

Figure 4.18	Point Cloud of Overall Route Taken by the Car Detected by The Radar Image	100
Figure 4.19	Car. (Camera View)	101
Figure 4.20	Doppler Velocity Distribution for a Car	102
Figure 4.21	Point Cloud of Overall Route Taken by the Car Detected by The Radar Image	103
Figure 4.22	Road Crossing Area at Location 1	106
Figure 4.23	Road Crossing Area at Location 2	106
Figure 4.24	Experimental Setup at Location 1	107
Figure 4.25	Experimental Setup at Location 2	107
Figure 4.26	Radar and Camera Field of View at Location 1	108
Figure 4.27	Radar and Camera Field of View at Location 2	108
Figure 4.28	Pedestrians - Point Cloud Projection onto Image	110
Figure 4.29	Pedestrians - Point Cloud Projection and Bounding Box Plot onto Image	110
Figure 4.30	Pedestrians - Scatterer Plot of Point Cloud Clusters	111
Figure 4.31	Pedestrians - Point Cloud Labelling at Location 1	112
Figure 4.32	Pedestrians - Point Cloud Labelling at Location 2	112
Figure 4.33	Cars - Point Cloud Projection onto Image	113
Figure 4.34	Cars - Point Cloud Projection and Bounding Box Plot onto Image	113
Figure 4.35	Cars - Scatterer Plot of Point Cloud Clusters	114
Figure 4.36	Cars - Point Cloud Labelling at Location 1	114
Figure 4.37	Cars - Point Cloud Labelling at Location 2	115
Figure 4.38	Bus - Point Cloud Projection onto Image	115

Figure 4.39	Bus - Point Cloud Projection and Bounding Box Plot onto Image	116
Figure 4.40	Bus - Scatterer Plot of Point Cloud Clusters	116
Figure 4.41	Bus - Point Cloud Labelling at Location 1	117
Figure 4.42	Bus - Point Cloud Labelling at Location 2	117
Figure 4.43	Truck - Point Cloud Projection onto Image	118
Figure 4.44	Truck - Point Cloud Projection and Bounding Box Plot onto Image	118
Figure 4.45	Truck - Scatterer Plot of Point Cloud Clusters	119
Figure 4.46	Truck - Point Cloud Labelling at Location 1	119
Figure 4.47	Truck - Point Cloud Labelling at Location 2	120
Figure 4.48	Custom Dataset Class Distribution	122
Figure 4.49	Confusion Matrix of 3 Layers Neural Network Model	125
Figure 4.50	Confusion Matrix of 4 Layers Neural Network Model	126
Figure 4.51	Confusion Matrix of 5 Layers Neural Network Model	127
Figure 4.52	Confusion Matrix – Benchmark Paper	134
Figure 4.53	Confusion Matrix of 4 Layers Neural Network Model	135
Figure 4.54	Example of Mislabelling of Target Due to Yolov7 Classification Predictions	139

LIST OF ABBREVIATIONS

ADAS	Adaptive Driver-Assistance Systems
CMOS	Complementary Metal-Oxide-Semiconductor
FMCW	Frequency Modulated Continuous Wave
MIMO	Multiple Input Multiple Output
4D	4 Dimensional
3D	3 Dimensional
2D	2 Dimensional
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
YOLO	You Only Look Once
EHF	Extremely High Frequency
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
PCA	Principal Component Analysis

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND OF THE STUDY

Over the past several years, there has been fast development in 77 GHz automotive radar technology. This technology serves as a cost-effective alternative to Light Detection and Ranging (LIDAR), in advanced driver assistance systems (ADAS) (Lim et al., 2019). It has been successfully deployed to functions such as autonomous emergency braking (AEB) and adaptive cruise control (ACC). Automotive radar offers not only affordable pricing but also demonstrates resilience in various lighting and weather circumstances. Moreover, it has the capability to accurately measure range, azimuth angles, and instantaneous velocity.

Many automotive markets are directing their research and development activities towards this field, resulting in progressively advanced levels of vehicle autonomy. In the present era, cars are mainly designed to be highly automated, enabling them to carry out precise driving functions with minimal human involvement. In order for a vehicle to attain advanced levels of independence, it is necessary that it acquire the capability to precisely perceive and comprehend its surroundings. The greater the extent and dependability of the data it can collect, the more informed choices it may make in diverse on-road and off-road situations.

Current vehicle sensor technology includes several kinds of technologies, such as radars, video camera, mono-cameras, and LIDAR (Cho et al., 2014). As vehicles undertake more complicated driving responsibilities, the need for strong and dependable systems becomes important. By integrating data from many sensor types, sensor fusion can greatly improve the performance and endurance of a system.

Although cameras and LIDAR are useful sensor technologies, they have limits, particularly in challenging weather and low-light scenarios. Camera sensor systems may have difficulties when faced with weather conditions limitations that scatter light and decrease signal quality. On the other hand, radar systems employ electromagnetic waves that have the ability to pass through scattering particles with little loss of energy, thereby ensuring its dependability in different situations. Radars provide numerous benefits compared to other sensors, such as the capacity to directly measure doppler velocities and robust performance in all lighting circumstances.

4D radar imaging is a modern sensor technology that offers precise and real-time information on an object's distance, speed, horizontal angle, and vertical angle. The acquisition of this high-resolution point cloud data is essential for precise environmental perception and object classification. The resilience of radar systems under harsh weather, along with its system to provide intricate data, allows 4D radar to be advantageous for self-driving vehicles.

Object classification is a vital problem for autonomous cars, particularly in high-risk situations like road intersections. Precise recognition and classification of entities, such as individuals on foot, bicyclists, and automobiles, enable the vehicle in question to make well-informed choices, hence improving safety and efficiency. Conventional radar imaging systems are limited in their ability to accurately classify objects due to restrictions in radar resolution and sparse point cloud data quality. 4D radar imaging offers high-resolution point cloud data that greatly improves the capability to differentiate between various objects and their motions in intricate settings such as traffic intersections.

In order to adopt 4D radar imaging for classifying objects at crossings, it is important to have a high-quality personalised dataset. This project includes the creation of a dataset by combining data from 4D radar imaging and other sensors using sensor fusion techniques. This customised dataset functions as the fundamental basis for training and evaluating a deep neural network, which is the essence of the object

classification system. It is essential to have a diverse dataset that includes many scenarios, item types, and environmental circumstances to ensure the neural network learns properly and can apply its knowledge to different situations.

Deep neural networks (DNNs) have shown outstanding results in diverse areas, including automotive applications. DNN is employed in this study to accurately classify items at road crossings by processing the fused sensor data between radar point cloud and camera image. In order to effectively train the DNN, it is necessary to have an extensive and broad dataset that can accurately represent the complexity of real-world settings. Subsequently, the DNN that has been training would undergo testing and validation to verify its precision and dependability in practical scenarios.

The purpose of this project is to utilise the current advanced technology of 4D radar imaging and deep neural network algorithms in order to create a system for classifying objects at road crossings. The project aims to investigate the possibility of using high-resolution 4D radar for safety and awareness of situations in road crossing areas. This will be achieved by combining data from radar and camera sensors to create a comprehensive dataset and train a deep neural network.

1.2. PROBLEM STATEMENT

In developing deep neural network object classification system using radar as part of monitoring system, there have been some challenges arises. Firstly, the challenges in classifying target using radar are due to the sparsity of the radar data (point cloud). Every individual data point within a radar point cloud corresponds to a reflection originating from either a surface or an object. The distribution of these points is generally characterised by sparsity, indicating an uneven dispersion within the surrounding environment. Sparse point clouds may appear due to various variables, such as the radar's limited range of view, the separation between objects, and the

resolution of the radar. Sparse point clouds offer a significant obstacle for accurate object recognition and classification due to the limitations of typical techniques that are designed to handle dense point clouds. These techniques often struggle to extract meaningful features from sparse data, hence limiting their effectiveness in this context. This study aims to proposed a method of sensor fusion for labelling sparse point cloud data according to its class.

Next, with the advantages of radar sensor to be able to maintain its performance effectively in adverse weather conditions and low light environment, the growth of radar-based monitoring systems has been showing a steady increase. Hence, there is a growing demand to achieve exceptional standards of accuracy in terms of multiple object detection. Even so, in past experiments and research conducted by others, radar detection accuracy is questionable when using for object detection in monitoring system. The accuracy of object detection is known to be decreased when the range of the object increase and as well the number of objects increases. This problem needs to be solved especially when crossing area may involve with many objects or target for instance cars and pedestrian. Regardless the number of objects or targets presence in a time, the performance of the radar must be maintained and stable. Hence, in this study, we aims to study the capability of the radar in maintaining its performance in detecting multiple target within extended range.

In addition, the task of creating a reliable deep learning system that can effectively classify radar-derived point clouds into specific item categories presents a significant and essential challenge. In order to ensure the dependable functionality of monitoring systems, it is essential that the system possesses the capability to not only identify the existence of objects but also differentiate between diverse object classifications, including but not limited to cars, pedestrians, cyclists and many more. The recognition process becomes more complex by the inherent noise, sparsity, and changes in object shape and size that characterise radar point cloud data. Moreover, it is frequently observed that real-life situations encompass dynamic and complex surroundings, necessitating the development of a solution that can effectively adjust to a wide range of variables. The challenge can be addressed through the utilisation of

sensor fusion, a technique that involves the integration of diverse sensor data to improve the categorization of point cloud data into their appropriate class types.

By integrating data obtained from radar point clouds with data acquired from other sensor modalities, such as camera, a multi-modal deep learning system can be developed. This study aims to enhanced precision in radar object classification by creating a custom dataset that fused radar point cloud data and camera data to train neural network model for radar point cloud object classification.

1.3 RESEARCH OBJECTIVES

1. To investigate the high-resolution detection capability of the 4D Radar Imaging with different types of targets vary in size considering the point cloud data behaviour.
2. To develop object detection and classification model using deep learning method to classify objects using 3 different models.
3. To validate the detection and classification model by using data measurement on road crossing areas.

1.4 RESEARCH METHODOLOGY

At the beginning of this project, a literature review on radar technology, 4D radar imaging, object clustering and classification, and other related studies based on previous related research was done. Radar point cloud classification techniques have been chosen in this work, which is based on sensor data fusion.

After a literature review, the initial work starts by testing the working performance of the 4D radar imaging on its capability to detect different types of targets of different sizes. The technique chosen here is experimental; it is conducted in an open area with controlled space. This step covered the first objective, which is to investigate the high-resolution detection capability of 4D radar imaging with different types of targets varying in size.

Next, an experiment is conducted at a road crossing area for the purpose of data collection. The techniques used here are also experimental. These experiments will be done at two different locations in order to create customs datasets of point cloud data. The dataset created for this work are based on sensor data fusion between radar and camera. DBSCAN techniques will be used for clustering point cloud data, while the YOLO algorithm will be used for image object classification. Hence, the data processed from both DBSCAN and YOLO will be integrated, creating a customised dataset that will be used for training and testing deep neural networks. The customs dataset is split into 80% for model training purposes and 20% for model testing purposes.

Moreover, in this project, we will design and develop the object classification system using Keras and Tensorflow using Python programming. Keras and Tensorflow are the frameworks used to develop the deep neural network models for object classification in this work. This step covered the second objective of this project which is to develop object detection and classification model using deep learning method.

Finally, based on the custom dataset, we already split 20% of it for the purposes of testing the deep neural network model. Hence, after the process of developing and training the model, we will proceed in testing the model to investigate the performance of the model in classifying the desired objects into its respective classes. Hence, using the dataset, we can test the performance of the deep neural network model and eventually portray objective number 3.

1.5 SCOPE OF RESEARCH

The scope of this project is limited to radar point cloud classification using 4D radar imaging at road crossing area. This research is conducted to develop a classification system using Retina-4F image radar and mono-camera as part of data sensor fusion. The radar is proposed to operate in frequency of 77 GHz which is in EHF frequency band respect to the specification of the Retina-4F radar. This study is conducted at road crossing area in two different places. Since the aim of this study is to develop object detection and classification system at road crossing area, the experiment is conducted at busy pedestrian crossing area which several objects can be collected including pedestrian, bicycle, car, bus and truck. Other types of objects that could be presented at road crossing area such as animals are not taken into consideration and this stage of the project and can be considered in future works of this study.

Moreover, the classification system using deep learning was developed only for the after-processing of point cloud data. This project's main work is creating our own custom dataset from the data collected at the road crossing area and developing a deep neural network. Using the custom dataset, we use it to train and test the deep neural networks for the classification system. However, this classification system will not be integrated into the radar system for real-time object classification. There are two main reasons for these limitations, one of which is due to the project time limit. Integrating the deep neural network into the radar system requires much more study and work, which will take more time. Next, integrating the neural network with the radar system requires access to the radar system itself, as this project is using a ready-to-use radar system. Access to the radar system is limited as the developer of the radar protects the system's confidentiality.

1.6 RESEARCH SIGNIFICANCE

This thesis explores the field of advanced sensor fusion, specifically the integration of high-definition radar technology running at a frequency of 77 GHz with image data for the purpose of deep neural network-based object classification. The present study signifies a significant progression in the area of monitoring systems and autonomous vehicle technologies. By utilising advanced radar technology commonly used in autonomous cars, we overcome the limitations of conventional monitoring systems. The integration of radar point cloud data and images not only introduces a novel methodology but also provides numerous opportunities for improving the accuracy of object classification and expanding its practicality in real-world scenarios.

This study addresses the task of effectively integrating two different data sources, with the ultimate goal of enhancing the accuracy of object classification. The increasing need for accurate object recognition and classification using radar sensors in intricate settings has led to a growing demand. Our research has the potential to significantly enhance the capabilities of monitoring systems. The utilisation of our approach, which is built on deep neural networks, holds the potential to greatly enhance the precision of object classification. Consequently, this advancement would make notable contributions to the safety, efficiency, and dependability of autonomous systems, surveillance, and a range of other applications. In a modern time defined by the increasing influence of technology on various aspects of human existence, the present study represents a significant advancement in the development radar point cloud classification for the application of monitoring systems.

1.7 THESIS ORGANIZATION

The present thesis is structured into five primary sections, which include an Introduction, Literature Review, Methodology, Results and Discussion, and

Conclusion and Recommendations. Chapter 1 provides an overview of the background studies pertaining to 4D radar imaging technology. This chapter encompasses the problem statements and research objectives that serve as the fundamental basis for this study. Furthermore, this chapter provides an explanation of the significance of this research work, as well as the limitations and scope of the study.

Chapter 2 of this thesis includes an in-depth review of past related research, focusing on the introduction to radar and its fundamental operational principles. This chapter also presents a comprehensive knowledge pertaining to 4D radar imaging. The topics covered include FMCW MIMO-Based radar imaging, the FMCW signal model, radar image resolution, and a comparative analysis between radar imaging and other sensors such as cameras and lidar. Furthermore, there are discussions regarding several types of deep learning methods, encompassing the topics of clustering and classification. Finally, this chapter will provide an overview of previous relevant literature that offers valuable recommendations for this study.

Chapter 3, known as research methodology presents a detailed flow chart of this research work. This chapter includes the use of tools, experimental setup and procedure, system parameters, hardware and software platforms, as well as data processing methodologies.

Chapter 4, which discusses results and discussion, will provide the overall outcomes of all the tests and experiments conducted through this research study. From the outcomes of the preliminary test, radar performance test, point cloud data processing, image data processing, point cloud data projection, point cloud data labelling, and the result of point cloud classification in the deep neural network. This chapter ends with the overall deep neural network model's performance.

Lastly, chapter 5 will include the research conclusion and future recommendations of this work.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

Since the beginning, radar technology has revolutionised our ability to understand and interact with our surroundings. With applications ranging from military surveillance to weather forecasting and autonomous vehicle navigation, radar systems have evolved continuously to meet diverse and complex challenges. This literature review discusses a thorough review of key radar domain elements, looking into fundamental concepts such as the radar equation and 4D imaging capabilities. It also explores the complex domain of the signal model of FMCW MIMO radar signals, revealing their potential for high-resolution imaging. In addition, the review examines the crucial task of object clustering and classification within imaging radar scenarios and reviews the techniques utilised for precise target discrimination. Throughout this review, a synthesis of related works is provided, highlighting the important contributions that have propelled radar technology to its current significance.

2.2 RADAR IMAGING TECHNOLOGY

Radar is a technology that is widely used in our daily life that provides extensions of sensory detection that beyond human capabilities. Mentioned by Oloumi (2016), Radar utilized the radio frequency signals which transmit into open air to detect an object or target which used the reflected echo from the target to estimate the range and distance. In other word, it is a detection system used to determine range, velocity and angle of an object. Basically, most of the radar system used the same principles of transmitting electromagnetic wave to illuminate the desired object which then the

reflected signal is detected using the receiver and further processed to obtain information (Peebles, 1998).

There are many types of radar exist around the world as it used for different types of application. Radar is mostly divided according to their characteristic of the radar systems itself. Few types of radar that is widely used around the globe such as monostatic radar, bistatic radar, MIMO radar, etc. Hence, these types of radar have become the basic technique implemented in developing radar system for many types of application. As different types of radar have its own pros and cons, it is usually why some cases of application used different techniques of radar system which are more suitable to the situation of the application. In the context of this work, we will be focusing on radar imaging and its technology advancement.

Radar imaging systems differ from typical radar systems in that they provide spatial information, allowing for precise visualisation of targets rather than just range and velocity data (Sun and Zhang, 2021). The application of this technology is common in several domains, such as remote sensing, environmental monitoring, medical imaging, and the automotive industry. In the context of remote sensing application, radar imaging technology used is commonly known as Synthetic Aperture Radar (SAR). SAR imaging is an advanced technology used in remote sensing to obtain detailed images of the land areas using radar signals from satellites, drones or aircraft (Drummer, 2016). Mentioned by Chan and Koo (2008), SAR is a microwave-based imaging radar that can penetrate clouds, rain, and trees to capture high-resolution images of the ground surface under any weather circumstances.

The primary characteristics of SAR imaging is that it's able to generate images with high spatial resolution, enabling precise differentiation of geography and landscape. Further mentioned by Drummer (2016), SAR does this by utilising a method known as synthetic aperture, which effectively enlarges the antenna size by shifting it along the flight trajectory while collecting data. This can be illustrated in the figure 2.1 below. The process of synthesising a very long antenna involves utilising

the forward motion of an existing antenna. A signal wave pulse is transmitted at every point, and resulting echoes received by the receiver and stored in an ‘echo store’. Each echo reflected from the ground exhibits a distinct signature in terms of Doppler frequency fluctuation. SAR processing includes the synchronisation of Doppler frequency fluctuations and the subsequent demodulation procedure, which involves modifying the frequency variation in the return echoes originating from each ground point (Chan and Koo, 2008).

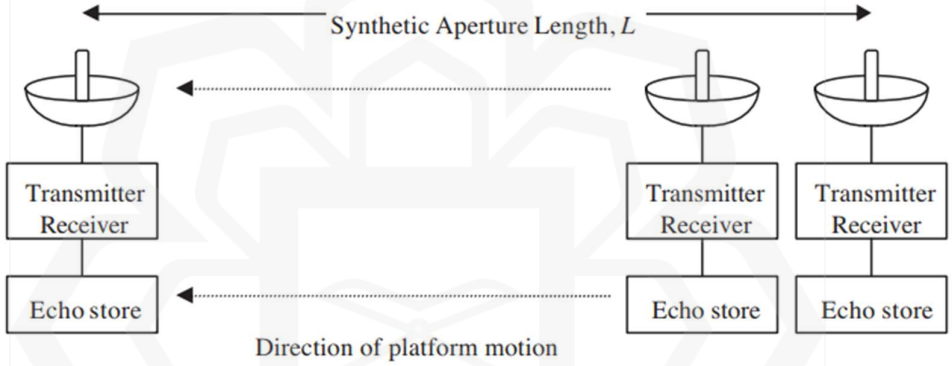


Figure 2.1 Synthetic Aperture Length (Drummer, 2016)

Another notable innovation in radar technology nowadays is ground-penetrating radar (GPR). GPR is a geophysical technique that uses radar pulses to create images of the underlying layers of the Earth's surface (Goodman and Piro, 2013). Further mentioned by Goodman and Piro (2013), this technology is extensively utilised in several domains, like archaeology, geology, environmental studies, and engineering. Hence, it is crucial to develop such technology to improve the imaging and analysis of prehistoric and historic sites, with the aim of identifying and mapping archaeological artefacts and optimising excavation strategies (Zhao et al., 2015). Geologists and environmental scientists use it to identify alterations in material characteristics, determine the presence of underground water, and oversee soil pollution. Moreover, engineers and researchers employ GPR to conduct infrastructure

inspections, which involve evaluating the state of roads, bridges, and buildings as well as identifying the positions of subsurface utilities such as pipelines and cables (Rasol et al., 2022)(Ékes and Neducza, 2012).

GPR operates by emitting high-frequency radio waves into the ground using an antenna. When these waves come across various subsurface materials, they bounce back to the surface and are detected by a receiver. The duration of the radar waves' round trip to the receiver is measured and utilised to provide a visual representation of the underground layers. The extent to which the radar waves may penetrate and the clarity of the image are determined by the frequency of the waves. Discussed by Everett (2013). Higher frequencies result in improved clarity but limited penetration, whilst lower frequencies allow for deeper penetration but less clarity. Operators can ascertain the location and composition of subsurface features by analysing the signals that are returned.

After review of the uses and features of GPR and SAR, it is understood that radar imaging technology is essential for both remote sensing and subsurface investigation. The adaptability of radar systems can be seen by the use of GPR in geological and archaeological research and SAR's capacity to produce high-resolution images of the Earth's surface from orbit. Both methods work on the fundamentals of electromagnetic wave propagation and reflection, despite their different applications. Beyond these fields, the automotive industry has also benefited greatly from the widespread use of radar technology. Advanced driver assistance systems (ADAS) and autonomous driving technologies rely more and more on automotive radar imaging, which uses the same fundamental radar concepts. Automotive radar systems are essential for improving vehicle safety and navigation because they make use of radar's capacity to detect and estimate the distance, speed, and angle of objects. To reflect back on the aim of this study, automotive radar imaging will be discussed in the next section

2.2.1 Automotive Radar Imaging

The number of projects done in recent years involving Automotive Radar Imaging is growing drastically. This is due to the demands in the automotive industry to move towards Adaptive driver-assistance systems (ADAS). Shared by Hasch (2015), the authors illustrate a few of the driver assistance features that will likely be available in the year 2014 and above as shown in figure 2.2. Radar sensors have become an integral part of ADAS, enabling crucial functionalities such as object detection, collision avoidance, adaptive cruise control and many more.

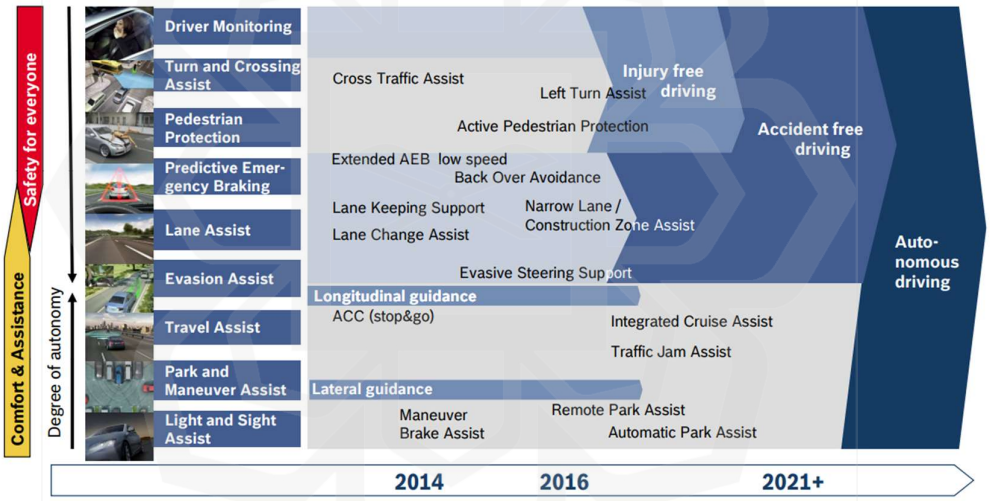


Figure 2.2 Roadmap of Driver Assistance Functions using Radar over Years (Hasch, 2015)

The widespread adoption of radar technology in the automotive industry highlights its reliability, robustness, and effectiveness in enhancing vehicle safety. The number of road accidents and incidents shows a very high rate as the number of vehicles on the road has increased. Based on world road accident statistic in 2022, 1.3 million people have died because of road accidents around the world, and about 20 to 50 million have

suffered road accident injuries (World Health Organization: WHO, 2022). Hence, to improve the safety of road users, researchers and manufacturers begin to support the automotive radar imaging technology (Waldschmidt et al., 2021) in autonomous vehicles, where the technology offers much higher reliability in providing high accuracy in object detection, road lane warning, blind spot detection, and an early potential crash warning system (Gao et al., 2019).

Automotive radar systems operate by transmitting radio waves, typically in the millimeter-wave frequency range (24 GHz to 77 GHz), and receiving the waves that reflect back from objects. The radar unit sends out a continuous wave or a series of pulses, and by measuring the time it takes for the reflected waves to return, the system calculates the distance to the object. Additionally, by analyzing the frequency shift of the reflected waves (Doppler effect), the radar can determine the speed of the object. Modern systems often use Frequency Modulated Continuous Wave (FMCW) technology, which provides high resolution and accuracy. Multiple Input Multiple Output (MIMO) technology is also employed to enhance the spatial resolution and provide detailed imaging of the surroundings.

One of the key for the advancement of radar for autonomous vehicles is provided by Complementary Metal-Oxide-Semiconductor (CMOS) technology, which has been integrated with automotive radar imaging and offers a much lower cost of radar on-chip and antenna on-chip mass production (Bilik et al., 2019). Moreover, CMOS technology plays an important role in the radar system to provide low power consumption and increased data processing capacity (Ragonese et al., 2022). This has led radar manufacturers to improve their radar chipset solutions for the automotive market. Today, radar imaging systems implement the cascading of multiple chips to enhance the sensor technology. Based on (Bilik et al., 2018), compared to a single chip, a radar system cascade with multiple chips allows the radar to operate with multiple transmit and receive antennas, which improves object detection.

The primary advantages of automotive radar imaging is its ability to operate effectively in various weather conditions. Unlike optical systems such as cameras and LIDAR, radar is not significantly affected by rain, fog, or dust, making it reliable in poor visibility conditions. Additionally, radar can accurately measure the speed and distance of moving objects, which is crucial for applications like adaptive cruise control and collision avoidance. The robustness and reliability of radar systems make them a vital component in ensuring the safety and efficiency of modern vehicles, contributing to enhanced driver assistance and paving the way for autonomous driving technologies.

Despite its numerous advantages, automotive radar imaging does have some limitations. One challenge is the potential for interference, both from other radar-equipped vehicles and from external sources, which can affect the accuracy and reliability of the radar signals. Additionally, while radar excels at detecting metallic objects, it can sometimes struggle with non-metallic or low-reflectivity objects, such as pedestrians or certain types of obstacles. Another limitation is the resolution; while millimeter-wave radar offers good resolution, it still lags behind the detailed imaging capabilities of LIDAR systems. These limitations necessitate the integration of radar with other sensor technologies to create a comprehensive and reliable perception system for vehicles.

It is evident that as the demands for safety, accuracy, and reliability continue to grow, more advanced radar imaging technologies are required. Among these advancements, Frequency Modulated Continuous Wave (FMCW) and Multiple Input Multiple Output (MIMO) radar systems stand out for their ability to enhance resolution and accuracy. These technologies enable the creation of 4D radar imaging, which provides detailed information about an object's range, velocity, angle, and elevation. By integrating FMCW and MIMO techniques, radar systems can achieve unprecedented levels of performance, making them indispensable for next-generation automotive applications.

2.3 MILLIMETER-WAVE FMCW MIMO-BASED 4D RADAR IMAGING

2.3.1 FMCW MIMO Millimeter-Wave Radar

4D radar imaging is based on millimetre-wave (mmwave) radar, which is a radar system that uses short-wavelength electromagnetic waves for the transmission and reception of signals. The transmit signal's wavelength is in the millimetre range, giving it the term “millimetre wave” as shown in figure 2.3 (Lawrence et al., 2017). Within the electromagnetic spectrum, this particular wavelength is classified as short. Based on Jankiraman (2018), modern radar technology utilises FMCW mmwave radar, which involves the continuous propagation of a frequency-modulated signal over time. This enables the precise determination of range, velocity, and angle. The radar transmitter emits a signal with a short wavelength in a specified direction. When this signal reaches an object, its surface reflects the electromagnetic wave back. The range, velocity, and angle of objects can be determined from the reflected signal obtained by the radar receiver system.

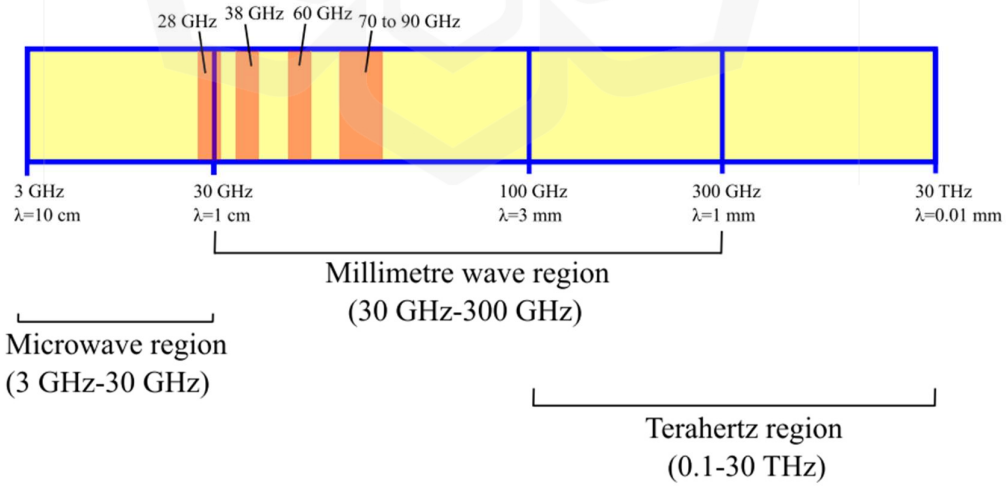


Figure 2.3 Millimetre Wave Signal (Lawrence et al., 2017)

The utilisation of mmwave radar technology gives an important advantage to the FMCW system, since it enables a significant reduction in the size of its components. Mentioned by Vavriv et al. (2015), the efficacy of mmwave radar can be attributed to the advantages offered by the millimetre wave spectrum. Mmwave radar show a compact design and improved mobility due to their small dimensions and reduced weight. Significantly decreased antenna size and weight can be achieved at millimetre waves, while maintaining equivalent antenna beam width and gain. Therefore, due to the relatively short wavelength of the signal, the antenna utilised for mmwave radar is compact in size. Moreover, another significant benefit of mmwave radar is by integrating it with MIMO system to significantly enhance the resolution and accuracy of the radar system (Venon et al., 2022). A signal with a shorter wavelength enhanced spatial resolution and the capability to identify objects in motion as small as a fraction of a millimeter. The utilisation of mmwave radar is deemed appropriate for applications that necessitate a remarkably precise level of detection and range capabilities. Hence, based on the given enhancement of mmwave radar with FMCW and MIMO technology, modern radar has become state-of-art for autonomous radar applications especially in achieving 4D imaging.

With the presence of MIMO radar system, this enables the radar to obtain wide field of view (fov) respect to azimuth and elevation angle. In conventional radar systems, the extent of azimuth angle coverage has always been constrained by the physical aperture of a single radar antenna (Kees et al., 1995). This implies that the detection and tracking capabilities of a single antenna are constrained within a finite angular range. MIMO radar systems employ a multiple of transmit and receive antenna elements, deliberately positioned to effectively encompass a wider angular range (Waldschmidt et al., 2021) and higher azimuth angle resolution (Kim et al., 2020). Through the intelligent processing of signals emanating from numerous antennas, MIMO systems have the capability to efficiently survey a wider azimuth angle, thereby offering a more comprehensive perspective of the surrounding environment.

Based on the discussion by Xu (2022), the authors illustrate the comparison between 4 receiver antennas and 8 receiver antennas in object detection accuracy. As

discussed above, MIMO radar improves object detection by enhancing angular resolution of radar imaging, and this is illustrated as shown in figure 2.4. Based on the angle-fft plot in the figure, when a radar detects an object, the amplitude will be high, which suggests that object is present. It was seen that the four receiver antennas failed to distinguish the object's presence at both -10° and +10°. As compared to 8 receiver antennas, the object can be easily distinguished from both locations, suggesting that a higher number of antennas could enhance object detection accuracy giving by the high angular resolution.

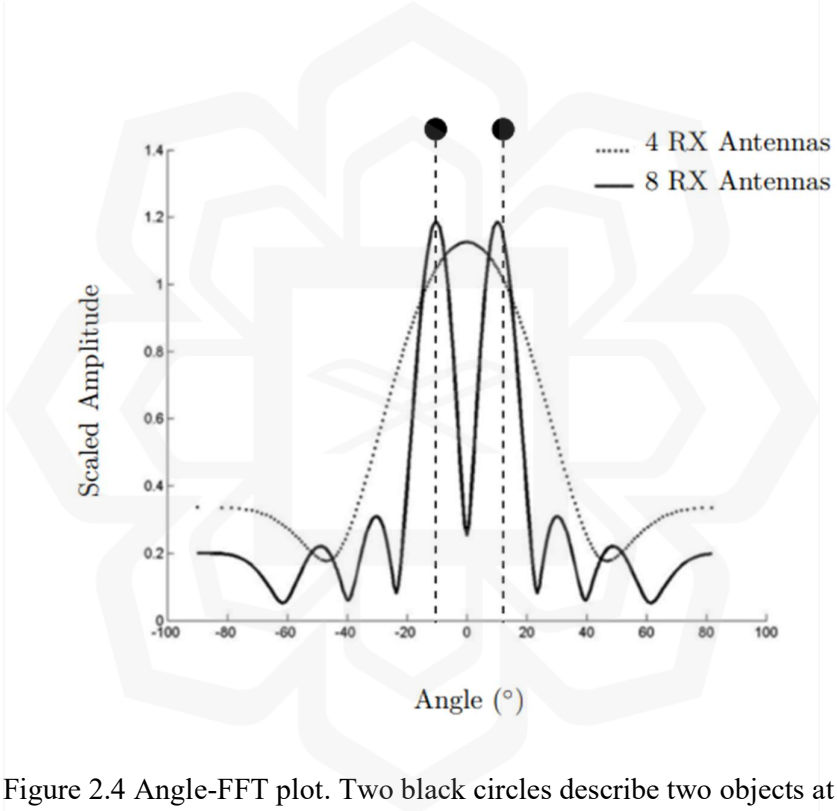


Figure 2.4 Angle-FFT plot. Two black circles describe two objects at ±10◦ individually (Xu, 2022)

Until today, FMCW MIMO radar has proven to be a reliable radar system, and the system is widely integrated into automotive radar applications. MIMO systems are used in radar imaging technology in order to improve angular resolution and accuracy, slow-moving target detection, and the percentage of object detection (Waldschmidt et al., 2021; Bilik et al., 2018). For this purpose, the presence of this MIMO system led

to high-resolution 4D Radar imaging, resulting in a denser point cloud output (Cheng et al., 2021). The concept of 4D radar imaging is to provide high resolution of azimuth, elevation, range, and Doppler information of a target (Sun & Zhang, 2021). However, to achieve highly advanced automated driving, a single sensor system is insufficient with today's radar technology. Based on Ragonese et al. (2022), ADAS implementation on automated vehicles usually combined several sensor technologies known as Lidar, Camera, and Ultrasound. Even so, radar still plays a vital role in automated driving, especially with 4D radar imaging systems enabling high accuracy of distance measurement, doppler velocity, and independence in any weather conditions as well as lighting exposure (Waldschmidt et al., 2021).

Furthermore, FMCW MIMO mmwave radar technology also enables radar imaging to operate in different operational modes in term of its range. The operational modes of these systems typically consist of three main categories: short range, medium range, and long range Stateczny et al., (2019). Every mode provides unique benefits customised to meet individual application needs, spanning from nearby object identification to distant object detection. Further discussed by Xu (2022), long range mode radar is employed for the purpose of measuring the distance and velocity of adjacent cars, medium range mode radar is utilised for a broader visual range field of view, and short-range mode radar is employed for determining the location of objects in close proximity to the vehicle. Different range mode of the radar is influenced by the characteristics of the antenna beam pattern. As figure 2.5 below, it illustrates the comparison between short range mode radar antenna beam pattern and the long-range mode radar beam pattern.

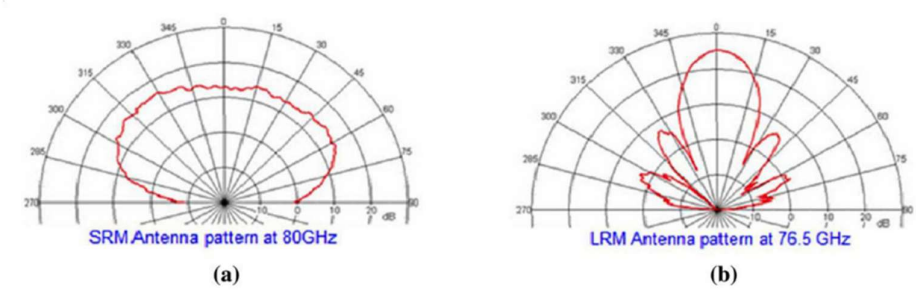


Figure 2.5 Antenna pattern of (a) Short Range Mode Radar (b) Long Range Mode Radar (Lin et al., 2016)

Short to medium range mode radar operate in 77 GHz frequency is the main focus of this work. Observe from the plot of the antenna pattern for short range mode radar, can be study that the radar system functions by utilising a narrower beam shape. The narrowed beam enables enhanced resolution and accuracy in identifying things in close proximity. The narrower beam pattern allows for the identification of adjacent objects and offers precise spatial data regarding their positions and motions. The utilisation of short-range mode can be found in several applications, including parking assistance, pedestrian identification, and collision avoidance systems, where the ability to accurately identify small objects within a limited distance. The choice of this range mode radar which is between short to medium range will be further justify and discussed in methodology section in Chapter 3.

The increasing popularity of autonomous radar imaging has led to a growing inclination towards the implementation of radar-based monitoring systems. As explained by prior research Skvortsov et al. (2012), this approach offers limited advantages, such as wide area coverage. The field of view of an automotive radar imaging system is typically capable of covering a broad area surrounding a vehicle, making it well-suited for monitoring applications. This capability enables the system to provide the extensive coverage desired for monitoring. Additionally, the authors also highlight the affordability and compactness of automobile radar imaging in the

current market. This characteristic makes it suitable for monitoring purposes, as it can be easily installed in a manner similar to that of a conventional surveillance camera.

In summary, the advancement of automotive radar imaging today is based on integration with cascading multiple chips, allowing the radar sensor to operate with the FMCW and MIMO system and enhancing the angular resolution for better object detection. Commercial automotive radar imaging opens new doors to monitoring applications, which in this work is to adopt automotive radar imaging as part of a road crossing monitoring system. The next section will touch more details on the FMCW signal model and radar resolution. It will also discuss the comparison between existing sensors on the market widely used for monitoring applications.

2.3.2 FMCW Signal Model

FMCW MIMO mmwave radar allows the advancement of radar imaging in terms of its object detection accuracy and angular resolution. Based on this work, linear FMCW radar system is used, hence, in this section, FMCW signal model will be briefly discussed. FMCW radar transmit linear and continuous signal known as chirp signal shown in figure 2.6. Then, the receiver antennas received signal from the target which is also linear and continuous. The instantaneous frequency difference between the transmitted (Tx) frequency and received (Rx) frequency can be known by frequency modulation (FM) ranging. Hence, with the frequency modulation, the target range and doppler shift can de determine (Tong et al., 2015).

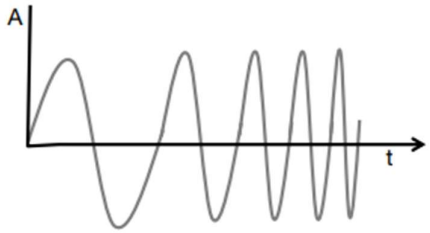


Figure 2.6 FMCW Chirp Signal with Amplitude as Function of Time

Based on (Li et al., 2021), the chirp signal can be characterized by the start frequency fc, time duration Tc, and signal bandwidth B. Hence, the sinusoidal FMCW chirp signal can be expressed as:

$$s(t) = \cos\left(2\pi f_c t + \pi k t^2\right) \quad (2.8)$$

with chirp rate as:

$$k = \frac{B}{T_c}$$

As the radar transmit signal, the signal is reflected by the target at distance r, to the receiver antenna, the signal is delayed and can be expresses as:

$$\tau = \frac{2\,r}{c} \quad (2.9)$$

where c is the speed of light.

The continuous chirp signal resulting in the presence of intermediate frequency (IF). IF signal also known as beat signal exist only when the time interval between the transmitted chirp and received chirp overlap each other. The sinusoidal IF signal can be expressed as

$$s_{IF}(t) = A\sin(2\pi f_{IF}\, t + \phi_{IF}) \quad (2.10)$$

Based on figure 2.7 can be observed that the f_{IF} remains constant as the distance between transmit and receive signal is fixed. This given that the IF frequency as:

$$f_{IF}(t) = f_T(t) - f_R(t) = k\,\tau \quad (2.11)$$

The phase of IF signal can be determined at the start time of the IF signal which given by:

$$\phi_{IF} = 2\pi f_c \tau \tag{2.12}$$

Given by the frequency and phase of the IF signal, hence, for an object with a distance r from the radar, sinusoidal IF signal can be expresses as:

$$s_{IF}(t) = A \sin(2\pi f_{IF}\, t + \phi_{IF}) \tag{2.13}$$

where IF frequency is:

$$f_{IF} = k\tau = \frac{2rk}{c} \tag{2.14}$$

and phase:

$$\phi_{IF} = 2\pi f_c \tau = \frac{4\pi r}{\lambda} \tag{2.15}$$

It is noted that the parameter λ, which represents the wavelength of the chirp signal, is denoted at the start frequency fc. From this, when the radar detects multiple objects that reflected the signal to the receiver, the IF signal is a linear combination of multiple signals with IF frequency and phase with respect to each object reflected signal.

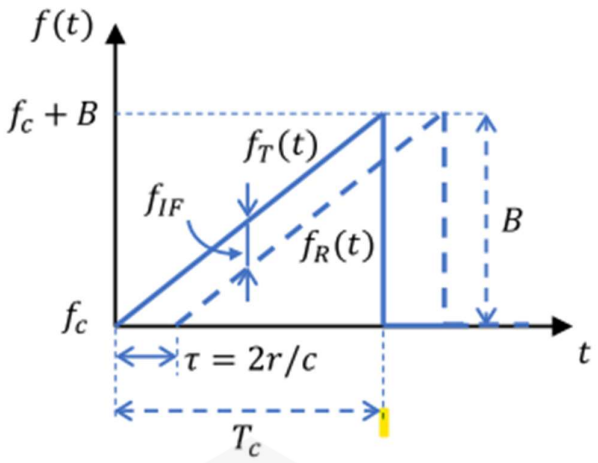


Figure 2.7 Constant Frequency IF Signal from the Transmitted and Received Chirp Signals (Li et al., 2021)

2.3.3 Resolution: Range, Velocity and Angular

Resolution is usually defined by the minimum separation between two or more targets or the smallest measurable interval that can be distinguish by a radar system. Range resolution can be defined as the smallest distance between a two or more objects that can be distinguish by the radar (Texas Instrument: TI, 2020). Recall back from equation (2.14), the frequency value is directly proportional to range, r, by performing range-FFT, the range resolution that is influenced by the frequency resolution is defined by the bandwidth of the transmitted signal, derived as:

$$\Delta r = \frac{c}{2\,k}\Delta f_c$$

$$\Delta r = \frac{c}{2\,kT_c} = \frac{c}{2B} \tag{2.16}$$

When multiple objects detected by the radar, each reflected chirp signal passed through the range-FFT and doppler-FFT which shows the peak of the individual reflecting objects spectrum. Hence, from the range-doppler fft, object with different

range and doppler velocity can be measured according to the angular frequency of discrete-time signal given by the condition of:

$$\Delta\omega \geq \frac{2\pi}{N} \tag{2.17}$$

where N is the number of chirps.

Doppler-velocity of the radar can be mathematically derived as:

$$\Delta\omega = \frac{4\pi\Delta v T_c}{\lambda}$$

$$\frac{2\pi}{N} = \frac{4\pi\Delta v T_c}{\lambda}$$

$$\Delta v = \frac{\lambda}{2NT_c} \tag{2.18}$$

for a frame period of $T_f = NT_c$, doppler-velocity resolution can be expressed as:

$$\Delta v = \frac{\lambda}{2T_f} \tag{2.19}$$

Angle estimation of an object by the radar is influenced by the direction of reflected signal received by the antenna at an angle which known as angle of arrival (AoA). When the reflected object wave received at two adjacent, it can be described as in figure 2.8.

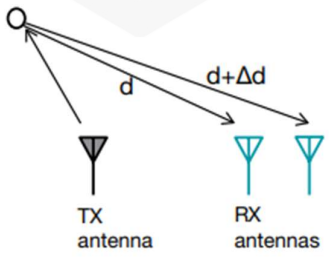


Figure 2.8 Reflected Object Wave Receive at Two Adjacent Antennas (TI, 2020)

For an object with AoA of θ, the discrete frequency can be express as:

$$\omega_{\theta 1} = \frac{2\pi d sin\theta}{\lambda} \quad (2.20)$$

while for the object with AoA of (θ + Δθ), the discrete frequency can be expressed by

$$\omega_{\theta 2} = \frac{2\pi \ (\sin(\theta+\Delta\theta))}{\lambda} \quad (2.21)$$

According to angular frequency discrete-time signal (2.17), angular resolution can be derived as:

$$\Delta\omega = \frac{2\pi d}{\lambda}(\sin(\theta + \Delta\theta)) - sin\theta)$$

$$\Delta\omega = \frac{2\pi d}{\lambda} cos\theta\Delta\theta$$

$$\frac{2\pi}{N} = \frac{2\pi d}{\lambda} cos\theta\Delta\theta$$

$$\Delta\theta = \frac{\lambda}{Ndcos\theta} \quad (2.22)$$

Assuming the spacing of the antenna is d = λ/2 and θ = 0, angle resolution is given by:

$$\Delta\theta = \frac{2}{N} \quad (2.23)$$

2.4 RADAR IMAGING VS CAMERA VS LIDAR

In this modern era, technologies have been emerged to provide range of solutions to problems related to sensors and safety. There are many types of commonly known sensors technology in the market such as Camera and LIDAR. These technologies have a long history from the first being design and develop as part to achieve safety enhancement same as Radar. In time, Camera and Lidar have offer high resolution of images which help in various application in industry. Most of these existing technologies are very promising and reliable however there are still certain limitations which causing another radar alternative need to be design and develop. This section will be focusing on comparing the technology of Camera, Lidar and Radar to understand the advantages and limitations that exist.

Camera is one of the commonly used technologies in many types of application such as security, surveillance, etc. In the older days, film camera has been used mostly of photography purposes, however, in time when there is a demand in industry, the technology has drastically improved which nowadays they are offering a very high-resolution of image and video. High-resolution camera images are very useful in surveillance system as the technology able to provide object detection, multiple target recognition, and face identification (Zhang et al., 2014). However, in terms of tracking object, camera is lacking behind compared to Lidar and Radar (Amaral et al., 2016). Moreover, camera range view is much limited compared to radar in terms of detecting objects by penetrating through walls or obstacles (Bassyouni, 2019). Mentioned by Krišto et al. (2020), RGB camera performance degrade when encounter with scenario that is low light such as night and adverse weather conditions resulting camera to become less robust than Lidar and Radar.

Next, Lidar is one of the advance technologies nowadays that have been used in many types of applications which involved smart and safety system. This is because, Lidar offer high-resolution imaging which provides dense point cloud (Wisultschew et al., 2021). Moreover, in terms of tracking, detecting, and classification, Lidar shows reliable performance (Zhang et al., 2020; Gao et al., 2018). However, compared to Radar, Lidar is much more expensive (Cai et al., 2020) which causes it not reliable for mass production. In addition, Lidar is not robust towards weather conditions as the technology depends on optics. Mentioned in (Govoni et al., 2015), Lidar performance in heavy rain and dense fog resulting in blind spot and image degradation.

Thirdly, researcher is moving forward towards developing radar imaging for autonomous technology system to cater the limitations of camera and lidar. According to Gao et al. (2019), millimetre wave radar is potentially to provide the ability to give high accurate location, doppler velocity and angular resolution of desired target. Most importantly, radar imaging is capable of detecting environment object such as pedestrian, bicycle and cars (Xu et al., 2018). Even so, in term of object recognition, there is still a lacking behind the recognition system due to the sparsity of the point cloud. Mentioned by Cho et al. (2014), the authors overcome the limitation of object

recognition by embedding few technologies together which is combining camera, lidar and radar. Hence, by combining the technology, they are able to improve the visual recognition of the object and tracking capability of the radar system. However, compared to both camera and lidar sensor technology, radar shows higher reliability and robustness.

Based on the discussion above, table 2.1 shows the summary of comparison between the three sensors technology used by today autonomous technology, Camera, Lidar and Radar.

Table 2.1 Comparison Between Camera, Lidar and Radar

	Camera	Lidar	Radar
Detection	Good	Medium	Good
Tracking	Poor	Medium	Goodr
Classification	Good	Medium	Poor
Speed Measurement	Poor	Medium	Good
Robustness	Poor (light and weather condition)	Good (light) Poor (weather condition)	Good (both light and weather condition)
Cost	Low	High	Medium

2.5 OBJECT CLUSTERING AND CLASSIFICATION USING DEEP LEARNING

2.5.1 Deep Learning

To understand the application of deep learning in radar imaging, it is important to grasp the foundations of machine learning, deep learning, and the distinction between supervised and unsupervised learning. Based on Janiesch et al. (2021), machine learning is a field within the subject matter of computers that focuses on the advancement of algorithms and models, with the aim of powering machines to acquire knowledge and engage in predictive or decision-making tasks without the need for explicit programming. The process includes the utilisation of statistical methodologies and algorithms for the purpose of examining and comprehending substantial quantities of data, enabling machines to recognise patterns, formulate predictions, and enhance their efficacy over the passage of time. The field has been a subject of ongoing research investigation, with researchers aimed at enhancing the precision and effectiveness of machine learning algorithms.

In addition, further discussed by Janiesch et al. (2021), deep learning, which is a subset of machine learning, has emerged as an effective methodology capable of extracting complex characteristics from data. Deep learning is an area of study within the domain of machine learning, defined by the utilisation of neural networks with several layers and parameters. The proposed methodology uses a logical structure of layers to effectively extract and manipulate features from the given dataset. The lower layers are responsible for acquiring elementary characteristics, while the higher layers are capable of acquiring complex characteristics that are derived from the lower layers. The utilisation of an ordered structure enables deep learning algorithms to effectively analyse and derive valuable insights from huge datasets, encompassing information obtained from diverse sources. Deep learning techniques encompass several models for instance a widely known neural network such as CNN and RNN. The application

of deep learning has been observed across several areas, encompassing classification and prediction tasks. This approach has demonstrated considerable potential in addressing complex issues and enhancing the precision of machine learning algorithms.

Moving forward with supervised learning, discussed by Alloghani et al. (2020), supervised learning refers to a machine learning technique that involves teaching a model by using labelled data, where the expected outcome is already known. The dataset provided by the user is split into two subsets: the training subset and the test subset. The purpose of this division is to enable the model to acquire knowledge from the training data, which it can then utilise to generate predictions or categorise fresh, previously unknown data. Decision tree, Naive Bayes, and Support Vector Machines are often employed algorithms in the field of supervised learning. Supervised learning algorithms possess the ability to deduce knowledge from a given dataset comprising labelled data, hence enabling them to make predictions. It is crucial to acknowledge that supervised learning needs the presence of instructions derived from prior experience or apparent trends within the dataset. This methodology is frequently employed in situations when the intended outcome is already established, and the objective is to generate exact predictions or classifications.

On the other hand, further discussed by Alloghani et al. (2020) Unsupervised learning refers to a machine learning methodology in which the model acquires knowledge of patterns and relationships within the data without the presence of labelled or planned output variables. The process entails the categorization or aggregation of comparable data elements according to their intrinsic qualities or attributes. Unsupervised learning methods are employed in situations when the objective is to uncover latent patterns, structures, or associations present within the dataset. These algorithms do not necessitate any prior information or explicit instruction regarding the desired conclusion, making them suitable for application to datasets without labelled data. Unsupervised learning methods are frequently employed in several applications, such as clustering, which involves the grouping of data points based on their similarities, and association mining, which entails the

identification of rules that describe associations between variables. Unsupervised learning algorithms are of the utmost significance in the process of acquiring information from extensive datasets, enabling the extraction of valuable insights and facilitating data-driven decision-making.

In short, both supervised and unsupervised learning approaches have their advantages and applications in radar imaging. Supervised learning is beneficial when labelled training data is available and can provide accurate object classification. Unsupervised learning, on the other hand, can uncover valuable insights and patterns in unlabelled radar data without relying on predefined categories. Figure 2.9 illustrates the categorization of machine learning classes, highlighting the distinctions between various forms of learning.

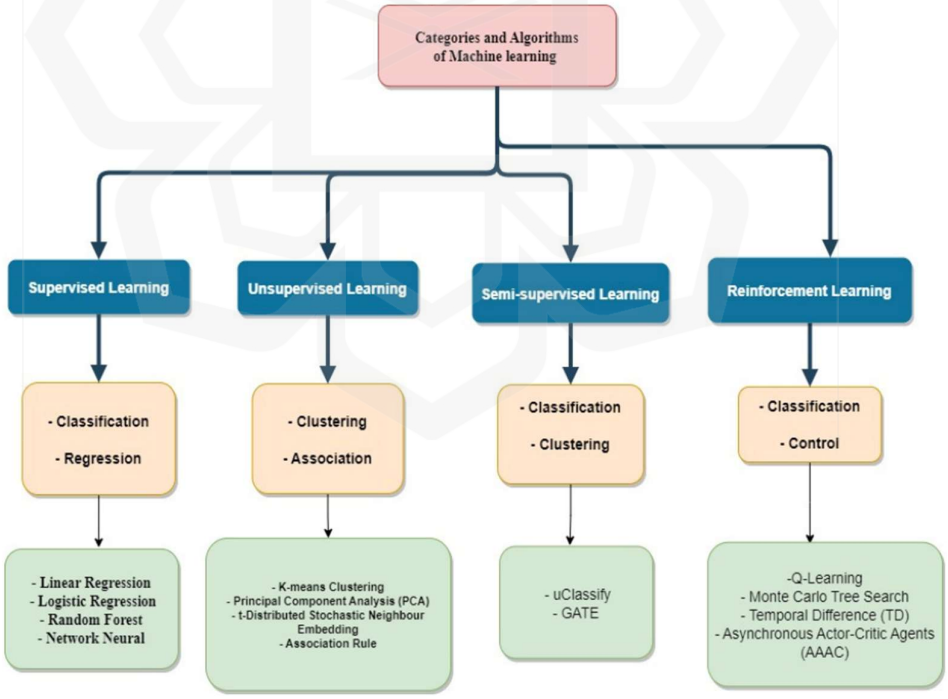


Figure 2.9 Supervised and Unsupervised Machine Learning (Taye, 2023)

Understanding the distinctions between supervised and unsupervised learning is essential in order to employ appropriate deep learning techniques for radar imaging tasks such as object clustering and classification.

2.5.2 Clustering for Point Cloud Data

Clustering is a key process in data analysis that plays a crucial role in obtaining significant insights from radar point cloud data. Moreover, clustering is a crucial procedure in this work, where radar point cloud processing involves grouping objects with similar properties based on their different traits. This section aims to thoroughly analyse and compare four well-known clustering methods: K-Means, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), and OPTICS (Ordering Points to Identify the Clustering Structure). This section begins with a brief overview of each methodological approach.

i. K-MEANS

K-means clustering is a clustering approach that is often used. The idea of the k-means clustering technique is to divide the radar data into a predetermined number of groups. Each data point is then given to the cluster that has the closest mean value (Jin et al., 2020). Further discussed by Reddy (2018), a common clustering algorithm called K-means divides a given dataset into K clusters. The initial centroids are chosen at random from a set of K representative points, and each data point is then assigned to the closest centroid using a proximity metric. After that, the centroids are updated by the algorithm using the mean of the data points allocated to each cluster. Until convergence, where the centroids stop changing or a relaxed convergence condition is satisfied, it iteratively repeats these two phases.

ii. DBSCAN

DBSCAN is another widely used clustering approach in the field of data processing. DBSCAN is based on the concept of density reachability. It groups together closely packed points based on a specified distance threshold, epsilon (eps) and minimum number of points (minPts) within that distance (Schubert, 2018). As mentioned by Reddy (2018), the DBSCAN algorithm is designed to detect areas of high point density within the data space and then create clusters by relating points located within these regions. The operational procedure involves the creation of a localised region surrounding each individual data point, followed by the identification of core spots that possess an adequate quantity of neighbouring points within a predetermined radius. DBSCAN generates clusters of various shapes by continuously expanding these core points to include their nearest neighbours (Reddy, 2018; Jin et al., 2020).

iii. HDBSCAN

HDBSCAN is a clustering algorithm that builds upon DBSCAN through the use of a hierarchical method (Campello et al., 2013). The process involves creating a hierarchical structure of clusters with varying densities and later identifying the most stable clusters from this structure. Based on Stewart and Al-Khassweneh (2022), the approach starts by calculating the core distance for the k nearest neighbours for every point in the dataset. This establishes a fundamental measure for hierarchy clustering based on mutual reachability distance (Malzer and Baum, 2020). Subsequently based on Stewart and Al-Khassaweneh (2022), a weighted graph is used to construct an extended minimum spanning tree, where the edges are determined by mutual reachability distances. The tree serves as the framework for building the hierarchical structure of the HDBSCAN algorithm. Afterwards, the HDBSCAN hierarchy is constructed, offering a thorough understanding of the links between clusters in the data. By utilising this hierarchy, major clusters are found, enabling a targeted study of the dataset's fundamental structure.

iv. OPTICS

OPTICS clustering is an algorithm that utilises the ordering process by storing the sequence in which the points are processed rather than assigning clusters. OPTICS and DBSCAN share the same eps and minPts hyperparameters, but the eps parameter is not considered essential (Plakalovic, 2023). In OPTICS, the only criteria that is maintained is the minPts condition, where a point is considered dense at a certain radius r if it has a minimum of minPts neighbours within that radius (Ankerst et al., 1999). The method's fundamental concept is to generate a reachability database that arranges points in a sorted manner according to their reachability distance (Plakalovic, 2023; Islam et al., 2021). Next, the method calculates the reachability distance between the two points and appends the point with the greatest reachability distance to the ordering list. The technique is to iterate over each subsequent point in the list until all the points have been processed (Schubert, 2018). Once the algorithm processes the points, it creates a plot that illustrates the reachability distance. The reachability distance plot shows the distance for each individual point in the dataset.

Gaining an understanding of the approach used by every method provides an initial understanding of how the algorithm of each clustering method operates. Therefore, it is essential to investigate the advantages and drawbacks of each method in the context of clustering radar point cloud data. In order to study and review which method suitable the most for this work, there are three main factors need to be considered which are the processing complexity, sparsity or noise of the point cloud data and the nature of the clusters.

i. Complexity for radar point cloud clustering processing

The K-means clustering method is widely used in several fields because of its simplicity and low computational complexity in handling clustering problems (Ikotun et al., 2023). K-means exhibits rapid clustering and excels in handling extensive datasets, giving it ideal for time-sensitive applications that require rapid clustering

(Jahwar and Abdulazeez, 2020). Nevertheless, K-means demands a predetermined number of clusters, which might provide a difficulty when dealing with radar point cloud data, as the number of clusters is not always known beforehand.

On the other hand, clustering methods that are based on density, such as DBSCAN, HDBSCAN, and OPTICS, are more suited for dealing with the intricacies of radar point cloud data. DBSCAN is capable of detecting clusters with irregular patterns and efficiently handling noise by classifying outliers. Nevertheless, the efficiency of the system could decrease when dealing with extensive datasets because of its O(n2) complexity (Bushra, 2021). HDBSCAN enhances the performance of DBSCAN by introducing a hierarchical clustering framework (Campello et al., 2013), which enhances its ability to handle different density levels and reduces its sensitivity to parameter configurations. However, the hierarchical structure of it can result in additional computing burden.

OPTICS enhances these capabilities by generating a reachability plot, which captures the clustering structure at various density levels. However, this comes at the expense of increased memory usage and a complexity of O(n2) due to the need to maintain a priority queue (Islam et al., 2021). The high requirement for memory may limit the real-time processing of extensive radar point cloud information. DBSCAN, HDBSCAN, and OPTICS are better suitable for radar point cloud clustering than K-means in dynamic and noisy situations due to their adaptation to shifting densities and robustness to noise.

Hence, while K-Means offers simplicity and low computational complexity, making it suitable for large datasets, it falls short in handling noise and complex cluster shapes. DBSCAN and HDBSCAN provide more robust clustering for noisy and varied-density data but at the cost of higher computational complexity, especially without efficient indexing. OPTICS, while offering detailed density-based cluster analysis, requires significant memory and computational resources. For radar point cloud data, the choice of algorithm depends on the specific requirements of cluster

robustness, shape complexity, and available computational resources. Leveraging the strengths of these density-based algorithms typically results in more accurate and effective clustering outcomes for complex radar data.

ii. Sparsity or noise of the radar point cloud data

In the context of clustering radar point cloud data characterized by sparsity or noise, various clustering algorithms exhibit distinct strengths and weaknesses. K-Means is notably sensitive to noise and outliers, as it assigns every point to a cluster, which can distort the cluster centroids and degrade clustering accuracy. Additionally, the requirement to specify the number of clusters in advance (K) poses a challenge, especially in scenarios where the cluster distribution is unknown (Jahwar and Abdulazeez, 2020). Moreover, K-Means assumes that clusters are spherical and of similar size, an assumption often violated in real-world radar data, leading to suboptimal performance.

On the other hand, DBSCAN is a density-based algorithm that excels in identifying clusters of arbitrary shape and handling noise effectively. DBSCAN's robustness to noise allows it to identify and exclude noise points, making it well-suited for radar point clouds with significant noise levels. It can discover clusters with arbitrary shapes, providing a significant advantage for complex radar data structures. Unlike K-Means, DBSCAN does not require the number of clusters to be specified in advance, instead relying on two parameters: ε (epsilon), the neighborhood radius, and MinPts, the minimum number of points needed to form a dense region. However, DBSCAN may struggle with datasets that contain clusters with varying densities, as a single ε value might not be suitable for all clusters within the data.

HDBSCAN improves upon DBSCAN by providing enhanced handling of clusters with varying densities and offering hierarchical clustering capabilities. HDBSCAN's ability to identify clusters with varying densities addresses one of DBSCAN's primary limitations. It generates a hierarchy of clusters, allowing for

flexible analysis at different levels of granularity. Additionally, HDBSCAN includes mechanisms for more automatic and adaptive parameter selection, reducing the need for extensive manual tuning. Similar to DBSCAN, HDBSCAN is robust to noise and can effectively exclude noise points from clusters, enhancing the clustering quality for low-quality radar data (Wang et al., 2021).

In addition, OPTICS creates an ordering of points that reflects the data's density-based clustering structure, which can be used to extract clusters with varying densities. Unlike K-Means, OPTICS does not force every point into a cluster, which helps in naturally excluding noise points. It also shares the ability to discover clusters of arbitrary shapes, similar to DBSCAN and HDBSCAN (Nanni and Pedreschi, 2006). However, while OPTICS does not require a global ε parameter, the selection of MinPts and other parameters still influences its performance and necessitates careful selection.

Density-based clustering algorithms such as DBSCAN, HDBSCAN, and OPTICS generally outperform K-Means for radar point cloud data, particularly in the presence of noise and sparsity. These algorithms' capabilities to handle noise, identify clusters of arbitrary shapes, and function without needing the number of clusters predefined make them more suitable for complex and low-quality radar data. K-Means, despite its computational efficiency, is less effective in such scenarios due to its sensitivity to noise, requirement for a predefined number of clusters, and assumption of spherical clusters. Leveraging the strengths of density-based clustering algorithms enables more accurate and robust clustering of radar point cloud data, thereby enhancing the overall quality of 4D radar imaging and analysis.

iii. Nature of the clusters (size, shape and density)

Radar point cloud data exhibits diverse characteristics, including varying shapes, sizes, and densities, which pose challenges for traditional clustering algorithms like K-Means and offer opportunities for density-based methods such as DBSCAN, HDBSCAN, and OPTICS.

K-Means operates under the assumption of spherical clusters with equal sizes (Ikotun et al., 2023), making it less suitable for radar point cloud data characterized by irregular shapes and varied densities. While K-Means offers a straightforward approach based on spatial proximity, its constraint on cluster shape and size may lead to suboptimal clustering results. This limitation becomes apparent in datasets where clusters exhibit uneven forms and densities, such as those found in radar data collected from road crossing areas. K-Means' reliance on predetermined cluster shapes and sizes renders it less effective in capturing the complexity inherent in radar point cloud data.

DBSCAN, on the other hand, offers flexibility in detecting clusters with arbitrary shapes and sizes (Nanni and Pedreschi, 2006). By defining clusters as dense regions separated by regions of low-density objects (Elbatta et al., 2012). DBSCAN can adapt to the diverse nature of radar point cloud data. However, DBSCAN's sensitivity to user-defined density parameters poses challenges, as selecting appropriate thresholds requires careful consideration. Additionally, DBSCAN may struggle with clusters of significantly different densities, as its global density field parameter (ε) might not effectively capture local density variations. Despite these limitations, DBSCAN's ability to handle noise, discover clusters of arbitrary shapes, and adapt to varying densities makes it a valuable tool for radar point cloud clustering.

HDBSCAN builds upon DBSCAN's density-based approach by introducing hierarchical clustering and density adaptation mechanisms. This enables HDBSCAN to address some of DBSCAN's limitations, particularly regarding clusters with varying densities. By self-adjusting its density field under different density conditions (Wang et al., 2021), HDBSCAN offers improved robustness to density variations in radar point cloud data. Additionally, HDBSCAN's hierarchical clustering capabilities provide insights into the data's clustering structure at multiple granularity levels, enhancing its utility in complex radar datasets.

OPTICS offers a nuanced approach to cluster detection by ordering points based on density reachability. Like DBSCAN, OPTICS can discover clusters with

arbitrary shapes and sizes, making it well-suited for radar point cloud data with diverse characteristics (Nanni and Pedreschi, 2006). Moreover, OPTICS addresses one of DBSCAN's major limitations by adapting to data with varying densities. This adaptability makes OPTICS widely used for clustering trajectories, where density variations are prevalent (Deng et al., 2015). However, OPTICS requires more memory due to its use of a priority queue, and its computational complexity may be higher compared to DBSCAN and K-Means.

While K-Means provides a simple and efficient approach to clustering radar point cloud data, its constraint on cluster shape and size limits its effectiveness. DBSCAN, HDBSCAN, and OPTICS offer more flexibility in handling the diverse characteristics of radar data, including arbitrary shapes, varied sizes, and densities. Despite their individual limitations, density-based clustering algorithms are better suited for capturing the complexity inherent in radar point cloud data, enhancing the quality of cluster analysis and interpretation.

In conclusion, for the scope of this work, utilizing DBSCAN proves to be a suitable choice for clustering the radar point cloud data. While it is acknowledged that DBSCAN may experience degradation in performance when handling extensive large datasets, the dataset utilized in this study is of a rational size, enabling DBSCAN to effectively manage it. Moreover, the sparsity of the radar point cloud falls within an optimal range, rendering DBSCAN a favorable option. Although HDBSCAN and OPTICS are alternative methods capable of handling sparsity, the high-dimensional nature of the radar utilized ensures that the sparsity level remains conducive for DBSCAN's application.

Furthermore, while it is true that DBSCAN may not accommodate varying densities of point clouds as adeptly as HDBSCAN and OPTICS, this limitation is mitigated by the specific task requirements of this study. The objective here is to classify certain objects while treating others as noise during the classification process.

Given this constraint, DBSCAN's capability to effectively cluster the point cloud data while focusing on the specified objects serves the purpose of this work adequately.

In summary, the rational dataset size, optimal sparsity level of the radar point cloud, and the task-specific nature of the classification requirements collectively affirm DBSCAN as a sufficient and appropriate choice for clustering in this study.

2.5.3 YOLO for Image Object Recognition

Object detection is a crucial task in computer vision, with applications ranging from autonomous vehicles to surveillance systems. Based on Padilla et al., (2021) and Olorunshola et al., (2023), numerous object identification methods have been developed, including Region-based Convolutional Neural Network (R-CNN), Fast Region-based Convolutional Neural Network (Fast R-CNN) and Single Shot Detector (SSD). Joseph Redmon et al., (2016) and his colleagues presented a pioneering method known as You Only Look Once (YOLO). YOLO simplifies the object detection process by considering it as a single regression problem, instead of using numerous steps like standard approaches.

YOLO reframes object detection as a single regression problem, predicting both bounding boxes and class probabilities directly from full images in one evaluation. This approach contrasts with traditional methods that rely on region proposals and multiple stages, thus offering a substantial boost in processing speed and making YOLO suitable for real-time applications. Figure 2.10 shows the releases timeline of YOLO versions.

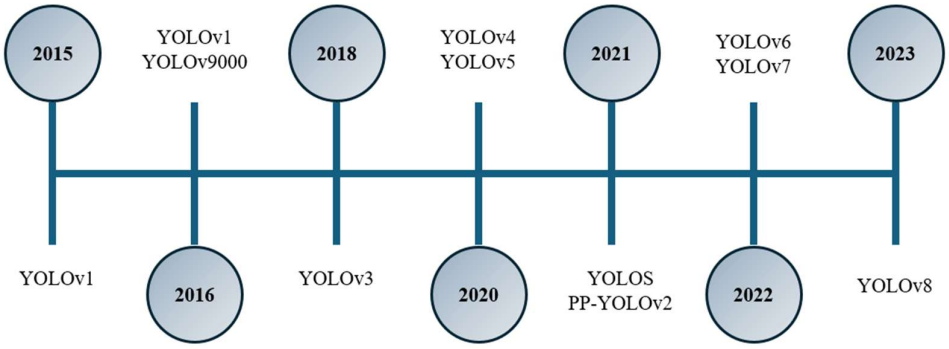


Fig. 2.10 YOLO timeline (Terven and Cordova, 2023).

The YOLO approach is distinguished by its ability to recognise objects in real-time process while maintains a consistently high degree of accuracy. As discussed by (Olorunshola et al., 2023; Redmon et al., 2016; Elavarasu and Govindaraju, 2024) This method generates a grid with dimensions of S×S based on the given image. An object is identified as a grid cell if its centre lies within the cell. Each grid cell will have expected confidence ratings and bounding boxes. If there is no item detected in that cell, the confidence ratings should be set to 0. Otherwise, the confidence score should be determined by calculating the intersection over union (IOU) between the projected box and the five predictions (x, y, w, h and confidence) that constitute each bounding box. The x and y image coordinates reflect the precise location of the centre of the box within the limits of the grid cell on the image while w and h represent the dimensions of the bounding box, specifically referring to width and height. The confidence prediction quantifies the IOU between the actual box and the anticipated box.

YOLO's architecture integrates a single neural network that processes the entire image in one pass. Described by Redmon (2016), the network design is based on the GoogLeNet model for image classification. The network consists of 24 convolutional layers, which are then followed by 2 fully connected layers. The architecture of YOLO is designed by the approach of using 1 × 1 reduction layers in combination with 3 × 3 convolutional layers rather than employing inception modules like GoogLeNet. This design allows YOLO to achieve remarkable speed while maintaining a good level of accuracy. The architecture of YOLO is shown in figure 2.11.

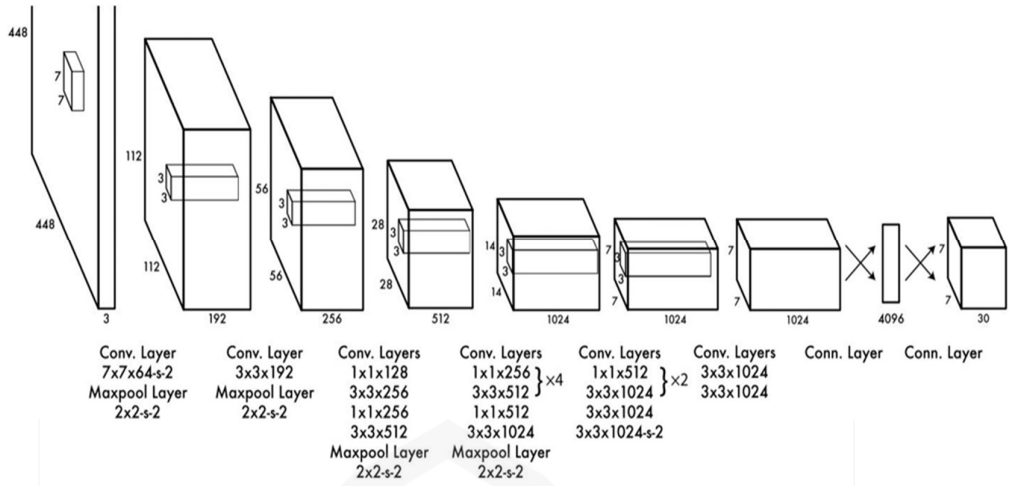


Fig. 4. YOLO network architecture that made up with Convolutional Neural Network (CNN) (Elavarasu and Govindaraju, 2024).

Next in this section, we will explore two primary variants of YOLO: YOLOv7 and YOLOv8. This version of YOLO is the most recent version released between the years 2022 and 2023.

i. YOLOv7

YOLOv7 significantly enhances the accuracy of object recognition in real-time, while simultaneously reducing the costs associated with inference. As demonstrated by the previous benchmarks, YOLOv7 surpasses other well-known object detectors by decreasing around 40% of the parameters and 50% of the compute needed for real-time object detections in modern settings (Olorunshola et al., 2023; Rangari et al., 2022). This enables it to make predictions more rapidly and with more precision in classification. Moreover, YOLOv7 achieves a frame rate of 30 FPS and a mean average precision (mAP) of 56.8% (Olorunshola et al., 2023; Elavarasu and Govindaraju, 2024). Additionally, it outperforms all other object detectors in the range of 5 to 160 FPS.

In addition, YOLOv7 has enhanced its layer aggregation networks (E-ELAN) to improve efficiency. It maintains the model's ideal structure and original features by implementing a compound scaling technique. E-ELAN employs expand, shuffle, and merge cardinality techniques to constantly improve the learning capability of the network while preserving the original progressive path (Wang et al., 2023). The E-ELAN system modifies the computational block design but leaves the architecture of the transition layer unmodified. E-ELAN not only preserves the original E-LAN design architecture, but also directs various groups of computational blocks to acquire a wider range of features.

YOLOv7 additionally includes model scaling to support concatenation-based models (Wang et al., 2023; Elavarasu and Govindaraju, 2024). The primary objective of model scaling is to modify some characteristics of the model and produce models of varying scales in order to accommodate varied requirements for inference speed. The proposed compound scaling method preserves the features and optimal structure of the model from its initial design.

ii. YOLOv8

The convolutional neural network employed by YOLOv8 is an advancement over previous YOLO algorithms. YOLOv8 surpasses YOLOv7 in terms of enhancement, mostly due to significant architectural enhancements and strategic modifications. By utilising cross-stage partial networks and CSPDarknet53 as the underlying architecture, the process of feature representation is enhanced, resulting in more advanced and distinct feature learning (Lakshmi and Sridhar, 2024). The CSPDarknet53 architecture, which forms the foundation of the system's structure, consists of 53 convolutional layers that are interconnected through cross-stage partial connections. These connections serve to improve the flow of information between the layers. The head of YOLOv8 consists of many convolutional layers followed by fully connected layers. This component is responsible for predicting the boundaries, object confidence scores, and class probabilities of recognised objects in an image.

The capacity of YOLOv8 to carry out multi-scaled object identification is another crucial feature. This is accomplished using a Feature Pyramid Network (FPN) (Elavarasu and Govindaraju, 2024; Lakshmi and Sridhar, 2024), which consists of layers with different specializations for finding objects in the image that are different sizes and scales. Moreover, FPN enhances its capacity to record data at various resolutions. As a result, YOLOv8 performs better in object detection tasks since it can recognize both big and tiny things with equal effectiveness.

One of the major enhancements in YOLOv8 is its anchor-free architecture. Like YOLOv7, it eliminates the need for predefined anchors, allowing for dynamic box prediction. Moreover, the anchor-free method decrease the quantity of prediction boxes and accelerate the process of non-maximum suppression for classification and regression tasks (Hussain, 2023). This change not only simplifies the model but also improves its flexibility in detecting objects of varying sizes.

iii. *Comparison YOLOv7 and YOLOv8*

The different architectural changes demonstrate the equilibrium between velocity, precision, and adaptability. YOLOv7 is designed to prioritise real-time efficiency by utilising a simplified and less complicated model. However, YOLOv8 is specifically designed to improve detection accuracy and expand the ability to detect objects at different scales. This is evident in its complex and advanced design. The trade-off in this case is between YOLOv7's priority of fast performance and YOLOv8's emphasis on improved accuracy and adaptability in recognising objects of different sizes.

Table 2.2 Table of Comparison between YOLOv7 and YOLOv8 (Lakshmi and Sridhar, 2024)

	Yolov7	Yolov8
Neural Network	Fully convolutional - ResNeXt	Fully convolutional
Backbone Network	CBS, E-ELAN, MP and SPPCSPC modules	CSP Darknet53, C2F modules
Loss Function	Focal Loss	DFL – cross entropy optimization , IoU – between predicted and bounding box
Neck	Features Pyramids	PAN – Path Aggregation Network
Training Techniques	Re-parameterized Convolution (RepConvN)	Swish Activation

The choice between YOLOv7 and YOLOv8 depends on the particular needs of the application. If the main priority is obtaining real-time performance, YOLOv7 would be the most appropriate option because of its optimised efficiency and reduced complexity. Alternatively, if the application requires a high level of precision and the capacity to operate extraordinarily well with items of different sizes, then YOLOv8 would be the more suitable choice. YOLOv8 is specifically developed to improve accuracy and the ability to detect things at several scales.

2.5.4 Radar Imaging Classification

Classification in radar imaging include the process of assigning predetermined names of class or categories to objects, depends on their features properties. Several approaches, such as deep learning algorithms, have been employed in the classification of radar images, resulting in great improvements in both accuracy and performance. There are several techniques for classification of radar images that is commonly used.

Deep neural networks demonstrate great accuracy in the classification of radar images. Mentioned by Ashtiani et al. (2022), these networks are made up of interconnected layers of artificial neurons that execute computational processes on input data. The input data is initially organised and afterwards processed through the neurons of the initial layer, which is then followed by intermediate (hidden) layers. The outcome of the classification process is observed at the output of the last layer. Further mentioned by the authors, deep neural networks provide the ability to learn complicated features and patterns from input data, thereby resulting with high-accuracy classification. Keras and TensorFlow are well-known platforms used for the implementation of deep learning models (Grattarola and Alippi, 2021). The integration of Keras with TensorFlow enhances the capabilities of this open-source machine learning framework, providing it with additional benefits (Abadi et al., 2016). TensorFlow functions as the underlying framework that offers efficient computational capabilities on both central processing units (CPUs) and graphics processing units (GPUs) for the purpose of training neural networks. On the other hand, Keras serves as the interface that provides a user-friendly and high-level application programming interface (API).

Next, Convolutional neural networks (CNN) have been frequently used in the field of classification of image. CNN have shown outstanding performance in various applications such as speech processing, image classification and face recognition due to its inherent capability of constructing spatial structures of data (Alzubaidi et al., 2021). Convolutional layers inside CNN are responsible for the application of filters

to the input data, hence enabling the capture of localised patterns and distinctive features. Further discussed by Alzubaidi et al. (2021), pooling layers are responsible for down sampling the feature maps, thereby decreasing the spatial dimensions while preserving crucial information. Fully linked layers establish connections between every neuron in the preceding layer and the subsequent layer, facilitating the neural network's ability to acquire complex representations and generate predictions.

Besides deep neural network and CNN, various other categorization methods have been used in the context of radar imaging. Support Vector Machines (SVMs) have been employed in the domain of radar image classification. SVMs are a classification algorithm designed to differentiate data points by creating hyperplanes in order to maximise the margin between distinct classes (Zhan and Yu, 2011; Wang et al., 2021). Linearly separable data is where support vector machines (SVMs) prove to be most advantageous. Moreover, Decision trees is another classification method that employ hierarchical tree structures to classify radar data by utilising feature thresholds. Discussed by Somvanshi et al. (2016), the model can be described as a hierarchical structure like a tree. In this structure, internal nodes correspond to features or attributes, branches represent decision rules, and leaf nodes provide class labels or predicted values. These methodologies provide the advantage of interpretability and provide the capability to handle both numerical and categorical variables.

It is important to mention that the selection of a classification methodology is dependent upon many elements, including the characteristics of the radar data, the complex nature of the classification objective, and the accessibility of labelled training data. Researchers persist in their works to investigate and advance novel algorithms and methodologies with the aim of enhancing the performance of radar image categorization and effectively tackling the unique obstacles encountered within this domain. In this work, we choose to work with deep neural network as the classification model to learn and train on point cloud due to its potential of handling complex data information. Moreover, deep neural network can be modified according to the need of this study to achieve the highest accuracy in classifying the radar point cloud data.

2.6 RELATED WORKS

Research in autonomous driving applications utilizing 77 GHz Imaging Radar has shown significant advancements, particularly in vehicle orientation estimation (Lim et al., 2022). Lim et al. proposed a method to estimate vehicle orientation using three regression techniques: Principal Component Analysis (PCA), Decision-Tree, and Convolutional Neural Network (CNN). Their study leveraged dense point cloud data from radar to accurately estimate orientation angles, with CNN demonstrating the best performance with a root mean square error (RSME) of 3.34°. This highlights the efficacy of CNNs in processing radar point clouds for precise angle estimation, setting a foundation for similar methodologies in object classification tasks.

In contrast to vehicle orientation, our project focuses on utilizing radar point cloud data to distinguish between various objects commonly found at road crossings, such as pedestrians, bicycles, cars, buses, and trucks. One challenge highlighted by Lim et al. is the higher point cloud sparsity for smaller objects like pedestrians and bicycles, potentially limiting CNN's effectiveness compared to larger objects like cars, buses, and trucks.

Sensor technology plays a crucial role in enhancing safety and monitoring in autonomous systems. Krišto et al. (2020) demonstrated the use of thermal cameras for pedestrian detection in adverse weather and low-light conditions, aiming to bolster border security against illegal activities. By complementing RGB cameras with thermal sensors, the authors achieved a high detection rate of up to 97.93% for human and non-human objects across varying weather conditions. Their adoption of YOLOv3 for object detection from thermal images underscores its potential applicability in our project, especially for scenarios requiring robust object classification under challenging environmental conditions.

Wen and Jo (2021) integrated LiDAR and RGB cameras to develop a lidar-camera-based sensor technology for 3D object detection. LiDAR provided point cloud

data for object localization, addressing challenges such as point cloud sparsity at long distances. Their approach enhanced accuracy by fusing image information from RGB cameras with point cloud data, demonstrating improved object detection reliability and system accuracy. This sensor fusion methodology aligns with our project's goal of leveraging multi-modal data to enhance object classification capabilities in dynamic environments.

Another significant area of research involves point cloud classification using (FMCW MIMO radar imaging. Wu et al. (2021) utilized 77GHz millimeter-wave MIMO radar to classify human gestures, employing the MMPointGNN algorithm developed by Gong et al. (2021). Their study achieved an accuracy of approximately 83% in gesture classification, highlighting the influence of radar point cloud data quality on classifier performance, particularly for small objects like humans. This underscores the importance of robust classification algorithms capable of handling sparse radar data effectively.

Kim et al. (2021) proposed an innovative approach using aggregated frames of point cloud images to enhance the accuracy of human activity classification with 77GHz FMCW MIMO radar. By leveraging 2D CNN and Deep Recurrent Neural Network (DRNN), they achieved a high classification accuracy of up to 97% for activities such as bowing, kicking, sitting, and standing. Their method of aggregating point cloud data frames effectively increased data density, facilitating more accurate classifier predictions. However, the study's limitation to an indoor environment where background point cloud data was filtered out suggests potential challenges in real-world applications where environmental complexities may affect classifier performance.

The reviewed literature showcases diverse approaches in radar and camera sensor fusion for object classification and gesture recognition, highlighting both advancements and challenges in the field. Lim et al. (2022) demonstrate the effectiveness of CNNs in processing dense radar point cloud data for precise vehicle

orientation estimation, laying groundwork for similar applications in object classification tasks. However, the higher sparsity of radar point clouds for smaller objects, as noted by Lim et al., presents a challenge in accurately distinguishing pedestrians and bicycles compared to larger vehicles.

Krišto et al. (2020) and Wen and Jo (2021) illustrate the importance of sensor fusion in enhancing detection accuracy under adverse conditions. The integration of thermal cameras by Krišto et al. (2020) and LiDAR by Wen and Jo (2021) with existing RGB camera systems demonstrates significant improvements in object detection reliability and robustness, aligning with the need for multi-modal data integration in our project.

The use of FMCW MIMO radar for gesture recognition, as explored by Wu et al. (2021) and Kim et al. (2021), highlights advancements in classification algorithms tailored for sparse radar point clouds. While MMPointGNN (Wu et al., 2021) and aggregated frame approaches (Kim et al., 2021) show promising results in gesture and activity classification, challenges remain in adapting these methods to real-world environments with varying data sparsity and environmental conditions.

Overall, these studies underscore the importance of robust sensor technologies and advanced classification algorithms in autonomous systems. The integration of radar and camera data fusion techniques, coupled with sophisticated deep learning models, holds promise for improving object classification accuracy and reliability across diverse operational scenarios.

Based on all of the related works discussed, can be observed that the are many methods have been used in order to classify point cloud data. Hence, all of the above methods and discussion are summarized in the table 2.2 below

Table 2.3 Table of Summary of Related Works

Author	Sensor type	Approaches	Contributions	Limitations
Lim et al., 2022	77 GHz FMCW MIMO Radar	This paper employed various regression procedures, such as principal component analysis (PCA), decision tree, and convolutional neural network (CNN), are utilised for the purpose of estimating the orientation angle.	This study shows the effectiveness of the proposed methodology, which utilises CNN framework, in accurately calculating the orientation angle. The results indicate a notably low root mean square error (RMSE) of 3.34°., highlighting the exceptional precision achieved by the method.	A limitation is seen in the employment of the PCA-based regression technique for the estimation of orientation angles. The underlying assumption of this methodology is the presence of a Gaussian distribution of data, which may not be applicable in every situation of the data. In circumstances where the data distribution is different from Gaussian, this assumption might result in poor regression performance.

Krišto et al., 2020	Thermal Camera	The authors employed various object detection models, namely Faster R-CNN, Cascade R-CNN, and YOLOv3, which were subsequently retrained on a dataset containing thermal images data taken from video sources.	This study explores the application of CNN models originally intended for RGB images to the task of person detection in thermal photos. Based on a comparative analysis of many CNN object detectors, it is evident that YOLOv3 stands out as a more efficient option while maintaining equivalent performance.	The research lacks a full analysis and comparison of different object detecting algorithms and only focuses on the result based on YOLOv3 for image object detection.
Wen and Jo, 2021	Lidar and Camera	The proposed method of this paper utilizes fully connected networks (FCNs). The research discusses the utilisation of two fully connected (FC) layers in order to modify the dimensions of the point-wise features that	This paper proposed an architecture for 3D multi-class object detection that the paper suggests combines camera and lidar data. The process extracts point-wise information from RGB images and fuses	The paper does not address the potential application or expansion of the approach to sensor modalities other than lidar. Due to the point cloud nature of both lidar and radar output, it is possible that the

		are retrieved from the lidar and camera data. This adjustment is necessary to ensure compatibility for fusion purposes. Following the fusion procedure, a fully connected (FC) layer is employed to further integrate the fused features and generate the final output.	them with the relevant point cloud data using a shared voxel-based backbone.	proposed method may not be suitable for the specific applications involving radar point cloud data.
Wu et al., 2021	77GHz FMCW MIMO Radar	This study introduces point cloud classification network, namely MMPointGNN, which is specifically designed for the categorization of millimetre wave radar point clouds. The neural network is employed for the purpose of classifying gestures performed by traffic	This study shows the experimental findings of four distinct gesture recognition tasks, namely stopping, turning right, turning left, and holding. The primary objective is to assess the efficiency of the suggested mmWave radar point cloud classification technique	The paper lacks a thorough assessment of the suggested MMPointGNN network approach. In order to evaluate the efficacy of these strategies, it is advantageous to use quantitative criteria such as accuracy, precision, and recall.

		police officers, and its training is conducted using a dataset specifically for traffic police gestures.	using MMPointGNN. The overall achievement of the classification achieves up to 83% of accuracy.	
Kim et al., 2021	77GHz FMCW MIMO Radar	This study proposes a classification methodology that integrates DRNN with 2D CNN.	This study introduces a fusion approach for classification, integrating DRNN with 2D CNN. DRNN is employed for the purpose of identifying time-varying patterns inside the point clouds. On the other hand, 2D CNN is utilised to extract distinctive characteristics from the point clouds, which are subsequently fed as input into the DRNN. The approach employed in this study results in a classification accuracy above 97%.	The experiment is done in an indoor environment which the background point cloud have been filtered out. Further investigation is required to assess the efficiency of the suggested methodology in real-world contexts and when applied to larger datasets.

2.7 SUMMARY

Based on the literature review from past related works, there are some factors that we can conclude. Firstly, in terms of sensor type used in past research, there have been several types of sensors discussed, which include 77 GHz FMCW MIMO radar, thermal cameras, RGB cameras, and lidar. All of the past related work has proven the reliability of this sensor for object classification in different real-life applications. However, considering the applications of this research which is object detection and classification at road crossing area, 77 GHz FMCW MIMO radar is the most suitable and reliable for this work.

Next, in terms of object classification methods, several methods have been discussed. It can be observed that in the application of image classification, most of the related works use CNN. This is understandable, as CNN provides a powerful algorithm for object classification using image data. Moreover, there is also the use of YOLO as part of the object classification method, and the foundation of the YOLO algorithm also adopts the CNN architecture. Hence, given the nature of this study, we proposed using YOLO to process image data to provide object classification information that includes the bounding box information and target class label.

Lastly, based on all the contributions and limitations discussed in the past related work, we proposed the use of sensor data fusion, which integrates radar and camera sensors, for the purpose of creating a customized dataset that can be used for training and testing deep neural network. This is because, based on past related work, the fusion of sensor technology could enhance object classification using deep neural networks. However, the integration between sensors is mostly done using lidar and cameras. Hence, in this work, we propose the fusion of radar and camera to create a customized dataset for development of deep neural network.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 INTRODUCTION

Hardware and software development of the surveillance system will be based on Smart Radar System Inc's products known as Retina-4F. Radar system have always been made up of hardware and software platforms hence the tools used are as mentioned earlier. In this chapter, there are three components that will be discussed which is the development approaches, details of hardware and software platforms as well as the details of parameter that is proposed to be used in this project.

3.1.1 Flow Chart

The project flow and development are based on research problem which become the motivation and purpose of this research as illustrated in figure 3.1. By understanding the root problem that need solutions, then the discussion of related work and existing technology will be analysed. This is important as the first step to start the research because all the idea and understanding will be developed here. Many resources and information were gathered to find articles, journals, and related papers regarding 4D FMCW MIMO radar, monitoring sensors technology, deep learning point cloud classification and application of 4D radar.

This workflow continues by hardware and software acquisition and installation. Initially, the radar needs to be acquired and familiarised first as well as the compatibility of the software as developing environment. The approaches here is to

adopt the 4D technology of the Retina-4F as the hardware platforms for data collection and Linux Ubuntu software environment as software platforms for data acquisition and processing. As both hardware and software platforms of Retina-4F and Linux have gone through the process of configuration and calibration, system validation will start by conducting a preliminary test.

If the system the found to have any problems, improvement will be made in line with preliminary test conducted. Throughout the experiment, all the observation and results will be collected to further analysed and evaluation of the system. As the initial phase have been satisfied, the work continues with actual experiment at a controlled and open-space area for radar performance test. This experiment is to understand the capability of the radar in detecting several targets and to study the behaviour of the point cloud of the targets. The data collected will be processed and manipulated for further analysed the performance of the radar imaging.

Then, an experiment will be conducted at a road crossing areas for data collection. The data collected at the experiment scene will be further processed for creation of custom dataset based on sensor data fusion method. This custom dataset created is use for training and testing deep learning algorithm and obtained the result of model performance for object classification at road crossing area. In line with working with object classification using deep learning based on the custom dataset, all the results and analysis obtained will be recorded and reported for journal and thesis writing. Lastly, as the final results obtained and satisfied, the work will be concluded and confirmed that the objectives have been achieved.

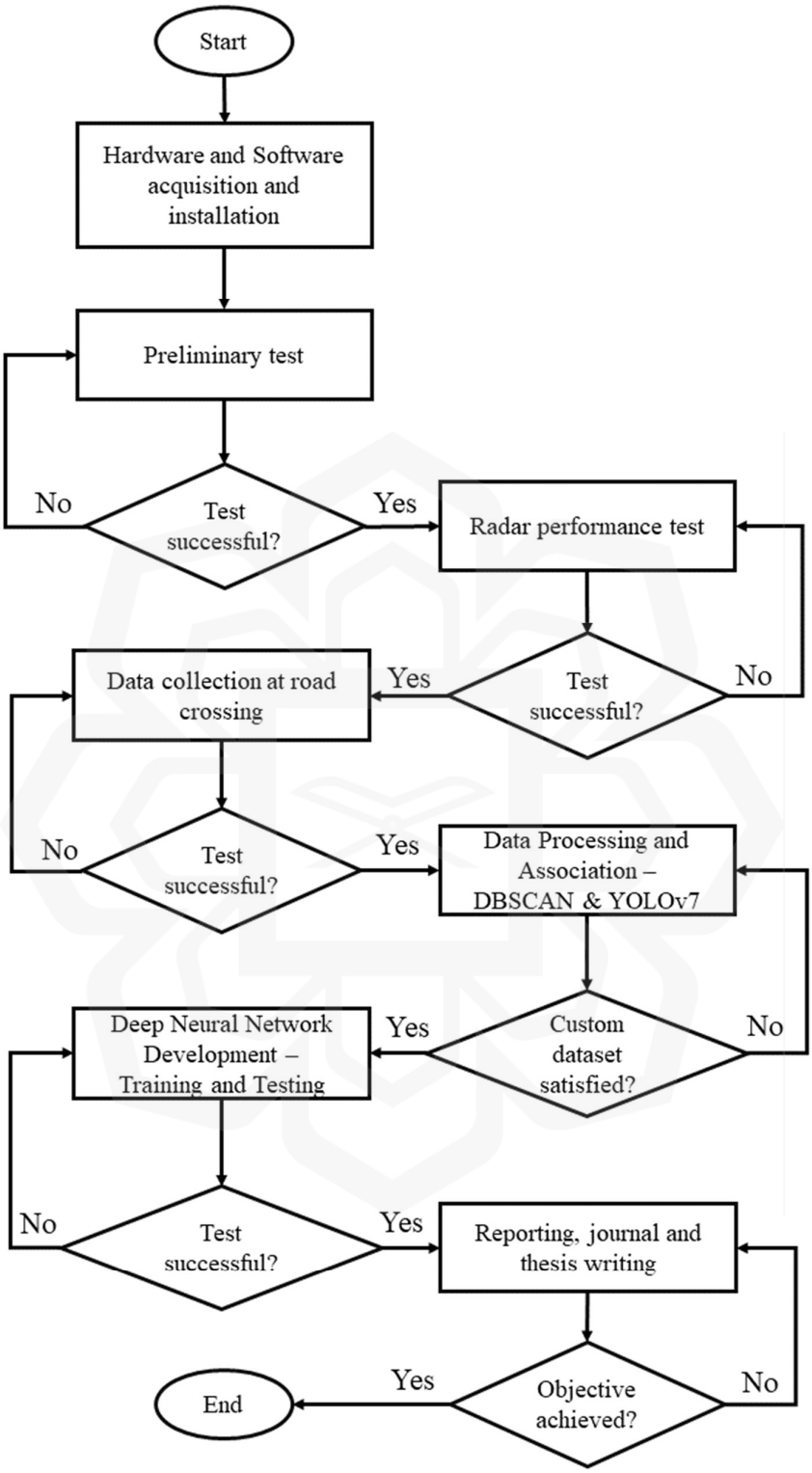


Figure 3.1 Flow Chart

3.2 RETINA-4F AND CAMERA

3.2.1 System Hardware Platform

The two sensors used in this project are 4D radar imaging known as Retina-4F and mono camera using ELP USB Camera.

i. Retina-4F Radar

4D radar used in this project is Retina-4F as presented in, developed by a company based in South Korea named Smart Radar Systems. The development of Retina-4F radar is powered by Texas Instruments' CMOS radar chipsets which are widely known for automotive and industrial purposes. The choice of using Retina-4F in this work is driven by several significant advantages that make it especially appropriate for precise detection, tracking, and classification of objects.

Primarily, in regard of its architecture, Retina-4F radar contained four chips embedded inside the hardware. Each chip contained onboard patch antennas, which have 3 transmitters (TX) and 4 receivers (RX) each. Hence, the maximum patch antennas contained inside the radar are 12 TX and 16 RX. The presence of multiple TX and RX allows the system to extract high-resolution data from the targeted object which is known as 4D (X, Y, Z, Velocity). Due to this, Retina-4F offers enhanced spatial resolution, enabling more accurate identification and positioning of objects. Precise detection of humans, cars, and objects is necessary in road crossing settings to ensure safety and efficiency.

Next, Retina-4F radar offer sufficient range operation for medium range object detection with a maximum detection range of 100m, 50m and 40m to detect car, bike, and person respectively. The current level of detection is adequate and suitable for this work environment, as it typically covers a road crossing area within a radius of 400

meters. The system's medium-range capabilities allow for early detection of objects, allowing sufficient time for processing and response.

In addition, Retina-4F could collect the precise distance and velocity of the target in the horizontal and vertical axis. Since the system uses high-resolution 4D technology that operates in the EHF frequency band, it can operate in any kind of weather condition such as fog, smoke, rain, snow, dust etc. The light condition of the surroundings will also not affect the operation of the radar since it can work well at night as well. Same with any common automotive radar, Retina-4F also transmits electromagnetic waves to detect targets. Hence, with the advanced chips presence along with high-resolution technology, Retina-4F manage to give the best results desired with 4D point cloud mapping. Figure 3.2 shows the Retina-4F Image Radar.

Figure 3.2 Retina-4F Imaging Radar

ii. ELP USB CAMERA

The utilisation of a mono camera, which is a single lens camera, for sensor fusion with radar offers greater benefits in comparison to utilising a multi-lens camera, like those commonly seen in smartphone. The main reason for this is the straightforwardness and precision of projecting point cloud data onto an image plane. Using a monocular camera simplifies the calibration procedure. There are two types of calibration used in this work known as Intrinsic Calibration and Extrinsic Calibration. Intrinsic

calibration, corrects the lens distortions and other camera-specific factors, only needs to be done once. Similarly, the process of extrinsic calibration, which determines the spatial relationship between the radar and the camera, is simpler when working with only one lens. The simplicity of calibration leads to increased precision and reliability in data fusion.

In this work, mono camera used is ELP USB Camera as presented in the figure 3.3. ELP USB camera can be easily adapted to any hardware medium due to its USB connection. As the camera sensor is mainly used to provide image data, any kind of mono camera can be adopted for this work as long as it can provide high-resolution and good quality image data. This camera in our work has 8 megapixels in HDMI format and is compatible with Ubuntu Linux operating systems.

Figure 3.3 ELP USB Mono-Camera

3.2.2 Radar System Parameter

The radar system will be using EHF frequency band in the range of 77 GHz based on the detailed product specification respectively with bandwidth of 3.8 GHz. Since this

project is aiming to be applied as monitoring system which need to cover an amount of area, medium range mode of the radar will be fully utilized. Hence, the parameter of the project research is proposed as table 3.1 below.

Four Texas Instruments (TI) AWR2243 chips have been integrated into this radar system, enabling Retina-4F to differentiate several objects without experiencing a significant loss in detection and localization accuracy. The radar uses four cascade chips, each of which has patch antennas with four transmitters (Tx) and three receivers (Rx), for a total of 12 transmitter antennas, 16 receiver antennas, and 192 virtual channels. This resulting in azimuth angle resolution of 2.0° and elevation angle resolution of 4.7°.

Table 3.1 Radar Operating System Parameter

Parameter	Value (Unit)
Frequency	77 (GHz)
Bandwidth	3.8 (GHz)
Range Resolution	0.5 (m)
Doppler Resolution	0.22 (km/h)
Azimuth Angle Resolution	2.0°
Azimuth FOV	100° (± 50°)
Elevation Angle Resolution	4.7°
Elevation FOV	24° (± 12°)
Point Cloud Output	max 6 114 per frame

3.2.3 System Software Platform

The software platform to be used in this project is Ubuntu by Linux. Ubuntu acts as the software environment that runs the Robotic Operating System (ROS) which can control the operation of radar and camera. ROS is a medium that can provide communication between components, interfaces and tools. As part of ROS system, there is a visualization tool known as ROS visualization (Rviz). Rviz visualises the sensor data from radar and camera in real-time which helps the user understand how the sensors are looking at the environment.

The output of the radar is mapped on a 3D map using Rviz. Rviz allows the data information received from the radar to be visualised in the form of point clouds. Moreover, Rviz provides interactive control to the user, through which the user can control the visualization environment of the 3D map. In addition, Rviz also can visualize real-time image data from camera. Figure 3.4 shows an example of Rviz tools that visualize point cloud data from 4D radar and video images from camera.

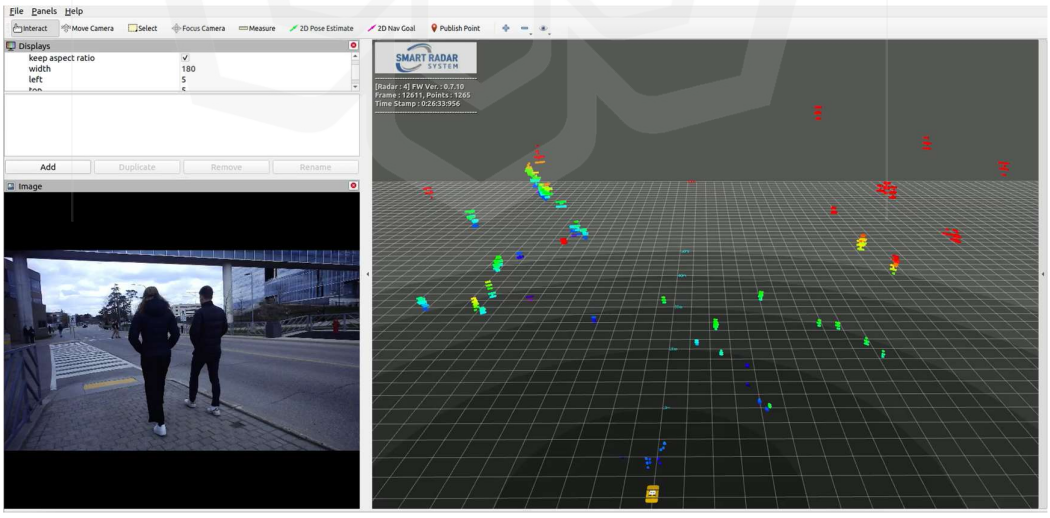


Figure 3.4 RVIZ Visualization Tool Environment

In this work, Rviz is mainly used only during the collection of the sensors' data to create a dataset to be used for classification. As it allows the user to visualize the data in real-time, it helps the user control the position and orientation of the camera and radar as closely as possible to ease the data fusion processing.

3.2.4 ROS Data Acquisition

There is a need to create a custom dataset for this project for the purpose of training and testing the deep neural network for object classification. Hence, the point cloud data and image data obtained by ROS need to be acquired and saved into our own computer system for further processing. More details explanations of the process involved will be discussed below.

In running an autonomous sensor, ROS allows users to communicate with the sensors to obtain data information for instance, images, object localization, doppler velocity, range measurement and power intensity. This information is established by ROS in the form of messages according to the sensor data formats. In this work, there are two nodes connected to ROS which are the radar node and the camera node. This node is controlled by ROS to run simultaneously according to the same timestamp. Once the nodes have been established by ROS, while running the nodes, ROS will establish the messages (data obtained from the sensors) under different channels known as ROS topics. ROS topics serve as channels of communication through which other ROS nodes are able to access their data.

Once the topics have been established, the user can obtain the data and save it to the computer system through another ROS tool known as ROSBag. ROSBag allows users to save data from all the nodes presence into one binary file with the ".bag' extension. Once this file has been saved into our computer system, it can be easily

replayed, processed, or analysed later. To process the data obtained from the sensors, we used python programming as the medium for data manipulation and processing.

3.3 EXPERIMENTAL SETUP AND WORKFLOW

In this section we will discuss the experimental setup and experimental workflow that involved in this work. Experimental setup focusing on the arrangement of the radar and camera within the measurement area at road crossing area. Experimental workflow discussing on the process involved starting from the data collected at road crossing area up until the labelling of point cloud data for customized dataset.

3.3.1 Experimental Setup for Data Collection

Pedestrians, cyclists, cars, buses, and trucks are the items of interest for classification in the experiment that involves gathering data at road crossings since they are the most frequent types of targets at these locations. The experimental setup for data collection at a road crossing is shown in figure 3.5, with the distance between the crossing and the radar set to be roughly 50m. All measurements were made outside at two different road crossings. Due to the radar's azimuth field of view (FOV), which is 100° ± 50°, and its location approximately 50 meters from the crossing, it is possible to observe and measure the motions of various objects, including cyclists and pedestrians, both before and after they cross the road. By having a much wider view than when the radar image is focused exclusively on the crossing, many more actions can be collected. For example, a car that is waiting for pedestrians to cross will have a very low, approaching zero, doppler velocity value. This information is useful to train the classifier to recognise that cars have doppler features with different ranges of doppler velocity.

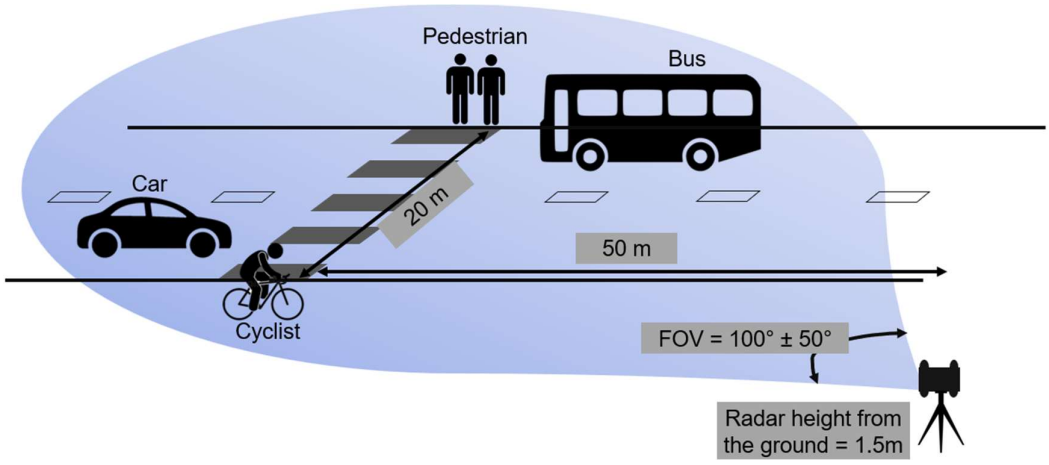


Figure 3.5 Experimental Setup for Data Collection at Road Crossing Area

Based on figure 3.6, both 4D Radar Imaging and a monocular camera are used for the measurement. Due to the extremely small and lacking dataset currently available for radar point clouds at road crossing areas, it is necessary to gather data before beginning to train a deep neural network for classification. This data must include the objects' features as well as their labels. The camera and radar are only combined for the deep neural network's training. The robotic operating system (ROS) environment in which Retina-4F operates makes data collection easier since it allows for simultaneous control over radar and camera nodes. As a result, the timestamp can be used to synchronise the point cloud data and image.

Figure 3.6 Radar and Camera Mounting Arrangement

3.3.2 Experimental Workflow for Point Cloud Dataset

Referring to figure 3.6, a very near mounting arrangement between the radar and camera is made to ease the data fusion process. In this project, the image data obtained from the camera is processed through the YOLOv7 object detection algorithm to produce bounding box predictions of objects present in the image. There are other types of information given by the YOLOv7 which includes the center point of the bounding box, height and width of the bounding box, type of class of the object and the confidence percentage of the class of the object. all of this information is crucial especially when we want to use it to label the point cloud of the target that situated inside the bounding box.

Meanwhile, radar data from Retina-4F is being processed using DBSCAN for point cloud clustering. DBSCAN helps in clustering the point cloud of the object as one cluster which will be useful when processing the data in python for labelling of the object. Based on the bounding box and cluster data gained from the processing, the next step is to associate both sensors' data using the data fusion method. The fusion of the sensors' data resulting a complete dataset of point clouds with class labels. This dataset is used to train a deep learning model for object classification, which is divided into two subsets known as the training set and the test set. The whole experimental workflow is presented by the block diagram in figure 3.7.

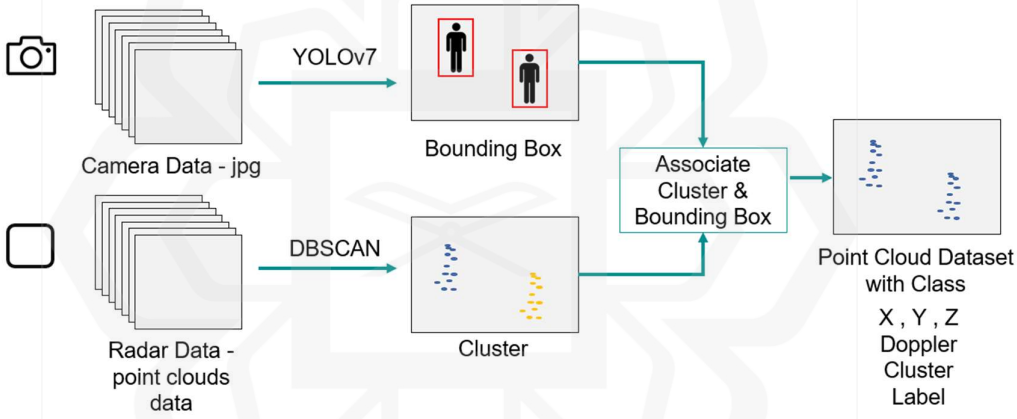


Figure 3.7 Experimental Workflow

3.4 DATA PROCESSING APPROACHES

To prepare the point cloud data and camera data collected from the sensors, both of the data need different processing approaches. Recall back from the experimental workflow block diagram, the fusion of the sensors data started after processing of the radar point cloud and image data. Point cloud data obtained from the radar sensor is passed through the DBSCAN for clustering. Image data obtained from mono camera

is pass to YOLOv7 for object classification and recognition. To classify the point cloud data after clustering, we proposed the use of point cloud projection onto an image. This part of the work involves the fusion of the camera data and radar data. The first step in this data fusion is to perform camera and radar calibration before taking measurements at the road crossings. Calibration is crucial as we want accurate position and orientation of both sensors with respect to the measurement scene. From here, the intrinsic and extrinsic parameters of the sensors can be acquired to help in the projection process; this calibration method is usually known as Extrinsic Calibration (Ji and Prokhorov, 2008; Ji et al., 2010; as cited in Oh et al., 2018). The significance of the calibration lies mainly in the misalignment between the camera data and radar data. More details explanation on the projection work will be further discussed below.

3.4.1 Clustering

Retina-4F point cloud data output has the following information: timestamp, Cartesian coordinates (X, Y, and Z), doppler, and Power Intensity. Based on this data information, it is known to be a feature of the point cloud and can be used for clustering. In lidar point cloud clustering, it is usually used for segmentation to distinguish different segments of the point cloud features, for instance, to segment the car point cloud between the tires and body of the car. However, in this work, the clustering process focuses on grouping the point cloud solely based on the type of objects. This means that the point cloud of each object belongs to different clusters. The segmentation of the point cloud is not necessary to ease the fusion process with image data. Moreover, the sparsity of the radar point cloud is still too much compared to the lidar point cloud, which makes it very complex for segmentation.

One of the commonly known methods for clustering non-linear data is DBSCAN. The algorithm used in DBSCAN is that it finds the nearest neighbour around each point of the point cloud and defines the cluster by connecting the points that are within the range. The region can be defined by adjusting DBSCAN parameters

known as Epsilon and Minimum points. The Epsilon is used to define the radius around a point, and the algorithm will use that radius to search for neighbouring points. Any points within the radius region will be part of the same cluster. Minimum Points refers to the minimum number of points that are within the radius of a core point to form a cluster. This is where the algorithm is used to define noise's presence in the point cloud. By identifying a core point and having a minimum number of points within the radius to form a cluster, if the minimum number of points is not satisfied, it will be considered noise.

The utilisation of DBSCAN in this study, as opposed to other clustering methods, is explained by three factors: the complexity of the clustering method, the sparsity of the point cloud data, and the nature of the point cloud cluster. DBSCAN and HDBSCAN offer greater benefits for implementation in this study compared to K-MEANS and OPTICS. However, in order to reduce computing complexity during the clustering process, DBSCAN has been selected for clustering the radar point cloud in this study. The reason for not using HDBSCAN in this work is the limited computer resources, as HDBSCAN demands intensive computational resources, particularly for large datasets. However, HDBSCAN is being looked into for future implementation in order to enhance the efficiency of the clustering process, as HDBSCAN is well-known for its higher efficiency compared to DBSCAN.

In this project, the optimal value for epsilon and minimum point chosen were 1.5 and 5 respectively. To most suitable epsilon and minimum points can be obtained by the process of trial and error. Next, after defining the epsilon and minimum point values, the point cloud data can be processed for clustering. For clustering, four features of the point cloud data are passed to DBSCAN: the Cartesian coordinates of the point (x, y, and z), and doppler velocity. DBSCAN's point cloud clustering produces a cluster of points; hence, it is insufficient to prepare the dataset for neural network classification algorithm training and testing. To aid the neural network in recognizing and predicting the type of class a point cloud belongs to, a further step in the process is required to provide a class label for the point cloud data. Therefore,

YOLOv7 is incorporated into the workflow to create a dataset of point clouds with class label.

3.4.2 YOLOv7 Object Detection

This work implements YOLOv7 as a component of image processing for object detection, aiming to obtain bounding box information about the targeted object. First introduced in 2022, YOLOv7 quickly gained popularity among users due to its high accuracy and efficiency in processing real-time data. The selection between YOLOv7 and YOLOv8 is dependent upon the specific requirements of the application. If the primary objective is to achieve real-time performance, YOLOv7 would be the most suitable choice due to its optimised efficiency and lower complexity. According to Olorunshola et al., (2023) and Boesch (2023), YOLOv7 can significantly reduce around 35% of the parameters and computation required for accurate and efficient object detection. Given its speed, accuracy, and low computational complexity, YOLOv7 is an ideal image object detection tool for this application.

To get the bounding box information for labelling the point cloud to its class, the image data collected by the mono-camera need to be processed with object detection algorithm. YOLOv7 model is pre-trained using the Microsoft COCO dataset (Wang et al., 2023) which have around 90 classes of objects. Hence, for objects that is in interest such as car, pedestrian, truck, and bus can be easily acquire from this model. By performing the object detection task on the image collected by the mono-camera, it gives out the information of bounding boxes, type of class and class probability on all the objects recognized in the image.

The neural network in YOLOv7 use the forward propagation method compared to other neural network which normally use both forward and backward propagation to perform predictions. Based on the input either image or video, the algorithm will

first find and locate object that present in the frame. This process usually called as object detection. Correspondingly, the algorithm will generate bounding box prediction as well as the label and confident threshold of the objects which this process is called as object recognition. However, these two processes happen together as YOLOv7 is based on YOLO one-stage architecture.

Based on the input, YOLO divides the image into grid cells and a grid cell is in charge of detecting an object if its centre falls within that grid cell. YOLO then predicts bounding box and confidence scores of the box that enclosed the object detected in the grid cell. The network defined the bounding boxes given by the center coordinate of the bounding box (x, y), the width (w) and height (h). Confidence score is defined by the probability of an object respect to their class label that present inside the bounding box. In order to get the prediction, YOLO network predicts the class label in each grid cell as the YOLO model already pre-trained with many types of objects that allows the neural network to be able to classify objects in the input image into its specific labels.

Due to the prediction done by YOLO network for each grid cell, multiple bounding boxes are generated. The network needs to assign just one bounding box predictor to be in charge of each object during training. Based on whose prediction has the highest current intersection over union (IOU) with the ground truth, YOLO designates one predictor for predicting an object. Using IOU method, this allows the network to improve it accuracy in detecting object with different proportions. Another approach that YOLO use to tackle the multiple bounding box generation is by implementing a post-processing step known as Non-Maximum Suppression (NMS). NMS is used mainly to eliminate redundant and overlapping bounding box by suppressing the bounding that have the largest IOU with the current highest probability bounding box. This guarantees that each object is only detected once and one bounding box is obtained. As a result, YOLO was able to produce information about object presence in an image, including information about the object's bounding box location and size, class label, and confidence score. The output of image object detection by YOLOv7 is illustrated by figure 3.8 below.

Figure 3.8 YOLOv7 Object Classification with Bounding Box

This work aims to obtain the bounding box information at the end of the image object detection process using YOLOv7. The image reveals that YOLOv7 surrounds all detected objects with bounding box information and provides a confidence level. Based on this information, the next step involves implementing the association between point cloud cluster data and object-bounding box information using point cloud projection.

3.4.3 Point Cloud Projection onto 2D Image

Figure 3.9 shows the block diagram of the steps taken to project point cloud data onto image. The point cloud data need to be transformed from radar cartesian coordinate (X_R, Y_R, Z_R) into camera coordinates system (X_C, Y_C, Z_C) and then to image coordinate. The coordinate system between radar coordinate and camera coordinate system is significantly different as presented in figure 3.10. The X-axis of the camera and the radar coordinate system remain the same, however, observed that Y-axis of

the radar coordinate system is facing the same axis as z-axis in the camera coordinate system. To transform Z-axis of the radar coordinate system into camera coordinate, it is facing downwards respect to the camera principal axis. From here on, the intrinsic matrix (K), rotation matrix (R), and translation matrix (t) are the main parameter to construct the camera projection matrix (Oh et al., 2018). Assuming that a scatterer point exists in an image coordinate given by (X_C, Y_C, Z_C) and we want to project the point into 2D image plane. Hence, there is a need to transform from image coordinate to pixel known as [u, v]. The transformation matrix can be obtained from the camera-radar calibration and represented as:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = K\ [R \mid t] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

K is the intrinsic parameter of the camera obtained by camera-radar calibration represented as:

$$K = \begin{bmatrix} f_x & 0 & c_x \\ f_y & c_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where fx and fy is the focal length of the camera along the xaxis and y-axis. This focal length determines the cameras zoom level and horizontal field of view. Moreover, cx and cy represents the principal points of x-axis and y-axis which optical axis and image plane intersect. This point is presented in pixels. Note that the Rotation and Translation, [R t], matrix are concatenated where it is known as extrinsic parameter obtained after calibration. Rotation matrix represented as:

$$R = R_x R_y R_z$$

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$$

$$R_y = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

$$R_z = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and translation matrix is represented as:

$$t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

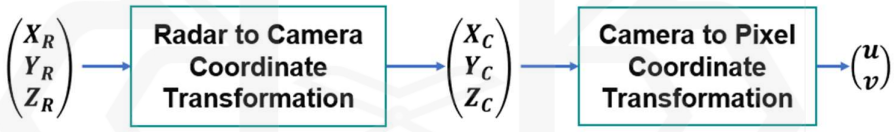


Figure 3.9 Radar to Pixel Coordinate System Transformation

Point cloud projection onto an image can be done after calibrating the camera-radar sensor using the Extrinsic Calibration method and converting the 3D point cloud coordinate into the 2D image coordinate. From here on, the point cloud cluster that is contained inside the bounding box may be retrieved using the bounding box information that we obtained using the YOLOv7 technique. The centre coordinate of the bounding box shows the location of the bounding box in 2D image coordinate along with the width and height. From simple data processing using python, after transforming 3D point cloud onto image, any point cloud cluster that within the bounding box can be extracted. In this stage, the point cloud cluster will be labelled in accordance with the type of class of target in the image included within the bounding box. This produced a dataset of radar point clouds with class labels after a few

processing steps, which may be utilised for neural network classification training and testing.

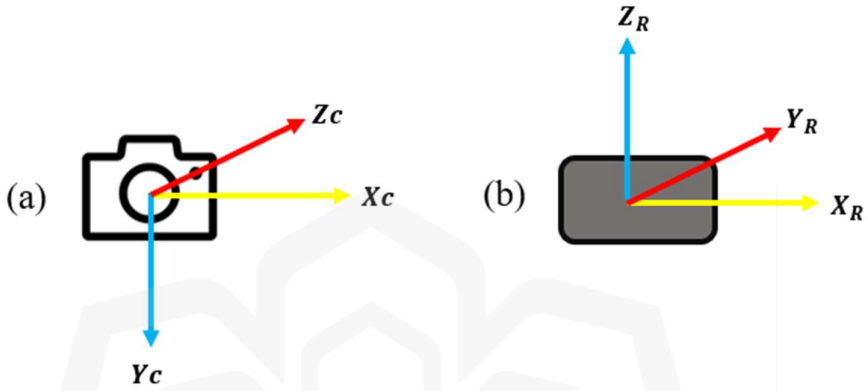


Figure 3.10 Difference Between Camera and Radar Coordinate System. (a) Camera Coordinate System. (b) Radar Coordinate System.

3.4.4 Point Cloud Classification

In this work, we develop a deep learning model by using an open-source machine learning library that is known as Keras. To be more specific, Keras is a neural network library that is built on top of TensorFlow (Boncolmo et al., 2021). Keras use Python programming language which made it more popular choice for Application Programming Interface (API) compared to any other neural network library for instance PyTorch. The integration between Keras and TensorFlow making this open-source machine learning library to be more advantages. TensorFlow work as the backend provides low-level computational complexity on either CPUs or GPUs to train neural network while Keras is the frontend that offer high-level and user-friendly API.

The network architecture of Keras as a deep neural network model is called as Feedforward Neural Network (FNN) (Si et al., 2019; Vinayakumar et al., 2019) as presented in figure 3.11. The network is made up of several layers consist of neurons in each layer and fully connected to each neuron in the following layers. In other words, the neurons in the first layer are connected to adjacent neurons in the following layers, but it is not connected to the neurons that is within the same layers. This explains why this network architecture is known as a feedforward neural network since it goes from the input layer to the output layer in a single direction.

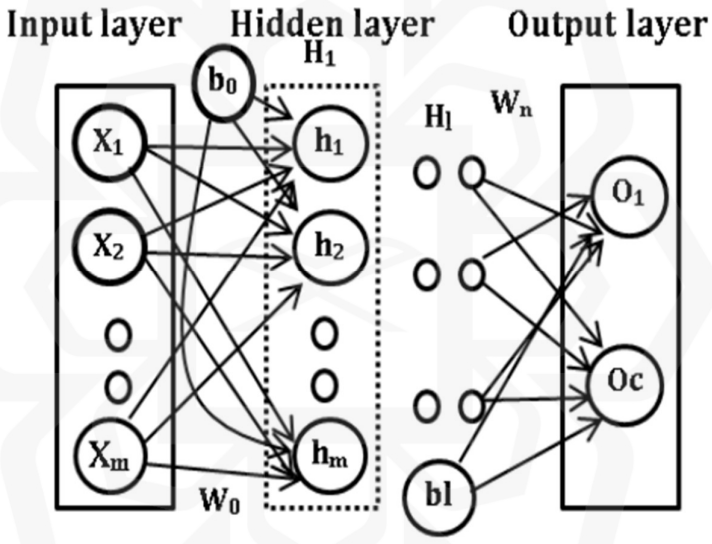


Figure 3.11 Feedforward Neural Network (Vinayakumar et al., 2019)

The feedforward neural network consists of three primary layers, including the input layer, hidden layer, and output layer. The input layer is responsible for receiving the input data provided by the user, as implied by its name. In this study, the point cloud features are fed into the input layer, namely the centre coordinate of the cluster, encompassing the variables x, y, and z, the doppler velocity, power intensity, and count of points within the cluster as shown in figure 3.12. The centre coordinate of the cluster offers location information regarding the target's position within the frame, while the

doppler velocity provides information on the target's doppler speed relative to the radar. The power intensity gives the target's reflectivity in relation to its range following signal processing using a two-dimensional fast Fourier transform (2D FFT). Finally, we provide the count of points that are contained within the cluster. One of the reasons for including count as one of the inputs is to enable the model to learn and distinguish the various point cloud features associated with distinct targets. When comparing buses and trucks to people and vehicles, it can be observed that buses and trucks have significantly higher point cloud densities. The density of the point cloud is significantly increased, leading to a correspondingly higher count of points within a cluster. Therefore, the inclusion of point cloud counts within a cluster has the potential to enhance the performance of deep neural networks in accurately identifying and distinguishing targets of varying sizes.

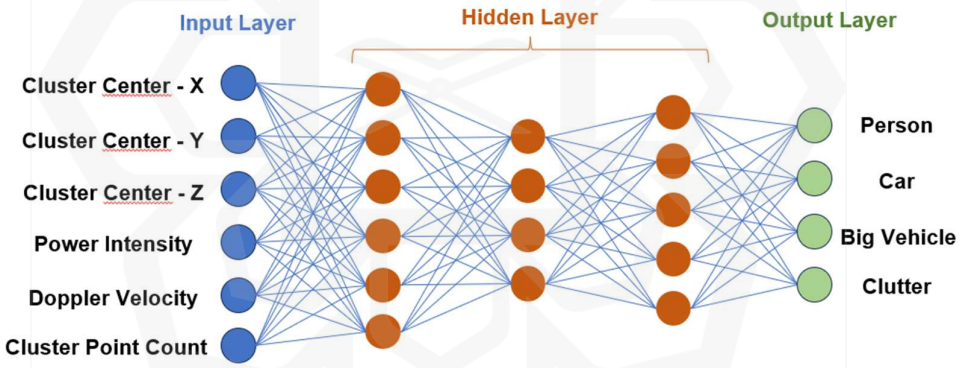


Figure 3.12 Proposed Feedforward Neural Network System Architecture

The presence of neurons in the input layer depends on the dimension of input as for instance in this work the dimension size is 6 hence there is 6 neurons in the input layer. Hidden layer refers to the layers that is between the input and output layers which performs computational task and nonlinear transformation of the input data. The neural network's final layer, the output layer, focuses mostly on providing the network's final prediction. The neurons in this layer are determined by the desired prediction by the user for the network's final result. Binary classification tasks and

multi-class classification are the two categories into which it is typically split. In the output layer of a binary classification task, there is often simply one neuron that provides the likelihood that the input belongs to one of the two classes. While more than one neuron is often included in the output layer of a multi-class classification task to determine which specific class the input data would belong to from more than two classes (Chen et al. 2021).

To handle the numerous kinds of classes we want the neural network to learn and predict in this study, we employ multi-class classification. As a result, there are five classes in our model's final output, including person, car, bus, truck and clutter. Since there are 5 classes that need to be distinguished, the final layer of the neural network will consist of 5 nodes from each class, as was previously stated.

The activation function is another crucial component of the neural network. The non-linear transformation that transforms the input data into output at the hidden and output layers is known as the activation function. By providing non-linearity, this activation function enables the network to learn complicated correlations between the data. In summary, this is how feedforward neural networks work: the user passes on the data features to the layers and each neuron in each layer of the neural network completes its task by weighing the inputs, applying activation functions, and sending the output that has been transformed to the following layer.

In this work, to start developing the neural network model, there are some parameters that need to be defined. First, defining the total number of units and number of layers in the network. Based on literature, the number of layers that is the most accurate neural network need to find through trial-and-error exercise. Based on (Si et al., 2019), they tested 9 different structures of models to identify the best model. Based on its assessment of their work, the neural network model with 32-64-64-64 with Nesterov-accelerated Adaptive Moment Estimation (Nadam) optimization is considered as the best model. Using this discovery, we construct several structures of

models to identify the best model fit for this work. The proposed models to be tested in this work is constructed with different number of layers; 32-16-4, 64-32-4, 64-32-16-4 and 64-64-32-16-4 These four models will be used for training and testing on the point cloud dataset. In defining the number of units for constructing the neural network, it depends on the complexity of the dataset used for training and testing. In this work, since we have a large dataset, the number of units is chosen in the multiple of eight and tested with different number of unit and layers to see which model performed the best. However, it is important to mentioned that the last number of unit is set 4 because the output of this classification is to classify the point cloud data mainly into 4 class which consists of pedestrian, car, big vehicle and clutter. The dataset created after the projection of the point cloud data is split into training set and testing set, each comprising 80% and 20% of the total.

3.5 SUMMARY

This chapter discussed the methodology used in this research work, which includes the hardware and software tools, experimental setup and workflow, as well as the data processing approaches. Each methodology procedure that is chosen in this research work has been planned in order to smooth the workflow of this project.

Referring back to the aim of this study, we adopt 77 GHz 4D Radar Imaging and deep neural network for radar point cloud object classification. The preprocessing of the data for training the deep neural network involve in using unsupervised learning which is DBSCAN for clustering radar point cloud data and YOLOv7 for image object classification to obtained bounding box information of the desired object. By using sensor data fusion, we used point cloud data projection method to associate the radar data and camera data for the purpose of creating a customized dataset. The dataset should include the desired object to be classify in this work such pedestrians, car, bus and truck.

The deep neural network models proposed to be developed in this work will have three different number of layers. The network architecture is powered by KERAS and TensorFlow frameworks consisting of multiple convolutional layers for feature extractions and fully connected layers for classification.

CHAPTER FOUR

RESULTS AND DISCUSSIONS

4.1 INTRODUCTION

In the results and discussion section, we give the comprehensive findings of our study, which encompass several phases of experimentation. These stages include conducting preliminary tests, radar performance test, collecting data at road crossings, and analysing the outcomes of point cloud clustering, point cloud projection onto a 2D image, and the generation of a point cloud dataset with target labels. In the end, the results of the deep neural network model in training and classifying point cloud data were presented.

The first stage of this study is the preliminary test which we conduct to study the working principle of Retina-4F radar imaging and examine the compatibility of the hardware and software platforms after the process of acquiring the hardware tools and installation of software tools. This is to ensure that both of the hardware and software platforms are compatible and function efficiently.

The second stage of this study involves examining the performance of 4D radar imaging in detecting targets of varying sizes inside a controlled open space environment. Moreover, this work is also to learn and study the behaviour of the point clouds for pedestrian, multiple pedestrians, motorcycle and car.

In the next stage of this study, the work has been focusing on collecting and processing image data and radar point clouds data for target classification. We trained the model with the dataset obtained after the data fusion processing. There are two road crossing scenes used for data collection, and the dataset is randomly divided into a training and testing set. The combination of both scenes gives a dataset consisting of more than 10,000 frames. The dataset created after the projection of the point cloud data is split into training set and testing set, each comprising 80% and 20% of the total.

4.2 TASK ONE: PRELIMINARY TEST

The Retina-4F radar imaging technology, developed by Smart Radar Systems, has been marketed as a high-resolution and robust solution for target detection in 4D applications. In this work, the radar is not being develop from scratch where the functionality and working parameter are within control. Therefore, a first experiment is performed to understand the operational principles of Retina-4F radar imaging utilizing the Linux Ubuntu operating system.

4.2.1 Experimental Setup

The preliminary test takes place at the Kulliyyah of Architecture and Environmental Design (KAED) of the International Islamic University Malaysia (IIUM). The purpose of this setup is to conduct a study of a stationary target that has been detected by the radar within its software environment, referred as RVIZ. The available space is restricted to a maximum of 5 square metres, and the chosen location is situated within an enclosed indoor environment. The configuration is shown in Figure 4.1.

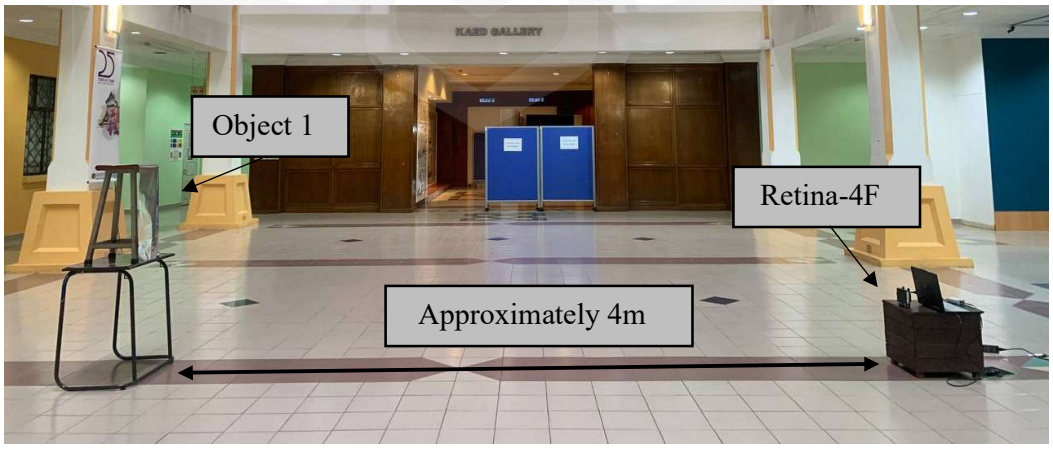


Figure 4.1 Indoor Testing of Retina-4F

4.2.2 Experimental Result

The primary objective of this test is to verify the functionality of both the hardware and software components of the radar system following the installation of the software environment. Additionally, this test aims to investigate the operational principles underlying the radar system. The purpose of implementing this measure is to prevent any potential issues that may develop in the future while conducting experiments and collecting data. During the duration of the experiment, the Retina-4f successfully demonstrated its capability to display point clouds and effectively detect stationary objects. Figure 4.2 illustrates the RVIZ software displaying a point cloud representation of a stationary object that has been successfully mapped onto a 3D plane. The plane in dispute is equipped with a 5-metre grid on each side, allowing for precise observation of objects within the designated circle.

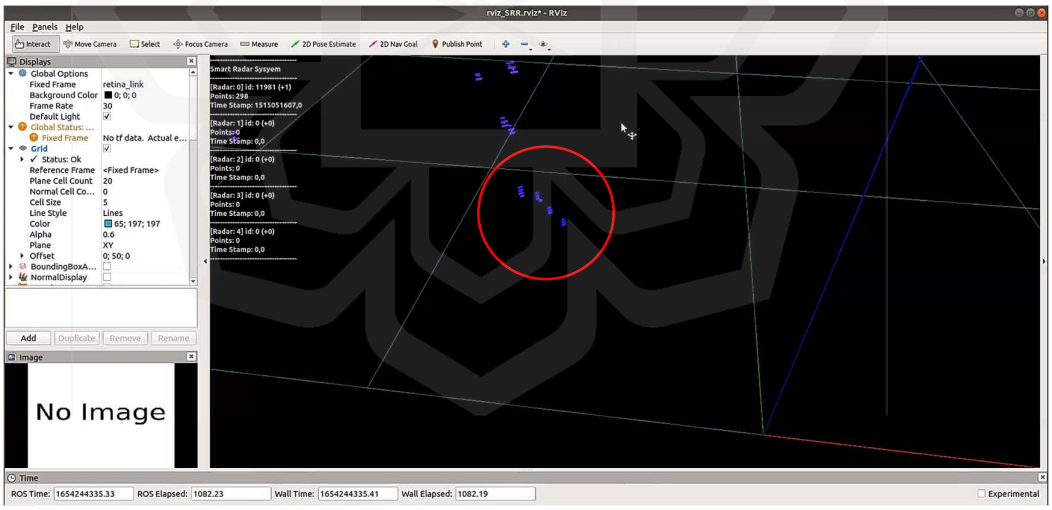


Figure 4.2 Point Clouds of Object 1 in RVIZ

It could be observed that the data output from the radar to the RVIZ is well sent, which allows for the viewing of the point cloud. Moreover, Retina-4F is capable

of extracting 3D information hence, the height of the object can be known by the colour indicator of the point cloud. Blue colours indicate that the object is in the 2m range of height, while red colours indicate the 5m and above range of height. Moreover, the 3D data information extracted by the radar consists of X, Y, and Z axis coordinates, Doppler velocity, and Power intensity. This 4D data information can be observed as illustrated in figure 4.3 below.

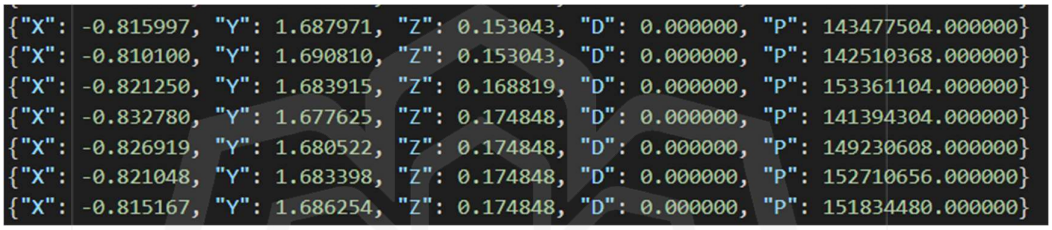


Figure 4.3 Radar Point Cloud Output Information

4.2.3 Discussion

The analysis of the point cloud data of the stationary object in RVIZ indicates that the hardware and software platforms are compatible, without any further issues. In addition, the radar is capable of generating point cloud data that can be utilised for subsequent data processing and manipulation tasks. This preliminary test is crucial as the radar is not developed in this project however, it is acquired directly from market. Due to this, this step to test the compatibility of the platform as well as the understanding how the data of radar could be acquired is important. Subsequently, the project can proceed to the next phase, which involves evaluating the radar's performance in identifying different objects such as pedestrian, multiple pedestrians, motorcycle, and car within an open space environment.

4.3 TASK TWO: RADAR PERFORMANCE TEST

A radar performance test was conducted to analysed it performance is detecting several targets which includes pedestrian, multiple pedestrians, motorcycle and car. Retina-4F system offers data related to the target, specifically its X, Y, and Z coordinates in radar point clouds with respect to its axis, as well as its doppler velocity and power intensity. The data provided is of most significance for utilisation in data processing and analysis, with the goal of developing a deep learning classification system. The analysis of the radar's initial performance is accomplished by utilising the data obtained from an experiment carried out within a controlled test environment. The objects present in this scenario include pedestrian, multiple pedestrians, motorcycle, and car, which are frequently observed targets at road crossings. The primary objective of this test is to investigate the characteristics and behaviour of the point cloud generated by the radar. The point cloud data was afterwards transformed into a three-dimensional representation using the Python programming language.

The colour bar shown in figure 4.4 presents an image showing the range of doppler velocity of the objects. By employing different patterns of colour, it is possible to determine the motion of an object's movement relative to the radar, whether it is travelling away from or towards it. The colour of an object can be used to signify its doppler velocity, with hotter colours such as red indicating higher velocities for object that is moving away from the radar. In contrast, the utilisation of cool colours, specifically blue, is used to indicate that as velocity increases for targets that is moving towards the radar. The analysis of a desired object in point cloud involves taking into account both the movement of the object and its velocity. Therefore, by utilising the Python programming language to generate a visual representation of a point cloud and remove unnecessary data points while concentrating only on the object of interest inside the area. A dataset including 5107 frames was acquired in this experiment to investigate the radar's performance. Hence, within this section, many scenarios are presented, and the outcomes are measured in order to analyse the performance of the radar system.

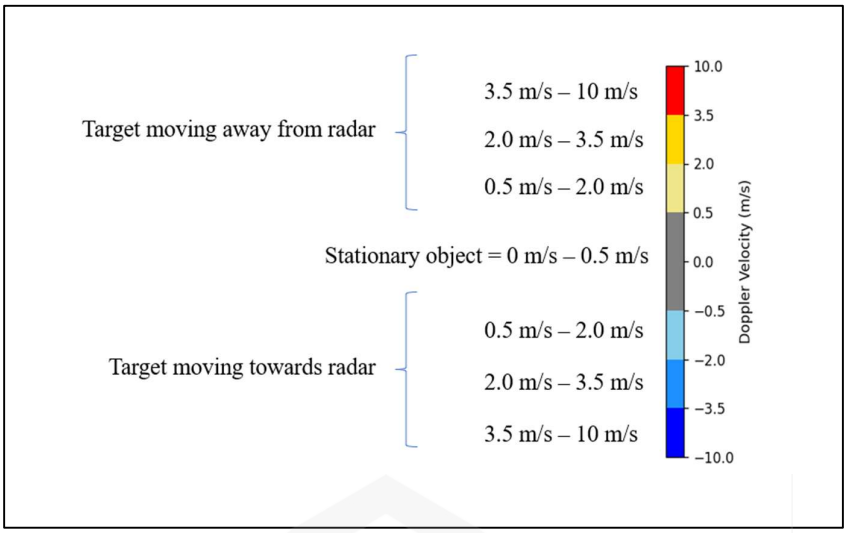


Figure 4.4 Doppler Velocity Colorbar Indicator

4.3.1 Experimental Setup

In this experiment, we conducted the performance test at IIUM marching field area that consists of roughly 90-meter width times 130-meter long. The reason for choosing this particular location for experimentation is based on its characteristics as an expansive, open-air space that is situated at a considerable distance from potential sources of interference, such as trees and vehicular activity. The location has strategic significance in ensuring optimal radar performance by minimising the presence of background noise that could potentially disrupt the experimental outcomes. Figure 4.5 shows the marching field area as observed in Google Maps.

Figure 4.5 IIUM Marching Field as Observed in Google Maps

We perform the measurement within a controlled setting as illustrated in figure 4.6. The radar system is attached to a tripod located at a height of 150 centimetres (cm) above ground level as shown in figure 4.7. The horizontal dimension of the area is defined as a width of 60 meters, while its length in the vertical direction is 50 meters. However, this range may vary depending on the specific testing scenario.

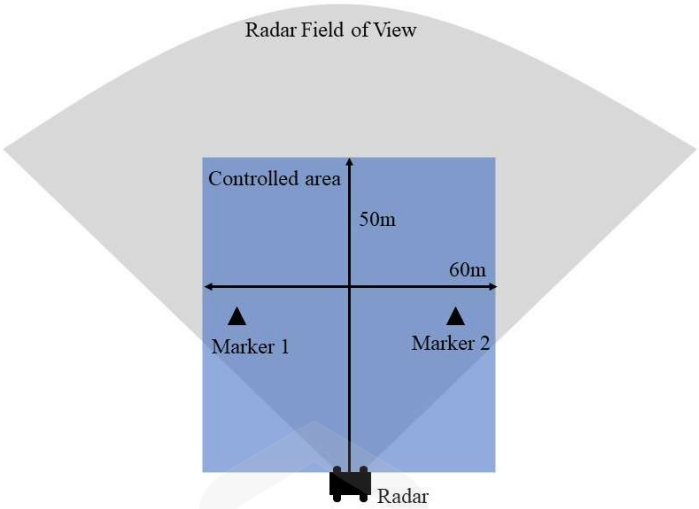


Figure 4.6 Experimental Setup of the Controlled Area

Figure 4.7 Radar Mounted at the Controlled Area

4.3.2 Experimental Results

We presented the outcomes of the performance test result for four scenarios which includes a pedestrian, multiple pedestrians, motorcycle and car.

4.3.2.1 Maximum Detection Range of a Pedestrian

The first scenario is to evaluate the maximum range of pedestrian detection inside a restricted area of 50 meters, adopting a short-range field of view. There are two fixed objects, acting as markers, positioned at distances of 9 m and 18 m, respectively as shown in figure 4.8. These objects are corner reflectors. The purpose of this corner reflector is to distinguish between a moving target and a static object in order to analyze the characteristics of the point cloud data associated with each object as detected by the radar.

Figure 4.8 Pedestrian. (Camera View)

Based on 15 measurements conducted, the mean value for the maximum range detection of pedestrian is determined to be 39.3 meters. This is shown in figure 4.9 where we run a python script to detect the maximum detection range of a pedestrian for every 15 measurements and calculate the mean.

Figure 4.9 Mean Average of Maximum Detectable Distance for a Pedestrian

The pedestrian is maintaining a consistent velocity while crossing the route. Observed from the range doppler plot, the velocity is maintained around 1.00 m/s to 1.50 m/s (3.6 km/h to 5.4 km/h) with the highest concentration around 1.20 m/s to 1.28 m/s (4.32 km/h to 4.61 km/h). This is illustrated, as shown in figure 4.10. Moreover, based on the range doppler velocity plot, while the pedestrian is maintaining its speed along the route, in the range of 26 m and above, it can be observed that the doppler velocity of the pedestrian detected by the radar is getting distorted.

Figure 4.10 Doppler Velocity Distribution for a Pedestrian

Due to this, we plot the 2D X-Y axis of the pedestrian at 32.8 m to observe the doppler velocity of the pedestrian. According to the 2D plot displayed in figure 4.11, it can be observed that at a distance of roughly 32.8 meters, the moving object displayed a velocity that was nearly identical to that of the stationary object (nearly zero). In this case, the Doppler velocity was found to be 0.119993 m/s (0.432 km/h). This information is based on the point cloud data retrieve from the radar as displayed in figure 4.12. This observation indicates that as the detection range for pedestrians reaches its maximum, the accuracy of the Doppler velocity measurements for moving pedestrians slightly decreases, however, the range is detection of this radar for a pedestrian is satisfactory.

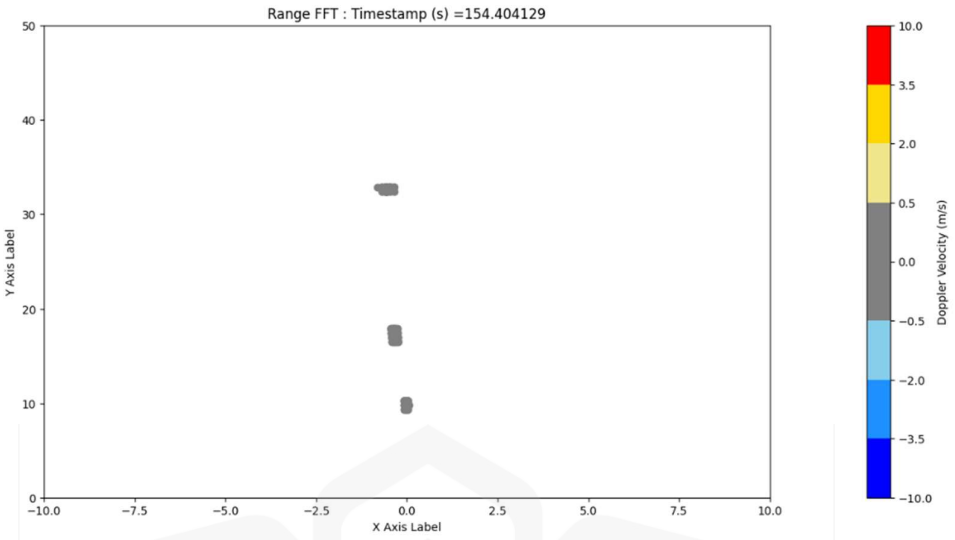


Figure 4.11 Pedestrian. (X-Y-axis 2D Plot)


```
82622  {"X": -0.546962, "Y": 32.385017, "Z": 2.963687, "D": 0.119993, "P": 290249.656250}
82623  {"X": -0.678202, "Y": 32.377033, "Z": 3.023189, "D": 0.119993, "P": 282160.000000}
82624  {"X": -0.565181, "Y": 32.379208, "Z": 3.023189, "D": 0.119993, "P": 289504.312500}
82625  {"X": -0.452153, "Y": 32.380981, "Z": 3.023189, "D": 0.119993, "P": 287678.843750}
82626  {"X": -0.339120, "Y": 32.382359, "Z": 3.023189, "D": 0.119993, "P": 276683.500000}
82627  {"X": -0.564529, "Y": 32.341846, "Z": 3.399782, "D": 0.119993, "P": 267628.218750}
82628  {"X": -0.481515, "Y": 32.865616, "Z": 2.804545, "D": 0.119993, "P": 142169.140625}
82629  {"X": -0.574328, "Y": 32.903206, "Z": 2.301167, "D": 0.119993, "P": 131062.476562}
82630  {"X": -0.459470, "Y": 32.905010, "Z": 2.301167, "D": 0.119993, "P": 132314.734375}
82631  {"X": -0.688570, "Y": 32.872005, "Z": 2.683907, "D": 0.119993, "P": 136529.546875}
82632  {"X": -0.573821, "Y": 32.874207, "Z": 2.683907, "D": 0.119993, "P": 141167.703125}
82633  {"X": -0.459066, "Y": 32.876011, "Z": 2.683907, "D": 0.119993, "P": 141268.265625}
82634  {"X": -0.344304, "Y": 32.877411, "Z": 2.683907, "D": 0.119993, "P": 136831.203125}
82635  {"X": -0.599011, "Y": 32.859062, "Z": 2.858177, "D": 0.119993, "P": 145113.562500}
82636  {"X": -0.802494, "Y": 32.835957, "Z": 3.066283, "D": 0.119993, "P": 133976.453125}
82637  {"X": -0.687870, "Y": 32.838558, "Z": 3.066283, "D": 0.119993, "P": 140885.375000}
82638  {"X": -0.573238, "Y": 32.840759, "Z": 3.066283, "D": 0.119993, "P": 143125.125000}
82639  {"X": -0.458598, "Y": 32.842560, "Z": 3.066283, "D": 0.119993, "P": 140695.703125}
82640  {"X": -0.343954, "Y": 32.843960, "Z": 3.066283, "D": 0.119993, "P": 133597.156250}
```


Figure 4.12 Radar Data for a Pedestrian at Range 32.9 meters

4.3.2.2 Multiple Pedestrians Detection

The second scenario as shown in figure 4.13 involves evaluating the radar's effectiveness in accurately detecting and differentiating multiple people who are in close range, specifically at a distance of less than 1 meter from each other. 15 measurements were taken in total.

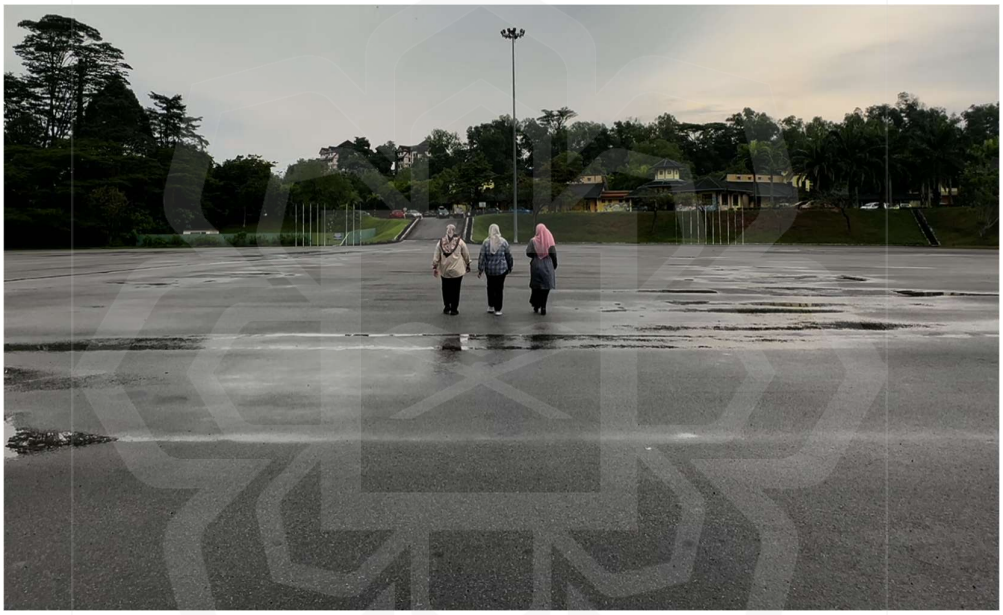


Figure 4.13 Multiple Pedestrians. (Camera View)

The pedestrians cross a controlled test area with a range of 50 meters, maintaining a constant pace. Can be observed the from the point cloud data collected and illustrated in figure 4.14, the pedestrians are maintaining their speed in the range of 1 m/s to 1.5 m/s (3.6 km/h to 5.4 km/h) with highest concentration speed around 1.02 m/s to 1.10 m/s (3.67 km/h to 3.96 km/h).

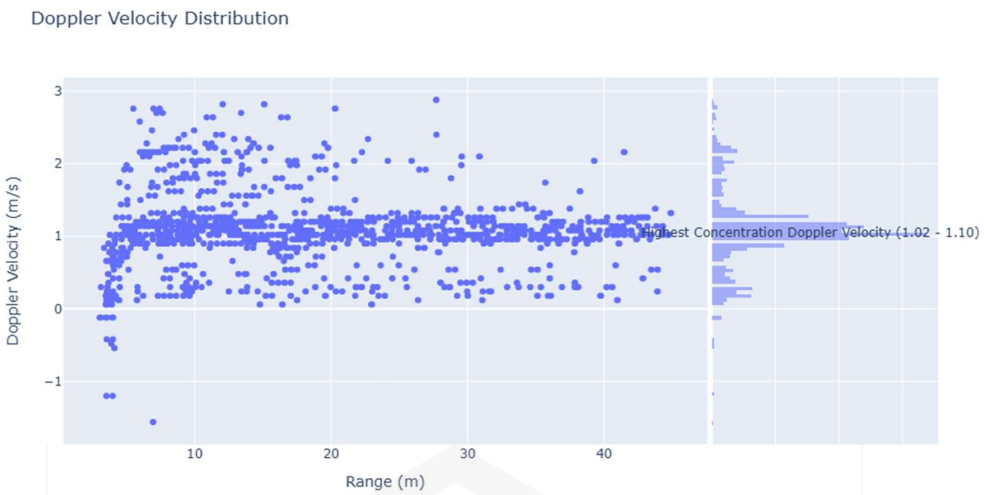


Figure 4.14 Doppler Velocity Distribution for Multiple Pedestrian

One of the main purposes of this test measurement is to study the radar's capability to distinguish multiple pedestrians across a controlled range of distance. Figure 4.15 presents the results of one of the conducted measurements. The method used in this test is point cloud clustering using DBSCAN. The purpose of using data clustering is to disguise between different group of point cloud data that represent the pedestrians as detected by the radar. As shown in figure 4.15, we plotted clusters to demonstrate this, each representing three pedestrians at varying ranges.

According to the figure, the radar performs satisfactorily in detecting multiple pedestrians within a range of 15 m and below. At the range of 15 m and above, the radar's ability to distinguish between pedestrians has deteriorated as the pedestrians' representation by the same clusters. This constraint highlights the significance of advancing radar technology to improve its ability to distinguish objects at longer distances, therefore enhancing its usefulness in monitoring. However, since the radar used in this work is limited to medium range, this result is satisfactory.

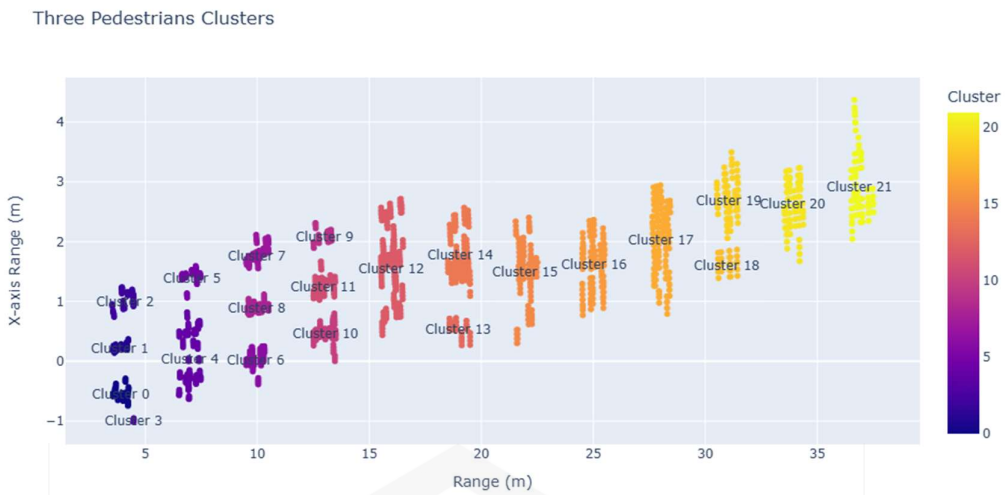


Figure 4.15 Multiple Pedestrians Clusters across the Range

At a distance of 15 meters, it is possible to identify multiple pedestrians. Range resolution is a crucial aspect of automotive radar systems as it signifies the system's capacity to differentiate between several targets. Based on this, we can observe that when the range resolution of the radar system is enhanced, the radar capability of distinguishing multiple targets with close distance to each other is high. Furthermore, environmental factors such as the amount of noise and the ruggedness of the test environment can also impact the radar's ability to accurately distinguish between pedestrians at much longer distances. This emphasises the intricate relationship between technological capabilities and the natural environment in applications such as radar imaging.

4.3.2.3 Motorcycle Detection

The third case scenario is displayed as shown in figure 4.16. The objective of this scenario is to examine the point cloud's behaviour in relation to the vehicle's orientation (slanted and horizontal orientation). The controlled test area has dimensions of 70 meters in length and 60 meters in width.

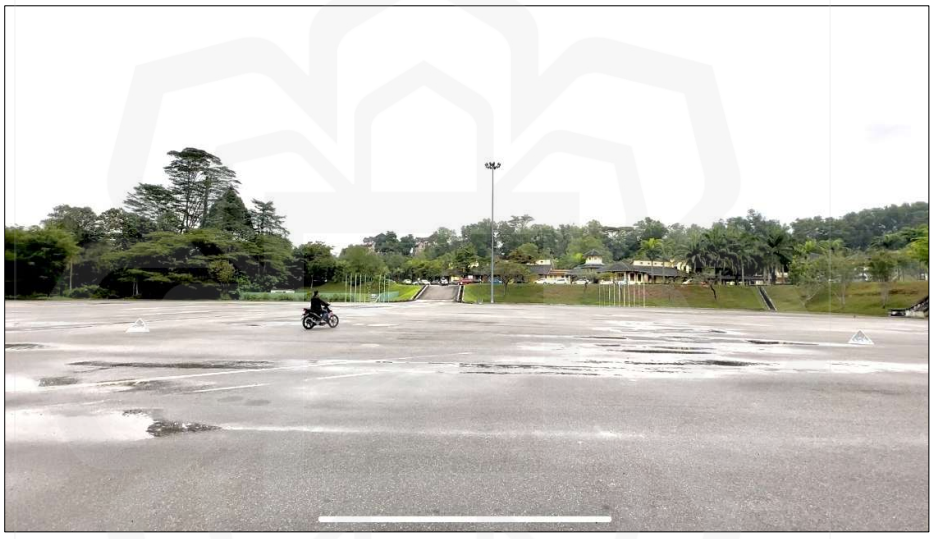


Figure 4.16 Motorcycle. (Camera View)

The motorcycle maintains a consistent velocity within the range of 5 km/h to 10 km/h (equivalent to 1.30 m/s to 2.80 m/s). Figure 4.17 clearly shows that the radar detects the doppler velocity of the motorcycle within the specific range. The maximum concentrations of doppler velocity are observed between 2.66 m/s and 2.70 m/s, which is equivalent to 9.58 km/h to 9.72 km/h.

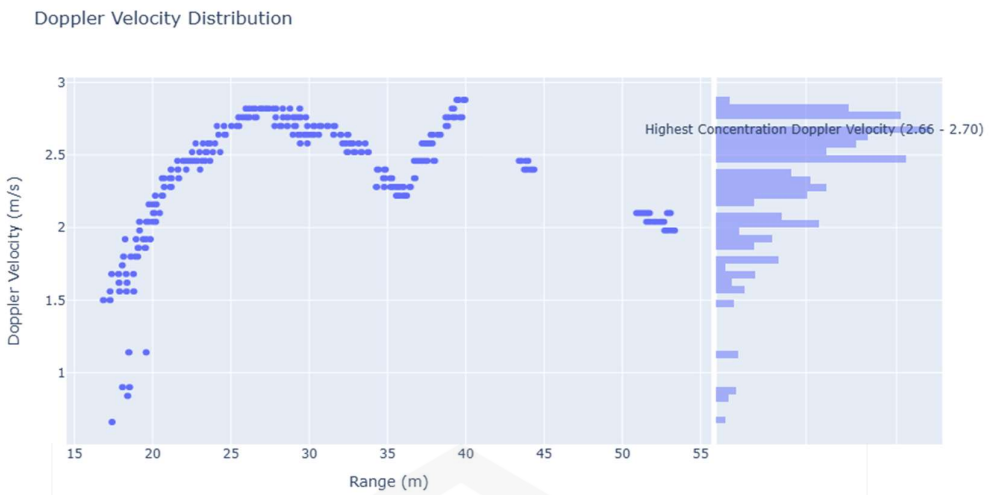


Figure 4.17 Doppler Velocity Distribution for a Motorcycle

According to figure 4.18, the trajectory followed by the motorbike has a diagonal pattern resembling that of a rectangle. From this test, we can study that as the motorcycle approaches a distance of 40 meters and beyond, the ability to detect the motorcycle gradually decreases. This behaviour can be observed from figure 4.17 above and figure 4.18 below. Due to this, certain parts of the point clouds are not represented in the plot. The potential cause of this happening can be attributed to the orientation of the vehicle and the resulting echoes from scattered objects. However, compared to pedestrian detection, it is noticeable that the doppler velocity of a motorbike going at a consistent range speed is effectively identified at distances of 40 meters and beyond. The yellow spots presented in figure 4.18 show this observation.

Next, it is seen that the movement of the target towards and away from the radar can be identified from the observed point cloud behaviour in term of its velocity. This is represented in figure 4.18, where the target's route is represented by a blue dotted point plot on the map. The situation is similar when comparing the target's movement away from the radar, namely at distances of 40 meters and beyond. In this case, the radar is unable to identify motorcycles that are approaching it in slanted

orientation. Nevertheless, within a distance of 40 meters or less, the motorbike exhibits a high level of detectability.

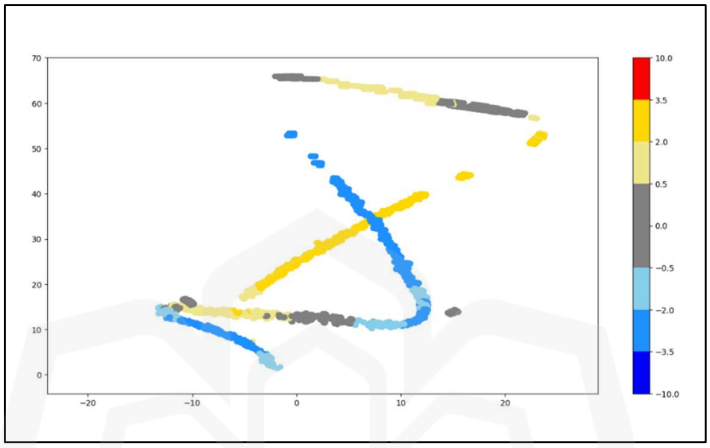


Figure 4.18 Point Cloud of Overall Route Taken by the Motorcycle Detected by The Radar Image

In addition, it is worth noting that, the doppler velocity of the motorcycle detected by the radar in horizontal orientation is slightly different as compared to slanted orientations. The doppler shifts observed when an item moves horizontally relative to the radar largely indicate the radial component of its velocity. This alignment enables more forward computation measurement of velocity and orientation by utilising the measured frequency shift. On the other hand, when an object moves at an angle or slanted direction in relation to the radar, measuring accurately the component of its velocity that aligns with the radar's line of sight becomes more challenging. The radar's observation of the Doppler shift will be affected by both the object's velocity components: the radial and transverse velocities relative to the radar. due to this, the doppler velocity of the object detected by the radar respect to its orientation are slightly different even though the object maintained it speed during the test.

4.3.2.4 Car Detection

In this fourth scenario, the experiment is carried out to investigate the radar's ability to detect a car as its target and the characteristic of the point cloud data of the car while identifying the capability of the radar system to identify cars at extended distances in a wide testing area. The dimensions of the controlled test area are 70 meters in length and 60 meters in width. The scenario is displayed as shown in figure 4.19.

Figure 4.19 Car. (Camera View)

The car moves while maintaining a velocity range of 10 km/h to 20km/h (2.78 m/s to 5.56 m/s). This can be observed from the radar data collected that the car is moving within the said speed range with highest concentration of velocity in between 3.36 m/s to 3.49 m/s (12.096 km/h to 12.564 km/h) as illustrated in figure 4.20 below.

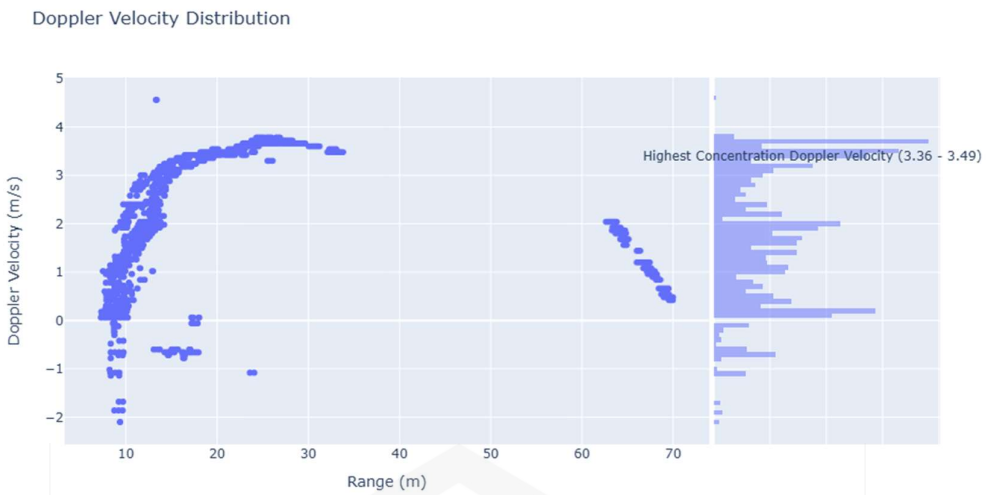


Figure 4.20 Doppler Velocity Distribution for a Car

According to the data plotted in figure 4.21, the trajectories taken the car is random without any fix orientation. This is because the main idea of this test is just to observe and study the detection of the car is extended distance compared to other objects in this test. However, it is noticeable that while the car is in slanted motion towards the right by a distance of 40 meters, the target is observed to be undetectable. This behaviour can be observed in both figure 4.20 above and figure 4.21 below.

In comparison to the detection of motorcycles, it has been observed that the detectability of motorcycles has decreased within a range of 40 meters. Nevertheless, within the domain of car detection, it has been observed that cars remain detectable within a range of approximately 40 meters to 70 meters when it is moving towards the radar. Therefore, in terms of a broad measuring range, the car can be detected at a distance of more 40 meters. Figure 4.21 shows the overall route taken by the car in this test.

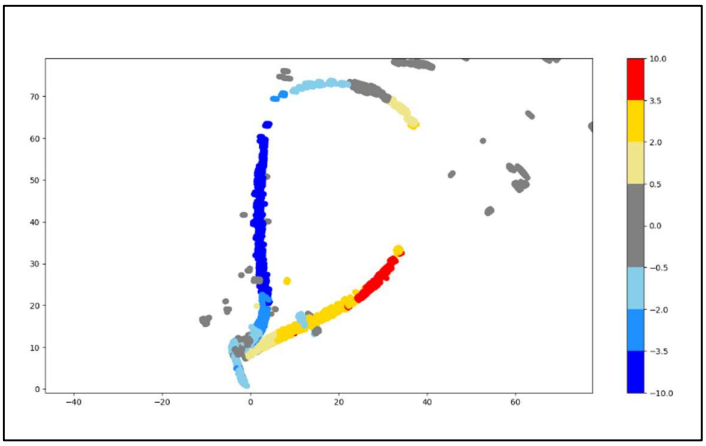


Figure 4.21 Point Cloud of Overall Route Taken by the Car Detected by The Radar Image

4.3.3 Discussion

The performance test of this project is presenting the radar's capability in identifying targets of varying sizes and forms. By analysing the results obtained from the radar performance measurements, we are able to study the characteristics of the point cloud representing the detected objects. This insight contributes to our comprehension of the radar's functionality as a sensor for object detection. There are several studies that we can study here which will be further discussed below.

The nature of the point cloud data of pedestrians, motorcycle, and car vary due to the reflection wave from the object being targeted to the radar being influenced by the size and reflectivity of the respective targets. Nevertheless, the radar's ability to detect targets over extended distances and wide areas has resulted in satisfactory outcomes, demonstrating its dependable performance.

Initially, when we measure a single pedestrian's maximum range, we observe that the mean average for the maximum detectable range is 39.3 metres. This result

indicates that in the next phase of this study, where we collect measurements at road crossing areas, the position of the radar and the road crossing area are no more than 40 metres ahead of each other. We aim to guarantee that the radar can effectively gather and identify pedestrians during the measurement. Since pedestrians primarily use the road crossing area, it makes sense to make sure the crossing areas are within their detection range.

Next, during the test of the radar's ability to detect multiple pedestrians, we discovered that the radar could distinguish multiple pedestrians within a range of 15 meters ahead. This too could be one of the reasons why, when mounting the radar in road crossing areas, the crossing areas must be within this range to ensure a good result for multiple pedestrians that are crossing the road. However, to ensure the collection of sufficient data surrounding the crossing areas, the area must be large enough, with a 40-meter range being sufficient for the maximum detectable range of pedestrians.

For the motorcycle and car tests conducted, we learned that motorcycles and cars are detectable at extended range as compared to pedestrians. However, the orientation of the moving car and motorcycle during the measurement has shown the difference in the doppler velocity of the objects. This is due to the doppler effect, which measures the frequency shift of signals reflected off moving objects in radar systems. From this insight, we propose that in the next phase of the measurement, the road crossing area will be positioned in a slanted orientation with respect to the position of the radar. This is to ensure that the doppler shift of the objects on the road and the crossing area during the measurement are taken into account.

Gathering all of the outcome in this test has leads the way for the subsequent phase of the project, which involves the measurement at road crossing area to create a custom dataset for the development of a deep neural network algorithm for target classification. This algorithm will utilise sensor fusion techniques, combining radar imaging and camera data.

4.4 TASK THREE: DATA COLLECTION AT ROAD CROSSING AREAS

This section will present data collected from two distinct places. The data collection process is conducted with the objective of collecting data from the sensors in order to create a customised dataset for utilisation in this study. The reason behind the creation of a customised dataset is due to the lack of readily available point cloud datasets that include target labels specifically for road crossing areas. Furthermore, within the framework of this research method including the integration of two different sensor technologies, we have generated a dataset by collecting data from both the Retina-4F sensor and the camera.

4.4.1 Experimental Setup

The measurement for data collection in this task is conducted at two road crossing areas located at University of Waterloo, Canada. Figure 4.22 shows the road crossing area at location 1 while figure 4.23 shows the road crossing area at location 2 in Google Maps. The red circle shape highlighting the road crossing in the images. The selection of this areas for measurement is based on the visibility of desirable targets, including people, cars, buses, and trucks.

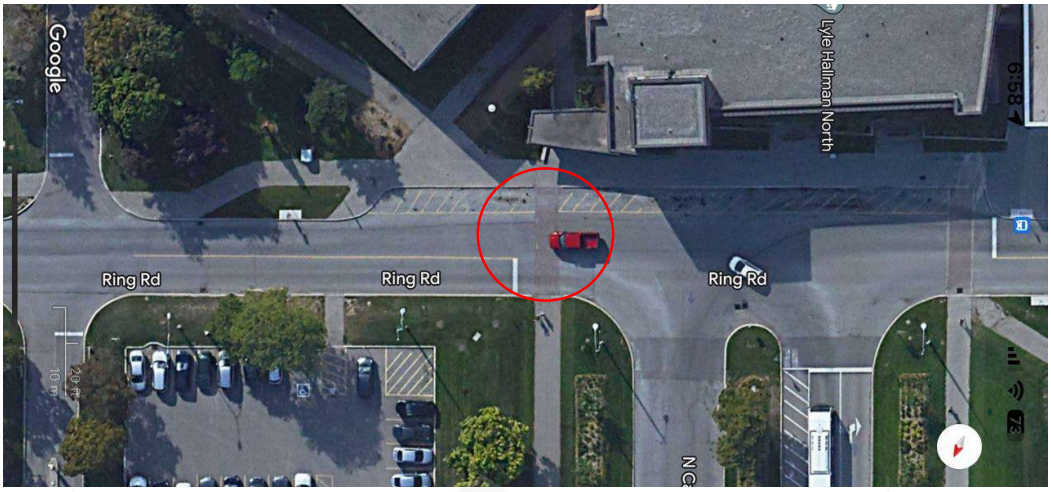


Figure 4.22 Road Crossing at Location 1

Figure 4.23 Road Crossing at Location 2

Figure 4.24 shows the scenario of experimental setup at road crossing location 1 and figure 4.25 shows the scenario of experimental setup at location 2. Both of figures shows the radar and camera attached to a tripod, positioned at a distance of 150 cm from the ground, which corresponds to the tripod's maximum height. Since this project is experimental, the measurements are carried out at a real-scenario road

crossing location at two different places. At both of the location, the pedestrian road crossing is located at a distance of approximately 40 m from the radar's location.

Figure 4.24 Experimental Setup at Location 1

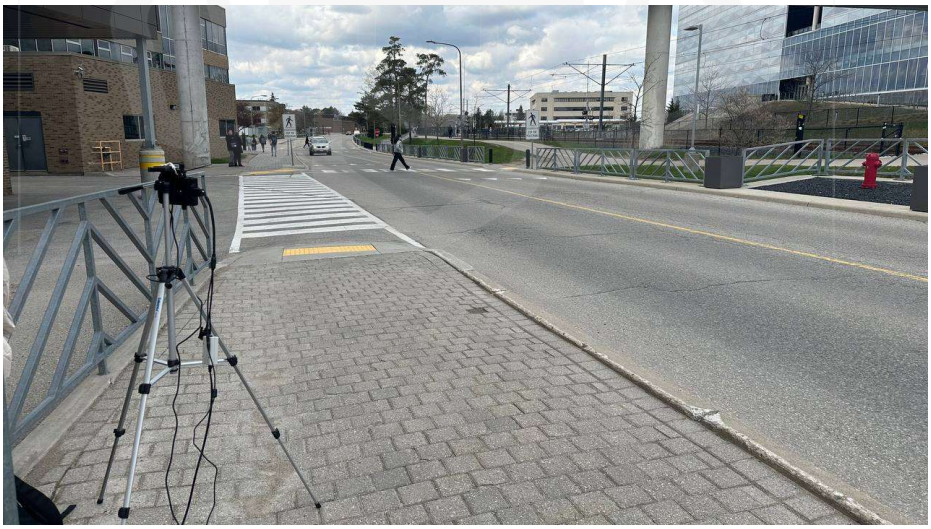


Figure 4.25 Experimental Setup at Location 2

Figure 4.26 and figure 4.27 displays the field of view that is being observed by the radar and camera. The radar has the azimuth field of view up to 100° ± 50°, hence, it is considered very wide view and sufficient to monitor the whole road crossing areas.

Figure 4.26 Radar and Camera Field of View at Location 1

Figure 4.27 Radar and Camera Field of View at Location 2

4.4.2 Experimental Results

In this section, we provided the outcomes of the sensor data fusion process following the steps of clustering, point cloud projection, and labelling for data collected at road crossing areas. All of the data processing procedures are executed using the capabilities of the python programming.

In our python data processing system, each target has been labelled using distinct colors to represent different classes. In this context, persons are associated with the color blue on the map, cars are represented by the color red, buses are denoted by the color orange, and trucks are represented by the color purple on the map. In addition, aside from the point cloud associated with the targets, all other point clouds of the background and uninterested targets are now assumed to be cluttered and colored grey.

4.4.2.1 Person

The result shown in figure 4.28 represents the outcomes of the data fusion processing, specifically the projection of point cloud data onto image data for pedestrian class. The acquired point cloud data went through processing and was afterwards projected onto image data through the use of python programming. The projection parameters required for converting the point cloud coordinates into image coordinates were obtained through the calibration of the camera and radar. Therefore, according to the calculation performed, the point cloud was effectively projected onto 2D image.

Figure 4.28 Pedestrians - Point Cloud Projection onto Image

Moving forward, as explained in Chapter 3, the image data undergoes processing by Yolov7 in order to extract the bounding box information related to the objects that are present within the image. Based on the provided bounding box information, we proceed by plotting the same bounding box onto the image that bounded the targets. The cluster of point cloud contained within the bounding box can be extracted from this point in time. The cluster of point cloud that has been extracted is assigned labels based on the bounding box labels. For example, if the bounding box is labelled as "person," the point clouds within that bounding box are now labelled as "person." Figure 4.29 presents the representation of point clouds projection and bounding boxes overlaid on the pedestrian image at both locations.

Figure 4.29 Pedestrians - Point Cloud Projection and Bounding Box Plot onto Image

The scatter plot in figure 4.30 displays the point cloud data after performing the clustering procedure using DBSCAN. Based on the analysis of the plot, it is seen that the clustering process of the point cloud focusing on grouping the point clouds, thereby transforming the point cloud into a cluster. In order to facilitate the training of a deep neural network in acquiring knowledge of the spatial relationships within a point cloud, it is important to assign appropriate labels to each cluster of the point cloud, thereby associating them with their respective classes. Therefore, the process of projecting a point cloud is the primary mechanism responsible for labelling clusters within a given dataset. Subsequently, a custom dataset of point clouds data and their related labels has been acquired.

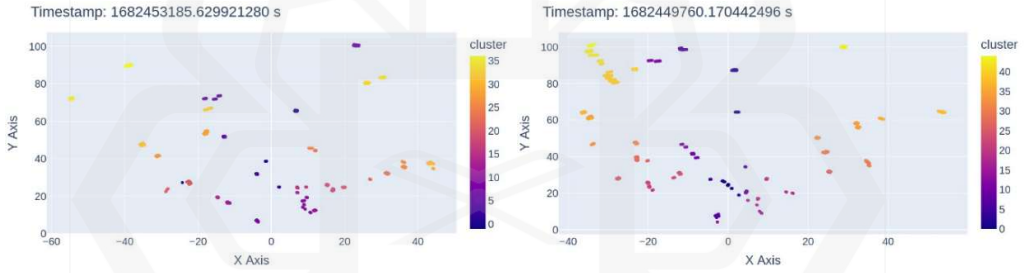


Figure 4.30 Pedestrians - Scatterer Plot of Point Cloud Clusters

Figure 4.31 and figure 4.32 shows point cloud projection onto 2D image with the bounding box alongside the scatterer plot of a frame of point cloud data with label from the customize dataset. In this plot, can be seen that the clusters of point clouds are now have been assigned with labels based on the objects' class type in relation to the bounding box. Figure 4.31 illustrates the situation at road crossing position 1, whereas figure 4.32 illustrates the scenario at road crossing 2.

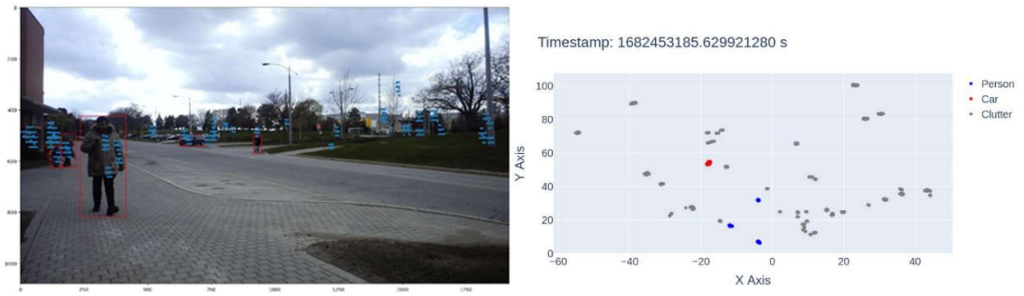


Figure 4.31 Pedestrians - Point Cloud Labelling at Location 1

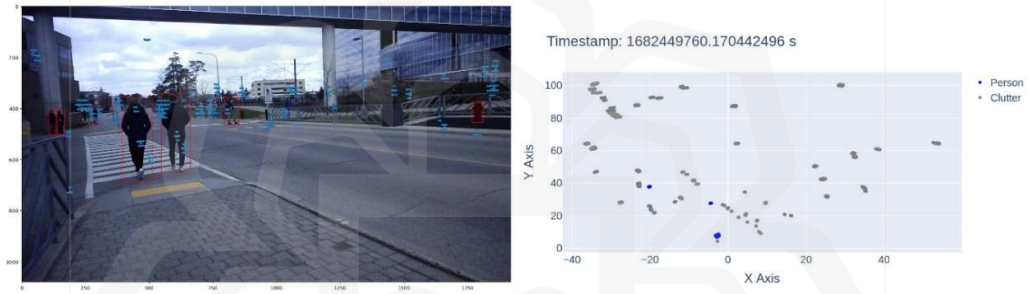


Figure 4.32 Pedestrians - Point Cloud Labelling at Location 2

4.4.2.2 Car

In this subsection, we want to present the outcomes of labelling point cloud data related to cars. The data processing steps are similar across all target classes. The projection of point cloud data obtained from radar onto picture data from a camera is shown in figure 4.33.

Figure 4.33 Cars - Point Cloud Projection onto Image

Following that, by utilising the bounding box data of the cars from YOLO, we proceed to plot the bounding box onto the image data and afterwards extract the point clouds cluster belonging to the cars that are contained in the bounding box. Figure 4.34 shows the presence of cars in the image with the bounding box and point clouds at both road crossing areas.

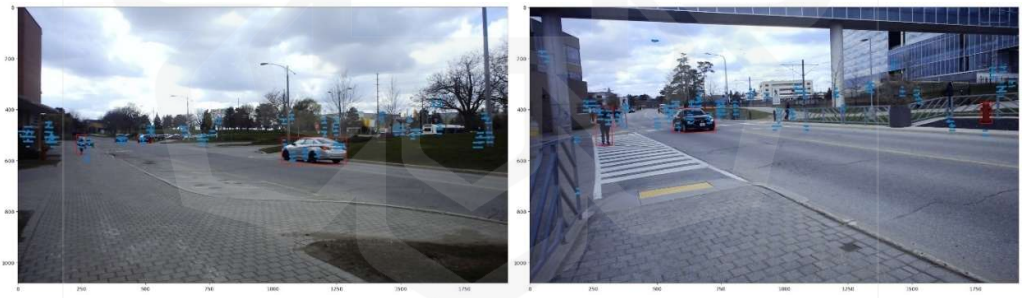


Figure 4.34 Cars - Point Cloud Projection and Bounding Box Plot onto Image

Similarly, the scatter plot representing the point cloud data after clustering is shown in figure 4.35. In contrast to the point cloud data representing pedestrians, it is observable that the reflectivity of cars results in denser point clouds in the plot.

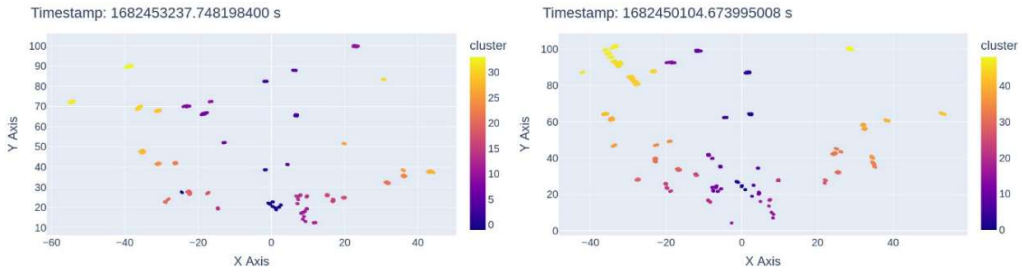


Figure 4.35 Cars - Scatterer Plot of Point Cloud Clusters

In figure 4.36, the presence of two cars is shown at road crossing location 1. The scatter plot of the point cloud data also shows the presence of two cars, which are visualized by the colour red. The procedure of labelling targets for cars have demonstrates a satisfactory outcome. The outcomes for cars at road crossing location 2 are shown in figure 4.37 that display the presence of a car along with two pedestrians.

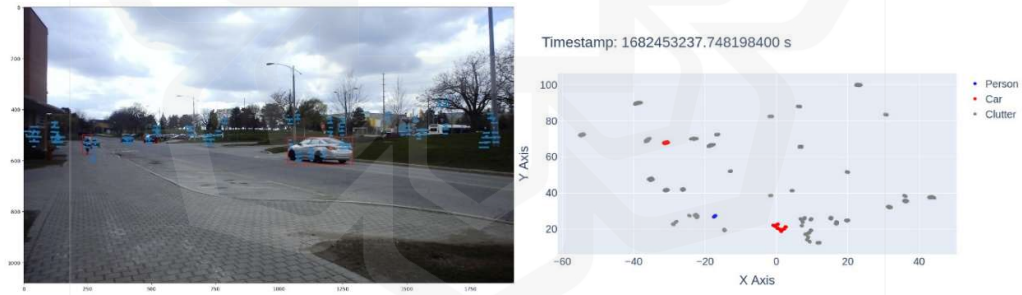


Figure 4.36 Cars - Point Cloud Labelling at Location 1

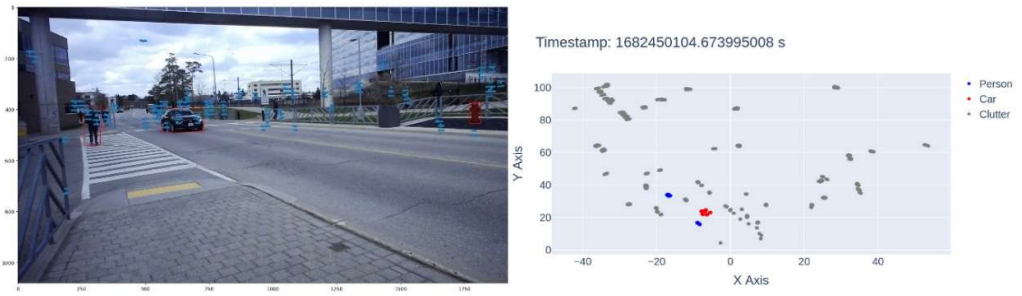


Figure 4.37 Cars - Point Cloud Labelling at Location 2

4.4.2.3 Bus

In this subsequent discussion, we will show the results obtained from the analysis of point cloud and image processing techniques, specifically focusing on identifying and classifying bus class types through target labelling. In comparison to both pedestrian and car classes, the bus class shows significantly greater signal reflection, leading to a notable increase in the density of point clouds, as depicted in figure 4.38.

Figure 4.38 Bus - Point Cloud Projection onto Image

By employing the YOLO algorithm for image data processing, we are able to extract the bounding box information relating to the bus. Subsequently, this information is added into the image as shown in figure 4.39, along with the projection

of the point cloud. Next, the cluster of point cloud is extracted and then linked to the label "bus".

Figure 4.39 Bus - Point Cloud Projection and Bounding Box Plot onto Image

Through the utilisation of scatter plots in Python, the visual representation of point clouds allows for the observation that the point cloud corresponding to buses has a higher level of recognizability in comparison to those of cars and persons. This distinction arises from the dense nature inherent in the point cloud of buses. This is illustrated in figure 4.40.

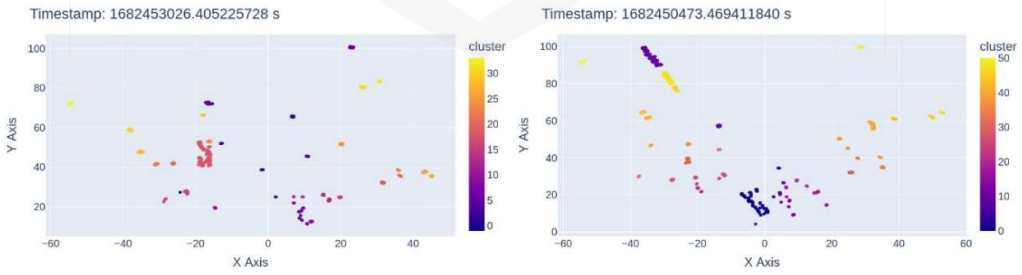


Figure 4.40 Bus - Scatterer Plot of Point Cloud Clusters

Once the cluster has been labelled with its corresponding class, the existence of bus labelling in the dataset can now be observed, as represented in figure 4.41 and figure 4.42. The bus is visually represented through the utilisation of the colour orange, while the surrounding noises are depicted as a colour of grey.

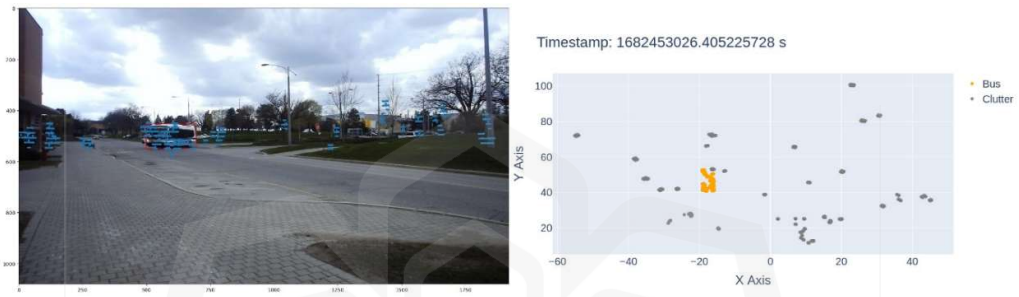


Figure 4.41 Bus - Point Cloud Labelling at Location 1

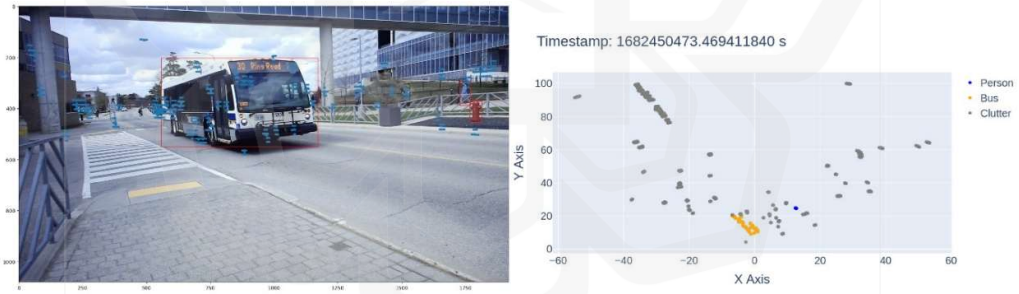


Figure 4.42 Bus - Point Cloud Labelling at Location 2

4.4.2.4 Truck

In this last section, we show the results relating to the category of trucks. Just like buses, trucks also show a notable density of point clouds in their data due to the reflective properties of these vehicles. Based on the data shown in figure 4.43, it is

noticeable that the density of projected points on the truck is significantly higher compared to the car located behind the truck, as depicted in the image.

Figure 4.43 Truck - Point Cloud Projection onto Image

In continuation of the analysis, the next step involves overlaying the bounding box over the image, in conjunction with the projection of the point cloud, as shown in figure 4.44. Then, the point cloud can be extracted and assigned appropriate labels corresponding to their classes.

Figure 4.44 Truck - Point Cloud Projection and Bounding Box Plot onto Image

The identification of truck clusters in both scenarios illustrated in figure 4.45 is facilitated by the dense nature of the vehicle's point cloud, as plotted by the outcomes of point cloud clustering after DBSCAN processing.

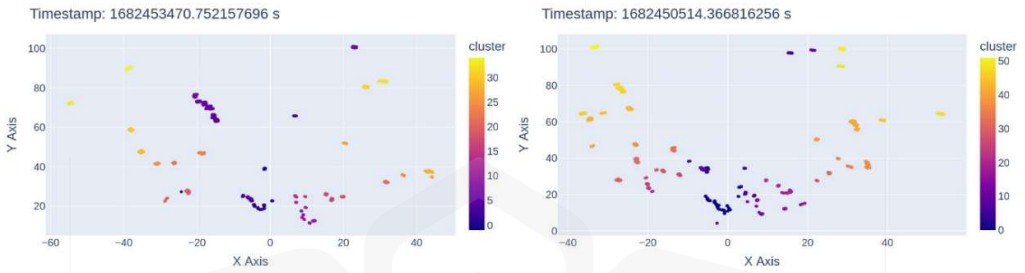


Figure 4.45 Truck - Scatterer Plot of Point Cloud Clusters

The visual representation of the custom dataset is shown in figure 4.46 and figure 4.47. These figures illustrate the point cloud of the truck, which has been appropriately labelled with the class type 'truck' and visually represented using the colour purple.

Figure 4.46 Truck - Point Cloud Labelling at Location 1

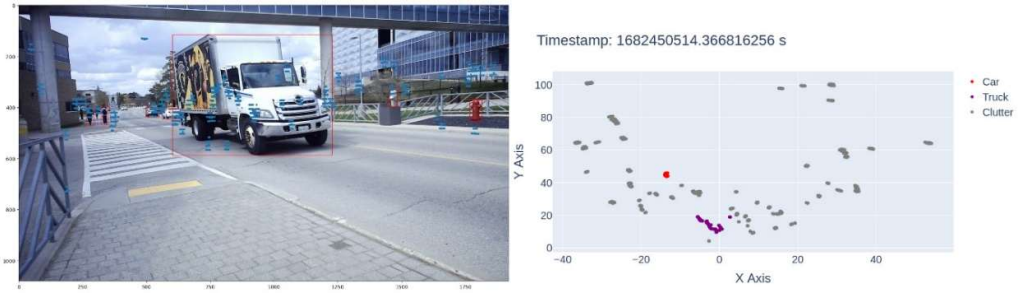


Figure 4.47 Truck - Point Cloud Labelling at Location 2

4.4.3 Discussion

Within this section, we provide an analysis of the results obtained through the process of projecting point clouds onto images, generating point cloud cluster plots, and applying point cloud labelling techniques by sensor data fusion. The findings of this study offer significant insights into the effectiveness of our methodology and underscore its potential for training and evaluating deep neural networks in the context of object classification. The point cloud projection approach employed in our study has exhibited remarkable precision in converting three-dimensional spatial data into two-dimensional visual representations. The projection of point cloud secures the preservation of valuable depth and location information of the targets, hence enhancing the input data for training deep neural networks.

The visually striking point cloud cluster plots attest to the effectiveness of our clustering algorithm. The well-defined clusters showcase the ability to discern objects within complex scenes accurately. This granularity is essential for classifier algorithm to distinguish objects and learn intricate features of the targets point cloud. Labelling of point clouds with object identifiers is a satisfying achievement. This facilitates supervised learning for neural network enabling them to train on labelled data, thereby improving their capacity for object recognition and classification. The satisfying object labelling results, lay the foundation for supervised deep learning applications. The enriched representations of point clouds offer a unique opportunity to leverage the

power of deep learning in complex, three-dimensional environments. This framework can enhance the performance of classification system in applications of road crossing monitoring.

In order to enhance the robustness of the neural network model, we conduct the collection of data from a different road crossing point. By conducting data collection in many locations, a more comprehensive range of real-world events, conditions, and variances can be represented. The presence of diversity is crucial in the process of training neural networks, since it enables them to efficiently generalise and function optimally over a wide range of scenarios.

In conclusion, our results underscore the potential of our point cloud projection methodology to drive advancements in deep neural network training and testing. The combination of accurate projection, cluster analysis, and labelling offers a valuable resource for researchers and engineers seeking to harness the capabilities of deep neural networks in the realm of three-dimensional data analysis and object recognition.

4.5 TASK FOUR: RADAR POINT CLOUD CLASSIFICATION

Upon completion of the data collecting and data processing stages, we successfully generated a custom dataset of radar point cloud data, encompassing point cloud cluster coordinates, doppler velocity, and corresponding labels. The dataset was divided into two separate subsets, namely the training set and the testing set. The training dataset will be utilised to train the deep neural network that we construct employing Keras, as elaborated upon in Chapter 3. The testing set is utilized to evaluate the efficacy of the neural network classifier in determining the cluster of the point cloud, specifically classifying it as either person, car, or big vehicle (bus and truck). The training set will make up 80% of the dataset, whereas the testing set will constitute 20% of the dataset.

The distribution of the data points for each class is presented in table 4.1 and the data distribution is depicted in figure 4.48 below.

Table 4.1 Number of Data Points for each Class

	Training	Testing	Total
Person	203 341	51 108	254 449
Car	183 126	45 632	228 758
Big Vehicle	93 085	23 248	116 333

Figure 4.48 Custom Dataset Class Distribution

It is observable that the targets within each class exhibit different proportions within the dataset. The dataset comprises pedestrians, accounting for 28% of the total, and cars, which constitute 25% of the dataset. Buses and trucks are classified as big vehicles and constitute only 13% of the dataset. Finally, the clutter class constitutes

34% of the dataset. This is due to the fact that many objects other than the intended objects presence in the point cloud data, such as background noise originating from trees which are classified as clutter. Based on an analysis of the dataset, it is obvious that there is an imbalance in the overall class distribution.

The imbalance in class frequencies observed in the point cloud dataset, which includes people, cars, buses, and trucks, can be linked to two primary sources. To begin with, it is common for the distribution of these classes in real-world scenarios to display a natural imbalance. In road crossing settings, there is often a greater majority of pedestrians and cars in comparison to buses and trucks. As a result, the imbalanced representation of the custom dataset may effectively reflect the real-world condition. Furthermore, it is important to acknowledge that data gathering processes have the potential to induce bias, which might subsequently result in class imbalance. This potential bias may arise due to several factors related to the data gathering process, including the selection of the timing of data collection. However, the overall custom dataset is satisfying enough for training and testing of deep neural networks.

In the subsequent subsection, we present the outcomes of the deep neural network classification and the overall performance of the model.

4.5.1 Target's Classification Performance

The objective of this work is to classify 4D point cloud data using deep learning method for the purpose of monitoring road crossings. Following the process of data fusion and the subsequent generation of a customised dataset, the next step involves training and testing the deep neural network. The model performed training with different number of layers for 10 epochs using a training dataset consisting of over 10,000 frames of point cloud data. The reason of using 10 epoch is due to the computational resource hardware limitation. 1 epoch have taken an average of 307

seconds making the training of deep neural network with large dataset requires to much time consume. Subsequently, predictions were made on the radar point cloud using the training set.

i. 3 layers neural network performance (64 – 32 – 4)

Figure 4.49 displays the confusion matrix, which provides an overview of the model's predictions for classifying point cloud clusters as either person, car, or big vehicle (consist of bus and truck). Based on an analysis of the confusion matrix, it can be observed that the model successfully predicts point cloud clusters of pedestrians as person with an accuracy of 97.77%. The model's predictions exhibit a misclassification rate of approximately 2.23 % in which the pedestrian cluster is erroneously identified as other objects. However, this level of misclassification is not deemed statistically significant.

Furthermore, the model demonstrates a high level of accuracy, accurately predicting the point cloud cluster of cars as cars with a precision of 97.67%. The outcome is satisfying for both the person and the car that is the target class. Despite the presence of misclassification of person and car class within the corresponding class, its impact is not deemed to be significant. In the classification of big vehicle which consists of bus and truck class types, a notable amount of misclassification is seen. The model for the big vehicle class type demonstrates a predictive accuracy of 75.25 %. Nevertheless, it is evident that a significant proportion of instances involve the incorrect classification of big vehicle as cars up to 24.41 %.

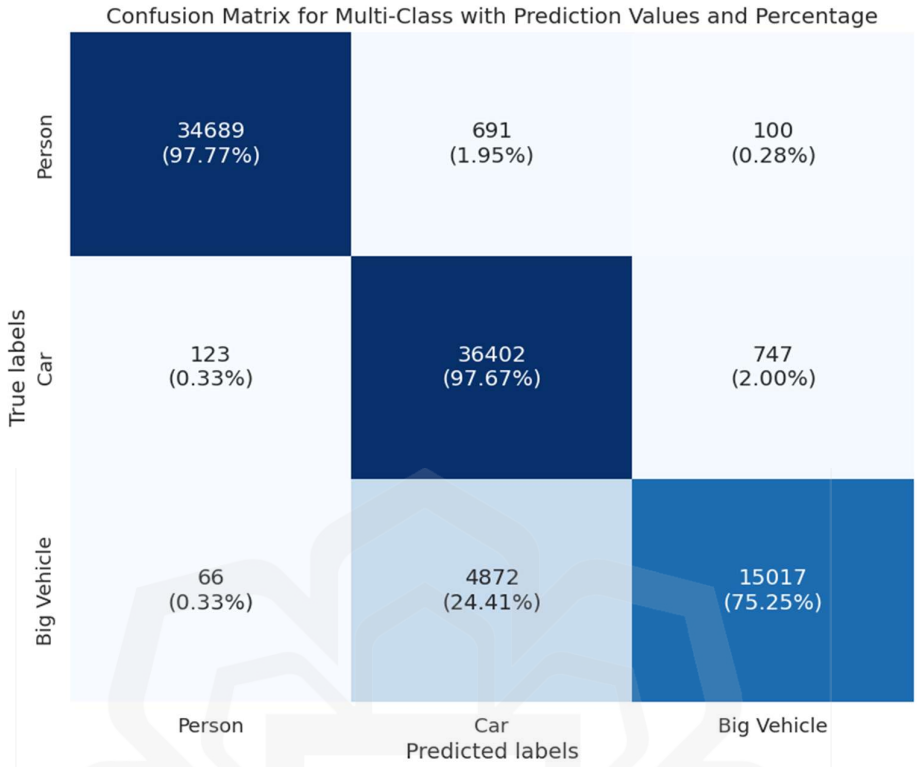


Figure 4.49 Confusion Matrix of 3 Layers Neural Network Model

ii. 4 layers neural network performance (64 – 32 – 16 – 4)

Next, we construct 4 layers of neural network to classify the object into respective classes. Figure 4.50 displays the confusion matrix, which provides an overview of the model's predictions for classifying point cloud clusters as either person, car, or big vehicle. Based on an analysis of the confusion matrix, it can be observed that the model successfully predicts point cloud clusters of pedestrians as person with an accuracy of 98.09%. The model's predictions exhibit a slightly higher classification accuracy as compared to the 3 layers neural network model. However, there are still a notable misclassification rate of approximately 1.91 % in which the pedestrian cluster is erroneously identified as other objects. Same as previous model, this level of misclassification is not statistically significant.

Next, the 4 layers model illustated a high level of accuracy, predicting the point cloud cluster of cars as cars with a precision of 94.79%. However, there is a significant different as compared to 3 layers neural network model with accuracy of 97.67%. Even so, the outcome is satisfying for both models of neural network in classifying cars class. Despite the presence of misclassification of person and car class within the corresponding class, its impact is not greatly affecting the performance of the model.

In addition, in the classification of big vehicle which consists of bus and truck class types, there are still a notable amount of misclassification is seen. The model for the big vehicle class type demonstrates a predictive accuracy of 78.41 %. Slightly higher compared to the previous model. Nevertheless, there is significant proportion of instances involve the incorrect classification of big vehicle as cars up to 21.08 %.

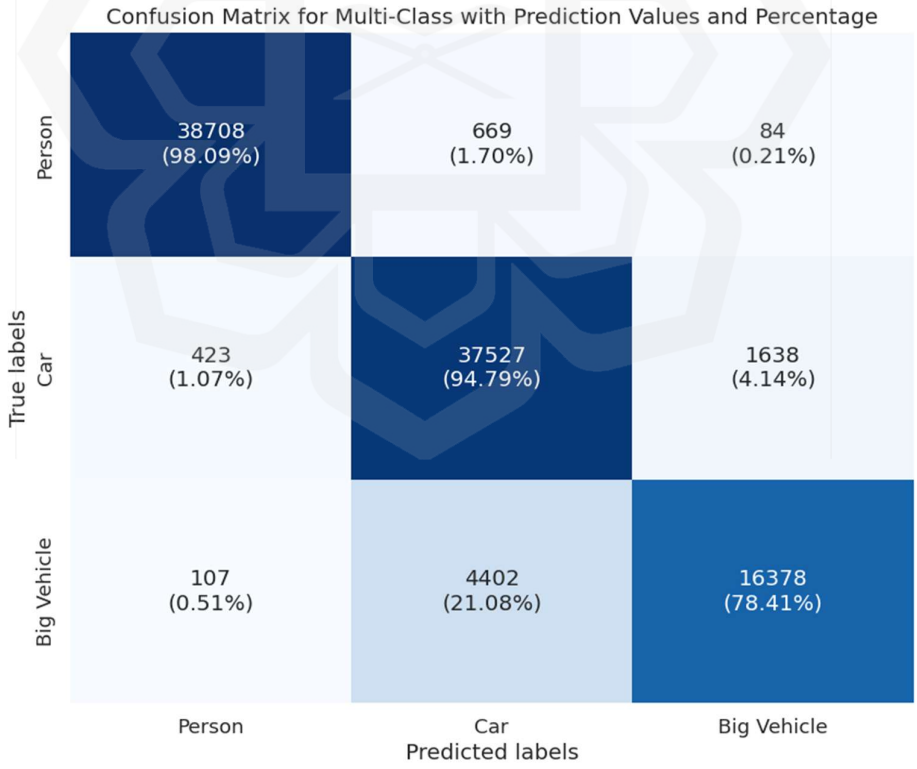


Figure 4.50 Confusion Matrix of 4 Layers Neural Network Model

iii. 5 layers neural network performance (64 – 64 – 32 – 6 – 4)

Lastly, we construct a 5-layer neural network model. Figure 4.51 displays the confusion matrix of the model's performance in classifying the objects into their respective classes. The confusion matrix reveals that the model accurately predicts pedestrian point cloud clusters as people, with an accuracy rate of 97.94%. The model's predictions misclassify pedestrians as other objects at a rate of approximately 2.06%. Furthermore, the model predicts the point cloud cluster of cars with an accuracy of 95.57%. This model exhibits a significant error in classifying cars as large vehicles, reaching up to 5.36%. This is the highest misclassification of cars compared to the other two previous models. Furthermore, this model has the highest accuracy in detecting large vehicles, up to 80% compared to previous models. Nevertheless, there are still significant misclassifications of large vehicles as cars, up to 18.91%.

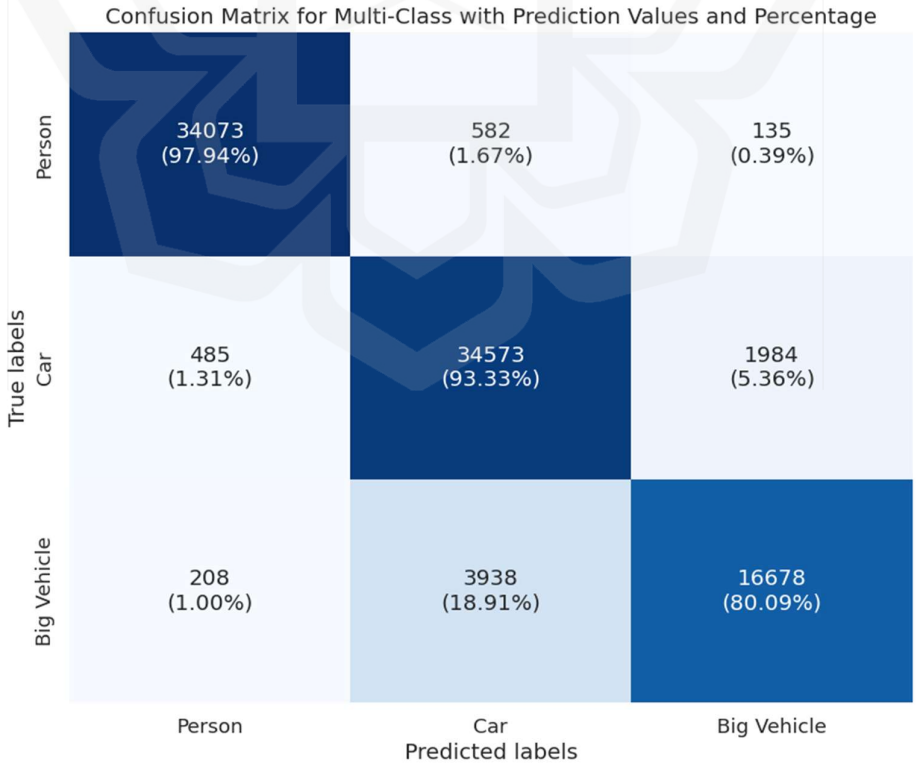


Figure 4.51 Confusion Matrix of 5 Layers Neural Network Model

4.5.2 Overall System Performance

The evaluation of a deep learning model's performance is made easier by the utilisation of a confusion matrix, which serves as a key element in gaining significant insights into the model's predictive capabilities. The confusion matrix provides valuable information for evaluating the success of a model. It allows us to calculate important metrics such as precision, accuracy, and F1 score, which provide a full assessment of the model's performance. Moreover, in the context of multi-class classifications, it is needed to consider the average metric that offers a comprehensive understanding of the performance demonstrated by a neural network model designed for the classification of multiple classes.

Firstly, precision is a quantitative measure that assesses the model's capacity to accurately identify true positives out of all positive predictions. The attribute of high precision signifies that the model exhibits a low frequency of misclassifying negative instances as positive ones. This characteristic has significant importance in situations where the occurrence of false positives is undesired. Next, recall is a metric that measures the model's capacity to accurately detect all instances that truly belong to a certain category. High recall becomes crucial in situations where it is important to record a significant number of true positive instances, even at the cost of accepting some false positive instances. Lastly, the F1-score is a measurement that effectively balances precision and recall. The metric offers a well-rounded evaluation of a model's efficacy, which is particularly advantageous in scenarios involving imbalanced datasets or where there is a requirement to achieve an equilibrium between precision and recall. F1-score is computed by taking the harmonic mean of precision and recall, providing a unified measure that reflects the balance between these two crucial elements of model efficacy.

Furthermore, based on the model evaluation table, the micro-average, macro-average, and weighted-average F1 scores is presence. These findings imply a considerable level of proficiency exhibited by the model in its entire form. The micro-

average score measures the effectiveness of the model in producing accurate predictions for all categories, without exhibiting bias towards any specific class. In this context, micro-average is the accuracy of the model performance. The macro-average F1 score is calculated by calculating the F1 score for each individual class and then taking the average. This metric provides evidence that the model exhibits consistent performance across different classes. The calculation of the weighted average score takes into account the issue of class imbalance by assigning a higher weight to classes that have a larger number of occurrences.

The tables below show the overall model evaluation metric that present the value of precision, recall and F1-score for each class and the average metrics for each model of neural network.

i. 3 layers neural network model evaluation report

Table 4.2 Model Evaluation Report (3 layers)

	Precision	Recall	F1-Score
Person	0.99	0.68	0.81
Car	0.87	0.80	0.83
Big Vehicle	0.95	0.65	0.77
Micro Average	0.93	0.72	0.81
Macro Average	0.94	0.71	0.80
Weighted Average	0.94	0.72	0.81

For the 3-layer model, which comprises layers with 64, 32, and 4 neurons respectively, the micro average F1-score achieved is 0.81. This model's relatively simple architecture suggests it is not overly complex, making it less prone to

overfitting and easier to train. However, the trade-off for this simplicity is that it might not be able to capture the complex patterns inherent in the data as effectively as deeper models. The model's ability to achieve a micro average F1-score of 0.81 indicates it is reasonably effective but might be missing some of the nuances in the data that could be captured with a more sophisticated architecture.

ii. 4 layers neural network model evaluation report

Table 4.3 Model Evaluation Report (4 layers)

	Precision	Recall	F1-Score
Person	0.99	0.76	0.86
Car	0.88	0.82	0.85
Big Vehicle	0.90	0.70	0.79
Micro Average	0.93	0.77	**0.84**
Macro Average	0.92	0.76	0.83
Weighted Average	0.93	0.77	0.84

The 4-layer model, which consists of layers with 64, 32, 16, and 4 neurons respectively, shows a significant improvement with a micro average F1-score of 0.84. The additional layer in this model helps to enhance its representational capacity, allowing it to capture more intricate features and relationships in the data. This higher score suggests that the model's increased depth enables better learning and generalization. The balance achieved in this model between complexity and performance indicates that it is well-suited for the given task, effectively leveraging the additional layer to improve predictive accuracy without introducing substantial overfitting.

iii. 5 layers neural network model evaluation report

Table 4.4 Model Evaluation Report (5 layers)

	Precision	Recall	F1-Score
Person	0.98	0.67	0.79
Car	0.88	0.76	0.82
Big Vehicle	0.89	0.72	0.79
Micro Average	0.92	0.71	0.80
Macro Average	0.92	0.71	0.80
Weighted Average	0.93	0.71	0.80

In contrast, the 5-layer model, with layers of 64, 64, 32, 16, and 4 neurons respectively, results in a micro average F1-score of 0.80, the lowest among the three models. Despite its more complex architecture, this model underperforms compared to its simpler counterparts. Several factors could contribute to this outcome. Firstly, the increased complexity can lead to overfitting, where the model captures noise in the training data rather than the underlying patterns, thus performing poorly on unseen test data. Moreover, deeper networks are more susceptible to the vanishing or exploding gradient problem, which can hamper effective training. This problem occurs when gradients used to update the weights during backpropagation become too small or too large, respectively, making the optimization process inefficient. Additionally, managing more parameters makes the optimization process more challenging, potentially leading to convergence issues or suboptimal training outcomes.

In summary, the 4-layer model's superior performance with a micro average F1-score of 0.84 suggests that it strikes the right balance between model complexity and performance. The additional layer allows it to learn more sophisticated features without overcomplicating the training process or leading to overfitting. The 3-layer

model, while simpler, performs reasonably well with a micro average F1-score of 0.81, indicating that it is effective but might benefit from a slight increase in complexity to improve performance further. On the other hand, the 5-layer model's performance, with a micro average F1-score of 0.80, suggests that its added complexity does not translate into better performance, possibly due to overfitting and optimization difficulties.

4.5.3 Discussion and Comparison with Other Method

In this section, comparing our study with previous work in the field, it becomes evident that there are several notable differences and improvements in our approach. The benchmark paper is a work done by Jin et al., (2020) where the work aims to evaluate the performance of radar-based object categorization in real-world traffic monitoring, demonstrating the efficacy of Gaussian Mixture Models (GMM) in processing radar point cloud data despite the difficulties caused by noise and clutter.

The benchmark study and this research contrast notably in terms of the radar devices utilised. The benchmark study utilised the Texas Instrument AWR1843BOOST radar chip board, but this study employed the TI AWR2243. This differentiation suggests that each research may gain advantages from varied capabilities and performance criteria that are inherent to their corresponding radar evaluation boards. Furthermore, the study done the experimental setup at different area: the benchmark research carried out its testing in a parking lot, positioning the radar equipment at an elevation of 3 m. This configuration has the potential to provide an alternative viewpoint and higher quality data compared to this study, which conducted experiments in a road crossing configuration with the radar positioned at a lower elevation of 1.5 m. This lower placement in this study is more commonly used for detecting and monitoring mobility modes at the pedestrian and vehicle level. However, as the limitation as discussed, this work employed the height up to 1m5 m above the ground due to maximum height of the tripod stand.

In terms of measurement resolutions, the benchmark study achieved a range resolution of 9 cm, a doppler resolution of 0.8 m/s, an azimuth angle resolution of 15°, and an elevation angle resolution of 28°. Conversely, this study proposed the used of radar with a coarser range resolution of 50 cm but significantly finer doppler resolution of 0.06 m/s, an azimuth angle resolution of 2°, and an elevation angle resolution of 4.7°. In addition, the types of objects considered in each study also vary. The benchmark study focused on pedestrians and cars, while this study expanded the scope to include buses and trucks. This broader range of transportation modes in this study indicates a more complex and challenging classification task, which could lead to more robust and versatile detection capabilities.

The effective detection area differed as well. The benchmark study had an effective radar detection area of up to 15 meters in range and 18 meters in cross-range. In contrast, this study extended the detection range to 40 meters, suggesting that radar system used in project is capable for longer-distance detection. This extended range could be crucial for detecting objects at a greater distance, providing an advantage in monitoring application. In terms of data collection, the benchmark study collected 8035 frames of training data over 13 minutes and 1222 frames of testing data over 2 minutes. However, in this study, the data is collected over 10,000 frames with a total measurement time of 2.5 hours, indicating a more extensive data collection process. This larger dataset could contribute to more robust training and testing, potentially leading to more reliable classification outcomes.

Data processing tools and methods also varied between the studies. The benchmark study utilized scikit-learn APIs to fit GMM for classification, an unsupervised learning method. In contrast, in this work, DBSCAN is used for clustering radar point cloud data and YOLOv7 for image processing and labelling, combining sensor fusion for data association. This approach is likely leads to more accurate and reliable classifications, as it combines unsupervised clustering with supervised labelling. Moreover, the classification methods employed were different. The benchmark study used GMM, which, as an unsupervised learning method, does not necessarily align predicted labels with ground truth. On the other hand, in this

work, we applied supervised learning using KERAS and TensorFlow for deep neural network models. This approach likely results in more accurate classifications, given the nature of supervised learning.

Given the differences in methodology and experimental setup the next part of the discussion is to evaluate the performance and outcomes of both studies. The result of the benchmark paper is compared with the result of the best performance model in this work which is the 4 layers model.

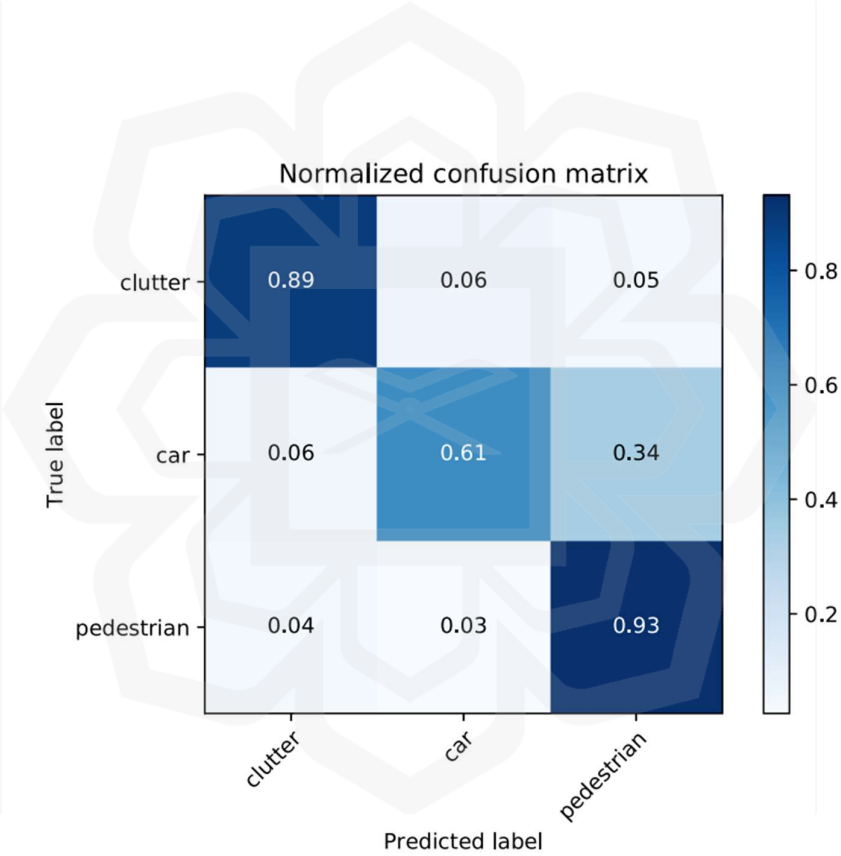


Figure 4.52 Confusion Matrix – Benchmark Paper

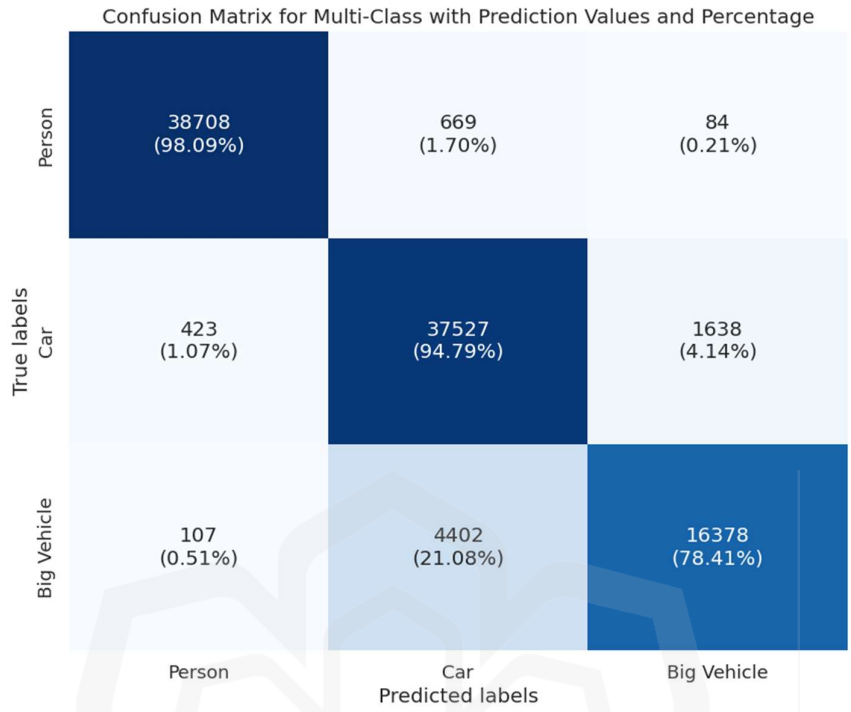


Figure 4.53 Confusion Matrix of 4 Layers Neural Network Model

The confusion matrix from the benchmark study presents the classification accuracy for three categories: clutter, car, and pedestrian. The true positive rates for clutter, car, and pedestrian are 89%, 61%, and 93%, respectively. The matrix reveals several key points. Firstly, the benchmark study exhibits a high accuracy in detecting clutter, with 89% of true clutter instances correctly classified. This high detection rate indicates that the model is effective in distinguishing irrelevant data points from actual objects. Secondly, the car detection accuracy is moderate at 61%. This lower accuracy suggests that the model has difficulty distinguishing cars from pedestrians, as evidenced by 34% of cars being misclassified as pedestrians. Lastly, pedestrian detection shows a high accuracy of 93%, indicating that the model performs well in identifying pedestrians within the radar point cloud data. Overall, the benchmark study's results indicate that while the model is proficient in identifying pedestrians and clutter, it struggles with accurately classifying cars, often confusing them with pedestrians.

In contrast, in this study's confusion matrix includes three categories: person, car, and big vehicle. The true positive rates for person, car, and big vehicle are 98.09%, 94.79%, and 78.41%, respectively. Several critical points emerge from this matrix. Firstly, the person detection rate is exceptionally high at 98.09%, demonstrating that the model is highly effective at identifying individuals within the radar data. Secondly, the car detection accuracy is also high at 94.79%, showing a significant improvement over the benchmark study. This improvement suggests that the use of YOLOv7 for image processing and DBSCAN for clustering, along with sensor fusion, has enhanced the model's capability to accurately classify cars. Lastly, detection of big vehicles (buses and trucks) shows a lower accuracy of 78.41%. This lower accuracy indicates that while the model performs well, there is still room for improvement in distinguishing between large vehicles and other categories.

In terms of performance and results, the confusion matrix in this study shows marked improvements in the detection of cars and persons compared to the benchmark study. Specifically, the car detection rate of 94.79% is significantly higher than the benchmark study's 61%, highlighting the effectiveness of this work model's approach. The person detection accuracy in this study (98.09%) also surpasses the pedestrian detection accuracy in the benchmark study (93%), indicating superior performance in this category as well. Although there is additional challenge of detecting big vehicles in this work, the accuracy rate of 78.41% suggests that while effective, there is still potential for further refinement in distinguishing large vehicles from other categories.

Table 4.5 Model Evaluation Report – Benchmark Paper

	Precision	Recall	F1-Score
Person	0.85	0.93	0.89
Car	0.88	0.61	0.72
Clutter	0.71	0.89	0.79

Table 4.6 Model Evaluation Report (4 layers)

	Precision	Recall	F1-Score
Person	0.99	0.76	0.86
Car	0.88	0.82	0.85
Big Vehicle	0.90	0.70	0.79
Micro Average	0.93	0.77	**0.84**
Macro Average	0.92	0.76	0.83
Weighted Average	0.93	0.77	0.84

In the benchmark study, the precision, recall, and F1-score for detecting persons were 0.85, 0.93, and 0.89, respectively. This indicates a high recall, suggesting that the benchmark model was effective at identifying most instances of persons, but the precision was slightly lower, meaning there were some false positives. For car detection, the precision was 0.88, recall was 0.61, and F1-score was 0.72, showing a significant drop in recall compared to precision. This suggests that the benchmark model struggled more with correctly identifying cars, possibly confusing them with other categories. Clutter detection had a precision of 0.71, recall of 0.89, and an F1-score of 0.79, indicating a higher tendency to identify clutter correctly but with a relatively lower precision, reflecting a notable number of false positives.

In comparison, this work demonstrated improved performance metrics across all categories. The precision, recall, and F1-score for person detection were 0.99, 0.76, and 0.86, respectively. While the recall is lower compared to the benchmark study, the precision is significantly higher, suggesting the model produces fewer false positives when detecting persons. For car detection, the model achieved a precision of 0.88, recall of 0.82, and F1-score of 0.85, showing a balanced and robust performance. This represents a substantial improvement in recall over the benchmark study, indicating the model's enhanced capability to correctly identify cars. For big vehicle detection, the precision was 0.90, recall was 0.70, and F1-score was 0.79. Although this category

was not included in the benchmark study, the metrics indicate a solid performance in identifying large vehicles, albeit with a lower recall compared to precision.

This study also reports micro, macro, and weighted averages for the classification performance. The micro average precision, recall, and F1-score are 0.93, 0.77, and 0.84, respectively. These metrics provide an overall performance evaluation considering all instances across categories, showing a strong general performance with high precision and good recall. The macro averages, which consider the performance across categories equally, were 0.92, 0.76, and 0.83 for precision, recall, and F1-score, respectively. These averages reflect the model's balanced performance across different categories. The weighted averages, which take into account the proportion of each category, were 0.93, 0.77, and 0.84, respectively, indicating that the model maintains high performance even when the category distribution is considered.

In summary, the comparative analysis shows that this study's classification model demonstrates superior precision across all categories compared to the benchmark study. This indicates that the model in this work produces fewer false positives, particularly for person and car detection. The recall for car detection in this work is significantly improved, highlighting the model's enhanced capability to correctly identify cars. While the recall for person detection is lower than the benchmark, the higher precision compensates for this by reducing false positives. The inclusion of big vehicle detection in this study adds a layer of complexity, with good performance metrics indicating the model's capability to handle more diverse categories. The overall averages further underscore the robustness and effectiveness of this work classification model, showing high precision and balanced performance across all categories.

4.5.4 Challenges

After examining the models' performance, it is clear that there are multiple elements that contribute to the accuracy challenges, specifically the confusion between car and big vehicle categories. Three main factors have been recognised as the key causes of this challenge:

i. Misclassification during dataset labelling using sensor fusion

In certain situations, during the processing of images using the YOLO algorithm, difficulties arise in accurately distinguishing between a car, bus, and truck, resulting in inaccurate predictions for the bounding box. This example of mislabelling due to YOLO can be illustrated as shown in figure 4.54. The presence of confusion among the classes can be attributed to the larger number of mislabelling in the custom dataset. Given the extensive dataset comprising over 10,000 frames of point cloud data, it is highly impractical to manually resolve mislabelling of point cloud clusters and assign them to their appropriate classes. Given the time constraints associated with this project, the mislabelling resulting from the YOLO algorithm will be addressed as a component of future enhancements in this study.

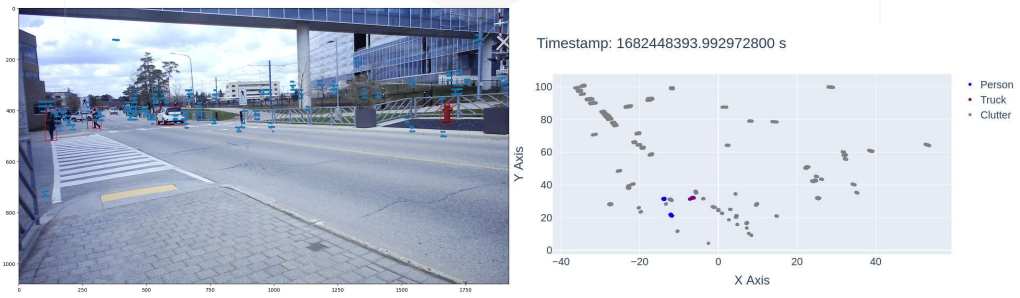


Figure 4.54 Example of Mislabelling of Target Due to YOLOv7 Classification Predictions

ii. Imbalance Dataset

The second factor is associated with the disparity in the dataset. The number of data points for trucks and buses is considerably lower in comparison to cars and pedestrians as displayed in figure 4.48. This disparity develops due to the reduced occurrence of trucks and buses in the area of the road crossing during the measurement. An unbalanced dataset can distort the learning process of the model, since it has a tendency of bias to the more commonly occurring classes. Consequently, the model's ability to accurately identify these often-occurring classes improves, while its performance declines when it comes to less commonly represented classes such as trucks and buses.

4.6 SUMMARY

This chapter provides the test outcomes conducted throughout this study, covering the preliminary test, radar performance test, collecting data, data processing and labelling, and object classification. The preliminary test was conducted in an indoor space area while the radar performance test was conducted at a controlled open space environment, and lastly, the data collection experiments were conducted at two different road crossing areas. The results of the conducted experiments have been carefully analysed and discussed.

The preliminary test conducted is a great starting point for this work. As the radar in this work is developed on our own, learning the compatibility of the radar hardware and the computer software is crucial to ensure the connectivity between the platforms are established. From here on, we can move forward in achieving the first object of this project which is to investigate the high-resolution detection capability of the 4D Radar Imaging with different types of targets vary in size.

This led to test two which is the radar performance test. In this work, we have discussed 4 different types of targets consist of a pedestrian, multiple pedestrians, motorcycle and car. The detection and performance of the radar in detecting the object have been discussed and shows great satisfaction. The main outcome of this test also contributes for the next phase of this work which is to conduct measurement at road crossing areas. From the radar performance test, we can propose and justify the position of the radar during the measurement and the position of the road crossing areas respect to the radar view. This part is crucial because, without proper measurement at the road crossing areas, we cannot have progress towards the other two main objective of this work.

From here on, test three which is data collection at road crossing areas have been conducted successfully. Based on this measurement, we can work towards data fusion that associate the radar point cloud data and camera data to create a customized dataset. The measurements outcome manage to produce a dataset with more than 10,000 frames of data which is sufficient for testing and training deep neural network.

The neural network developed using Keras and Tensorflow framework are used for training the dataset. We developed three different models of neural network using three different layers in each model (consist of 64-32-4, 64-32-16-4 and 64-64-32-16-4). From this step, we manage to achieved the second objective of this work which is to develop object detection and classification model using deep learning method.

Lasty, after the training of the model have been accomplished, we used the testing dataset to study the performance of the neural network models. The outcomes of the testing have shown that the second model with 4 layers of neural network performed the best with accuracy of 84% as compared to the other two models. this is a very satisfactory result for this work, and we have managed to achieve the third objective which is to test the detection and classification model by using data measurement on road crossing areas.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 CONCLUSION

This work is highly motivated to solve challenges relating to point cloud classification using a deep neural network as part of a monitoring system. With the presence of high-resolution radar imaging in the automotive market, given its advantages, there is a trend in research work towards developing a radar-based monitoring system with object classification capability. However, researchers face the challenges of object classification using radar imaging due to the sparsity characteristic of the point. In this work, we propose a method to solve this challenge through sensor data fusion. We integrate the usage of radar imaging and cameras to create a customised dataset of point clouds and use it to train a deep neural network. In this dataset, all the spatial relationships and information of the radar point cloud data are reserved. Hence, the deep neural network employed in this work could learn the complex relationships of the point cloud data and make predictions.

This study has demonstrated the efficacy of custom dataset by integrating 4D radar imaging and camera data for classifying radar point cloud data in road crossing environments. Utilizing the TI mmWave radar evaluation board AWR2243, combined with the ROS and Python programming for data processing and a USB camera for video reference, we were able to capture and process high-resolution radar point cloud data. Our setup, positioned at 1.5 meters high, with high-resolution radar imaging capabilities: range resolution of 50 cm, Doppler resolution of 0.06 m/s, azimuth angle resolution of 2 degrees, and elevation angle resolution of 4.7 degrees, due to the implementation of FMCW waveform design and MIMO direction-of-arrive (DOA) algorithms.

There are also questions raised in terms of the capability of radar-based monitoring systems to maintain their accuracy for multiple target detections. In the application of monitoring, especially in road crossing areas, the number of target presences at a time may be very high. Due to this, the capability of a radar sensor to detect all the targets must be accurate and efficient. Based on the experimental work done at the road crossing area, the radar imaging employed in this work has shown satisfactory performance in detecting all the targets in the area. This also includes multiple targets of different sizes at a time. The radar manages to maintain its accuracy and performance throughout the experiments and produces a great radar point cloud dataset.

Moreover, as part of the monitoring system, the radar imaging employed is expected to be able to detect and classify a variety of objects presence, which is not limited to pedestrians but also cars and other vehicles. Based on the outcomes of the classification system using the deep neural network deployed in this work, the algorithm has shown great performance in making predictions for all class targets. Even though there is notable confusion between cars, buses, and trucks, it can be enhanced by collecting more data for the imbalanced distribution of the customised dataset used in this work and improving on the mislabelling of targets in the dataset due to YOLOv7.

In comparison, the benchmark study employed a Gaussian Mixture Model (GMM) for classification, which, while effective, did not match the adaptability and accuracy of our deep neural network approach. Our use of YOLOv7 for object detection and DBSCAN for clustering was critical in accurately associating radar and camera data, providing a robust and scalable solution for real-time applications. The benchmark's confusion matrix indicated a higher misclassification rate for cars and pedestrians, whereas our approach demonstrated superior precision and recall across all classes, underscoring the effectiveness of our method.

Based on all of this discussion, we can conclude that all of the objectives of this work have been achieved. Firstly, to investigate the high-resolution detection capabilities of the 4D radar imaging with different types of targets. The performance of 4D radar imaging in detecting multiple targets of different sizes and classes in the controlled area as well as at the road crossing areas has shown the capability of the radar in providing extensive information about the targets, including their localization and doppler velocity. Next, develop a detection and classification model using deep learning. Powered by Keras and TensorFlow, an open-source deep learning platform, this work manages to build a feedforward neural network that takes point cloud data input and makes predictions to classify the point data into its respective target class. Lastly, to test the detection and classification model using data measurement on road crossing areas. Based on the outcomes of the model's evaluation performance, the model manages to achieve up to 84% overall accuracy. However, there is still room in this work that can be used for further improvement in future research on 4D radar imaging object detection and classification using deep neural networks.

5.2 RECOMMENDATIONS FOR FUTURE RESEARCH

Recommendations are crucial for directing future research and identifying areas that need more investigation and development. In this thesis, there are three recommendations that need to be highlighted for the purposes of improving this project in future work.

i. Improve labelling accuracy of customized dataset.

The primary area for future enhancement in this study, aimed at enhancing the precision of the deep neural network model's predictions, involves improving the accuracy of dataset labelling. Special attention should be given to rectifying mislabelling problems arising from YOLOv7 bounding box predictions. The

prevention of problems related to misclassification in object detection systems can be significantly improved by doing research and deploying new algorithms for image object recognition for example YOLOv8, as well as by refining the labelling process of the customised dataset. These measures have the potential to enhance the overall performance of the object detection system.

ii. Dataset expansion and diversification.

In order to enhance the efficiency of the object classification model, it is important to broaden the dataset. The proposed extension should involve expanding the frequency of occurrences for particular categories, such as cars, buses, and trucks. In addition, the development of a deep neutral network will greatly benefit from the creation of a more diverse dataset that covers a greater variety of target types, such as motorcycles and bicycles.

iii. Real-time radar integration

A promising potential for future research includes the integration of real-time radar systems with object classification system at road crossing areas. The integration of radar imaging and deep neural network algorithm has the potential to greatly boost the capabilities of object classification and tracking, particularly in complex crossing situations. By integrating radar data with pre-existing object classification systems, researchers can investigate potential enhancements in system performance and safety protocols at junctions, hence enhancing their resilience and dependability for practical implementations.

REFERENCES

Lim, T. Y., Ansari, A., Major, B., Fontijne, D., Hamilton, M., Gowaikar, R., & Subramanian, S. (2019, December). Radar and camera early fusion for vehicle detection in advanced driver assistance systems. In Machine learning for autonomous driving workshop at the 33rd conference on neural information processing systems (Vol. 2, No. 7).

Cho, H., Seo, Y. W., Kumar, B. V., & Rajkumar, R. R. (2014, May). A multi-sensor fusion system for moving object detection and tracking in urban driving environments. In 2014 IEEE international conference on robotics and automation (ICRA) (pp. 1836-1843). IEEE.

Oloumi, D. (2016). Ultra-wideband synthetic aperture radar imaging: Theory and applications

P. Z. Peebles, Radar principles. A Wiley-Interscience publication, New York, NY: Wiley, 1998. OCLC: 833158540.

S. Sun and Y. D. Zhang, "4D Automotive Radar Sensing for Autonomous Vehicles: A Sparsity-Oriented Approach," in IEEE Journal of Selected Topics in Signal Processing, vol. 15, no. 4, pp. 879-891, June 2021

Drummer. (2016, March 8). Synthetic Aperture Radar (SAR) and Inverse SAR (ISAR) Enable an Amazing Range of Remote Sensing Applications | The Lyncean Group of San Diego. https://lynceans.org/all-posts/synthetic-aperture-radar-sar-and-inverse-sar-isar-enable-an-amazing-range-of-remote-sensing-applications/

Chan, Y. K., & Koo, V. (2008). An introduction to synthetic aperture radar (SAR). Progress In Electromagnetics Research B, 2, 27-60.

Goodman, D., & Piro, S. (2013). GPR remote sensing in archaeology (Vol. 9, p. 233). New York: Springer. (GPR PUNYA)

Zhao, W., Tian, G., Forte, E., Pipan, M., Wang, Y., Li, X., ... & Liu, H. (2015). Advances in GPR data acquisition and analysis for archaeology. Geophysical Journal International, 202(1), 62-71.

Rasol, M., Pais, J. C., Pérez-Gracia, V., Solla, M., Fernandes, F. M., Fontul, S., ... & Assadollahi, H. (2022). GPR monitoring for road transport infrastructure: A systematic review and machine learning insights. Construction and Building Materials, 324, 126686

Ékes, C., & Neducza, B. (2012, June). Robot mounted GPR for pipe inspection. In 2012 14th International Conference on Ground Penetrating Radar (GPR) (pp. 160-164). IEEE

Everett, M. E. (2013). Near-surface applied geophysics. Cambridge University Press

Hasch, J. (2015, April). Driving towards 2020: Automotive radar technology trends. In 2015 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM) (pp. 1-4). IEEE.

World Health Organization: WHO. (2022). Road traffic injuries. www.who.int. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Waldschmidt, C., Hasch, J., & Menzel, W. (2021). Automotive radar—From first efforts to future systems. IEEE Journal of Microwaves, 1(1), 135-148.

Gao, X., Xing, G., Roy, S., & Liu, H. (2019, November). Experiments with mmwave automotive radar test-bed. In 2019 53rd Asilomar conference on signals, systems, and computers (pp. 1-6). IEEE.

Bilik, I., Longman, O., Villeval, S., & Tabrikian, J. (2019). The rise of radar for autonomous vehicles: Signal processing solutions and future research directions. IEEE signal processing Magazine, 36(5), 20-31.

Ragonese, E., Papotto, G., Nocera, C., Cavarra, A., & Palmisano, G. (2022). Cmos automotive radar sensors: mm-wave circuit design challenges. IEEE Transactions on Circuits and Systems II: Express Briefs, 69(6), 2610-2616.

Bilik, I., Villeval, S., Brodeski, D., Ringel, H., Longman, O., Goswami, P., ... & Liu, S. (2018, April). Automotive multi-mode cascaded radar data processing embedded system. In 2018 IEEE Radar Conference (RadarConf18) (pp. 0372-0376). IEEE.

Lawrence, N. P., Ng, B. W. H., Hansen, H. J., & Abbott, D. (2017). 5G terrestrial networks: Mobility and coverage—Solution in three dimensions. IEEE Access, 5, 8064-8093.

Jankiraman, M. (2018). FMCW Radar Design. Artech House.

Vavriv, D. M., Bezvesilniy, O. O., Volkov, V. A., Kravtsov, A. A., & Bulakh, E. V. (2015, April). Recent advances in millimeter-wave radars. In 2015 International Conference on Antenna Theory and Techniques (ICATT) (pp. 1-6). IEEE.

Venon, A., Dupuis, Y., Vasseur, P., & Merriaux, P. (2022). Millimeter wave FMCW radars for perception, recognition and localization in automotive applications: A survey. IEEE Transactions on Intelligent Vehicles, 7(3), 533-555.

Kees, N., Schmidhammer, E., & Detlefsen, J. (1995, May). Improvement of angular resolution of a millimeterwave imaging system by transmitter location multiplexing. In IEEE NTC, Conference Proceedings Microwave Systems Conference (pp. 105-108). IEEE.

Kim, H., You, S., Jeong, B. J., & Byun, W. (2020, October). Azimuth angle resolution improvement technique with neural network. In 2020 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 1384-1387). IEEE.

Xu, R. (2022). Using 3D Imaging Radar for Indoor Localization and Mapping (Doctoral dissertation, Carnegie Mellon University Pittsburgh, PA).

Cheng, Y., Su, J., Chen, H., & Liu, Y. (2021, June). A new automotive radar 4D point clouds detector by using deep learning. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8398-8402). IEEE.

Sun, S., & Zhang, Y. D. (2021). 4D automotive radar sensing for autonomous vehicles: A sparsity-oriented approach. IEEE Journal of Selected Topics in Signal Processing, 15(4), 879-891.

Stateczny, A., Kazimierski, W., Gronska-Sledz, D., & Motyl, W. (2019). The empirical application of automotive 3D radar sensor for target detection for an autonomous surface vehicle's navigation. Remote Sensing, 11(10), 1156.

Lin, J. J., Li, Y. P., Hsu, W. C., & Lee, T. S. (2016). Design of an FMCW radar baseband signal processing system for automotive application. SpringerPlus, 5, 1-16.

Skvortsov, V., Lee, K. M., & Yang, S. E. (2012, November). Inexpensive radar-based surveillance: Experimental study. In 2012 9th International Conference & Expo on Emerging Technologies for a Smarter World (CEWIT) (pp. 1-6). IEEE.

Tong, Z., Renter, R., & Fujimoto, M. (2015, October). Fast chirp FMCW radar in automotive applications. In IET International Radar Conference 2015 (pp. 1-4). IET.

Li, X., Wang, X., Yang, Q., & Fu, S. (2021). Signal processing for TDM MIMO FMCW millimeter-wave radar sensors. IEEE Access, 9, 167959-167971

Instruments, T. (2020). The fundamentals of millimeter wave radar sensors.

Amaral, V., Marques, F., Lourenço, A., Barata, J., & Santana, P. (2016). Laser-based obstacle detection at railway level crossings. Journal of Sensors, 2016.

Zhang, S., Wang, C., Chan, S. C., Wei, X., & Ho, C. H. (2014). New object detection, tracking, and recognition approaches for video surveillance over camera network. IEEE sensors journal, 15(5), 2679-2691.

Bassyouni, A. M. (2019, October). A Mobile 4D Imaging Radar System Network for Home Land Security. In 2019 IEEE International Symposium on Phased Array System & Technology (PAST) (pp. 1-7). IEEE.

Krišto, M., Ivasic-Kos, M., & Pobar, M. (2020). Thermal object detection in difficult weather conditions using YOLO. IEEE access, 8, 125459-125476.

Zhang, J., Xiao, W., Coifman, B., & Mills, J. P. (2020). Vehicle tracking and speed estimation from roadside lidar. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 5597-5608.

Gao, H., Cheng, B., Wang, J., Li, K., Zhao, J., & Li, D. (2018). Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment. IEEE Transactions on Industrial Informatics, 14(9), 4224-4231.

Cai, H., Li, F., Gao, D., Yang, Y., Li, S., Gao, K., ... & Huang, Z. (2020, October). Foreign objects intrusion detection using millimeter wave radar on railway crossings. In 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 2776-2781). IEEE.

Govoni, M., Guidi, F., Vitucci, E. M., Degli Esposti, V., Tartarini, G., & Dardari, D. (2015). Ultra-wide bandwidth systems for the surveillance of railway crossing areas. IEEE Communications Magazine, 53(10), 117-123.

Xu, S., Wang, J., & Yarovoy, A. (2018, September). Super resolution DOA for FMCW automotive radar imaging. In 2018 IEEE Conference on Antenna Measurements & Applications (CAMA) (pp. 1-4). IEEE.

Cho, H., Seo, Y. W., Kumar, B. V., & Rajkumar, R. R. (2014, May). A multi-sensor fusion system for moving object detection and tracking in urban driving environments. In 2014 IEEE International Conference on Robotics and Automation (ICRA) (pp. 1836-1843). IEEE.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. Electronic Markets, 31(3), 685-695.

Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A systematic review on supervised and unsupervised machine learning algorithms for data science. Supervised and unsupervised learning for data science, 3-21.

Taye, M. M. (2023). Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. Computers, 12(5), 91.

Jin, F., Sengupta, A., Cao, S., & Wu, Y. J. (2020, April). Mmwave radar point cloud segmentation using gmm in multimodal traffic monitoring. In 2020 IEEE International Radar Conference (RADAR) (pp. 732-737). IEEE.

Reddy, C. K. (2018). Data clustering: algorithms and applications. Chapman and Hall/CRC.

Schubert, E., & Gertz, M. (2018, August). Improving the Cluster Structure Extracted from OPTICS Plots. In LWDA (pp. 318-329).

Campello, Ricardo J. G. B. and Moulavi, Davoud and Sander, Joerg, “Density-based clustering based on hierarchical density estimates,” in Advances in Knowledge Discovery and Data Mining. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 160–172.

Stewart, G., & Al-Khassaweneh, M. (2022). An implementation of the HDBSCAN* clustering algorithm. Applied Sciences, 12(5), 2405.

Malzer, C., & Baum, M. (2020, September). A hybrid approach to hierarchical density-based cluster selection. In 2020 IEEE international conference on multisensor fusion and integration for intelligent systems (MFI) (pp. 223-228). IEEE.

Plakalovic, A. (2023, July 5). Clustering Algorithms: DBSCAN vs. OPTICS. Atlantbh Sarajevo. https://www.atlantbh.com/clustering-algorithms-dbscan-vs-optics/

M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander. “OPTICS: Ordering Points To Identify the Clustering Structure”. In: Proc. SIGMOD. 1999, pp. 49– 60.

Islam, M. R., Jenny, I. J., Nayon, M., Islam, M. R., Amiruzzaman, M., & Abdullah-Al-Wadud, M. (2021, August). Clustering algorithms to analyze the road traffic crashes. In 2021 International Conference on Science & Contemporary Technologies (ICSCT) (pp. 1-6). IEEE.

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. Information Sciences, 622, 178-210.

Jahwar, A. F., & Abdulazeez, A. M. (2020). Meta-heuristic algorithms for K-means clustering: A review. PalArch's Journal of Archaeology of Egypt/Egyptology, 17(7), 12002-12020.

Bushra, A. A., & Yi, G. (2021). Comparative analysis review of pioneering DBSCAN and successive density-based clustering algorithms. IEEE Access, 9, 87918-87935.

Wang, L., Chen, P., Chen, L., & Mou, J. (2021). Ship AIS trajectory clustering: An HDBSCAN-based approach. Journal of Marine Science and Engineering, 9(6), 566.

Nanni, M., Pedreschi, D.: Time-focused clustering of trajectories of moving objects. J. Intell. Inf. Syst. 27, 267–289 (2006)

Elbatta, Mohammad & Bolbol, Raed & Ashour, Wesam. (2012). A Vibration Method for Discovering Density Varied Clusters. ISRN Artificial Intelligence. 2012. 10.5402/2012/723516.

Deng, Z., Hu, Y., Zhu, M., Huang, X., & Du, B. (2015). A scalable and fast OPTICS for clustering trajectory big data. Cluster Computing, 18, 549-562.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

Terven, J., & Cordova-Esparza, D. (2023). A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond. arXiv preprint arXiv:2304.00501.

Padilla, R., Passos, W. L., Dias, T. L., Netto, S. L., & da Silva, E. A. (2021). A comparative analysis of object detection metrics with a companion open-source toolkit. Electronics, 10(3), 279.

Olorunshola, O. E., Irhebhude, M. E., & Evwiekpaefe, A. E. (2023). A comparative study of YOLOv5 and YOLOv7 object detection algorithms. Journal of Computing and Social Informatics, 2(1), 1-12.

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7464-7475).

Elavarasu, M., & Govindaraju, K. (2024). Unveiling the Advancements: YOLOv7 vs YOLOv8 in Pulmonary Carcinoma Detection. Journal of Robotics and Control (JRC), 5(2), 459-470.

Rangari, A. P., Chouthmol, A. R., Kadadas, C., Pal, P., & Singh, S. K. (2022, December). Deep Learning based smart traffic light system using Image Processing with YOLO v7. In 2022 4th International Conference on Circuits, Control, Communication and Computing (I4C) (pp. 129-132). IEEE.

Lakshmi, Mrs & D, Sridhar. (2024). AN ANALOGY ANALYSIS OF THE OBJECT DETECTION ALGORITHMS USING YOLOV5, YOLOV7, AND YOLOV8.

Hussain, M. (2023). YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. Machines, 11(7), 677.

Ashtiani, F., Geers, A. J., & Aflatouni, F. (2022). An on-chip photonic deep neural network for image classification. Nature, 606(7914), 501-506.

Grattarola, D., & Alippi, C. (2021). Graph neural networks in tensorflow and keras with spektral [application notes]. IEEE Computational Intelligence Magazine, 16(1), 99-106.

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). {TensorFlow}: a system for {Large-Scale} machine learning. In 12th USENIX symposium on operating systems design and implementation (OSDI 16) (pp. 265-283).

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 8, 1-74.

Zhan, Q., & Yu, L. (2011, June). Objects classification from laser scanning data based on multi-class support vector machine. In 2011 International Conference on Remote Sensing, Environment and Transportation Engineering (pp. 520-523). IEEE.

Wang, J., Li, G., Jiao, J., Zhao, Z., & Li, J. (2021, December). Vehicle Classification via Multi-dimension Feature Extraction with Millimeter Wave Radar. In 2021 CIE International Conference on Radar (Radar) (pp. 1336-1339). IEEE.

Somvanshi, M., Chavan, P., Tambade, S., & Shinde, S. V. (2016, August). A review of machine learning techniques using decision tree and support vector machine.

In 2016 international conference on computing communication control and automation (ICCUBEA) (pp. 1-7). IEEE.

Lim, S., Jung, J., Lee, B. H., Choi, J., & Kim, S. C. (2022). Radar Sensor-Based Estimation of Vehicle Orientation for Autonomous Driving. IEEE Sensors Journal, 22(22), 21924-21932.

Wen, L. H., & Jo, K. H. (2021). Fast and accurate 3D object detection for lidar-camera-based autonomous vehicles using one shared voxel-based backbone. IEEE access, 9, 22080-22089.

Wu, J., Zhu, Z., & Wang, H. (2021, September). Human Detection and Action Classification Based on Millimeter Wave Radar Point Cloud Imaging Technology. In 2021 Signal Processing Symposium (SPSympo) (pp. 294-299). IEEE.

Gong, P., Wang, C., & Zhang, L. (2021, July). Mmpoint-gnn: Graph neural network with dynamic edges for human activity recognition through a millimeter-wave radar. In 2021 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE.

Kim, Y., Alnujaim, I., & Oh, D. (2021). Human activity classification based on point clouds measured by millimeter wave MIMO radar with deep recurrent neural networks. IEEE Sensors Journal, 21(12), 13522-13529.

Ji, Z., & Prokhorov, D. (2008, June). Radar-vision fusion for object classification. In 2008 11th International Conference on Information Fusion (pp. 1-7). IEEE.

Ji, Z., Luciw, M., Weng, J., & Zeng, S. (2010). Incremental online object learning in a vehicular radar-vision fusion framework. IEEE Transactions on Intelligent Transportation Systems, 12(2), 402-411.

Oh, J., Kim, K. S., Park, M., & Kim, S. (2018, November). A comparative study on camera-radar calibration methods. In 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV) (pp. 1057-1062). IEEE.

Boesch, G. (2023, November 21). YOLOv7: A Powerful Object Detection Algorithm (2024 Guide). Viso.ai. https://viso.ai/deep-learning/yolov7-guide/#:~:text=As%20previously%20shown%20in%20the

Boncolmo, S. D., Calaquian, E. V., & Caya, M. V. C. (2021, November). Gender Identification Using Keras Model Through Detection of Face. In 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) (pp. 1-6). IEEE.

Si, M., Tarnoczi, T. J., Wiens, B. M., & Du, K. (2019). Development of predictive emissions monitoring system using open source machine learning library–keras: A case study on a cogeneration unit. IEEE Access, 7, 113463-113475.

Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al-Nemrat, A., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. IEEE Access, 7, 41525-41550.

Chen, S. H., Hung, C. S., Wang, J. Y., Chen, C. H., & Hsu, K. C. (2021, December). The Implementation of Hybrid Electric Vehicle Battery Fault and Abnormal Early Warning System Using Keras Neural Network Technology. In 2021 9th International Conference on Orange Technology (ICOT) (pp. 1-6). IEEE.

LIST OF PUBLICATIONS

1. Abd Aziz, L. S. W., Isa, F. N. M., Abd Rahman, F., Narayanan, A. H., Alghooneh, A. R., & Shaker, G. (2025). 4D Radar Imaging and Camera Fusion for Road Crossing Detection and Classification Using Deep Learning. *IIUM Engineering Journal*, *26*(1), 217-239.