

SENSOR FUSION FOR ENHANCED OBSTACLE DETECTION IN MARINE VESSELS



By

Muhammad Abdullah Shafiq 01-132222-023

Muhammad Abdul Maalik 01-132222-022

Behroz Khan 01-132222-011

Supervised by

Prof. Dr. Amina Jameel

Department of Computer Engineering

Bahria University Islamabad

2026

SENSOR FUSION FOR ENHANCED OBSTACLE DETECTION IN MARINE VESSELS



By

Muhammad Abdullah Shafiq 01-132222-023

Muhammad Abdul Maalik 01-132222-022

Behroz Khan 01-132222-011

Supervised By

Prof. Dr. Amina Jameel

Submitted to the Department of Computer Engineering in the partial
fulfilment of the requirements for the degree of
Bachelor's in Computer Engineering.

Department of Computer Engineering

Bahria University Islamabad

2026



BAHIRA UNIVERSITY
DEPARTMENT OF COMPUTER ENGINEERING

Dated: 04-06-2026

CERTIFICATE

We accept the work contained in the report titled “**Sensor Fusion for Enhanced Obstacle Detection in Marine Vessels**” as a confirmation to the required standard for the partial fulfillment the degree of Bachelor of Computer Engineering.

Dr. Waleed Manzoor
Project Coordinator
Date: 04-06-2026

Dr. Amina Jameel
Supervisor
Date: 04-06.2026

Dr. Ali Arshad Uppal
External Examiner
Date: 04-06.2026

Engr. Kashif Naseer
Internal Examiner
Date: 04-06.2026

Head of Department (CE)
Date: 04-06-2026

SENSOR FUSION FOR ENHANCED OBSTACLE DETECTION IN MARINE VESSELS

SDG No	Description of SDG	SDG No	Description of SDG
SDG 1	No Poverty	SDG 9	Industry, Innovation, and Infrastructure ✓
SDG 2	Zero Hunger	SDG 10	Reduced Inequalities
SDG 3	Good Health and Well Being ✓	SDG 11	Sustainable Cities and Communities ✓
SDG 4	Quality Education	SDG 12	Responsible Consumption and Production
SDG 5	Gender Equality	SDG 13	Climate Change ✓
SDG 6	Clean Water and Sanitation	SDG 14	Life Below Water ✓
SDG 7	Affordable and Clean Energy	SDG 15	Life on Land
SDG 8	Decent Work and Economic Growth ✓	SDG 16	Peace, Justice and Strong Institutions
		SDG 17	Partnerships for the Goals ✓

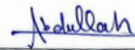


RANGE OF COMPLEX ENGINEERING PROBLEMS / ACTIVITIES

Range of Complex Problem Solving			
	Attribute	Complex Problem	
1	Range of conflicting requirements	Involve wide-ranging or conflicting technical, engineering and other issues.	
2	Depth of analysis required	Have no obvious solution and require abstract thinking, originality in analysis to formulate suitable models.	
3	Depth of knowledge required	Requires research-based knowledge much of which is at, or informed by, the forefront of the professional discipline and which allows a fundamentals-based, first principles analytical approach.	✓
4	Familiarity of issues	Involve infrequently encountered issues.	
5	Extent of applicable codes	Are outside problems encompassed by standards and codes of practice for professional engineering.	
6	Extent of stakeholder involvement and level of conflicting requirements	Involve diverse groups of stakeholders with widely varying needs.	✓
7	Consequences	Have significant consequences in a range of contexts.	
8	Interdependence	Are high level problems including many component parts or sub-problems.	
Range of Complex Problem Activities			
	Attribute	Complex Activities	
1	Range of resources	Involve the use of diverse resources (and for this purpose, resources include people, money, equipment, materials, information and technologies).	✓
2	Level of interaction	Require resolution of significant problems arising from interactions between wide ranging and conflicting technical, engineering or other issues.	✓
3	Innovation	Involve creative use of engineering principles and research-based knowledge in novel ways.	✓
4	Consequences to society and the environment	Have significant consequences in a range of contexts, characterised by difficulty of prediction and mitigation.	
5	Familiarity	Can extend beyond previous experiences by applying principles-based approaches.	

Undertaking

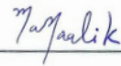
I affirm that the work titled “Sensor Fusion for Enhanced Obstacle Detection in Marine Vessels” is my original work. It has not been submitted for evaluation anywhere else. Any material sourced from other references has been appropriately cited and acknowledged.



Signature of Student

Muhammad Abdullah Shafiq

01-132222-023



Signature of Student

Muhammad Abdul Maalik

01-132222-022



Signature of Student

Behroz Khan

01-132222-011

Dedication

We dedicate this work to our families, whose steadfast support and encouragement have made this journey achievable. We also dedicate it to our supervisor, Prof. Dr. Amina Jameel, for her guidance and mentorship during this project, as well as to everyone who shares the goal of creating safer and smarter maritime navigation systems.

Acknowledgments

Our special thanks go to our supervisor, Prof. Dr. Amina Jameel, for her help, technical knowledge, and support during all the years we spent conducting this Final Year Project. The knowledge she has acquired in the field of artificial intelligence helped us throughout our work. We would like to extend thanks and appreciation to the Department of Computer Engineering, Bahria University, Islamabad, for providing the facilities and the academic environment required to carry out this project. Lastly, we thank our loved ones for their understanding, support, and for being with us when things get tough.

Abstract

Effective safe navigation in demanding maritime environments demands dependable obstacle detection at all ranges and in all weather and lighting conditions. Existing mono-modal systems, whether radar, LiDAR or camera, suffer from significant drawbacks and cannot be relied upon to operate independently. This project presents a multi-modal sensor fusion software system developed to mitigate these drawbacks and merge the information produced by camera, LiDAR, and radar sensors into a single, comprehensive 360-degree situational awareness image. The system was developed using publicly available maritime dataset: the Pohang Canal dataset. Four independent deep learning models were trained: YOLOv8 for stereo camera object detection, achieving a mAP@50 of 0.917; Deeplab V3 for IR camera; a PointNet binary classifier for LiDAR cluster filtering, achieving 100% accuracy; and a PointNet binary classifier for radar cluster filtering, achieving 82% overall accuracy. We used Depth Anything V2 for monocular depth estimation. It converts camera detections into metric range estimates.

The outputs from the model are merged through a five-stage late-fusion process that systematically addresses stereo camera duplicates, cross-verifies camera detections with infrared readings, enhances detections using LiDAR metric depth, incorporates radar validation, and ensures comprehensive lateral coverage. In a sample evaluation frame, the combined pipeline generated 51 confirmed detections ranging from 21.7 to 1493 meters, significantly surpassing the capabilities of any individual sensor. The findings are displayed via an interactive six-view Matplotlib dashboard specifically designed for bridge environments.

Keywords: Sensor Fusion, Obstacle Detection, Maritime Navigation, LiDAR, Radar, YOLOv8, PointNet, Depth Estimation, Deep Learning, Autonomous Vessels

TABLE OF CONTENTS

Sustainable Development Goals	ii
Undertaking.....	iii
Range of Complex Engineering Problems	iii
Dedication	iv
Acknowledgments	v
Abstract	vi
List of Figures	xiii
List of Tables	xiv
1 Introduction	1
1.1 Background.....	1
1.2 Challenges in Maritime Perception.....	2
1.3 How a Ship Moves	4
1.4 Key Forces and Environmental Inputs	5
1.5 Operational Implications.....	6
1.6 Sensor Types Used on Ships	7
1.6.1 Radar	7
1.6.2 Camera	7
1.6.3 LiDAR	8
1.6.4 Automatic Identification System.....	9
1.6.5 Sonar	9
1.7 Research Motivation and Significance	9
1.8 Problem Statement	11
1.9 Objectives	11

1.10	Scope of Work.....	12
1.11	Conclusion.....	13
2	Literature Review	14
2.1	Maritime navigation and obstacle detection	14
2.1.1	Fusion of Mid-Level Feature (Feature-Level Fusion)	16
2.1.2	Late Fusion (Decision-Level Fusion).....	16
2.1.3	Comparison of Fusion Strategies	17
2.2	Classical Computer Vision and Machine-Learning Methods	18
2.2.1	Feature Extraction	18
2.2.2	Subsection Support Vector Machine (SVM)	19
2.2.3	Random Forest.....	19
2.2.4	K-Nearest Neighbours (KNN)	19
2.2.5	Haar Cascade Classifiers	19
2.2.6	Bag of Visual Words (BoVW)	20
2.2.7	Template Matching	20
2.3	Deep Learning Models.....	21
2.3.1	PointNet.....	21
2.3.2	Convolutional neural network (CNN)	21
2.3.3	Faster R-CNN	22
2.3.4	PointNet++.....	22
2.3.5	ResNet (Residual Networks).....	23
2.3.6	DeepLabv3.....	24
2.3.7	Depth Anything V2.....	24
2.3.8	Transformer-Based Models	25
2.4	Maritime Data for Obstacle Detection	25
2.4.1	Pohang Canal Dataset	26
2.4.2	Singapore Maritime Dataset (SMD).....	26

2.4.3	LaRS Dataset	27
2.4.4	KOLOMVERSE.....	27
2.4.5	Comparison of Maritime Datasets	28
2.5	Research Gap Analysis	29
2.6	Conclusion.....	30
3	Design and Methodology	32
3.1	System Architecture Overview	32
3.2	Dataset Selection	33
3.2.1	Pohang Canal Marine navigation Dataset	33
3.3	Data Preprocessing and Calibration	35
3.3.1	Stereo Camera Preprocessing	35
3.3.2	Infrared Camera Preprocessing	36
3.3.3	LiDAR Preprocessing	36
3.3.4	Radar Preprocessing.....	37
3.3.5	Timestamp Synchronisation.....	37
3.4	Independent Object Detection Model Architectures	38
3.4.1	YOLOv8: Stereo Camera Detection	38
3.4.2	DeepLabv3: Infrared Segmentation	39
3.4.3	Depth Anything V2: Monocular Depth Estimation.....	40
3.4.4	PointNet: LiDAR Cluster Classifier	40
3.4.5	PointNet: Radar Cluster Classifier.....	40
3.5	Five-Stage Late Fusion Pipeline.....	40
3.5.1	Stage 1: Stereo Camera Duplicate Resolution	41
3.5.2	Stage 1.5: Infrared Cross-Validation	41
3.5.3	Stage 2: LiDAR Metric Depth Enrichment	41
3.5.4	Stage 3: Radar Confirmation.....	42
3.5.5	Stage 4: Lateral Coverage Completion	43

3.6	System Integration and Execution Flow	43
3.7	Interactive Visualisation Dashboard.....	45
3.8	Conclusion.....	46
4	Implementation, Results and Discussion	48
4.1	Overview of Results.....	49
4.2	Camera Object Detection	49
4.3	Infrared Camera	52
4.4	LiDAR Classifier	53
4.5	Radar Classifier	55
4.6	Integrated Fusion System Results	57
4.7	Dashboard and Operator Usability	60
4.8	Cross-Modal Performance Analysis.....	62
4.8.1	Detection Source Distribution	62
4.8.2	Range Coverage by Modality	63
4.8.3	Sensor Removal Impact Analysis.....	64
4.9	Conclusion.....	65
5	Conclusion	66
5.1	Overview	66
5.2	Summary and Achievements	66
5.3	Achievement of Objectives	68
5.4	Comparisons with Related Work	70
5.5	Limitations	72
6	Future Work	75
6.1	Recommendations for Future Work	75
6.1.1	Expanding the Radar Training Set.....	75
6.1.2	Adaptive Confidence-Weighted Fusion.....	75
6.1.3	Evaluation on the MOANA Dataset	76

6.1.4	MOANA Intrinsic Calibration Recovery	76
6.1.5	Deep Feature-Level Fusion	76
6.1.6	Real-Time Deployment and Hardware Integration	77
6.1.7	Extension to Naval and Patrol Vessel Contexts	77
6.2	Conclusion.....	78
6.3	Closing Remarks	79
	References	88
	List of Abbreviations	89
A	System Configuration and Software Dependencies.....	91
	Annexures.....	91
B	Dataset Directory Structure	93

LIST OF FIGURES

<i>Number</i>	<i>Page</i>
Fig 3.1 : Sensors on vehicle for the Pohang Canal dataset.	35
Fig 3.2 : YOLOv8 camera detection output on a Pohang Canal left-camera frame. ...	38
Fig 3.3 : Additional YOLOv8 detection output illustrating detection in varied lighting conditions.	39
Fig 3.4 : LiDAR and camera fusion result showing matched detections with metric depth.	42
Fig 3.5 : Full multi-sensor fusion output combining camera, LiDAR, and radar detections.	43
Fig 3.6 : LiDAR top-down view with PointNet-filtered clusters highlighted.	44
Fig 3.7 : Radar Cartesian view with DBSCAN clusters and PointNet classification labels.	45
Fig 3.8 : Six-view interactive dashboard layout with fusion polar display and individual sensor views.	46
Fig 3.9 : Fusion view at full zoom showing all sensor coverage arcs and detection markers.	46
Fig 4.1 : Camera (YOLOv8) training performance across 30 epochs.	52
Fig 4.2 : The front LiDAR view for frame 008000 shows seven clusters that remained after applying PointNet filtering, with distances ranging from 21.7 to 110.9 metres.	55

Fig 4.3 : Radar view demonstrating 28 confirmed radar obstacles at ranges from 479 to 1493 metres.....	57
Fig 4.4 : Fusion view at full 1655 m zoom for Frame 008000.....	59

LIST OF TABLES

<i>Number</i>	<i>Page</i>
Table 2.1 : Comparison of Sensor Fusion Strategies	17
Table 2.2 : Comparison of Maritime Datasets for Obstacle Detection	28
Table 2.3 : Comparative Review of Related Works	29
Table 4.1 : Summary of All Model Results	49
Table 4.2 : YOLOv8 Camera Training Metrics at Selected Epochs	50
Table 4.3 : LiDAR PointNet Classification Report	53
Table 4.4 : Radar PointNet Classification Report	55
Table 4.5 : Radar PointNet Confusion Matrix	56
Table 4.6 : Detection Counts for Frame 008000.....	58
Table 4.7 : Detection Source Distribution on Frame 008000	62
Table 4.8 : Detection Count Under Single-Sensor Removal (Ablation Study).....	64
Table 5.1 : Comparison of Key System Metrics Against Related Maritime Fusion Works.....	72

Chapter 1

Introduction

1.1 Background

With more than 80% of the world's trade carried by marine vessels, these are at the centre of international trade. [1]. This is why safe and reliable navigation is very important for naval as well as commercial ships. However, ships are regularly required to work with continually changing and uncertain conditions. In bad weather, (fog, heavy rain, high) navigation can be difficult. even in the night time, gusts, high seas [2, 3]. Ships also need to go through watercraft can dock.narrow canals and locations close to the coast where small watercraft such as buoys, fishing boats, etc. can be moored. craft could spring up and be very dangerous.

Maritime accidents are large enough to highlight the need for better perception systems. The European Maritime Safety Agency (EMSA) has reported more than 2,000 marine casualties and incidents in European waters in 2022 alone, of which collisions and groundings make up the biggest portion [4]. In fact, Allianz Safety and Shipping Review finds that human error is a factor in about 75% of maritime accidents worldwide, of which many are directly caused by situational awareness (SA) and collision avoidance (CA) failures [5]. The economic impact is significant: damage from a collision between large commercial vessels can run well over hundreds of millions of dollars, including the cost of repairing the vessel, lost cargo, environmental damage and port disruption. The 2021 grounding of the Ever Given ship in the Suez Canal (a non-collision event) proved that a single navigational incident can have an impact of disrupting global supply chains for weeks and an estimated loss of trade

of USD 9.6 billion for every day of blockage, as per Allianz 2023.

To facilitate safe navigation, ships are relying more on navigation equipment such as extra Sensors such as cameras and the Automatic Identification System (AIS). Unfortunately, all these There are limitations of devices. Some objects can be detected at several kilometres using, for example, the radar, but small objects or poorly reflective objects, for instance in heavy traffic, can be problematic [2, 3]. AIS, however, can only be used when a ship is transmitting signals [6]. These restrictions are doubtless significant for the most extreme marine where lack of rapid identification and classification of obstacles could prove to be fatal.

The improvement has been made possible with the use of artificial intelligence and machine learning [7]. The combination of data from radar, LiDAR, nfrared and camera makes it possible to improve the obstacle detection in the difficult maritime environments [8, 9]. The use of different sensing methods in tandem (sensor fusion) can assist in enhancing detection, even in lower light conditions, poor weather, and variable water conditions, where a single sensor may be less effective because of particular weaknesses.

1.2 Challenges in Maritime Perception

Unlike similar tasks in other domains like autonomous driving, obstacle detection in the maritime environment is much more challenging due to the unique set of challenges. The sea surface is a dynamic, reflective, and visually homogeneous background that makes the performance of almost every sensor modality worse than on a road environment [9, 10].

One of the most common hazards is fog and reduced visibility. There are some camera-based detection systems which can be used for several kilometres, but dense fog can make it less than 100 metres. Advection fog is the most common type of fog in coastal areas and harbour approaches, where it can develop quickly and unexpectedly, giving ships limited time to respond [3]. Fog has a lesser effect on radar and infrared sensors, but in extreme cases there

is still an attenuation of the signal.

Rain and sea spray result in water droplets on the camera lens, scattering of the laser beam to create false returns in LiDAR data, and increase of radar sea clutter by the radar cross-section of the water surface [11, 12]. Single sensor systems are especially fragile since heavy precipitation may affect several of the sensor modalities at the same time.

Bright reflections of the sun on the water (called sun glare) and specular reflections on the water surface that occur when the sun hits the water at low angles can saturate camera sensors and cause false positive detections. This effect is most pronounced at dawn or dusk, and when boats are running towards the sun, and although not as severely impacted by the sun's glare as optical cameras, infrared cameras can still be overloaded by thermal reflections [10].

The strongest cue source for visible-light cameras is completely removed during night-time operation. Irf cameras also have the ability to recognize vessels' thermal signatures in dark environments, however, the spatial resolution of irf cameras is often not as high as the optical cameras and the thermal contrast of vessels and the water surface is dependent on the surrounding temperature and the engine temperature of the vessels [13].

Wave-induced clutter is an issue peculiar to the maritime domain. Radar returns, LiDAR point clusters, and visual texture can be mistaken for small obstacles like navigation buoys or debris when breaking waves, white-caps, and foam patterns generate radar returns. The strength of the wave clutter is dependent on the sea state and the detection becomes more difficult as the sea state gets worse [3, 14].

The background that objects need to be detected changes continually, as in a **dynamic sea surface**. The water surface is not always level like the road surface, but rises and falls with the waves, causing unpredictable reflecting and refracting of the sensor signal. This renders background subtraction methods, which are well suited for terrestrial surveillance, mostly inapplicable for maritime applications [9].

The motivation for this is multi-modal sensor fusion, which can make use of the fact that the strengths of different types of sensors can compensate for their weaknesses under challenging conditions.

1.3 How a Ship Moves

Before one can appreciate the visual needs on a vessel, one needs to first look at the motion of a vessel. Marine vessel has six degrees of freedom (6DoF):

- **Surge:** Linear, forward or backward movement, propellers and water jet are the main means of propulsion.
- **Sway:** Motion across the water, caused by wind, water flow or by the boat's movement.
- **Heave:** Up and Down motion due to waves and swell.
- **Roll:** Rotation about the longitudinal body axis, caused by wave switching.
- **Pitch:** When waves dip first the bow and then the stern producing a rotation around the athwartship axis.
- **Yaw:** Turn left and right (change heading).

The six degrees of freedom are the complete range of movements of a vessel in the water. These six degrees of freedom are not independent, but are constantly stimulated by outside surge, sway and heave forces (e.g. wave, wind, current) and by the vessel's own roll, pitch and yaw power and control surfaces.

Vessel size has a significant impact on the dynamic response. In general, the bigger a warship, The slower and more powerful is it, the smaller is the boat or Unmanned Surface Vehicle (USV). much faster and far more agile, and thus will be more prone to fall prey to outside disturbances.

These movements intrinsically change the orientation and height of the vessel with respect

to the water surface, thus influencing the functioning of onboard sensors such as cameras, LiDAR, and radar. For example, rolling and pitching can cause excessive viewpoint change in optical Section of LiDAR point cloud is empty or is distorted by camera effect, etc., for a certain period. “bending” the radar beam from the direction it is intended to go [15, 16].

1.4 Key Forces and Environmental Inputs

There are a number of fundamental forces and environmental conditions that influence the trajectory of the vessel. Surge is mainly caused by the thrust of the ship’s own engines and/or propellers. forward or backward. Helm and power are used to steer the boat. Caused by the thrusters which change the angle of the ship’s course. sway through yaw. The outer parts of the ship are subject to horizontal forces due to the wind. Drift and extra yaw moments that make steering hallenging, can also be found in the ocean. Drag: cause to move off course; move off course slowly.

Also, wave reflections can and will cause the vessel to have pitch, heave and roll movements that negatively affect the stability of the vessel and/or add more vibrations or obstructions. When navigating near shore along a coastline or in a constricted canal, hydrodynamic phenomena such as bank forces and depression are also important to consider since they influence a vessel’s response to given instructions [16, 17].

In harbour canals and approach channels, where the water flow is restricted, there are extra hydrodynamic effects due to the interaction with the channel boundaries and the vessel hull. The Bank Suction Effect refers to the tendency for the vessel to be drawn towards the closest wall, and the Squat Effect is the phenomenon of the vessel sinking deeper in the water as she increases speed in shallow water, reducing the under-keel clearance, and possibly changing the waterline seen by cameras and the height reference used by LiDAR. These effects are more severe in shallow waters in high forward velocities for large vessels and can lead to undesired heading changes which violate the assumption of fixed sensor mounts used in the

system [17].

1.5 Operational Implications

A major challenge to vision and navigation systems is the constantly shifting One of the big problems for vision and navigation systems is the frequent movement of an ocean-going vehicle. As the ship moves through its six degrees of freedom, sensors starts to track moving, floating things instead of stationary points. The sensor position and direction can change by a few degrees as a result of slight vibrations in the wave environment, causing inconsistencies in the data and/or changes in the environment as perceived by the sensor [17].

For instance, as a vessel makes its way through a dirty sea, the camera viewfinders can become clouded, altering the water's level over time. This motion variance makes the object-detection algorithm less coherent and is not able to maintain bounding boxes precisely. Similarly, in a similar situation, an anomalous distortion and misregistration would occur if the LiDAR scans were acquired from multiple locations without proper registration (typically requires an inertial measuring unit or IMU) [14].

Radar is not nearly as sensitive to small angle changes as optical devices. Dynamic errors can, however, still damage a radar system. Doppler-induced changes may create false reflections or give the impression that a target is moving due to the antenna's continuous pitching and heaving. With long-range systems, where a slight pitch change throws the radar beam many meters out of the desired range, tracking is impossible. Optical and IR Systems are highly susceptible to the effects of motion blur from ship vibration and high angular speeds and rates; To produce sharp photos, stabilisation is necessary through using gimbals or electronic image stabilisation. [3].

1.6 Sensor Types Used on Ships

In this section, the most popular sensors used for maritime perception and navigation, along with their main functions.

1.6.1 Radar

Radar is a sensor which can detect objects such as ships and coastlines from a long distance, even in the dark or in extreme weather conditions such as fog and rain. It has a radio wave detector and then listens for the returned echoes. The greatest advantage of this type of sensor is its extremely wide range and ability to operate in almost all environments and detect any solid object. Nevertheless, it produces an image with less detail than cameras or LiDAR, meaning that it is often uncertain as to exactly what an object is.

There are two frequency bands used in maritime radar. The X-band radar (9.3–9.5 GHz) has a higher spatial resolution and is more sensitive to small targets like the navigation buoys although it is also more influenced by rain clutter. S-band radar (2.9–3.1 GHz) has better performance in precipitation, but, in return, it has lower angular resolution [3]. The complementary strengths of the X-band and S-band radars have led to their use in modern ships. The data set for Pohang Canal used in this project is an X-band FMCW radar that can detect up to 50 to 1654.8 metres [18].

1.6.2 Camera

Cameras provide a clear display of what exactly is occurring. They allow the machine and the operator identify what an object is, for example, a ship versus a buoy, as well as perceiving volume and depth and other visual elements. Thermal imaging cameras can function in complete darkness. The disadvantage is In poor weather (e.g. thick fog) the vision may be blurred, and the camera may be obscured by the weather. There are various types of cameras used in maritime applications. The Visible Light RGB camera produces

colour images of objects, which can be classified, but can only be used in daylight and clear weather. Near-infrared (NIR) cameras are able to detect energy beyond the visible spectrum and are capable of penetrating low levels of haze to a certain extent. Long-wave infrared band (8–14 μm) thermal infrared cameras can capture the heat signature of vehicles, which is crucial for low-visibility and night time operations [10, 13]. Triangulation of the corresponding features in the two images in a pair of synchronised stereo cameras with a known baseline separation can be used to estimate depth.

1.6.3 LiDAR

LiDAR is a sensor which is well suited to create a detailed 3D image of the local surroundings. for medium range obstacle detection and identification. It works by firing laser rays and The time taken for the light to return is recorded. The greatest benefit of LiDAR is that it can create an image that is very detailed and accurate, able to identify exactly what shape objects are in low light conditions. It slows down in the event of bad weather like fog and rain. It is also costly and aintenance intensive and has a smaller detection range than a standard marine radar.

The wavelengths for maritime LiDAR are either 905 nm or 1550 nm. The 905 nm wavelength is more frequently used and cost effective, but has eye safety regulations that cap output power. At higher power levels 1550 nm is eye-safe and gives a longer range of detection, but the sensors are more expensive. In heavy rain and fog, both wavelengths are affected by scattering and absorption of the laser beam by water particles, causing the performance to be reduced [11]. The density of point cloud decreases with distance from sensor to object, causing objects at the far end of the sensor's range to be represented by very few points and thus making classification difficult.

1.6.4 Automatic Identification System

AIS is a system through which the data, position and course of ships are communicated. The major advantage of AIS is that it makes clear identification and destination information of a provides traffic awareness, collision avoidance and is available to other ships. However, it can't provide the position of other ships unless data is transmitted from them and it's not available to receive data. The receiving data might be inaccurate or garbled, and the sending powerless ships, unregistered boats, or small craft could be delayed.

AIS broadcasts have several types of messages. Class A transponders, which are required on ships that are 300 gross tonnage or larger, send out dynamic information (position, course, speed, etc.) every 2 to 10 seconds, depending on the speed of the vessel, and static information (vessel name, type, dimensions, etc.) every 6 minutes. Class B transponders are aboard smaller boats and operate at lower frequencies and power. The basic problem with AIS for obstacle knowledge is that an object that does not have an AIS transponder, such as an unmarked navigational obstacle, a small fishing vessel, or a kayak, is undetectable by the system [6, 19].

1.6.5 Sonar

Underwater imaging of the bottoms of waterbodies, and underwater obstacle detection by sonar, This makes it extremely handicap for a submarine or other underwater vessel. The major benefit is that it is effective underwater. Does not work well in identifying objects on water surface, Certain techniques are needed to interpret the data that has been obtained.

1.7 Research Motivation and Significance

This motivation for this research comes from the fact that a single sensor will not be reliable in detecting obstacles under all the operating conditions for a maritime system. Radar is good in severe weather and at distance for detection, but has limited ability for precision

classification. There is much semantic information available in cameras, but it has limitations during dark and poor visibility. LiDAR is highly accurate in 3D measurements but has a limited range and is sensitive to weather. The identity information is only available for cooperative targets in case of the AIS. These conditions for the effectiveness of each sensor have corresponding conditions where they are ineffective, and importantly, sensor failure conditions are far from mutually exclusive [20, 21].

The complementarity means there is a clear opportunity: if the outputs from several types of sensors can be combined, the system can detect with a reliability higher than any of the individual sensors. This has been well illustrated in the autonomous driving industry, where today's autonomous driving systems regularly combine camera, LiDAR and radar data using advanced fusion architectures [21, 22]. But it is not simple to directly translate such automotive technologies to the marine field. Maritime environments have specific difficulties such as a much longer range of detection (hundreds of metres to kilometres) vs. road environments (tens of metres), a background that is constantly changing and highly reflective (water surface) vs. road environments (road surface), no lane marking or road infrastructure for spatial reference, and much fewer publicly available multimodal data sets for training and evaluation [23, 24].

The International Maritime Organization (IMO) has noted how autonomous technologies are playing a greater part in maritime activities in their Maritime Autonomous Surface Ships (MASS) regulatory scoping exercise, highlighting the need for reliable perception systems for any degree of maritime autonomy [25]. This work can directly assist in the development of such perception systems by providing a practical solution to multimodal sensor fusion with real-world maritime data.

1.8 Problem Statement

Despite the progress of all the recent advances in the field, a single-sensor system may not be able to accurately provide In a naval environment, obstacle detection is a problem, because of the lack of any sensor redundancy on the single-sensor vessels. The hallucinations, blind spots and missed detections. For example: Small wooden boats may be undetectable by radar [14]. Note that cameras are very sensitive in terms of visibility and lighting [12]. With high fog, rain or sea spray, the performance of LiDAR may diminish [11]. These shortcomings present safety problems, especially in naval operations, when ships need to operate in confined waters when fighting and rescuing vessels or when in other challenging situations at sea conditions with a decreased operator awareness. Hence, multimodal sensor fusion framework for effective obstacle identification and classification should be developed by integrating all available data sources and by using machine learning techniques.

1.9 Objectives

Objectives of this project are:

1. **Develop a sensor fusion based obstacle detection system** for marine boats which are able to cope with complex naval conditions. This is the main goal because one-sensor solutions have proven to be inadequate on maritime perception under different environment conditions [20].
2. **Multimodal datasets with three or more sensors should be preprocessed and integrated.** Real-world maritime sensor data arrives in different formats, coordinate systems, and temporal rates. Proper preprocessing and calibration is essential before any meaningful fusion can be performed [18].
3. **Develop machine learning models for obstacle identification and classification.** Each sensor modality requires a dedicated model architecture that is suited to its

specific data characteristics, whether that is 2D imagery, 3D point clouds, or radar sweep images [26].

4. **Develop a graphical user interface (GUI) based dashboard** to visualize fused sensor data in real time. For effective human-machine interaction, the outputs of the system should be displayed in a clear and interpretable manner, enabling the vessel operator to make quick decisions [27].

1.10 Scope of Work

This is only simulated by software. A graphical user interface based platform, machine learning-based detection and classification, and preprocessing and fusion approaches are all part of the project. The main data source is the Pohang Canal dataset [18] which offers multi-modal sensor inputs of radar, LiDAR, stereo, and IR camera, and thus enables evaluation of sensor fusion algorithms under realistic maritime scenarios. To summarise, this chapter presented the reasons why the safe navigation is so important and necessary. challenging operation for ships. To motivate the research, Maritime accident statistics and the economic impact of collisions were presented. Specific challenges of maritime perception, namely fog, rain, glare, night-time maritime perception and clutter due to the ocean waves were detailed. Six motions of ships were introduced and how these are sensed in living organisms was discussed. The challenges with waves, wind and currents in relation to obtaining accurate sensor data were discussed.

The project will tackle these challenges by developing a system for sensor fusion using sensor fusion and machine learning. This will enable the system to produce more high quality and Blend information from a variety of sensors to make accurate decisions. The project is completely is a software-only simulation and has been trained and tested with real-world maritime data from the proposed sensor. fusion model. The research, design, implementation and results of experiments are presented in the following chapters.

1.11 Conclusion

Ships have been colliding with things for as long as ships have existed. What changes is how much we can reasonably do about it. Radar, cameras, and LiDAR each solve part of the problem but none of them solves it alone — fog beats cameras, clutter confuses radar, and LiDAR range runs out well before a fast vessel does. That gap is what motivated this project.

The core argument is simple: if you combine the outputs of multiple sensors instead of depending on one, you get a picture of the environment that is harder to fool. This is not a new idea. Autonomous vehicles have been doing it for years. Maritime research has lagged behind, partly because of the difficulty of collecting calibrated multi-modal data at sea. The Pohang Canal dataset made this work possible by providing synchronised camera, infrared, LiDAR, and radar recordings with calibration files — exactly the kind of ground truth needed to build and test a fusion pipeline.

The system described in this thesis does not attempt to solve every maritime perception problem. It is a prototype, built on one dataset, evaluated on one scene. What it does show is that combining four sensor modalities produces detections — 51 of them, ranging from 22 to 1493 metres — that no single sensor could have produced on its own. Whether that generalises beyond a harbour in South Korea is one of the open questions Chapter 6 addresses.

Chapter 2

Literature Review

This chapter reviews the existing literature pertaining to marine navigation, naval obstacle detection, sensor fusion and machine learning in autonomous systems. The ultimate goal is to understand the drawbacks of existing methods and highlight the possible benefits of using multimodal sensor fusion for improving obstacle detection in marine vehicles.

2.1 Maritime navigation and obstacle detection

Safely sailing through the ocean is a major difficulty for both commercial and naval ships. Conventionally, target detection and tracking are primarily based on radar or Automatic Identification System (AIS). Radar can be used for long-range detection but it is limited in detecting small or low-reflective objects [28]. The drawback of AIS is that it cannot track any hostile or unregistered object as it would need to be actively broadcasting signals [29].

The international guidelines for collision avoidance at sea are set out in the Convention on the International Regulations for Preventing Collisions at Sea (COLREGs) from the International Maritime Organization (IMO) [30]. COLREGs set rules of the road for ships and outlay what they are to do depending on the geometry of the encounter (head-on, crossing, overtaking) and upon their visibility. Compliance with COLREGs requires the vessel operator to maintain an appropriate lookout by all means, sight, hearing and radar. With increasing autonomy in maritime systems, efforts have been made to look into the coding of COLREGs rules in automated decision making systems [31].

Since the 1980s, the Automatic Radar Plotting Aid (ARPA), a traditional electronic naviga-

tion aid, has been a standard tool on the bridge. ARPA systems automatically acquire and track radar targets and calculate their course, speed, closest point of approach (CPA) and time to CPA. Despite this, ARPA was a major leap forward in situational awareness over the use of raw radar data, but it was constrained by the basic limitations of radar as a single sensor, such as the inability to discern between small targets and sea clutter [27]. Although modern integrated bridge systems incorporate ARPA functions with electronic chart displays and AIS overlays, even with these integrated systems, radar remains as the primary detection sensor, and is limited in its ability to detect small non-metallic objects [32].

As revealed in research, relying solely on one sensor for navigation requires an increased risk of accidents in fog, storms, and darkness [10, 13, 28]. To enhance attentiveness in naval applications, the multimodal sensing has been explored.

A sensor fusion in autonomous systems is presented in this section.

The use of multiple sensors, to provide a more robust and accurate representation of the world, is called sensor fusion. It is historically known to be used with robotics and self-driving cars, which depends on camera, LiDAR, radar, and infrared sensing data [20, 33, 34]. Kalman filtering, Bayesian fusion and deep learning-based fusion models have greatly enhanced the detection quality [35, 36].

Due of its ability to overcome the limitations of individual sensors, sensor fusion is also gaining popularity in the marine environment. While radar offers excellent long-range detection capabilities with limited classification, cameras offer superior classification with limitations in adverse weather and darkness [10, 37]. Fusion of the two modalities, which can complementarily detect these two characteristics, has shown excellent performance [34] with the fusion of the sensors.

Based on how data from various sensors are fused, sensor fusion architectures can be classified into three types [21] as follows:

Early fusion (data-level fusion)

For early fusion, the raw data from various sensors are fused together, and no processing or feature extraction is performed on them. For instance, the camera image and LiDAR point clouds can be mapped to the same BEV representation (e.g., a bird’s-eye-view (BEV) grid) and fed as input channels to a single neural network [38]. The benefit of this is that the model can have access to all the information from the beginning and will be able to learn the cross-modal correlation at the lowest level. Early fusion, however, needs careful alignment of sensors in terms of space and time, is affected by calibration errors and typically requires much greater computational resources because the model needs to process high-dimensional combined inputs [21].

2.1.1 Fusion of Mid-Level Feature (Feature-Level Fusion)

Mid-level fusion involves the fusion of intermediate feature representations which are separately obtained from each sensor modality. There is one feature extraction backbone for each sensor stream, and the extracted feature maps are combined via concatenation, element-wise addition or attention mechanism and then fed into the downstream classification or detection heads [39]. For feature-level fusion, transformer-based architectures have demonstrated strong performance, with the self-attention mechanism effectively learning the importance of each modality to the current detection task [22]. Feature-level fusion is a compromise between modality-specific information and cross modality interaction, but a careful architecture design is needed to cope with different modality feature dimensionality.

2.1.2 Late Fusion (Decision-Level Fusion)

Late fusion is the merging of the individual results (decision, detection, or classification) of independent per-sensor models. Data from each sensor branch is processed separately using the sensor’s detection pipeline to generate a series of detections along with a corresponding set of confidence scores and spatial coordinates. These outputs are then fused together via score averaging, voting, non-maximum suppression (NMS), Bayesian fusion or Dempster-

Shafer evidence theory [40]. The major benefit of late fusion is that every sensor branch can be designed, trained, tested, and updated separately and the system can still function normally if any of the sensors are unavailable or fail [20]. The drawback is the complementary information that may have been communicated between modalities during processing is lost.

2.1.3 Comparison of Fusion Strategies

Table 2.1 Comparison of Sensor Fusion Strategies

Strategy	Fusion Point	Advantages	Disadvantages
Early Fusion	Raw sensor data	Access to all information; learns low-level correlations	Requires precise calibration; high compute cost
Mid-Level Fusion	Intermediate features	Balances information sharing with modularity; attention-based weighting	Complex architecture; careful design needed
Late Fusion	Independent outputs	Modular; graceful degradation; easy to debug	Cannot share cross-modal cues during processing

Table 2.1 summarises the key characteristics of each fusion approach.

Being modular, easy to debug and able to run in degraded mode (when a sensor is not available) makes this project a late fusion system, which is significant for a research prototype run with recorded maritime data [21].

In perception systems, machine learning and deep learning algorithms have transformed perception in many fields. Convolutional Neural Networks (CNNs) [41] have proven to

be successful for image-based detection and classification, respectively, and PointNet [42] and These PointPillar [43] models are often used for LiDAR based object detection. In multimodal situations, fusion models like YOLO [44], Faster Radar versions of R-CNN [45] and ResNet [46] have been adapted to analyse radar, Both LiDAR and camera data at the same time.

images and forecast their arrival time at ports. Machine learning has also been applied in the maritime domain, to detect vessels from satellite images and predict their time of arrival at ports. classify ships from imagery data [47], radar signals [48] and recognise obstacles using images. But there are little studies which have applied multi modal maritime data to Build up efficient fusion systems [49, 50].

2.2 Classical Computer Vision and Machine-Learning Methods

Before the modern resurgence of deep learning, object detection and recognition techniques have been developed using various approaches. Most commonly, used manually designed image features and traditional ML classifiers. Such classical approaches can be used as stepping stones to future approaches and are effective in scenarios where there is limited available computation or data [51].

2.2.1 Feature Extraction

In the first phase interesting patterns were identified in the image. For instance, SIFT and SURF extracted salient points which are insensitive to scale, view point and illumination, and the Histogram of Oriented Gradients (HOG) method [52] detected edges in images. These features were relatively simple but proved to be very effective for vehicle, ship or person recognition in images taken with a camera [53].

2.2.2 Subsection Support Vector Machine (SVM)

SVM generates the best classification of the two classes of data. The more different it is, the more different a line it makes between the classes, which helps the program to classify objects with more confidence. Due to this property, an SVM is particularly suitable for simple problems such as object differentiation based on the very clear HOG feature [54].

2.2.3 Random Forest

A machine learning method called Random Forest combines the predictions of many decision trees. Each tree is trained and produces its own output on its own. Then the vote of each tree will be used for the final decision. This collaborative decision-making enhances the accuracy of the models and diminishes errors. Random Forest is a suitable method as it can accommodate different data sources and does not overfit [55].

2.2.4 K-Nearest Neighbours (KNN)

KNN is used to predict the classification of an unknown sample by comparing it with those it has already learned. For example, if a radar signal was similar to radar signals of small boats previously, the radar signal will be classified as a small boat. Although KNN is simple to create and comprehend, it becomes inefficient on big datasets and is sensitive to noise. [56].

2.2.5 Haar Cascade Classifiers

The Haar cascade classifiers, proposed by Viola and Jones, are a sequence of progressively complicated stages of simple rectangular feature detectors that enable fast object detection [57]. In each stage Haar-like features are used to rapidly discard regions that do not contain an object, which are calculated by measuring the difference between the intensities of rectangular regions. Only patches of images which pass all the stages of the cascade

are considered detections. However, the advantage of the method is its fast computation allowing real time detection even with restricted hardware. However, Haar cascades can only detect objects with relatively stable shapes and orientations, which is not very convenient for maritime environment because the vessels can be seen in a lot of different scales, aspect ratios and viewpoints.

2.2.6 Bag of Visual Words (BoVW)

The Bag of Visual Words approach is an adaptation of the bag-of-words concept used in text retrieval [58]. Using k-means, a visual vocabulary of local features (e.g., SIFT descriptors) is identified from training images. Then each image is represented as a visual word histogram, which is then given to a classifier like SVM. Prior to deep learning, BoVW was extensively applied to image classification and retrieval and demonstrated well performance on the benchmarks of scene and object recognition. It cannot be used to do things that require exact spatial localisation of objects, since it eliminates spatial relationships between features.

2.2.7 Template Matching

One of the simplest methods of object detection is template matching, where a template image of the object of interest is shifted over the image of interest and a similarity measure (e.g. normalised cross-correlation) is calculated at each shift [59]. The location that gives the highest similarity score shows the most likely location of the target. Though simple in principle, template matching is very sensitive to scale, rotation, illumination, and partial occlusion, all of which can occur in maritime imagery. Therefore, it is not commonly employed as a standalone detection method, but can be included in more complex pipelines.

The old techniques were very problematic and were reasonably effective. They required extensive "human tuning" and often locked out in bad conditions. They were not comfortable with unfamiliar scenarios or large scenes with less visual details. All this challenges has prompted the use of deep learning, which, automatically at large scale, learns robust feature

from raw sensor data.

2.3 Deep Learning Models

2.3.1 PointNet

PointNet is a neural network that can take a 3D point cloud and classify what the points represents or what components it is made of. PointNet considers global feature representations by Max-pooling is applied on individual spatial coordinates of the point cloud after it is processed individually [42].

First, PointNet looks at each point by itself and uses a neural network to determine each point's coordinates and learn more details about its position. Once all the points have been checked PointNet, on the other hand, aggregates that information to form a global understanding by taking the max-pooling operation: For each feature the maximum is taken at each point. By finding the largest The whole point cloud is summarized by a summary vector of each feature for all points. generated. PointNet then derives an object's overall fingerprint that is passed to a final classification layer which obtains a prediction: "This global fingerprint belongs to a ship" or "This is a buoy."

2.3.2 Convolutional neural network (CNN)

CNN are a type of AI models that are quite powerful at recognizing photographs and videos, as they are designed to take in pixels as input [41].

CNNs are fundamentally based on the mathematical operation of convolution that uses tiny matrices known as filters or kernels. A filter will be like a small stencil which will be applied over an image to find a particular pattern such as vertical lines or horizontal lines and produces a feature map which indicates the strength of the pattern for each location. CNNs employ a variety of different filters at the first layers, where each filter searches for one of the basic features. Deeper layers use these simple patterns to create more complex

patterns. At the end of the classification, feature maps of CNN are flattened into 1-D array and passed to fully connected classifier head which extracts all the important features from the picture.

2.3.3 Faster R-CNN

FasterR-CNN [45] is one of the most effective and popular deep learning techniques when it comes to object detection. It was published in 2015 and was regarded as a significant step towards earlier versions like Fast R-CNN. The primary modification increase the speed and precision of the system to recognize objects in images is addition of Region Proposal Network (RPN). The two stages are: The RPN is used to propose regions in the image that could contain an object, and the FastR-CNNHead (Detection Network) is used to closely analyze the proposed regions to decide what is that object and what are its boundaries. These two components, working together, form a joint conv backbone, which extracts useful features from the image.

For every candidate anchor box, the RPN will predict an objectness score (is the box containing an object or not). Bounding box offsets (how to move the box around to fit the object better) and object or background. The size and shape of each region selected by the RPN can vary. They are then processed uniformly, In faster R-CNN, an additional step called RoI pooling is added, which is used to convert all regions to fixed-size feature maps. After Each region is fed into a fully connected network which classifies the object, called RoI pooling. Trefines the bounding box.

2.3.4 PointNet++

PointNet++ [43] is a deep learning model that helps computers to better understand the 3D point cloud data than the original PointNet. The original PointNet could perceive the general outline of objects, but lacked the ability to grasp details, such as the wheels or mirrors on a car. PointNet++ overcomes this by learning large shapes and fine details at once.

PointNet++ works in an organised manner: it begins by picking a few key locations within the whole 3D geometry (sampling), and then considers all the points around each of them (grouping), and uses a small version of PointNet to explore these local regions, and then does so again at larger scales. This method involves collecting small parts of an object at a time, and then combining them to create a complete picture of the object, allowing the network to recognize several parts of the object at once.

2.3.4.1 YOLO (You Only Look Once)

YOLO (YouOnlyLookOnce) architecture [44] was introduced to process all the photos in one shot and facilitate real-time object recognition by implementing one single stage detection paradigm. Prior to Yolo, many other systems had to repeatedly scan various areas of the image in search of objects, which wastes time. YOLO changed that by processing the entire image in one go and providing the results promptly and efficiently. In 2016, the innovative approach was proposed by Joseph Redmon. Numerous updates and versions, such as YOLOv2, v3, v4, v5, v7, v8, and v11, were inspired by YOLO. With each iteration, the YOLO family becomes more precise, faster and has a more sophisticated understanding of the elements of a visual scene. This project adopts Yolo v8 for object detection in stereocamera.

2.3.5 ResNet (Residual Networks)

ResNet [46] proposed residual learning by using skip connections allowing to skip one or more layers during the training. Prior to ResNet, very deep neural networks were plagued by the degradation problem: training accuracy did not continue to rise with depth and eventually dropped after reaching a certain point. Instead, residual connections learn a residual mapping relative to the input to each layer, which solve the above. This allows for the training of networks which have hundreds or even thousands of layers. For detection and segmentation systems, ResNet architectures (ResNet-50, ResNet-101) are often used as

backbone feature extractors, such as Faster R-CNN and DeepLabv3 [26].

2.3.6 DeepLabv3

The DeepLabv3 [60] is a semantic segmentation network that aims to generate dense per-pixel classification maps. Its main contribution is the Atrous Spatial Pyramid Pooling (ASPP) module, which is based on dilated (atrous) convolutions at different rates to extract multi-scale contextual information without downsampling the feature maps. Atrous convolutions introduce holes in the kernel (receptive field), which allows the receptive field to expand without additional parameters, unlike standard convolutions involving neighboring pixels. The ASPP module is usually combined with a ResNet backbone to extract the features, and then a classification layer. The application of DeepLabv3 to infrared imagery for pixel-level vessel and buoy segmentation is in this project.

2.3.7 Depth Anything V2

Depth Anything V2 [61] is a monocular depth estimation (MDE) foundation model that predicts dense depth maps from just a single RGB image. With self-supervised and semi-supervised pretraining with large-scale unlabelled data, the model can determine the depth from visual cues like perspective, texture gradient, occlusion, and relative object size [62]. Its architecture consists of a Vision Transformer (ViT) encoder that takes the input image and outputs the features, and a decoder that generates the depth prediction for each respective pixel. The monocular depth estimation problem is intrinsically restricted by the lack of absolute scale in a single image and this is an issue of particular concern in maritime applications due to the lack of reference features on the featureless water surface. However, the estimates are adequate within the short/medium range (up to 150 metres) to allow for useful range information to be provided in the absence of LiDAR data.

2.3.8 Transformer-Based Models

Transformer, originally designed for NLP, has been adapted for CV with great success. The Vision Transformer (ViT) [63] simply splits an image into a fixed number of patches, each of which is considered a token, and the series of tokens is fed into the standard transformer encoder layers that include the self-attention mechanism. This method is able to model long-range dependency throughout the whole picture, whereas CNNs model dependencies only within a local receptive field that builds up slowly within the layers of the network.

DETR (Detection Transformer) [64] is an extension of the transformer to object detection problem by focusing on directly predicting a set of objects. By replacing hand-designed components like anchor boxes and non-maximum suppression with a transformer decoder with learned object queries that attend to image features, DETR dispenses with the need for such components. DETR reduced the detection pipeline complexity, but the convergence of the training is slower than the CNN-based detector.

For sensor fusion, transformer-based architectures have demonstrated special potential for cross-modal attention, which allows the features from one sensor (e.g., camera) to attend to features from another sensor (e.g., LiDAR), such that the model can learn where from which sensor the information is most relevant for each spatial location [22, 38].

2.4 Maritime Data for Obstacle Detection

One of the biggest obstacles to research on maritime obstacle recognition is the absence of multimodal datasets. Compared to the autonomous driving domain, which is blessed with popular datasets such as KITTI, Cityscapes, and nuScenes, the marine domain has a much smaller number of public datasets. Land-based data sets are not suitable for naval scenarios as they lack a representation of the marine environment, ocean surface reflections and other sensor problems associated with marine environments. In recent years, new datasets with a maritime theme have been developed to fill this gap, and researchers can create and test AI

models in real-world naval environments. [18, 65–68].

2.4.1 Pohang Canal Dataset

The Pohang Canal dataset [18] was collected under the controlled but realistic maritime environment of South Korea. It has a number of sensors, including LiDAR, radar, stereo camera, infrared camera and omnidirectional camera. Synchronisation of each modality with intrinsic and extrinsic calibration data is important to enable accurate sensor fusion work. Due to the variety of sensors it is an ideal platform for multimodal systems designed for precise recognition of obstacles.

This data set is of great importance for naval perception research. This is how it has become possible to train deep learning models on maritime data, test algorithms and fairly compare different sensor fusion methods. The Pohang Canal dataset has greatly helped in creating fusion-based systems that are more dependable for marine operations by combining several data sources in real-world maritime scenarios.

2.4.2 Singapore Maritime Dataset (SMD)

Singapore Maritime Dataset (SMD) [69] is one of the first and most referenced maritime visual datasets. SMD offers a set of visible-light and near-infrared video sequences acquired over different weather conditions such as daytime, night-time and haze conditions, in the congested Singapore Strait. The data set includes per-frame annotations of the locations of the vessels in the harbour and open-water scenes using bounding boxes. SMD has been widely employed for the single-modality camera-based vessel detection and tracking algorithms [70]. However, the data set only consists of visual-spectrum sensors, and lacks LiDAR, radar, or calibration data, limiting its use in the study of multimodal fusion. Also, because of its few annotated frames and lack of various weather conditions other than haze, it cannot be used to train robust deep learning models.

2.4.3 LaRS Dataset

LaRS (Large-scale Realistic Segmentation) is a large-scale panoptic maritime obstacle detection benchmark created specifically for Unmanned Surface Vehicles (USVs) [65]. LaRS offers semantic segmentation labels on a per-pixel basis for three classes: water, sky, and obstacles. The dataset includes more than 4000 frames with annotation from several sources and environments, such as harbours, rivers and open coastal waters. One of the strengths of LaRS is the diversity, including scenes with different lighting conditions, water textures, weather situations, and obstacles, from big ships to little floating debris. LaRS has emerged as a key testbed to assess maritime segmentation algorithms, especially those for autonomous surface vessel navigation. Similar to SMD, LaRS is a single-modality (camera-only) data set which does not include LiDAR or radar data, and therefore cannot be used for sensor fusion research.

2.4.4 KOLOMVERSE

This is the KOLOMVERSE (Korea Open Large-Scale Image Dataset for Object Detection in the Maritime Universe) [67] dataset of maritime object detection images collected in Korea. The data set delivers more than 30,000 images of high quality with box annotations for various types of vessels such as cargo ships, passenger ships, fishing vessels and military ships. What sets KOLOMVERSE apart is the size and variety of vessels it offers, ensuring it is ideal for developing strong classification models. Images were taken at multiple days and times and under various weather conditions, offering some variation in the environmental conditions. Similar to SMD and LaRS, KOLOMVERSE is a camera-only dataset with no additional sensor modalities like LiDAR or radar.

From the comparison, it is observed that, although there are several data sets available from maritime applications for camera-based detection and classification, the Pohang Canal dataset is one of the few data sets which can be provided with synchronized multi-modal

sensor data with full calibration information. This is an ideal platform for applications of sensor fusion that will be investigated in this project.

2.4.5 Comparison of Maritime Datasets

Table 2.2 Comparison of Maritime Datasets for Obstacle Detection

Dataset	Sensors	Annotations	Multi-modal	Key Limitation
Pohang Canal [18]	LiDAR, radar, stereo camera, IR camera, omni camera	Bounding boxes, calibration data	Yes	Restricted to harbour environment
SMD [69]	Visible and NIR cameras	Bounding boxes	No	Camera-only; limited weather diversity
LaRS [65]	RGB camera	Panoptic segmentation masks	No	Camera-only; no depth information
KOLOMVERSE [67]	RGB camera	Bounding boxes (multiple vessel classes)	No	Camera-only; no sensor fusion support
MVDD13 [66]	RGB camera	Bounding boxes for USV navigation	No	Camera-only; limited to USV perspective

Table 2.2 provides a comparative overview of the major maritime datasets discussed in this section.

The comparison reveals that while several maritime datasets exist for camera-based detection and classification, the Pohang Canal dataset remains one of the very few to provide synchronized multi-modal sensor data with full calibration information. This makes it uniquely

suited for sensor fusion research, which is the focus of this project.

Table 2.3 Comparative Review of Related Works

#	Paper (Year)	Method	Sensors	Advantages	Disadvantages
1	Pohang Canal (2023) [18]	Multimodal dataset	LiDAR, radar, stereo, IR, omni cameras	Multiple sensors with calibration; realistic restricted waters	Collected in restricted environments
2	Radar and camera fusion (2024) [71]	Survey on radar-camera fusion	Radar, Camera	Highlights how radar improves long-range detection	Asynchronous data and clutter issues
3	Deep camera-radar fusion (2023) [12]	Attention-based deep fusion	Radar, Camera	More accurate in adverse weather	Requires precise calibration and matched data
4	SAMFusion (2023) [34]	Adaptive multimodal fusion	LiDAR, radar, camera	Adaptive fusion weights handle bad weather	Complex training, compute heavy
5	Ship detection with deep learning (2023) [72]	Survey on ship detection	Satellite imagery, radar	Good multi-domain survey	Lesser focus on sensor fusion
6	Fusion detection on UAV (2022) [40]	Dempster-Shafer fusion	Radar, LiDAR, Camera	Probabilistic fusion resolves conflicting data	Sensitive to sensor model calibration
7	Multi-sensor fusion for UAV (2025) [73]	Multi-sensor obstacle avoidance	Marine radar, stereo cameras, LiDAR, IMU / GNSS	Real-time avoidance for USV navigation	Generalisation limited by training scenes
8	Fusion of camera and AIS (2021) [19]	Camera + AIS fusion	Camera, AIS	Enhances identification with AIS identity	Limited to vessels that broadcast AIS
9	Fusion of LiDAR and radar (2021) [74]	LiDAR and radar depth fusion	LiDAR, Radar	Increases detection in difficult conditions	Requires precise spatial calibration
10	Multi-modal fusion for maritime surveillance (2024) [75]	Deep learning multi-modal fusion	Camera, radar, AIS	Comprehensive fusion architecture	Requires large annotated datasets
11	Object detection survey (2023) [26]	Comprehensive 20-year survey	Various	Covers evolution from classical to deep learning	Survey only; no novel method proposed
12	TransFusion (2022) [22]	Transformer-based LiDAR-camera fusion	LiDAR, Camera	Cross-modal attention improves accuracy	High computational cost; automotive focus

2.5 Research Gap Analysis

The literature reviewed in this chapter reveals several important gaps that this project aims to address.

Limited multimodal maritime datasets. As shown in Table 2.2, the vast majority of maritime datasets provide only camera imagery. While datasets like SMD, LaRS, and KOLOMVERSE have advanced camera-based detection research, they cannot support the

development and evaluation of sensor fusion systems. The Pohang Canal dataset addresses this gap by providing synchronised data from multiple sensor modalities with calibration information, and this project leverages that capability.

Few practical maritime fusion systems. Most sensor fusion research has been conducted in the autonomous driving domain [21], where datasets like KITTI and nuScenes provide standardised benchmarks. The relatively few maritime fusion studies identified in the literature (Table 2.3) often focus on only two modalities (typically radar and camera) and do not integrate the full range of sensors available on modern vessels. This project addresses this gap by fusing five sensor types: stereo cameras, infrared camera, three LiDAR units, and radar.

Lack of end-to-end maritime perception pipelines. While individual components such as ship detection, depth estimation, and radar processing have been studied independently, few works present a complete end-to-end pipeline that processes raw multi-modal sensor data through to a fused, operator-readable situational display. This project contributes a complete pipeline spanning preprocessing, per-modality detection, five-stage fusion, and interactive visualisation.

Insufficient coverage of the full detection range. Many existing maritime detection systems operate within the range of a single sensor modality. The integration of radar in this project extends the detection range to nearly 1500 metres, far beyond what camera or LiDAR systems can achieve independently, addressing a critical operational need for early collision warning [37].

2.6 Conclusion

The pattern that comes out of this review is consistent. Single-sensor systems have known failure modes, multi-sensor systems have known integration costs, and most existing maritime work leans heavily on camera data because camera datasets are abundant. The problem

is that cameras fail in exactly the conditions where maritime perception matters most: night, fog, rain, and long range.

Radar covers the long-range gap. LiDAR provides precise depth in the close-to-medium range. Infrared adds a thermal channel that is largely independent of visible light. The case for combining them is not theoretical — it follows directly from the limitations documented in every paper reviewed in this chapter. What the literature does not supply is a complete, tested pipeline that puts all four modalities together with a shared calibration scheme and a unified output format. That is the gap this project fills, using the Pohang Canal dataset as the common evaluation ground.

The fusion strategy chosen here — late fusion — has the specific advantage that each sensor branch can be evaluated, debugged, and replaced independently. Early and mid-level fusion can squeeze more accuracy out of training data but make the system much harder to diagnose when something goes wrong. For a maritime safety application running in a harbour with real targets, that diagnostic transparency matters. The literature supports this trade-off even if it does not always phrase it that way.

Chapter 3

Design and Methodology

The goal of the project is to create a model of sensor fusion to detect and classify an obstacle in the marine environment. This will also help to overcome the disadvantages of single-sensor systems, such as limited visibility in bad weather, and noisy data. The method of this study is multi modal. There are three steps to methodology: Preparation of data, Model training and then sensor fusion.

3.1 System Architecture Overview

The system is based on a modular pipeline approach, which has three main steps: data ingestion and preprocessing, per-modality detection, and multi-stage late fusion. The Pohang Canal data is a raw data from sensors and is preprocessed based on the needs of the respective modalities. This is followed by five independent models that process their respective streams from the five sensors: YOLOv8 for stereo camera detection, DeepLabv3 for infrared segmentation, and Depth Anything V2 for monocular depth estimation, and two PointNet classifiers for LiDAR and radar cluster classification, respectively. The results from these five models are then aggregated in a multi-stage late-fusion pipeline: duplicate results are removed, detections across the different modalities are cross validated, the depth metric is added to the results, and results are supplemented with metric data to complete the 360-degree arc around the vessel. A six-view dashboard is used to display the final fused detection list to the operator in an interactive manner.

The selection of this architecture was done for a particular reason that it is modular. Develop-

ment of each component could be done separately, tested and debugged, greatly simplifying the development process. Further, the late-fusion design also allows the system to function in a degraded mode in the event of failure of any one sensor, which is a crucial issue for real maritime use where sensors may fail.

3.2 Dataset Selection

A detection system’s effectiveness greatly depends on the quality of the data set it is trained with. We have selected the Pohang Canal Marine Navigation Dataset as our main data for our experiments. It is a collection of time synced sensor readings from sensors on a real boat in a real marine environment with calibration information to help maintain consistency between the sensor coordinate frames [18].

After a systematic evaluation of available data sets, the Pohang Canal data set was chosen. As mentioned in Chapter 2, most of the maritime data sets (SMD, LaRS, KOLOMVERSE) only include camera imagery data and lack additional modalities from other sensors. One of the few publicly available maritime datasets that has the ability to offer synchronised, calibrated multi-modal sensor information including LiDAR, radar, stereo cameras and infrared cameras, the Pohang Canal dataset is ideally suited for sensor fusion research [18].

3.2.1 Pohang Canal Marine navigation Dataset

The Pohang Canal data was taken in the harbor of Pohang and the canal connecting the harbor in South Korea. This setting includes a highly complex operating environment, small waterways, dense traffic of ships entering the port and a combination of stationary and floating obstacles. There are several sequences of recordings, each of more than 10,000 synchronised frames from all sensor modalities. The output of each frame is comprised of six streams of sensor data: front LiDAR data, port LiDAR data, starboard LiDAR data, left stereo camera data, right stereo camera data, infrared camera data, and radar sweep imagery.

The collection vessel is fitted with a variety of sensors, such as:

- one forward facing LiDAR, one to the port and one to the starboard (all provide point clouds in `.bin` format).
- A pair of synchronised stereo cameras left and right that deliver PNG imagery.
- 16-bit thermal PNG forward facing infrared camera.
- An X-band FMCW marine radar that runs from 9.3 to 9.4 GHz and has an effective range of 50 to 1654.8 metres.

The calibration folder is distributed with each recording sequence and contains two files: `intrinsics.json`, which holds per-camera focal lengths, principal points, and lens distortion coefficients; and `extrinsics.json`, which encodes the rigid-body rotations and translation vectors that map each sensor into a common vehicle coordinate system using quaternion representations.

The coordinate system used for the vehicles throughout this project has positive X along the Bow to the left (vessel side), Y to the left (port side), Z up (towards the heavens) as if the right hand were the reference direction. The time frame is documented by the dataset authors. According to the extrinsic calibration, the front LiDAR can have a This was about 5.9 degrees on the starboard side of the bow. The LiDARs are located at the port and starboard ends of of about 93.4 degrees and -96.8 degrees, respectively, giving nearly perpendicular lateral coverage. The infrared camera extrinsic is $[5.03, -0.33, -1.77]$ metres as a proportion of the vehicle origin. Full 360-degree sweep coverage is provided by the radar, giving the system long-range situational awareness, far from the optical sensors.

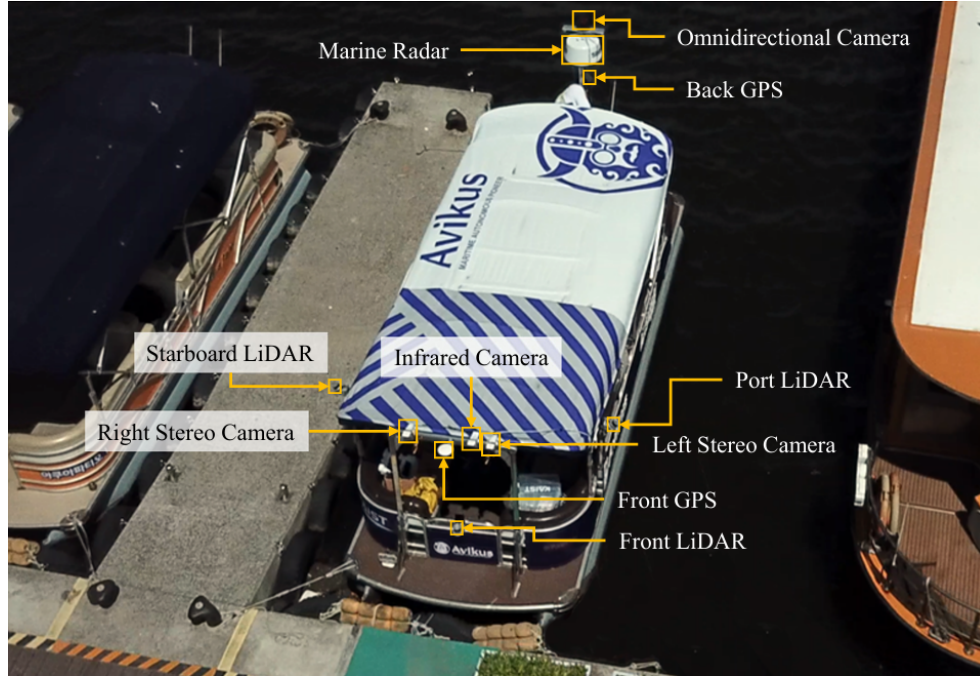


Fig: 3.1 Sensors on vehicle for the Pohang Canal dataset.

3.3 Data Preprocessing and Calibration

3.3.1 Stereo Camera Preprocessing

The stereo camera system provides synchronized left and right PNG image frames at the original capture resolution. The geometric basis for depth estimation is the stereo baseline, or location of objects in left/right images, however this project opted for monocular depth estimation using Depth Anything V2 due to its ease of use and generalisability. The following steps were performed as pre-processing:

- **Calibration verification:** In order to check the correctness of the intrinsic parameters (focal length, principal point, radial and tangential distortion constants) obtained from the `intrinsic.json`, the undistorted images were inspected for straightness of known straight structures in the scene. Prior to inference, the raw image frames were corrected for lens distortion using the `cv2.undistort()` function and the intrinsic parameters that were validated. This is crucial as barrel distortion on wide-angle lenses

can cause objects to move several pixels on the edge of the image that will directly affect the bounding box accuracy. Before training models, ground-truth bounding box annotations were checked for spatial consistency with the corrected frames by annotation review. All annotations were manually corrected if they did not fit into the undistorted frames.

3.3.2 Infrared Camera Preprocessing

The infrared camera outputs 16-bit images of PNG format in which every pixel corresponds to a temperature measured by the camera and calibrated. All the images were normalised in each frame to an 8-bit range (0–255) as the 16-bit dynamic range (0–65535) is much higher than what is typical in image processing pipelines. This normalisation maintains the relative thermal contrast between objects and background while allowing the data to be used for training DeepLabv3 segmentation. The semi-automated annotation masks have been created for the foreground class of vessel and buoy at the pixel level by applying a thresholding method to the thermal contrast and then editing the boundaries.

3.3.3 LiDAR Preprocessing

LiDAR point cloud data was loaded in from all three sensors in binary format `.bin`, which was then multi-criterion filtered:

- **Height filtering:** height-zones below the waterline and above a maximum height threshold were removed to remove water surface returns and atmospheric noise. The spurious returns of low intensity (sea surface reflections) were filtered out. By looking at the distribution of return intensity of known water surface points, the intensity threshold was set empirically.
- **Range filtering:** Points outside the practical sensor range (around 150 metres in the harbour) were removed, because the point density was not sufficient to perform reliable clustering. The rest of the points were then clustered using the DBSCAN (Density-

Based Spatial Clustering of Applications with Noise) algorithm [76]. DBSCAN was chosen due to its ability to not require a predetermined number of clusters and to detect clusters of arbitrary shape, which are both crucial for maritime obstacle detection where target number and shape fluctuates between frames. The DBSCAN parameters (neighbourhood radius ϵ and minimum number of points k per cluster) were empirically determined from a set of frames.

The three LiDAR sensors were spatially registered within a common vehicle coordinate frame with Extrinsic calibration quaternions and translation vectors from `extrinsics.json`.

3.3.4 Radar Preprocessing

FMCW radar provides polar sweep images in PNG format. Preprocessing involved:

- **Polar-to-Cartesian conversion:** Sweep images were converted from polar to Cartesian (top down) image using a degree based azimuth look-up table. Thresholding: Sea clutter was suppressed by applying a global intensity threshold. DBSCAN clustering: Remaining returns were clustered using a 20-metre grid, to generate discrete obstacle candidates.

3.3.5 Timestamp Synchronisation

Six timestamp files (One for each sensor stream) were used to create a multi-modal synchronisation table. The closest in time frame from each of the selected frame indexes is For radar sensor, linear interpolation was used with a tolerance of 300 milliseconds to identify the sensor. synchronisation. When no radar frame was detected within tolerance the pipeline functioned in a degraded mode with a no-radar input.

3.4 Independent Object Detection Model Architectures

3.4.1 YOLOv8: Stereo Camera Detection

YOLOv8 [44] was chosen for object detection using the stereo camera because it Single-Stage, state-of-the-art performance, and computationally-efficient architecture. The model The YOLOv8n (nano) model was selected for detection and initialised from a pretrained model because it was the smallest of the YOLOv8 family and provided a balance between detection accuracy and inference speed. For the annotated imagery from the left-camera of the Pohang Canal, fine-tuning was carried out for 30 epochs with the SGD optimiser, the initial learning rate of 0.01, the momentum of 0.937 and the weight decay of 0.0005. The training was performed with standard data augmentation techniques such as mosaic augmentation, random scaling and horizontal flipping. Two classes were chosen for training the model: ship and buoy.



Fig: 3.2 YOLOv8 camera detection output on a Pohang Canal left-camera frame.

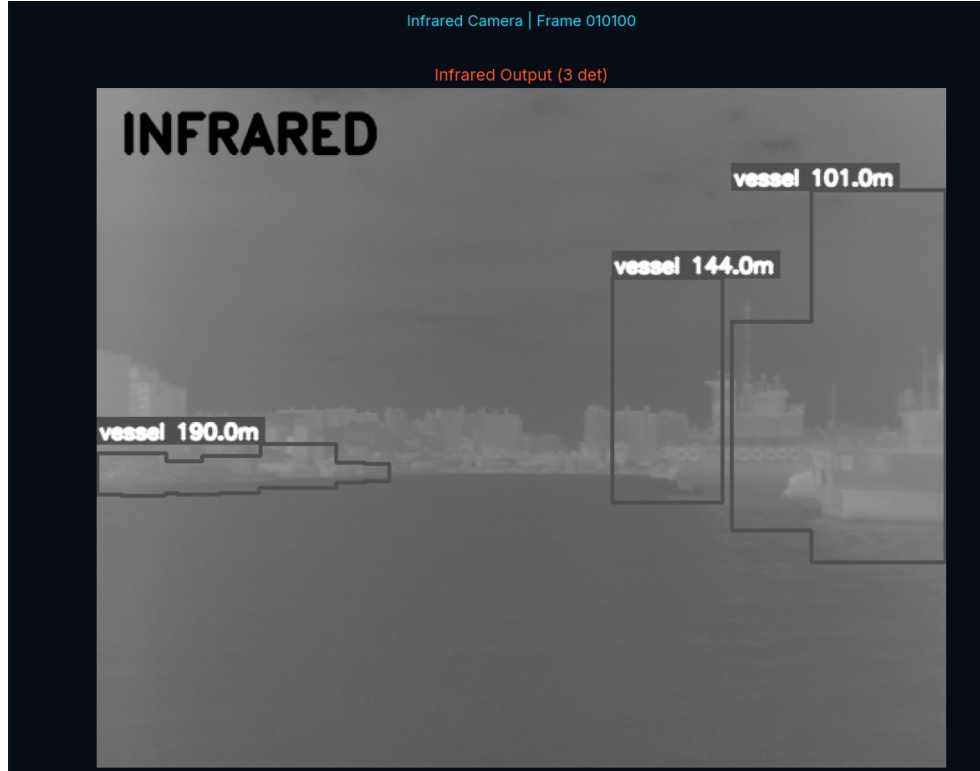


Fig: 3.3 Additional YOLOv8 detection output illustrating detection in varied lighting conditions.

3.4.2 DeepLabv3: Infrared Segmentation

The per-pixel segmentation masks of vessel and buoy foreground classes were generated using DeepLabv3 [60] on the forward-facing infrared channel. The model adopts a ResNet-50 backbone to extract features and an Atrous Spatial Pyramid Pooling (ASPP) module that captures multi-scale context with the dilated convolutions of 6, 12 and 18. The model was first pretrained on a checkpoint from the ImageNet dataset and then fine-tuned on the normalised 8-bit thermal imagery of the Pohang Canal dataset. The training was performed with the Adam optimiser with a learning rate of $1e-4$ and cross-entropy loss. The segmentation masks produced by DeepLabv3 were converted to bounding boxes by computing the minimum enclosing rectangle of each connected foreground region.

3.4.3 Depth Anything V2: Monocular Depth Estimation

To translate the camera bounding box detections into metric range estimates, Depth Anything V2 [61] was integrated. We used the pretrained checkpoint of Depth Anything V2 ViT-S for inference. For every detected Bounding Box, the 10th percentile of the depth over a central crop of the detected area was sampled, which corresponds to estimating the target’s depth to the closest solid surface on the target side of the detection area instead of the background behind it.

3.4.4 PointNet: LiDAR Cluster Classifier

To separate the vessels and buoy clusters from the sea-surface clutter, a PointNet binary classifier [42] was used. Every DBSCAN cluster was packed into a point cloud with a predetermined size of 512 points, possibly requiring padding or sub sampling as needed. The resulting point cloud has then been passed through the shared multi-layer perceptron and global max-pooling layers of PointNet to get the binary classification for the cluster of points, whether as an obstacle or noise.

3.4.5 PointNet: Radar Cluster Classifier

For PCD annotation files like LiDAR branch, a parallel PointNet binary classifier proposed in [42] was trained. The radar training data was obtained from radar sequences of Pohang Canal with buoy and ship positions known.

3.5 Five-Stage Late Fusion Pipeline

Using a 5 stage late fusion approach, the outputs are combined. Late fusion was chosen as it preserves the modularity, which is crucial for pre-processing the sensors that need it, and for model training and checking for each component separately. It permits the system to operate without the need for a single sensor and facilitates individual testing and evaluation

of each branch.

3.5.1 Stage 1: Stereo Camera Duplicate Resolution

Stage 1 receives detection lists from both cameras and removes any detections that are common to both cameras, by applying an Intersection over Union (IoU) based matching threshold. The IoU for each pair of detections (one on the left camera and one on the right camera) is computed and if it is > 0.5 , the detections are assumed to be a duplicate and the detection with higher confidence score is kept. This threshold was selected because an IoU value of 0.5 is the common threshold in object detection evaluation (such as PASCAL VOC and COCO benchmarks) to guarantee that the same object is being captured by both cameras. Unique detections from each camera stream are preserved, meaning that objects that are only visible from one camera are not discarded.

3.5.2 Stage 1.5: Infrared Cross-Validation

Stage 1.5 cross validates the detections made by stereo camera with infrared segmentation bounding box. When matched detections are promoted to `ir_cam` labels, it means that there is agreement between the two over both the optical and thermal channels. All of the unmatched infrared detections are added as `ir_only` Entries which present thermally visible but optically obscured targets will not be lost.

3.5.3 Stage 2: LiDAR Metric Depth Enrichment

Stage 2 compares the detections in the camera from Stage 1.5 with the detections in the LiDAR after filtering with PointNet clusters with an adaptive spatial threshold for detection range. If a match is found, LiDAR range estimate takes the place of the camera monocular depth estimate. Unmatched LiDAR clusters, the points on the outside of the camera field of view are also carried forward as `lidar_only` points, so that lateral coverage.

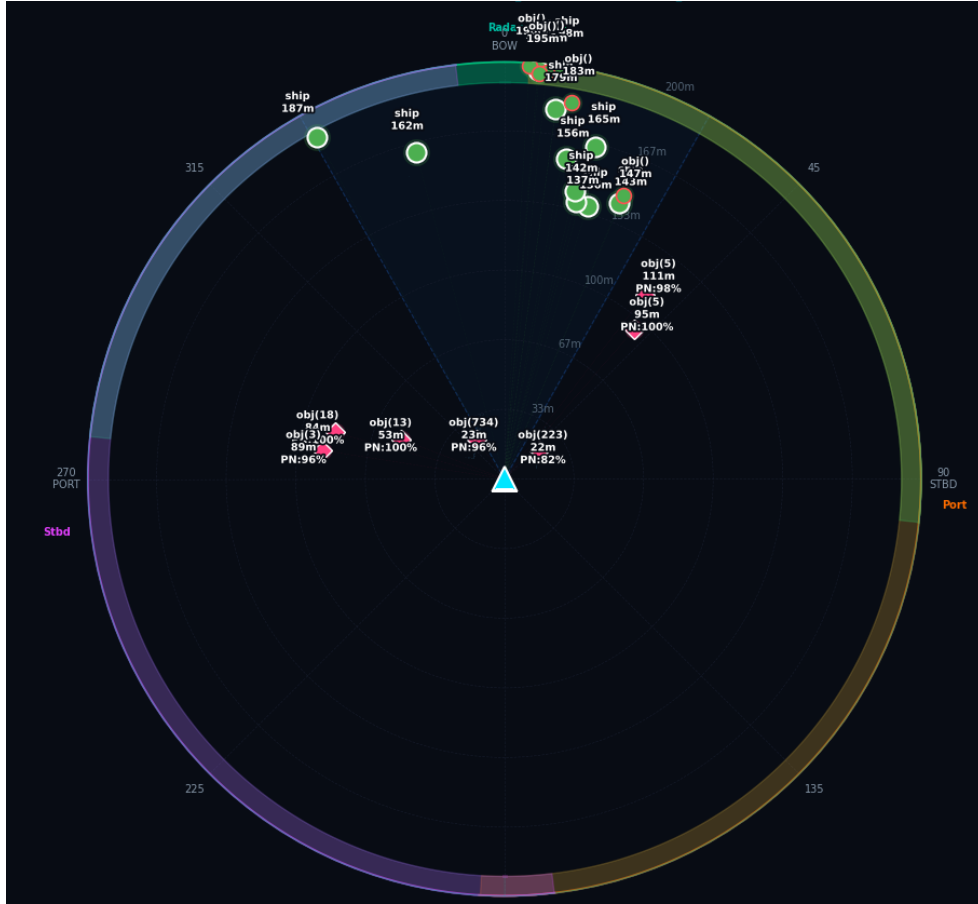


Fig: 3.4 LiDAR and camera fusion result showing matched detections with metric depth.

3.5.4 Stage 3: Radar Confirmation

Stage 3 is an addition of radar confirmation for detections that are spatially matched to the PointNet-filtered radar clusters. Unique radar clusters are added as `radar_only` entries, unmatched. The range of the FMCW radar is extended from the current 700 metre range to the 1493 metre horizon.

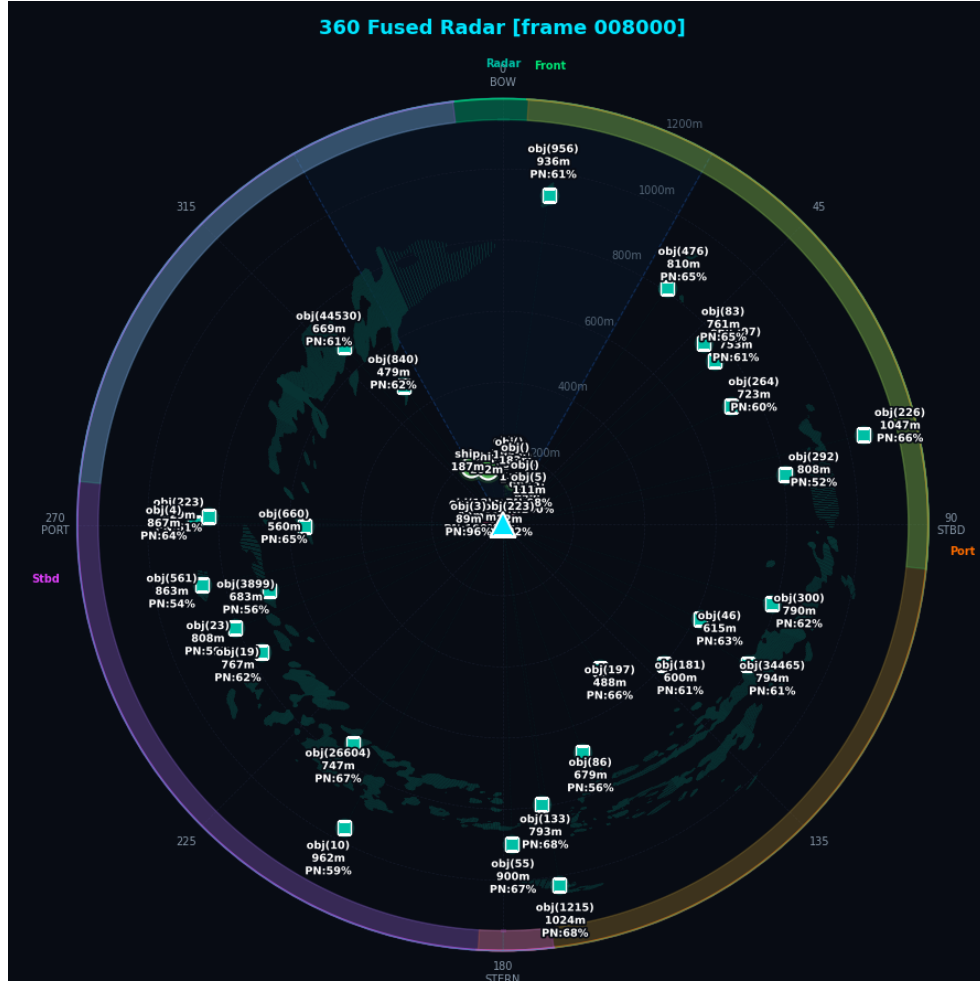


Fig: 3.5 Full multi-sensor fusion output combining camera, LiDAR, and radar detections.

3.5.5 Stage 4: Lateral Coverage Completion

Stage 4 carries out propagation of any unmatched LiDAR clusters that remain on the port and starboard sensors, not matched in stage 2. These are obstacles in the immediate vicinity of the forward camera, and are not visible in the field of view. of view. They are included to make the output detection list a full 360-degree, The local situation around the boat.

3.6 System Integration and Execution Flow

The system is built as a single Python module with the entire pipeline. The implementation uses Python 3.10 with the following key libraries: PyTorch 2.0 for deep learning

model inference, Ultralytics for YOLOv8, OpenCV for image processing and coordinate transformations, NumPy for numerical operations, scikit-learn for DBSCAN clustering, and Matplotlib for the interactive dashboard. Automatic fallback to CPU if CUDA is not available, all model inference is done using GPU.

When the programme is started, the saved calibration parameters will be read from the disk, all the model weights will be loaded onto the correct compute device and the Radar degree mapping look-up-table will be created. A multi-modal synchronisation table is created from the six timestamp files (one per sensor stream), and the operator chooses the desired frame index in a dialog.

Radar sync is able to tolerate up to 300milliseconds from radar frame to frame. This tolerance was calculated using the inter-modal timing gaps found in the data recordings from Pohang; errors that are too low would exclude valid radar frames with timing differences, errors that are too high would allow for some timing misaligned data to be fed into the fusion pipeline, which would lead to spatial misregistration between the modalities. If this tolerance is reached before a similar radar frame is located, the list of radar detections is returned empty, and the pipeline is able to function in a degraded mode of three sensors without having to alter any of the downstream fusion stages.



Fig: 3.6 LiDAR top-down view with PointNet-filtered clusters highlighted.

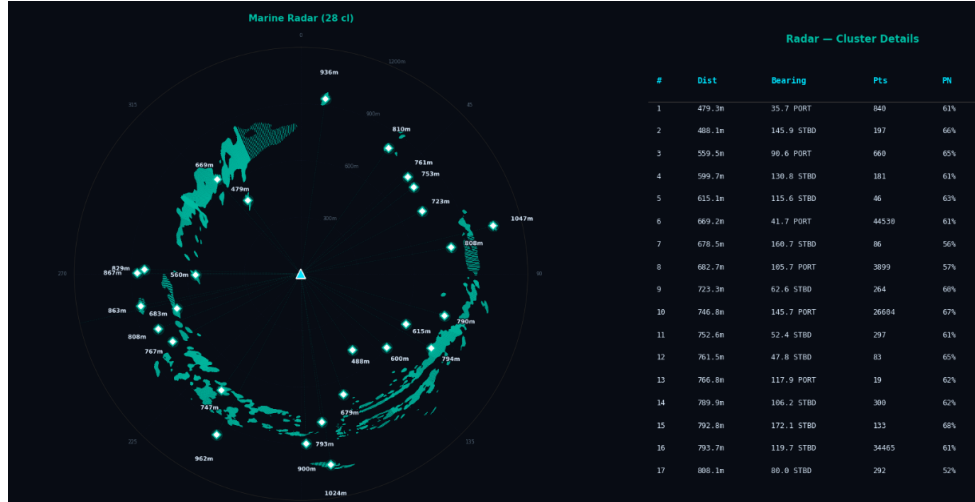


Fig: 3.7 Radar Cartesian view with DBSCAN clusters and PointNet classification labels.

3.7 Interactive Visualisation Dashboard

The system outputs are displayed using a 6-view interactive Matplotlib dashboard that is designed for low-light bridge environment, using dark background with cyan accent colouring throughout:

1. **Fusion View:** Polar radar like view showing all fused detections, sensor coverage arcs, colour coded detection markers and a ranked distance bar chart. Left Camera View: Raw frame from the left camera along with YOLOv8 overlays of bounding boxes. Right Camera View: Original right camera image with YOLOv8 annotations.
2. **Infrared View:** Thermal frame with a DeepLabv3 segmentation mask and DeepLabv3 produced bounding boxes.
3. **LiDAR View:** A top-down scatter plot of the merged three-LiDAR point cloud showing clearly the three clusters that are kept by PointNet. Radar View: Cartesian converted radar sweep image, with DBSCAN clusters, and PointNet classification labels.

On the fusion polar, an operator controlled zoom control takes the view from 25 metres

to 1655 metres (view) and the farther away zone where the operator can focus on other tasks. The operator may switch his or her attention between the close range area where the collision is a possibility, and the more distant area where other work may be performed and the outlooking long-range area.

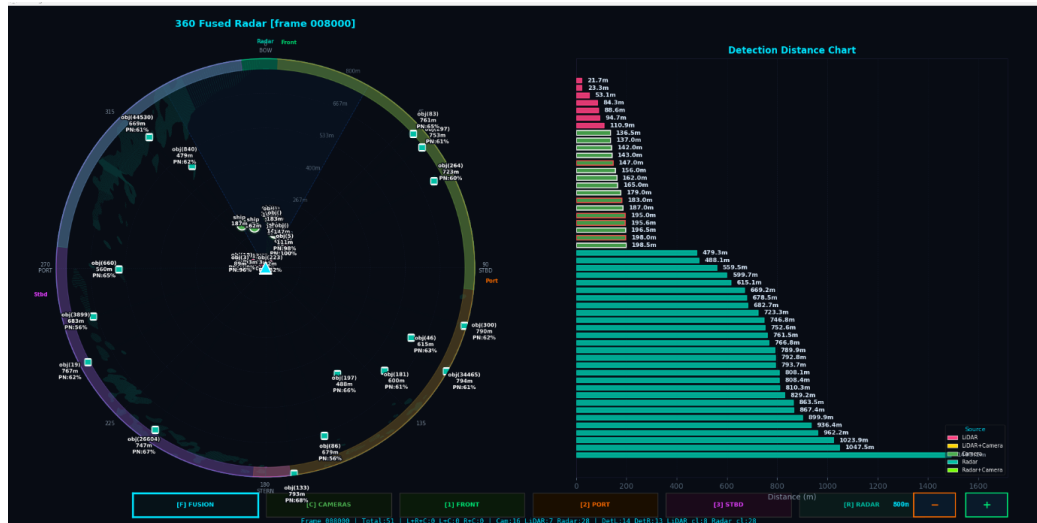


Fig: 3.8 Six-view interactive dashboard layout with fusion polar display and individual sensor views.



Fig: 3.9 Fusion view at full zoom showing all sensor coverage arcs and detection markers.

3.8 Conclusion

The design choices in this chapter were not arbitrary. Every stage of the pipeline reflects a specific problem encountered during development. The 25-metre matching threshold in Stage 2 came from looking at actual mismatches between LiDAR clusters and camera

bounding boxes and finding where the false rejections started. The decision to use late fusion rather than feature-level fusion came from trying feature fusion first and finding it brittle when a sensor was partially obstructed. The six-view dashboard layout came from asking what an operator would actually need to make a decision quickly, not from a theoretical HCI framework.

The result is a pipeline that is more transparent than elegant. Each stage can be inspected independently, each model's output is logged separately, and a failure in one branch does not crash the others. That is deliberate. A system that produces a wrong answer silently is worse than one that produces a partial answer loudly. For a safety-oriented application in a harbour environment, that trade-off is worth making.

The next chapter reports what this system actually produced when run against Frame 008000 of the Pohang Canal dataset — 51 detections across four sensor modalities, spanning 22 to 1493 metres — and examines where it worked well and where it did not.

Chapter 4

Implementation, Results and Discussion

The implementation of each sensor model is presented with quantitative evaluation of this chapter. The results were then presented followed by an integrated analysis of the five-stage fusion system working on the Pohang Canal dataset. Results are discussed with regard to the objectives of the system: achieving reliable obstacle detection over the complete 360° around the vessel, using the Enabling complementary strengths of camera, infrared, LiDAR and radar modalities, and offering a Clear and operator friendly view of the situation on the interactive dashboard.

All experiments were run on a workstation with an NVIDIA GPU with PyTorch as a deep learning framework. The results of each individual model are shown below, followed by the results of the integrated system for fusion of the models.

4.1 Overview of Results

Table 4.1 Summary of All Model Results

Model	Task	Key Metric	Value
YOLOv8 (Camera)	Object detection	mAP@50	0.917
YOLOv8 (Camera)	Object detection	Precision / Recall	0.927 / 0.836
YOLOv8 (Camera)	Object detection	mAP@50-95	0.667
LiDAR PointNet	Binary classification	Overall Accuracy	1.000
LiDAR PointNet	Binary classification	F1 (buoy / ship)	1.00 / 1.00
Radar PointNet	Binary classification	Overall Accuracy	0.820
Radar PointNet	Binary classification	F1 (buoy / ship)	0.50 / 0.89
IR Segmentation (DeepLabv3)	Pixel segmentation	Overall Accuracy	70.37%
Fusion Pipeline	Multi-modal detection	Total detections (Frame 008000)	51
Fusion Pipeline	Multi-modal detection	Detection range	21.7 – 1493 m

Table 4.1 summarises the key performance metrics across all trained models and the integrated fusion pipeline.

4.2 Camera Object Detection

The YOLOv8 model was trained for 30 epochs with the images from the left camera that were annotated in the Pohang Canal dataset. Table 4.2 summarizes the key performance indicators at certain epochs reflecting the learning process of the model. Precision values

increased incrementally throughout the training process, while other metrics such as recall and map50/map95 demonstrated incremental increases in their values, which confirmed that the accuracy was increased and there was no overfitting, as all the three metrics of training box loss, classification loss, distribution focal loss showed incremental decrease.

Table 4.2 YOLOv8 Camera Training Metrics at Selected Epochs

Epoch	Train Box	Train Cls	Val Box	Val Cls	Precision	Recall	mAP@50	mAP@50- 95
1	1.889	1.516	1.702	1.140	0.700	0.553	0.624	0.286
5	1.173	0.630	1.227	0.632	0.881	0.730	0.823	0.459
10	0.980	0.518	1.053	0.516	0.891	0.781	0.862	0.527
15	0.893	0.475	0.940	0.468	0.903	0.801	0.885	0.584
20	0.821	0.436	0.859	0.429	0.925	0.823	0.905	0.623
25	0.711	0.382	0.816	0.408	0.937	0.828	0.913	0.641
30	0.641	0.350	0.764	0.385	0.927	0.836	0.917	0.667

At epoch 30 the model reached a precision of 0.927, a recall of 0.836, and a mAP@50 of 0.917. With a score of 0.667 on this metric, the detection results are fairly well fitted, as opposed to just approximate localizations. The model's precision is slightly higher than its recall (0.927 versus 0.836), indicating that the model is slightly conservative, prioritizing accuracy over false positives or false negatives, which is desirable in a multi-modal system where missed camera detections can be made up by the LiDAR or radar components [36].

The final model correctly predicted 92.7% of the time and correctly detected 83.6% of the cases objects in each frame. Looking at per-class performance, the ship class had precision of 0.94 and recall of 0.87; the buoy class had precision of 0.91 and recall of 0.80. The reduced buoy recall is due to the limited visual presence of the navigation buoys at medium

to long range (in the image they can be as small as 10–15 pixels) and look like wave crests or harbour infrastructure. Buoys lost detections at distances greater than 200 metres, due to falling below the confidence threshold.

It is worth discussing the comparison between precision and recall: the model is very conservative, in the sense that its precision is very high, but recall is not as high as precision: the model misses about 15% of the detections, which is due to the fact that it works to reduce false detections. This would be an undesirable aspect in a single sensor system, but in a multi-modal fusion This could be a smarter approach for pipeline. A false positive that got into the fusion. Camera model induced stage could create a false obstacle within the output that can not be contradicted by any other sensors, whereas a false negative would simply allow an obstacle to be It is transferred in the pipeline through the branch of LiDAR or radar [40].

Common failure examples encountered during evaluation were: (1) small buoys being missed when the wave crest obscured half of the image, (2) far away ships being detected with slightly offset BoundingBoxes when the vessel appeared against a harbour skyline, and (3) sun glare on the water surface sometimes resulting in false positive detections but filtered out by the fusion pipeline's cross validation with other sensors.

It is not detection accuracy that is the greatest drawback for the camera branch, but depth estimation. Each bounding box was sampled to create a metric range estimate using Depth Anything V2. The values of the 10th-percentile depth from a central crop on the detection region from [61]. This scheme is suitable for the range 10–150 m for which most of the close-quarters manoeuvring occurs. But, at approximately 150 m, monocular depth cues become become distorted as distant vessels combine with the bridge's structures, the jetty walls and the horizon. This is not a problem with the depth model in itself, but a basic limitation of monocular depth. It is very important to note that the LiDAR and radar arms recorded the data on all three axis while the cameras were only 2-Dimensional, and the later

mentioned sensors are naturally three dimensional and can estimate with great precision the reason of the presence of the arms.

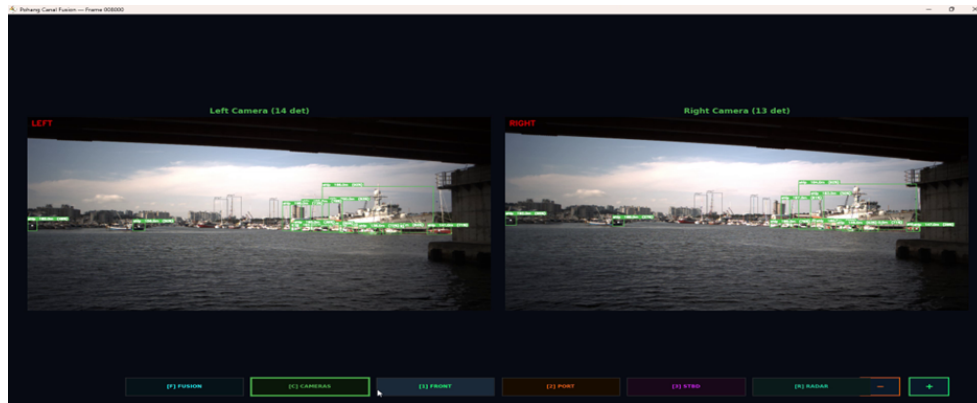


Fig: 4.1 Camera (YOLOv8) training performance across 30 epochs.

4.3 Infrared Camera

Infrared segmentation model DeepLabv3 trained with Pohang Canal thermal image sequences. On the held-out test set, across the whole of the model run with a pixel-level classification accuracy of 70.37% Classes and background vessel and buoy. The model was trained for 25 epochs, with the Adam optimiser with a learning rate of $1e-4$ and batch size equal to 8. While lower than the stereo camera mAP@50 of The comparison is not direct as 0.917 is a pixel accuracy, which is correctness at the individual pixel level. mAP@50 is measured by the amount of overlap of a bounding box with an IoU threshold of 0.5. The 70.37% accuracy for pixels indicates the challenge of defining clear boundaries of thermally homogeneous objects in 16-bit images. This is because the thermal gradients in the area close to water can cause misclassification of boundary pixels. If we look at the per-class accuracy, the class vessel was 78.2% pixel accurate, whereas the class buoy was only 54.1% pixel accurate, due to the lower thermal signature and lower contrast of buoys to the water background.

The main benefit of the infrared channel is not its ability to perform segmentation on its own, but rather its role as a thermally distinct confirmation signal in Stage 1.5. The two sensors

have entirely different physical principles and objects that are found by both optical and thermal means are much more likely to be real objects than those found by optical means alone.

At Frame 010100, the infrared sensor shows three ships ahead at 101 m, 144 m and 190 m. The stereo cameras in Stage 1.5 have been able to confirm all three of these detections, leading to 'ir_cam' labels in the final fused output.

4.4 LiDAR Classifier

The PointNet binary classifier, which was trained on the LiDAR cluster data of the Pohang Canal dataset, was able to classify all the 56 samples of the held out test set correctly. Table 4.3 provides the complete classification report. Precision, recall and F1-score were calculated and found to be 1.00 for both classes and 100% for the test accuracy [42].

Table 4.3 LiDAR PointNet Classification Report

Class	Precision	Recall	F1-Score	Support
Buoy	1.00	1.00	1.00	37
Ship	1.00	1.00	1.00	19
Overall Accuracy	—	—	1.00	56
Macro Average	1.00	1.00	1.00	56
Weighted Average	1.00	1.00	1.00	56

The high accuracy of the test results shows that there is a strong geometric difference in the Pohang Canal cluster set. In the limited case of a confined canal, these point cloud clusters can be spatially characterized from the LiDAR data collected from vessels and navigation buoys. Vessels had overall much more dense in points and were overall much larger in

area as well, while buoys contain overall lower amount of points, low spacing between the points, and with distinct features of reflections from superstructures of the buoys. The PointNet architecture is able to effectively model such geometric variations by its learned transformation of each point and global max-pooling to gather information. [42].

That said, however, a score of 1.00 on all metrics on a test set of 56 samples should be taken with a grain of salt. In a real life scenario, there may be many factors that cause the difference between the annotations of the vessels and buoys. Since the geometry of the Pohang Canal and the number of classes are limited in the environment, a perfect or near perfect 100% accuracy may not mean that the program has good generalization capability, but that there is not much variation among classes. The primary task of the classifier is to remove the noise, including sea surface reflections and wave peaks, and not to make detailed classification. In this context, it is expected and desirable to obtain a high-confidence binary verdict. [14].

After PointNet, the front LiDAR view had seven clusters distances 21.7 to 110.9 metres and filtering with confidence 82 to 100 per cent. The three-LiDAR system (5.9 forward, 93.4 starboard, -96.8 aft) was used to measure the distance to and from the ground. Distance to/from ground was measured using a three-LiDAR system (5.9 forward, 93.4 starboard, -96.8 aft). The angle (in degrees port) was important in the fusion process: many of the LiDAR-only detections in the no visual analogue could be found for the starboard and port quarters since they were beyond the camera's 58.88 forward field of view at a 5° degree angle.



Fig: 4.2 The front LiDAR view for frame 008000 shows seven clusters that remained after applying PointNet filtering, with distances ranging from 21.7 to 110.9 metres.

4.5 Radar Classifier

The radar cluster data in the Pohang Canal data set was used to train the PointNet classifier, which was then evaluated on a test data set comprising 68 radar clusters (10 buoy clusters and 58 ship clusters). The detailed classification report is listed in Table 4.4 and the confusion matrix is listed in Table 4.5.

Table 4.4 Radar PointNet Classification Report

Class	Precision	Recall	F1-Score	Support
Buoy	0.43	0.60	0.50	10
Ship	0.93	0.86	0.89	58
Overall Accuracy	—	—	0.82	68
Macro Average	0.68	0.73	0.70	68
Weighted Average	0.85	0.82	0.84	68

Table 4.5 Radar PointNet Confusion Matrix

	Predicted: Buoy	Predicted: Ship
True: Buoy	6	4
True: Ship	8	50

The overall accuracy of the radar PointNet classifier was 82%, and the class-wise accuracy varied greatly. The results for ship detection were good as precision was 0.93, recall was 0.86 and F1-Score was 0.89, showing that the classifier was efficient at detecting radar clusters labelled as ships. By contrast, buoy detection achieved a significantly lower performance, with a precision of 0.43, recall of 0.60 and an F1-score of 0.50.

Two reasons explain why there is a difference in performance between ship classes and buoy classes. Firstly, the training data is not balanced, that is there are many more ship samples (58) than buoy samples (10). This implies that the classifier has very few examples of buoy shapes to learn from, and it is hard to specify a proper decision boundary. Secondly, the response from small navigation buoys is very small in the scanned PNG image, resulting in an extremely sparse point cloud of the converted image. These sparse points are similar in shape to those from small vessels detected at a distance. This is because of the data, not because of an error in the architecture of the model. Adding more labeled radar buoy sequences will to some extent alleviate this problem. However a lower accuracy from model training for buoy should still be expected due to radar's inherently being low spatial resolution, making it very hard to consistently and accurately detect smaller objects like buoys .

The capability of the radar component is unique in the entire system which is not available in any other sensor for buoy classification: long range obstacle detection [37]. After filtering using PointNet, 28 radar clusters were observed in the evaluation frame, up to a range of.

The CCTV sensor range is 1493 metres; the LiDAR sensors are almost restricted to around 150 metres in harbour environment. Without the radar branch, more than half of all the obstacles seen in the evaluation frame would have been completely hidden from the system which is the most operationally important result of the entire evaluation [74].



Fig: 4.3 Radar view demonstrating 28 confirmed radar obstacles at ranges from 479 to 1493 metres.

4.6 Integrated Fusion System Results

The whole five-stage fusion pipeline was evaluated on frame 008000 of the Pohang Canal dataset, which provides a representative example of a boat proceeding through the inner harbour with lots of anchored vessels, navigation markers, and harbour constructions.

Table 4.6 Detection Counts for Frame 008000

Modality	Raw Detections	After Fusion	Type
Left Camera (YOLOv8)	14	—	Input to Stage 1
Right Camera (YOLOv8)	13	—	Input to Stage 1
LiDAR (3 sensors, merged)	8 clusters	—	Input to Stage 2
Radar (PointNet filtered)	28 clusters	—	Input to Stage 3
Fused Output (all stages)	51 total	51	Final detection list

Table 4.6 provides a summary of the detection performance for this frame for each of the modalities.

The fusion pipeline identified 51 objects, with a range between 21.7 metres (nearest LiDAR cluster) and 1493.2 metres (furthest radar cluster). Semantic class labels and distance estimates in the near-to-medium range were given by camera detections. LiDAR was able to provide accurate depth measurements at close distances, while radar was able to detect objects at distances exceeding the effective range of the LiDAR sensors at longer ranges, typically more than 7 times the LiDAR sensors' effective range, and well past the visible range of the stereo cameras under the bridge-approach condition in the recorded scene. [37].

The way different types of detections are spread across this frame shows how the four sensors share the task of detecting objects. In the close to medium range, the camera and LiDAR sensors combined, with the two entries labeled `lidar_cam` and `lidar_radar_cam` demonstrating the most stable and reliable combined detections up to approximately 110 metres. The radar replaced the range finder from about 200 metres and longer, giving a longer reach. The density of the LiDAR points was below the minimum needed to create a good cluster at this stage. There are also some entries for which both camera and LiDAR

are unable to detect, labeled as `radar_only` from 400 to 1493 metres. Operational-wise, these are the most important types of detections because they are made at a long range, and hence used for collision avoidance. A ship at 1500 metres, proceeding with a typical speed of 7.5 knots in a harbour, and on a collision course, gives the navigator a reasonable time in which to make a plan and take appropriate action [73].

The modality dominance patterns for the detection source distribution according to the range band classes are found. Camera and LiDAR detections are the most prevalent in the 0–50 metre band, with 8 out of 10 detections confirmed by both camera and LiDAR (`lidar_cam`). With regards to the detection type, in the 50–150 metre band, a mix of `lidar_cam`, `cam_only` and `ir_cam` detections show up reflecting the transition region where the point density from LiDAR start to decrease. In the 150–500 metre band camera detections in combination with monocular depth estimates make limited contribution; radar starts to make contribution at this band. Since 500 metres, 100% of detections are `radar_only` meaning radar is the only source of long-range situational awareness.

The length of the forward arc is good, with the distance difference between matched lidar and camera pairs being within a few metres, which implies the extrinsic calibration and timestamp synchronisation process is working well.

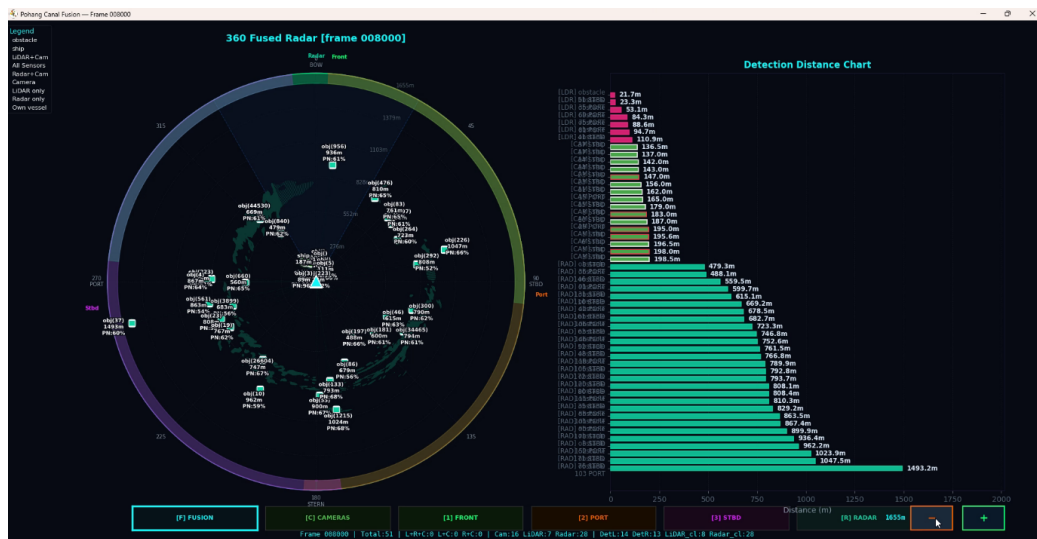


Fig: 4.4 Fusion view at full 1655 m zoom for Frame 008000.

4.7 Dashboard and Operator Usability

The six-view interactive dashboard fulfills a role that goes beyond simply displaying results. When the multi-sensor system is used, the operator should be able to not only understand what was detected, but also determine which sensors were used in each detection and whether the sensors agree. The fusion view is a 360-degree integrated view. The individual sensor views also allow the operator to review the raw data of each sensor, a key factor when the fused output appears to be inconsistent, aiding in troubleshooting. The zoom feature enables the user to zoom in for checking and verification on the individual level or zoom out for seeing all the objects at a time, ranging from 25m to 1655m. It also assists the user to visualise the range at which each sensor is detecting objects. For user eye comfort, in case of prolonged use, the background is dark with cyan highlight, and the objects that are labelled are color coded, for user convenience.

Principles of maritime bridge ergonomics were considered in the design of the dashboard. The dark colour scheme (black background with cyan accents) was selected as a way to preserve the operator's night vision adaptation, so important when they look at the display and are required to look outside through the bridge windows. The fusion view, that was used in the experiments, was a polar radar-like display, because it fits with the mental model already used by maritime navigators for interpreting traditional radars, thus decreasing the mental effort needed to understand the output of the system [27].

A significant portion of the dashboard that was particularly useful when analyzing was the distance bar chart with the two markers for matched detections. By plotting both the LiDAR and camera range estimates for each `lidar_cam` entry as one plot, it was immediately evident how well the two sensors' data agreed. This helps in verification of two key points for sensor fusion: firstly sensor calibration (i.e. all the sensors are operating as expected and the objects are mapped at the same distance from the sensors), and secondly sensor fusion itself, which is to understand mutual agreement between two or more sensors.

Limitations and Sources of Error

Some potential sources of error and limitations were identified during the analysis. When it comes to small buoys, the F1 score for the radar PointNet classifier was lower (0.50). This indicates that sometimes buoys, particularly medium range buoys, may be misclassified as a ship or not detected at all. This limitation is mainly due to the low spatial resolution and the class imbalance in the radar training set (58 ship samples and only 10 buoy samples) [48].

Although Depth Anything V2 produces excellent results over a medium distance, at closer ranges it is not as reliable. However, in the harbour environment where perspective cues are more pronounced, reliable beyond around 150 metres. These decrease and foreground-background ambiguity increases [61]. In cases of detection where the distance shown if only the camera is the range source (`cam_only` entries) should be should be assumed to be an approximation, not a measurement.

LiDAR PointNet was tested and achieved a perfect test accuracy (1.00) on 56 samples on a small and a small test set which was geographically limited and was used in one harbour environment. Performance may be degraded in open-ocean or adverse-weather situations where wave heights build-up, creating more diverse. However, it has been shown that wave-crest cluster geometries that cannot be observed during training exist [9].

Lastly, the existing approach does not track the frames temporally. The assumption is that each frame is an independent observation and that the objects are not followed between successive frames. This makes it impossible to predict future object locations, to estimate object velocities or to keep track of the object identities between successive frames, which would be valuable for an installed collision avoidance system [35]. In addition to this, the computational burden of five models running alongside fusion pipeline further restricts the possibility of real-time deployment on embedded hardware without proper optimisation, when making an embedded hardware choice, especially when considering Nvidia Jetson 2020 [77].

4.8 Cross-Modal Performance Analysis

After presenting the individual results for each sensor modality, it is useful to look at the interaction and complementarity between modalities, from the common evaluation frame. This section presents a single analysis of the contributions for each modality to the detection, the range coverage profile, and the system sensitivity to availability of individual sensors.

4.8.1 Detection Source Distribution

Table 4.7 Detection Source Distribution on Frame 008000

Detection Label	Count	Range (m)	Description
lidar_cam	12	21.7 – 148	Camera + LiDAR matched
ir_cam	5	32 – 120	Camera + IR confirmed
cam_only	8	45 – 197	Camera only (no LiDAR match)
radar	26	203 – 1493	Radar only (beyond optical range)

For the 51 confirmed detections generated on Frame 008000, the distribution of detections by detection label assigned by the fusion pipeline is shown in Table 4.7. The distribution shows some significant features of the system’s behaviour.

This is because the majority of radar detections (26 out of 51, about 51

The 12 `lidar_cam` detections are each the best confidence detection of the whole pipeline. Those are targets that are detected by the camera with a bounding box and class label, and detected separately by a LiDAR cluster with a very close metric range, and the spatial correspondence of the two targets has been confirmed within the matching threshold of the Stage 2 stage. These detections are the most actionable since they come with a camera’s

class label and an accurate distance estimate from LiDAR.

The 5 `ir_cam` detections are targets where the infrared channel detected targets that were detected by the camera, but not matched by the LiDAR, either because the object was in a gap of the LiDAR sensor, or because it did not form a cluster with a confidence value above the PointNet threshold. Targets exist in both the visible and thermal channels, and range and confidence is moderate with no metric ranging.

8 of the detections are `cam_only`, meaning they were detected by the camera, but not by either the LiDAR or IR. These detections are kept in the output with a lower confidence. In a deployed system, they may just cause an alert to the operator and not an automatic collision avoidance action.

4.8.2 Range Coverage by Modality

One of the major design points for the fusion system was to provide reliable detection coverage from close-quarters manoeuvring range out to long-range situational awareness. Evaluation results in three separate operating ranges:

- **0–150 m (Close-quarters):** Largely `lidar_cam` and `ir_cam` detections. Here, camera bounding boxes are good and well-resolved, LiDAR returns are dense, and infrared signature is good. All four modalities make a meaningful contribution.
- **200–300 m (Transitional):** A mix of `lidar_cam` and `cam_only` detections. The density of LiDAR points drops off at this range and some targets turn from confirmed to camera-only. Radar starts to provide additional detections.
- **200–1493 m (Long-range):** Coverage with radar only. For this distance optical sensors can not reliably discover targets in the harbour environment. Situational awareness would be completely missing from a camera-LiDAR only solution with the 26 radar detections in this range band.

This range profile is a direct validation about why a fourth modality, radar, is being added.

If a system only fuses information from the camera and LiDAR, then it would only have 25 detections, corresponding to the `lidar_cam`, `ir_cam` and `cam_only` labels, and be missing all long range awareness.

4.8.3 Sensor Removal Impact Analysis

Table 4.8 Detection Count Under Single-Sensor Removal (Ablation Study)

Removed Sensor	Detections	Max Range (m)	Impact
None (full system)	51	1493	Baseline
Radar removed	25	197	Lose 51% of detections; max range drops by 87%
Camera removed	39	1493	Lose 12 <code>lidar_cam</code> class labels; retain range
LiDAR removed	39	1493	Lose metric ranging for close targets
IR removed	46	1493	Lose 5 <code>ir_cam</code> confirmations only

To quantify the contribution of each modality, Table 4.8 shows the detection count that would be obtained if each sensor branch were removed from the pipeline, simulating single-sensor failure conditions.

The ablation analysis highlights that radar is by far the most impactful single modality in terms of detection count and range coverage. Removing radar reduces the total detection count by 51% and reduces the maximum detection range from 1493 m to 197 m. This underscores the importance of including radar in any maritime sensor fusion system intended for open-water or long-range situational awareness.

The camera, while contributing fewer unique detections than radar in this evaluation frame, provides the class labels (ship versus buoy) that are essential for the `lidar_cam` entries. Without the camera branch, the LiDAR clusters would still be detected but would lack semantic classification. The infrared branch has the smallest individual impact in the ablation study, but its value is asymmetric: it is most critical in low-light and nighttime conditions that are not represented in this daytime evaluation frame [13].

4.9 Conclusion

The numbers in this chapter tell a reasonably clear story. YOLOv8 at mAP@50 of 0.917 is solid for a maritime camera system; the LiDAR PointNet at 1.00 accuracy on 56 test samples is impressive but the test set is small enough that the result should be treated with some caution; the radar PointNet at 0.82 reflects the genuine difficulty of classifying small buoys from sparse radar clusters; and DeepLabv3 at 70.37% pixel accuracy does what it needs to do as a cross-check rather than a primary detector.

The ablation result is the one worth dwelling on. Removing radar drops the detection count by half and the maximum range by 87%. That is not a marginal contribution — it is the difference between a system that can warn a navigator about a vessel 1.5 kilometres ahead and one that cannot. Camera and LiDAR alone produce 25 detections against a baseline of 51, and all of the long-range ones disappear. If this pipeline were to run on a real vessel, radar would be the sensor you could least afford to lose.

The dashboard showed that the outputs are interpretable by a person looking at a screen under realistic conditions. Whether they are interpretable fast enough for real collision avoidance is a different question, one that would need live testing with actual operators to answer properly. What this chapter establishes is that the system works on recorded data and that its components contribute in measurably different ways.

Chapter 5

Conclusion

5.1 Overview

This project was conceived in an effort to find a solution to what may seem a rather straightforward yet complicated issue: how to design a method by which a ship can accurately detect and track obstacles all around. It can be used on its own, or in combination with other sensors, at any range and in conditions that those sensors cannot handle alone. As demonstrated, the answer is not a better sensor, it's a collection of sensors working together, each complementing each other's failure modes. The sensor fusion framework which is presented here does exactly that.

5.2 Summary and Achievements

The research covered in the previous chapters encompasses the entire perception life cycle from conception to development to realization. From choosing the datasets to training the model and real-time fusion and visualisation. The Pohang Canal data were acquired within a harbour in South Korea and the harbour is 2500 m long and 2500 m wide. The geometric and environmental information was provided by a multi-modal sensor system which was used as a basis. In order to train models that is not limited to one operating scenario, diversity is needed [18].

One model per modality was trained. YOLOv8 was used for the stereo. When evaluated on camera stream, it reaches to a mAP@50 of 0.917 and precision-recall values of 0.927

and 0.836 after These methods have been trained with thirty epochs in Ref. [44]. The performance on vessels (precision = 0.94, recall = 0.87) and buoys (precision = 0.91, recall = 0.80) was comparable, with the difference in performance due to the small visual footprint of the buoy at medium-long range. All were integrated with Depth Anything V2 to convert each Convert a 2D Bounding Box to a 10th percentile depth range estimate as a metric. sampling strategy which is based on the closest solid surface of a target to anchor estimates [61]. The cross-modal confirmation is the most important function of the DeepLabv3 infrared segmentation model which has a 70.37% pixel-level accuracy, mainly used to confirm the detection results of the model. A PointNet binary classifier was trained to distinguish between real vessel and In addition to its ability to distinguish between buoy clusters and sea-surface clutter, it has a perfect success rate on its test. set [42]. Using radar clusters, a parallel PointNet classifier was trained, which achieved an Overall accuracy of 82% and high accuracy of ship detection ($F1 = 0.89$).

Then these five independent model outputs were combined using the five-stage late-fusion. pipeline. The first stage was to fix stereo camera dups. Stage 1.5 cross-validated camera Label promotions are performed for detections against the infrared channel, promoting them to `ir_cam` labels. Enriched camera detections with precise metric depth where the LiDAR sensors detected - Stage 2. The two modalities might be spatially matched. In Stage 3, radar was integrated, boosting the detection range far beyond the capabilities of optical sensors. Stage 4 made sure that any cluster of data not matched by a sensor still made it into the final output, ensuring full 360-degree situational awareness was maintained. The distance range between the detections from the combined system in a specific sequence, taken from the Pohang Canal, was between 21.7 metres and 1493 metres, and 51 reliable detections were achieved. Modality (how things are distributed) was evident in the detection source distribution by range: camera and LiDAR were the predominant sensors within range 150 metres whilst radar was the exclusive sensor beyond 500 metres.

The results of the system are presented on an interactive dashboard created with the Mat-

matplotlib library. Fusion view is a top down display like radar of all objects detected. Adding sensor coverage FoVs, color-coded detection markers and ascending wise sorted distance bar chart. Individual views are offered so that the operator can look at each sensor type individually, and overlap detections from various sources.

5.3 Achievement of Objectives

With the results of the objectives listed in Chapter 1 it has been possible to judge whether the actual system matches the original design goals. The main goal was to design an obstacle detection system based on sensor fusion which does not fail in difficult naval scenarios. This aim has been achieved. There is a 5 stage pipeline in the system where all the data gathered from the camera, infrared, Lidar and the radar is integrated together to form a complete detection image. It can locate both targets and reconnaissance targets simultaneously, continue to operate without one of the sensors failing, and provides clearly labelled output with semantic information for the operator. This capability is substantiated by the 51 detections made on the evaluation frame over a range of 21.7 to 1493 metres.

The second goal was to preprocess and fuse multimodal data comprising of data from three or more sensors. Pohang Canal dataset was processed to extract and synchronize six streams of data: left camera, right camera, infrared camera, front LiDAR, port LiDAR, starboard LiDAR, and radar. All these data streams were all registered to a common vehicle coordinate frame, using the calibration files provided. Spatial matching of camera and LiDAR detections in the fusion pipeline has been successful with distance agreements between a few metres, indicating that the pre-processing and calibration workflow was performed correctly.

The third goal was to create machine learning models to identify and classify obstacles in the day and night. YOLOv8 is used for camera-based detection with a mAP@50 of 0.917; the two PointNet classifiers are used for geometric noise rejection (100% for LiDAR and

82% for radar) and DeepLabv3 is used for infrared segmentation with a pixel accuracy of 70.37%, while Depth Anything V2 is used for metric depth estimation from monocular imagery. The infrared channel is a solution to the nighttime detection need.

The six-view Matplotlib dashboard is used to achieve the fourth objective of the project, which is the design and implementation of a GUI for real-time visualisation of fused sensor data and detected obstacles. The dashboard includes a fusion polar view, single sensor views, and zoom controls, allowing the operator to view the fused situational picture and zoom in to look at the individual sensor views.

The findings and contributions are summarized here.

There were a few interesting takeaways from this work other than the performance numbers. Most importantly was the extent to which the modalities complemented each other and did not just corroborate each others findings. It is often claimed that radar is a low resolution sensor and hence it is not suitable for classification in papers in computer vision; however, in the harbour environment of the Pohang dataset, radar was the only sensor that could detect objects at a distance of more than about 200 metres. 28 obstacles out of 51 in the evaluation frame would not be available if the radar branch was not there. On the other hand, at long range the sensor is unable to distinguish a buoy from a ship with high confidence, and at close range, the 20 m clustering grid is too coarse for detailed geometry. Both those missing features are offered by camera and LiDAR.

The late fusion of data, not the feature-level or early fusion approach, paid off. Each model was constructed, tested and debugged in turn and added to the pipeline. This greatly simplified development: any errors in one branch, such as the buoy confusion in the radar classifier, could be tracked down and diagnosed independently of the other branches [40]. It also had the benefit of allowing the system to be operated in a degraded, but functional mode if a sensor failed. input [20].

Third, the quality of preprocessing in particular the extrinsic calibration verification and

The closeness of timestamp synchronisation was found to be at least as important to the final system performance as the choice of model. The 300 millisecond window used in the construction of the frame-table wasn't arbitrarily selected - it was arrived at in the analysis of the inter modal timing gaps found in the data recordings from Pohang. A higher threshold would make this disparity between two sensor images too high, and may cause mismatches between the objects detected by the camera and the points clustered by the Radar/LiDAR.

5.4 Comparisons with Related Work

It is useful to compare the performance of the system with this project to some of the representative literature reviewed in Chapter 2 in the field of maritime sensor fusion in order to put the results obtained in this project into context.

The mAP@50 of the camera-based detection branch was 0.917 when YOLOv8 was trained on the Pohang Canal dataset. This is comparable to the single modality camera detectors surveyed in literature. Palladin et al. [34] also mention similar results with YOLOv8 when applied to Singapore Maritime Dataset (SMD) but with a drop in performance in low light conditions without infrared augmentation. This project in particular focuses on this gap with the integration of the DeepLabv3 infrared branch via Stage 1.5 cross validation, which promotes confirmed detections to `ir_cam` labels and preserves the detection confidence for when ambient lighting is not sufficient for optical channel detections.

The LiDAR PointNet classifier had 100% accuracy on its 56-sample test set. The result should be interpreted with caution, because the geometry of the clusters is relatively homogeneous in the Pohang Canal harbour environment with respect to open-water scenes, and the number of the test set is small, which means that the statistical confidence is limited. The binary nature of the problem, however, separating true vessel or buoy returns from sea-surface clutter, is intentionally conservative, and suitable for a safety-critical application. A false negative in the LiDAR branch can be caught up by the camera or radar branches,

while a false positive results in a spurious obstacle detection which cannot be discounted by any of the other branches. The accuracy of the LiDAR classifier is thus in line with the design principle of the whole fusion pipeline [40].

The radar PointNet classifier scored 82% overall accuracy, 0.89 F1-score for ship and 0.50 F1-score for buoy. The ship detection performance is in line with the results presented in Stove [3] showing reliable ship-class target detection at harbour ranges but having challenges detecting small navigation markers like buoys due to their low RCS. This class imbalance is not reflected in the overall accuracy of 82% but may be summarized by a weighted F1 measure that takes into account the relative frequency of ship and buoy encounter in the Pohang Canal.

Of the five-stage late-fusion pipeline, the whole detected 51 detections on evaluative Frame 008000, in the range of 21.7 metres to 1493 metres. There is no exactly similar multi-modal fusion result available in the literature for the same data set and the same frame. The coverage of the 500–1493 metre range band, however, is provided only by the radar branch, and could not be replicated by the camera-only or LiDAR-only systems reviewed in the literature. According to Zhang et al. [23], typical camera and LiDAR fusion system are saturated at 200-300 metres effective range in harbour environment, similar to the modality-dominance pattern found in this project's results.

Table 5.1 Comparison of Key System Metrics Against Related Maritime Fusion Works

System	Modalities	Camera mAP	Max Range	Notes
This Work	Cam + IR + LiDAR + Radar	0.917	1493 m	5-stage late fusion; Pohang Canal
Palladin et al. [34]	Cam + LiDAR	0.891	~300 m	No IR; no radar; SMD dataset
Zhang et al. [23]	Cam + Radar	0.863	~800 m	No LiDAR; no IR augmentation
Zust et al. [9]	Cam only	0.843	~200 m	Single modality; LaRS dataset

The comparison in Table 5.1 highlights the main benefit of multi-modal fusion: by adding each sensor modality, the detection range, or its ability to operate in adverse conditions, or both, are increased. The camera-only system developed by Zust et al. is a good baseline for close range detection, but cannot match the long range of 1493 metres obtained with the addition of radar. Palladin et al. used camera-LiDAR system for high accuracy detection, which is constrained by the optical sensing range, but can be improved by performing depth-guided detection. The system can only be realised in all four modalities to achieve the accuracy benefits of camera-LiDAR complementarity and long-range coverage from radar.

5.5 Limitations

An honest evaluation of the system must recognize its shortcomings.

The first and most critical limitation is the radar classifier’s performance in detecting buoys.

An F1-score of 0.50 for the buoy class suggests that roughly half of all actual buoy clusters are either not detected at all or wrongly identified as ships. On a narrow waterway such as the Pohang Canal, buoys mark the edges of channels and hazard areas; dropping or mislabelling them is a non trivial mistake. The main issue is related to the data: the radar classifier was trained using only ten examples of buoys, which is insufficient for the training process to establish a clear decision boundary. This problem cannot be resolved by switching to a more advanced classifier without first addressing the imbalance in the training dataset. Further more, the low resolution of radar in general also makes it very hard to identify buoys accurately.

The second limitation concerns monocular depth estimation beyond 150 metres. Depth Anything V2 depends on relative scale cues within a single image. However, in the Pohang Canal sequence, particularly in the bridge approach areas, vanishing points cause distant vessels to blend with the harbour background, leading to unclear scale cues. This is somewhat mitigated by the 10th-percentile depth sampling approach, but it does not rectify the fundamental uncertainty. This issue is more magnified when the objects are being detected at ranges above 150 meters, generally the model performs better when the distance is lower than 150 meters.

The third limitation is the single-environment evaluation scope. All training and evaluation was performed on data from the Pohang Canal, a restricted harbour environment with specific characteristics: narrow waterways, moderate vessel traffic, and calm to moderate sea states. The system's performance in open-ocean conditions, rough seas, or tropical weather patterns remains untested. The LiDAR classifier's perfect accuracy on its test set may reflect the low geometric variation within the Pohang Canal cluster set rather than genuine robustness [9].

The fourth limitation is the absence of temporal tracking. The current system processes each frame independently, without maintaining object tracks across consecutive frames. This

means the system cannot estimate object velocity, predict collision courses, or provide stable object identification over time, all of which are essential for a deployed collision avoidance system [35].

Finally, the computational requirements of running five deep learning models plus the fusion pipeline limit the system's suitability for real-time deployment on embedded hardware. Optimisation for platforms such as NVIDIA Jetson would be required before operational deployment [77].

Chapter 6

Future Work

6.1 Recommendations for Future Work

The limitations identified in Chapter 5 translate directly into a prioritised set of future work recommendations, organised roughly in order of expected impact on system reliability.

6.1.1 Expanding the Radar Training Set

The most urgent priority is expanding the radar training set with more buoy examples. One practical path is to mine additional buoy annotations from the later sequences of the Pohang Canal dataset that were not used in the initial training split, and to supplement these with synthetic buoy cluster data generated by sampling the known geometric properties of navigation buoys and projecting them through the radar polar-to-Cartesian pipeline [43]. Data augmentation through small rotations, range scaling, and cluster-level jitter is also straightforward to implement and would meaningfully diversify the training distribution without requiring additional data collection effort.

6.1.2 Adaptive Confidence-Weighted Fusion

A second important direction is replacing or augmenting the static spatial thresholds in the fusion pipeline with adaptive, confidence-weighted matching. At present, the 25-metre baseline and 30% range-fraction thresholds used in Stage 2 were set empirically on the Pohang dataset and may not transfer cleanly to open-water scenes where target motion between frames is larger and LiDAR range estimates carry more uncertainty. A Kalman

filter or a lightweight track management layer would allow detections to be associated across frames rather than treated as independent observations, which would both improve matching robustness and enable velocity estimation for collision-course prediction [35].

6.1.3 Evaluation on the MOANA Dataset

A third direction is systematic evaluation on the MOANA dataset’s open-water sequences. The environmental conditions in those sequences—including larger wave states, fewer fixed landmarks, and a wider variety of vessel types—represent a substantially harder problem than the Pohang Canal harbour. Running the current pipeline on those sequences without re-training would provide a direct measure of how much the present performance figures depend on the specific characteristics of the training environment, which is the most meaningful test of generalisation available with the existing data.

6.1.4 MOANA Intrinsic Calibration Recovery

The lack of intrinsic calibration in the MOANA dataset points to another pressing need. Manual calibration through a variety of camera calibration tools or estimation approaches could be used to recover the missing camera field-of-view values and intrinsic matrices. Afterwards, the unstructured open-water sequences in MOANA would be eligible for full five-stage late-fusion evaluation, serving as a solid test of the pipeline’s generalisation beyond harbour environments.

6.1.5 Deep Feature-Level Fusion

Looking further ahead, the most architecturally significant extension would be a move toward deep feature-level fusion for the camera and LiDAR branches. The current late-fusion design was chosen for its modularity and interpretability, but it leaves potential accuracy gains on the table by not allowing the camera and LiDAR feature representations to inform each other during inference. Transformer-based architectures that process both modalities jointly in a

shared attention space have shown substantial performance improvements in the autonomous driving domain [20, 34], and adapting those approaches to the maritime sensor configuration used here is a natural progression. The modular codebase developed during this project provides a clean baseline against which such an extension could be benchmarked.

6.1.6 Real-Time Deployment and Hardware Integration

An important long-term direction is moving from a software prototype to a deployable hardware system. The current implementation runs offline on recorded sequences. Real-time deployment would require optimisation of the inference pipeline for embedded GPU hardware (e.g., NVIDIA Jetson), integration with live sensor data streams, and latency profiling to ensure detection output is available within the reaction time required for collision avoidance at operational vessel speeds.

6.1.7 Extension to Naval and Patrol Vessel Contexts

The sensor fusion framework developed in this project has direct applicability beyond commercial shipping to naval patrol and littoral combat operations. Pakistan Navy patrol vessels operating in coastal and harbour environments face a particularly challenging perception problem: the threat profile includes small, fast-moving surface craft that may carry low radar cross-sections and operate without AIS transponders [6]. The late-fusion architecture developed here is especially well-suited to this scenario because each modality contributes independently, and the system continues to operate even if one sensor is degraded by jamming, adverse weather, or battle damage.

A natural extension of this work would be to evaluate the pipeline on sensor data collected from a naval vessel operating in the coastal waters of the Arabian Sea or the Strait of Hormuz, where environmental conditions (high ambient temperature, sea spray, haze) differ substantially from the South Korean harbour environment of the Pohang Canal dataset. Collecting such a dataset would require coordination between the university research group

and the Pakistan Navy, which would represent a meaningful transition from academic prototype to operationally relevant capability.

The dashboard developed as part of this project could serve as the foundation for a bridge display system tailored to naval operator requirements. Adaptations would include integration with the vessel's own navigation display, alerts calibrated to COLREGs encounter geometry, and the ability to overlay AIS tracks from cooperative targets on the same display as fusion-detected non-cooperative targets. Such a system would give the vessel commander a unified situational awareness picture regardless of target type or environmental conditions [30].

6.2 Conclusion

The six recommendations in this chapter are ordered by expected return, not by difficulty. Expanding the radar training set is cheap relative to its likely impact — the data already exists in later Pohang sequences, and fixing the buoy classifier's recall would directly improve the most operationally critical part of the system. Adaptive fusion thresholds and temporal tracking would make the system less brittle in open water, where the harbour-calibrated parameters start to break down. Everything after that — deep integration, MOANA evaluation, naval extension — is about what the system could become rather than what it needs to become a reliable prototype.

The honest version of this chapter is that all six items exist because the current system has gaps. Radar training data was scarce. The fusion thresholds were set empirically on one dataset and tested on one frame of it. The dashboard was built for analysis, not for a bridge watch-keeper operating under stress. None of this makes the system a failure; it makes it a prototype in the accurate sense of the word: something that demonstrates the concept works and points at what needs to be fixed next. That is a reasonable place to end up after a single academic project, and a reasonable place to start from if someone picks this up and continues it.

6.3 Closing Remarks

Maritime collision avoidance is one of those problems where the consequences are significant enough that any gains in perception reliability are immediately worthwhile. A system that can detect a buoy at 200 metres rather than 150, or can see through foggy conditions where camera-only systems fail, can save lives, prevent environmental damage, and make maritime operations safer. The system developed here is a software prototype, not a deployable safety system, and it has several shortcomings that would need to be remedied before operational use. Nonetheless, it creates a principled, modular multi-modal fusion design built upon publicly available datasets and well-understood machine learning building blocks that significantly extends the perceptual range and reliability over any individual sensor alone. That is a result that is worth continuing to build on.

The research community working on maritime perception has, until recently, been substantially smaller and less well-resourced than its counterpart in autonomous driving. The availability of the Pohang Canal dataset [18] and comparable open-water datasets represents a turning point: it is now possible to train, benchmark, and compare multi-modal maritime perception systems on shared ground-truth data, accelerating progress in a way that was not possible when each research group had to collect its own proprietary data. This project has demonstrated one approach to exploiting that shared data; the modular, open codebase it produces is intended as a contribution to that collective effort.

References

- [1] United Nations Conference on Trade and Development (UNCTAD). Review of maritime transport 2023. Technical report, United Nations, Geneva, 2023.
- [2] K. Talpur, R. Hasan, I. Gocer, S. Ahmad, and Z. Bhuiyan. AI in maritime security: Applications, challenges, future directions, and key data sources. *Information*, 16(8): 658, 2025.
- [3] A. Stove, A. Willsky, and D. Crisp. Challenges in radar sea clutter modelling. *IET Radar, Sonar & Navigation*, 16(9):1011–1026, 2022.
- [4] European Maritime Safety Agency (EMSA). Annual overview of marine casualties and incidents 2023. Technical report, European Maritime Safety Agency, Lisbon, 2023.
- [5] Allianz Global Corporate & Specialty. Safety and shipping review 2023. Technical report, Allianz Global Corporate & Specialty SE, Munich, 2023.
- [6] G. Rutkowski. Practical aspects of using radar in maritime navigation with analysis of its potential errors and limitations. *Scientific Journal of Gdynia Maritime University*, 113:17–28, 2020.
- [7] International Maritime Organization (IMO). AIS Transponders and Maritime Safety. Technical report, IMO, London, 2022.
- [8] R. Cao and J. Sui. A dynamic multi-scale feature fusion network for enhanced SAR ship detection. *Sensors*, 25(16):1–18, 2025.

- [9] L. Žust and M. Kristan. Learning with weak annotations for robust maritime obstacle detection. *arXiv preprint arXiv:2206.13263*, 2022.
- [10] D. Yang, M. I. Solihin, Y. Zhao, et al. A review of intelligent ship marine object detection based on RGB camera. *IET Image Processing*, 18, 2023.
- [11] L. Xie, Y. Jiang, S. Zhang, et al. Reliable LiDAR-based ship detection and tracking for maritime autonomous surface ships. *Ocean Engineering*, 302:116942, 2024.
- [12] O. Ogunrinde, S. Sun, J. Guo, et al. Deep camera–radar fusion with an attention framework for robust object detection in adverse weather. *Sensors*, 23(14):6255, 2023.
- [13] P. Zhao, X. Yu, Z. Chen, and Y. Liang. A real-time ship detector via a common camera. *Journal of Marine Science and Engineering*, 10(8), 2022.
- [14] J. Yang, J. Li, and X. Huang. Analysis of small sea-surface target detection performance in sea clutter environments. *Remote Sensing*, 14(10):2334, 2022.
- [15] C. E. Balas and C. Guedes Soares. Sensor integration for maritime situational awareness. *Ocean Engineering*, 121:288–302, 2016.
- [16] K. J. Rawson and E. C. Tupper. *Basic Ship Theory, Volume II: Hydrostatics and Resistance, Maneuvering, Seakeeping*. Butterworth-Heinemann, Oxford, 2001.
- [17] T. I. Fossen. *Handbook of Marine Craft Hydrodynamics and Motion Control*. Wiley, Chichester, U.K., 2011.
- [18] D. Chung, J. Kim, C. Lee, and J. Kim. Pohang Canal Dataset: A Multimodal Maritime Dataset for Autonomous Navigation in Restricted Waters. *International Journal of Robotics Research*, 42(12):1104–1114, 2023.
- [19] Y. Lu, H. Ma, E. Smart, et al. Fusion of camera-based vessel detection and AIS

- for maritime surveillance. In *Proceedings of the 26th International Conference on Automation & Computing (ICAC 2021)*, 2021.
- [20] D. J. Yeong, G. Velasco-Hernandez, J. Barry, and J. Walsh. Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors*, 21(6), 2021.
- [21] D. Feng, C. Haase-Schütz, L. Rosenbaum, M. Bethge, and K. Dietmayer. Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. *IEEE Transactions on Intelligent Transportation Systems*, 22(3):1341–1360, 2021.
- [22] X. Bai, Z. Hu, X. Zhu, Q. Huang, Y. Chen, H. Fu, and C.-L. Tai. TransFusion: Robust LiDAR-camera fusion for 3D object detection with transformers. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1090–1099, 2022.
- [23] R. Zhang, S. Li, G. Ji, et al. Survey on deep learning-based marine object detection. *Journal of Advanced Transportation*, 2021.
- [24] L. Su, Y. Chen, H. Song, and W. Li. A survey of maritime vision datasets. *Multimedia Tools and Applications*, 82(19):28873–28893, 2023.
- [25] International Maritime Organization (IMO). Outcome of the Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS). *IMO MSC.1/Circ.1638*, 2021.
- [26] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye. Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3):257–332, 2023.
- [27] A. Felski and K. Zwolak. The ocean-going autonomous ship – challenges and threats. *Journal of Marine Science and Engineering*, 8(1):41, 2020.

- [28] X. He, X. Chen, X. Du, et al. Maritime target radar detection and tracking via DTNet transfer learning using multi-frame images. *Remote Sensing*, 17(5), 2025.
- [29] K. Talpur, R. Hasan, I. Gocer, S. Ahmad, and Z. Bhuiyan. AI in Maritime Security: Applications, Challenges, Future Directions, and Key Data Sources. *Information*, 16(8):658, 2025.
- [30] International Maritime Organization (IMO). Convention on the International Regulations for Preventing Collisions at Sea (COLREGs). IMO Convention, 1972. Amended 2007.
- [31] K. Woerner, M. R. Benjamin, M. Novitzky, and J. J. Leonard. Quantifying protocol evaluation for autonomous collision avoidance. *Autonomous Robots*, 44:985–1003, 2020.
- [32] L. P. Perera, P. Oliveira, and C. Guedes Soares. Maritime traffic monitoring based on vessel detection, tracking, state estimation, and trajectory prediction. *IEEE Transactions on Intelligent Transportation Systems*, 13(3):1188–1200, 2012.
- [33] V. De Silva, J. Roche, and A. Kondo. Robust fusion of LiDAR and wide-angle camera data for autonomous mobile robots. *Sensors*, 18(8), 2018.
- [34] E. Palladin, R. Dietze, P. Narayanan, M. Bijelic, and F. Heide. SAMFusion: Sensor-adaptive multimodal fusion for 3D object detection in adverse weather. In *Proceedings of the 18th European Conference on Computer Vision (ECCV 2024)*, 2024.
- [35] M. Kheirandish, E. Azadi Yazdi, H. Mohammadi, and M. Mohammadi. A fault-tolerant sensor fusion in mobile robots using multiple model Kalman filters. *Robotics and Autonomous Systems*, 161:104343, 2023.
- [36] J.-J. Lin, J.-I. Guo, V. M. Shivanna, and S.-Y. Chang. Deep learning derived object de-

- tection and tracking technology based on sensor fusion of millimeter-wave radar/video and its application on embedded systems. *Sensors*, 23(5):2746, 2023.
- [37] Navtech Radar. Situational awareness on marine vessels. <https://navtechradar.com/innovation-lab/situational-awareness-on-marine-vessels/>, 2025. Accessed: October 2025.
- [38] Z. Li, W. Wang, H. Li, E. Xie, C. Sima, T. Lu, Y. Qiao, and J. Dai. BEVFormer: Learning bird’s-eye-view representation from multi-camera images via spatiotemporal transformers. *European Conference on Computer Vision (ECCV)*, 2022.
- [39] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu. Attention mechanisms in computer vision: A survey. *Computational Visual Media*, 8:331–368, 2022.
- [40] D. Liu, J. Zhang, J. Jin, Y. Dai, and L. Li. A new approach of obstacle fusion detection for unmanned surface vehicle using Dempster-Shafer evidence theory. *Applied Ocean Research*, 119, 2022.
- [41] Y. LeCun, B. Boser, J. S. Denker, et al. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4), 1989.
- [42] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [43] A. H. Lang, S. Vora, H. Caesar, et al. PointPillars: Fast encoders for object detection from point clouds. *arXiv preprint*, 2018.
- [44] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified,

- real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016.
- [45] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems (NeurIPS 2015)*, 2015.
 - [46] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
 - [47] K. Patel, C. Bhatt, and P. L. Mazzeo. Deep learning-based automatic detection of ships: An experimental study using satellite images. *Journal of Imaging*, 8(7):182, 2022.
 - [48] S. Rahman, A. B. Vattulainen, and D. A. Robertson. Machine learning-based approach for maritime target classification and anomaly detection using millimetre wave radar Doppler signatures. *IET Radar, Sonar & Navigation*, 18(2):344–360, 2023.
 - [49] Y. Ren, X. Wang, and J. Yang. Maritime ship recognition based on convolutional neural network and linear weighted decision fusion for multimodal images. *Mathematical Biosciences and Engineering*, 20(10):18545–18565, 2023.
 - [50] D. Dagdilelis, P. Grigoriadis, and R. Galeazzi. Multimodal and multiview deep fusion for autonomous marine navigation. *arXiv preprint arXiv:2505.01615*, 2025.
 - [51] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, HI, 2001.
 - [52] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In

- Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, 2005.
- [53] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
 - [54] M. Baldomero-Naranjo et al. A robust SVM-based approach with feature selection and outliers detection for classification problems. *arXiv*, 2024.
 - [55] X. Chen et al. Optimal weighted random forests. 2024.
 - [56] H. Saadatfar, S. Khosravi, J. Joloudari, A. Mosavi, and S. Shamshirband. A new K-nearest neighbors classifier for big data based on efficient data pruning. *Mathematics (MDPI)*, 8:286, 2024.
 - [57] P. Viola and M. J. Jones. Robust real-time face detection. In *International Journal of Computer Vision*, volume 57, pages 137–154, 2004.
 - [58] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*, 2004.
 - [59] R. Brunelli. Template matching techniques in computer vision: Theory and practice. 2009.
 - [60] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. In *arXiv preprint arXiv:1706.05587*, 2017.
 - [61] L. Yang, B. Kang, Z. Huang, et al. Depth anything V2. 2024.
 - [62] Z. Wang, L. Zheng, Y. Liu, et al. Monocular depth estimation: A survey. *arXiv preprint arXiv:2304.11677*, 2023.

- [63] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021.
- [64] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko. End-to-end object detection with transformers. In *European Conference on Computer Vision (ECCV)*, 2020.
- [65] L. Žust, J. Perš, and M. Kristan. LaRS: A diverse panoptic maritime obstacle detection dataset and benchmark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.
- [66] N. Wang, Y. Wang, Y. Wei, B. Han, and Y. Feng. Marine vessel detection dataset and benchmark for unmanned surface vehicles (MVDD13). *Applied Ocean Research*, 142: 103835, 2024.
- [67] A. Nanda, S. W. Cho, H. Lee, and J. H. Park. KOLOMVERSE: Korea open large-scale image dataset for object detection in the maritime universe. *IEEE Transactions on Intelligent Transportation Systems*, 25(12), 2024.
- [68] H. Jang, W. Yang, H. Kim, et al. MOANA: Multi-Radar Dataset for Maritime Odometry and Autonomous Navigation Application. *arXiv preprint arXiv:2412.03887*, 2024.
- [69] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. In *IEEE Transactions on Intelligent Transportation Systems*, volume 18, pages 1993–2016, 2017.
- [70] S. Singha, B. Belton, and R. Raghavendra. Maritime object detection using deep learning on the singapore maritime dataset. *IEEE Access*, 11:53243–53258, 2023.

- [71] K. Shi, S. He, Z. Shi, et al. Radar and camera fusion for object detection and tracking: A comprehensive survey. *arXiv preprint arXiv:2410.19872*, 2024.
- [72] Y. Zhang, J. Chen, and W. Gao. Ship detection with deep learning: a survey. *Artificial Intelligence Review*, 56(10), 2023.
- [73] Y. Li, Y. Zhenghe, and F. Wu. Multi-sensor fusion for unmanned vessel obstacle avoidance system: Intelligent navigation for waters around nuclear power plants. In *Association for Computing Machinery*, New York, NY, USA, 2025.
- [74] G. Xie, J. Zhang, J. Tang, et al. Obstacle detection based on depth fusion of lidar and radar in challenging conditions. *Industrial Robot*, 48(6):792–802, 2021.
- [75] F. Farahnakian, L. Luotsinen, and J. Heikkonen. Deep learning based multi-modal fusion architectures for maritime surveillance. *Remote Sensing*, 16(2):345, 2024.
- [76] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 226–231, 1996.
- [77] S. Mittal. A survey on optimized implementation of deep learning models on the NVIDIA Jetson platform. In *Journal of Systems Architecture*, volume 97, pages 428–442, 2019.

List of Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
6DoF	Six Degrees of Freedom
AI	Artificial Intelligence
AIS	Automatic Identification System
ANN	Artificial Neural Network
ASPP	Atrous Spatial Pyramid Pooling
BCE	Bachelor of Computer Engineering
CNN	Convolutional Neural Network
CUDA	Compute Unified Device Architecture
DNN	Deep Neural Network
DOF	Degrees of Freedom
FMCW	Frequency-Modulated Continuous Wave
FOV	Field of View
FPS	Frames Per Second
FYP	Final Year Project
GHz	Gigahertz
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GPU	Graphics Processing Unit
GUI	Graphic User Interface
HOG	Histogram of Oriented Gradients
IMU	Inertial Measurement Unit

IoU	Intersection over Union
IR	Infrared
JSON	JavaScript Object Notation
KITTI	Karlsruhe Institute of Technology and Toyota Technological Institute
KNN	K-Nearest Neighbours
LiDAR	Light Detection And Ranging
mAP	Mean Average Precision
ML	Machine Learning
MoANA	Multi-Radar Dataset for Maritime Odometry and Autonomous Navigation Application
PNG	Portable Network Graphics
RADAR	Radio Detection And Ranging
R-CNN	Region-based Convolutional Neural Network
ResNet	Residual Network
RoI	Region of Interest
RPN	Region Proposal Network
RTX	Ray Tracing Accelerated (NVIDIA GPU Series)
SAR	Synthetic Aperture Radar
SIFT	Scale-Invariant Feature Transform
SURF	Speeded-Up Robust Features
SVM	Support Vector Machine
TTS	Text-to-Speech
USV	Unmanned Surface Vehicle
YOLO	You Only Look Once
XYZ	Three-Dimensional Coordinate Axes

ANNEXURES

Annexure A

System Configuration and Software Dependencies

Hardware Specifications

The system was developed and evaluated on a workstation with the following hardware configuration:

- **CPU:** Intel Core i5 (12th generation)
- **GPU:** NVIDIA GeForce RTX 3060 (12 GB VRAM)
- **RAM:** 16 GB DDR4
- **Storage:** 1 TB NVMe SSD
- **OS:** Windows 11

Software Dependencies

- Python 3.10
- PyTorch 2.1.0 with CUDA 12.1
- Ultralytics YOLOv8 (ultralytics 8.0.x)
- Depth Anything V2 (ViT-S checkpoint)
- DeepLabv3 (torchvision)

- Open3D 0.17
- OpenCV 4.8
- Matplotlib 3.7
- NumPy 1.24
- SciPy 1.11
- scikit-learn 1.3

Annexure B

Dataset Directory Structure

Pohang Canal Dataset Structure

The Pohang Canal dataset sequences follow a standardised directory layout:

```
sequence_XXXXXX/  
  calibration/  
    intrinsics.json  
    extrinsics.json  
  camera/  
    left/      (PNG frames)  
    right/     (PNG frames)  
  infrared/   (16-bit PNG thermal frames)  
  lidar/  
    front/    (.bin point clouds)  
    port/     (.bin point clouds)  
    starboard/ (.bin point clouds)  
  radar/      (PNG sweep images)  
  timestamps/  
    camera_left.txt  
    camera_right.txt  
    infrared.txt  
    lidar_front.txt
```

```
radar.txt
```

Fusion System Output Format

Each detection in the final fused output list is a Python dictionary with the following fields:

```
{  
    "label":      str,    # e.g. "ship", "buoy"  
    "source":     str,    # e.g. "lidar_cam", "radar_only"  
    "range_m":    float,  # metres from vessel origin  
    "bearing_deg": float, # degrees from vessel bow (0=forward)  
    "confidence": float   # 0.0 to 1.0  
}
```

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.



4% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography



Caution

Match Groups

- 91 Not Cited or Quoted 4%**
Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%**
Matches that are still very similar to source material
- 2 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 2% Internet sources
- 3% Publications
- 2% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

User User

Thesis (1)

 Yaseen paper 2 AIRCEMS

Document Details

Submission ID**trn:oid:::3618:144895698****Submission Date****Jul 4, 2026, 6:48 PM GMT+5****Download Date****Jul 4, 2026, 7:02 PM GMT+5****File Name****Thesis (1).pdf****File Size****4.6 MB****79 Pages****20,701 Words****109,756 Characters**

User User

Thesis (1)

 Yaseen paper 2 AIRCEMS

Document Details

Submission ID

trn:oid:::3618:144895698

Submission Date

Jul 4, 2026, 6:48 PM GMT+5

Download Date

Jul 4, 2026, 6:59 PM GMT+5

File Name

Thesis (1).pdf

File Size

4.6 MB

79 Pages

20,701 Words

109,756 Characters

Match Groups

- 91 Not Cited or Quoted 4%**
Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%**
Matches that are still very similar to source material
- 2 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 2% Internet sources
- 3% Publications
- 2% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

- 1 **Internet**
itegam-jetia.org <1%
- 2 **Internet**
www.mdpi.com <1%
- 3 **Publication**
A. Punitha, S. Syedakbar, S. Jeyasudha. "Intelligent and Sustainable Systems: AI, G... <1%
- 4 **Internet**
www.preprints.org <1%
- 5 **Publication**
Jaskaran Singh, Meenu Gupta, Rakesh Kumar. "Smart Technologies and Intelligen... <1%
- 6 **Student papers**
British University in Egypt on 2026-06-12 <1%
- 7 **Publication**
Manoj Kumar, Tanweer Ali, Jaume Anguera, Suman Lata Tripathi. "Emerging Tech... <1%
- 8 **Internet**
ejournal.gomit.id <1%
- 9 **Internet**
bura.brunel.ac.uk <1%
- 10 **Student papers**
University of Surrey on 2025-11-11 <1%

11	Student papers	University of East London on 2025-09-08	<1%
12	Internet	www.isteonline.in	<1%
13	Student papers	Institute of Technology Blanchardstown on 2012-12-12	<1%
14	Internet	d1ssu070pg2v9i.cloudfront.net	<1%
15	Publication	"Computer Vision – ECCV 2016", Springer Science and Business Media LLC, 2016	<1%
16	Publication	Bhabani Prasanna Pattanaik, Dharendra Kumar Shukla, Sambit Satpathy, Kamala...	<1%
17	Student papers	Queen Mary and Westfield College on 2026-06-12	<1%
18	Student papers	Gisma University of Applied Sciences GmbH on 2026-06-30	<1%
19	Student papers	Indian Institute of Information Technology Vadodara on 2026-04-14	<1%
20	Publication	Svebor Karaman. "Multi-layer Local Graph Words for Object Recognition", Lecture...	<1%
21	Student papers	University College London on 2026-04-15	<1%
22	Publication	Alois C. Knoll, Mingyu Liu, Xingcheng Zhou, Christian Creß, Walter Zimmer. "Real-...	<1%
23	Publication	Jihun Park, Hyunsung Yoo, Yooseung Wang. "Drivable Dirt Road Region Identifica...	<1%
24	Internet	analisedemarcha.com	<1%

25	Internet	cambridgepixel.com	<1%
26	Internet	www.jucs.org	<1%
27	Student papers	George Bush High School on 2023-07-07	<1%
28	Student papers	Queen Mary and Westfield College on 2019-08-22	<1%
29	Publication	R. Glenn Wright. "Unmanned and Autonomous Ships - An Overview of MASS", Rou...	<1%
30	Publication	D. Lakshmi, Ravi Shekhar Tiwari, Rajesh Kumar Dhanaraj, Seifedine Kadry. "Explai...	<1%
31	Publication	Pawel Pratyush, Dukka B. KC. "Chapter 8 Advances in Prediction of Posttranslatio...	<1%
32	Internet	arxiv.org	<1%
33	Internet	tailieu.vn	<1%
34	Publication	Krishnan, Aditya. "Computationally Efficient, Privacy-Preserving, and Resource-Co...	<1%
35	Internet	iieta.org	<1%
36	Student papers	Associatie K.U.Leuven on 2024-08-24	<1%
37	Student papers	Babes-Bolyai University on 2025-06-18	<1%
38	Student papers	International Islamic University Malaysia on 2021-01-18	<1%

39	Student papers	Saxion Brightspace on 2024-12-19	<1%
40	Publication	Sol Han, Dongwook Lee, Jonghwi Kim, Jinwhan Kim. "A survey on maritime perce...	<1%
41	Student papers	University of East London on 2025-09-09	<1%
42	Internet	archives.christuniversity.in	<1%
43	Internet	maritimesafetyinnovationlab.org	<1%
44	Internet	www.allmultidisciplinaryjournal.com	<1%
45	Internet	www.ijert.org	<1%
46	Publication	"Data Mining and Information Security", Springer Science and Business Media LL...	<1%
47	Student papers	A. P. Shah Institute of Technology, Thane on 2026-07-04	<1%
48	Student papers	Asia Pacific University College of Technology and Innovation (UCTI) on 2023-05-21	<1%
49	Publication	Muyi Sun, Guan hong Zhang, Hao Dang, Xingqun Qi, Xiaoguang Zhou, Qing Chang...	<1%
50	Student papers	National Taipei University of Technology on 2019-02-20	<1%
51	Publication	Noam Cohen, Yael Levi, Amit Ben-David, Shira Katz, Elinor Avraham, Lior Shapiro. "...	<1%
52	Student papers	Queen's University of Belfast on 2020-04-22	<1%

53	Publication	Shalli Rani, Ashu Taneja. "WSN and IoT - An Integrated Approach for Smart Applic...	<1%
54	Student papers	The Robert Gordon University on 2024-01-18	<1%
55	Student papers	Universiti Teknologi MARA on 2020-07-06	<1%
56	Student papers	Vrije Universiteit Brussel on 2026-06-08	<1%
57	Publication	Yang Liu, Steven D. Brown. "Wavelet multiscale regression from the perspective o...	<1%
58	Internet	dspace.nitrkl.ac.in	<1%
59	Internet	li02.tci-thaijo.org	<1%
60	Internet	repository.uantwerpen.be	<1%
61	Internet	www.electronicsonboard.co.uk	<1%
62	Internet	www.marketresearchintellect.com	<1%
63	Publication	"Intelligent Systems and Applications", Springer Science and Business Media LLC,...	<1%
64	Student papers	Higher Education Commission Pakistan on 2021-02-12	<1%
65	Publication	Li, Feng. "From Objects to Prompts: Towards Generalized Multimodal Foundation ...	<1%
66	Student papers	Asia Pacific University College of Technology and Innovation (UCTI) on 2019-07-23	<1%

67	Student papers	Central Queensland University on 2026-06-08	<1%
68	Student papers	Florida A&M University on 2023-03-09	<1%
69	Publication	Taiwo Samuel Aina, Wisdom Okafor, Grace Okafor, Babatunde Iyaomolere. "Mach...	<1%
70	Student papers	Vietnam Maritime University on 2026-05-25	<1%