

# Traversability Mapping in Post-Flood Environment

Thesis approved by  
the Department of Computer Science  
University of Kaiserslautern-Landau  
for the award of the Doctoral Degree  
Doctor of Engineering (Dr.-Ing.)

to

**Hannan Ejaz Keen**

**Date of the defense** : 01.07.2024  
**Dean** : Prof. Dr. Christoph Garth  
**Reviewer** : Prof. Dr. Karsten Berns  
**Reviewer** : Prof. Dr. Muhammad Moazam Fraz



# Acknowledgment

First and foremost, I express my gratitude to Almighty Allah for providing me with the strength, patience, and opportunity to successfully complete my PhD. I am thankful for His guidance throughout this journey.

I would like to extend my sincere appreciation to Professor Karsten Berns for his unwavering support, invaluable advice, and expert supervision that helped shape my research and personal development. Thanks to Frau Rita Broschart for her administrative efforts. I also want to thank my wife Sundus and my kid Muhammad Mustafa, for their constant support and sacrifices that gave me the strength and peace of mind to navigate the challenges of this PhD journey. Their love and understanding have been my sanctuary. I'd also like to thank my family in Pakistan, especially my parents, whose endless prayers and expectations for this day have been a source of immense motivation, inspiring me to persevere through all hurdles. I am grateful to my colleagues, especially Dr. Patrick Fleischmann, Dr. Thorsten Ropertz, Dr. Patrick Wolf, Dr. Alexander Köpper, Axel Vierling, Hamza Jan, and Sascha Steffens, whose collaboration, insights, and encouragement have been essential in overcoming the various challenges and achieving this milestone. Finally, I want to thank our project partners and sponsors whose financial support played a crucial role in the completion of this project. Their belief in our work and their contributions have been fundamental to our success.

This thesis stands on the collective effort and support of everyone mentioned, and many others who have contributed in various capacities. I am deeply thankful to everyone who has been a part of this journey.

Hannan Ejaz Keen

Kaiserslautern, March 2024



# Abstract

The increased frequency of floods due to global warming has posed formidable challenges for rescue operations worldwide. Standard mapping tools, including Google Maps and OpenStreetMap, are infeasible for post-flood scenarios due to the destruction of known structures and road networks. While response teams often use satellite and drone imagery to aid flood relief efforts, they typically lack the ability to detect underwater obstacles, making it unsafe for rescue boats to navigate. Furthermore, the complex nature of post-flood environments, such as highly variable depth, random unstructured obstacles, and extremely turbid water due to sand particles, increases the need for robust environment perception and mapping systems tailored to the complexities of post-flood environments. Additionally, a significant challenge in developing such a system arises from the unavailability of comprehensive datasets about flooded environments, a limitation that has constrained previous research efforts.

The primary objective of this thesis is to provide a robust surface and underwater perception system that uses rich multi-modal sensory knowledge and provides traversability information for safe navigation. This perception system involves comprehensive understanding of obstacle's characteristics and categorizing their threat levels by integrating surface and underwater sensory data. Additionally, this research attempts to devise a versatile system capable of seamless reconfiguration across various surface water vehicles. This adaptability benefits rescue teams equipped with boats featuring diverse kinematics and motion models.

Consequently, this thesis proposed a novel Shallow Water traversability Mapping (SWiM) architecture, which integrates multiple sensory modalities to create a lightweight 2.5-dimensional traversability map covering both surface and underwater modalities. Enhancements in obstacle detection within the underwater environment, utilizing low Signal-to-Noise Ratio (SNR) sonar imagery, are achieved through dedicated image enhancement and depth estimation modules. Concurrently, accurate object distinction from water is facilitated by deploying various deep-learning-based object detection and segmentation techniques on camera images. The fusion of camera and LiDAR data through inverse-perspective mapping enhances the certainty of obstacle detection. By combining obstacle maps from both modalities, the system can compute essential features of obstacles, including their threat level and whether they are floating or sinking. System validation encompasses a diverse data bank comprising data from state-of-the-art datasets and novel multi-modal MASTER dataset, captured on three distinct boats in various water bodies. Additionally, a comprehensive post-flood simulation is presented using generative adversarial networks (GANs) to replicate realistic sensory noise models, enabling rigorous testing in complex scenarios.



# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>  | <b>1</b>  |
| 1.1      | Challenges in Post-Flood Environment . . . . .   | 2         |
| 1.2      | Objective and Contribution . . . . .   | 4         |
| 1.3      | Outline . . . . .  | 5         |
| <b>2</b> | <b>Surface Water Vehicles in Post-Flood Scenarios: A Review</b>                            | <b>7</b>  |
| 2.1      | Traversability Mapping in Aquatic Environment . . . . .                                    | 8         |
| 2.2      | State-of-the-art Surface Water Vehicles . . . . .  | 9         |
| 2.2.1    | Vehicles for Bathymetry . . . . .  | 10        |
| 2.2.2    | Research Vehicles . . . . .  | 12        |
| 2.2.3    | Urban Waterways Vehicles . . . . .   | 14        |
| 2.2.4    | Environmental Missions . . . . .   | 17        |
| 2.3      | Comparative Analysis . . . . .   | 20        |
| 2.4      | Research Gap . . . . .   | 23        |
| <b>3</b> | <b>Conceptual Overview of the Surface Water Traversability Mapping (SWiM) Architecture</b> | <b>25</b> |
| 3.1      | SWiM Architecture . . . . .  | 26        |
| 3.1.1    | Underwater Traversability . . . . .  | 28        |
| 3.1.2    | Surface Water Traversability . . . . .   | 30        |
| 3.1.3    | Fused Mapping . . . . .  | 31        |
| <b>4</b> | <b>Datasets and Test Environments</b>  | <b>33</b> |
| 4.1      | Natural Waterbodies . . . . .  | 34        |
| 4.1.1    | State-of-the-art Datasets . . . . .  | 34        |
| 4.1.2    | MASTER Dataset: . . . . .  | 40        |
| 4.2      | Simulation . . . . .   | 43        |
| 4.2.1    | Post-flood Water . . . . .   | 43        |
| 4.2.2    | Flooded Environments . . . . .   | 45        |
| 4.2.3    | Surface Vehicles Models . . . . .  | 47        |
| 4.2.4    | Sensor System . . . . .  | 48        |
| 4.2.5    | Simulated Sensory Data . . . . .   | 69        |
| 4.3      | Discussion . . . . .   | 73        |

---

|  |            |
|--|------------|
| <b>5 Underwater Traversability</b>                             | <b>75</b>  |
| 5.1 Challenges in Underwater Sensing . . . . .                 | 76         |
| 5.1.1 Comparative Analysis of Commercial Sonars . . . . .      | 76         |
| 5.1.2 Issues in Sonar Imagery . . . . .                        | 79         |
| 5.2 Underwater Traversability Mapping Framework . . . . .      | 82         |
| 5.2.1 Image Filtration . . . . .                               | 84         |
| 5.2.1.1 Image Enhancement . . . . .                            | 84         |
| 5.2.1.2 Depth Estimation . . . . .                             | 91         |
| 5.2.2 Obstacle Filtration . . . . .                            | 93         |
| 5.2.2.1 Obstacle Detection . . . . .                           | 93         |
| 5.2.2.2 Obstacle Refinement . . . . .                          | 97         |
| 5.2.3 Depth and Obstacle Maps . . . . .                        | 100        |
| 5.3 Discussion . . . . .                                       | 103        |
| <b>6 Surface Water Traversability</b>                          | <b>105</b> |
| 6.1 Challenges in Surface Water Sensing . . . . .              | 106        |
| 6.1.1 Comparative Analysis between Cameras and LiDAR . . . . . | 109        |
| 6.2 Surface Water Traversability Mapping Framework . . . . .   | 110        |
| 6.2.1 Image Processing . . . . .                               | 112        |
| 6.2.2 Point cloud Processing . . . . .                         | 122        |
| 6.2.3 Obstacle Map . . . . .                                   | 129        |
| 6.3 Discussion . . . . .                                       | 132        |
| <b>7 Fused Traversability</b>                                  | <b>135</b> |
| 7.1 Fused Traversability Mapping Framework . . . . .           | 136        |
| 7.1.1 Buoyancy Estimator . . . . .                             | 138        |
| 7.1.2 Fused Local Threat Map . . . . .                         | 139        |
| 7.1.3 Traversability Map . . . . .                             | 140        |
| 7.2 Discussion . . . . .                                       | 142        |
| <b>8 Conclusion</b>  | <b>143</b> |
| 8.1 Key Findings . . . . .                                     | 144        |
| 8.2 Future Work and Outlook . . . . .                          | 145        |
| <b>A System and Specification</b>                              | <b>147</b> |
| <b>B Surface Water Vehicles</b>                                | <b>157</b> |
| <b>C Framework for Intelligent Robot Control</b>               | <b>161</b> |
| <b>D Unreal Engine and Flood Simulation</b>                    | <b>167</b> |
| <b>E Deep Learning Architectures</b>                           | <b>173</b> |
| <b>Bibliography</b>  | <b>187</b> |



# 1. Introduction

Global warming significantly contributes to the rise of natural disasters such as floods, hurricanes, and droughts worldwide. This phenomenon has far-reaching consequences for the earth's climate system, including changes in precipitation patterns and increased frequency of extreme weather events. According to Guha et al. [Guha-Sapir 04], the number of natural disasters tripled compared to thirty years ago. In addition, the Intergovernmental Panel on Climate Change (IPCC) report [Masson-Delmotte 18] has warned that the earth's temperature has risen by nearly 1.5 degrees compared to pre-industrialized levels. The rising global temperature increases atmospheric moisture, alters precipitation patterns, and melts more glaciers, causing increased rainfall. This, in turn, leads to higher water levels in rivers and oceans, increasing the risk and severity of floods. For example, in 2017, Hurricane Harvey caused severe flooding in Houston, Texas, displacing over 30,000 people and causing economic losses of over \$125 billion. The storm was fueled by warmer ocean temperatures due to global warming, which led to increased rainfall and worsened flooding. Furthermore, in 2021, total losses of 280 billion USD <sup>1</sup> occurred due to natural calamities around the globe. Among those calamities, a significant portion (90 billion USD <sup>2</sup>) occurred due to floods. In addition, several studies by the Intergovernmental Panel on Climate Change (IPCC) indicate a likely rise in extreme flash floods and river floods. Consequently, radical measures and funding are required to deal with the flood estimation systems and post-flood rescue operations.

The impact of flooding is undeniable and leads to significant loss of life and property damage, leaving many people stranded without access to help or resources. In flood relief operations, assistance is delivered via helicopters or boats due to the loss of the road network. Since aerial assistance is often limited in scale, boats are the major contributor to such operations. These boats rely on accurate and up-to-date maps that provide critical information on the affected areas, including topography, infrastructure, and evacuation routes. However, floods often result in the destruction or loss of maps due to water damage

---

<sup>1</sup><https://www.munichre.com/en/risks/natural-disasters-losses-are-trending-upwards.html>, last accessed: (09.01.2023)

<sup>2</sup><https://www.munichre.com/en/risks/natural-disasters-losses-are-trending-upwards/floods-and-flash-floods-underestimated-natural-hazards.html>, last accessed: (09.01.2023)

or structural collapse. This poses a significant challenge for rescue boat by hindering their ability to navigate and plan operations efficiently. Without reliable maps, rescue operations may be delayed, and the safety of both rescuers and affected populations can be compromised.

Standard maps, such as Google Maps and OpenStreetMap, while valuable for everyday navigation, are often inadequate for rescue operations in post-flooded environments. Several factors contribute to their limited feasibility in such scenarios; primarily, standard maps rely on pre-existing data and may not account for rapid and dynamic changes caused by the flood, such as altered topography, floodwater depth, submerged infrastructure, or random debris. Consequently, these maps may provide inaccurate or outdated information, leading to potential navigational challenges for rescue boats that impede rescue boats' ability to identify safe and accessible routes. Moreover, standard maps heavily rely on satellite imagery and aerial surveys, which may be limited or inaccessible during and immediately after a flood due to cloud cover, poor visibility, or disrupted infrastructure. This lack of updated imagery can hinder the ability to assess the current conditions and plan rescue operations effectively. Therefore, there is a dire need for a novel post-flood traversability mapping system that utilizes new sensory modalities rather than satellite imagery to understand the surface and underwater surroundings; efficiently map the significantly changed environment.

## 1.1 Challenges in Post-Flood Environment

A flooded environment faces drastic changes due to fallen trees, opaque water, floating cattle, and destructed houses, as depicted in Figure 1.1. These changes destroy the road network to the affected areas and produce several navigational problems for rescue operations. Additionally, these extreme changes limit the use of standard navigational maps for such an environment. Consequently, post-flood rescue operations suffer from several challenges due to a limited understanding of the destruction level. Hence, an operation with manual boats and aerial assistance is traditionally performed to gather updated environmental information, which usually takes several days to cover the whole area and faces several inherent challenges, as discussed below.

- *Unstructured Environment*: The flood introduces several random objects, such as fallen trees, debris, or destroyed infrastructure, creating significant problems in rescue operations. These objects could have been defined structures in the past, such as bridges or other buildings, but now are demolished by flood and cannot be detected with state-of-the-art object detection algorithms.
- *Water Turbidity*: Apart from solid objects, the flood also carries a significant amount of sand and soil, making the underwater environment entirely opaque for boat drivers. Moreover, the limited visibility in muddy water is challenging for underwater cameras to detect objects. Therefore, active sensors are more practical to use in such a scenario.
- *Multi-Modal Environment*: Due to the different physical properties of air and water, similar sensory modalities cannot be used for object detection over and under the surface of the water environment. For instance, LiDAR is often used to detect surface



**Figure 1.1:** The post-flood images from a devastating flood in Pakistan in 2022. This flood affected more than 70 percent of land and 10 million lives <sup>3</sup>. An estimated loss of 1800 casualties and 15 million USD. The unstructured nature of the destroyed environment is a great challenge for developing a traversability mapping system that can be helpful for flood response forces.

obstacles, but light rays are impenetrable in water; hence LiDAR is not feasible for underwater obstacle detection. Therefore, other sensors are commonly used, which lead to a challenge in the fusion of different sensory modalities.

- *Shallow Waterbed:* The depth of flood water may reach up to 3 meters, and due to the heavy loads over rescue boats, a few centimeters of the boat's body may sink under the water. However, the sudden variations in water depth (less than 30 cm) pose problems for the safe navigation of boats.
- *Slower Response:* Typically, multi-sensor systems gather a massive amount of data. Consequently, such systems suffer slower response times due to high computations. Therefore, besides safe navigation, the adequate response time of the rescue system is also highly critical.
- *Unavailability of Relevant Dataset:* The post-flood environment is usually unique, and collecting datasets during natural calamities is often prohibited due to safety

<sup>3</sup><https://www.unicef.org/emergencies/devastating-floods-pakistan-2022>

concerns. Resultantly, the unavailability of data makes it challenging to validate any proposed system. Hence, it leads to the need for more multi-sensor post-flood datasets in the relevant community.

The challenges aforementioned immensely affect rescue operations in a flooded environment; therefore, many human lives and global economies suffer each year. These are inherent challenges of a destructed environment due to flood; hence, a need arises for a mapping system that can cater to those challenges robustly and reliably.

## 1.2 Objective and Contribution

The primary goal of this thesis is to develop a post-flood traversability mapping architecture that can use rich sensory information about the post-flood destructed environment which is unavailable in standard maps such as Google Maps and Open Street Maps. Such a system is essential for safe navigation and planning rescue operations. The proposed system should be: **robust**; i.e., it should be adaptive to environmental changes, **reconfigurable**; it should be easily re-configurable over several different boats, **reliable**; it should provide an accurate understanding of the environment, **responsive**; it should not only compute but also respond to the parametric changes in real-time.

The contributions in the presented thesis are as follows:

**Arbitrary features modeling** The destructed post-flood environment faces challenges in the semantic understanding of the underwater scenario. Moreover, state-of-the-art machine-learning based object detection algorithms are not feasible due to the unavailability of essential training features in an unstructured surrounding. However, the robust feature modeling of arbitrary objects improves the overall mapping accuracy.

**Object Categorization** In a flooded environment, the objects can be categorized as 1) Floating/sinking (based on underwater footprint) and their 2) Threat Level. This information helps in defining the reliability factor of the traversability map. For instance, if an object is classified as floating, it can be concluded that its position may change in the future due to water flow. Furthermore, the threat level of an obstacle can be an essential property to consider either an obstacle is traversable or not. For instance, floating leaves are safer than a standing tree.

**Multi-modal fusion** Due to significant differences in surface and underwater environments, different sensors help gather distinct feature sets. For example, sensors such as cameras and laser scanners are not feasible for underwater object detection; hence, only sonar is helpful within the water. However, the converse is true for surface water information. Thus, the multi-sensor fusion technique fuses the surface and underwater features, consequently increasing the performance of the object classifier.

**Multi-level segmentation** Understanding the destruction level is fundamental for planning a post-flood rescue operation. Therefore, the multi-level map segmentation, i.e., based on water depth, the object category (floating or submerged), or the vehicle's kinematics, helps to develop an easily understandable interface for rescue boat drivers.

**Simulated and realistic multi-boat validation** The realistic system validation includes rigorous testing over multiple boats [Meckel 22], including novel Pontoon Boat <sup>4</sup>, Basilisk <sup>5</sup>, and Astrider <sup>6</sup> in several water bodies. Each boat has different dimensions and a kinematic model. Moreover, the simulated validation provides similar multi-boat testing over a flooded environment and physics-based sensor plugins that present a time efficient and robust system validation.

## 1.3 Outline

The dissertation addresses the complexities of post-flood environment and presents a novel SWiM (Shallow Water traversability Mapping) architecture that segments the unstructured random objects present in both surface and underwater flooded environments, facilitating the generation of an up-to-date traversability map. Notably, the architecture is adaptable to various boats with distinct kinematic models. The thesis is structured as follows, providing a systematic outline of its components and objectives.

**Chapter 2** provides a comprehensive review of the latest advancements in surface water vehicles specifically designed for marine applications. It evaluates the advantages and disadvantages inherent in these systems, offering a critical analysis of each. Furthermore, this chapter delivers a comparative study of the approaches proposed in the existing literature, highlighting their distinct features and operational capabilities. Moreover, it concludes by identifying and discussing the research gaps evident in the current state of technology. Identifying unaddressed areas and limitations within the existing body of work is the primary motivation driving the research presented in this thesis.

**Chapter 3** introduces the novel concept of Shallow Water Traversability Mapping (SWiM) architecture, a cutting-edge approach designed to navigate through the unique challenges of post-flood environments. This chapter expands on the complexities of the SWiM architecture, delving into the challenges such environments present for traversability mapping. It methodically breaks down and explores the three core modules constituting the SWiM architecture, providing a detailed examination of their functions and interrelationships. Through this comprehensive discussion, the chapter lays a foundational understanding of SWiM's operational framework and its applicability in post-flood scenarios.

**Chapter 4** presents the datasets and test environments integral to developing and evaluating the architecture outlined in this thesis. It introduces the innovative Multi-Modal Shallow Water (MASTER) dataset, a comprehensive collection encompassing multi-sensory and multi-modal data gathered from various water bodies including Hamburg port, Technisches Hilfswerk (THW) Hoya test site and Gelterswoog Lake. This chapter not only details the composition and characteristics of the MASTER dataset but also discusses the development of a novel simulated environment using Unreal Engine. It elaborates on the design process of sonar plugin within simulation, employing Generative Adversarial Networks (GANs) to enhance the simulation's fidelity and effectiveness. This comprehensive discussion underscores the chapter's focus on the critical tools and resources used in the architecture's validation.

---

<sup>4</sup>A big transporter boat specifically build for post-flood environment by consortium of several industrial partners in a *Pontonboot* project

<sup>5</sup>A commercial drone boat

<sup>6</sup>An in-house built raft

**Chapter 5** provides a detailed exploration of the complexities of underwater sensing, explicitly focusing on the comparative analysis of various sonars based on their operating frequencies. This chapter delves into the prevalent issues in sonar imagery, such as speckle noise and low signal-to-noise ratio (SNR), offering a thorough examination of these challenges. Additionally, it presents an in-depth look at the Underwater Traversability Mapping (UTM) framework, elaborating on the outcomes yielded at each stage of this module's operation. The chapter concludes with the computation of the underwater obstacle map, which effectively illustrates the precise location of each obstacle as detected within the sonar imagery.

**Chapter 6** addresses the specific challenges associated with surface water sensing, offering a refined analysis of two key sensing modalities employed over water surfaces: camera and LiDAR. It thoroughly examines the advantages and disadvantages of each modality, providing a balanced perspective. Additionally, this chapter extensively discusses the Surface Water Traversability Mapping (SWTM) framework, detailing its operational framework and methodology. It highlights the results achieved at each stage of the SWTM process, particularly emphasizing the effectiveness of semantic segmentation and object detection achieved through advanced deep learning techniques. The chapter concludes by presenting a comprehensive obstacle map, showcasing the module's capability to map the surface water environment.

**Chapter 7** discusses the complex process of integrating surface and underwater obstacle maps, a crucial aspect in post-flood scenario analysis. It begins by categorizing the various obstacles encountered in post-flood conditions, laying a foundational understanding for the subsequent discussion. The chapter then introduces a novel approach within the Fused Traversability Mapping (FTM) module, discussing each step of this methodology in detail. It thoroughly examines the results and accuracy achieved at every stage of the FTM process. The chapter concludes by presenting fused traversability maps of several environments to provide a holistic view of post-flood obstacles.

**Chapter 8** concludes the thesis by summarizing the contributions, outcomes and their evaluation regarding the thesis objectives. Moreover, it discusses future research and development prospects in an unstructured environment due to natural calamity.

## 2. Surface Water Vehicles in Post-Flood Scenarios: A Review

In the field of robotics, traversability mapping plays a pivotal role, as it assists robots navigate complex and dynamic environments safely and effectively. Through the utilization of advanced sensing modalities, such as LiDAR and cameras [Keen 20] [Zaheer 21], robots have the capability to generate detailed maps that accurately represent terrain characteristics and identify obstacles. These highly detailed and reliable maps offer essential information about the navigability of different regions, allowing robots to navigate their paths with precision and adaptability. Moreover, traversability mapping also helps robots with the ability to recognize potential threats or challenging terrain segments, enabling them to quickly modify their routes and prevent potential accidents. This technique fundamentally enhances the operational efficiency of robots, enabling various applications across diverse industries, including search and rescue missions, agricultural operations, warehouse logistics, construction tasks, and even medical procedures. The remarkable precision with which robots can now navigate complex terrains facilitates tasks such as crop health monitoring [Shafi 20], indoor mapping [Wang 19b], and complicated medical surgeries [Lv 22], emphasizing the potential of traversability mapping. Ultimately, this technique optimizes the performance of autonomous robotic systems and ensures the safety of humans and machines, making it an indispensable tool for advancing the field of automation and robotics.

The significance of precise and updated maps becomes particularly crucial in circumstances where conventional mapping services, such as Google Maps and OpenStreetMap, are unable to provide accurate and reliable information. This limitation is especially evident in the aftermath of natural disasters, including but not limited to hurricanes, earthquakes, and floods, where the environmental landscape suffers from widespread destruction, leading to the increase of unstructured and unpredictable features or objects that are not accurately represented in standard maps. In such environments, the ability of autonomous robots to navigate with adaptability is of utmost importance. Therefore, precise and updated traversability maps are essential in such calamities to ensure that rescue and relief operations can be carried out effectively.

## 2.1 Traversability Mapping in Aquatic Environment

Studying the aquatic environment is extremely important as water makes up around 71% of the Earth <sup>1</sup>. Historically, research on aquatic environments was primarily limited to military and defense applications, focusing on gaining tactical superiority. However, with the advancement of robotics and its proven effectiveness in terrestrial and off-road environments, scientists and researchers working in aquatic domains have become increasingly interested in exploring this field. This development observed the beginning of water robotics in the latter half of the 20<sup>th</sup> century.

Water robotics is significant in addressing aquatic environmental challenges, including offshore energy systems inspection, underwater mine detection, oil spillage detection, understanding marine life, and supporting various maritime activities. The significance of water robotics lies in its ability to operate in complex and unsafe underwater environments where human access is limited. While advancing further, in the last decade, several researches have been done on water robotics for commercial usage. There is a significant increase in water robotic applications [tor 16] [Galceran 12]; consequently, several underwater and surface water vehicles have been recently designed, such as Roboat I/II, Oceanalpha, and Mayflower Autonomous Ship MAS400. These vehicles are designed to undertake various tasks, including underwater mapping, surveillance, and search and rescue operations, signifying a remarkable expansion in the role and capabilities of water robotics in addressing contemporary challenges and advancing scientific and industrial initiatives in aquatic environments.

In water robotics, traversability mapping plays an essential role similar to its terrestrial counterpart, facilitating the efficient and secure navigation of robots in complex aquatic environments, whether on the surface or underwater. It is significant for various aquatic vehicles, including surface and underwater vehicles (SVs/UVs), remotely operated vehicles (ROVs), and even human divers, ensuring they can navigate challenging and dynamic underwater environments. Traversability mapping proves invaluable across various aquatic applications, from offshore oil exploration and infrastructure inspection to the scientific endeavors of marine biology research and environmental monitoring. To achieve this objective, researchers use a range of exploration sensors, such as LiDAR, sonar, cameras, and other sensors that can capture relevant data. The data collected by these sensors is subjected to complex algorithms that analyze and process it, ultimately generating highly accurate maps with detailed information about surface or underwater obstacles encompassing rocks, coral reefs, and other potential threats. These maps enable safe and efficient navigation and exploration of aquatic domains. Beyond these benefits, the information gathered from these sensors can deepen our understanding of aquatic life, their habitats, and the impact of human activities on underwater ecosystems, enriching our understanding of the aquatic world's complexities.

Similarly, accurate and up-to-date maps are crucial for search and rescue teams in a post-flood environment to navigate shallow and murky waters, where these teams face numerous challenges operating in the affected areas. It is essential to have reliable traversability maps even before the rescue operations commence. This is especially critical for rescue teams that use relatively larger boats, that face significant navigational challenges in

---

<sup>1</sup><https://www.usgs.gov/special-topics/water-science-school/science/how-much-water-there-earth>



environments which include shallow water and submerged obstacles. Therefore, having access to accurate maps can play a vital role in ensuring successful rescue operations.

Unmanned surface water vehicles have the potential to play a pivotal role in addressing this challenge. These autonomous vehicles can be deployed to the affected areas, serving as advanced survey units. Equipped with a multi-sensor system, these vehicles are capable of surveying and comprehensively assessing the environment. By integrating data from various sensors, including sonar, LiDAR, and cameras, these vehicles can generate precise and updated traversability maps. These maps not only provide vital information about submerged and surface obstacles but also highlight safe pathways and potential threats.

Essentially, unmanned surface water vehicles act as a first response team, mapping out the conditions and challenges in the post-flood environment. This information is invaluable for the safe and efficient navigation of the larger rescue boats, enabling them to make informed decisions and avoid potential threats. By providing accurate and updated maps, these autonomous vehicles enhance the overall effectiveness of search and rescue operations, ultimately contributing to the safety of both rescue personnel and flood-affected populations.

## 2.2 State-of-the-art Surface Water Vehicles

Unmanned surface vehicles (USVs) are integral to aquatic research, environmental monitoring, and various aquatic applications. These vehicles are designed to operate on the surface of water bodies, such as rivers, lakes, and oceans, and perform various tasks. Furthermore, USVs offer several advantages, including versatility, mobility, and the ability to navigate through different water conditions. These vehicles are equipped with sensors, cameras, and other instruments, enabling researchers to collect data on water quality, temperature, turbidity, and other environmental parameters. They play a vital role in studying and monitoring aquatic ecosystems, assessing the impact of human activities, and understanding the effects of climate change on water bodies. USVs can also be used for search and rescue operations, disaster response, and environmental cleanup efforts. They can be remotely operated or autonomous, providing flexibility and adaptability in diverse scenarios. The integration of advanced technologies, such as artificial intelligence, machine learning, and sensor fusion, enhances the capabilities of these platforms, enabling real-time data analysis and decision-making. With their ability to access remote or unsafe areas, these vehicles contribute to scientific research, environmental management, and the sustainable use of water resources.

In literature, a distinct categorization of unmanned vehicles is notably lacking, and in many instances, these vehicles are classified based on their specific applications, as highlighted by Liu et. al. [Liu 16]. The primary applications encompass areas such as 1) Bathymetry, 2) Scientific Research, 3) Urban Waterways, 4) Marine Exploration, 5) Military, and 6) Other Applications. Primarily, this thesis focuses the post-flood environment which is categorized under shallow water environment, thereby specifically targeting ports, lakes, and shallow rivers. Consequently, unmanned vehicles designed for operating in shallow water environments predominantly fall within the domains of scientific research or urban waterways. Within these specified domains, individual vehicles can be further categorized based on their sensory modalities, distinguishing between surface water and underwater capabilities.

The subsequent subsections of this thesis explore existing surface water vehicles, providing an in-depth examination of their mechanical designs, sensor systems, and the diverse range of applications they serve. A comprehensive comparative analysis of these vehicles is conducted, shedding light on their respective strengths and limitations. Furthermore, the analysis identifies existing gaps in the state-of-the-art surface water vehicles, pinpointing areas where advancements or innovations are required to enhance their overall efficacy and performance.

### 2.2.1 Vehicles for Bathymetry

Bathymetry is the study of waterbed of waterbodies including the ocean, rivers, streams, and lakes. This study includes the depth and shape of waterbed. It is an important aspect of marine science and engineering, as accurate bathymetric data is essential for a wide range of applications, including navigation, oil and gas exploration, and environmental monitoring.

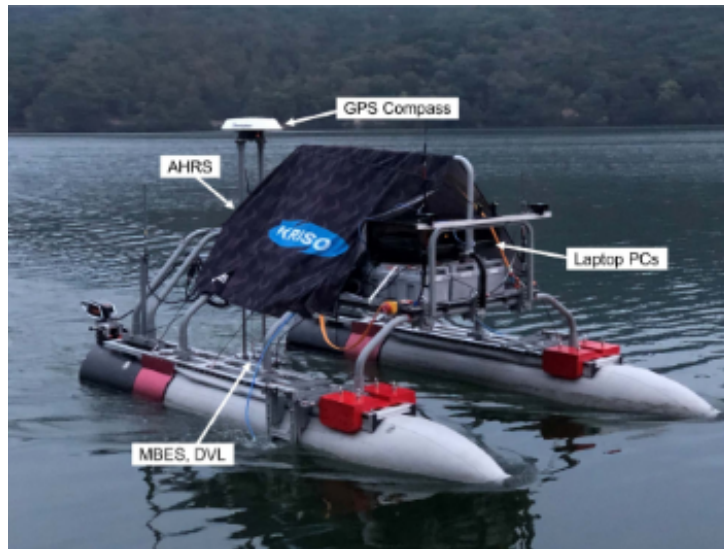
In recent years, there has been increasing interest in using bathymetry in surface water vehicles, such as boats and autonomous underwater vehicles (AUVs). These vehicles are equipped with various sensors and instruments that can be used to collect bathymetric data, including sonar systems, echo sounders, and laser scanners. One of the main advantages of using surface water vehicles for bathymetry is their ability to collect data over large areas in a relatively short amount of time. This is particularly useful in cases where the bathymetric data needs to be updated frequently, such as in areas where tides, currents, and sedimentation rates are constantly changing. Additionally, surface water vehicles are helpful to collect data in shallow waters and near shore areas, where traditional methods such as ship-based surveys may not be practical. These areas can be difficult to access due to their proximity to the shore or the presence of obstacles such as rocks and reefs.

#### ASV by KRISO

The Korean Research Institute of Ships and Ocean (KRISO) has developed an ASV, shown in Figure 2.1. The primary goal of this ASV is to serve as a versatile platform for testing and validating advanced navigation and control algorithms in complex surface water environments. The ASV's unique catamaran (length: 4.1 m; width: 2.4 m) configuration provides flexibility in accommodating a wide range of variable payloads.

To enhance the ASV's navigational capabilities, a multi-sensor system have been integrated to measure its motion parameters in real-time. The sensor system includes a GPS compass, an Attitude Heading Reference System (AHRS), and a Fiber Optic Gyro (FOG). These sensors provide precise and accurate information about the ASV's position, orientation, and heading. In addition to the sensor array for motion data, the ASV presents a robust situational awareness system. Optical sensors, such as a stereo camera and LiDAR, have been integrated to detect and identify objects floating on the water surface. These sensors help the ASV to perceive and respond to its environment effectively.

Furthermore, the ASV is equipped with three distinct acoustic sensors: Multibeam Echo Sounder (MBES), Doppler Velocity Log (DVL), and an acoustic modem. These acoustic devices operate at different frequencies that do not overlap, enabling concurrent usage without interference. The MBES provides detailed bathymetric information about the



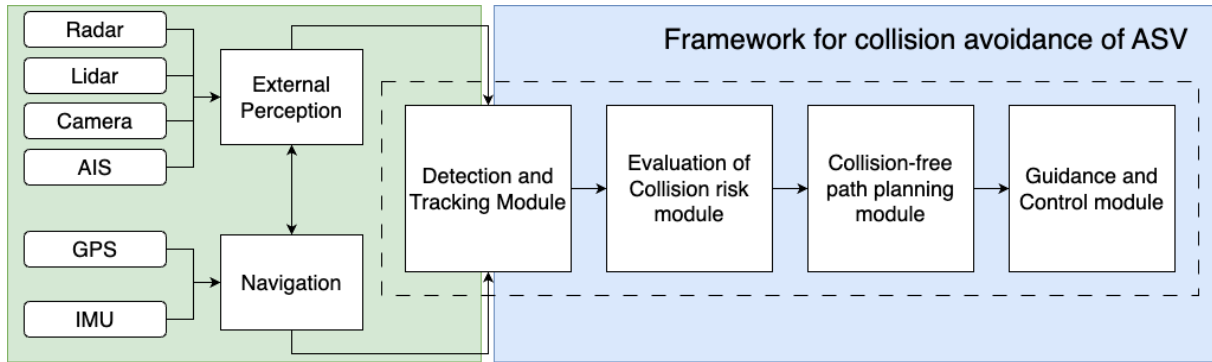
**Figure 2.1:** Surface water vehicle developed by KRISO. The vehicle is equipped with GPS Compass, Attitude Heading Reference System (AHRS), Multibeam echosounder (MBES) and Doppler Velocity Log (DVL). [Jung 22]

water column and seafloor, the DVL aids in velocity measurements and underwater navigation, and the acoustic modem facilitates communication and data exchange in underwater scenarios.

The navigation system of ASV comprises two distinct methodologies: acoustic-based and terrain-based. In acoustic-based navigation, the ASV uses acoustic sources as underwater reference points. These stationary sources emit specific acoustic signals periodically, which are received and interpreted by a pair of hydrophones installed on the ASV. The received acoustic data is then employed to refine the ASV's position and orientation information using the Extended Kalman Filter (EKF) Simultaneous Localization and Mapping (SLAM) algorithm. This acoustic-based navigation strategy enhances the ASV's underwater positioning accuracy.

In the terrain-based navigation approach, the ASV uses a pre-surveyed underwater terrain map generated using a multibeam sonar. This map is a prior reference for the ASV's navigation and is integrated into a particle filter system, which allows the ASV to continuously update its position and motion based on real-time terrain data. Fusing this terrain-based information with acoustic-based navigation enhances the ASV's ability to navigate effectively in diverse underwater environments.

Moreover, the ASV is equipped with an autonomous navigation system that has waypoint guidance and collision avoidance mechanisms. The guidance approach simplifies the ASV's behavior into a three-degree-of-freedom (3-DOF) kinematic model. The motion model employed in guidance and control strategies emphasizes minimizing line-of-sight (LOS) and cross-track errors, enhancing the ASV's ability to follow its intended path precisely. This is achieved by implementing a conventional Proportional-Derivative (PD) controller designed to regulate the ASV's course, ensuring it progresses toward its goal waypoints accurately. On the other hand, the collision avoidance framework (shown in figure 2.2) within the ASV's navigation system comprises three core modules: an evaluation of collision risk module, a collision-free path planning module, and a guidance and control module. This



**Figure 2.2:** Collision Avoidance Framework [Choi 20b]

system works collectively to assess potential collision risks, compute safe and obstacle-free paths, and guide the ASV along these paths, minimizing the possibility of collisions with underwater obstacles.

### 2.2.2 Research Vehicles

Scientific research-oriented vehicles designed for shallow water environments often emphasize precision and data collection. These vehicles leverage advanced sensor systems, including high-resolution cameras, LiDAR, and sonar, to capture detailed information about the water surface and its underlying topography. The mechanical design of such vehicles incorporates features that enhance stability and maneuverability, allowing them to navigate challenging environments with agility. Applications of these vehicles range from ecological studies to environmental monitoring, showcasing their versatility in contributing valuable data for scientific endeavors.

#### Water Strider

The Water Strider is an autonomous surface vehicle developed by Fraunhofer IOSB with a specific mission: to create accurate and up-to-date maps of shallow water environments above and below the water's surface. These maps ensure safe and efficient navigation of waterways and surrounding infrastructure.

The Water Strider, shown in Figure 2.3, is a significantly large vehicle. It measures 6000 mm in length, 6000 mm in width, and 4000 mm in height and weighs around 500 kilograms. It is highly agile and versatile, with a top speed of 3.3 meters per second. The vehicle is built on the Ducky 19 catamaran platform, which provides stability and maneuverability for operations in aquatic environments. Furthermore, the ASV has a comprehensive array of sensors, such as multibeam sonar, LiDAR, cameras, Inertial Measurement Unit (IMU), and Global Navigation Satellite System (GNSS). This sensor system is vital in facilitating navigation, obstacle detection and avoidance, and mapping tasks above and under the water's surface.

The vehicle uses a 3D LiDAR sensor and two mast-mounted cameras. The LiDAR sensor, combined with the vehicle's motion data (ego-motion), is employed to dynamically construct a map of the surrounding environment, including identifying obstacles within the vehicle's operational space. The dynamic map has short-term memory, allowing the vehicle



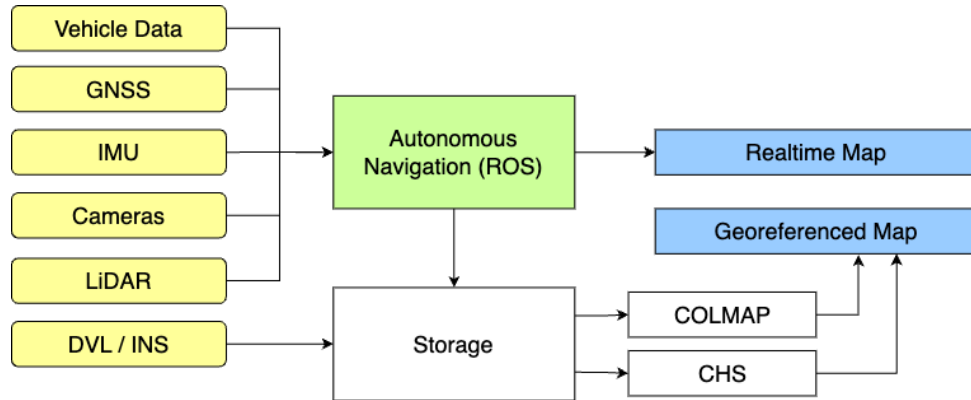
**Figure 2.3:** Water Strider during the competition final in Kalamata, Greece

to handle static and dynamic obstacles in real-time. Simultaneously, the cameras capture images processed through photogrammetric 3D reconstruction techniques. These images are transformed into high-resolution 3D reconstructions of the above-water surroundings, providing valuable visual data. For underwater mapping, the Water Strider relies on sonar technology. Specifically, it utilizes a Bathyswath sonar transducer coupled with the Sonardyne Inertial Navigation System (INS) for georeferencing bathymetric data. The sonar data is then processed using the Bathyswath Swath Processor and CARIS HIPS&SIPS software. This approach involves the combination of overlapping scans of the seafloor or riverbed, culminating in the creating of a comprehensive depth map. This underwater mapping capability results in a holistic view of the submerged environment.

The Water Strider employs a control strategy that combines autonomous control for path following and obstacle avoidance with the ability for operator intervention, ensuring safe and effective operations during shallow water mapping missions (shown in figure 2.4). The control system's multi-layered approach includes velocity profile computation, path control, and the Line-of-Sight (LOS) algorithm, which determines the necessary vehicle speed and heading. The system relies on a PI controller for speed control, managing the rotation of electric motors, while a P controller is used for heading control, computing the required rudder angle. Notably, these control methods do not rely on a kinematic or dynamic model of the ASV.

The vehicle can operate autonomously within a predefined area of interest. Users define a polygon representing the desired operational area, and the ASV can then autonomously navigate within that defined polygon. A path planning algorithm generates an optimized path for covering the polygon, factoring in the vehicle's starting position. The ASV follows the designated path at a user-specified speed. It also can halt or adjust its path in response to failures or obstacles.

Furthermore, the Water Strider offers an operator intervention mode. It features an interactive digital map table that serves as the interface between the operator and the



**Figure 2.4:** Dataflow and Software Architecture of Water Strider. [Zube 22]

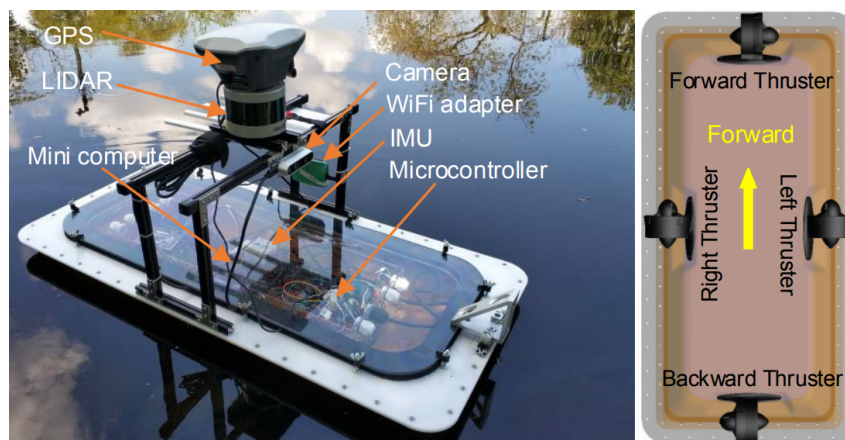
vehicle. Operators can specify missions, evaluate system performance in real-time, and oversee the ASV's operations. The digital map table streamlines mission definition and assessment, offering an intuitive means of interacting with the vehicle.

### 2.2.3 Urban Waterways Vehicles

Surface water vehicles, such as boats and autonomous underwater vehicles (AUVs), can be used in urban waterways for a variety of purposes, including transportation for many coastal and riverside cities, water quality monitoring, and trash collection. Urban waterways, such as rivers and canals, are often subject to high levels of pollution and environmental degradation, making them an important area of focus for environmental monitoring and management.

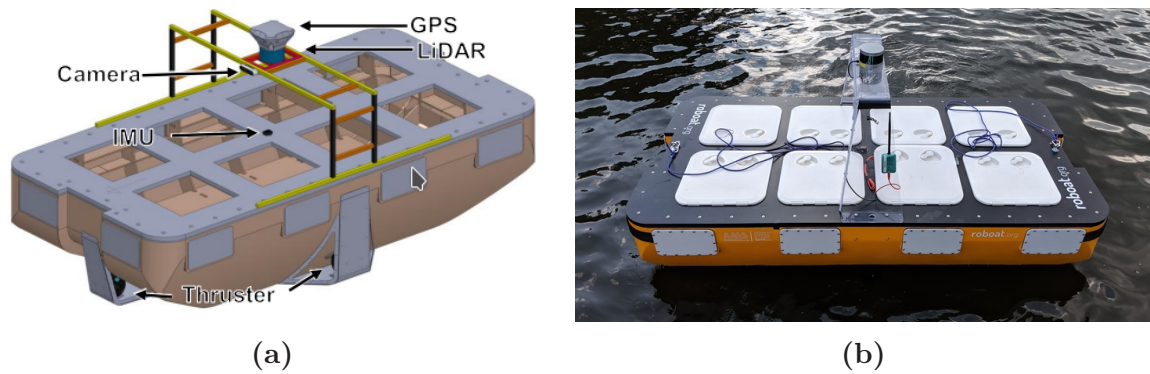
#### ROBOAT

ROBOAT is an advanced project in the field of autonomous surface vehicles aimed at reducing urban traffic congestion by utilizing water-based transport networks. This innovative initiative is a collaboration between prestigious institutions such as the Massachusetts Institute of Technology (MIT), Delft University of Technology, and Wageningen University



**Figure 2.5:** Roboat : Quarter Scale Prototype. (Left) Figure shows the sensor system. (Right) Shows the position and orientation of four thrusters.

& Research, with the first deployment in Amsterdam, Netherlands. These self-navigating watercraft perform various tasks, including passenger transportation, logistics such as garbage collection and package delivery, water infrastructure surveying, and monitoring water quality.



**Figure 2.6:** Roboat : Half Scale Prototype. (a) Simulation model shows the position of sensor system and thrusters (b) Completed prototype floating in Amsterdam’s urban waterways. [Wang 20b]

The Roboat has undergone significant advancements since its inception in 2017. The inaugural prototype, Roboat I (in figure 2.5), was comparatively smaller, weighing a mere 15kg and measuring 900 x 450mm. It was powered by an 11.1V battery with an operational time of 3 hours and featured four T100 thrusters, in addition to a suite of sensors such as 16-line 3D lidar, a depth camera, GPS, and IMU. The second prototype, Roboat II shown in figure 2.6, was introduced the following year. It had a larger hull measuring 2000 x 1000 mm, was made of wood and fiberglass, and weighed around 80 kg. It offered an operational window of 2 hours and retained the sensor configuration of its predecessor. The design centered around simplicity of construction and interconnectivity among vessels, allowing for multiple units to be assembled in parallel or perpendicular orientations to create larger configurations. The thruster configurations of both boats (Roboat I/II [Wang 19a] [Wang 20b]) allow holonomic movements. This is especially valuable for autonomous docking and latching for platform building.

In 2019, the Roboat project achieved a significant milestone by creating the most advanced version of the technology. This full-scale version (figure 2.7) had dimensions of 4000 x 2000 mm and a symmetrical design that enabled it to maneuver in tight spaces, regardless of the bow’s orientation. Each boat had a replaceable top deck, making it more adaptable and versatile. The navigational capabilities of the boats were highly precise and accurate to within decimeters. This makes them proficient in operating on water surfaces, navigating complex urban waterways, and carrying out various tasks with minimal human intervention.

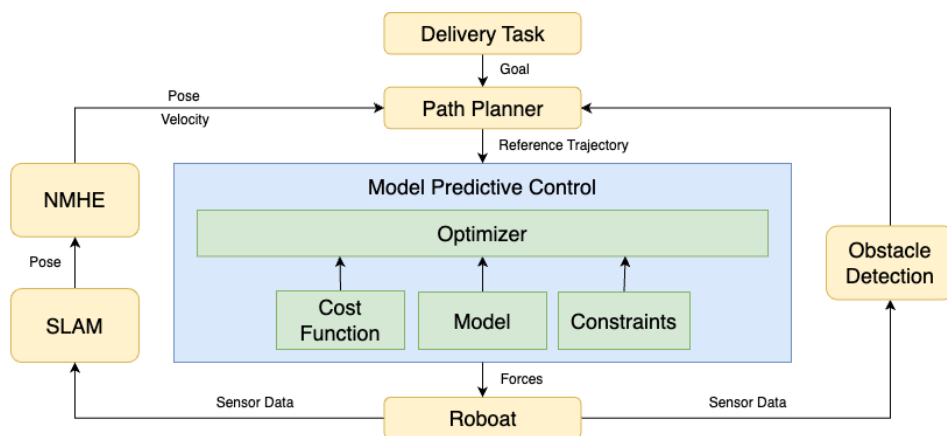
Accurate navigation and successful docking within urban waterways hinge on precise localization. To achieve this, the authors have implemented a novel system that fuses data from LiDAR, cameras, and GPS sensors, providing the system with a decimeter-level localization precision. These sensors are adept at detecting and avoiding static and dynamic obstacles positioned above the water surface, significantly enhancing the system’s safety and navigation capabilities. Roboat employs a factor graph to facilitate the fusion of data from multiple sensors and to formulate a solution for the simultaneous localization



**Figure 2.7:** Roboat : Full Scale Prototype. (a) Final Roboat version as urban transportation in Amsterdam. (b) The top module of Roboat can be replaced for other applications including load transportation and garbage collector. [Wang 20b]

and mapping (SLAM) problem. This integrated system introduces three crucial factors into the factor graph framework: lidar odometry factors, GPS factors, and loop-closure factors, each contributing to the robustness of the localization and mapping processes.

The system further enhances its capabilities by computing two distinct sets of features: edge features and planar features. Feature extraction is conducted by transforming the lidar point cloud into a range imagery, where each pixel corresponds to a specific point within the point cloud dataset. This transformation is essential in managing the massive data generated by the lidar scans. To streamline data processing, the proposed technique selects new keyframes based on predefined thresholds related to the vehicle's pose changes. For example, in the case of Roboat II, the chosen thresholds encompass a 1-meter threshold in position change and a  $10^\circ$  threshold in rotation. The scans that lie between these keyframes are ignored, and the transformation between the two keyframes is subsequently computed. The transformed keyframes are combined into a voxel map, contributing to the system's mapping and localization strategies.



**Figure 2.8:** Autonomy framework of Roboat II. It includes a planner, a SLAM module, a Nonlinear Model Predictive Control (NMPC) tracking module, a Nonlinear Moving Horizon Estimation (NMHE) state estimator, and an object detector. [Wang 20b]

Both vehicles within the Roboat project are equipped with Intel NUCs, and their control



software is founded upon the Robot Operating System (ROS), a widely recognized platform for the development of robotic applications. The complex dynamics of aquatic environments necessitate the utilization of an online nonlinear model predictive controller (NMPC), with the dynamical model being experimentally estimated to achieve satisfactory results in tracking control. Simultaneously, state estimation is conducted via a nonlinear moving horizon estimation (NMHE) algorithm. The combination of these strategies helps in real-time mapping and localization, execute path planning, and accurately track the computed trajectory, making it a complex and adaptable aquatic robotic platform. Figure 2.8 illustrates the overall autonomy framework of Roboat I/II. It is noteworthy that the system does not incorporate sonar sensors for detecting obstacles under the water's surface.

## 2.2.4 Environmental Missions

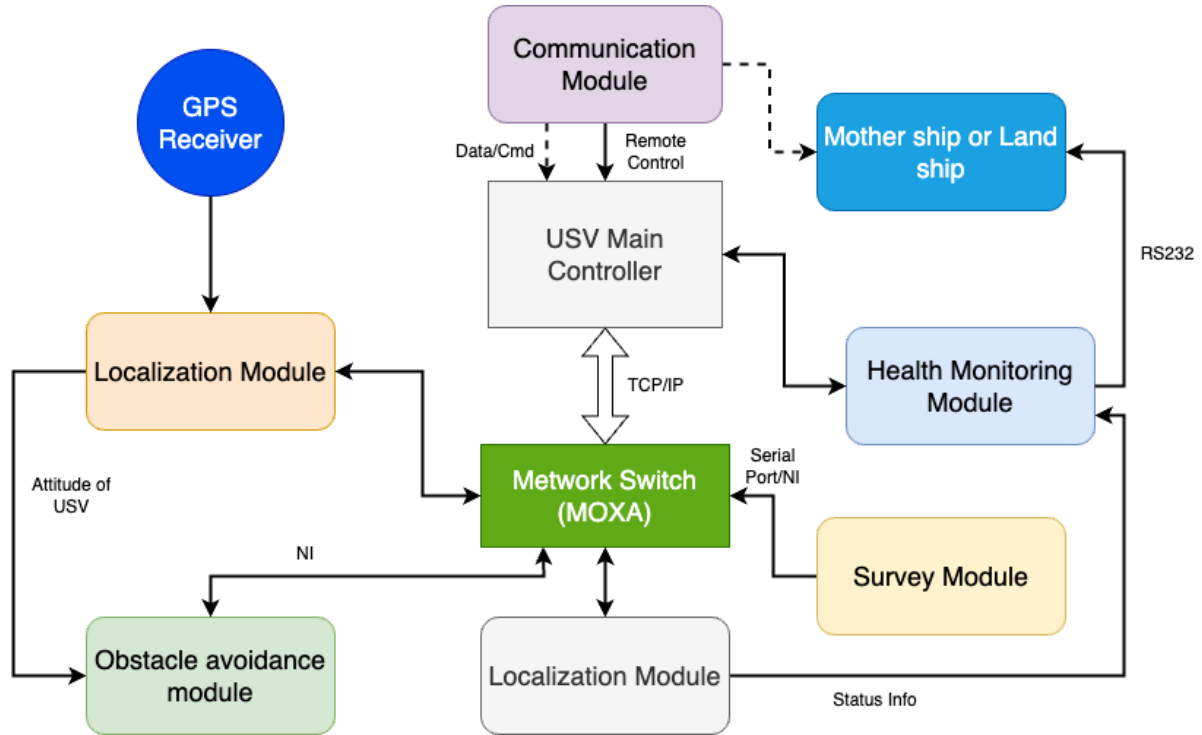
### JingHai-I

The JingHai-I is an autonomous surface water vehicle developed by Shanghai University, explicitly designed to provide a versatile platform for testing and validating a wide range of algorithms. Primarily intended for hydrographic surveying and charting in nearshore shallow waters, this USV (Unmanned Surface Vehicle) offers robust capabilities, including long-range missions, collision avoidance, and the ability to navigate shallow waters with a minimal draft. Figure [Li 22] shows the USV a during a test survey.



**Figure 2.9:** JingHai-I surface water vehicle. The vehicle is equipped with multi-sonar system, including echosounder, bathymetry, side-scanner, and forward-looking sonar. [Li 22]

In terms of physical dimensions, the JingHai-I measures 6.28 meters in length and 2.86 meters in width, with a total system weight of approximately 2300 kilograms, taking into account the installed equipment. The vehicle is equipped with a Global Positioning System (GPS) and an attitude heading reference system (OCTANS) to precisely determine its location and orientation, encompassing heading, roll, and pitch. To enhance its navigation and obstacle avoidance capabilities in real-time, the USV features a laser range finder and a forward-looking sonar. In addition to these features, the JingHai-I has a suite of sensors, including single-beam bathymetry, multibeam sonar, side-scan sonar, forward-looking sonar, and an acoustic Doppler current profiler (ADCP) to collect important



**Figure 2.10:** System Architecture of JingHai-I [Peng 17]

environmental data. These sensors provide critical information for monitoring the marine environment during surveying and mapping operations. The complete architecture of JingHai-I is shown in figure 2.10.

The control system of the JingHai-I comprises two main components: the mother ship and the USV itself. The USV's communication system enables it to establish real-time connections with the mother ship or land station using a radio frequency modem or maritime satellite communication. This facilitates the exchange of data and control commands. The communication capability ensures seamless coordination between the USV and its operators or base stations. It utilizes Inmarsat (Fleet Broadband) for communication, primarily when the vehicle operates at a distance. Furthermore, an R/F modem is employed for remote control when the USV is beyond 2 kilometers from the base station.

The mother ship can remotely operate or plan the USV's path, offering control via a remote handle or portable control box. It can also receive data from the USV. On the other hand, the USV carries a camera, sonars, and a hydrological measurement interface. A dedicated port supports image transmission, allowing large video data streams to be transmitted. Microwave communication technology is applied for digital control signals and data acquisition transmission.

The USV's control system further comprises an embedded industrial control computer, a serial extension board (DSP), and a network switch. The industrial control computer is responsible for collecting, processing, and storing data from the measurement devices. It possesses the capability to safeguard data in the event of communication issues, reducing the risk of data loss.

## CHARLIE

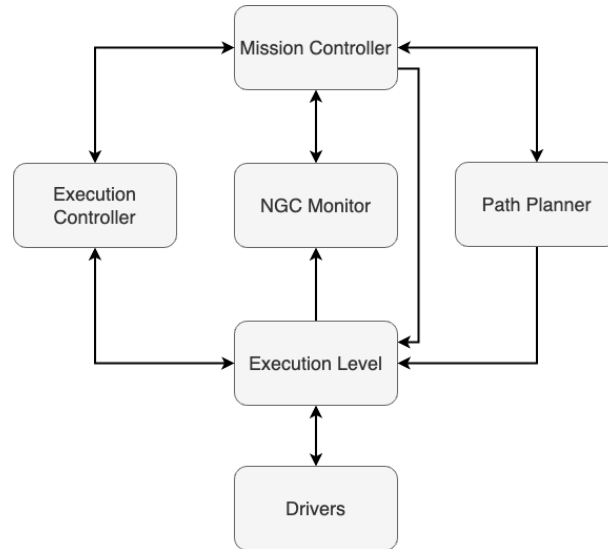
The Charlie Unmanned Surface Vehicle (USV) is a versatile testbed that promotes collaboration among diverse research groups dedicated to advancing the field of unmanned marine vehicles. Its primary purpose is to experimentally validate the algorithms and methodologies proposed by these research groups. As a result, it provides a valuable platform for investigating various aspects of unmanned marine technology; for instance, figure 2.11 shows Charlie gathering data at Genova harbor. The research it supports covers topics such as embedded real-time platforms for robotics and industrial automation, modeling and identification, fundamental navigation techniques, guidance, and control, path following in straight-line and generic trajectories, cooperative mission control, and path tracking.



**Figure 2.11:** Charlie testing in Genova harbor. [Caccia 07]

In terms of its physical attributes, the Charlie USV integrates on-board computing capabilities tailored for navigation, guidance, and control. Its catamaran-like design imparts stability, redundancy in hull buoyancy, and the capacity for transporting payloads. It presents a propulsion system driven by DC motors with two propellers, a steering mechanism featuring dual rudders, and a power supply structure comprising lead batteries supplemented by solar panels. To ensure precise positioning and accurate heading information, it is equipped with a GPS and an azimuth gyro. Communication with remote control and supervision stations is achieved seamlessly through either a wireless LAN or a radio modem.

The control system architecture of the Charlie USV is designed in a hierarchical structure as shown in figure 2.12, which consists of different levels. At the lowest level, there are software modules termed "Drivers" that handle input/output operations, managing actuation devices (like thrusters and rudders) and sensing instruments (such as GPS, compass, and echosounder). Above this is the "Execution Level", responsible for executing fundamental Navigation, Guidance, and Control (NGC) tasks. These tasks interact through shared variables, enabling coordinated and cohesive performance. This level communicates with the Drivers, providing reference values to actuation devices and receiving instrumental measurements from sensors.



**Figure 2.12:** Hierarchical control architecture of Charlie [Caccia 09]

The "Execution Controller" is responsible for the activation, deactivation, and initialization of guidance, control, and estimation tasks. It manages conflicts and dependencies concerning task resources, notably variables, to ensure system consistency. This level is event-driven and uses Petri nets to describe the connections among estimation and control tasks. Unlike the Execution Level, which primarily operates at the continuous-time layer, the Execution Controller transitions states in response to commands from other architectural components or human operators. These commands may encompass explicit requests for task activation, notifications from the Execution Level, or signals from a higher-level planner layer. The integration of a "Mission Control" module improves the control architecture, entailing the development of modules with the capacity for online re-planning of the vehicle's motion, known as the "Path Planner." Additionally, it involves signaling specific events linked to the evolution of continuous-time variables, a function expertly managed by the "NGC Monitor." The Path Planner calculates and manages a collision-free path to a predetermined destination, taking into account constraints related to waypoints, local tangent, and curvature. The NGC Monitor cooperates closely with the Mission Controller, essentially serving as the interface layer that bridges continuous-time-driven modules with their event-driven counterparts.

In essence, the Charlie USV serves as a robust framework for integrating a spectrum of algorithms and methodologies, streamlining the testing and evaluation of different guidance and control architectures. This innovative architecture enhances the adaptability and ease of experimentation in the field of unmanned marine vehicle technology.

## 2.3 Comparative Analysis

To keep the discussion brief, the preceding section only focuses on a few well-known surface water vehicles (Table 2.1). However, it is acknowledged that many other water vehicles exist in the literature. The rationale behind this choice is that the discussed vehicles are recently published and provide detailed architectural information relevant to the presented thesis. Nevertheless, other research efforts are included in this comparative analysis that

**Table 2.1:** Existing surface water vehicles and their essential characteristics. The table provides a comparative overview of different surface water vehicles, listing their names, structural design, dimensions, weight, sensor outfitting, and distinctive operational capabilities. It includes details about the type of vessel, ranging from monohull to catamaran, and the array of onboard equipment such as navigational aids and sensory devices. Salient features of each vehicle are also highlighted, emphasizing their unique functionalities and suitability for specific tasks in maritime operations.

| Name                               | Type                  | Dim (mm)                                | W (kg)          | Sensors  | Salient Features   |
|------------------------------------|-----------------------|---|-----------------|--|--|
| ROBOAT<br>[Wang 19a]<br>[Wang 20b] | Monohull<br>Catamaran | 900 × 450<br>2000 × 1000<br>4000 × 2000 | 15<br>80<br>150 | LiDAR,<br>GPS, IMU,<br>Depth Camera                          | Obstacle Avoidance,<br>Automated Docking<br>Path Planning                        |
| KRISO<br>[Choi 20b]                | Catamaran             | 4100 × 2400                             | 85              | GPS, MBES,<br>DVL,<br>Acoustic Modem                         | Waypoint Navigation<br>Acoustic-based Navigation<br>Terrain-based Navigation     |
| SWAMP<br>[Odetti 20]               | Catamaran             | 1230 × 700                              | 38              | GPS, IMU   | Underwater Mapping   |
| Water Strider<br>[Zube 22]         | Caterman              | 6000 × 4000                             | 500             | Camera, LiDAR,<br>Sonar,<br>GPS                              | Path Generation<br>Obstacle Avoidance<br>Vehicle Control                         |
| JimHai-I<br>[Li 22]                | Water-jet Propulsion  | 6280 × 2860                             | 2300            | GNSS, LiDAR,<br>Radar, Camera,<br>Attitude Heading Reference | Collision Avoidance<br>Path Generation   |
| Charlie<br>[Caccia 07]             | Caterman              | 2400 × 1800                             |                 | GNSS, Compass<br>Echosounder Camera                          | Coast Autonomous Navigation<br>Underwater Anti-intrusion System<br>Path planning |

are relevant to surface and underwater sensor approaches and strategies, even if they do not comprise complete vehicle architectures. This approach has allowed for a more comprehensive examination of relevant research and has highlighted the diversity and extent of contributions in the field. The analysis is categorised based on the comparison of environment and system of available architectures.

**Environment-wise** The previously mentioned autonomous vehicles have become increasingly popular in aquatic environments. However, most systems are designed for open water bodies or structured surroundings. For instance, the Roboat (as discussed in Section 2.2.3) is purposefully designed to navigate through structured urban environments. Similarly, other vehicles in this domain are primarily intended for marine research where there are very few random obstacles in the water. Moreover, the reduced complexity of the aquatic environment in terms of surface or underwater obstacles simplifies the task of object detection and path planning, making it possible for applying various algorithms and strategies from the field of deep learning.

Moreover, the existing literature focuses on natural water bodies characterized by clear and relatively unpolluted water conditions [Mogstad 19]. This is an essential aspect as turbid water in a flooded environment introduces substantial variations in the raw data obtained from sensory modalities, particularly an issue for underwater object detection and the generation of traversability maps. In this context, many autonomous systems have been developed to employ vision sensors in underwater applications. Notably, works by [Zhang 22] and [din 17] have put forth innovative techniques in image enhancement specifically tailored for underwater environments, leveraging camera-based imagery. Similarly, research contributions by [pen 21] and [Trucco 06] have offered novel methods for image restoration, primarily aimed at underwater object detection. However, it is noteworthy that these methods operate under the fundamental assumption of clear water conditions.

One significant factor that affects the choice of sensor system in the aquatic environment is the depth of the water. The water's depth is critical in the selection criteria for underwater sensor systems. For instance, in the context of sonar-based systems, it is essential to note that sonar data can be affected by multiple echoes originating from the waterbed, which lower the object-to-background ratio, consequently leading to a reduced object detection rate. Consequently, in such scenarios, it becomes important that the choice of sensor, along with its specific attributes and characteristics, should be dependent on the available water conditions, depth, and illumination levels to ensure optimal performance in underwater sensing and perception tasks.

**System-wise** – The architectural designs discussed in the earlier sections are specifically tailored to typical types of watercraft. This means these systems lack the flexibility to adapt to different vehicle dimensions and kinematic characteristics. Moreover, it is noteworthy that most of these systems exhibit a mono-modal nature, implying that they are primarily designed for either surface water or underwater operations, with limited cross-functionality. For instance, [Han 19] implements the SLAM using radar data of coastal surface in gps restricted areas and [Sinisterra 17] presents the USV for ship detection over the surface with camera imagery. Contrarily, [Raber 19] designed low-cost USV to map coral reefs using underwater sonar. Similarly, a complex

system is presented in [Karoui 15] to generate mosaics of underwater scenes using similar forward-looking multibeam sonar, but it does not consider the surface water environment. In instances where they do incorporate multi-modal capabilities (in case of water strider in section 2.2.2), the data processing and mapping procedures are distinctly segregated between the different sensor modalities. This lack of reconfigurability and integration between surface and underwater domains represents a noteworthy constraint in scenarios that demand versatile and adaptive aquatic platforms. In order to achieve interoperability among various types of boats and comprehensive mapping and situational awareness, more versatile and adaptable architectures in aquatic autonomous systems are needed. Future advancements may address these limitations to enable more agile, cross-domain operations and efficient data fusion techniques.

## 2.4 Research Gap

The increasing impact of global warming has resulted in a significant increase in the occurrence and strength of natural disasters across the globe. Among these calamities, floods have become one of the most common and environmentally damaging disasters. When a flood occurs, it causes significant changes to the environment, leading to complex challenges in rescue and relief operations such as unknown underwater environment, random obstacles and insufficient maps. These challenges highlight the need for innovative technological solutions to improve the speed and accuracy of these operations. Specifically, surface water vehicles equipped with multi-sensor system can play a pivotal role in this context. These vehicles are well-suited to perform essential functions, primarily focused on surveillance and mapping of the flooded areas.

The previous section presents a detailed comparison of existing surface water vehicles in the commercial and research domains and highlights their operating environments and system characteristics. However, it is worth noting that, to date, there is a significant gap in the literature related to the development of similar systems designed explicitly for disastrous areas, particularly in post-flood environments. This absence of relevant research or commercial solutions for post-flood environment highlights a critical research and development opportunity.

In such disaster-prone scenarios, where traditional means of transportation and data collection are often compromised or rendered ineffective, the demand for specialized surface water vehicles that can operate in the aftermath of floods is particularly demanding. These specialized systems must be equipped to navigate and collect data in complex and dynamic aquatic environments, facilitating rapid response, disaster assessment, and recovery efforts.

The existing systems encounter notable challenges when applied within post-flood environments. These challenges are discussed in detail below:

**Presence of Unstructured and Random Obstacles** – Floods result in widespread destruction, introducing various unstructured and scattered objects such as fallen trees, submerged vehicles, and damaged bridges, which hinder effective navigation. The current systems are ill-equipped to detect and respond to these obstacles, making them unsuitable for post-flood operations.

**Opacity of Floodwater** – Floodwater is often filled with high concentrations of sand and mud, making it opaque and reducing water clarity. This turbidity complicates the effectiveness of vision-based cameras as they struggle to capture meaningful imagery, making it difficult to assess the environment.

**Variable Water Depth** – The depth of water in a flooded area can vary significantly, making it challenging for sonar-based imaging systems. The varying depths affect the behavior of sonar signals, making it difficult to generate accurate imagery. As a result, currently available systems may not provide precise depth information in such dynamic environments.

**Comprehensive Traversability Mapping** – Effective navigation in post-flood scenarios requires a comprehensive understanding of both the surface and underwater terrain. However, existing systems often lack the capability to create detailed traversability maps that encompass both surface and underwater features. Such integrated mapping is essential for optimizing the navigation of rescue boats and vehicles in the intricate and obstacle-laden post-flood landscape.

Given these challenges, it is evident that dedicated and novel strategies are required to develop surface water systems specifically designed for post-flood environments. The systems should have enhanced adaptability and versatility to effectively address the complexities and uncertainties inherent in such disastrous scenarios. By tackling these challenges, future systems can play an essential role in augmenting the efficiency and safety of rescue and relief efforts following floods and similar natural disasters.



### **3. Conceptual Overview of the Surface Water Traversability Mapping (SWiM) Architecture**

Traversability mapping is a vital component for autonomous and remotely operated vehicles to navigate safely and efficiently through unknown environments. Evaluating traversability involves determining whether a path is safe and feasible for a vehicle to travel through while accomplishing its mission objectives. The traversability assessment considers various factors, such as terrain topography [Seraji 03], potential obstacles [Zhenpeng 09] [Ye 07] (such as rocks, wrecks, or dynamic obstacles), and environmental conditions [Yang 18] (such as currents and turbulence in water). Researchers and engineers can create traversability maps by analyzing these factors, which provide valuable insights into safe navigation areas and those that may pose risks or complications. These maps are essential in guiding vehicles, helping them plan optimal paths, avoid collisions, and make real-time decisions to adapt to changing conditions. Understanding traversability is essential for successful missions with various applications, including search and rescue [Dube 16], marine exploration [Jain 15], environmental monitoring [Ruetz 22], and underwater infrastructure inspection [Teixeira 16].

Traversability mapping is significant not only in terrestrial or aerial but also in aquatic environments [Oskard 89]. It enhances navigational safety, facilitates the exploration of underwater natural resources, and assists in disaster preparedness and response. In particular, the aquatic environment is categorized as shallow or deep, based on the water depth. Shallow water, typically less than 30 meters deep and located near coasts and riverbeds, is affected by weather and tides and has muddy waterbeds, which can impact underwater sensing such as sonar imagery due to the variable speed of sound and reflections. In contrast, deep water, usually greater than 200 meters deep and found in the middle of oceans, is cooler and less temperature variable, leading to less change in the speed of sound. These differences significantly influence the physical characteristics of an environment and sensory modalities including sonar.

This thesis proposes a novel architecture – SWiM (Shallow Water traversability Mapping) for surface water platforms to navigate safely in shallow flooded environments that are less than 2 meters deep. Currently available architectures are suitable for freshwater or marine environments, which are much easier to navigate compared to post-flood environments. For instance, Roboat I/II [Wang 20a] is specifically intended for use in urban waterways, where the environment is well-structured and open, and depth changes are gradual. In contrast, post-flood environments are typically congested, highly unstructured, and have abrupt depth changes. Furthermore, urban waterways usually have maps that provide information about the structure of lakes and upcoming bridges. However, these maps are no longer relevant in post-flood environments due to the destruction, hence necessitate new techniques to create reliable and usable traversability maps. These maps are particularly difficult to generate because flooded water has limited visibility, and the water depth can be highly variable, often reaching as low as 50cm in extreme scenarios. Moreover, the very shallow water environment may create extreme problems in underwater sensing. Hence, in post-flood situations, producing accurate traversability maps is crucial for rescue teams to respond efficiently.

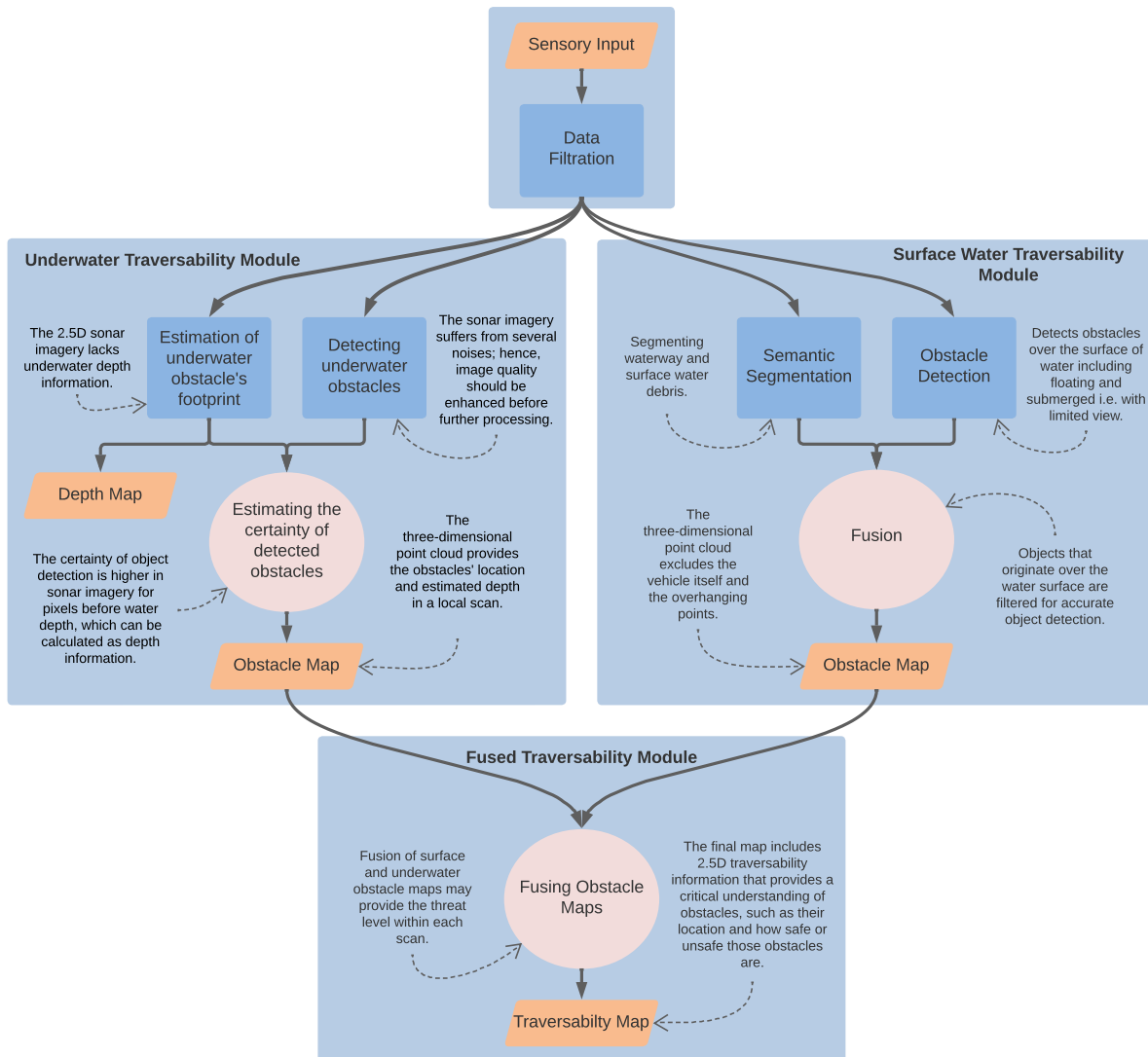
The proposed SWiM architecture is designed to meet environment and application-specific requirements as presented below:

- It detects potential obstacles with more than 80% accuracy in the unstructured environment.
- It analyzes the underwater and surface water footprints to classify obstacles as floating or sinking.
- It combines multi-domain and multi-modal information into one map (within  $\pm 25cm$  accuracy).
- It undergoes rigorous testing on multiple vehicles with varying dimensions and motion principles to ensure reconfigurability.

### 3.1 SWiM Architecture

The SWiM (Shallow Water traversability Mapping) architecture is specifically designed for generating traversability maps in very shallow water, more specifically a post-flood environment. This architecture recognizes the unique characteristics of flooded environments, which comprise two primary mediums: underwater and surface water. Due to the distinct physical properties of air and water, different sensory modalities are required for effective mapping in these two mediums. For instance, in underwater environments, sound waves are found to be more efficient than light waves due to water's high acoustic impedance, which hinders the propagation of light. Consequently, sonar emerges as the preferred sensor in an underwater environment, outperforming cameras or LiDAR scanners. On the other hand, in surface water environments, cameras and LiDAR sensors are more practical because light can travel through air with minimal energy loss compared to sound.

The SWiM architecture addresses the challenges posed by these distinct physical media by dividing the traversability mapping process into two primary modules: underwater



**Figure 3.1:** Block Diagram of the Shallow Water Traversability Mapping (SWiM) Architecture for Traversability Analysis. The figure shows how the system processes underwater and surface water sensory inputs to detect obstacles and generate maps that identify safe and unsafe areas for navigation, integrating these into a final map that offers a detailed view of the environment's traversability.

traversability and surface water traversability. Each module focuses on the unique characteristics and requirements of its respective medium. Within each primary module, the architecture generates detailed obstacle maps tailored to the specific medium. These maps capture information about obstacles and their locations. To provide a comprehensive view of the environment, a fusion strategy is employed to merge the information from both modules. This fusion strategy takes into account whether obstacles are floating or sinking, a critical distinction in water environments.

The resulting fused map is then transformed into a global coordinate system using GPS/GNSS data, allowing precise georeferencing of the map. Safety criteria are applied based on the surface and underwater footprint of each obstacle, contributing to a more accurate assessment of navigational safety. Additionally, the underwater traversability

module generates a depth map that enhances the understanding of underwater terrain. This depth map is integrated with the fused obstacle map in the 2D traversability map module, further enriching the map.

The architecture's methodology and workflow are visually represented in Figure 3.1. Subsequent sections of the thesis deals with the specific challenges, requirements, and abstract methodologies of each module, providing a comprehensive understanding of the SWiM architecture's design and functionality.

### 3.1.1 Underwater Traversability

#### Challenges

The underwater traversability brings several inherent challenges that must be addressed individually for safe navigation. The underwater module specifically addresses those challenges for the optimal performance of the overall architecture. These challenges can be broadly categorized into two groups: domain-specific and application-specific. Each of these categories is briefly discussed below.

**I. Domain-Specific** Traditional sensors such as LiDAR, cameras, and radars are ineffective underwater; hence, limited sensor options exist compared to those available for terrestrial or aerial use. This limitation is primarily due to the challenges posed by water's properties, such as its opacity and acoustic characteristics. Among the available sensor options, certain sensors like sonar prove to be more effective in underwater navigation, especially in opaque water environments. Sonar systems use sound waves to detect underwater features and hazards, making them a preferred choice in such scenarios. However, even advanced sensors including sonar have their own limitations. Sonar data can be affected by issues like high speckle noise, low signal-to-noise ratio, and low contrast between obstacles and the underwater terrain. These limitations must be addressed to ensure accurate and reliable navigation. Furthermore, evaluating the accuracy and performance of new underwater navigation techniques can be challenging due to the absence of standardized benchmarks and datasets. Creating benchmark datasets for underwater environments is crucial for testing and validating navigation systems.

**II. Application-Specific** In addition to the domain-specific challenges, the post-flood environment introduces a set of application-specific challenges that further complicate safe navigation, particularly in rescue operations. These challenges are due to the complicated and continually changing conditions of such environments. A large amount of sand particles lead to poor underwater visibility, making it difficult to navigate and identify objects. These sand particles also contribute to interference and noise in sonar data. Additionally, strong currents and fluctuating water levels can make it challenging for water vehicles to maintain their position and avoid collisions with debris or submerged obstacles. Moreover, the post-flood environment is often highly cluttered and unstructured due to random obstacles like submerged vehicles, fallen trees, and other debris that pose a safety hazard for water vehicles.

## Proposed Concept

The presence of significant noise within sonar imagery negatively impacts the detection precision and overall robustness of the underwater traversability module. As a result, it becomes imperative to incorporate noise removal and image enhancement techniques [Keen 23] to improve mapping accuracy downstream in the process significantly. Numerous denoising approaches have been documented in the literature, with many of them relying on a fusion of conventional image filtration methods, including Gaussian blur, image averaging, and adaptive thresholding. It is noteworthy that there is no universally superior technique, as most of these methods are tailored to specific applications and require parametric tuning.

Furthermore, sonar imagery typically provides a flat, two-dimensional view of the underwater environment, with each pixel representing the intensity of the reflected sound signal. However, this representation lacks depth information about obstacles and waterbed, essential for determining whether an obstacle is a safety concern and also for helping vehicles to navigate effectively. To overcome the limitation in an underwater environment, the depth of the waterbed can be used to estimate the depth of obstacles, allowing for safer and more precise navigation.

Moreover, as the distance from the sonar sensor increases, the resolution of the forward-looking sonar imagery decreases. This degradation in resolution occurs because of the forward-looking sonar's physical principle, in which sound waves spread out as they travel longer distances, resulting in less detailed images. To enhance the precision and effectiveness of detection and tracking algorithms, it is possible to employ a data quality assessment method similar to the one detailed in [Ropertz 17]. This technique allows for the identification and removal of pixels with low probability, ensuring that the data used for analysis is of higher quality and is more reliable. By implementing this approach, the system can focus on the most relevant and trustworthy information, ultimately improving the accuracy of its algorithms.

The limited underwater datasets poses a challenge to the implementation of deep learning-based object detection and segmentation techniques in underwater environments. Consequently, traditional image processing methodologies become essential in detecting underwater obstacles within sonar imagery. Techniques such as Canny edge detection, contour identification, and filtering are pivotal for detecting these obstacles. Further refinement of detection results is achieved through non-maximum suppression, which significantly enhances the accuracy of the identified obstacles. Object tracking plays an integral role in this process by assigning a probability of detection to each object, based on its relative movement to the vehicle. This probabilistic measure is critical for the subsequent integration of detected obstacles into traversability maps, where it significantly contributes to the evaluation and depiction of the certainty levels associated with each obstacle. Additionally, the development of a detailed simulated environment using Unreal Engine 4 offers a novel approach to augment the accuracy of obstacle detection. This simulated environment enables the testing and fine-tuning of the object detection algorithms, providing a controlled setting that mimics real-world post-flood conditions. This combination of conventional image processing and advanced simulation not only enhances detection accuracy but also bolsters the reliability of the traversability maps used in underwater navigation and exploration.

### 3.1.2 Surface Water Traversability

#### Challenges

In surface water vehicles, such as boats and autonomous marine platforms, the selection of sensors plays a pivotal role in achieving a comprehensive understanding of the surface water environment. In this context, two widely used sensor types are cameras and LiDAR scanners, each offering unique strengths and encountering distinct limitations when applied to water surfaces. Cameras are renowned for providing a visual perspective of the environment. They excel in capturing high-resolution images, encompassing color and texture details that are valuable for visual recognition and analysis. Furthermore, camera images are well-suited for applying deep learning techniques, enabling sophisticated object detection and classification. However, when operating in water environments, cameras face several formidable challenges. The reflective properties of water surfaces can give rise to glare and reflections within the captured images, effectively obscuring objects. This phenomenon poses a significant limitation to detecting objects in such conditions accurately. In contrast, LiDAR scanners are unaffected by reflections from water surfaces. Additionally, LiDAR exhibits superior performance in low-light or night-time conditions, making it a valuable choice for 24/7 operation. However, LiDAR data inherently lacks color information, limiting its ability to provide visual context. Furthermore, while cameras excel in detecting vegetation or surface features over water, they struggle to provide precise information about the distance and height of objects from the watercraft. This limitation can be particularly problematic when accurate measurements of objects' positions and dimensions are essential. In practice, relying solely on one of these sensors may be beneficial in specific scenarios but is often insufficient, especially in challenging environments like post-flood scenarios.

#### Proposed Concept

In scenarios where precision in distance and height measurements, as well as detailed information about object characteristics like color and texture, is of paramount importance, employing a combination of sensors becomes essential. The integration of both cameras and LiDAR sensors facilitates a holistic understanding of the surrounding environment, enabling surface water vehicles to navigate effectively and make well-informed decisions even in the face of complex and ever-changing water conditions. This approach, known as sensor fusion, harnesses the unique strengths of each sensor type while mitigating their individual limitations. By doing so, it ensures a higher degree of reliability and robust performance in diverse surface water conditions. However, implementing sensor fusion has its complexities. Challenges arise due to differences in data formats, the need for calibration, and synchronization requirements between LiDAR and camera systems. These issues demand advanced data processing techniques and precise alignment of sensor outputs to achieve successful fusion. In the proposed architecture, a 3D LiDAR sensor is integrated with a stereo camera system for surface water traversability mapping. To guarantee the integrity of data, techniques similar to those presented in [Ropertz 17] are incorporated. The initial stage involves data filtering, building on the foundations laid by previous research [Keen 20]. This is followed by the deployment of various deep learning architectures to execute object detection and segmentation from camera imagery. The fusion of both object detection and segmentation significantly enhances the reliability and accuracy

of obstacle detection. These detected obstacles are then precisely superimposed onto a LiDAR-generated point cloud, facilitating the determination of their three-dimensional coordinates. Subsequently, this comprehensive 3D obstacle map, rendered as a point cloud, is conveyed to the next module for advanced processing and informed decision-making.

### 3.1.3 Fused Mapping

#### Challenges

The post-flood environment often presents many challenges characterized by random and unstructured obstacles. These obstacles encompass a wide range of objects, including submerged cars, fallen trees, and assorted debris. This scenario is particularly complex because these obstacles can exhibit static and dynamic behaviors. Static obstacles remain fixed, while dynamic obstacles may move due to water flow. The presence of dynamic obstacles introduces a dynamic element into the environment. Consequently, the fused map representing the post-flood landscape must evolve with time as these obstacles shift and relocate in response to the water's movement. This dynamic aspect must be considered when assessing the navigability and safety of the environment. Additionally, there are several situations encountered where relying solely on surface water traversability information proves insufficient for understanding the obstacles in the area. For instance, small floating obstacles or thin vegetation may not pose significant challenges to a surface water vehicle regarding navigation. However, relying solely on surface water information may not provide an accurate picture of the surroundings.

#### Proposed Concept

To address the complexities associated with navigating and operating effectively in post-flood environments, a fusion strategy is proposed in this technical thesis. This strategy involves the integration of information from two critical data sources: surface water obstacle maps and underwater obstacle maps. By combining these maps, the goal is to achieve a more precise, comprehensive, and robust understanding of the environmental obstacles that can impact the safety and functionality of surface water vehicles.

The fusion strategy begins by categorizing the obstacles into two main factors:

**Floating or Sinking** This categorization helps differentiate between objects that remain on the water's surface and those that have submerged underwater.

**Threat Level** This factor differentiates between obstacles based on the level of threat each obstacle presents on the safe navigation of boat. For instance, floating leaves are safer for navigation as compared to a standing tree.

To ensure a comprehensive understanding of the environment, the fused map is transformed into global coordinates using GPS/GNSS (Global Positioning System/Global Navigation Satellite System) data. This transformation provides georeferenced information, enabling precise navigation and location-based decision-making.

The primary output of this architecture is a safe traversability map, which serves as a valuable resource for surface water vehicles. This map not only identifies the presence of

obstacles but also assesses the threat level they pose based on their surface and underwater footprints. This threat level information is crucial for determining the navigability and safety of specific routes and areas.

To enhance the reusability and facilitate fast transmission of the output map, it is computed as a multi-class color coded map. In this representation, each 2D location within the map corresponds to an object, and the color code value at that location indicates the threat level posed by the object. This traversability map format optimizes data storage, transmission, and processing while retaining essential information about the environment's safety and obstacles.



## 4. Datasets and Test Environments

The importance of datasets and test environments in robotics cannot be overstated. In order to develop and evaluate robotic systems, it is crucial to have access to such datasets and test environments that accurately reflect the real-world scenarios in which the robots operate. Test environments provide a controlled setting in which researchers can systematically evaluate the performance of their robotic systems. Standard datasets, on the other hand, enable researchers to train and test machine learning algorithms that are used in robotic systems.

Developing robust robotic systems required rigorous testing across both realistic and simulated environments. In the specific domain of post-flood environments, conducting extensive real-world testing is inherently challenging. This difficulty arises from various factors, including restricted access to destructed areas, potential interference with ongoing rescue operations, and the substantial financial risks associated with unpredictable conditions in flood-affected regions. Consequently, the shortage of available datasets directly linked to post-flood scenarios increases the challenge. While datasets related to shallow water bodies like rivers or ports may facilitate testing in simpler scenarios, they often lack the complexity inherent in post-flood environments. This creates a paradoxical situation wherein systems must be designed to operate effectively in post-flood conditions despite the impracticality of evaluating them in these real-world scenarios.

Simulated environments emerge as a pivotal solution to address the limitations of real-world testing in post-flood scenarios. Simulations provide a controlled and reproducible setting for testing robotic systems under a spectrum of conditions that imitate the challenges posed by post-flood environments. However, for simulations to be effective, they must strike a delicate balance between reality and controllability. The simulated environment should be sufficiently realistic to capture the complexities of post-flood scenarios while remaining manageable for testing various system functionalities.

Within the framework of Shallow Water Traversability Mapping (SWiM) architecture, this chapter emphasizes the significance of test environments, particularly those that are simulated, and the critical role played by datasets in advancing the field of robotics. The objective is to examine the complexities associated with the development and utilization of these environments, with a particular focus on the challenges posed by post-flood scenarios.

## 4.1 Natural Waterbodies

The absence of datasets directly associated with post-flood environments introduces a significant challenge in the comprehensive evaluation of robotic systems tailored for such scenarios. The lack of relevant datasets restricts the ability to thoroughly test and refine these systems under conditions that mirror the complexities of post-flood environments. Consequently, there is an urgent and compelling need for datasets that can serve as valuable resources in enhancing the accuracy, adaptability, and robustness of robotic systems designed for post-flood scenarios.

Existing datasets, although not explicitly focused on post-flood environments, provide a foundation for evaluating certain aspects of robotic functionalities. Datasets captured in natural water bodies such as rivers, ports, or urban waterways, offer valuable insights, particularly in assessing simpler scenarios. These scenarios may include the detection of structured surface water objects, semantic segmentation of a scene, or the detection of controlled underwater objects. While these datasets contribute to the evaluation of specific functionalities, their applicability to the complexities of post-flood environments is limited.

In the subsequent section, a comprehensive overview of state-of-the-art datasets is presented. These datasets, although not originally designed for post-flood scenarios, have been extensively utilized in this work to evaluate and refine various aspects of the robotic system. Their utility lies in their capacity to facilitate the assessment of fundamental functionalities such as object detection, segmentation, and scene understanding. However, the limitations of these datasets become apparent when attempting to replicate the dynamic and unstructured nature of post-flood environments.

To address the gaps in existing datasets and accommodate the specific requirements of post-flood scenarios, a novel dataset is introduced. This dataset is thoroughly annotated to capture the complexities of post-flood environments, offering a unique and tailored resource for evaluating robotic systems in these challenging conditions. The development of this dataset is based on a critical analysis of existing datasets, identifying the limitations and areas where improvements are necessary to align with the complexities of post-flood scenarios.

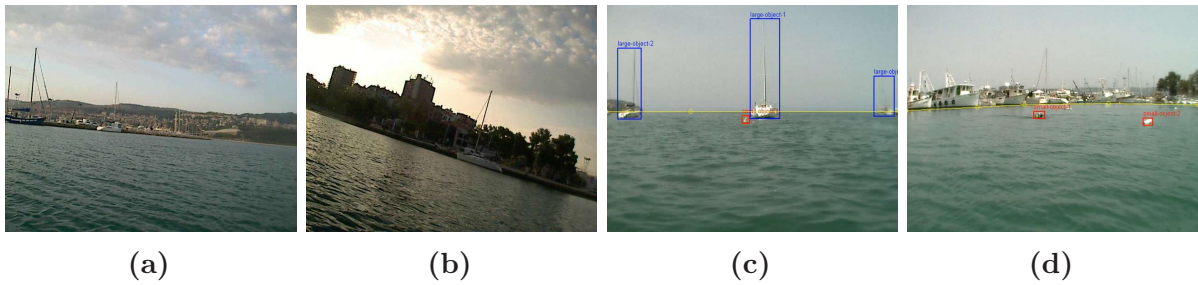
### 4.1.1 State-of-the-art Datasets

In this section, state-of-the-art surface and underwater datasets are presented that have been comprehensively used to train and test SWiM architecture. These datasets are discussed below:

**MODD / MODD2 Dataset:** Kristan et al. [Kristan 15] present a Marine Obstacle Detection Dataset (MODD) that comprises a collection of 12 video sequences represented as frames. The dataset stands out for its annotations, providing a total of 4454 fully annotated frames, each with a resolution of  $640 \times 480$  pixels. Similarly, Bovcon et. al. [Bovcon 18] have presented an extension of MODD dataset named as MODD2 that has 28 time-synchronized sequences (11675 frames at resolution of  $1278 \times 958$  pixels) gathered using on board stereo camera, IMU and GPS. This dataset is considered as the largest benchmark for evaluating obstacle detection algorithms in aquatic environments.

The data collection process involved the utilization of a small unmanned surface vehicle (USV) with a length of approximately 2.2 meters. The USV was equipped with an Axis-207W<sup>1</sup> camera positioned at a height of approximately 0.7 meters above the water surface. The camera had a field of view of 55°, capturing a comprehensive view of the surrounding aquatic environment. The USV traversed the test site, situated at the port of Koper, Slovenia, at a maximum speed of 2.5 m/s. The datasets were carefully gathered across different periods of the day and diverse weather conditions, reflecting the variability inherent in real-world scenarios.

The sequences within these datasets include a variety of environmental challenges. For instance, in MODD dataset, the first ten sequences depict images captured under normal weather conditions, providing a foundational benchmark for obstacle detection algorithms. In contrast, the last two sequences deliberately introduce extreme environmental challenges, with the sun positioned directly facing the vehicle, as illustrated in Figure 4.1. This deliberate variation in environmental conditions ensures that the dataset covers a broad spectrum of challenges, allowing researchers to evaluate the robustness and generalization capabilities of obstacle detection algorithms under diverse scenarios.



**Figure 4.1:** Example Data from MODD Dataset. (a-b) Diverse dataset captured at different light conditioning in the coastal waters of Koper, Slovenia for MODD dataset (c-d) Annotations include large object, small object, and sea edge (boundary between the sea and the horizon) [Kristan 15].

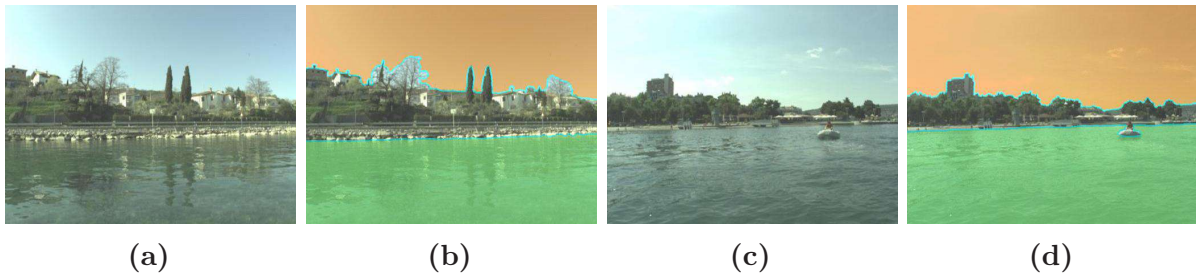
An invaluable addition to these datasets is the inclusion of sea edge annotations, denoting the boundary between the sea and the horizon. This supplementary annotation enriches the datasets' utility by providing a reference point for detecting obstacles floating specifically over the water surface. By including the sea edge annotations, MODD/MODD2 offer more distinct and context-aware datasets, allowing researchers to focus specifically on obstacles within the aquatic domain.

**MaStr1325:** MaStr1325 [Bovcon 19] represents a large-scale marine semantic segmentation dataset, specifically designed to facilitate the advancement of segmentation methods tailored for deployment on small-sized coastal Unmanned Surface Vehicles (USVs). This dataset consist of three core components: Images, Ground Truth Masks, and Inertial Measurement Unit (IMU) data, collectively presenting a comprehensive environment for algorithmic development and evaluation. Comprising a total of 1325 diverse images, MaStr1325 is a product of complex data collection efforts spanning

<sup>1</sup><https://www.axis.com/products/axis-207w/support>

a two-year duration, during which a USV was manually navigated in the Gulf of Koper, Slovenia.

Each image within the MaSTr1325 dataset has undergone manual annotation by human annotators. Figure 4.2 shows the annotated masks that enrich the dataset with semantic segmentation information. The annotations contain three distinct classes: sea, sky, and environment. This hierarchical segmentation provides a granular understanding of the visual scene, allowing algorithms to perceive between critical elements such as water surfaces, the sky, and the broader environment. The manual annotation process ensures high-quality ground truth data, serving as a reliable benchmark for evaluating the performance of segmentation algorithms.



**Figure 4.2:** Example Data from Mastr1325 Dataset. (a, c) Diverse image dataset captured in the coastal waters of Koper, Slovenia for Mastr1325 Dataset (b, d) Mask classes include obstacle, water, sky, and unknown [Bovcon 19].

**ROSEBUD:** The dataset presented by Lambert et al. [Lambert 22] contains an imagery captured along Sugar Creek and the Wabash River in the state of Indiana, USA. The dataset is derived from video recordings, with 249 distinct image frames extracted from over 2 hours of footage recorded on the Wabash River on July 15, 2021, and 300 images obtained from over 3 hours of video recorded on Sugar Creek on September 30, 2021. The collective dataset, denoted as ROSEBUD, provides a total of 4392 images, each available in high resolutions of 1920 x 1440 and 512 x 384 pixels. The set includes 549 original images, along with binary classification masks, seven-classes segmentation masks, and ground-truth image labels for each image.

The video recordings were conducted at a consistent rate of 30 frames per second (FPS) using a GoPro Hero4 camera mounted approximately half a meter above the waterline. The Wabash River images were captured using an Unmanned Surface Vehicle (USV) developed at the Mahmoudian lab’s BREAM, while Sugar Creek images were acquired from a Canoe operated by a human pilot. The incorporation of both USV and human-operated boat data adds an extra layer of diversity to the dataset, considering the potential variations in data acquisition approaches.

For thorough annotation, 549 images were manual labeled by human annotators. The annotation process accounts for seven distinct fluvial classes: (1) water, (2) exposed river shore/bank, (3) bridge, (4) boat, (5) flora, (6) debris (encompassing logs, trash, rocks), and (7) sky. This annotation process provides a detailed understanding of the environmental context, catering to diverse applications in remote sensing, environmental monitoring, and machine learning algorithm development.

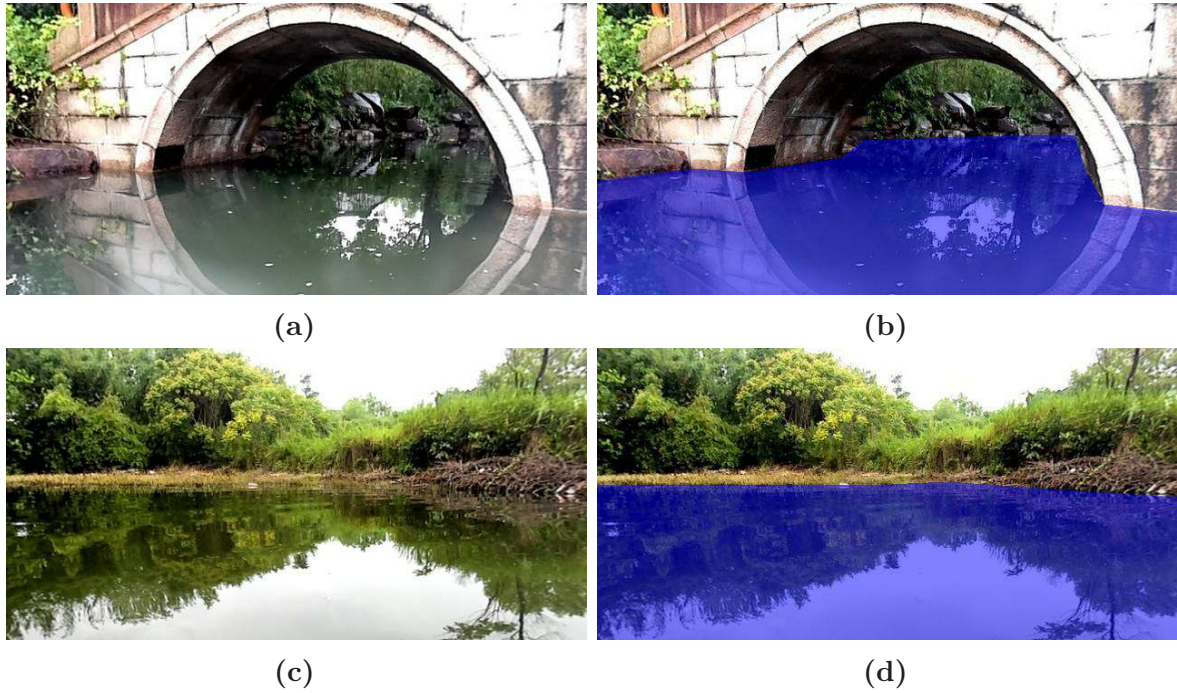


**Figure 4.3:** Example Data from Rosebud Dataset. (a) Images captured in a shallow lake (b) Segmentation Mask [Lambert 22].

The distinct origin of images from both a USV and a manually operated canoe introduces variability in navigational perspectives, water conditions, and environmental contexts, enhancing the dataset’s generality. Such diversity is critical for training and evaluating algorithms that can generalize across different water bodies, boat types, and operational scenarios. Furthermore, the provision of images at two resolutions (1920 x 1440 and 512 x 384 pixels) caters to the diverse computational requirements of researchers and practitioners, allowing for flexibility in dataset utilization based on specific application needs and resource constraints.

**USVInland Dataset:** The dataset presented by Cheng et al. [Cheng 21] contains a diverse range of sensor data, incorporating simultaneous localization and mapping (SLAM)/odometry, stereo matching, groundtruth data derived from LiDAR, and water segmentation data from camera. This comprehensive collection is derived from real-world scenes of urban waterways, capturing various waterway landscapes and weather conditions at different times. The SLAM/odometry sub-dataset offers enriched information, including the time of data collection, the duration of data acquisition, trajectory length, and prevailing weather conditions for each sequence.

The dataset integrates an array of sensors, contributing to its richness and versatility. A LiDAR sensor with a wavelength of 905 nm, 16 beams, a horizontal field of view (HFOV) of 360°, and a vertical field of view (VFOV) of 32° operates at 10 Hz, with a maximum range of 100 m, a range resolution of 2 cm, and a horizontal angular resolution of 0.18°. A stereo camera, synchronized with a built-in Inertial Measurement Unit (IMU), captures RGB images at resolutions of 640 × 400 or 1280 × 800 at a rate of 20 Hz. The camera has a horizontal fov of 95°, a vertical fov of 50°, and a diagonal fov of 112°. The IMU, operating at 200 Hz, is 6-axis. Three radar sensors, each operating at 10 Hz, provide a horizontal fov of 94°, a vertical fov of 50° (-6 dB), a range resolution of 0.045 m, a maximum unambiguous range of 18.08 m, a maximum radial velocity of 3.96 m/s, and a radial velocity resolution of 0.25 m/s. An Inertial Navigation System (INS) with a 50 Hz sampling rate and two GPS receivers (U-Blox ZED-F9, 5 Hz) with RTK signal reception capabilities



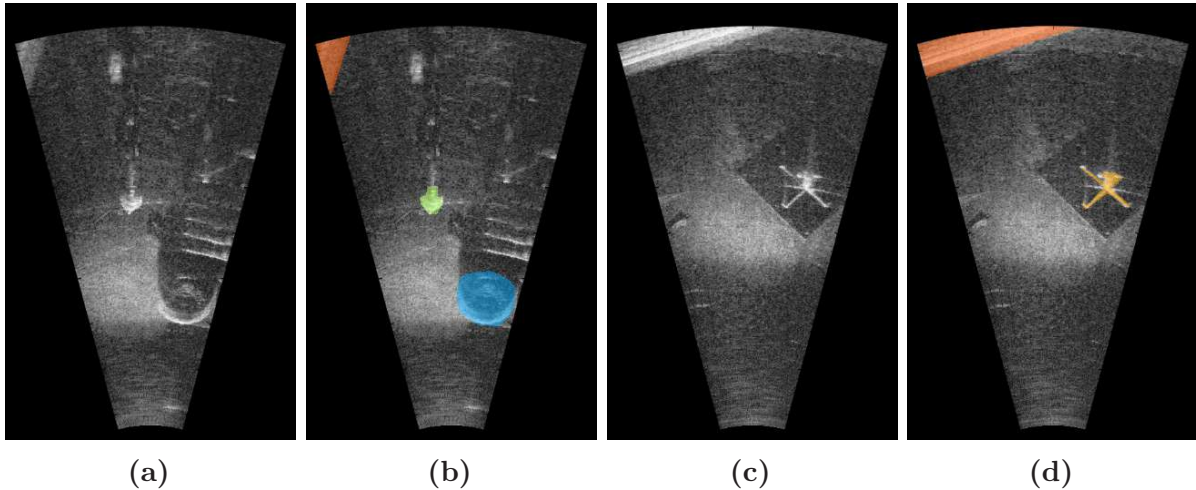
**Figure 4.4:** Example Data from USVInland Dataset. (a, c) Images captured in an urban waterway (b, d) Mask class include water [Cheng 21].

enhance the dataset’s geospatial accuracy.

The stereo matching sub-dataset comprises 324 image pairs, providing valuable data for evaluating stereo vision algorithms. The groundtruth data, generated from LiDAR point clouds, enhances the dataset’s utility for precision applications, enabling validation against ground truth for sensor calibration and algorithmic refinement. In the water segmentation sub-dataset, 518 relatively low-resolution images ( $640 \times 320$ ) and 182 relatively high-resolution images ( $1280 \times 640$ ) are accompanied by annotation data. Annotated images in the water segmentation sub-dataset enable the discrimination between water reflections and physical objects. This annotation data serves as a valuable resource for training and evaluating algorithms aimed at enhancing the accuracy of water segmentation, a prerequisite for reliable aquatic sensing and navigation.

**Marine Debris Dataset:** The Marine Debris Dataset introduced by Singh and Valdenegro-Toro [Singh 21] addresses a critical gap in underwater analysis by focusing on semantic segmentation using forward-looking sonar (FLS) data. This dataset is a valuable contribution to the field as it establishes a baseline for creating and testing algorithms designed to detect and classify objects in sonar images.

The dataset is generated through the deployment of an ARIS Explorer 3000 forward-looking sonar attached to an autonomous underwater vehicle (AUV) within an artificial water tank with dimensions  $(W, H, D) = (3 \times 2 \times 4) m^3$ . The ARIS Explorer 3000 sonar operates at a high frequency of 3 MHz, featuring a set distance range of 0.7-5 meters. It incorporates 128 acoustic beams, offering a field of view of  $30^\circ \times 15^\circ$  and a spatial resolution that varies from 2.3 mm per pixel in proximity to 10 mm per pixel in the distant range, thereby enabling a diligent representation



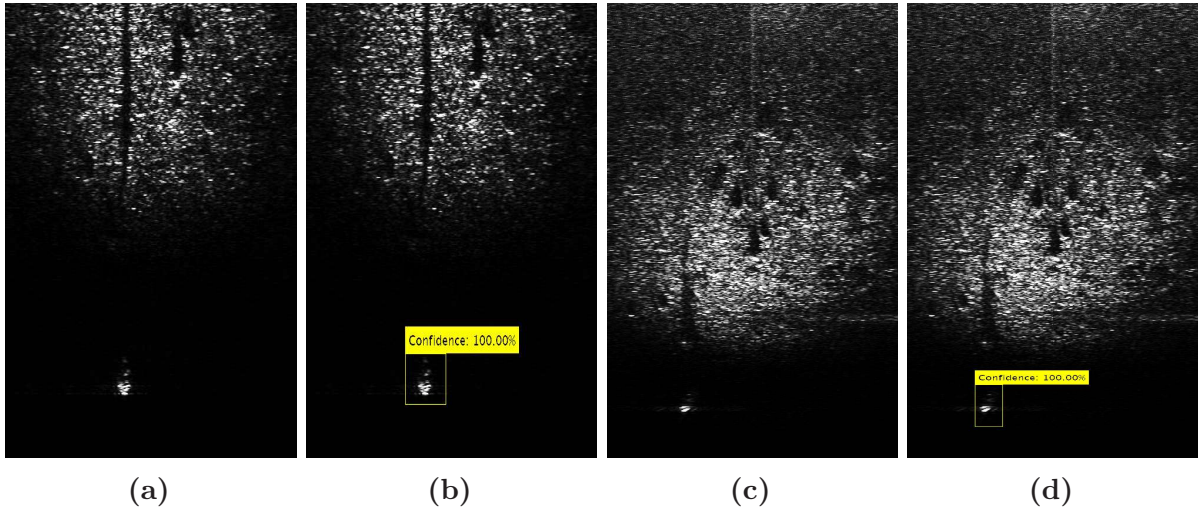
**Figure 4.5:** Example Data from Marine Debris Dataset. (a, c) Sonar imagery captured in a water tank (b, d) Mask class include Background, Bottle, Can, Chain, Drink-carton, Hook, Propeller, Shampoo-bottle, Standing-bottle, Tire, Valve, and Wall [Singh 21].

of the underwater environment. To avoid temporal correlation between frames, a strategic approach involves ignoring five frames between any two consecutive frames. This methodology ensures the dataset’s integrity, providing a robust foundation for further research and algorithmic development in the domain of underwater robotics and sonar-based image analysis.

The dataset includes a variety of objects, both typical maritime and domestic items, creating a realistic scenario for data collection within the artificial tank. Comprising annotated forward-looking sonar imagery (1868 images in total) acquired in diverse underwater conditions, this dataset encompasses eleven object types, including plastic bags, bottles, and submerged items, making it a comprehensive resource for semantic segmentation tasks. The dataset is divided into training and testing sets, with each image featuring detailed annotations. Environmental factors such as varying water depths, turbidity levels, and lighting conditions are considered, providing a diverse and realistic dataset for algorithm evaluation.

**UATD Dataset:** Xie et al. [Xie 22] introduced a novel dataset focused on advancing underwater object detection, featuring multibeam forward-looking sonar (FLS) imagery. The dataset, encompassing 10,000 sonar images, was collected from diverse locations and depths, each image precisely annotated with bounding boxes and corresponding labels. The annotated dataset, comprising 9200 images, was categorized into training (7600 images), test1 (800 images), and test2 (800 images). A noteworthy distinction is made between the marine debris dataset, portraying FLS images in a sector representation, and the UATD dataset, which was processed to eliminate exterior black regions around the sonar part, enhancing uniformity but potentially compromising human readability. The data collection, initiated in 2020 at Maoming and Dalian in China, occurred in a natural setting, particularly in a shallow water lake.

The data acquisition utilized a Tritech Gemini 1200ik sonar, a multibeam forward-looking sonar with dual frequencies of 720 Hz and 1200 Hz for long and short ranges,



**Figure 4.6:** Example Data from UATD. (a, c) Sonar imagery captured in a shallow water lake (b, d) Annotated obstacles. [Xie 22].

respectively. A specialized mechanical structure was devised to affix the sonar to the boat, enabling image capture at varying depths and tilt angles. This depth and tilt variation strategy significantly expanded the dataset size by accounting for the impact of slight changes in sonar or object positions. The dataset features ten distinct object types, including cylinders, tires, cages, human body models, and others.

#### 4.1.2 MASTER Dataset:

The standard datasets discussed above do not fulfill the multi-modal requirement of SWiM architecture due to their narrow focus on either surface or underwater environments. To ensure a comprehensive evaluation of the multi-modal SWiM architecture, a novel dataset termed as Multi-modal Shallow waTER (MASTER) has been compiled from a variety of aquatic environments, containing ports, lakes, and shallow areas of urban rivers. This extensive dataset represents temporal data sequences, spanning between 2019 to 2021. Its uniqueness lies in the incorporation of multi-sensor and multi-modal data, offering a comprehensive understanding of the water environment. The array of sensors integrated into the dataset comprises stereo cameras, LiDAR, GPS/GNSS, and SONAR, collectively providing a rich and dynamic dataset capturing various aspects of the surface and underwater environment.

Recorded across multiple surface water vehicles, each characterized by distinct dimensions and motion models, the MASTER dataset stands out for its versatility. These boats, featured in the dataset, demonstrate diverse geometries and navigational behaviors, ensuring the dataset’s applicability across a range of scenarios. For detailed information regarding the individual surface water vehicles used in data acquisition, the reader is directed to the corresponding content in the appendix B.

**Gelterswoog Lake:** The dataset was collected using BASILISK and ASTRIDER from the beginning of 2021 to mid-2023 in a shallow water lake near RPTU Kaiserslautern



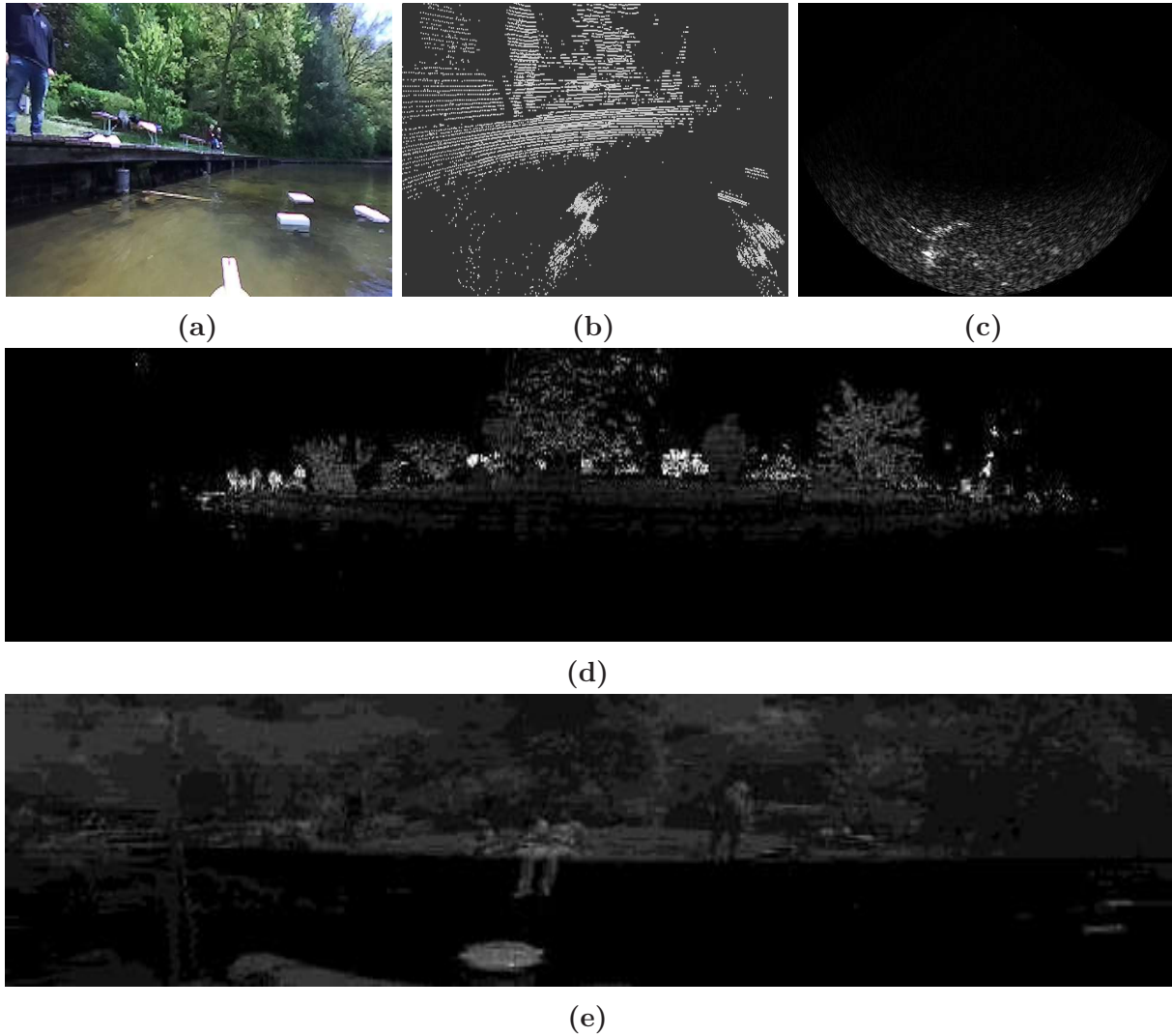
Campus. The data was recorded in various weather conditions, and the depth of the lake varied depending on the temperature. During the cold or rainy seasons, the depth of the lake ranged from 30 cm at the corners to 4 meters, while in the hot season, it was between 10cm to 2.5 meters. The lake has a few obstacles, such as water volleyball goal posts, concrete structures, floating balls, and a docking area. Additionally, to analyze underwater sonar functioning and the noise in its data, several known obstacles were thrown into the water. The lake is situated at global coordinates of 49.394278, 7.694304.

**THW test site:** Technisches Hilfswerk (THW) is Germany’s federal civil and disaster protection organization. The dataset was recorded in the first two weeks of November 2021 at the test site (urban river) of THW near Hoya. This is a river with very low water current and average depth approximately 5 meters. The dataset was recorded on BASILISK and PONTONBOOT. The test site is situated at global coordinates of 52.8268684, and 9.1992187.

**Hamburg Port:** As a collaborative partner of the project, the port authority has provided a test site at Hamburg Port. The dataset was recorded in the last two weeks of November 2021 using BASILISK, during an exhibition of the presented system. It was approximately 10 meters deep waterbody with very high water current. The port’s test site is situated at global coordinates of 53.506565, 9.968251.

This thesis contributes in the field of benchmarks by introducing a novel and comprehensive multi-modal shallow water dataset recorded across distinct surface vehicles, varying in size and motion model. Here, the shallow water means the water depth less than 10 meters and more importantly less than 2 meters. Figure 4.7 provides the sensor data, explaining the complex details of the dataset’s composition. The dataset encompasses a total of 1200 images and point cloud frames, each annotated with bounding boxes and segmentation masks. The inclusion of bounding boxes is particularly noteworthy, providing a labeled representation of obstacles within the dataset. It’s essential to highlight that the bounding boxes exclusively refer to a single class of *obstacles*, a strategic choice that streamlines the dataset’s focus for specific applications, such as traversability mapping and obstacle detection. The segmentation masks further enhance the dataset’s utility by offering semantic annotations, providing a more granular understanding of object boundaries and shapes. Figure 4.8 shows the annotated mask from MASTER dataset. Following are the salient features of the dataset:

1. *Temporal Diversity:* Recorded at different times of the day and across various weather conditions throughout the entire years of 2019-2021, the dataset encapsulates the dynamic nature of real-world scenarios.
2. *Multi-Modal:* The dataset comprises rich data modalities of both surface and underwater environments. The dataset presents surface water information from stereo camera, 3D LiDAR, and GPS/GNSS, whereas underwater information from multibeam forward-looking sonar, ensuring a multi-modal representation of the environment.
3. *Multi Waterbodies:* The dataset’s environmental diversity is emphasized by recordings in a variety of waterbodies, including lakes, ports, and rivers. To further enhance its



**Figure 4.7:** Example data from MASTER Dataset. (a) Camera Image, (b) Pointcloud, (c) Sonar Imagery (d) Reflectivity image, and (e) Noise image from LiDAR.

diversity, the dataset captures varying water depths ranging from 50 centimeters to 10 meters, presenting an understanding of different aquatic environments.

4. *Multi Boat*: The dataset was recorded on three different boats having distinct motion model and kinematics.

This contribution significantly addresses the gap in benchmark datasets specifically tailored for shallow waterbodies, offering a resource that aligns with real-world challenges encountered in aquatic environments. The recording across various parameters enhances the dataset’s relevance for a wide range of applications. Researchers can leverage this dataset to advance algorithms and models in areas such as perception, navigation, and object recognition within the context of surface water vehicles. The dataset’s open availability and comprehensive documentation further enhances collaborative research efforts, empowering the scientific community to collectively push the boundaries of knowledge in the domain of shallow waterbodies exploration.



**Figure 4.8:** Example data from MASTER Dataset. (a) Camera Image, (b) Segmentation Mask.

## 4.2 Simulation

The natural environment proves valuable for validating the proof of concept, allowing the testing of simple scenarios related to the detection and mapping of random objects. However, these environments fall short when it comes to assessing the complexities of post-flood scenarios. To overcome this limitation, a comprehensive simulation has been developed to facilitate extensive testing of post-flood scenarios. This simulation is designed within Unreal Engine, a sophisticated physics-based simulator renowned for its application in developing highly graphical games and simulated environments.

The simulation in Unreal Engine features the replication of few realistic environments and three distinct surface vehicles (SVs), namely ASTRIDER, BASILISK, and PONTONBOOT. Each of these vehicles is equipped with a multi-sensor system, mirroring the configurations of their real counterparts (provided in Appendix B) used in practical testing. Unreal Engine, being a sophisticated and physics-based simulator, provides an ideal platform for creating a realistic virtual environment that closely mimics the challenges and complexities encountered in post-flood scenarios.

The integration of Unreal Engine with the SWiM architecture is an essential aspect of this simulation. To establish this connection, custom APIs within the FINROC framework have been developed in-house. These APIs serve as a bridge, facilitating seamless communication between the SWiM architecture and Unreal Engine. This integration allows for real-time data exchange, enabling the SWiM architecture to control and receive feedback from the simulated SVs within Unreal Engine. Such a connection ensures that the behavior of the SVs and the performance of the SWiM architecture are rigorously evaluated in a controlled yet realistic setting.

### 4.2.1 Post-flood Water

The simulation of post-flood environments in this research is designed and implemented using the Unreal Engine’s water plugin as shown in Figure 4.9. Using this powerful plugin, the environment’s water parameters can be tuned to meet the specific requirements of the study. While the default water provided by the Unreal plugin is clear and moving, the focus of this work lies in replicating the conditions of still post-flood water. In post-flood scenarios, water movement is often minimal, and the environment is suffered by the destruction due flood. Therefore, adjustments are performed to the water parameters,

particularly in terms of turbidity and color, to accurately simulate the conditions of a post-flood environment.



**Figure 4.9:** Water Plugin - Unreal Engine. The image shows water placed within the simulation using *oceanwaterbody* blueprint.

One critical aspect of post-flood waters is the presence of high turbidity, caused by the influx of sediment and debris. In the Unreal Engine simulation, this turbidity level was intentionally increased to mirror the challenging conditions of navigating through turbid post-flood waters. The introduction of increased turbidity as shown in Figure 4.10 serves the purpose of simulating the real-world scenario where the water's clarity is compromised, posing challenges to sensors such as cameras, LiDAR and sonar mounted on surface vehicles (SVs). By replicating this environmental characteristic, the simulation provides a realistic testing ground for SVs, allowing for the evaluation of their ability to navigate effectively in challenging post-flood conditions.

The generation of the dataset within the Unreal Engine simulation is a complex process aimed at capturing the diversity of post-flood conditions. Different water depth levels are incorporated into the simulation to emulate the varying depths commonly encountered in post-flood environments. This deliberate variation in water depth adds a layer of complexity to the dataset, allowing for a comprehensive evaluation of the SV's performance across different depths. The scenarios generated at different water depths contribute to a more diverse dataset, enhancing the adaptability and versatility of the SV's navigation capabilities.

In essence, the Unreal Engine simulation, augmented by the water plugin, serves as a sophisticated platform for replicating the complex characteristics of post-flood environments. The intentional adjustment of turbidity and water color, coupled with the incorporation of diverse water depths, ensures that the simulation captures the complexities of navigating through post-flood waters. The generated dataset becomes a valuable resource for training and evaluating SVs, offering a realistic and controlled environment for systematically testing their capabilities. Through this simulation-based approach, the study achieves a high level of fidelity in replicating post-flood conditions, contributing to the robust development and assessment of SVs for effective post-flood rescue operations.



**Figure 4.10:** Representation of Visibility in a Simulated Underwater Environment. This figure contrasts two conditions in a simulated underwater setting: (a) depicts the environment without any sediment, offering clear visibility of an obstruction, and (b) shows the same scene with sand sediments, demonstrating the reduced visibility and the challenge it poses for object detection.

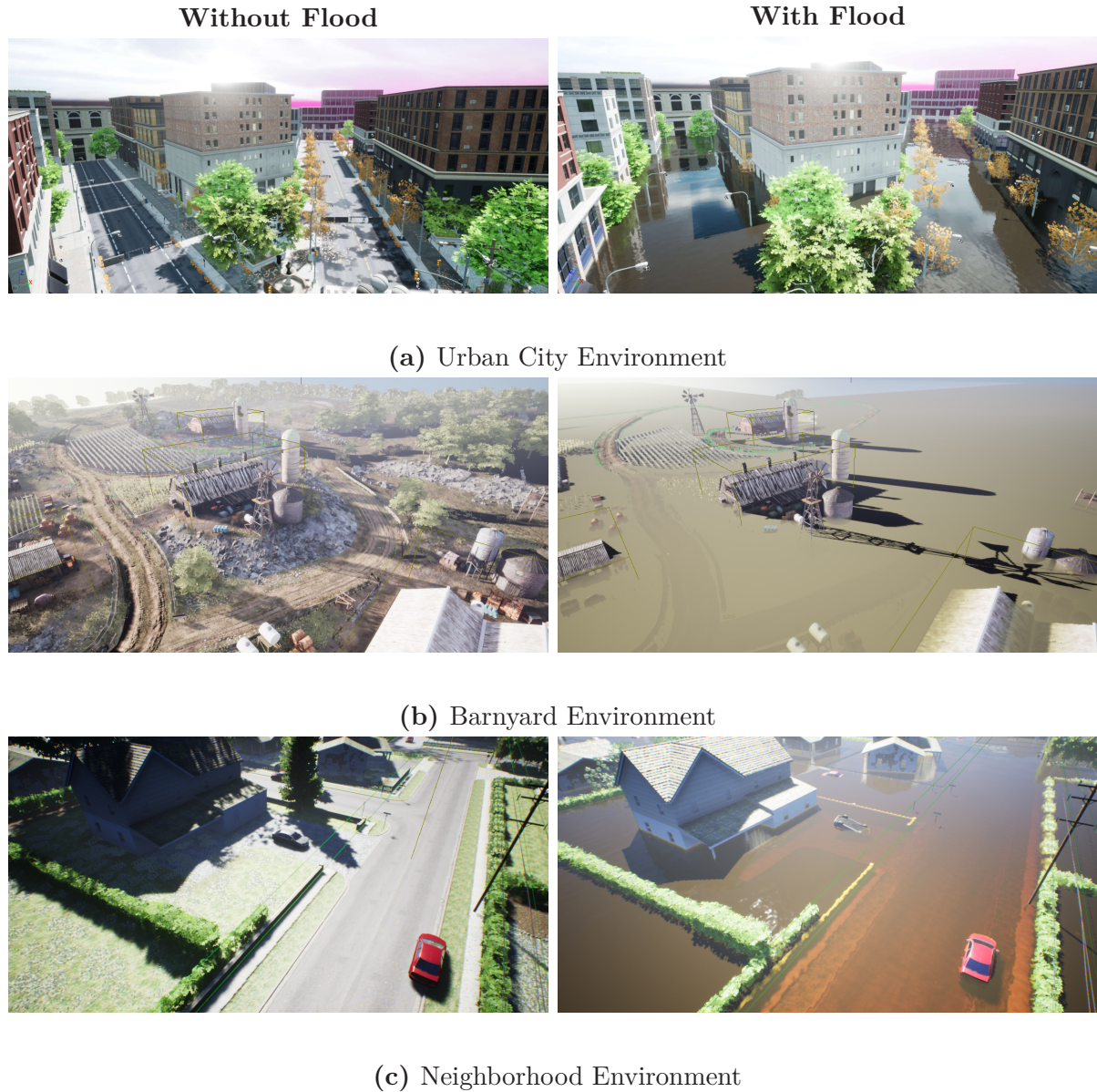
### 4.2.2 Flooded Environments

Flooding is a natural disaster that impacts diverse environments, ranging from remote and rural areas to densely populated urban cities. However, each environment presents distinct challenges, and the nature of these challenges significantly varies between remote and urban settings. In remote areas, the effects of flooding manifest in vast fields, potentially impacting animals, and the landscape is generally less congested. Conversely, urban city environments pose a unique set of challenges, characterized by high population density, congested buildings, and a multitude of random objects such as vehicles, poles, and garbage cans strewn across the landscape. Traversability is notably equally challenging in any flooded environment as each environment poses different nature of challenges.

Therefore, in the context of this research, several unreal environments are used ranging from barnyard to city areas (shown in Figure 4.11). The objective is to develop and test surface vehicles (SVs) capabilities in scenarios that closely resemble the challenges encountered in real post-flood environments. Hence, various scenarios have been precisely crafted, aiming to replicate issues commonly observed in actual post-flood situations. These scenarios include submerged cars, floating wooden planks, fallen trees, and scattered debris, among others. Each scenario is designed to capture the diversity of challenges faced by SVs when navigating through flooded landscapes. The choice to simulate sinking cars, fallen trees, and other obstacles is informed by their frequent occurrence in post-flood scenarios and their significant impact on traversal and rescue operations.

Numerous simulation runs have been executed within the simulated environment, generating a comprehensive and highly realistic dataset representative of post-flood conditions. The simulation dataset serves as a valuable resource for evaluating and refining the capabilities of SVs in response to the challenges posed by flooded environments. The simulations provide a dynamic and controlled platform for systematically testing the SV's ability to detect, navigate around, and respond to various obstacles encountered in the aftermath of a flood.

The simulated scenarios not only replicate physical obstacles but also account for environmental factors such as varying water depths, turbidity levels, and lighting conditions. These factors introduce additional layers of complexity, mirroring the unpredictable and dynamic nature of real post-flood environments. By incorporating these elements into the



**Figure 4.11:** Flooded environments simulated in Unreal Engine. The figure shows the environments of Unreal Engine that are flooded with turbid water. Each environment poses distinct challenges. Buoyancy component is added in several objects within this environment to show a realistic behavior due to water. In (c) a white car with buoyancy is visible.

simulation, the dataset becomes a rich and diverse repository that enhances the fidelity of the testing environment.

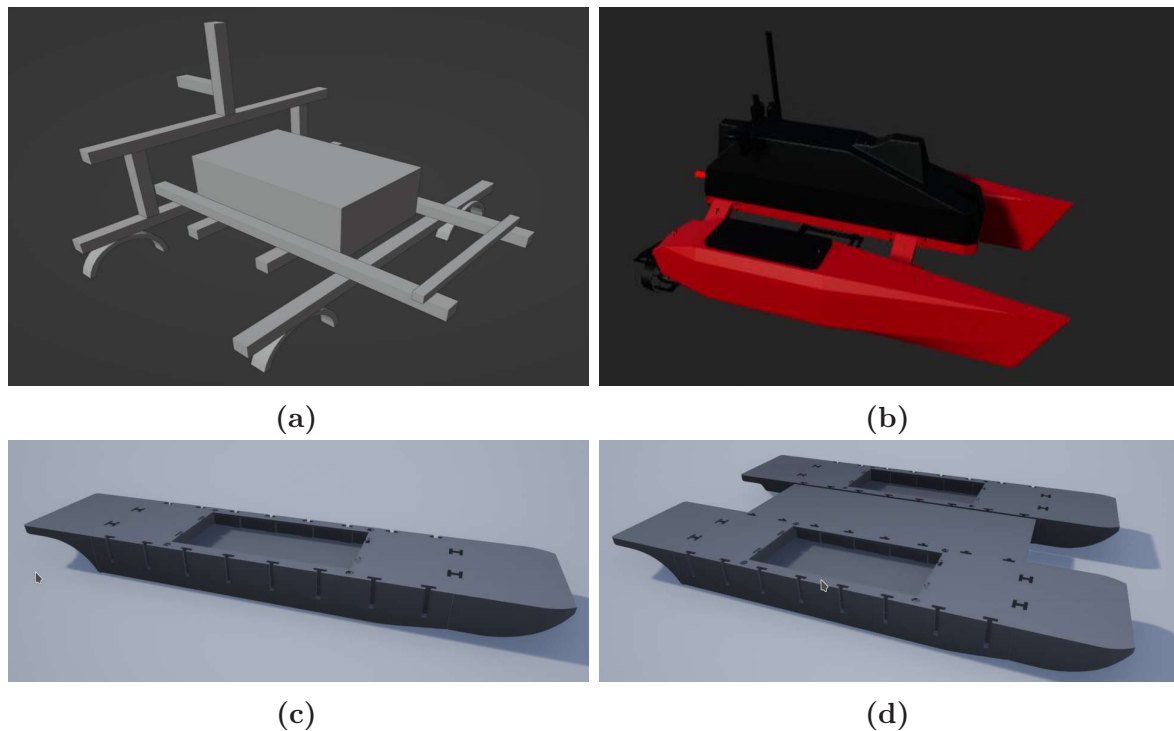
The dataset derived from these simulations serves a dual purpose. Firstly, it facilitates the training and validation of algorithms within the SV’s perception and decision-making systems. The SV’s sensors, including cameras, LiDAR, and sonar, are exposed to scenarios mimicking real-world post-flood challenges. This exposure enhances the adaptability and robustness of the SV’s sensing mechanisms, enabling it to effectively navigate through complex and unpredictable urban flood scenarios.

Secondly, the dataset becomes an invaluable resource for benchmarking and evaluating

the SV's performance. The dataset provides a baseline against which the SV's capabilities can be measured, ensuring that it meets the demanding requirements of post-flood rescue operations. The dataset's diversity contribute to a thorough evaluation, offering insights into the SV's strengths, limitations, and areas for improvement.

### 4.2.3 Surface Vehicles Models

In simulating the post-flood scenarios, the Unreal Engine proves to be a robust environment for designing and implementing highly detailed models of surface vehicles. Specifically, three distinct surface vehicles crucial to this study — the ASTRIDER, BASILISK, and PONTONBOOT — have been accurately modeled within the Unreal Engine framework. This modeling is an integral component of the research, contributing to the creation of a realistic and immersive simulated environment for testing and evaluating the performance of these surface vehicles in post-flood conditions.



**Figure 4.12:** Surface vehicle models designed in UNREAL Engine (a) ASTRIDER (b) BASILISK (c) PONTONBOOT Einzelboot Configuration. (d) PONTONBOOT Transporter Configuration.

The modeling process within Unreal Engine involves a comprehensive recreation of each surface vehicle's physical characteristics and structural components. For the ASTRIDER, BASILISK, and PONTONBOOT, the modeling includes the comprehensive details of their hulls, propulsion systems, sensor mounts, and any other components relevant to their functionality in post-flood environments.

One of the primary objectives in the design process is to capture the unique kinematics and features of each surface vehicle. The ASTRIDER, characterized by its agility and compact design, is represented in the Unreal Engine model. Similarly, the BASILISK, being a remote-controllable catamaran with distinct maneuverability, is intricately modeled to

replicate its real-world attributes. The PONTONBOOT, a modular and reconfigurable vessel, is digitally recreated to showcase its adaptability in post-flood scenarios. These models are not only visually accurate but are also embedded with the dynamic attributes that define the behavior of each surface vehicle in a simulated environment.

In Figure 4.12, the detailed models of the ASTRIDER, BASILISK, and PONTONBOOT are presented, offering a glimpse into the complexities of the design process. These visual representations serve as a reference for the integrity achieved in the digital recreation of the surface vehicles. The accuracy of the models is not limited to their external appearance; it extends to the inclusion of sensor systems, propulsion units, and any other components crucial to their functionality.

#### 4.2.4 Sensor System

The sensor system integral to the simulation environment encompasses a comprehensive suite, featuring LiDAR, stereo camera, GPS/GNSS, and a forward-looking sonar. Within the Unreal Engine framework, the LiDAR, camera, and GPS plugins have been developed in-house by FINROC developers, showcasing a tailored integration for this specific simulation. The upcoming sections will delve into the complexities sensor plugins, illustrating their functionalities and contributions to the simulated environment.

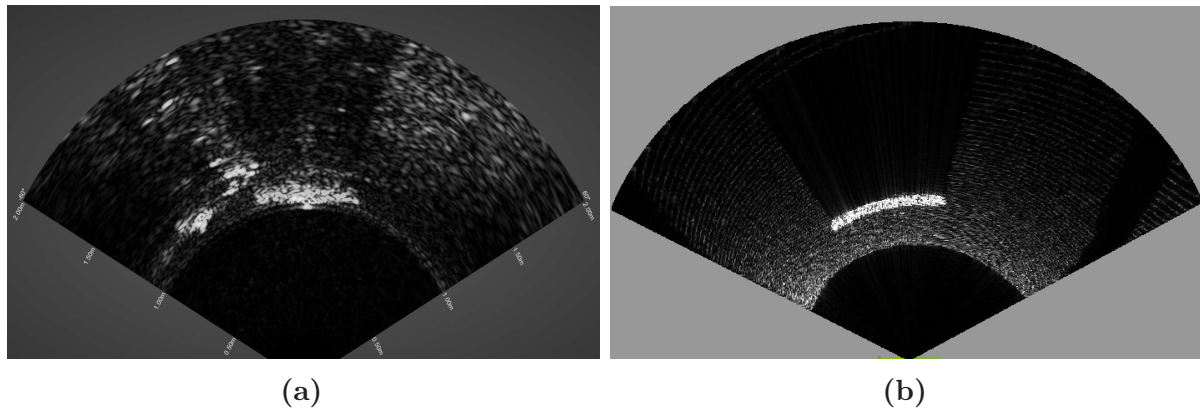
Of particular significance in this work is the introduction of a novel sonar plugin designed explicitly to simulate the forward-looking sonar utilized across all three surface vehicles. This sonar plugin represents an addition to the simulation toolkit, allowing for a detailed and realistic emulation of the sonar system's behavior. In subsequent sections, the discussion will delve deeply into the development and functionality of this sonar plugin, shedding light on its role in enhancing the reliability of the simulated post-flood environment. The existing sensor plugins including LiDAR, camera and localization sensors are discussed in appendix D. Sonar plugin, a novel contribution by this thesis, is explained in the upcoming section.

#### Sonar Plugin

The development of a sonar plugin within the Unreal Engine posed a substantial challenge due to the complex nature of modeling sound wave interactions with the environment. Despite the availability of pre-existing simulation models of forward-looking sonar, they exhibited unrealistic features, failing to accurately represent the complex noise within sonar imagery.

In a foundational work by DeMarco et al. [DeMarco 15], a basic sonar model for forward-looking sonar was simulated, employing Gazebo and ray tracing. While this model successfully simulated the shadow and speckle noise of sonar, the resulting object was unrealistically clear, lacking the complex noise patterns observed in real sonar imagery. It is revealed that the acoustic properties in this model were either hard-coded or omitted, lacking detailed speckle noise simulation. A subsequent effort by Cerqueira et al. [Cerqueira 17] introduced a GPU-based sonar simulation, encompassing mechanically scanned imaging sonar (MSIS) and forward-looking sonar (FLS). This model accurately represented the acoustic features and sound propagation characteristics, utilizing parameters such as pulse distance, echo intensity, and field-of-view to render the camera as a





**Figure 4.13:** Comparative Visualization of Sonar Data: (a) depicts a real sonar image acquired from underwater environment, characterized by a significant level of ambient noise, which is inherent to real-world underwater acoustic imaging. (b) presents a simulated sonar image, generated through computational modeling techniques as referenced in [Choi 21].

sonar device. The authors conducted physical tests to compare the simulation results with those obtained from a physical imaging sonar. The simulation achieved a high level of realism but fell short in addressing challenges arising from rather lower sonar frequencies (in kHz frequency). The authors primarily focused on very high frequencies (in the MHz range), resulting in simulated imagery that was sharp and high resolution.

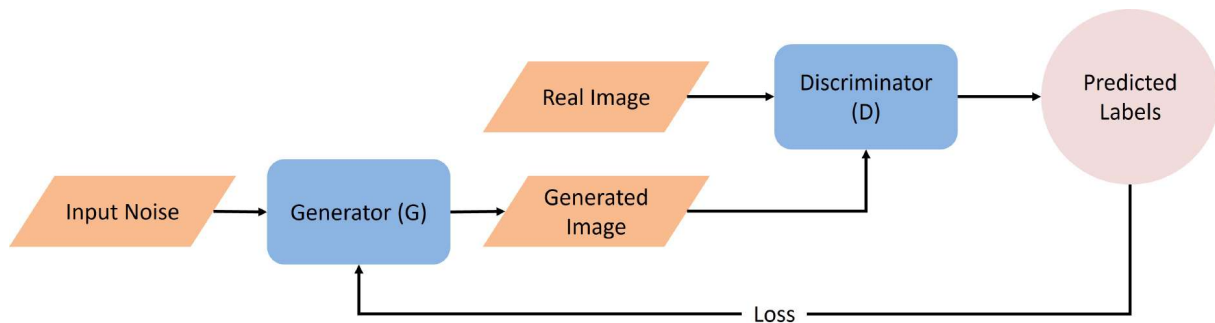
Choi et al. [Choi 21] contributed a ray-based multibeam sonar plugin, designed to consider the phase and reverberation physics of acoustic signals. This plugin provided raw sonar intensity-range data (A-plot) using the point scattering model, incorporating considerations for time and angle ambiguities, as well as speckle noise. The time and angle ambiguity in this context referred to the point spread function of the coherent imaging system, involving side lobes due to matched filtering and beamforming. While Choi’s work demonstrated good results, it still showcased less reverberation noise compared to real underwater environments (Figure 4.13), highlighting the challenges in achieving a fully realistic simulation.

The complexity of underwater acoustics, including reverberation noise, remains a significant challenge. As the Figure 4.13 shows, existing models often fall short of reproducing the rich and precise noise patterns observed in real-world underwater scenarios. These limitations prompted a novel approach in the present study, focusing on the application of Generative Adversarial Networks (GANs) to generate realistic sonar imagery, a process that will be explained in subsequent sections. This shift in methodology aims to overcome the shortcomings of existing models and contribute to the creation of a more authentic and accurate sonar plugin within the Unreal Engine.

## Generative Adversarial Networks

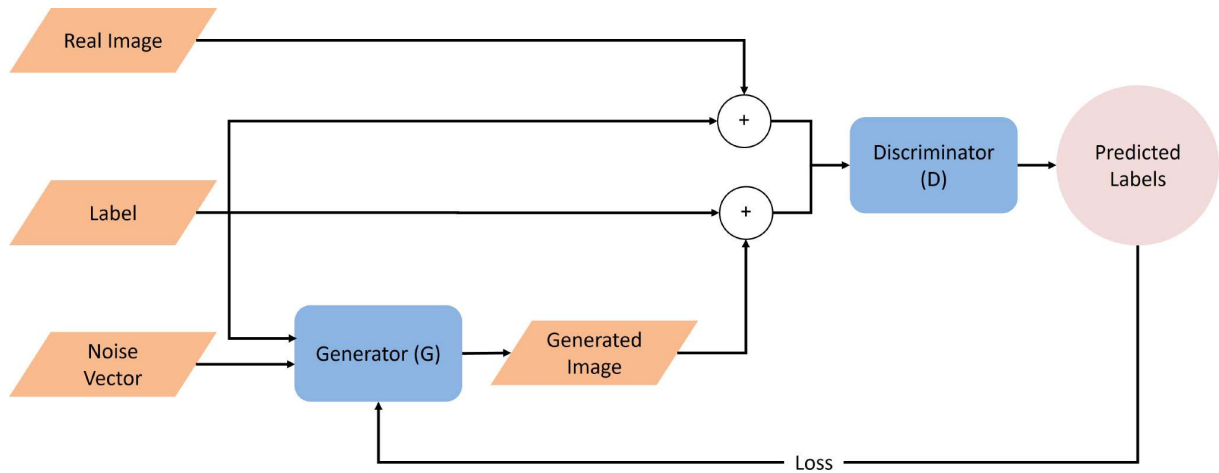
Generative Adversarial Networks (GANs) represent a class of artificial intelligence algorithms used in unsupervised machine learning, implemented by a system of two neural networks contesting with each other in a zero-sum game framework. This technique was introduced by Ian Goodfellow et. al. [Goodfellow 20] in 2014 and has since revolutionized

the field of generative modeling. GANs consist of two parts: the generator and the discriminator. The generator creates data that is similar to the input data but not identical, attempting to generate passable synthetic images, while the discriminator evaluates the authenticity of the generated data against the real data as shown in Figure 4.14. Through iterative training, where the generator continuously learns to produce more accurate synthetic images and the discriminator becomes better at identifying them, the system progressively improves until the discriminator cannot easily differentiate real data from synthetic. This adversarial process results in the generator producing high-quality, realistic synthetic data, which can be used in various applications, such as image generation, video creation, and voice synthesis. GANs have significant implications in the field of computer vision, natural language processing, and beyond. They have been used to create photorealistic images, simulate artistic styles for art generation, generate realistic human voices, and even create virtual environments for training autonomous systems. The technology's ability to understand and replicate complex data distributions makes GANs a powerful tool for researchers and developers, pushing the boundaries of artificial creativity and data synthesis.



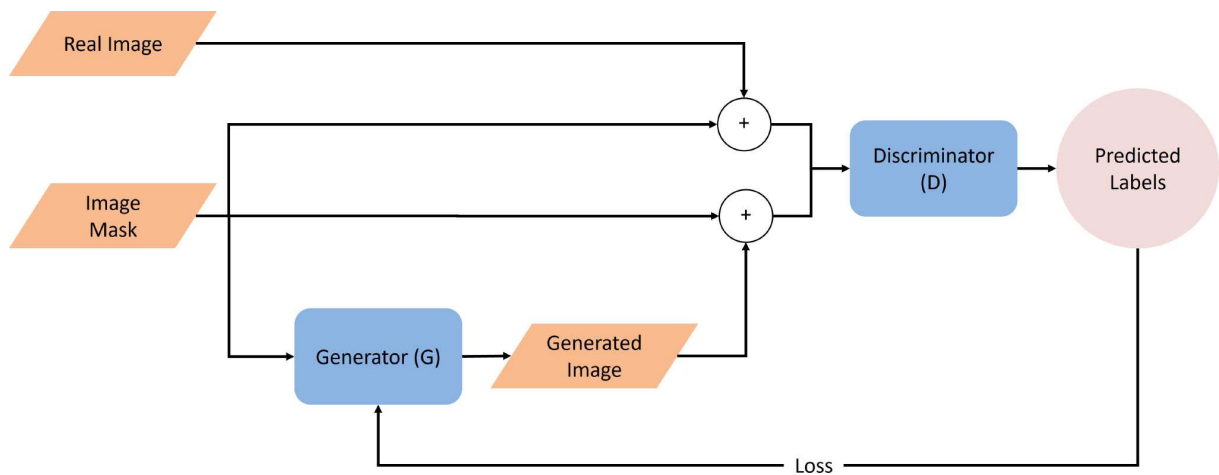
**Figure 4.14:** Architecture of a Generative Adversarial Network (GAN) - This figure depicts the structure of a GAN, where the Generator (G) creates images from random noise, and the Discriminator (D) assesses them against real images to determine their authenticity. The training process is governed by a loss function that guides the improvement of both networks, with the ultimate goal of the generator producing images indistinguishable from actual ones to the discriminator.

It is interesting to note that while standard GAN architecture can generate realistic imagery, they are not context-specific and often lack relevance in a specific domain, such as sonar imagery. In contrast, conditional GANs (cGANs) have the ability to produce better results in context-specific tasks by incorporating conditional variables into the generative process. By including conditional inputs such as labels or images, cGANs can generate images that conform to specified conditions, making them more applicable in tasks like style transfer, photo enhancement, and even the transformation of sketches into photorealistic images. The architecture of cGANs involves two key components: a generator and a discriminator. The generator, which is trained to produce images, receives both a random noise vector and a conditional input (Label) and generates an output image that corresponds to this condition. Meanwhile, the discriminator evaluates the authenticity of the generated images against real images, taking the conditional input into account (shown in Figure 4.15). This modification allows cGANs to be more context-specific and produce better results in specific domains.



**Figure 4.15:** Architecture of a Conditional Generative Adversarial Network (cGAN) - This figure showcases the structure of a cGAN, where the Generator (G) creates images based on input noise and labels, and the Discriminator (D) assesses their authenticity using the same labels. The process aims to refine the generator's output to closely mimic real images, as determined by the discriminator, with the goal of minimizing loss and enhancing the realism of the generated images.

To train a cGANs for generating sonar imagery, the approach used in this thesis involves utilizing obstacle masks from state-of-the-art datasets as an input to Generator (G) and Discriminator (D). This variant of cGAN is known as Pix2pix GAN that uses image pair and implements image-to-image translation as shown in Figure 4.16. This strategy



**Figure 4.16:** Architecture of a Pix2Pix Conditional Generative Adversarial Network (cGAN) - This diagram illustrates the Pix2Pix GAN model, which is designed for image-to-image translation tasks. The network comprises a Generator (G) that takes an image mask as input and generates a synthetic image aiming to resemble a corresponding real image. The Discriminator (D) receives both the generated image and the real image, concatenated with their respective image masks, to judge the authenticity of the synthesized image. The adversarial training process is governed by a loss function, which evaluates the discriminator's accuracy, providing the necessary feedback for iterative improvement of the generator's performance.

is tested using two distinct datasets, namely the Marine Debris dataset (Section 4.1.1) and the UATD dataset (Section 4.1.1). The Marine Debris dataset is equipped with segmentation masks and object detection annotations, serving as valuable resources for the GANs training. In contrast, the UATD dataset exclusively provides bounding box annotations. Generating segmentation masks manually can be a laborious task; hence, a method is devised to automatically generate binary masks using the bounding boxes. Pixels within these bounding boxes are considered as objects, marked in white within the mask, while the rest of the imagery remained black. This approach facilitated the creation of masks similar to those found in the Marine Debris dataset.

Both datasets capture sonar images through video sequences, leading to inherent similarities among the images. To introduce diversity into the training data, we selectively curated dissimilar images from each dataset by leveraging metrics such as structural similarity index (SSI), peak signal-to-noise ratio (PSNR), and mean square error (MSE). The values of metrics selected as threshold to filter the similar images are presented in Tables 4.1.

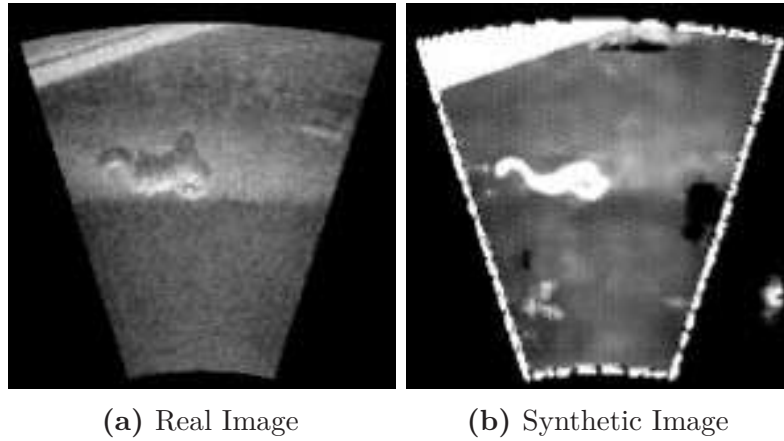
**Table 4.1:** Threshold Metrics for Image Selection in Video Sequences. This table presents threshold values for SSI, PSNR, and MSE to filter out similar images, ensuring that only dissimilar images are used for efficient GAN network training..

| SSI  | PSNR | MSE |
|------|------|-----|
| 0.63 | 25.4 | 178 |

Recognizing the essential role of image augmentation in enhancing the robustness of deep learning architectures, various augmentation techniques have been implemented including mirror, zoom, rotate, random brightness, random jitter, and random contrast adjustments. However, the experimentation suggested that only random jitter, and mirror techniques are succeeded in generating a realistic sonar imagery. Other techniques introduced undesirable artifacts, as illustrated in Figure 4.17. The applied augmentations, illustrated in Figure 4.18, showcase the impact of these techniques on the sonar images and underscore the significance of careful augmentation strategies in achieving high-quality results.

This data preprocessing, encompassing dataset selection, dissimilarity curation, and effective image augmentation, lays the groundwork for training GANs to generate realistic synthetic sonar images. The success of GANs depends on the richness and diversity of the training data, and the approach ensures a carefully curated dataset that encapsulates the complexity and variability inherent in real-world underwater environments.

The Generator, utilizes architecture of the U-Net model due to its unique configuration, featuring a contracting path to capture context and an expansive path to enable precise localization, that makes it exceptionally suited for detailed image synthesis tasks. It's this symmetry and the use of skip connections that allow the network to preserve important spatial information throughout the network, resulting in highly accurate synthetic image outputs that retain the complexity of the original input features. Besides, the Discriminator is tailored to evaluate the integrity of the generated images, operating with varying levels of granularity by altering its receptive field size. The discriminative process ranges from the granular PixelGAN, which analyzes the image at the pixel level, to the broader Image GAN with a receptive field of 286x286 pixels. This spectrum includes Patch GANs with receptive



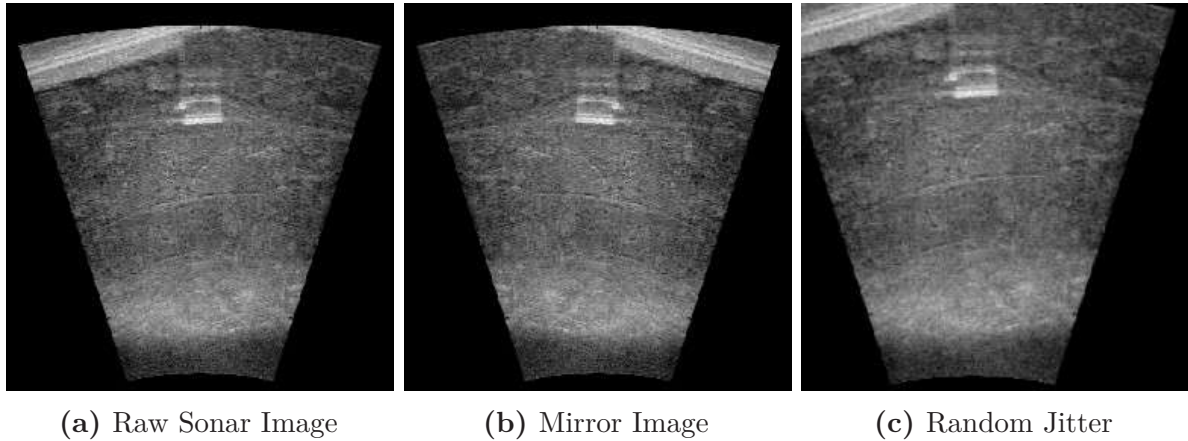
**Figure 4.17:** Effect of Augmentation Techniques over Synthetic Images generated using Pix2Pix cGAN. (a) displays a real sonar image captured from an underwater environment, showcasing natural acoustic reflections. (b) presents a synthetic counterpart produced through various image augmentation techniques such as mirroring, zooming, rotation, random brightness adjustment, jitter, and random contrast modulation. Despite the synthetic image in (b) being the result of a substantial training, encompassing 500 epochs, it is critical to observe that it still contains several undesired artifacts that are not present in the original real image in (a).

fields of 16 and 70 pixels, respectively, providing an approach to assessing the quality of the generated images. The choice of receptive field directly impacts the Discriminator's ability to discern local image features versus larger patterns and textures, thus influencing the training dynamics and the quality of the generated images.

To compute the performance of these varying Discriminator configurations, the L1 loss combined with the conditional GAN (cGAN) loss is utilized. The L1 loss ensures the generated images are close to the ground truth in pixel space, promoting structural similarity in the output, while the cGAN loss encourages the generated images to be indistinguishable from real images in the eyes of the Discriminator. This hybrid loss function is pivotal in steering the Generator towards producing images that are not only accurate in a pixel-wise sense but also convincing in terms of their overall composition and realism. In the subsequent sections, the architectural complexities of the GANs networks, their performance, and the comparative analysis against sonar images are comprehensively explored.

## Comparative Analysis - Standard Image Quality Metrics

Evaluating the performance of image synthesis algorithms such as Generative Adversarial Networks (GANs) is complex and multifaceted, relying on several key metrics that provide the quality of generated images. These metrics, which include the Structural Similarity Index (SSI), Peak Signal-to-Noise Ratio (PSNR), and Mean Squared Error (MSE), offer unique insights into different aspects of image quality. SSI evaluates the perceived changes in structural information, texture, and contrast of images, while PSNR measures the fidelity of a synthetic image to an original, and MSE calculates the pixel-wise variance between them. These metrics are indispensable for a comprehensive analysis of the image generation models, allowing for a holistic assessment of their performance.



**Figure 4.18:** Implementation of Selective Augmentation Techniques for Enhanced Pix2Pix cGAN Performance. This figure displays two effective augmentation techniques used for Pix2Pix cGAN to generate synthetic sonar images: mirroring and random jitter. These methods are chosen to enhance the quality of generated images while maintaining feature integrity, successfully addressing the problem of artifacts seen in Figure 4.17.

**Structural Similarity Index (SSI)** - The Structural Similarity Index (SSI), also known as the Structural Similarity Index Measure (SSIM), is a metric used to quantify the visual similarity between two images. It evaluates the visual impact of three characteristics of an image: luminance, contrast, and structure. SSI compares local patterns of pixel intensities normalized for luminance and contrast, thus focusing on changes in structural information, texture, and contrast. The SSI value ranges between -1 and 1, with 1 indicating perfect similarity. This metric is particularly useful for assessing the perceptual quality of images because it aligns closely with human visual perception. Higher SSI values indicate greater similarity between the generated image and the reference, which is typically the desired outcome in image generation tasks.

**Peak Signal-to-Noise Ratio (PSNR)** - Peak Signal-to-Noise Ratio (PSNR) is a commonly used metric that measures the quality of an image by comparing it to a reference image. PSNR is expressed in decibels (dB) and provides a numerical value representing the ratio of the peak signal strength to the noise introduced by compression or other factors. Higher PSNR values indicate a lower level of distortion, and thus, better image quality. It is widely employed in image processing to assess the fidelity of generated images.

**Mean Square Error (MSE)** - Mean Square Error (MSE) is a metric that calculates the average squared difference between the pixel values of the generated and reference images. MSE provides a quantitative measure of the overall pixel-wise difference between the two images. A lower MSE indicates less discrepancy and, theoretically, better image quality. However, MSE is sensitive to outliers and may not always align with human perception, making it important to interpret results in conjunction with other metrics.

### Variation in Network Performance and Dataset Complexity

To determine the performance of various Pix2Pix architectures, a comparative analysis is conducted on two sonar imagery datasets: the Marine Debris Dataset and the Underwater Acoustic Target Detection (UATD) Dataset (already explained in Section 4.1.1). The Marine Debris Dataset, gathered in a controlled pool environment, allows for examination under stable conditions, whereas the UATD Dataset, compiled from a dynamic river environment, presents a more challenging scenario with varying acoustic properties. The evaluative focus on SSIM, PSNR, and MSE metrics shed light on how different GAN models perform under varied conditions (Table 4.2). For instance, PixelGAN demonstrates proficiency with the Marine Debris Dataset, suggesting its effectiveness in less complex scenarios. Conversely, ImageGAN shows a higher SSIM value for the UATD Dataset, indicating its robustness in maintaining structural details within a more complex setting. The PSNR results across all networks for the UATD Dataset suggest that the models are well-tuned to reconstruct sonar images with high fidelity, while the MSE scores point to a precise pixel-level reconstruction in these diverse conditions.

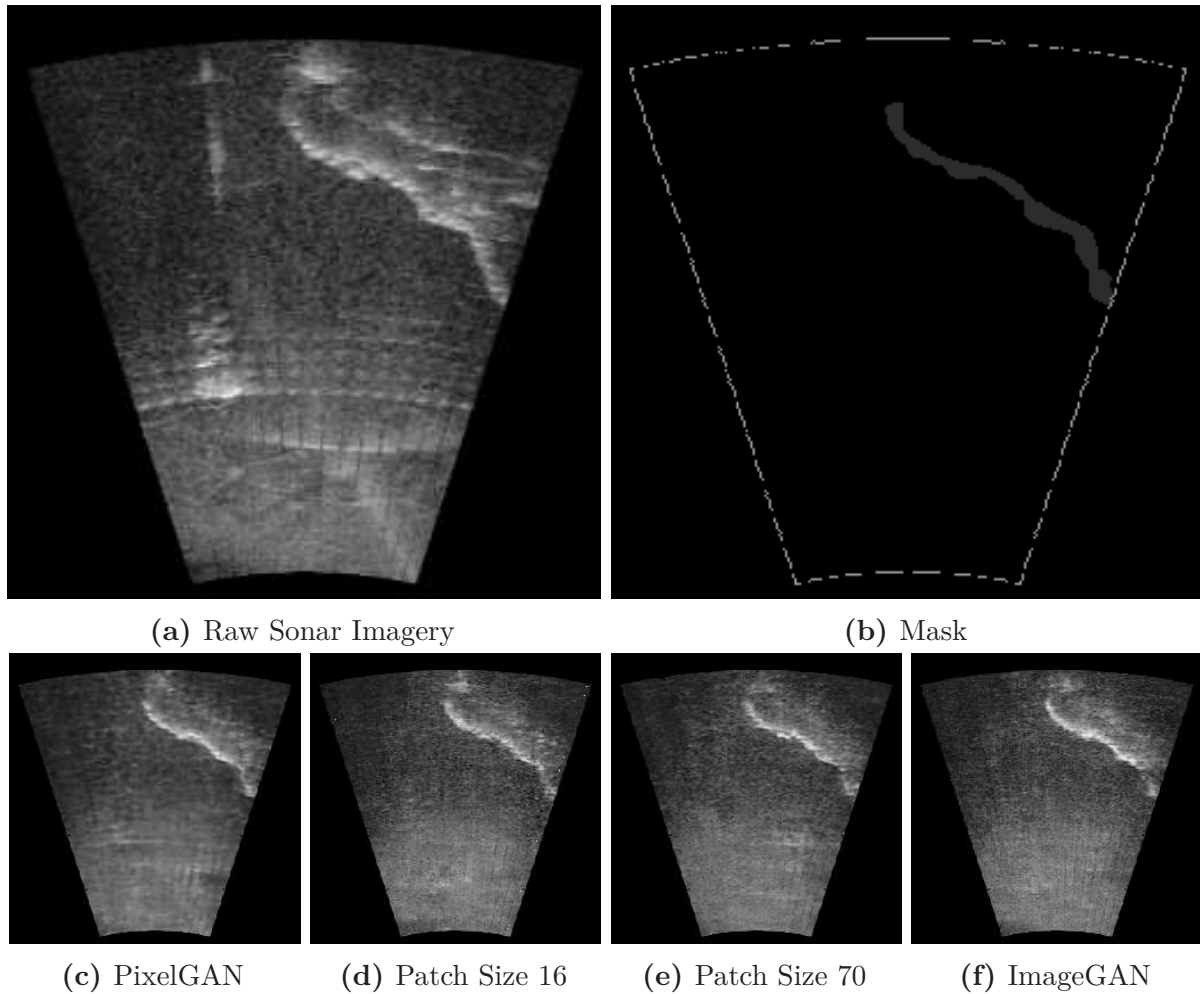
**Table 4.2:** Evaluation Metrics for Generated Images Across Different Networks and Datasets. This table presents a comparison of the Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Mean Squared Error (MSE) for images generated by various GAN networks, namely PixelGAN, PatchGAN with patch sizes (PS) of 16 and 70, and ImageGAN. The metrics are calculated for two distinct datasets: the Marine Debris Dataset and the UATD Dataset.

| Networks         | Marine Debris Dataset |               |               | UATD Dataset |               |               |
|------------------|-----------------------|---------------|---------------|--------------|---------------|---------------|
|                  | <i>SSIM</i>           | <i>PSNR</i>   | <i>MSE</i>    | <i>SSIM</i>  | <i>PSNR</i>   | <i>MSE</i>    |
| PixelGAN         | <b>0.586</b>          | <b>24.265</b> | <b>49.532</b> | 0.755        | 31.845        | 21.078        |
| PatchGan (PS:16) | 0.546                 | 23.090        | 52.079        | 0.729        | 30.995        | 23.759        |
| PatchGan (PS:70) | 0.561                 | 23.558        | 51.209        | 0.749        | 31.331        | 23.014        |
| ImageGan         | 0.548                 | 23.192        | 52.174        | <b>0.760</b> | <b>32.034</b> | <b>20.800</b> |

The performance variations among the networks within each dataset indicate the adaptability of GAN architectures to specific environmental factors. PixelGAN’s higher performance in the Marine Debris Dataset could be due to its efficiency in handling less complex sonar imagery, while ImageGAN’s success in the UATD Dataset may stem from its advanced ability to interpret the more complex river environment.

### Qualitative Results and Indirect Evaluation

However, when inspecting the generated images visually, a discrepancy between the quantitative metrics and qualitative assessment becomes apparent. Figures 4.19 and 4.20, showcasing images from the Marine Debris and UATD datasets, reveal that despite PixelGAN and ImageGAN’s numerical superiority, the images they generate appear blurrier than those from PatchGAN. This blurriness could obscure vital details crucial for accurate underwater object detection and classification. This observation suggests that standard metrics like MSE, PSNR, and SSIM may not wholly capture the necessary perceptual and contextual quality for sonar imagery analysis. As a result, indirect evaluation methods, such as the performance of object detection algorithms on generated images, might offer a more significant measure of a GAN model’s practical utility.



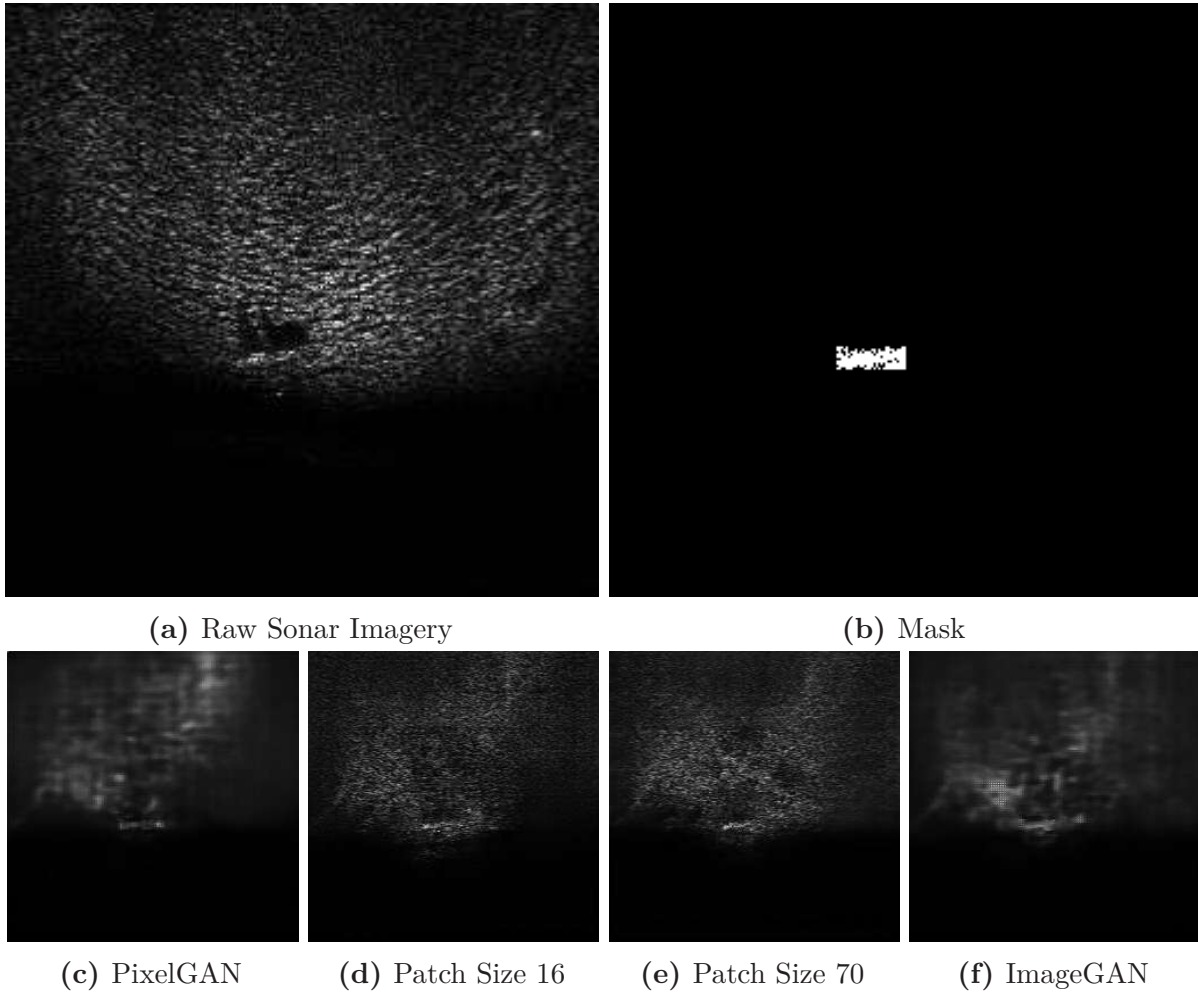
**Figure 4.19:** Generated Imagery from Marine Debris Dataset at different field sizes. This figure illustrates the images generated at different receptive field sizes from the image mask of the Marine Debris Dataset. According to Table 4.2, PixelGAN has better SSI, PSNR, and MSE values, but PixelGAN has generated visually blurred images.

Such indirect evaluations could include the ability of object detection algorithms to identify and localize objects within images accurately, relying on the preservation of edge information and textural details. Therefore, while PixelGAN and ImageGAN may statistically outperform other models, the practical needs of downstream tasks, such as object detection, highlight the need for reevaluating the metrics used to measure GAN success in sonar image generation. This insight emphasizes the importance of aligning evaluation processes with the end-use of generated images and indicates that including indirect evaluation methods could yield a more thorough understanding of a GAN's performance and relevance.

### Comparative Analysis - Object Detection Metrics

The evaluation of Generative Adversarial Networks (GANs) in the context of image quality assessment often relies on metrics such as Structural Similarity Index (SSI), Peak Signal-to-Noise Ratio (PSNR), and Mean Squared Error (MSE). While these metrics provide valuable insights into image fidelity, they have limitations that necessitate consideration.





**Figure 4.20:** Generated Imagery from Marine Debris Dataset at different field sizes. This figure illustrates the images generated at different receptive field sizes from the image mask of the UATD Dataset. According to Table 4.2, ImageGAN has better SSI, PSNR, and MSE values, but ImageGAN has generated visually blurred images.

SSI primarily focuses on structural information but may not fully account for complex perceptual aspects of human vision, including higher-level semantics and subtle differences critical for image understanding. Moreover, it can be sensitive to specific image transformations and may vary in effectiveness depending on the nature of the generated images. Similarly, both PSNR and MSE are pixel-wise metrics, lacking consideration for structural or semantic content in images and potentially prioritizing insignificant pixel differences. These metrics are sensitive to small pixel value changes, making them less robust to certain image transformations, such as translations or rotations, which might not significantly impact perceptual quality. To address these limitations, an indirect evaluation method involving object detection metrics is introduced.

### Object Detection Model Configuration

The indirect evaluation has utilized three object detection models: YoloX S, YoloX L, and Faster R-CNN. Each model is carefully configured to suit the nature of the sonar images. YoloX models, known for their speed and efficiency in real-time detection scenarios, are

adjusted in their small (S) and large (L) variants to balance between speed and accuracy. Faster R-CNN, renowned for its precision in object detection tasks, is calibrated to identify subtle features within the marine debris context. The Detr model, included in the study, showed minimal performance and thus, its results are not discussed. The detail of each model is presented in Appendix E

Each object detection model is independently trained on the GAN-generated images with corresponding annotations of obstacle locations. This training aimed to equip the models to recognize and localize obstacles within the synthetic imagery. The evaluation of the models' performance is conducted using two primary metrics: mean Average Precision (mAP) and Recall. The mAP metric assessed the precision of the models across various threshold levels, while Recall measured the models' ability to detect all relevant instances of debris. For readers understanding the mAP and Recall are elaborated below.

**Recall:** Recall, also known as sensitivity or true positive rate, measures the ability of an object detection algorithm to find all relevant instances of objects in an image. It is calculated as the ratio of true positive detections to the sum of true positives and false negatives. It ranges from 0 to 1, where 1 indicates that the algorithm has found all instances of objects in the dataset. Recall is particularly crucial in scenarios where missing an object is more critical than having false positives. For instance, in applications like surveillance or medical imaging, high recall is often prioritized.

**Mean Average Precision (mAP):** Mean Average Precision is a widely used metric in object detection tasks. It combines precision and recall across multiple levels of confidence thresholds to provide a comprehensive evaluation of an algorithm's performance. The mAP is calculated by averaging the precision values at different recall levels. This means that an algorithm achieving high precision at various recall levels will receive a higher mAP score. Higher mAP values indicate better object detection performance, reflecting both high precision and recall across different confidence thresholds.

### **Training and Evaluating Object Detection Models on Marine Debris Dataset**

To determine the quality of the images generated using the marine debris dataset, three object detection models are trained and evaluated. This includes examining each model's performance using both quantitative metrics and qualitative assessments. The findings are documented in detail in the subsequent section, which provides insights into the strengths and limitations of each model. After the evaluation, the generated images are tested using the trained models to demonstrate the performance and accuracy of the methodology in generating sonar imagery that closely mirrors real-world conditions. This comprehensive evaluation process aims to validate the effectiveness of the presented approach in producing highly realistic sonar imagery and enhancing the understanding and capabilities in underwater evaluation and testing.

### **Object Detection Analysis of Generated Images on Marine Debris Dataset**

In this section, the quality and realism of the generated images using an indirect approach will thoroughly evaluated. This analysis uses object detection models that are trained on

real data to test the generated imagery through evaluation metrics. The evaluation will cover both quantitative (Table 4.3) and qualitative (Figure 4.21) analyses, which will offer a comprehensive assessment of the performance and integrity of the generated images.

**Detailed Analysis of Table 4.3 Results** - The analytical cornerstone of this study resides in the results presented in Table 4.3, which encapsulates the performance of various GANs when interfaced with different object detection models on a marine debris dataset. The table outlines the mean Average Precision (mAP) and Recall metrics, serving as quantitative indicators of each model’s detection accuracy and comprehensiveness, respectively. These metrics are essential for understanding the practical viability of employing synthetic imagery in real-world marine debris detection applications.

**Table 4.3:** Indirect Evaluation of Generated Images in Marine Debris Dataset. This table summarizes the performance of various generative adversarial networks (GANs) when applied to marine debris image datasets and evaluated using different object detection frameworks.

| Networks                | YoloX S     |               | YoloX L    |               | Faster RCNN |               |
|-------------------------|-------------|---------------|------------|---------------|-------------|---------------|
|                         | <i>mAP</i>  | <i>Recall</i> | <i>mAP</i> | <i>Recall</i> | <i>mAP</i>  | <i>Recall</i> |
| PixelGAN                | 0.76        | 0.54          | 0.67       | 0.46          | 0.66        | 0.43          |
| <b>PatchGAN (PS:16)</b> | <b>0.81</b> | <b>0.56</b>   | 0.76       | 0.51          | 0.66        | 0.43          |
| PatchGAN (PS:70)        | 0.81        | 0.55          | 0.72       | 0.48          | 0.67        | 0.42          |
| ImageGAN                | 0.79        | 0.56          | 0.71       | 0.49          | 0.61        | 0.45          |

**PixelGAN** - The results for PixelGAN reveal a performance characterized by moderate mAP and Recall across YoloX-s and YoloX-l, with a noticeable decrease in performance when interfaced with YoloX-l. PixelGAN’s approach to pixel-level image generation seems to be more compatible with the YoloX-s model, which could be attributed to the model’s focus on agility and speed, possibly favoring the detailed textures that PixelGAN generates. However, the performance dip in YoloX-l suggests that the larger model, which potentially focuses on broader image features, does not leverage PixelGAN’s strengths as effectively. When observed in conjunction with Faster R-CNN, PixelGAN’s mAP and Recall metrics indicate an average performance, suggesting a moderate level of compatibility between PixelGAN’s synthetic imagery and the high precision demands of Faster R-CNN.

**PatchGAN (PS:16)** - PatchGAN, with a patch size of 16, demonstrates a remarkable synergy with the YoloX-s model, achieving the highest mAP score of 0.81 within this configuration. This suggests that the fine-grained detail captured by the small patch size translates well into detectable features for YoloX-s. The Recall metric, while not the highest, is sufficiently robust, indicating a balanced performance between detecting most debris instances and maintaining a high precision rate. The results with YoloX-l also show high mAP values, indicating that the detailed texture information provided by PatchGAN remains valuable across different scales of the YoloX architecture.

**PatchGAN (PS:70)** - Expanding the patch size to 70 with PatchGAN presents an intriguing outcome. The mAP score slightly decreases when compared to the PS:16 variant, suggesting that larger patch sizes may not capture the granularity required

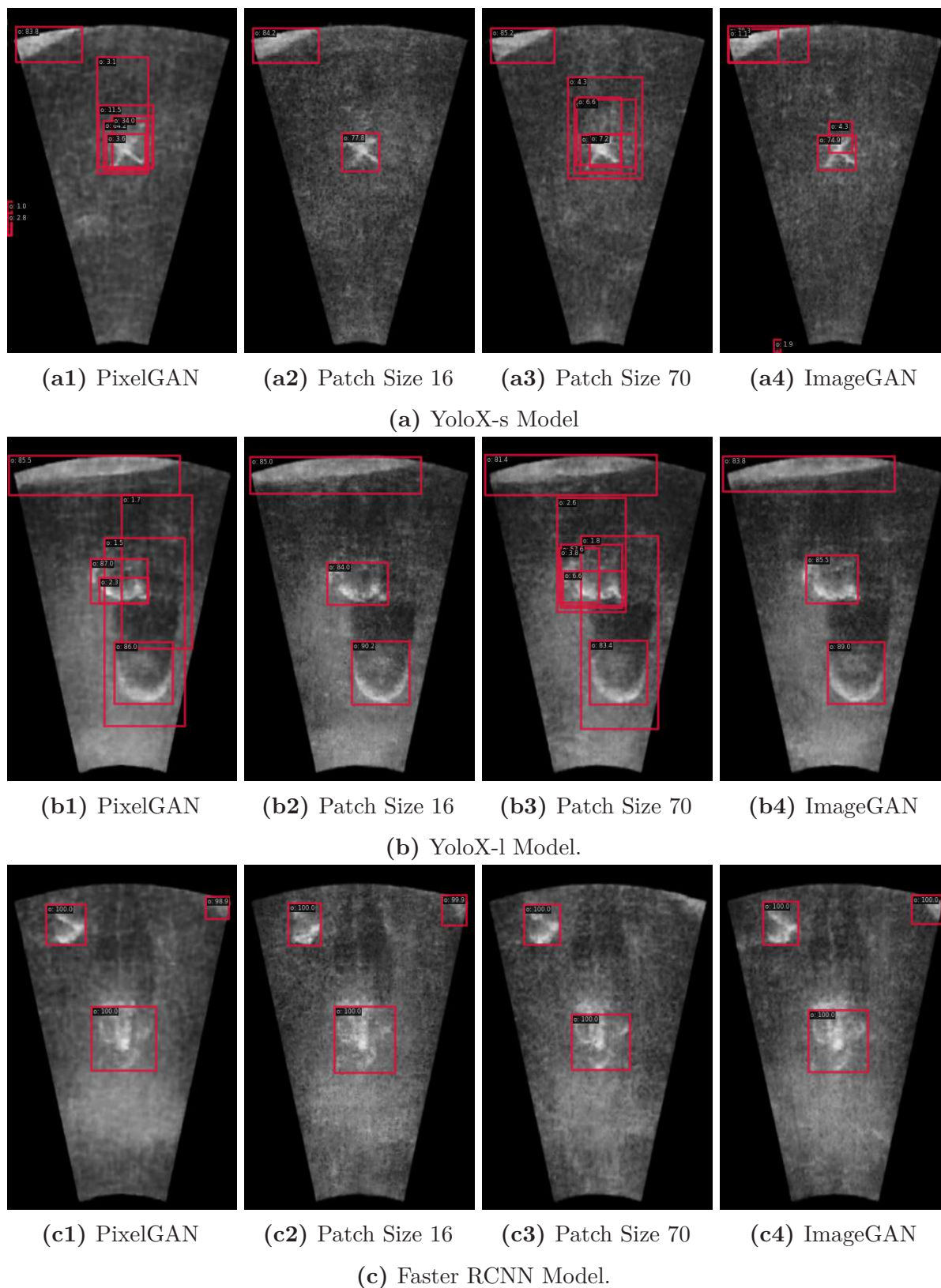
by YoloX-s as effectively. However, the Recall metric is consistent with its PS:16 counterpart. This might indicate that while the precision of detection might decrease slightly with increased patch size, the ability to detect all relevant objects remains stable. This trend is slightly more pronounced with YoloX-l, where the larger patch size seems to have a more substantial impact on precision, highlighting the delicate balance between patch size and the object detection model's ability to utilize the generated features.

**ImageGAN** - ImageGAN, with its broader focus on image-to-image translation, presents results that are slightly lower in mAP across the board when compared to PatchGAN (PS:16), yet still competitive. This might indicate that while ImageGAN can generate contextually rich images, the finer details crucial for high precision in object detection may be lacking. Nonetheless, the Recall metrics for ImageGAN across YoloX s and l, and Faster R-CNN, are indicative of a GAN that still performs well in detecting the presence of objects across varying scales and complexities of models.

**Visual Analysis of Figure 4.22 Results** The quantitative data presented in Table 4.3 is compelling, but it is the visual evidence from Figure 4.21 that truly contextualizes the performance of the GANs in conjunction with object detection models. This figure provides a comparative display of the detection results from four distinct GANs when analyzed using the YoloX-s, YoloX-l, and Faster R-CNN models. The red bounding boxes overlaid on the sonar images symbolize the detected marine debris, giving a qualitative assessment of each model's detection proficiency.

**YoloX-s Model (Figure 4.21a)** - The YoloX-s model's detection results demonstrate varying degrees of accuracy and completeness. With PixelGAN, the bounding boxes are fewer and less concentrated, suggesting that while some debris is detected, there may be a notable amount of debris that is not being captured by the model. This correlates with the moderate mAP and Recall values reported in Table 4.3. In contrast, the PatchGAN with a patch size of 16 exhibits a denser cluster of bounding boxes around the debris. This visual clustering corroborates the high mAP score from Table 4.3 and suggests that the model is not only detecting a greater number of debris items but is also doing so with a higher degree of precision. The similar performance is noted with PatchGAN (PS:70), where the bounding boxes are still abundant, although slightly more dispersed, which might be indicative of the slight drop in mAP observed in the table. ImageGAN shows a moderate density of bounding boxes, aligning with the slightly lower mAP and consistent Recall scores. It appears that ImageGAN's synthetic images provide enough context for the YoloX-S model to identify debris effectively, though perhaps with less precision than PatchGAN (PS:16).

**YoloX-l Model (Figure 4.21b)** - The detection outcomes using the YoloX-l model elucidate a different aspect of the GANs' performance. The YoloX-l model, being a larger variant, is expected to integrate more contextual information for detection. The visual results from PixelGAN show sparser detection, which is consistent with the lower mAP score seen in the table. This could suggest that the images generated by PixelGAN may not scale as well with the increased complexity of the YoloX-l model. The PatchGAN results, both for PS:16 and PS:70, show a good number



**Figure 4.21:** Comparison of Different GANs using YoloX-s, YoloX-l, and Faster RCNN Detection Model on Marine Debris Dataset. This Figure illustrates the object detection results on a marine debris dataset using the YoloX-s, YoloX-l, and Faster RCNN models in conjunction with four different generative adversarial networks (GANs): PixelGAN, PatchGAN with a patch size of 16, PatchGAN with a patch size of 70, and ImageGAN.

of accurately placed bounding boxes, though the former appears to have a slight edge in detection density and precision, visually affirming the quantitative findings. The ImageGAN detection results are also robust, though they seem to lack the granularity of the PatchGAN (PS:16), which is expected given the differences in their image generation approaches.

**Faster R-CNN Model (4.21c)** - For the Faster R-CNN model, the detections exhibit a trend that aligns with the quantitative results, but with additional insights. The visual analysis allows us to discern not just the presence of bounding boxes but also their appropriateness in terms of size and placement relative to the debris items. PixelGAN and ImageGAN show a spread of bounding boxes that, while indicative of debris detection, also suggest potential false positives, which would not have been as evident from the numerical data alone. PatchGAN (PS:16) continues to show strong visual alignment with the debris, with bounding boxes that are well-placed and tightly fit around the items, minimizing false detections. The PatchGAN (PS:70) results are similar, though the larger patch size seems to result in slightly less precise bounding box placement, which is an interesting observation that complements the mAP and Recall figures reported.

### Training and Evaluating Object Detection Models on UATD Dataset

As discussed in preceding sections, MASTER dataset (Section 4.1.2) uses the similar forward-looking sonar as that of UATD dataset (4.1.1). Due to less number of images

To expand the dataset, we integrated the underwater acoustic target detection (UATD) dataset (refer to Section 4.1.1), which has been acquired through the utilization of the Tritech Gemini 1200i sonar system, capable of dual-frequency operation at 720 kHz and 1200 kHz. While the UATD dataset shares similarities with MASTER dataset, the UATD dataset is noteworthy for featuring multiple classes, suited for the recognition of different underwater entities. Given the primary objective of object detection, and the inherent challenges posed by the random objects found in flooded environments, the several distinct classes within the UATD dataset are transformed into a single unified *object* class.

In total, the deep networks are trained using over 2000 images, drawn from both the MASTER and UATD datasets. This extensive dataset facilitated a rigorous evaluation of different network architectures, and the ensuing detection results are comprehensively presented in the sections below. The utilization of these datasets, tailored to the specific demands of object detection in underwater sonar imagery, has enabled us to enhance the robustness and performance of our algorithms, ultimately contributing to the advancement of this vital field of research.

**Detailed Analysis of Table 4.4 Results** - The data presented in Table 4.4 provides a comprehensive view of the performance metrics for various object detection algorithms across two distinct datasets: UATD and a combined dataset of MASTER and UATD. Here the mean Average Precision (mAP) and Recall metrics are presented to understand the strengths and potential limitations of each object detection network when applied to the task of underwater acoustic target detection, as well as their performance in a multimodal shallow water environment when combined with UATD.

**Table 4.4:** Performance Metrics of Object Detection Algorithms on Different Datasets. This table compares the average precision (mAP) and recall of various object detection networks trained and tested on two distinct datasets: UATD (Underwater Acoustic Target Detection) and a combined set of MASTER (Multi-modal Shallow waTER) and UATD.

| Networks      | YoloX S     |               | YoloX L    |               | Faster RCNN |               |
|---------------|-------------|---------------|------------|---------------|-------------|---------------|
|               | <i>mAP</i>  | <i>Recall</i> | <i>mAP</i> | <i>Recall</i> | <i>mAP</i>  | <i>Recall</i> |
| UATD          | <b>0.91</b> | <b>0.50</b>   | 0.87       | 0.48          | 0.86        | 0.42          |
| MASTER + UATD | 0.66        | 0.33          | 0.74       | 0.42          | <b>0.78</b> | <b>0.47</b>   |

**YoloX-s** - The YoloX-s model showcases a notably high mAP of 0.91 on the UATD dataset, indicating a superior precision in identifying targets within this specific set. This performance, however, is not mirrored on the combined MASTER + UATD dataset, where the mAP drops to 0.66. The significant reduction in precision could be indicative of the increased complexity or variability within the combined dataset, which may introduce challenges that the YoloX-s model is less equipped to handle. The Recall metric sees a decrease from 0.50 in the UATD dataset to 0.33 in the combined dataset, suggesting that the model's ability to detect all relevant instances is compromised when faced with the diversity of the combined dataset.

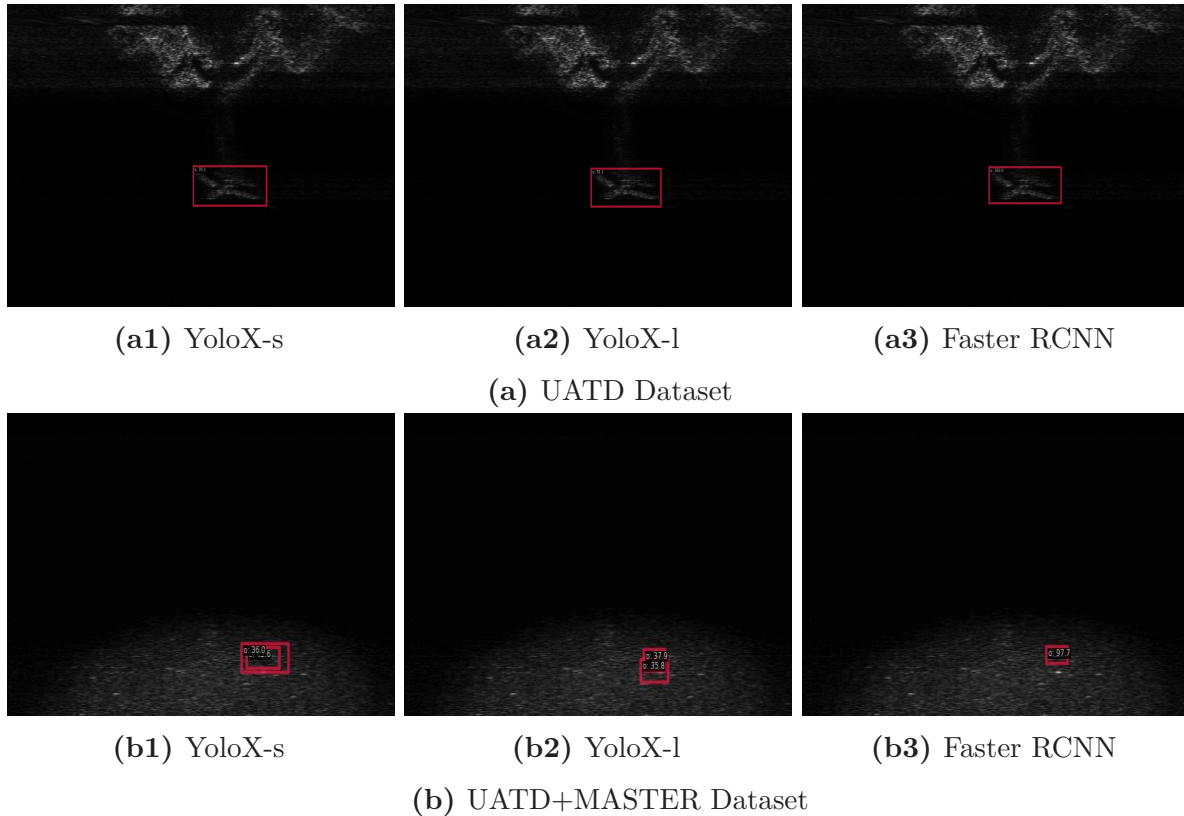
**YoloX-l** - The larger YoloX-l variant follows a similar trend to its smaller counterpart, with a decrease in mAP when transitioning from the UATD dataset (0.87) to the combined dataset (0.74). While the drop in precision is less pronounced than in the YoloX-s model, it nonetheless signifies that the YoloX-l model also struggles with the increased complexity of the combined dataset. The Recall metrics indicate a comparable decrease, reinforcing the notion that the combined dataset presents a more challenging detection environment for the YoloX-l model.

**Faster R-CNN** - The Faster R-CNN model, which is generally known for its high precision and slower detection speeds, shows less variation in performance between the two datasets compared to the YoloX models. The mAP on the UATD dataset is robust at 0.86, slightly decreasing to 0.78 on the combined dataset. This indicates that the Faster R-CNN model maintains a relatively stable precision across both datasets. The Recall metric also decreases slightly from 0.42 to 0.47, which suggests that while the model's overall detection rate is lower, it remains somewhat consistent across the two datasets.

**Comparative Analysis** - The comparative analysis of the object detection algorithms on the UATD and combined datasets provides valuable insights. The performance on the UATD dataset across all models is generally higher, indicating that the models are better suited or have been better optimized for this dataset. The combined MASTER + UATD dataset appears to introduce factors that reduce both precision and recall, which could be due to a variety of reasons including but not limited to increased noise, different types of targets, and a broader range of environmental conditions. Moreover, one notable observation is the relative stability of the Faster R-CNN model's performance across both datasets, which may point to its robustness as an object detection framework in varying conditions. Conversely, the larger fluctuations seen in the YoloX models' performance suggest that these models may

require further tuning to maintain high precision and recall in more complex or diverse datasets.

**Visual Analysis of Figure 4.22 Results** Figure 4.22 offers a side-by-side visual comparison of the performance of various object detection models on two distinct datasets: UATD (Underwater Acoustic Target Detection) and a combined dataset of UATD and MASTER (Multi-modal Shallow waTER). This visual assessment complements the numerical data from Table 4.4 by illustrating the actual detection outputs of the YoloX-S, YoloX-L, and Faster R-CNN models.



**Figure 4.22:** Comparative Object Detection on UATD and UATD+MASTER Datasets. This set of images displays the detection results of different object detection models including YoloX-s, YoloX-l, and Faster RCNN.

#### UATD Dataset (Figure 4.22a)

**YoloX-S** - On the UATD dataset, the YoloX-S model appears to localize the target accurately, as evidenced by the tight and well-placed bounding box. This visual confirmation aligns with the high mAP score of 0.91 noted in Table 5.3, indicating that YoloX-S is highly precise in this controlled environment. The solitary bounding box also reflects the Recall score, suggesting that while precise, the model may not be capturing all present targets.

**YoloX-L** - The YoloX-L model also displays a single, well-defined bounding box, supporting the high mAP of 0.87. The larger model seems to retain most of the precision advantages of the smaller variant, yet the visual output suggests a similar trend in potentially missed detections, as indicated by the Recall metric.



**Faster R-CNN** - Faster R-CNN's visual result is consistent with the other two models, showcasing one bounding box indicative of a high-confidence detection. This corresponds with the robust mAP of 0.86, but the image also visually substantiates the model's lower Recall score, echoing the possibility of undetected targets.

#### **UATD+MASTER Dataset (Figure 4.22)**

**YoloX-S** - In the combined dataset, the YoloX-S model's detection is represented by a less confidently placed bounding box, which is smaller and less centered compared to its UATD-only counterpart. This visual observation supports the lower mAP score of 0.66 and suggests a decline in the model's ability to generalize to the more complex combined dataset.

**YoloX-L** - The YoloX-L model exhibits a similar detection pattern on the combined dataset, with a bounding box that appears less precise than on the UATD dataset. This detection aligns with the reduced mAP of 0.74 and provides a visual testament to the model's decreased precision when faced with the added complexity of the combined dataset.

**Faster R-CNN** - Faster R-CNN shows a relatively small and imprecisely placed bounding box on the combined dataset. While the visual detection is present, it seems to confirm the numerical drop in mAP to 0.78, suggesting that while Faster R-CNN is still relatively stable in precision, its performance is somewhat affected by the dataset complexity.

**Comparative Visual Analysis** - The visual comparison across both datasets elucidates a clear trend: all models perform with higher precision on the UATD dataset compared to the combined UATD+MASTER dataset. The visual evidence suggests that the additional data complexity in the combined dataset impacts the models' ability to detect with the same level of confidence and precision. Furthermore, the visual analysis also highlights the absence of multiple detections in both datasets, which is particularly telling in the context of the Recall scores. While the models demonstrate high precision in their detections, as seen by the presence of bounding boxes, their ability to detect all relevant targets is not visually represented, indicating an area for improvement in target detection comprehensiveness.

### **Object Detection Analysis of Generated Images on UATD Dataset**

Similar to Section 4.2.4, this section thoroughly evaluates the quality of generated images using UATD dataset. The evaluation covers both quantitative (Table 4.5) and qualitative (Figure 4.23) analyses to provide a comprehensive assessment of the performance of the generated images.

**Detailed Analysis of Table 4.4 Results** - Building upon the analytical framework used to dissect Table 4.3, the performance data presented in Table 4.5 is discussed, which involves an indirect evaluation of various GANs within the UATD (Underwater Acoustic Target Detection) dataset. The table systematically records the mean Average Precision (mAP) and Recall metrics across different object detection models, providing a quantitative measure of each combination's performance in detecting marine debris.

**Table 4.5:** Indirect Evaluation of Generated Images in UATD Dataset. This table summarizes the performance of various generative adversarial networks (GANs) when applied to marine debris image datasets and evaluated using different object detection frameworks.

| Networks                | YoloX S    |               | YoloX L     |               | Faster RCNN |               |
|-------------------------|------------|---------------|-------------|---------------|-------------|---------------|
|                         | <i>mAP</i> | <i>Recall</i> | <i>mAP</i>  | <i>Recall</i> | <i>mAP</i>  | <i>Recall</i> |
| PixelGAN                | 0.62       | 0.29          | 0.61        | 0.30          | 0.55        | 0.25          |
| PatchGAN (PS:16)        | 0.67       | 0.33          | 0.71        | 0.36          | 0.54        | 0.26          |
| <b>PatchGAN (PS:70)</b> | 0.72       | 0.34          | <b>0.74</b> | <b>0.36</b>   | 0.58        | 0.29          |
| ImageGAN                | 0.68       | 0.34          | 0.69        | 0.33          | 0.64        | 0.32          |

**PixelGAN** - PixelGAN, a model known for its pixel-level image translation capabilities, shows moderate performance with YoloX s and l, and Faster R-CNN in the UATD dataset. The mAP scores hover around the lower sixties for both YoloX variants, with slightly better performance with YoloX l. This could suggest that PixelGAN’s generated images contain features that are marginally better captured by the larger YoloX variant. However, the Recall rates are comparatively low, indicating that while the model can detect debris with a reasonable degree of precision, it may miss a significant number of relevant instances.

**PatchGAN (PS:16)** - PatchGAN with a patch size of 16 exhibits improved performance over PixelGAN, particularly with the YoloX L model where it achieves an mAP of 0.71. This increment in mAP may be indicative of PatchGAN’s ability to generate synthetic images with details that are more attuned to the features YoloX L is designed to capture. The Recall metric also sees an increase, but it remains modest, suggesting that while precision is enhanced, the model’s comprehensive detection capability does not scale equally.

**PatchGAN (PS:70)** - When the patch size is increased to 70, PatchGAN’s performance shows a discernible improvement in mAP for both YoloX S and L, peaking at 0.74 with YoloX L. This increase might be attributed to the larger patches capturing more contextual details of the debris, which aligns well with the object detection algorithms. Interestingly, the Recall metric remains relatively steady as compared to the PS:16 variant, potentially highlighting a threshold in the trade-off between patch size and the model’s ability to detect all instances of debris without a significant gain in comprehensive detection.

**ImageGAN** - ImageGAN’s performance in the UATD dataset is quite notable, with mAP scores closely trailing those of PatchGAN (PS:70) and surpassing PixelGAN across all object detection models. Its Recall metrics are consistent with the highest scores observed for other GANs, suggesting that ImageGAN’s approach to image-to-image translation yields a balance between precision and recall. The consistently strong performance across different object detection frameworks may point to the robustness of ImageGAN’s generated images in terms of feature representation and variability.

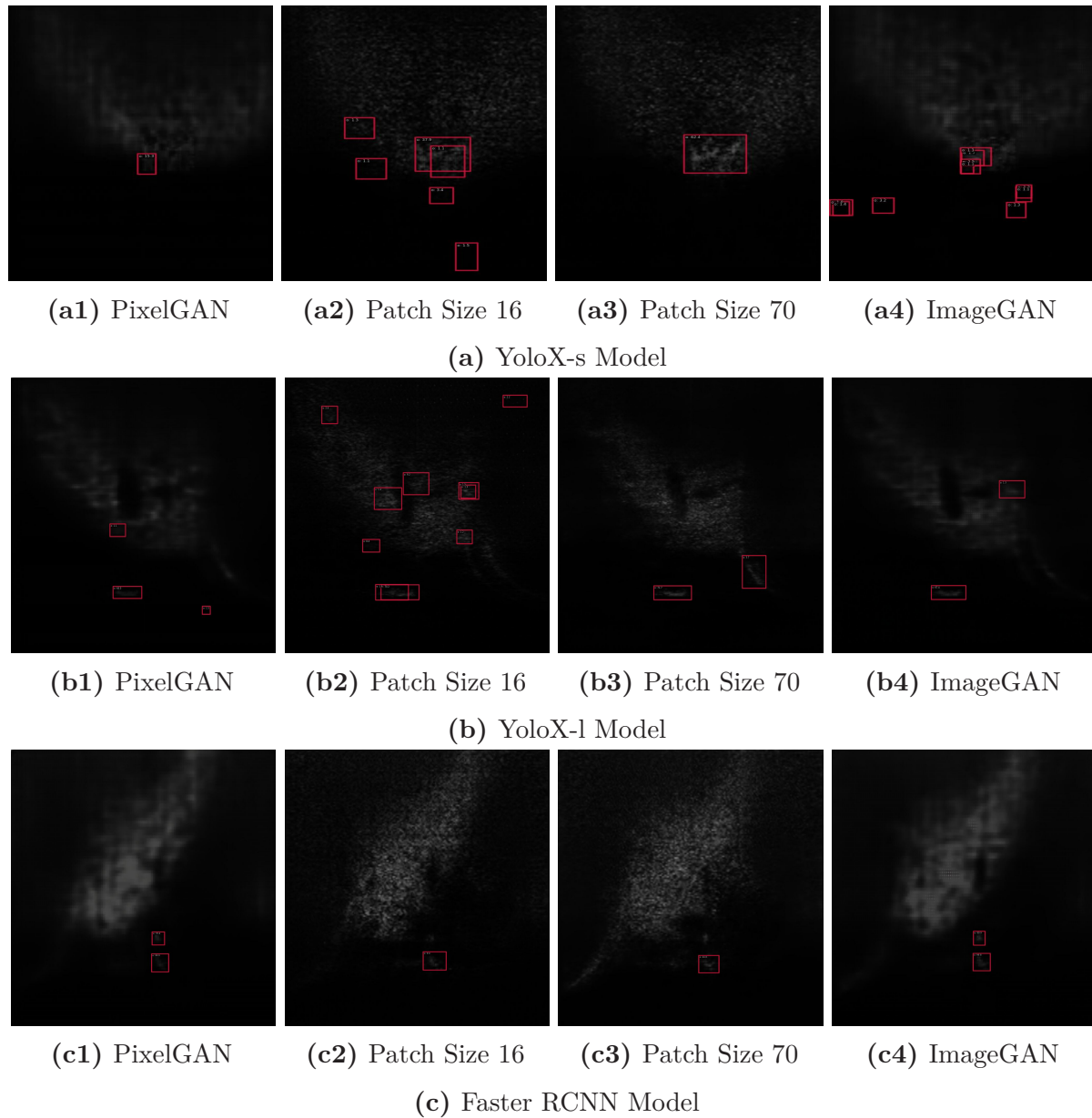
**Visual Analysis of Figure 4.23 Results** The Figure 4.23 offers a qualitative examination of the detection capabilities of various GANs using the YoloX-S, YoloX-L, and Faster R-CNN models on the UATD dataset. Each subfigure presents detection instances

within the underwater acoustic imagery, highlighted by red bounding boxes that signify the identified targets.

**YoloX-s Model (Figure 4.23a)** - The YoloX-s model, designed for speed and agility, reveals intriguing detection patterns across the different GANs. With PixelGAN, the detections are sparse and somewhat isolated, which may align with the lower mAP and Recall scores presented in Table 4.5. It suggests that while PixelGAN is capable of generating detectable features, it may not encompass the full scope of debris types present in the dataset. PatchGAN with a patch size of 16 displays a more concentrated grouping of bounding boxes around certain areas, indicating a higher detection rate which is in line with its higher mAP score. The visual output suggests an effective synthesis of detail that resonates with the YoloX-S model's detection algorithm. Meanwhile, PatchGAN (PS:70) shows even broader detection coverage, which visually supports the highest mAP score reported for this GAN and model combination in Table 4.5. ImageGAN presents a balanced spread of detections, neither as sparse as PixelGAN nor as densely clustered as PatchGAN (PS:16). This intermediate performance visually validates the quantitative data, suggesting that ImageGAN is capable of a consistent detection rate across varied debris representations.

**YoloX-l Model (Figure 4.23b)** - For the YoloX-L model, which is expected to capture more contextual information, the visual results are somewhat consistent with the YoloX-s detections but suggest differences in how each GAN interacts with the larger model's capabilities. PixelGAN again shows fewer detections, possibly reflecting its limited ability to provide the YoloX-l with the necessary contextual cues for detection in this dataset. PatchGAN (PS:16) and (PS:70) both show an increased number of detections with the latter indicating a higher rate of detection across the imagery. This visual evidence is congruent with the quantitative findings where PatchGAN (PS:70) garners the highest mAP with YoloX-l, suggesting that the larger patch size might be more aligned with the broader feature recognition capabilities of YoloX-l. ImageGAN, similar to its performance with YoloX-s, shows a moderate detection pattern, again supporting the mAP and Recall scores that indicate its balanced performance between precision and coverage.

**Faster R-CNN Model (Figure 4.23c)** - The visual results for the Faster R-CNN model add a different dimension to the analysis. Known for its precise detection capabilities, the Faster R-CNN model presents an interesting distribution of detections with each GAN. PixelGAN appears to underperform visually, with very few detections that may represent either a lack of comprehensive feature generation by PixelGAN or a mismatch with the detection algorithm of Faster R-CNN. PatchGAN (PS:16) and (PS:70) both yield a higher number of detections, with PS:70 seemingly providing a greater detection breadth, which could be attributed to the model's ability to utilize the larger contextual cues provided by the bigger patches. This visual observation is in line with the quantitative data, suggesting a stronger performance of PatchGAN (PS:70) in terms of mAP. ImageGAN continues to show a consistent pattern of detections across the board, with a balanced number of bounding boxes that seem to accurately encapsulate target areas, reinforcing its stable performance as reflected in the numerical analysis.



**Figure 4.23:** Comparison of Different GANs using YoloX-s, YoloX-l, and Faster RCNN Detection Model on UATD Dataset. This Figure illustrates the object detection results on a UATD dataset using the YoloX-s, YoloX-l, and Faster RCNN models in conjunction with four different generative adversarial networks (GANs): PixelGAN, PatchGAN with a patch size of 16, PatchGAN with a patch size of 70, and ImageGAN.

## Discussion on the Implications of the Results

### Implications for Marine Debris Detection

The performance metrics and visual results indicate that the choice of GAN significantly affects the object detection model's ability to identify marine debris. PatchGAN, particularly with a patch size of 70, consistently achieved high mAP scores across both datasets, suggesting its potential as a robust tool for generating synthetic training images. This finding is significant for marine debris detection as it implies that employing PatchGAN (PS:70) could enhance the precision of automated detection systems, leading to more effective identification and potential retrieval of debris from marine environments. Moreover, the consistently high Recall scores of ImageGAN across different models hint at its ability to produce diverse and contextually rich synthetic images that aid in comprehensive debris detection. For marine conservation efforts, this suggests that ImageGAN could be particularly useful in scenarios where the detection of all debris, irrespective of size and type, is critical.

### Implications for UATD Datasets

The UATD datasets pose unique challenges due to the acoustic nature of the imagery, which often lacks the visual clarity of optical images. The results demonstrate that GANs capable of capturing finer details and contextual information, such as PatchGAN (PS:70), can significantly improve object detection performance in UATD scenarios. The higher mAP scores suggest that PatchGAN (PS:70) can enhance the model's ability to discern targets amidst the noisy background inherent to acoustic imaging. Furthermore, the consistent performance of ImageGAN in both marine debris and UATD datasets reinforces its utility across different environmental conditions and detection frameworks. This consistency is crucial for developing versatile detection systems capable of operating in diverse underwater conditions.

### Cross-Contextual Implications

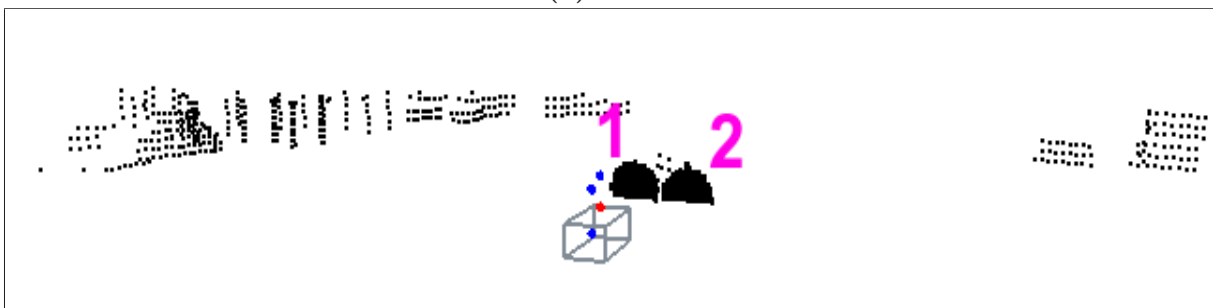
The cross-analysis of the results from both marine debris and UATD datasets reveals a compelling trend: while certain GANs are universally effective, the specific characteristics of the dataset significantly influence the performance of different GANs and object detection model combinations. This insight is vital for the development of tailored detection systems that can adapt to the characteristics of the target dataset. For instance, the higher precision of PatchGAN (PS:70) in UATD scenarios suggests that models requiring high precision in noisy conditions might benefit from training on images with larger patches. Conversely, the balanced Recall of ImageGAN in marine debris detection implies that scenarios requiring a broader detection scope might favor the use of ImageGAN-generated images.

## 4.2.5 Simulated Sensory Data

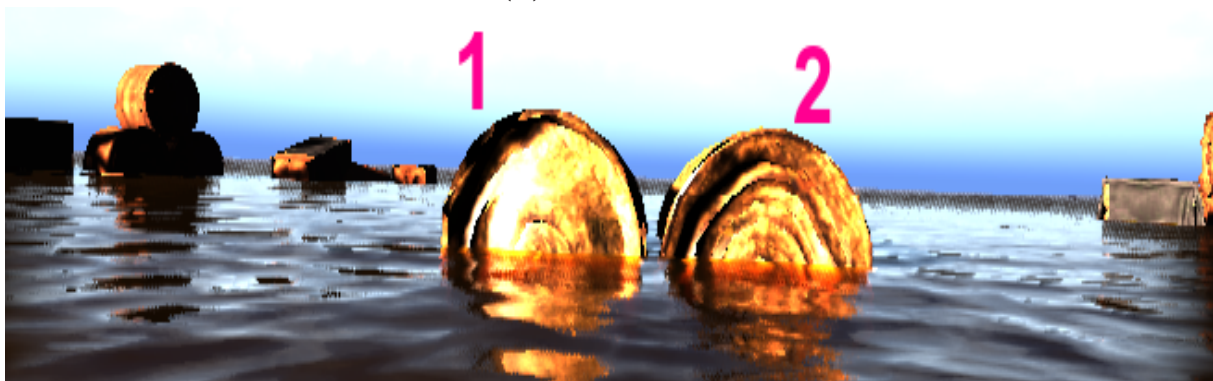
In the domains of autonomous navigation and environmental mapping, particularly within the context of flooded environments, the synthesis of multi-sensor data is essential for creating accurate representations of scenario. To that end, the exploration of varied scenarios within such environments provides invaluable insights into the challenges and potential applications for this technology. By employing a comprehensive simulation that mimics a flood-affected barnyard scenario, the adaptability and robustness of multi-sensor



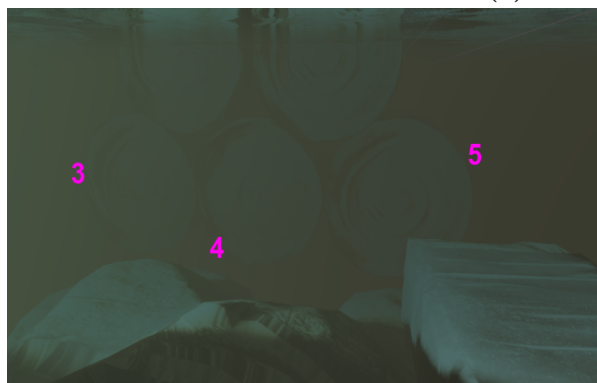
(a) Scene



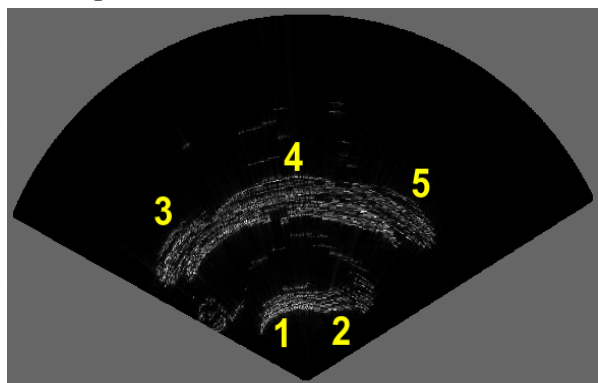
(b) LiDAR Scan



(c) Camera Image



(d) Underwater Scene



(e) Sonar Image

**Figure 4.24:** Simulation Data - Scenario 1. The figure illustrates a stack of hay bayles. LiDAR data shows two of them (1 and 2) in point cloud and sonar data shows all five because some part of top two hay bayles is within the water.



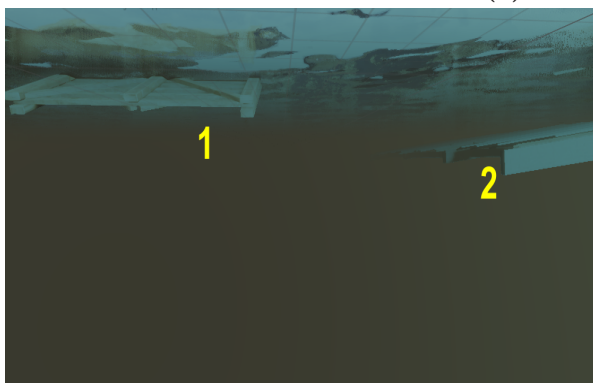
(a) Scene



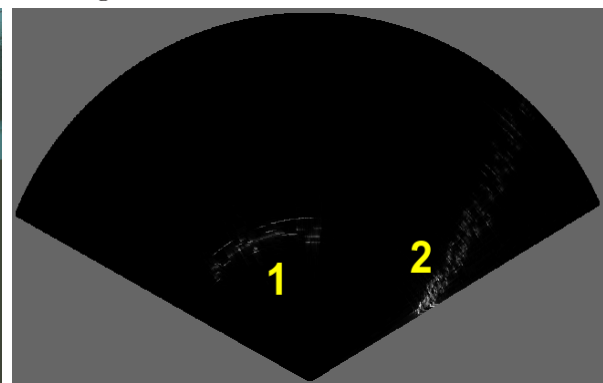
(b) LiDAR Scan



(c) Camera Image



(d) Underwater Scene

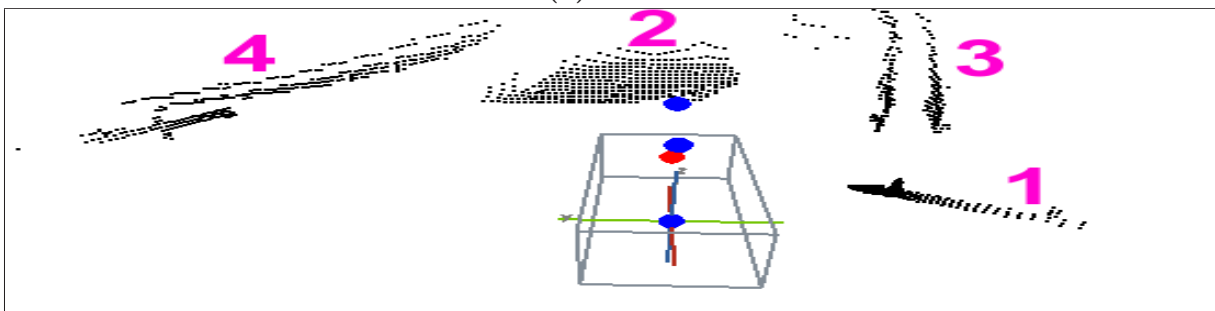


(e) Sonar Image

**Figure 4.25:** Simulation Data - Scenario 2. The figure illustrates two floating wooden fences. Fence (1) is half submerged and therefore partly visible in LiDAR and sonar data. Fence (2) has more weight therefore, is more underwater the water. Hence, the LiDAR data shows only few points.



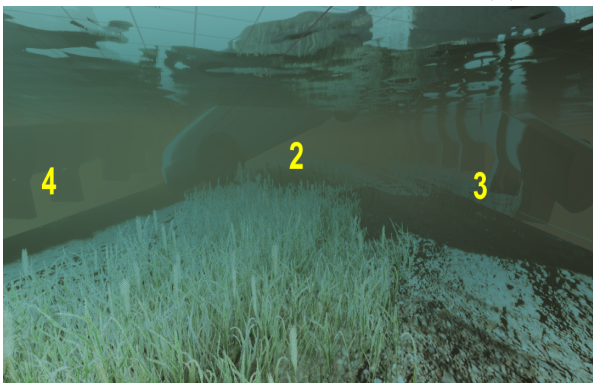
(a) Scene



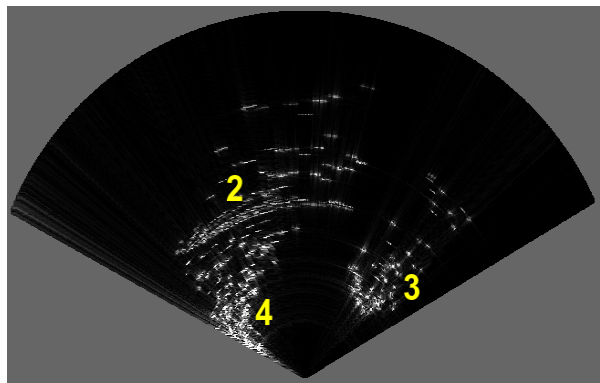
(b) LiDAR Scan



(c) Camera Image



(d) Underwater Scene



(e) Sonar Image

**Figure 4.26:** Simulation Data - Scenario 3. The figure illustrates the submerged car over the fenced road. Sonar data shows high reflections from car and road boundaries due to similar reflection from both surfaces.



data fusion in mapping complex environments can be understandable. For example, in the first scenario depicted in Figure 4.24, a simulation of a stack of hay bale, arranged in two-on-three fashion, provides an interesting LiDAR and sonar data captured above-water and submerged objects respectively. LiDAR data renders the hay bales that are partially above water, whereas the sonar image extends this representation to include the submerged portions, showcasing the complementary nature of these sensor technologies.

Moreover, the second scenario (shown in Figure 4.25) introduces dynamic floating objects, in this case, wooden fences, to illustrate how sensors respond to objects that exhibit vertical movement due to water currents. Here, the distinction between data captured by LiDAR and sonar becomes particularly important; the former delineates the portion of the fence above water, while the latter offers a complete image, capturing the submerged segments. This duality of perception underscores the value of integrated sensor systems for comprehensive environmental mapping.

The third scenario (shown in Figure 4.26) increases in complexity, featuring a partially submerged vehicle on a flooded roadway. The sonar data in this scenario reveals high reflectivity from both the road's surface and the lateral fences, demonstrating the sensor's sensitivity to different materials and angles of reflection. It highlights a common issue in underwater detection: the challenge of distinguishing objects among reflective interference from the environment. Such scenarios serve not only to test and validate the performance of the sensors but also to advance our understanding of their limitations and the need for sophisticated algorithms capable of processing complex sonar signatures.

Across these scenarios, the fusion of multi-sensor data is evident, as each sensory modality brings its strengths to bear on the task of mapping. The simulated environment acts as a testbed for methodologies that will later be applied to real-world settings, where the accuracy and reliability of these systems are of foremost importance. The data produced by this simulated environment will be instrumental in the forthcoming chapter, which delves into mapping methodologies for flood-impacted environments. Through such simulations, the approaches defined in upcoming chapters can be further refined, ensuring that when deployed in actual flooded environments, the resultant maps are both accurate and actionable, thereby enhancing safety and operational efficiency in aquatic navigation and disaster response efforts.

## 4.3 Discussion

The rigorous evaluation of any framework requires an extensive testing environment and diverse scenarios. Therefore, this chapter discusses several datasets and simulated environments ranging between both surface water and underwater domains. An important and relevant contribution, in this chapter, is the novel Multi-modal Shallow Water (MASTER) dataset. The dataset is rich with diverse surface and underwater modalities, encompassing LiDAR point clouds, stereo camera, and sonar imagery. This dataset is further enhanced by the inclusion of bounding box labels and segmented masks, making it an invaluable contribution for the fine-tuning and validation of detection and segmentation algorithms.

The MASTER dataset contributes significantly to aquatic and post-flood application research. Although providing abundant data for analysis over simpler post-flood environment, it does not include complex scenarios. To bridge this gap, this chapter introduces an

extensive simulation developed within the Unreal Engine framework. The simulation replicates the complex dynamics of post-flood scenarios span from barnyard settings to complex urban environment, with remarkable detail, offering a test bench for the methodologies discussed herein.

Moreover, this chapter presents another breakthrough achievement: simulating underwater sonar data. It addresses the acoustic challenges presented while emulating sonar imagery in the Unreal Engine by a plugin based on the Generative Adversarial Networks (GANs) principles and demonstrating more than 80% detection accuracy. The training of the network is performed by using both the Marine Debris and UATD datasets to train the Pix2pix GAN at different patch sizes, of which PatchGAN 16 and PatchGAN 70 performed excellently, respectively, hence validating the plugin.

The sonar plugin, while ample for a specific sonar device used in this study, requires very large scale datasets for learning features. This has always been one of the problems in the underwater acoustic domain. Therefore, alternative approaches that rely less on extensive data are essentially required, such as a more complex simulation of acoustic wave physics. Additionally, the scope of this thesis is defined by a selection of critical scenarios designed to emulate the conditions of a flooded environment. These scenarios serve useful for preliminary exploration, indicating the need for a broader spectrum of complex scenarios to ensure a robust evaluation of any proposed system within the challenging and variable conditions of post-flood environments.

## 5. Underwater Traversability

Flooding often results in the damage or destruction of critical infrastructure including bridges and roads. Such infrastructure are essential for transportation and connectivity in any area. Moreover, the floodwaters carry and deposit random objects into the affected environment. These objects include debris, fallen trees, and various items swept away by the floodwaters. The combined impact of these factors is transforming the affected area into a highly unstructured and unpredictable environment. In such conditions, traditional mapping solutions, including well-known platforms like Google Maps and OpenStreetMap, prove ineffective. These conventional maps are designed for structured and stable environments and rely on pre-existing data and infrastructure. In post-flood environment, this data becomes obsolete due to the destruction and environmental changes. As a result, rescue teams and disaster responders operating in post-flood scenarios face a significant and pressing challenge: they need access to accurate and up-to-date maps of the affected areas. Without reliable mapping data, planning and executing safe navigation and effective disaster response becomes extremely difficult. This highlights the importance of developing alternative mapping strategies that can adapt to the unpredictability and lack of structure inherent to such situations, ultimately aiding rescue and recovery efforts.

Modern sensory modalities come to the forefront to overcome the mapping challenges posed by flooding events. Satellite imagery, aerial drones, and remote sensing modalities are highlighted as pivotal tools for rapidly generating updated maps of post-flood environments. These modalities can produce high-resolution maps that provide critical information about the extent of flooding, infrastructure conditions, and the locations of crucial facilities like hospitals and shelters. However, these sensory modalities have limitations, particularly when understanding the underwater aspects of post-flood environments. Factors such as variable water levels and high turbidity due to sand particles can limit the effectiveness of satellite and drone imagery in accurately mapping the underwater environment.

Given these limitations, there is a pressing need to address the challenges associated with underwater mapping in post-flood scenarios. Improving our ability to map the underwater environment is a crucial area of focus for enhancing disaster response and rescue efforts in the aftermath of flooding events.

## 5.1 Challenges in Underwater Sensing

Underwater sensing poses unique challenges due to water's opaque nature. Unlike air or land, water does not readily transmit light, making it difficult for optical sensors such as cameras or LiDARs to operate efficiently in aquatic environments. The interaction between water molecules and suspended particles leads to optical phenomena such as scattering, absorption, and visibility loss. Therefore, traditional cameras are ill-equipped to capture clear underwater images, limiting visual clarity. This makes it challenging to gather accurate and detailed visual data, particularly in deep water or areas with high levels of turbidity or sediment. Hence, in contrast to terrestrial or aerial sensing, only a limited number of feasible options are available for underwater sensing.

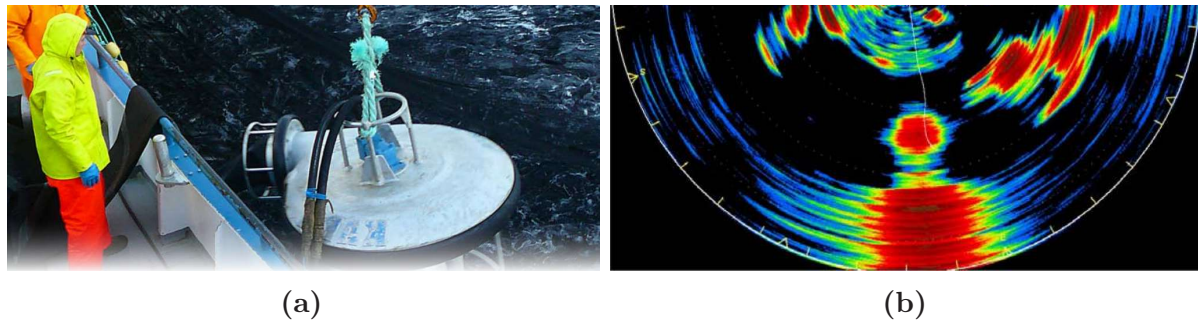
Acoustic sensors are optimal for detecting objects submerged in water due to the better transmission of sound waves through liquids compared to air or solids. Amongst such sensors, the Sonar (Sound Navigation and Ranging) stands out for its ability to emit and receive sound waves, enabling precise detection and identification of objects, underwater topography mapping, and distance calculation. Such features make Sonar highly effective in underwater environments.

### 5.1.1 Comparative Analysis of Commercial Sonars

There are several types of sonars commercially available as per requirement. The selection is purely based on several specifications such as resolution, field of view (FOV), range, and application. All of these specifications are based on one core factor which is the operating frequency of the sonar. Sonars are typically categorized into three types based on their operating frequency.

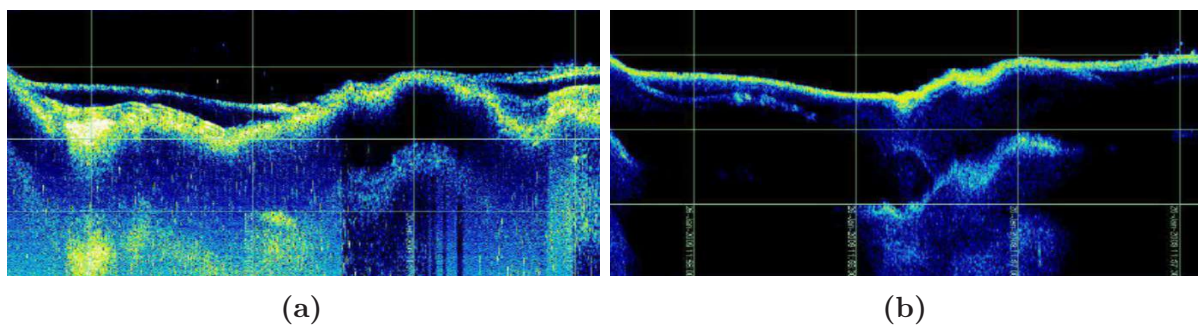
**Low-frequency Sonar:** Low-frequency sonars, typically working within the frequency range of 1 kHz (kilohertz) to 100 kHz, offer specific advantages and limitations. The lower operating frequency allows these sonars to achieve extended detection ranges, making them particularly well-suited for deep-water applications. This is because lower frequencies can propagate through water more effectively and are less prone to rapid attenuation at depth. Consequently, low-frequency sonars are commonly employed in deep-water exploration scenarios. However, there are trade-offs associated with low-frequency sonars as low frequency results in wider beam width. While this is advantageous for long-range detection, it diminishes the resolution of the resulting sonar imagery. In other words, the imagery produced by low-frequency sonars tends to be less detailed and may not provide fine-grained information about underwater objects. Due to their characteristics, low-frequency sonars find their primary utility in detecting large objects from considerable distances. These applications include the detection of substantial objects such as ultra-quiet diesel submarines, shipwrecks, and exploration of deep-water environments. The ability to cover a wide area and detect large objects at extended ranges makes them a valuable tool in scenarios where high resolution is not a critical factor, and the focus is on the detection of significant underwater features or structures.

One typical example of commercially available low frequency sonar is Simrad SX90 which has 20 to 30 kHz of operating frequency with 50-4500 meters range. Figure 5.1b shows the detection of a cargo vessel at depth of about 100 meters.



**Figure 5.1:** Detection of cargo vessel using Simrad SX90. (a) The part of sunken 35k ton cargo vessel. The wreck lays at about 100 meters depth. (b) The sonar data of the wreck. The sonar is operated at 23 kHz using "FM Auto" and "Narrow Beam".

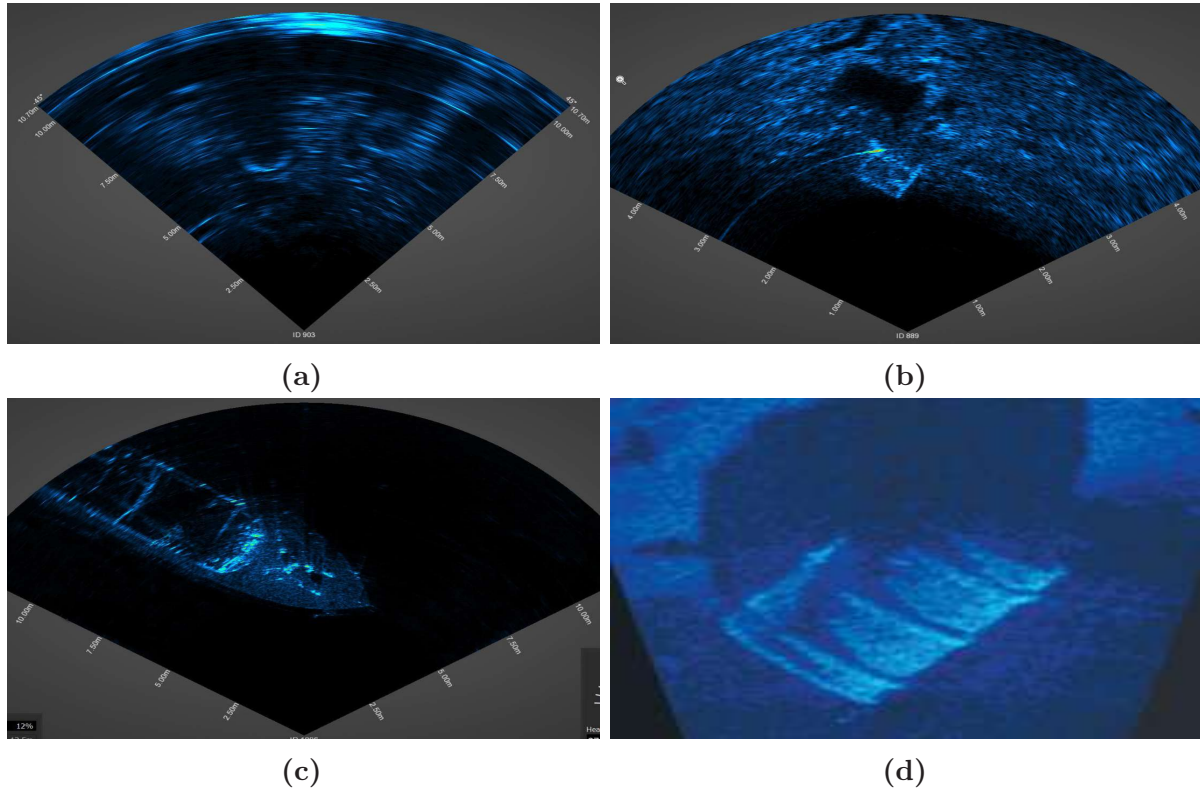
**Mid-frequency Sonar:** Medium-frequency sonars are powerful and versatile sensors that operate within the frequency range of 100 kHz to 500 kHz. With extended operational range and enhanced image resolution, they prove to be incredibly valuable in various fields. One such application is in fisheries, where they offer great assistance in locating and tracking schools of fish. The balance between range and resolution makes medium-frequency sonar highly effective in discerning fish populations beneath the water's surface. Additionally, these sonars play a critical role in underwater exploration and seafloor mapping. The SeaKing sub-bottom profiler by Tritech is a sonar that uses two different frequencies, 20kHz (shown in figure 5.2a) and 200kHz (shown in figure 5.2b), to map the sea floor. With its 20kHz pulse, the system can even penetrate the seabed and detect structural differences that cannot be seen by conventional echosounders.



**Figure 5.2:** Dual Frequency SeaKing-Bottom profiler. SeaKing Sub-Bottom profiler provides an accurate measurement of the seabed depth and its underlying structure. The profiler uses dual-frequency sound waves to create temporal imagery of the seabed. (a) When operating at a frequency of 20kHz, the imagery reveals the nature of the material beneath the seabed as low-frequency sound waves easily penetrate into the seabed. (b) On the other hand, at 200kHz, the imagery shows a clearer picture of the seabed's surface. This is because most of the sound waves reflect from the seabed surface due to the high frequency of operation.

**High-frequency Sonar:** High-frequency sonars typically operate at frequencies above 500 kHz. They are specifically designed to produce high-resolution images and are commonly used in shallow-water environments. High-frequency sonar is ideal for tasks that require detailed imaging, such as object recognition, mine detection, and

marine biology research. The high-frequency sound waves they produce result in clear and sharp images of underwater objects, making them invaluable for tasks that require precision. Figure 5.3 presents the qualitative analysis of images captured by different sonars. It is worth nothing that with the increase in frequency, the resolution of images significantly enhances.



**Figure 5.3:** Multibeam Imaging sonars with different operating frequencies and beamwidths (a) Mikron Gemini works over 720kHz with 2.34° of beamwidth. The range of the sensor ranges from 0.2m to 50m. (b) Gemini 720ik works over 720kHz with 1° of beamwidth. The range of the sensor ranges from 0.5m to 120m. (c) Similarly, Gemini 1200ik is a dual-frequency sonar (720kHz and 1200kHz) with upto 0.6° of beamwidth. The range of the sensor ranges from 0.1m to 120m. (d) ARIS Explorer 3000 works over 3MHz frequency with 0.25° of beamwidth.

**Selection Criteria:** This research has chosen the Tritech Gemini 720ik sonar system for its high-frequency capabilities. This decision was based on several reasons, one of which is the specific requirements of the application. The system needs to detect obstacles in very shallow waterbeds, which low-frequency sonars are not suitable for due to their inherent characteristics. Low-frequency sound waves generate wider beamwidths, which makes low-frequency sonars apt for detecting large objects from substantial distances. However, this characteristic is not ideal for detecting small and randomly positioned objects in flooded environments.

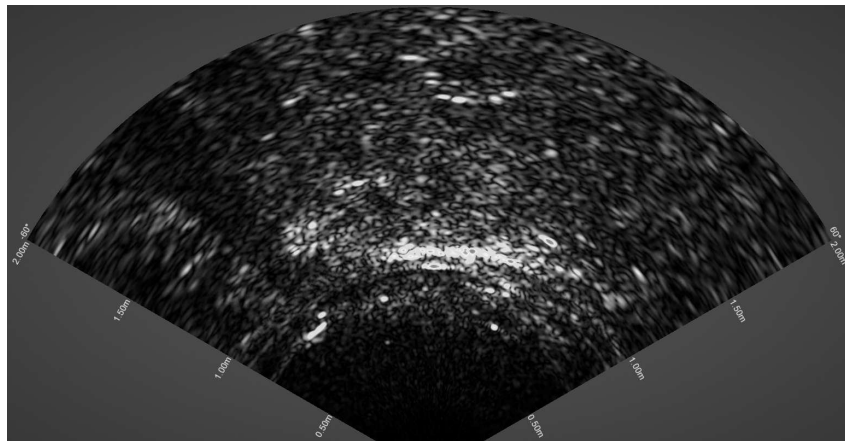
Conversely, mid-frequency sonars are often deployed for mapping the composition of the waterbed but lack the capability required for detecting small, unstructured objects in shallow waters. High-frequency sonars, on the other hand, are preferred for applications in shallow water environments, primarily due to their superior resolution and better object detection capabilities. The Tritech Gemini 720ik sonar,

in particular, has a  $1^\circ$  beamwidth, with horizontal field of view (HFOV) and vertical field of view (VFOV) spanning  $120^\circ$  and  $20^\circ$ , respectively. Additionally, this particular sonar is a cost-effective solution in comparison to higher-frequency sonars.

For a better understanding, an example image captured by the chosen sonar is presented in figure 5.3b. The characteristics and capabilities of the high-frequency Tritech Gemini 720ik sonar align seamlessly with the objectives of our research, making it an optimal choice for the detection of objects in the specific environmental conditions in very shallow waterbeds.

### 5.1.2 Issues in Sonar Imagery

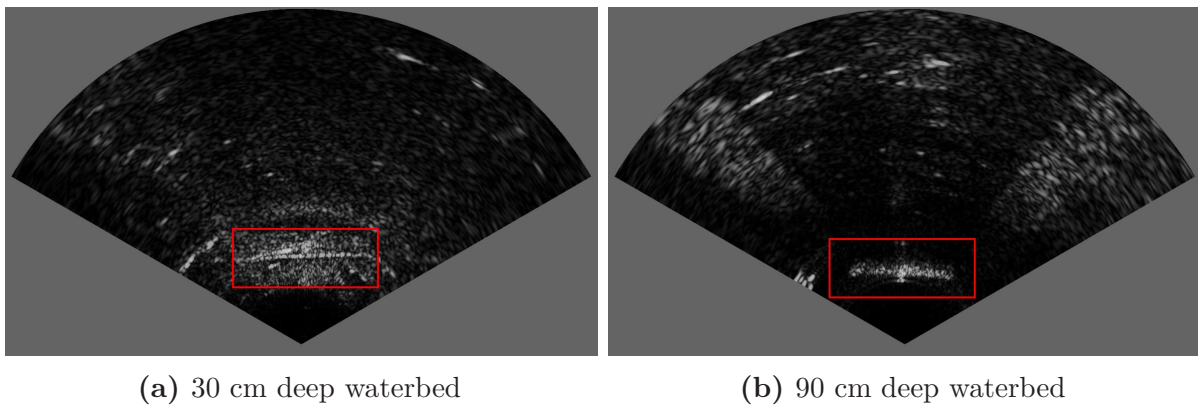
Standard sonar imagery faces several inherent issues that require careful consideration prior to its application in underwater mapping. Furthermore, the post-flood environment introduces additional complexities and challenges specific to sonar imagery, which are thoroughly discussed in the following sections.



**Figure 5.4:** Speckle Noise in real sonar imagery. The figure demonstrates speckle noise and reverberations generated within underwater sonar imagery due to an object and waterbed. The image shows an aluminium object 110 cm away from sonar housing within 70 cm deep water.

**Speckle Noise** One of the common challenges associated with sonar images is the presence of speckle noise [Chaillan 07]. Speckle noise manifests as granular and high-contrast patterns that can degrade fine details and make accurate interpretation challenging as shown in Figure 5.4. It is caused by the interference of the coherent acoustic waves reflected from different underwater objects, resulting in constructive and destructive interference patterns. This noise phenomenon is particularly pronounced in medical ultrasound, underwater navigation, and underwater mapping applications where the coherent nature of sonar signals is common. Speckle noise can significantly inhibit the ability to detect and identify objects of interest, such as underwater terrain features or unsafe obstacles. Several methods have been developed to reduce speckle noise in sonar images, including filtering techniques and statistical models. The choice of method depends on the specific application and the desired level of noise reduction. However, effectively reducing speckle noise while preserving essential image features remains a complex challenge.

**Low Obstacle-to-Background Contrast** In a very shallow water environment, when obstacles are in the same range as the waterbed, it becomes challenging to distinguish between the two. It is because the waterbed of an extremely shallow environment reflects high-intensity echos [Yuan 21], which decrease the obstacle-to-background contrast and lead to difficulty for sonar systems in detecting and identifying obstacles. This, in turn, increases the risk of collisions and navigation errors. In figure 5.5b, for instance, the waterbed is about 1 meter deep, meaning there is no echo from the waterbed at ranges shorter than 1 meter. This absence of background noise or reflection makes detecting obstacles within this range easier. However, in figure 5.5a, the object is in range with the waterbed, and there is low contrast between the two.

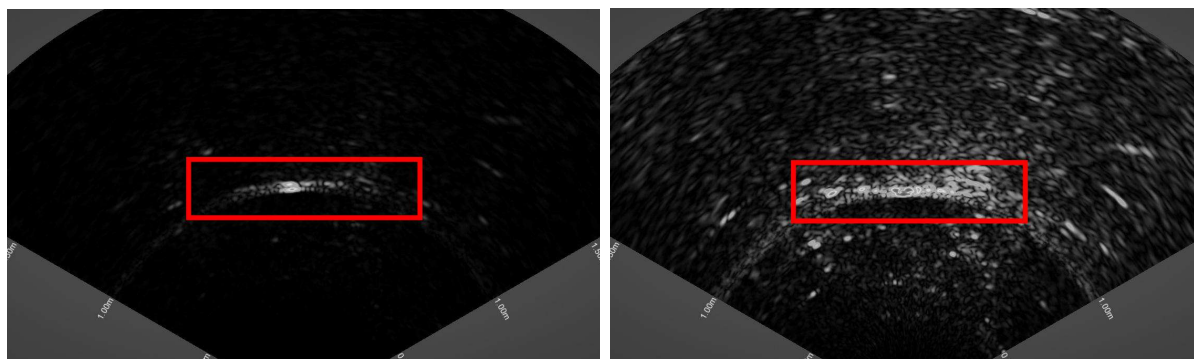


**Figure 5.5:** Comparison of sonar imagery at different waterbed depth level. The figure shows visual comparison of sonar imagery, showcasing an obstacle situated 50cm from the sonar housing, illustrating the effect of different waterbed depths. In this comparison, it is evident that reflections from the waterbed can diminish the obstacle-to-background contrast. This reduction in contrast adversely impacts the ability to detect obstacles underwater.

**Turbidity and Low Signal-to-Noise Ratio** The sound waves also suffer high-intensity echoes from the suspended particles within the water environment. Therefore, in a flooded environment, high amounts of suspended materials, such as sediment, silt, debris, and organic matter, affect the visibility and performance of sonar. This results in difficulty in determining the water depth and location of obstacles, consequently affecting the safety and efficiency of operations. The high-intensity echos can be reduced at the sensor level by reducing the sensor gain. However, the low signal-to-noise ratio of sonar imagery [Caimi 08] also reduces the obstacle intensities. Figure 5.6 shows two sonar images with the same object at the same range (1 meter) from sonar but at different gains. The low gain reduces waterbed reflections, but also weakens obstacle intensity. In contrast, high gain makes obstacle detection difficult.

**Unstructured Random Obstacles** In flooded environments, unstructured random obstacles can significantly challenge navigation and mobility. These obstacles may include debris, vegetation, rocks, and other objects randomly distributed throughout the flooded area. These unstructured and random obstacles manifest as irregular patterns with varying sizes, shapes, and acoustic characteristics. They disrupt the safe movement of underwater vehicles and introduce considerable uncertainty in mapping and object recognition tasks. However, the random distribution of obstacles





(a) Gain level at 60

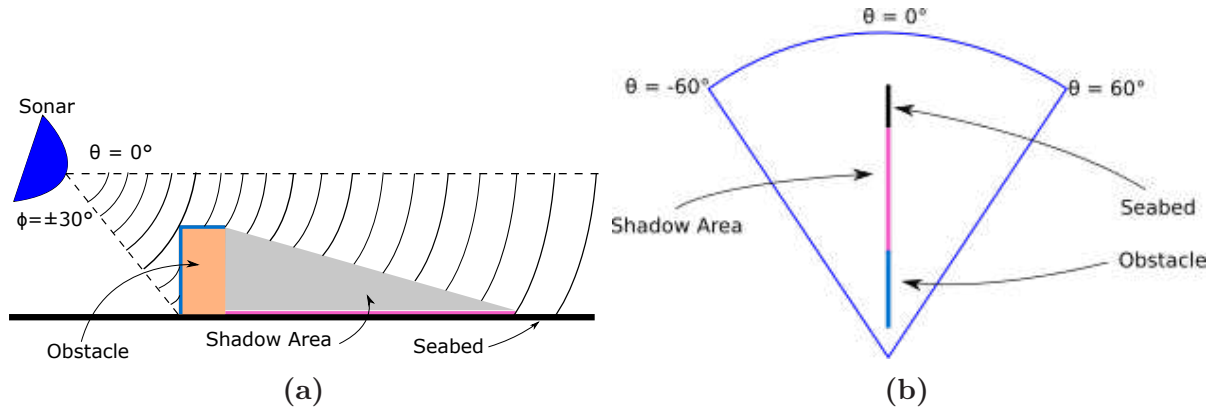
(b) Gain level at 95

**Figure 5.6:** Comparison of sonar imagery at different gain level. The figure displays an obstacle positioned 1 meter from the sonar, demonstrating the impact of varying gain levels. At a lower gain setting, there is a noticeable reduction in undesirable reflections from the seabed, but this also leads to a diminished reflection from the obstacle, potentially impairing its detection. Conversely, a higher gain setting amplifies the reflections from the obstacle, enhancing its visibility. However, this increased gain also results in a heightened reflection from the seabed. *The gain level ranges between 1-100*

can make it difficult to determine the dimension of an object and require frequent parametric adjustments to avoid collisions.

**Limited Depth Information** The absence of depth information in forward-looking sonar images constitutes a significant challenge in underwater imaging and navigation. Unlike traditional profiler, that provide depth information by sending a focused sound wave directly toward the waterbed and measuring the depth by calculating the time it takes for sound waves to bounce off objects and return to the sensor, forward-looking sonar primarily produces two-dimensional images without depth information as illustrated in Figure 5.7. This limitation inhibits the viewer’s ability to measure the depth of objects and accurately perceive the three-dimensional structure of the underwater environment. Additionally, due to the  $20^\circ$  VFOV, the depth uncertainty of object increases with range as shown in Table 5.1. The lack of depth information can lead to difficulties in object localization, obstacle avoidance, and overall scene understanding, which are crucial for safe and effective underwater operations. Overcoming the missing depth information remains a critical focus area in underwater imaging and robotics to ensure safer and more reliable operations in submerged environments.

**Unavailability of Baseline Benchmark** Standardized datasets and performance metrics, known as benchmarks, are crucial in developing and evaluating algorithms and systems for various applications. However, the lack of standard benchmarks in sonar imaging for underwater environments, specifically for post-flood scenario, it is difficult for to assess and compare the algorithms and robotic systems effectively. Overcoming this challenge requires collaboration among the scientific community to collect, annotate, and share benchmark datasets that reflect the diversity and complexity of underwater environments. These benchmarks would not only accelerate



**Figure 5.7:** Physical principle of sonar 2D imagery. The figure demonstrates the mapping of 3D underwater environment into 2.5D multibeam imaging sonar image. (a) Each beam of sonar has 30 deg vfov. The arrows are only for visualization purpose. In reality the wavepattern is generated which strikes at each point and return its echo. (b) The HFOV of sonar is 120 deg and the figure elaborates the mechanism of mapping at 0 deg.

innovation but also ensure the robustness and reliability of sonar-based systems in flooded environments.

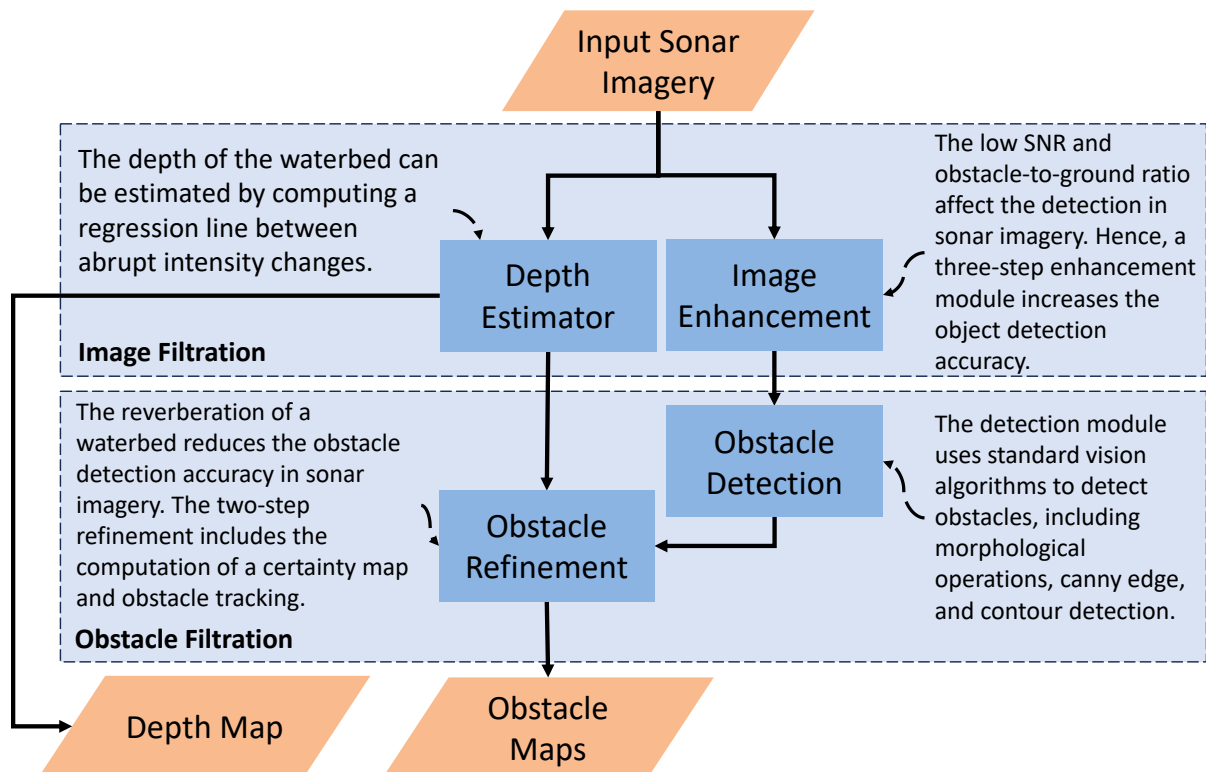
**Table 5.1:** Range vs Height Uncertainty of obstacle. The table illustrates the correlation between the distance of an obstacle and the uncertainty in its height due to unavailability of z-axis in sonar imagery.

| Obstacle Range<br>(cm) | Height Uncertainty<br>(cm) |
|------------------------|----------------------------|
| 50                     | 18.19                      |
| 80                     | 29.12                      |
| 120                    | 43.68                      |
| 200                    | 72.80                      |
| 400                    | 145.56                     |

## 5.2 Underwater Traversability Mapping Framework

The Underwater Traversability Mapping (UTM) framework constitutes an essential component within the SWiM architecture. It serves multiple functionalities aimed at delivering a comprehensive underwater obstacle detection system and enhancing the precision of underwater mapping. Several advanced techniques are available in the literature that are designed for object detection in underwater environment. These techniques prove to be invaluable in scenarios characterized by relatively clearer water but serve as a promising foundation for the development of improved object detection systems designed for the complexities of post-flooded environments. For instance, Mingzhe et al. [Wei 21] introduced a notable approach that combines the Particle Swarm Optimization (PSO) algorithm with Mutual Information (MI) for image registration. This methodology is implemented to identify sensitive regions within underwater images, thereby enhancing the precision of

object detection. Likewise, Long et al. [Long 23] explore the application of deep learning architectures, including Yolov3 and Faster R-CNN, to detect underwater objects. These techniques show promising results, but it is important to note that they were evaluated using the high-resolution Aris 3000 sonar imagery, which may not be available in all scenarios. Additionally, Henson et al. [Henson 19] present a similar approach involving image registration for seabed mapping, primarily relying on the Didson 300 system, a high-frequency sonar with exceptionally high resolution. Consequently, these techniques may not be as effective when applied to low-resolution images obtained from other sonar systems, such as the Tritech Gemini 720ik sonar.



**Figure 5.8:** Workflow of the Underwater Traversability Mapping (UTM) framework. This block diagram illustrates the comprehensive process flow of the UTM, designed to enhance sonar imagery for improved detection and characterization of underwater obstacles.

The Underwater Traversability Mapping (UTM) framework in SWiM (detailed illustration in Figure 5.8) presents a depth estimation technique which is a significant feature that accurately measures waterbed depth. This technique ensures a high degree of precision in depth estimation, which is fundamental for mapping underwater environments. Furthermore, to achieve high precision in underwater mapping, the UTM processes data using image filtration techniques that rectify common issues encountered in underwater imaging. These techniques include noise reduction and image enhancement by eliminating unwanted waterbed reflections, which can significantly improve the quality of sonar images.

Once the sonar imagery is enhanced, the UTM's object detection algorithms identify and categorize obstacles in the underwater environment based on multiple criteria, including reflection coefficients and the presence of shadows. The framework introduces a certainty map, which calculates detection probabilities depending on the depth and range of the

object. This strategy enhances the accuracy of obstacle detection within sonar imagery, leading to more reliable and precise underwater mapping.

### 5.2.1 Image Filtration

Sound waves are prone to several challenges (presented in section 5.1.2) that can significantly compromise the quality of sonar imagery, consequently diminishing the likelihood of successfully detecting objects within the architectural framework. As such, the initial and fundamental step in any sonar-based architecture revolves around image filtration. The next section deals into the strategies employed for image enhancement, offering a comprehensive evaluation of these techniques one-on-one with several state-of-the-art methodologies. Moreover, the results show that these enhancements present better noise reduction and object shape, ultimately amplifying the architecture's object detection capabilities.

Furthermore, a novel depth estimation algorithm is presented specifically to compute the depth of the waterbed using the forward-looking sonar system. This algorithm proves to be particularly advantageous, as it helps the architecture to accurately estimate the depth of waterbed. The fusion of image enhancement strategies and the depth detection algorithm lays the foundation for a more robust and effective sonar-based architecture, enhancing the overall object detection probability within the system. This work represents a substantial advancement in the field of underwater object detection and contributes to improving the reliability and precision of object detection within sonar imagery.

#### 5.2.1.1 Image Enhancement

Image enhancement is a fundamental technique in the field of digital image processing that involves the use of advanced algorithms and techniques over an image to improve its visual quality for human perception or machine analysis. The motivation for image enhancement arises from the need to improve the quality of images that are degraded due to factors, for instance, poor lighting, noise, or other environmental factors. Furthermore, image enhancement is essential in various applications such as underwater mapping, medical imaging, remote sensing, and surveillance, where the quality of image data is critical for decision making. The standard process of image enhancement is comprised of a series of steps including denoising, contrast adjustment, and sharpening, which aim to correct image degradation and optimize specific visual features.

Image enhancement is of paramount importance in the context of forward-looking sonar (FLS) imagery due to the unique challenges presented in section 5.1.2. FLS imagery is integral to various maritime applications which operate in shallow and murky waters, but is prone to issues such as signal attenuation, noise, and low visibility, resulting in poor image quality. Image enhancement techniques, such as filtering, contrast enhancement, and noise reduction, can improve the visibility of objects and structures in the underwater environment. For instance, filtering techniques can remove unwanted noise and artifacts from the image, while contrast enhancement can improve the visibility of objects and structures in the underwater environment. Furthermore, noise reduction techniques can help enhance the signal-to-noise ratio, resulting in better image quality. This in turn, can significantly enhance situational awareness, aiding in navigation, object detection, and search and rescue operations.

Numerous denoising techniques for sonar imagery have been documented in the existing literature. Jia et al. [Jia 19] have introduced a spatial domain image enhancement method utilizing a 3x3 Gaussian smoothing filter to reduce the speckle noise. Additionally, they have computed three gray-scale regions based on statistical features. However, this approach exhibits enhanced segmentation performance when applied to high-resolution Synthetic Aperture Sonar (SAS) images and its effectiveness diminishes when employed on low-resolution forward-looking sonar images, which is an inherent limitation of this approach. Similarly, [Choi 20a] has implemented a wavelet denoising approach for ultrasound images in the context of medical imaging. While this technique effectively reduces noise, it has an adverse impact on image contrast, subsequently leading to a decline in obstacle detection accuracy. Huang et. al. [Yuan 21] devised a methodology that combined dictionary learning with wavelet transform to filter noise from sonar images, while Oliveira et al. [Oliveira 21] conducted a comprehensive comparative study evaluating the SURF, ORB, BRISK, and SURF-Harris algorithms on sonar images. These studies predominantly focused on static and easily identifiable input images, which might not necessarily reflect real-world underwater scenarios. The authors in [Choi 20a] demonstrated that noise reduction achieved considerable results when the sonar was in a static position (lock-on phase). However, it introduced a blurring effect in the resultant image when the sonar was in motion (searching phase) due to the successive averaging of input images. Furthermore, the works in [Chen 19] and [Chen 21] exhibited effective noise reduction in deep-water environments but was less suitable for shallow-water conditions. The above-mentioned denoising techniques highlights the diversity of approaches to reduce noise in sonar imagery, each with its specific strengths and limitations, necessitating careful consideration of the particular characteristics and requirements of the application at hand.

The image enhancement techniques presented in proposed SWiM architecture serves the purpose of noise reduction in sonar images, amplifying the contrast between objects and their backgrounds, and segmenting potential obstacles to facilitate subsequent object detection processes. This research distinguishes itself from existing literature through the following noteworthy distinctions:

**Detection Challenges in Shallow Water Detection:** The detection of obstacles in shallow water environments, characterized by depths of less than 2 meters, is particularly complex due to the several critical issues. These include the substantial reflections and reverberations of sonar waves from the waterbed, which reduce the object to background contrast, hence significantly complicate the detection process. The work at hand acknowledges and addresses these challenges, striving to enhance obstacle detection under such demanding conditions.

**Low-Resolution Sonar Imaging:** In contrast to the high-resolution images often considered in the majority of the literature, the forward-looking imaging sonar typically used for obstacle detection provide images with lower spatial resolution. This specific characteristic necessitates the development of specialized image enhancement techniques according to the unique features and constraints of low-resolution sonar imagery. As such, this research takes into account the distinct requirements posed by low-resolution sonar imaging, which is essential in effectively addressing the real-world conditions and limitations encountered in practical underwater applications.

Sonar images are subject to abrupt variations in intensity due to reflections from the waterbed, resulting in the emergence of a granular pattern known as *Speckle noise* (as discussed in Section 5.1.2). It is essential to recognize that speckle noise is a multiplicative form of noise, therefore requiring a logarithmic transformation to convert it into additive Gaussian noise, as illustrated by the equations below:

$$\hat{y} = \log(I \times n) \quad (5.1)$$

$$\hat{y} = \log(I) + \log(n) \quad (5.2)$$

Here,  $I$  denotes the image, and  $n$  represents the noise. The logarithmic operation effectively transforms the multiplicative nature of speckle noise into an additive form. Consequently, all the subsequent analyses presented in this study are conducted on images that have undergone this pre-transformation.

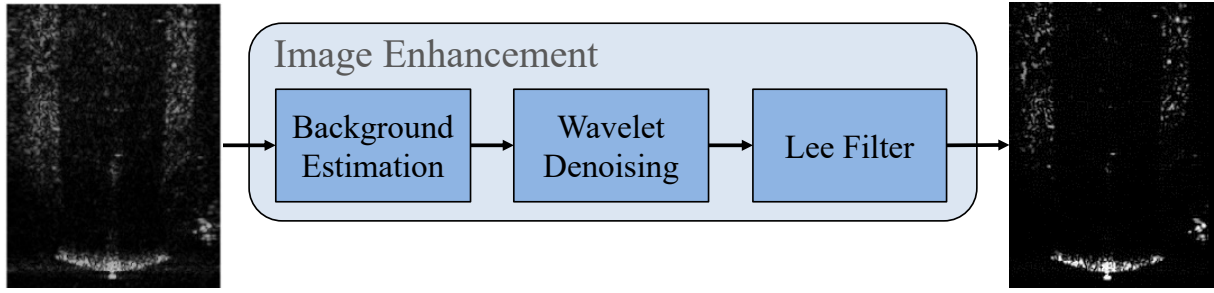
Typically, the Gaussian filter is a good starting point for analyzing the image denoising. However, it has the opposite effect on sonar images as it introduces a blurring effect and expands the spikes to further surrounding pixels. Similarly, several state-of-the-art kernel-based filters are available, which have proved to be better on sonar images than the Gaussian filter. For example, the median filter takes the median value within the kernel and limits the spikes with respect to its surrounding. This technique reduces the spikes but removes edges due to the blurring effect. Furthermore, all kernel-based filters often suffer from two significant problems. 1) Square kernel rounds the corner of the feature if the corner is not perpendicular. 2) They inhibit the smoothing near the edge instead of enhancing the edges.

Other filters are categorized into two categories. 1) Statistical Filters. 2) Wavelet Denoising Filters. Statistical filters such as Lee [Lee 81], Kuan [Kuan 85], and Frost [Frost 82], filter the noise by computing the mean, variance, and weights of pixels within the local window. These filters, specifically Lee, have shown an enhancement in edges. On the other hand, wavelet denoising filters remove noise by decomposing an image in different frequency sub-bands and performing soft or hard threshold operations. The high threshold value may remove the essential information whereas low threshold value may not remove the spikes.

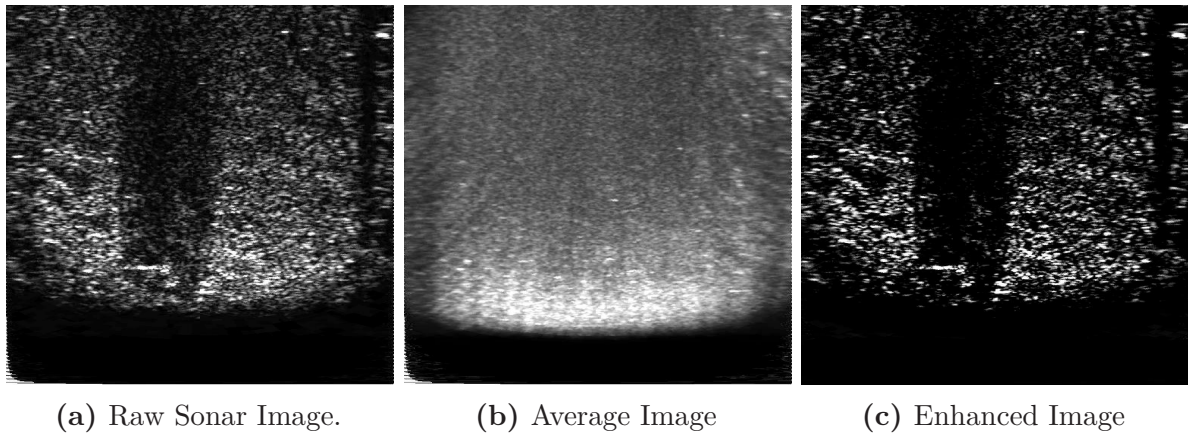
The methodology presented in Figure 5.9 aims to overcome the challenges of speckle noise caused by reflections and improve object detection rates by segmenting images. To achieve this, the denoising approach used in this methodology takes inspiration from previous studies such as [Galceran 12] and [Hurtós Vilarnau 14]. These studies have introduced a novel technique that involves the computation of the sonar insonification pattern by averaging multiple consecutive sonar images. The derived insonification pattern is then applied to each new sonar image, resulting in a clearer and distortion-free image with significantly reduced speckle noise. It can be mathematically formulated as:

$$I_{mean} = \frac{\sum_1^N I_{in}}{N} \quad (5.3)$$

$$I_{insonify} = I_{in} - I_{mean} \quad (5.4)$$



**Figure 5.9:** Sequential Procedure for Enhancing Underwater Imagery. This flow diagram showcases an image enhancement process for underwater sonar imagery, starting with a noisy original image and filtering in a clearer, noise-reduced output. The process involves background estimation, wavelet denoising, and Lee filtering to suppress noise and enhance object visibility within the image.

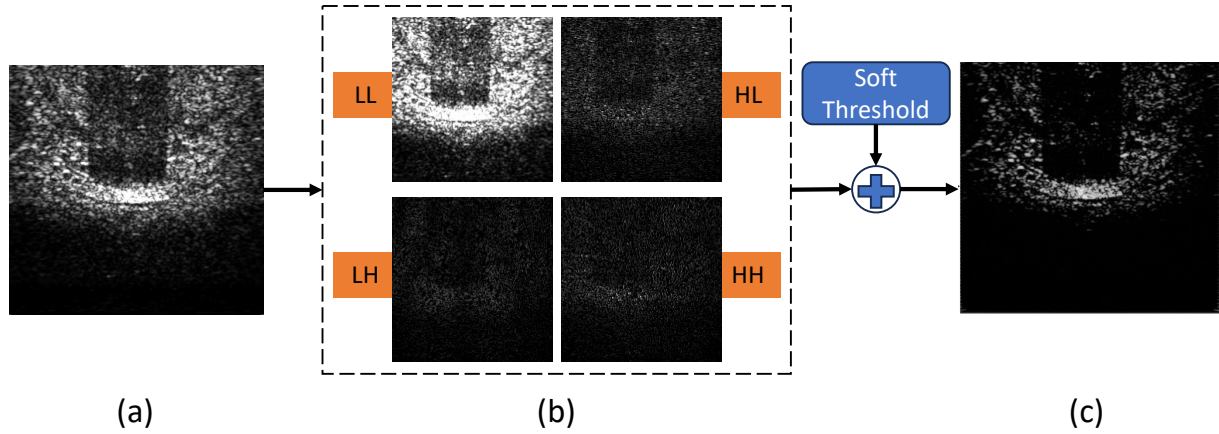


**Figure 5.10:** Background Estimation Technique Applied to Sonar Imagery. The figure demonstrate the background estimation method. This series of images illustrates the steps of a background estimation method for sonar imagery enhancement. It averages of multiple frames (100 in above case) for noise reduction, and provides an enhanced image for clearer visualization of underwater objects (*The contrast is adjusted in the images for better visualization only.*)

Figure 5.10 shows the insonification pattern of 100 consecutive sonar frames and result of estimating the background of the sonar image in a 2-meter deep lake.

Additionally, the presence of additive noise in the sonar images is effectively addressed through a box (mean) filter. This filtering technique helps to reduce abrupt intensity changes within the image, resulting in a smoother and less noisy appearance. However, the box filter may not fully eliminate random noisy points, such as waterbed reflections, that often appear in sonar images, especially at high frequencies. Hence, a more advanced approach is used involving the decomposition of the sonar image into sub-band images, each capturing different frequency components and orientations. This decomposition is made possible through the application of the Discrete Wavelet Transform (DWT) technique. DWT is a mathematical method that allows the analysis of a signal at various frequencies and resolutions. The DWT  $W(\tau, s)$  can be mathematically written [Choi 20a] as:

$$W(\tau, s) = \frac{1}{\sqrt{s}} \sum_{m=0}^{p-1} f[t_m] \lambda\left[\frac{t_m - \tau}{s}\right] \quad (5.5)$$



**Figure 5.11:** Discrete wavelength decomposition of sonar imagery. Figure shows the discrete wavelength decomposition in LL, HL, LH, HH intensity frequency bands. (a) Raw input image, (b) Different intensity frequency bands and (c) Output image after soft thresholding.

where  $f$  is signal evaluated at the discrete time index  $t_m$  and  $\lambda$  is known as wavelet act as window function on a mother wavelet.  $\tau$  is translation parameter,  $s$  is scale parameter which is inverse of frequency and can be written as:

$$\tau = k2^{-j} \quad (5.6)$$

$$s = 2^{-j} \quad (5.7)$$

where  $j$  is the scale index, and  $k$  is the wavelet transform signal index.

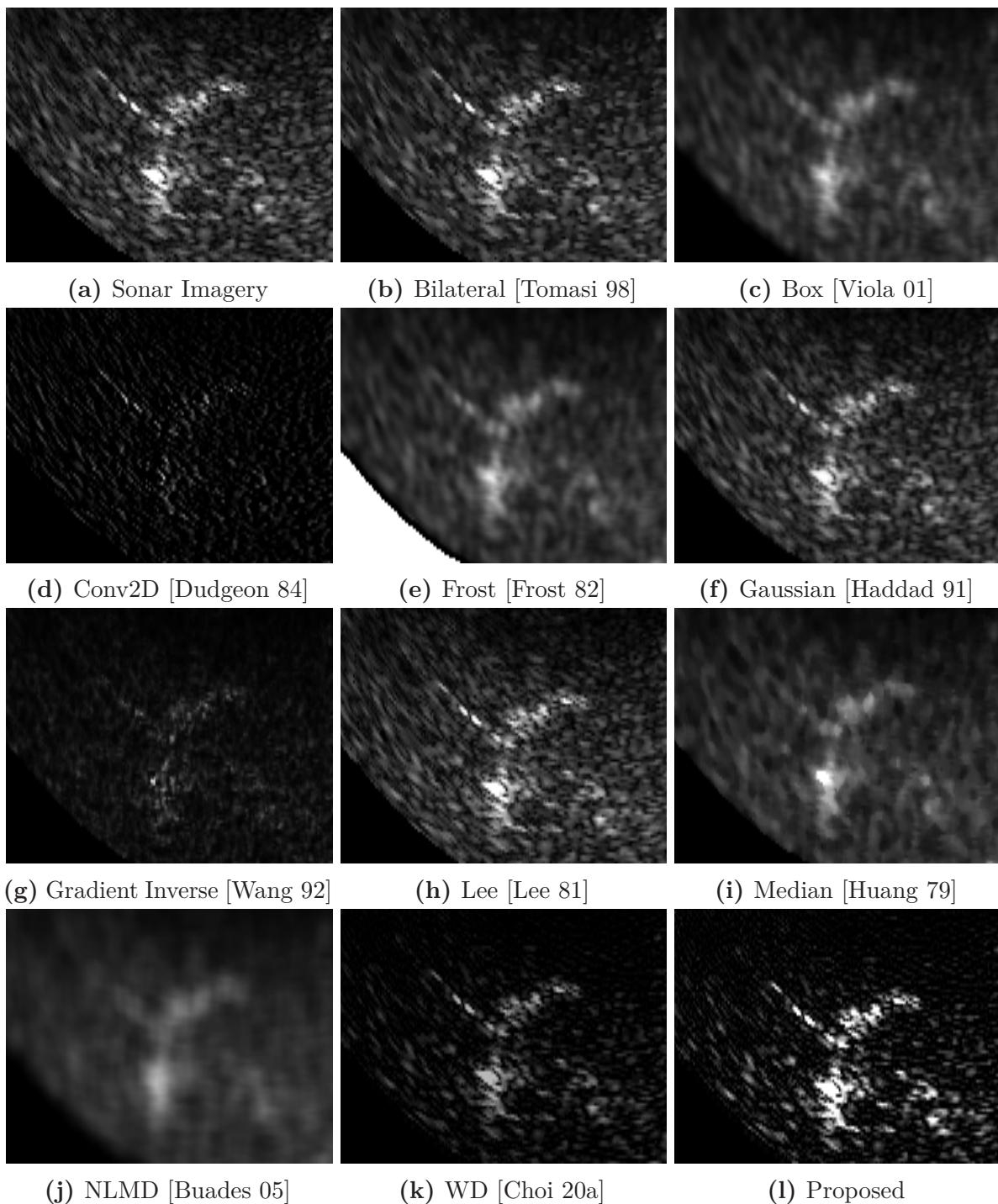
Figure 5.11 shows the single-level DWT on the sonar image with the soft threshold set to 0.6. The image includes an obstacle with its shadow.

The Discrete Wavelet Transform (DWT) has a significant impact on sonar images by removing high-frequency noise points from the waterbed and obstacle reflections. However, DWT can also reduce the intensity of important reflections (obstacles), which is a challenge. To improve the image quality and distinguish between waterbed sediments and obstacles, a local statistical analysis using the Lee filter is used. The Lee filter is known for its ability to enhance edges and fine details within the image, which results in improved contrast and reduced noise. This enhanced image quality ultimately improves the system's ability to accurately detect and segment underwater objects.

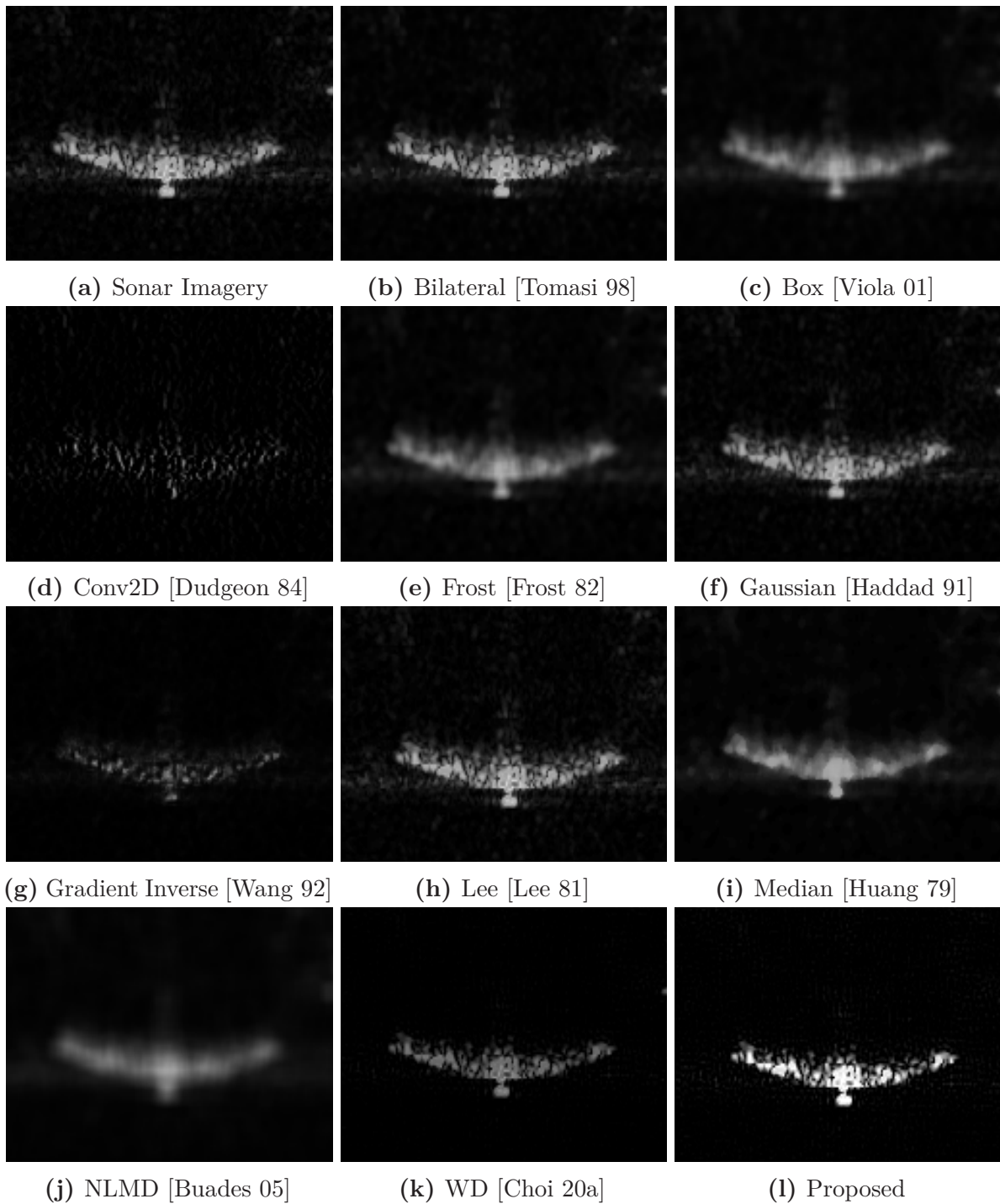
The presented image enhancement technique is systematically evaluated by comparing it with ten state-of-the-art methods that are well-documented in the existing literature. This evaluation is essential for assessing the effectiveness of proposed method in enhancing sonar images. To provide a comprehensive overview of the technique's performance, qualitative comparisons with methods are presented in Figures 5.12 and 5.13. These visual comparisons demonstrate that the proposed technique excels in producing enhanced images characterized by high object-to-background contrast.

However, conducting a direct quantitative analysis of this technique is challenging due to the absence of a ground truth image that could serve as a reference for objective evaluation.





**Figure 5.12:** Comparative Analysis of Filtering Techniques on Sonar Imagery for Enhancement (1). This figure showcases a variety of filtering methods applied to sonar imagery to evaluate their effectiveness in image enhancement. (l) features the proposed technique which, upon visual assessment, yields the most distinct contrast between the background and the obstacles and presents the least amount of blurriness. This clarity and contrast are crucial for the accurate identification and interpretation of underwater features and objects, demonstrating the proposed method's superiority over conventional and advanced filtering techniques in enhancing sonar imagery. NLMD stands for Non-local Mean Denoising whereas WD stands for Wavelet Denoising.



**Figure 5.13:** Comparative Analysis of Filtering Techniques on Sonar Imagery for Enhancement (2). This figure showcases a variety of filtering methods applied to sonar imagery to evaluate their effectiveness in image enhancement. (l) features the proposed technique which, upon visual assessment, yields the most distinct contrast between the background and the obstacles and presents the least amount of blurriness. This clarity and contrast are crucial for the accurate identification and interpretation of underwater features and objects, demonstrating the proposed method's superiority over conventional and advanced filtering techniques in enhancing sonar imagery. NLMD stands for Non-local Mean Denoising whereas WD stands for Wavelet Denoising.

To address this limitation, an indirect quantitative analysis is performed, focusing on the impact of each technique presented in Figure 5.12 on object detection accuracy.

Table 5.2 offers a detailed comparative analysis, providing the performance of each technique in terms of true positives and false negatives during the detection phase. Furthermore, each algorithm is compared with respect to the integrity of object detection i.e. either the object is detected as a complete object or multiple separate objects. This comparative analysis reveals that the proposed filtration technique consistently outperforms other standard filters in object detection. While the absence of a ground truth image necessitates this indirect assessment, the results clearly indicate the superiority of the proposed method in enhancing object detection accuracy within sonar imagery.

**Table 5.2:** Performance metrics of object detection using various filtering techniques on sonar imagery. This table presents the efficacy of different filtering methods applied to sonar imagery, evaluated by their impact on object detection results. It enumerates the percentage of True Positives and False Positives, and the integrity of object detection i.e. detected completely or in parts, categorized as *Mostly*, *Rarely*, or *Never* intact. The proposed technique has higher true positive detection and lower false positive detection. The *Object Intact* is an important criteria to avoid detecting an object as multiple object.

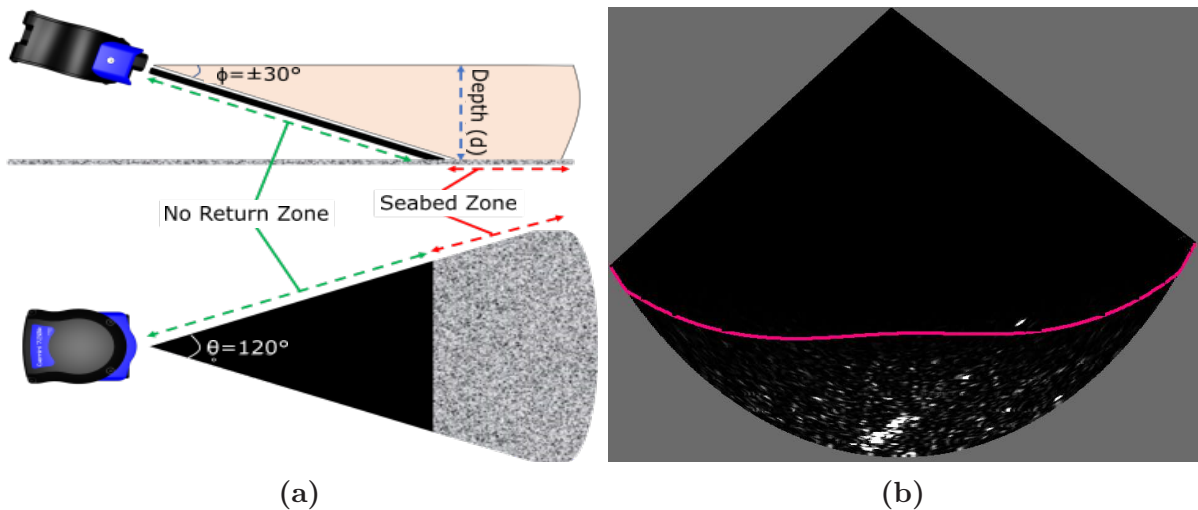
| Filter           | True Positive | False Positive | Object Intact |
|------------------|---------------|----------------|---------------|
| Bilateral        | 48.33         | 51.67          | Rarely        |
| Box              | 71.01         | 29.99          | Mostly        |
| Convolution 2D   | 59.13         | 40.87          | Never         |
| Frost            | 20.54         | 79.46          | Never         |
| Gaussian         | 63.11         | 36.89          | Never         |
| Gradient Inverse | 80            | 20             | Never         |
| Lee              | 76.03         | 23.97          | Never         |
| Median           | 68.3          | 31.7           | Mostly        |
| Non-local Mean   | 87.55         | 12.45          | Mostly        |
| Wavelet          | 62.1          | 37.9           | Rarely        |
| <b>Proposed</b>  | <b>92.2</b>   | <b>7.8</b>     | <b>Mostly</b> |

### 5.2.1.2 Depth Estimation

Accurate depth estimation is fundamental for safe and efficient navigation within aquatic environment. It enables surface and underwater vehicles to determine water depth, identify potential underwater obstacles, and ensure safe passage in real-time. Moreover, in scenarios such as search and rescue operations or archaeological surveys, the precise localization of submerged objects and the seabed is essential. Depth estimation aids in mapping underwater terrains, identifying shipwrecks, or locating missing persons by providing critical information about the spatial arrangement of objects in the underwater environment. Furthermore, depth estimation is closely related with obstacle detection and avoidance, a critical aspect of aquatic operations. By accurately estimating the depth of underwater objects, the proximity of obstacles, including rocks, reefs, or submerged structures can be detected. This information is crucial for collision avoidance and safe navigation, especially in shallow or complex aquatic terrains.

As discussed in Section 5.1.2, it is evident that forward-looking sonar imagery introduces specific constraints that necessitate careful consideration. A salient limitation lies in the inherent nature of such a sonar, which predominantly provides 2.5-dimensional data. This characteristic implies that the acquisition of precise depth information regarding underwater obstacles and the seabed presents a notable challenge. The significance of this limitation affects various underwater applications, particularly in the context of surface vehicles and the assessment of potential navigational obstacles.

While there are several 3D sonar systems commercially available, it should be noted that they are often extremely costly and impractical where the cost effective system is required. In such environments, the high intensity reverberations due to waterbed further complicates the acquisition of reliable depth information. Consequently, the accurate estimation of water depth is essential when utilizing forward-looking sonar. The ability to navigate safely, assess potential threats, and make informed decisions in underwater environments depends significantly on the development and application of techniques that can compensate for the inherent limitations of 2.5D forward-looking sonar data.



**Figure 5.14:** Sonar Imaging for Seabed Analysis. (a) illustrates the use of sonar for seabed mapping, showing the FOV and the calculation of seabed depth using the sonar’s angles. (b) The red line on the sonar image represents the seabed contour, providing a visual representation of its shape and depth in relation to the sonar’s position.

The SWiM architecture has introduced a novel depth estimation module that uses a regression-based approach to accurately measure the depth of a waterbed. To comprehend this module, it’s crucial to understand the principles of sonar imagery. Sound waves generate high-intensity patterns in sonar images when they interact with the waterbed. Any area without an echo will appear as zero pixel intensity. If the initial point of contact between sonar waves and the waterbed is 15 cm apart, the sonar imagery before this point will exhibit minimal reflection. As depicted in Figure 5.14a, this leads to a significant difference in pixel intensity values before and after the interaction point. Furthermore, to ensure the accurate estimation of waterbed depth within the sonar frame of reference and to enhance the precision of object detection, a polynomial fit regression line is estimated. This regression line is computed by examining the pixel values at the intersection of two

distinct segments, namely the 'no return zone' and the 'waterbed zone,' as illustrated in Figure 5.14b.

However, it is crucial to emphasize that the regression line, when computed in the sonar frame of reference, primarily provides estimated depth of the waterbed rather than its absolute depth. To ascertain the waterbed's depth accurately, an adjustment is required, involving repositioning the sonar at an angle such that the highest point within its vertical FOV aligns precisely with the water's surface, as depicted in Figure 5.14a. This specific geometric adjustment facilitates the application of appropriate trigonometric principles to estimate the depth, as represented by the following equation.

$$d = R \sin \phi \quad (5.8)$$

where  $R$  is the range of each pixel in regression line and  $\phi$  is the vertical opening angle of the sonar. It is worth noting that the sonar provides raw data in the form of scan points where each scan point has its own range and angle. This range data is then transformed into image. Therefore, the range of each pixel in the image can easily be calculated.

## 5.2.2 Obstacle Filtration

Underwater obstacle detection through sonar images can be a challenging task, as variations in water depth and environmental conditions may affect the image quality. Reliable obstacle detection is crucial for generating accurate and precise maps of the underwater terrain. To tackle this challenge, the Underwater Traversability Mapping (UTM) framework has been specifically designed to integrate conventional and novel strategies, combined to enhance the accuracy and reliability of obstacle detection in the dynamic and often unpredictable underwater environment. By fusing those strategies, UTM offers a comprehensive solution to the complex problem of obstacle detection, ultimately elevating the capabilities of traversability mapping in underwater scenarios. In the following subsections, this thesis comprehensively discusses the obstacle detection and refinement algorithms available in UTM, as well as the subsequent results.

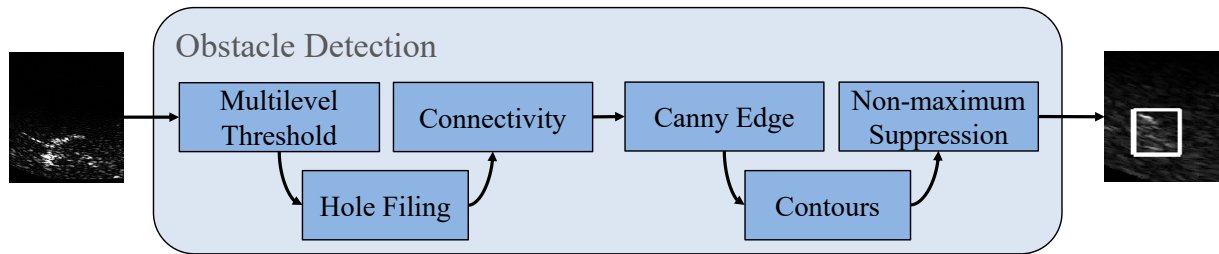
### 5.2.2.1 Obstacle Detection

The image filtration phase discussed in section 5.2.1 plays a key role in reducing a substantial amount of speckle noise, a common issue in sonar imagery. Nonetheless, sonar images often contain unwanted reflections from the waterbed, which can complicate the accurate detection of objects. In applications where the primary objective is to detect objects with maximum completeness or shape, the elimination of these waterbed reflections is essential. To tackle this challenge, a series of techniques have been deployed to detect objects from the waterbed reflections, with the aim of achieving a robust object detection within each sonar image.

Within this thesis, it has been determined that difference between objects and the waterbed in sonar imagery can be computed through two fundamental physical principles. First, objects typically exhibit a higher reflection coefficient compared to the seabed. This variance in reflection coefficients enables the differentiation of objects from the waterbed based on their higher intensity levels in contrast to the waterbed. Second, in cases where the reflection coefficient difference between an object and the seabed is minimal, object

detection can be helpful by identifying the shadows cast by objects on the seabed, as shown in figure 5.16a. Detecting such shadows can be accomplished through the utilization of standard image processing algorithms or more advanced deep learning architectures. Each of these approaches offers its distinct advantages and limitations. For instance, standard algorithms do not necessitate extensive datasets for learning but may require fine-tuning parameters dependent on the specific underwater seabed conditions. Conversely, deep learning architectures do not necessitate frequent parameter adjustments, but they do require substantial and diverse datasets to enhance their robustness across varied environmental contexts.

In our application, we have adopted a strategy by implementing both standard algorithms and deep learning architectures. By fusing the outcomes of these approaches, we aim to enhance the detection rate of the architecture, effectively using the strengths of each methodology while reducing their respective weaknesses. This fusion approach seeks to optimize object detection in sonar imagery and enhance the overall performance of the system.

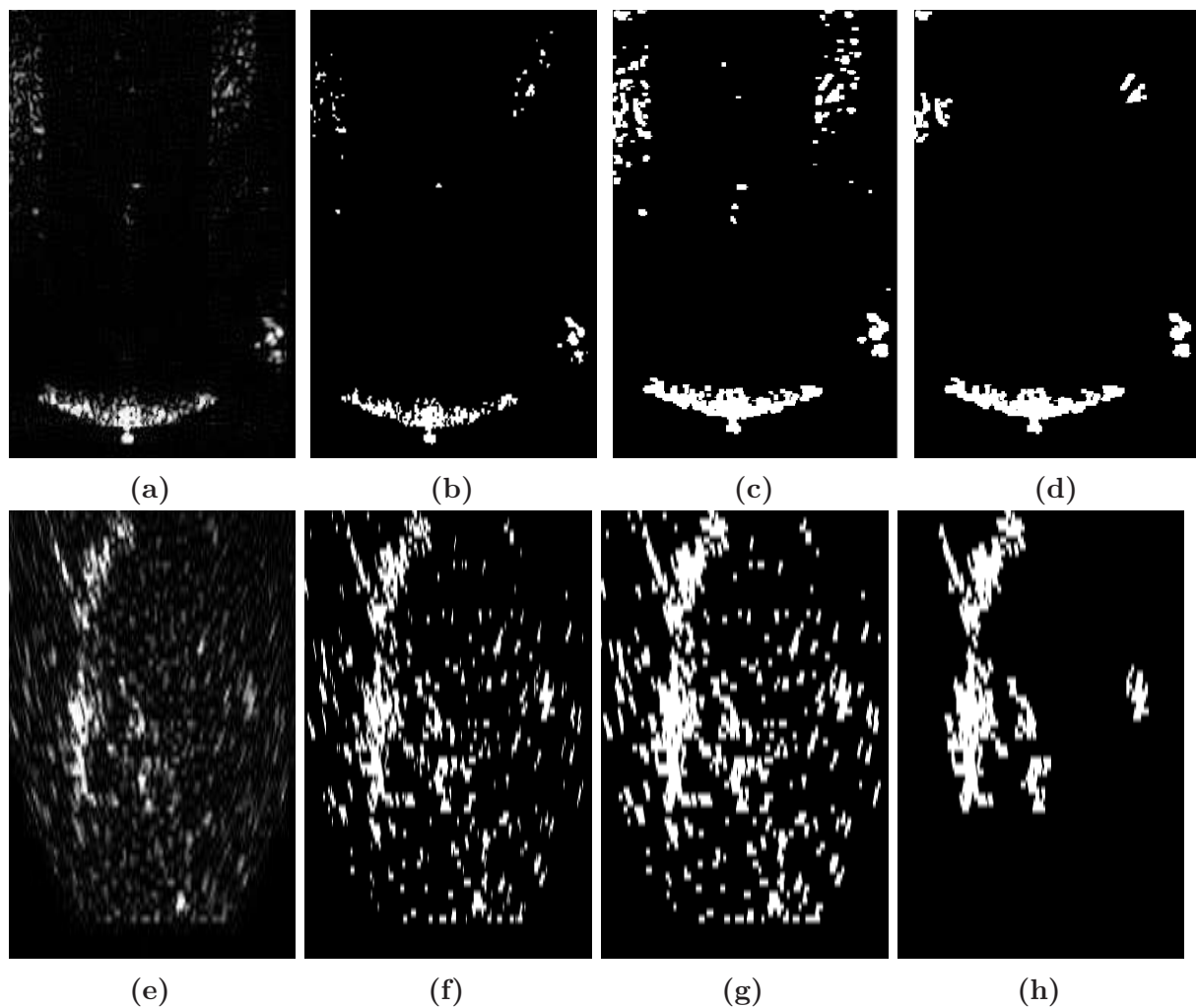


**Figure 5.15:** Obstacle Detection Flow Diagram. The figure displays a process flow for obstacle detection in sonar images, starting with multilevel thresholding and followed by connectivity analysis, hole filling, Canny edge detection, contour extraction, and concluding with non-maximum suppression for crisp obstacle detection.

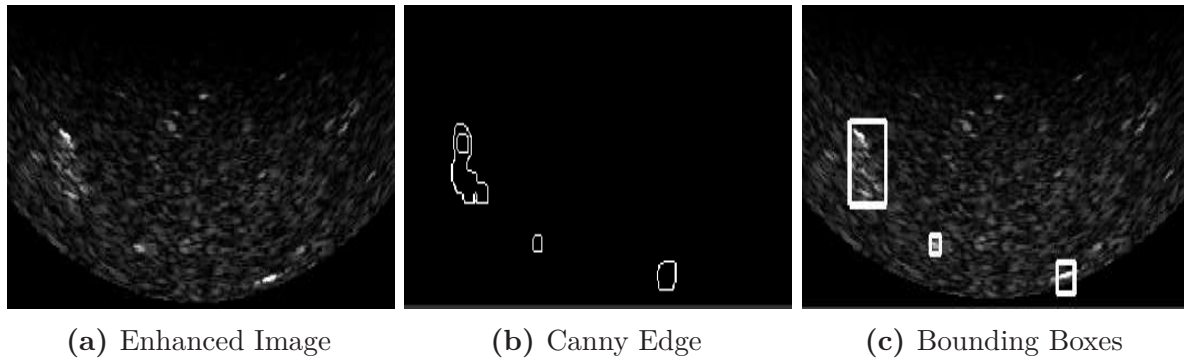
Several standard vision algorithms are implemented and tuned in order to detect obstacles in real-time. The comprehensive sequence of obstacle detection steps is outlined in figure 5.15, detailing the complex process by which the presented architecture independently detects the intensity and shadows of objects. This is achieved by using the same image to a shared module, with distinct parameters. Ultimately, the outcomes of both intensity and shadow detections are fused to filter accurate obstacles, which serve as the foundation for subsequent processing stages.

In the initial phase of this module, a multilevel thresholding technique is used to differentiate intensity levels. In the context of obstacle detection (excluding shadows), the low-intensity pixels are systematically removed. This process results in the image depicted in figure 5.16b and 5.16f, where a specific threshold value, determined as 0.29, has been applied. Nevertheless, it is essential to acknowledge that the thresholding step may eliminate important points within the imagery related to obstacle. To rectify this, a hole-filling procedure is introduced to connect and restore any fragmented components of the obstacles. In this context, we have assume a minimum obstacle size of 25 cm. The connectivity analysis is executed by examining groups of points against this criterion, facilitating the removal of non-obstacle points. The final qualitative results of this phase are presented in Figure 5.16c and 5.16g.

Moving forward, the well-known Canny Edge detection [Canny 86] technique is employed to identify the boundaries of the obstacles. Subsequently, contours are identified around each obstacle, enabling the computation of bounding boxes that encapsulate the detected objects as shown in figure 5.17. An additional refinement step involves filtering the contours using a non-maximum suppression strategy, which serves to merge bounding boxes belonging to closely situated objects, defining them as a singular obstacle. At this point, the certainty map assumes a pivotal role in allocating detection probabilities to each object. This integration significantly contributes to include true obstacles within the traversability maps, enhancing the precision and reliability of the entire process.



**Figure 5.16:** Three-Step Morphological Processing for Obstacle Detection in Sonar Imagery. This series of images demonstrates a morphological operation sequence designed to enhance obstacle detection in sonar data. (a, e) present the raw enhanced sonar image. The process begins with (b, f) applying a two-level threshold to mitigate the presence of high-intensity noise points which could be mistaken for actual obstacles. Next, (c, g) hole filling is employed to ensure that real obstacles are perceived as complete shapes without internal gaps. Finally, (d, h) connectivity analysis is performed to isolate discrete obstacle entities. By systematically reducing noise and reinforcing the continuity of object structures, this method effectively prepares the data for more robust obstacle detection in subsequent analysis stages.



**Figure 5.17:** Two-Step Canny Edge and Contour-Based Object Detection in Sonar Imagery. The figure displays a method for detecting objects in sonar images, using enhancement to clarify features, Canny edge detection to outline objects, and contour finding to place bounding boxes around them, streamlining the process from raw data to identifiable objects.

### Why not Deep Learning?

Deep-learning approaches, despite their widespread success in various domains of image processing and object detection, face significant limitations when applied to underwater object detection using sonar imagery. The unique challenges presented by underwater environments, such as variable acoustic properties, diverse noise characteristics, and the scarcity of labeled datasets, impede the effectiveness of deep-learning models. Sonar images are inherently noisy due to the medium through which sound waves travel, and they suffer from artifacts like speckle noise, multipath propagation, and varying degrees of signal attenuation. These issues are added up by the fact that sonar devices themselves can introduce device-specific noise, further complicating the training process for deep-learning algorithms. Unlike in standard vision tasks where enough diverse, and high-quality datasets are available, the underwater domain lacks such extensive datasets, making it difficult to train deep-learning models that can generalize well across different underwater environments and sonar systems. Moreover, the dynamic nature of underwater environments, with constantly changing conditions and the presence of random objects, poses additional challenges for deep-learning models, which require vast amounts of data to capture such variability effectively.

In contrast, standard vision algorithms, which often include traditional image processing techniques such as thresholding, edge detection, and morphological operations, have shown to be more adept at handling the specific challenges of sonar imagery. These algorithms can be fine-tuned to understand the effects of noise and improve the signal-to-noise ratio, enhancing the detectability of underwater objects without the need for extensive datasets. Moreover, these techniques can be adapted to exploit the inherent characteristics of sonar signals, such as their reflection patterns and the acoustic shadows cast by objects, to identify and classify underwater features more reliably. The adaptability, lower dependency on large datasets, and the ability to incorporate domain-specific knowledge make standard vision algorithms particularly suitable for the task of underwater object detection using sonar. This approach aligns with the findings and methodologies explored in the preceding sections, emphasizing the application of standard vision techniques over deep-learning models in the context of sonar-based underwater mapping and object detection.

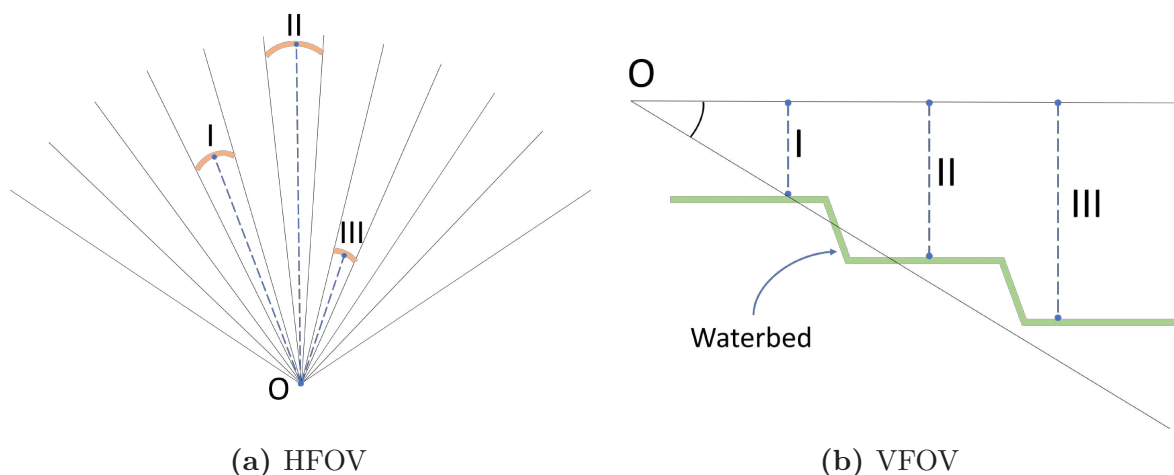


### 5.2.2.2 Obstacle Refinement

The information about the obstacle is further refined to increase the certainty of the detection based on two criteria. First, the sonar's physical constraints are considered, and a certainty map is generated, which determines the probabilistic certainty of each pixel in the image. Secondly, each pixel's motion with respect to vehicle motion is determined. Those pixels which depict the realistic motion are considered true obstacles, while the remaining pixels are neglected. Each criterion is further explained in the following sections.

#### Certainty Map

The primary challenge in sensor processing strategies revolves around addressing the inherent disturbances and uncertainties present in raw sensory data. In the context of multibeam forward-looking sonar systems, these uncertainties frequently present in a highly random fashion, making them particularly complex to model effectively. However, integrating models that account for the physical constraints of such sensors can significantly enhance the accuracy of object detection procedures. Within the domain of forward-looking sonar, these physical constraints encompass parameters such as angular resolution and vertical beam width, which play a pivotal role in determining the system's capabilities.



**Figure 5.18:** Illustration of field of view (FOV) and Distance Impact on Sonar Resolution. (a) Horizontal field of view (HFOV) demonstrates how the sonar beams fan out from the origin point (O), indicating three distinct areas (I, II, III) where the resolution decreases with distance. The closer beams have a higher resolution, which diminishes as the distance from the origin increases. (b) Vertical field of view (VFOV) shows the side profile of the sonar beams intersecting with the waterbed. Similar to the HFOV, the resolution is higher closer to the origin and deteriorates with distance, as indicated by the varying distances between the dashed lines (I, II, III). This visualization helps understand the challenges of detecting objects at different depths and distances, with the decrease in resolution potentially affecting the detection accuracy.

Angular resolution, for instance, denotes the angle between two consecutive sonar beams. It is important to note that a larger angular separation between beams results in an overall lower resolution for the sonar system. Furthermore, even when the angular separation between two adjacent beams remains constant, the arc distance between these beams increases as the range from the sonar source expands as illustrated in Figure 5.18a. This

increase in arc distance reduces the resolution, posing a substantial challenge for precise object detection.

Similarly, vertical beam width signifies the angular extent within which an object can be effectively detected by the sonar. This angular width expands with increasing distance from the sonar source. Consequently, a greater distance leads to a larger vertical arc width (illustrated in Figure 5.18b), increasing the uncertainty associated with estimating the depth of submerged objects. It is essential to recognize that these factors, including angular resolution and vertical beam width, are of paramount importance and must be considered to facilitate more accurate and reliable object detection in the context of forward-looking sonar systems. By comprehensively addressing these physical limitations, it becomes feasible to significantly enhance the precision and effectiveness of object detection processes.

**Table 5.3:** Range vs Detection Accuracy. The table correlates the detection performance of the Faster R-CNN model with varying depths and sonar ranges in an underwater environment. The data illustrates a general decline in both accuracy and recall as the distance and depth of the objects increase, indicating the challenges faced by object detection models in sonar-based range assessments. This table is critical for understanding the limitations of current detection technologies in relation to object proximity and depth in underwater scenarios.

| Depth<br>(m) | Sonar Range<br>(m) | O. Range<br>(m) | O. Depth<br>(m) | Accuracy | Recall |
|--------------|--------------------|-----------------|-----------------|----------|--------|
| 1.2          | 3                  | 0.6             | 0.20            | 0.82     | 0.53   |
| 1.2          | 3                  | 1.5             | 0.42            | 0.67     | 0.44   |
| 1.2          | 3                  | 2               | 0.42            | 0.69     | 0.44   |
| 2.8          | 3                  | 1.3             | 0.68            | 0.79     | 0.51   |
| 2.8          | 5                  | 2.2             | 1               | 0.81     | 0.57   |
| 2.8          | 5                  | 4               | 1.1             | 0.64     | 0.41   |
| 4.3          | 5                  | 2.7             | 2.9             | 0.73     | 0.49   |
| 4.3          | 10                 | 5.3             | 3.2             | 0.61     | 0.37   |

Furthermore, Table 5.3 presents an evaluation of object detection performance as a function of varying sonar ranges and object depths, utilizing the Faster R-CNN model. The dataset is segmented based on the range of detected objects, and the table reports on two key metrics: Accuracy and Recall. Depth categories are provided alongside corresponding sonar range distances, object range (O. Range), and object depth (O. Depth). Each row offers a snapshot of how detection accuracy and recall rates fluctuate with changes in object distance and depth, reflecting the model's capacity to recognize objects under different underwater sonar conditions. Results show a trend where detection accuracy and recall typically decrease as the object range and depth increase, suggesting challenges in detecting objects at greater distances and deeper in the water column.

The Certainty Map (CMAP) is an invaluable strategy used to calculate the probability associated with each pixel in the sonar imagery, taking into account a fusion of static and dynamic factors. When mapping the 3D underwater environment to 2.5D sonar data, a level of uncertainty arises regarding the precise height or location of an obstacle's collision point. This uncertainty primarily comes from the vertical sonar FOV, which undergoes changes as obstacles approach the sonar device. Consequently, the data tends to be more

reliable and certain about the exact location of an object's collision point as it moves nearer to the sonar source. To address this inherent uncertainty, a fused CMAP is constructed, taking into consideration the following factors:

**Static CMAP** The static component is dependent on the VFOV and the range of the sonar sensor. It assigns variable certainty to pixels based on their proximity to the sonar. Pixels that are farther from the sonar are considered as less certain due to the wider VFOV and the inherent limitations in acquiring precise information about the depth or exact location of an obstacle. For example, an obstacle located 2 meters away may exist within a 70 cm vertical arc, as given in Table 5.1.

**Dynamic CMAP** Conversely, the dynamic CMAP relies on depth values provided by the *Depth Estimation* phase (section 5.2.1.2). This component calculates the closest point on the waterbed based on the depth value, effectively identifying the point beyond which pixels exhibit fewer reflections, appearing primarily darker as illustrated in Figure 5.14. Consequently, the region before this point, where obstacles possess a high contrast with the background, is categorized as a high-probability area, while the region beyond is designated as low-probability.

The fused CMAP is represented mathematically as follows:

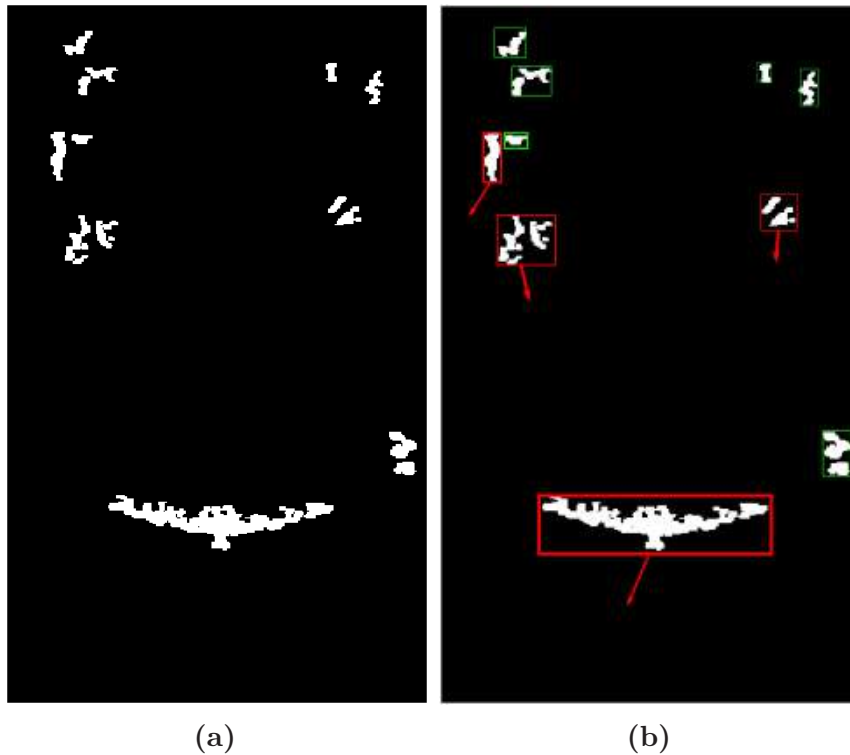
$$P_{ptx}^S = \begin{cases} 1 - \left(\frac{0.4}{d_{water}} \times d_{ptx}\right), & \text{if } d_{ptx} \leq \frac{d_{water}}{\cos(90^\circ - VFOV)} \\ 0.6, & \text{otherwise} \end{cases} \quad (5.9)$$

Here,  $P_{ptx}^S$  denotes the probability associated with each sonar point,  $d_{water}$  represents the water depth obtained from the *Depth Estimation* phase, and  $d_{ptx}$  is the distance of each pixel in the Cartesian coordinate system from the sonar's origin. The vertical field of view of the sonar (in our case  $20^\circ$ ) is denoted by  $VFOV$ . The factor  $\frac{d_{water}}{\cos(90^\circ - VFOV)}$  is the nearest point of the waterbed at a specific height. The conditions set forth in the equation above establish that the probability of a pixel closer to the sonar is higher than that of pixels farther away, reflecting the relationship between depth, distance, and the sonar's FOV.

### Motion-Based Tracking

The previously identified obstacles, as filtered from the certainty map phase, undergo further refinement through the implementation of a motion-based tracking algorithm. This particular algorithm uses the predictive capabilities of the Kalman filter [Salhi 12] to track the motion of each obstacle. The filter computes the trajectory of each obstacle, represented as a group of pixels within each image frame, and effectively maintains its track. The refinement process considers an obstacle as a potential threat based on two distinct criteria: firstly, the number of consecutive frames an object is tracked, and secondly, its motion in relation to the vehicle's motion.

Under the first criterion, it is acknowledged that the tracking algorithm may occasionally lose track of an obstacle's trajectory. Such a track can be re-established if the tracker manages to reacquire it within a window of three frames. However, if the object's trajectory is reacquired after a gap of three frames, it's treated as a new track. The number of



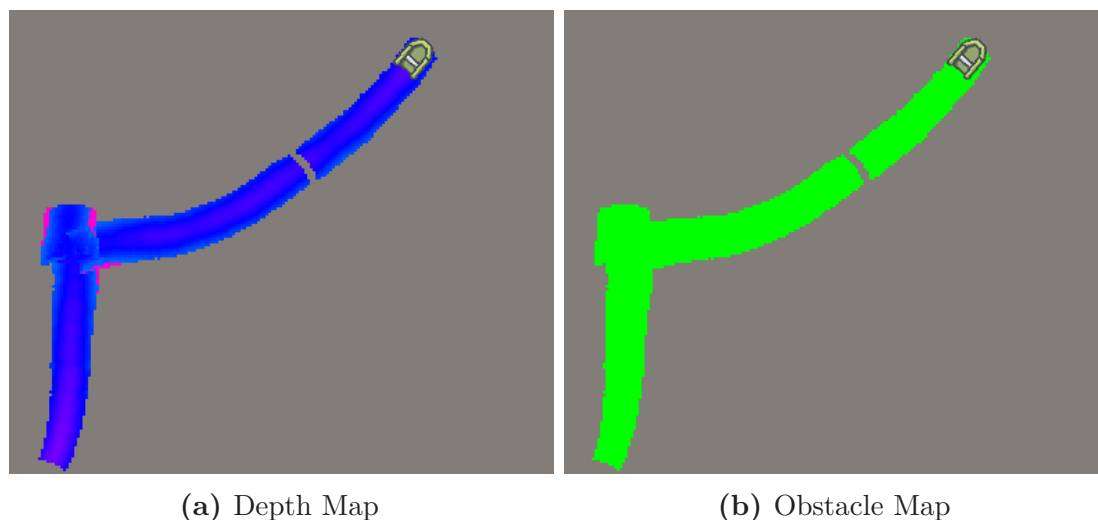
**Figure 5.19:** Application of Object Tracking in Sonar Image Processing for Obstacle Detection. (a) demonstrates the initial identification of potential obstacles in sonar imagery, where various shapes are detected. (b) applies object tracking algorithm to distinguish between moving obstacles (red bounding boxes), which are potential threats, and stationary objects (green bounding boxes), which may be noise. This approach filters out irrelevant static noise, enhancing the accuracy of obstacle detection in dynamic underwater environments.

consecutive frames during which the object remained trackable serves as one of the refinement criteria.

The tracked obstacle is then analyzed in associated with the vehicle's motion. An object is considered eligible for further consideration if its motion is detected as being in the opposite direction to that of the vehicle and falls within a specified range. Tracks that do not meet these criteria are rejected. For illustrative purposes, refer to figure 5.19, where obstacles highlighted in red signify those eligible for tracking due to their counter-directional motion to the vehicle, while those marked in green are rejected and not regarded as potential threats.

### 5.2.3 Depth and Obstacle Maps

The Underwater Traversability Mapping (UTM) framework constitutes a critical tool for assessing underwater environments, providing a pair of essential maps: a depth map and an obstacle map. These maps are derived from depth and obstacle point clouds, which undergo a series of transformations to enable mapping in global coordinates. Initially, GPS coordinates are leveraged to establish a global reference frame, ensuring consistent alignment of the point cloud data with it across various time stamps. Subsequently, these individual point clouds are organized and stored in grid maps to construct a comprehensive representation of the underwater terrain.

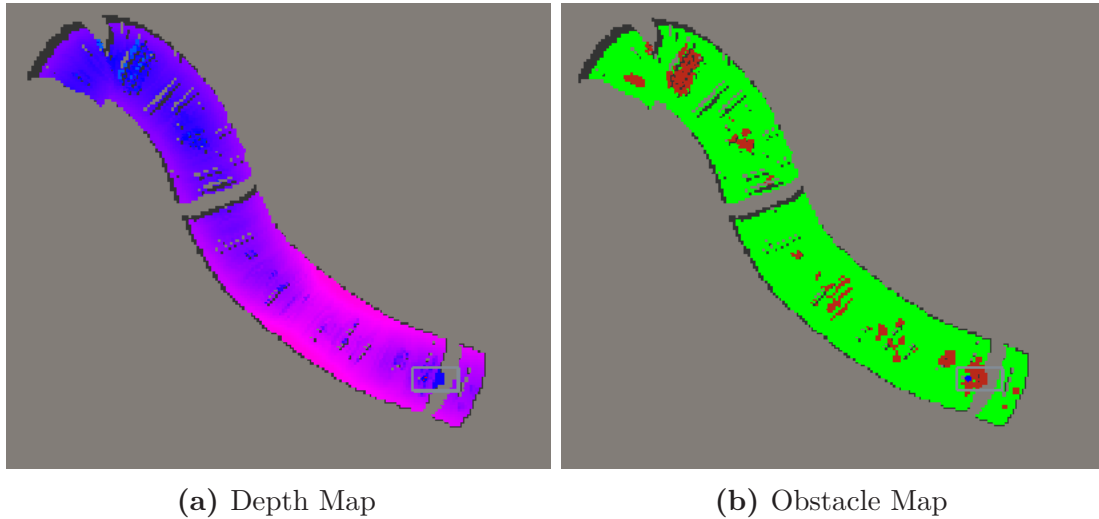


**Figure 5.20:** Underwater Mapping Results - Scenario 1. The boat is driven in Gelterswoog Lake. The sonar range is set to 2 meters due to very shallow lake. (a) shows the depth map of lake, where the low to high depth is demonstrated with blue to pink color, respectively. The average depth of lake is 1.2 meters with  $\pm 23$  cm of variance. (b) shows the obstacle map with waterbed in green color and obstacles in red. No red means there are no substantive obstacles found.

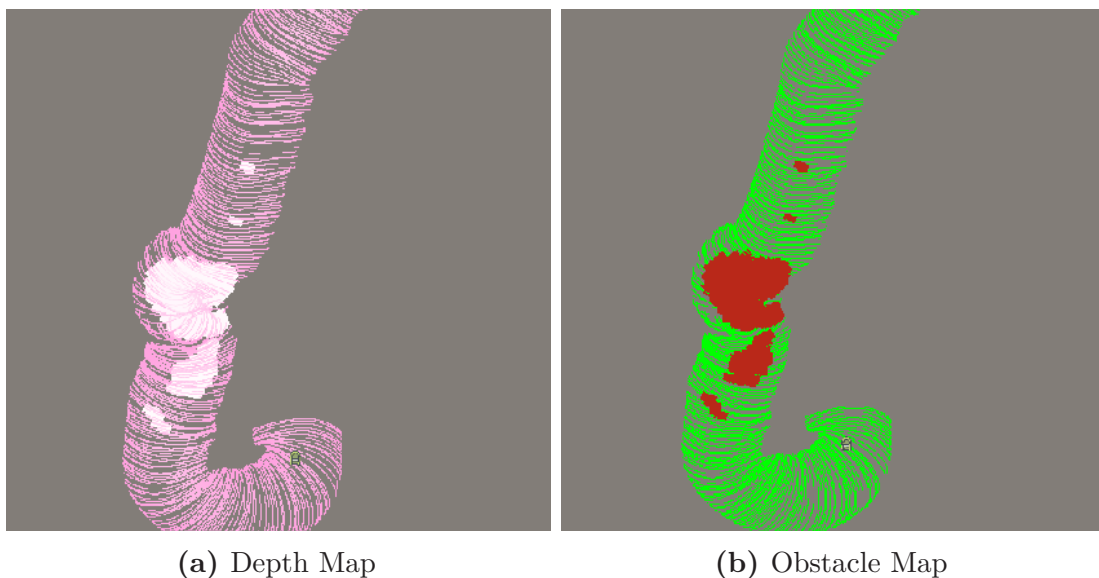
In the context of real-world testing, the UTM is evaluated across three distinct natural environments: a lake, a river, and a port. Each environment presents unique characteristics in terms of depth and obstacles, necessitating tailored assessments. The resulting depth and obstacle maps (demonstrated in Figures 5.20-5.22) provide valuable insights into the underwater topography and potential navigational challenges encountered in these environments. The depth map offers a visualization of the water body's depth within the sonar's range, with a color palette ranging from dark blue (indicating shallow depths) to dark pink (representing the sonar's maximum range). Meanwhile, the obstacle map highlights detected obstacles within the sonar's vicinity, with green denoting the sonar data's proximity and red indicating detected obstacles.

For instance, consider Figure 5.20, which illustrates the maps generated in a shallow public lake with a known depth of approximately 1.5 meters. The boat is driven at 0.9 m/s speed with the 2.1 km/h wind on testday. Here, the UTM captures the lake's shallow depths, depicted by blue points in the depth map. Notably, the absence of obstacles in the obstacle map indicates a clean lake devoid of navigational hazards. Conversely, Figure 5.21 showcases maps generated in a river near Hoya (THW test site), where the known depth is 3.5 meters. Here, the boat is driven at 1.1 m/s speed with 4.3 km/h wind on testday. In this scenario, the depth map reveals variable depths ranging from 0.9 to 3.2 meters, reflecting the river's dynamic nature. Additionally, the obstacle map identifies debris, represented in red, further highlighting potential obstacles to navigation.

Similarly, Figure 5.22 presents maps of the Hamburg port test area, characterized by a known depth of approximately 9 meters. Due to high wind (approximately 11 km/h) on test day, the boat is driven with 1.7 m/s speed. Here, the depth map depicts deeper water bodies in light pink, with areas containing mud debris that significantly alter the water depth, as indicated by white regions. The obstacle map mirrors these variations, detecting



**Figure 5.21:** Underwater Mapping Results - Scenario 2. The boat is driven within a river near Hoya (THW Test site). The sonar range is set to 4 meters due to known depth of site. (a) shows the depth map of river, where the low to high depth is demonstrated with blue to pink color, respectively. The average depth of the site is 2.9 meters with  $\pm 82$  cm of variance. (b) shows the obstacle map with waterbed in green color and obstacles in red. The map has found some underwater debris.



**Figure 5.22:** Underwater Mapping Results - Scenario 3. The boat is driven in Hamburg Port area. The sonar range is set to 10 meters due to known depth of site. (a) shows the depth map of river, where the low to high depth is demonstrated with blue to pink color, respectively. The average depth of the site is 8.1 meters with  $\pm 42$  cm of variance. The map shows a deep water site (b) shows the obstacle map with waterbed in green color and obstacles in red. The map has found some underwater debris.

mud debris as obstacles.

Overall, the UTM's capability to generate accurate depth and obstacle maps across diverse underwater environments underscores its utility in facilitating safe and efficient navigation.

By providing detailed insights into underwater topography and potential hazards, the UTM enhances decision-making processes and aids in mitigating risks associated with underwater traversal. Through its comprehensive mapping capabilities, the UTM contributes to advancing our understanding of underwater terrains and supports various applications, ranging from scientific exploration to search and rescue operations.

## 5.3 Discussion

The sound waves have proved to be an important modality for underwater sensing. The accurate underwater sensing depends on several factors, most importantly frequency of sonar. This chapter highlights the difference between several sonar categorized based on their operating frequency.

Moreover, sonar imagery suffer from several challenges including low contrast and low signal-to-noise ratio. To cater these challenges, this chapter has provided a Underwater Traversability Mapping (UTM) framework that includes several noise reduction and image enhancement techniques that have proved to be more robust and less noisy compared to standard filters. These technique further enhances the accuracy of obstacle detection in underwater environment. It is crucial to acknowledge, however, that the deployment of deep-learning architectures in the domain of underwater obstacle detection remains challenging. The reasons are manifold, including the limited underwater sonar data and most importantly distinct sonar imagery of every device due to the noise generated by the sonar device's internal hardware and the diverse acoustic reflections off varied waterbed compositions, such as rocky or muddy terrains.

The UTM framework contributes with a novel depth estimation technique that measures the depth of water bed by comparing darker and brighter regions in the sonar imagery. This technique accurately estimates the depth within the variance of  $\pm 23\text{cm}$  in shallow waterbodies and  $\pm 82\text{cm}$  in deeper waterbodies. This technique helps in detecting the non-traversable depths essentially required due to abrupt depth variations in post-flood environment.

The final results of the chapter are the underwater depth and obstacle maps. These maps, which demonstrates the framework's success, exhibit a obstacle detection accuracy surpassing 72%. This accuracy is not merely a statistic but a proof of the potential for technology to overcome the inherent challenges of underwater navigation and mapping. Through rigorous research and methodical application of image processing techniques, this work makes the way for advancements in underwater exploration and monitoring.





## 6. Surface Water Traversability

The safe and efficient navigation of ships and boats in various water environments necessitates a comprehensive understanding of the surroundings. The process of safe navigation involves not only avoiding obstacles but also understanding the dynamic nature of the water environment, which may include factors such as water currents, variations in depth, and the presence of submerged or floating objects. Traditional navigation systems often relied on single sensors, such as radar or sonar, to provide limited insights into the immediate vicinity. However, advancements in sensor technologies have enabled the integration of diverse sensors, each contributing unique information to create a more precise and accurate representation of the environment. Huntsberger et al. [Huntsberger 11], present a mapping technique to integrate multi-sensor data that is imperative for enhancing the perceptual capabilities of water vehicles and optimizing navigation strategies.

In the context of a post-flooded environment, generating traversability information is a fundamental aspect of safe navigation. Traversability information refers to the assessment of whether a particular region of the water surface is navigable without encountering obstacles or hazards. Traversability information is crucial for rescue teams for planning optimal paths and avoiding potential dangers. The deployment of multiple sensors facilitates a comprehensive assessment of traversability by capturing various aspects of the environment [Meckel 22]. For instance, visual sensors such as cameras can provide high-resolution images, aiding in the identification of surface-level obstacles, while depth sensors, including sonar, contribute information about underwater topography and potential submerged hazards. The effectiveness of multi-sensor fusion extends beyond real-time navigation to mapping the water environment. Creating detailed maps of the water surface and underwater terrain is essential for planning routes, identifying potential risks, and facilitating post-flood analysis.

Hence, there is a dire need to develop a surface water traversability mapping architecture capable of addressing the abovementioned challenges and producing a robust and accurate map. The deployment of a robust mapping architecture has the potential to empower rescue teams with the capability to plan and navigate efficiently within the extensively destructed post-flood environment. In essence, creating accurate traversability maps is a pivotal solution to facilitate effective rescue operations in complex and dynamic post-flood scenarios.



**Figure 6.1:** Destruction due to post-flood environment. *The illustrations are only for concept generated by AI model*<sup>1</sup>.

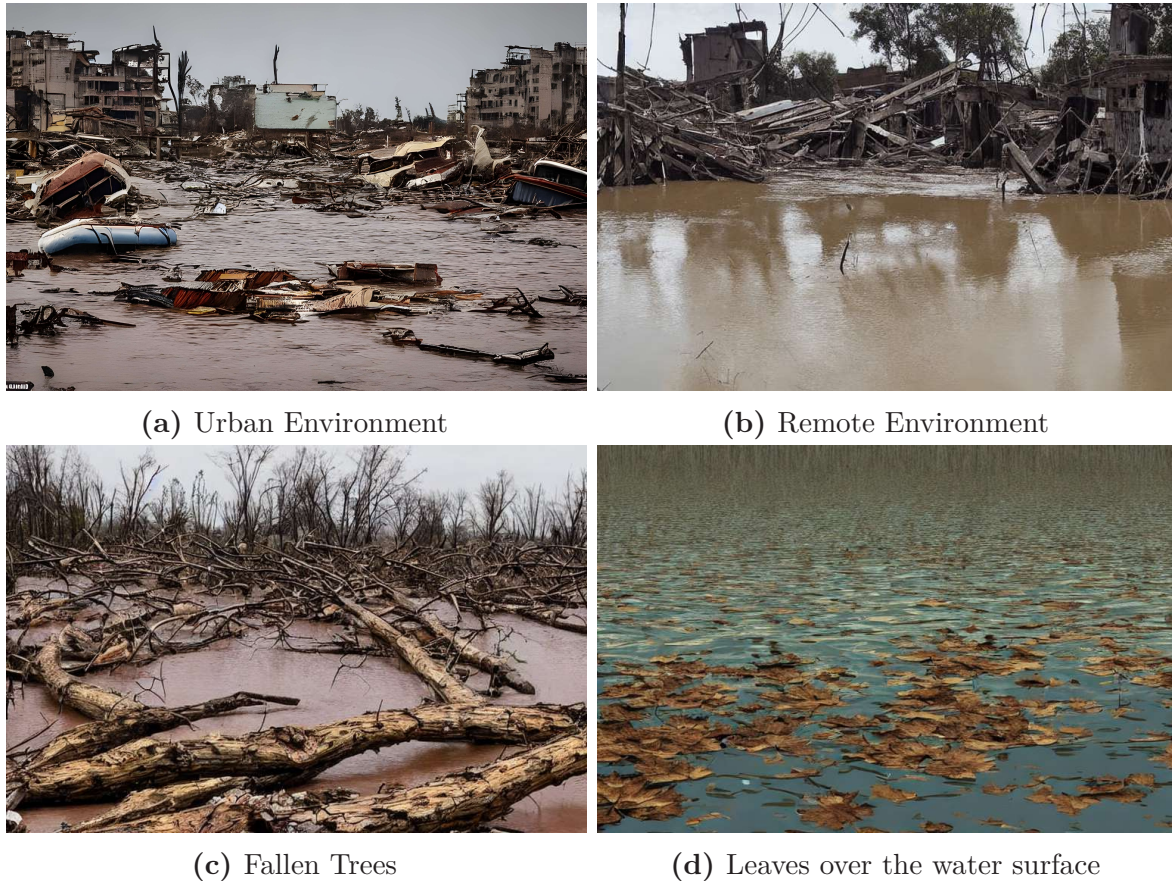
This chapter undertakes a thorough exploration of these challenges, offering a profound understanding of the complexities involved. The assessment encompasses factors ranging from environmental dynamics to the limitations of existing sensor technologies in the context of surface water navigation. To comprehend the surface water environment comprehensively, it is imperative to evaluate the advantages and disadvantages of available sensory modalities. This chapter critically examines these modalities, discusses their capabilities and shortcomings in capturing the diverse and often unpredictable nature of surface water obstacles.

A focal point of this chapter is the introduction of the Surface Water Traversability Mapping (SWTM) framework, an integral component within the broader Shallow Water Traversability Mapping (SWiM) architecture. The SWTM is designed to address the complex task of detecting and mapping surface water obstacles, contributing significantly to the creation of a high-precision traversability map. Section 6.2 provides the details of the SWTM, mentioning its underlying techniques and methodologies. By doing so, it aims to provide not only a conceptual understanding but also practical insights into how the SWTM enhances the accuracy of traversability maps in the challenging domain of surface water navigation.

## 6.1 Challenges in Surface Water Sensing

**Satellite and Aerial Imagery:** The challenges associated with mapping the underwater environment have been comprehensively addressed in section 5.1, underscoring the inherent limitations of satellite and aerial imagery. These imaging modalities, depicted in Figure 6.1, exhibit a notable constraint in their ability to solely capture surface water information, as they lack to penetrate the water's surface. For the specific task of surface water mapping in post-flood environment, the use of satellite or aerial imagery introduces several significant issues, thereby rendering them inadequate for capturing essential details concerning submerged or floating objects. A primary issue in publicly available satellite data lies in its slow update frequency, due to the continuous orbital movement of satellites around the Earth. This slow update rate,

<sup>1</sup><https://deepai.org/machine-learning-model/text2img>



**Figure 6.2:** Debris due to post-flood environment. *The illustrations are only for concept generated by AI model*<sup>2</sup>..

at times extending to several days or even months, poses a substantial limitation in providing real-time data crucial for dynamic environmental conditions. Conversely, the utility of aerial drone imagery is constrained by the limited flight time and the absence of long-distance data transfer capabilities in drones, thereby restricting their applicability for sustained surveillance or comprehensive data recording. Notably, commercially available aerial drones typically exhibit an average flight time of 15-20 minutes, a temporal constraint further increased by variable wind conditions.

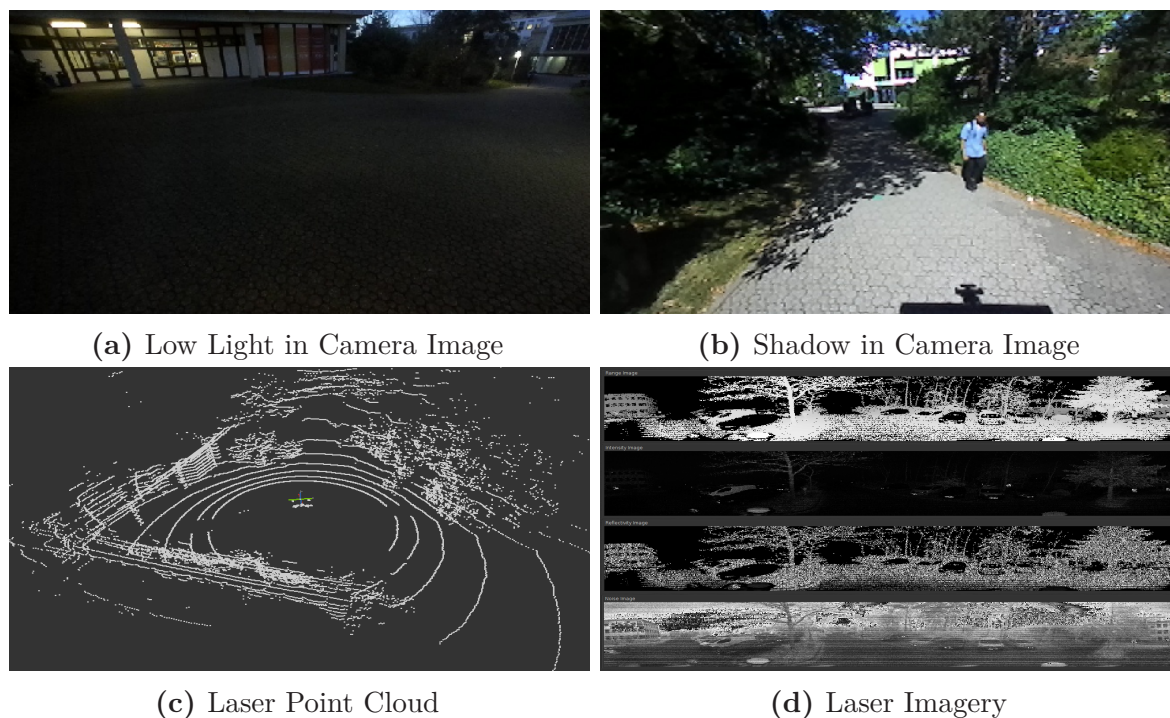
Satellite and aerial imagery encounter challenges in aligning with underwater sensing in post-flood environments, primarily due to different perspectives. Achieving alignment between data captured above and below water surfaces becomes a complex task, suffered by misalignment issues arising from different observation angles. The complex process of detecting common features between these different perspectives to register data introduces significant complications, limiting the seamless integration of satellite or aerial imagery with underwater sonar data. This misalignment poses a substantial obstacle in creating a cohesive representation of the surface and underwater environment during post-flood scenarios. Additionally, satellites and drone imagery can be used for topological path planning but not for local navigation and collision avoidance.

<sup>2</sup><https://deepai.org/machine-learning-model/text2img>

**Random Objects:** In a post-flood environment, rescue teams face numerous challenges arising from the presence of unstructured and unknown objects floating over the surface of water. The complexity is amplified by low visibility and the variable nature of these objects, impeding smooth navigation and effecting rescue efforts. An essential factor for navigating these challenging conditions is a comprehensive understanding of the distinct characteristics presented by these floating objects. This information is pivotal for devising effective navigation strategies and implementing obstacle avoidance measures. Notably, submerged fallen trees and fractured bridges as shown in Figure 6.2 emerge as high-risk elements, necessitating accurate navigation planning, while comparatively smaller objects like floating leaves and wooden blocks may present lesser immediate threats.

**Lack of Map:** Obtaining accurate and up-to-date traversability information of an affected area can be a challenging task, primarily due to the unavailability of standard maps, including widely used platforms such as Google Maps and OpenStreet Maps. These platforms, while valuable for routine navigation, lack the adaptability required to capture and represent the complexities of a highly changing environment. In a post-flood scenario, the absence of predefined routes and the sudden emergence of new obstacles defy the static nature of conventional maps. The immense scale and unpredictability of environmental changes demand real-time information, which these standard maps fail to provide. Consequently, rescue teams find themselves navigating through destructed scenarios, without reliable mapping resources. In this challenging context, their reliance shifts to their prior knowledge and experience gathered through training and past missions. While valuable, this reliance on experiential knowledge has inherent limitations, as flood-induced alterations often introduce unprecedented challenges. The inadequacy of standard mapping platforms and high structural variations to encapsulate the rapidly evolving post-flood environment underscores the critical need for alternative approaches in acquiring accurate and up-to-date traversability information.

In the subsequent section, a comprehensive comparative analysis of cameras and LiDAR (Light Detection and Ranging) sensors unfolds, with a specific focus on their applicability in the complex context of post-flood environments. This analysis extends beyond technical specifications to consider the real-world implications of using cameras and LiDAR sensors in post-flood scenarios. The efficiency of each modality in detecting and accurately mapping unstructured objects is assessed, considering factors like object size, environmental conditions, and the overall reliability of the generated data. Furthermore, the upcoming analysis serves not only to discuss the strengths and limitations of cameras and LiDAR sensors in post-flood environments but also to guide the selection of the most suitable sensing mechanisms for specific operational contexts. The findings contribute valuable insights to the ongoing discourse on optimizing sensing technologies for effective navigation and mapping in the challenging aftermath of floods. Ultimately, the choice between cameras and LiDAR sensors lies on understanding of the environmental conditions, operational requirements, and the desired level of mapping precision in post-flood scenarios.



**Figure 6.3:** Example data from TUK Campus Dataset. The dataset includes multi-sensor data captured at the RPTU Kaiserslautern-Landua campus at different times of day for the autonomous bus project. The sensors include Localization sensors (GNSS, IMU), Stereo cameras, and LiDAR scanners. [Keen 21]

### 6.1.1 Comparative Analysis between Cameras and LiDAR

Environmental sensing, obstacle detection, and mapping in surface water environments necessitate the utilization of diverse sensory modalities. Cameras and LiDAR emerge as prominent choices for object detection in this context. Cameras, serving as passive sensors, capture two-dimensional images, making them adept at detecting visual attributes such as color and texture. The widespread use of cameras in this domain is attributed to their versatility and the wealth of information conveyed through images, which includes valuable textural details absent in LiDAR data. In contrast, LiDARs operate as active sensors by emitting laser beams and measuring the time taken for the light to return, generating two or three-dimensional point clouds. This methodology enables LiDARs to excel in detecting object depth, shape, and distance from the sensor. While cameras are more prevalent in the field due to their extensive exploration in conventional vision and deep-learning object detection, LiDARs provide crucial advantages in terms of precise distance and size measurement.

Despite the widespread use of cameras, challenges persist, particularly in variable light conditions. Low-light situations as shown in Figure 6.3a can compromise object detection via camera imagery, posing a limitation in certain operational scenarios. Similarly, Figure 6.3b shows the variable light conditioning in a single image effecting the overall camera vision. Moreover, camera-based detection systems, such as the one introduced by Sinisterra et al. [Sinisterra 17], may exhibit constraints in identifying randomly positioned destructed obstacles lacking prior knowledge of their characteristics.

LiDARs, on the other hand, offer significant benefits in generating detailed three-dimensional point clouds (shown in Figure 6.3c) that are instrumental in detecting object distance and size. In post-flood environments, where the identification of destroyed infrastructure is critical for mapping traversability, LiDARs provide indispensable insights. However, the maturity of deep learning and object detection algorithms for LiDAR data lags behind that of camera images, partially due to the absence of texture information in LiDAR datasets.

Recent advancements in LiDAR technology have introduced intensity and reflectance imagery as shown in figure 6.3d, resembling features found in camera images. While this facilitates the application of deep learning methodologies to LiDAR data, the inherent limitation of lacking color information remains. Objects are primarily differentiated based on distance, highlighting the ongoing challenges and potential avenues for improvement in LiDAR-based object detection, especially in complex and dynamic environments.

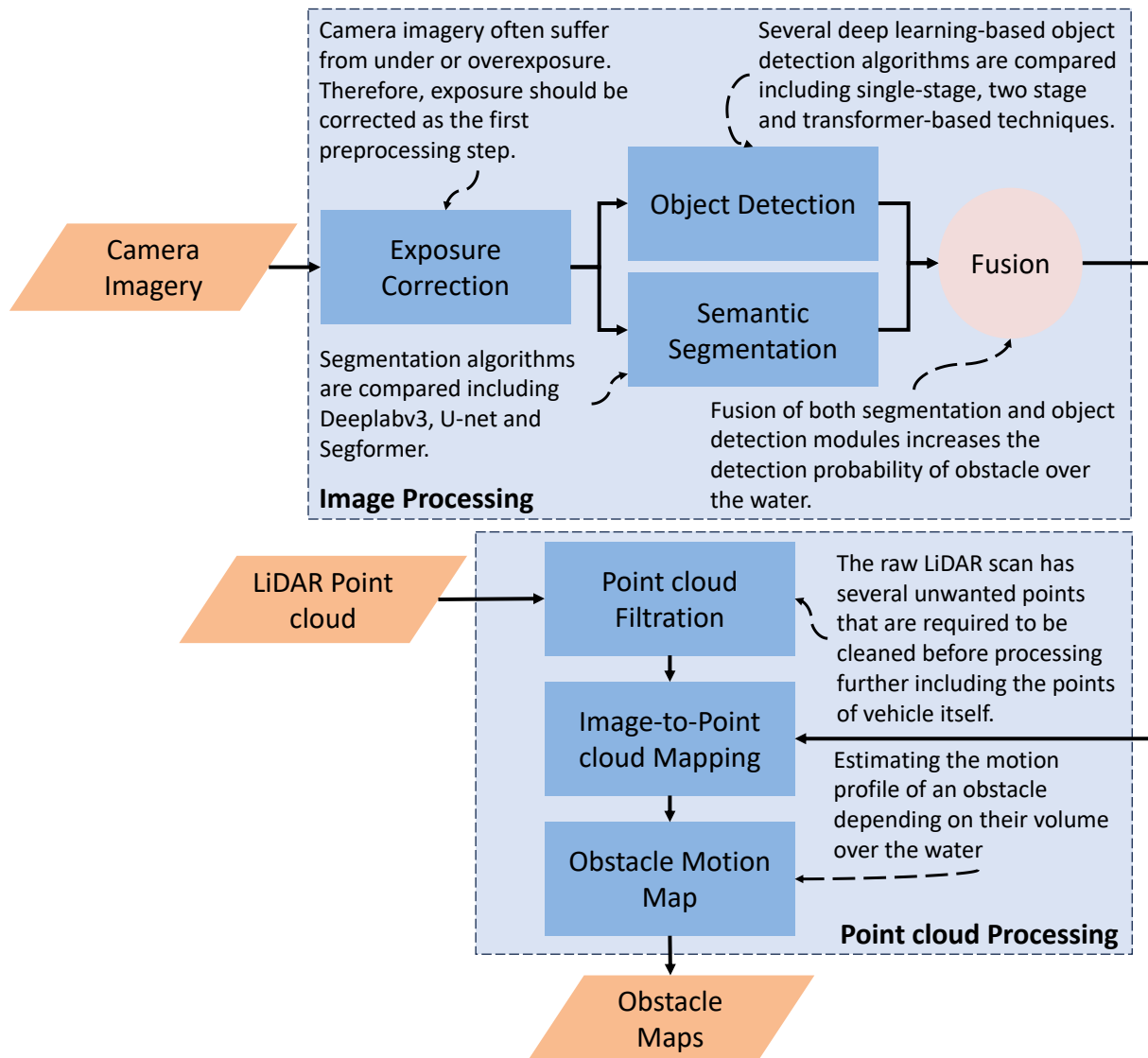
To address these challenges, further research is imperative to develop robust object detection systems that accurately identify obstacles on water surfaces. While cameras maintain their primacy in detection systems, integrating LiDAR point clouds with camera imagery holds promise in enhancing overall effectiveness. This integration can provide a more comprehensive understanding of the environment, leveraging the strengths of both modalities to overcome the limitations inherent in each, ultimately advancing the state of object detection technology in highly dynamic aquatic scenarios.

## 6.2 Surface Water Traversability Mapping Framework

The SWiM (Shallow Water Traversability Mapping) architecture, as proposed, introduces an innovative approach to address the challenges of detecting and mapping surface water objects that pose threats to the safe navigation of surface vehicles. Central to this architecture is the Surface Water Traversability Mapping (SWTM) framework, a sophisticated system designed to fuse the camera and LiDAR data. Unlike conventional approaches that treat data from each sensor in isolation, the SWiM architecture recognizes the need for a holistic fusion of information from both sensors, a crucial consideration in the context of post-flood environments.

As illustrated in Figure 6.4, the SWTM leverages the fusion of Camera and LiDAR data to extract fundamental features of objects present on the water's surface. While existing techniques in the literature often treat data from these sensors separately, such an approach proves inadequate in the face of the unstructured and random objects characteristic of post-flood environments. These objects lack distinct shapes and exhibit random, nonlinear features that effect straightforward learning by state-of-the-art deep learning detection algorithms. Recognizing this challenge, the SWTM adopts a strategy of segmenting image data into basic classes and subsequently filtering objects using point cloud data.

Standard methodologies including state-of-the-art deep learning object detection algorithms may not be optimal in flooded environments. The random and unstructured nature of objects on the water's surface defies the capabilities of such algorithms, which often rely on learned patterns and distinct shapes. In this regard, the SWTM proposes a more adaptive approach, focusing on basic class segmentation and leveraging point cloud data for object detection. This strategy aligns with the unique demands of post-flood scenarios, where traditional algorithms may prove infeasible.



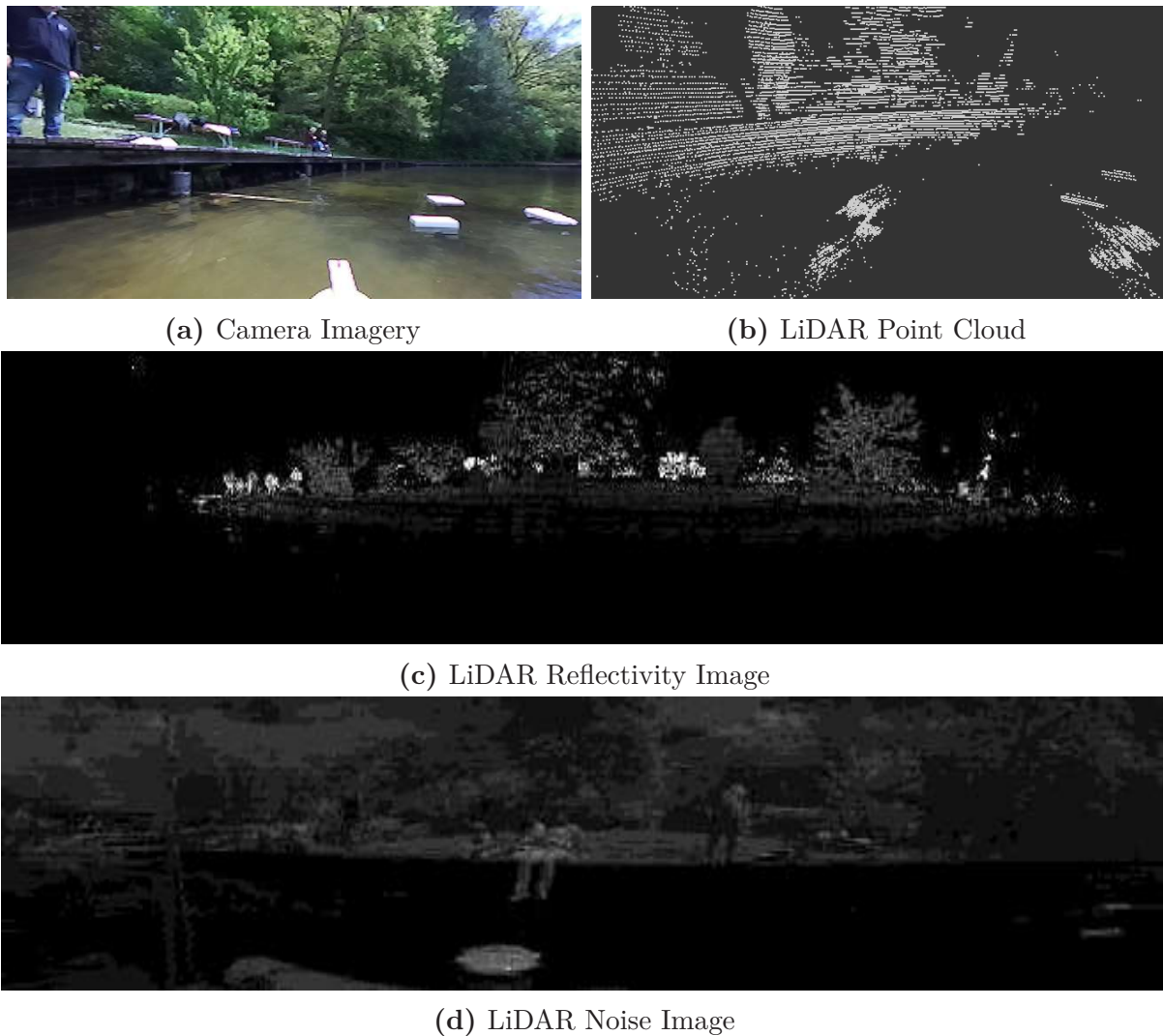
**Figure 6.4:** Surface Water Traversability Mapping (SWTM) Framework - Block diagram of the proposed methodology. The methodology uses camera imagery and LiDAR point cloud to detect obstacles over the water surface. The obstacles detected in camera imagery are mapped in the LiDAR point cloud.

Building upon prior research, the SWiM architecture draws inspiration from noteworthy works in the field. Wolf et. al. [Wolf 22], introduce a behavior-based perception system tailored for off-road vehicles. In this work, the authors employ YOLO (You Only Look Once) detection on camera images to identify potential obstacles. Subsequently, utilizing inverse-perspective techniques, they correlate the detected objects in point cloud data. This approach demonstrates a paradigm for integrating information from multiple sensors to enhance obstacle detection and environmental perception.

Another relevant work [Wang 20a], contributes to the fusion of modalities in the context of surface water navigation. Here, the authors similarly employ YOLO detection on camera images to identify potential obstacles. The essential novelty lies in their subsequent efforts to locate the same objects within point cloud data. This recognition of the importance

of fusing data from different sources aligns with the principles underlying the SWiM architecture.

The SWiM architecture represents an innovative effort in optimizing traversability mapping in post-flood environments. By integrating camera and LiDAR data and adopting a distinct approach to object detection, the SWTm within the SWiM architecture overcomes the limitations of existing methodologies. The fusion of modalities, inspired by prior research, offers a comprehensive solution to the challenges posed by unstructured and random objects in flooded environments. As the SWiM architecture advances, its impact on the efficacy of surface water vehicle navigation and disaster response in post-flood scenarios holds significant promise. Figure 6.5 shows the sensory data required by SWTm.



**Figure 6.5:** Example data from MASTER Dataset. The dataset includes multisensor and multidomain data captured in different lakes, rivers, and port areas. The sensors include GNSS, a stereo camera, 128 lines of 3D LiDAR, and an underwater forward-looking imaging sonar.

### 6.2.1 Image Processing

Camera imagery stands out as the most crucial sensory modality in the domain of robotics, serving as the primary source of visual information for robots. Despite its paramount



importance, several challenges, with noise being a predominant concern, confront camera imagery. This noise can originate from various factors, including suboptimal lighting conditions, camera quality, and the compression techniques employed during image transmission. While there exist numerous techniques and advanced sensors for mitigating noise related to camera quality and compression, addressing challenges associated with lighting conditions remains a persistent issue.

One of the significant challenges arising from variable lighting conditions is the occurrence of over and underexposure in images. When the camera is directed towards a prominent light source, such as the sun, the high intensity of light floods the sensor, resulting in overexposed images. Conversely, if the light source is positioned behind the camera, images tend to be underexposed due to inadequate light reaching the sensor. Hence, achieving exposure correctness becomes the foremost step in any image processing module. This correction is pivotal for maintaining image quality and ensuring that subsequent analysis and processing are founded on accurately exposed visual data. Notably, exposure correctness is a multifaceted problem, especially in scenarios where lighting conditions fluctuate dynamically, as in outdoor environments or during night-time operations.

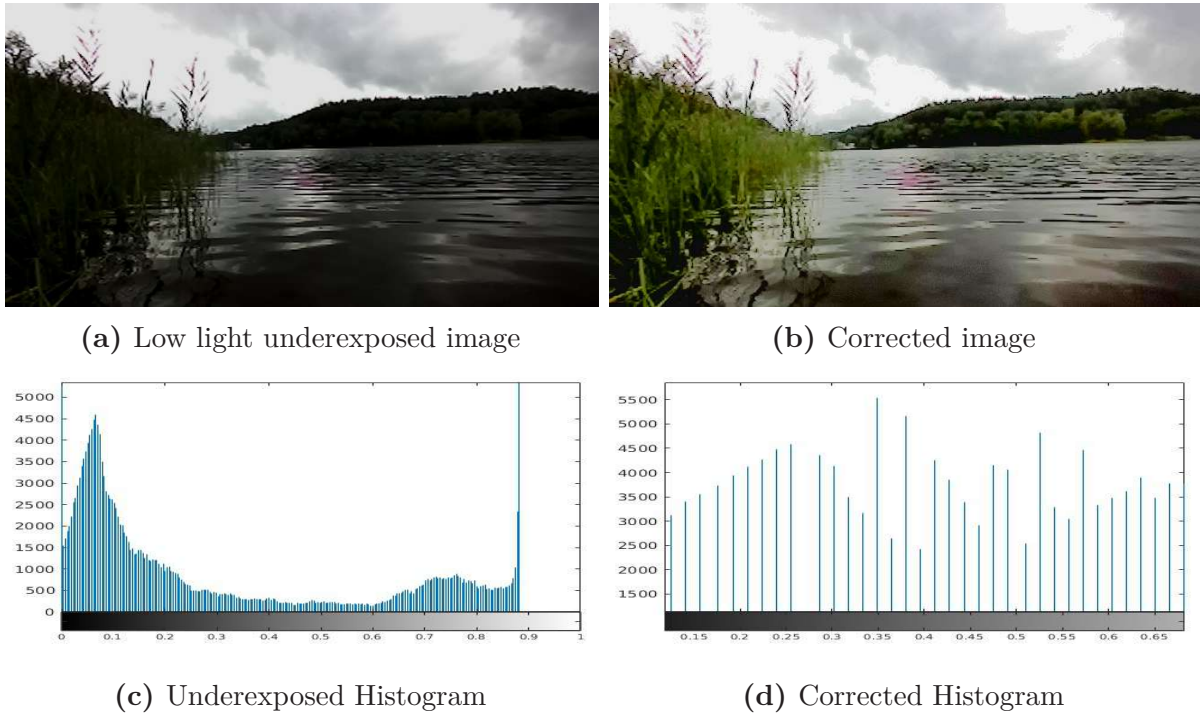
In the context of night-time image processing, overcoming the challenges posed by low-light conditions becomes particularly crucial. Night-time operations intensify the inherent difficulties associated with variable lighting. To address this, sophisticated image processing techniques tailored for low-light scenarios are imperative. Algorithms designed for night-time image processing often involve enhancing contrast, reducing noise, and adapting exposure settings dynamically. These measures collectively contribute to improving the visibility and interpretability of images captured under challenging lighting conditions, ensuring that robotic systems can navigate and operate effectively in diverse environments.

Furthermore, the SWTMM integrates advanced object detection and segmentation algorithms, aiming to identify and characterize unstructured objects within the visual field. Object detection involves detecting and locating specific entities or features in an image, while segmentation defines the boundaries of these objects, providing a clear information for subsequent analysis. The robust implementation of these algorithms within the SWTMM enables the system to comprehend and interact with its environment effectively.

### **Exposure Correction:**

Capturing camera images from the water surface introduces several exposure issues, comprising both overexposure and underexposure scenarios, thereby impeding their effectiveness in object detection applications. The challenges are notably increased by the reflective properties of water, as sunlight can produce overexposed images that obscure details crucial for object detection. Conversely, under low-light conditions, the camera may capture images with insufficient information, compromising the accuracy of subsequent object detection processes. These exposure anomalies necessitate accurate filtering of images based on their exposure levels to enhance their utility for object detection. Establishing a defined threshold range for image exposure becomes imperative for facilitating exposure correction and ensuring optimal conditions for subsequent analytical processes.

Figure 6.6 and 6.7 provides a visual representation of the exposure challenges encountered in camera images captured at Gelterswoog Lake near RPTU Kaiserslautern Landau. The



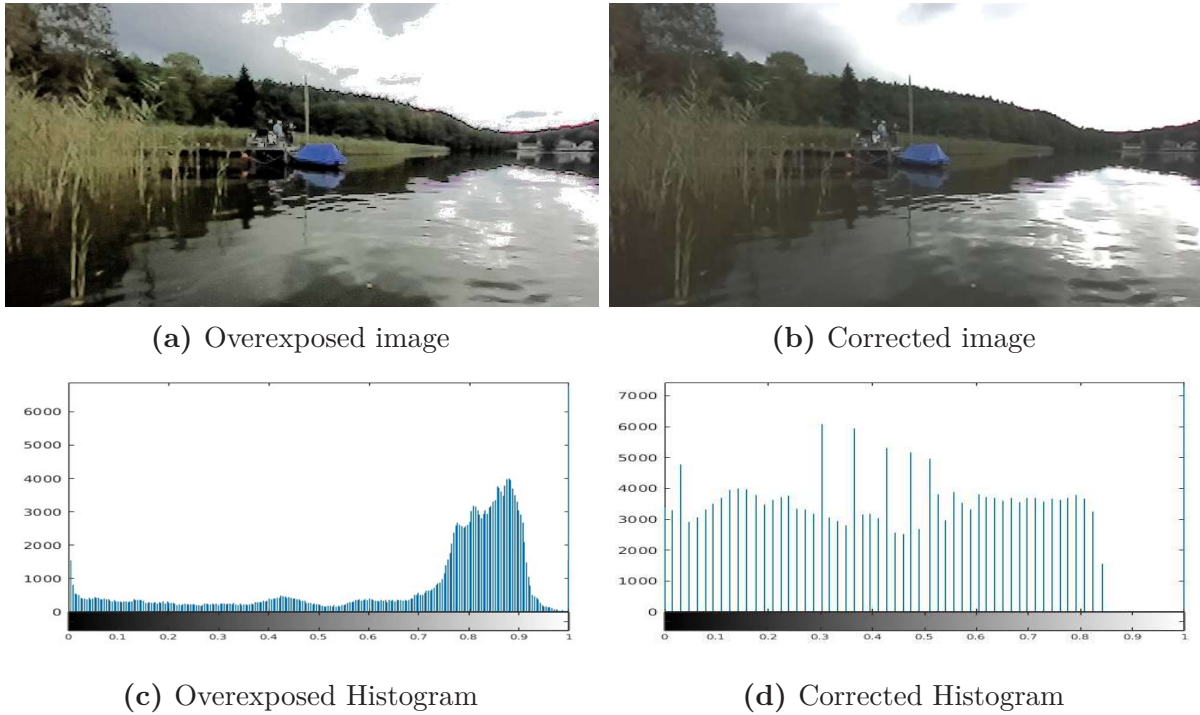
**Figure 6.6:** Exposure Correction in Underexposed Image. (a) shows an original image from the MASTER dataset characterized by low light and underexposure, resulting in a dim and detail-obscured scene. (b) displays the same image after applying exposure correction, yielding a visually clearer and more balanced image with enhanced details. The histograms in (c) and (d) represent the distribution of pixel intensities before and after correction, respectively. The initial histogram (c) skews left, indicating underexposure, while the corrected histogram (d) shows a more even distribution across the range, confirming the successful correction of the image.

images showcase instances of overexposure and low-light conditions, highlighting the real-world complexities inherent in water environments. Analyzing these images underscores the importance of implementing exposure-based filtering mechanisms to enhance the quality and utility of the captured data for subsequent analytical tasks.

To systematically address exposure-related challenges, captured images are selected and forwarded for further processing based on their exposure level. This entails establishing a predefined threshold range within which the exposure of images should fall to be considered suitable for subsequent processing. This threshold range serves as a criterion for exposure correction, ensuring that the images utilized for object detection are within optimal exposure conditions.

The assessment of image exposure levels involves a detailed analysis of histograms. Initially, the RGB image undergoes conversion into the Hue Saturation Value (HSV) domain using equation 6.1 - 6.3, where the Value (V) component governs color intensity.

$$H = \begin{cases} \cos^{-1} \frac{(R - \frac{G}{2} - \frac{B}{2})}{\sqrt{R^2 + G^2 + B^2 - RG - RB - GB}}, & \text{if } G \geq B, \text{ or} \\ 360 - \cos^{-1} \frac{(R - \frac{G}{2} - \frac{B}{2})}{\sqrt{R^2 + G^2 + B^2 - RG - RB - GB}}, & \text{if } B > G \end{cases} \quad (6.1)$$



**Figure 6.7:** Exposure Correction in Overexposed Image. (a) shows an original image from the MASTER dataset affected by high intensity from opposing sun, resulting in an overexposed image. (b) displays the same image after applying exposure correction, yielding a visually clearer and more balanced image with enhanced details. The histograms in (c) and (d) represent the distribution of pixel intensities before and after correction, respectively. The initial histogram (c) skews right, indicating overexposure, while the corrected histogram (d) shows a more even distribution across the range, confirming the successful correction of the image.

$$S = \begin{cases} 1 - \frac{\max(R,G,B)}{\min(R,G,B)}, & \text{if } M > 0 \\ 0, & \text{if } M = 0 \end{cases} \quad (6.2)$$

$$V = \frac{\max(R, G, B)}{255} \quad (6.3)$$

Subsequently, the histogram of the Value component is computed to evaluate the distribution of intensities. This computation includes two critical ranges: 1) the *Feasible Range* (FR) and 2) the *Satisfactory Range* (SR). The FR is pivotal in the preliminary filtration of images, delineating the threshold beyond which inadequate information impedes effective histogram equalization, thereby obstructing enhanced contrast. Images with a mean and variance of the histogram beyond the FR are systematically excluded from subsequent processing due to the inherent challenges associated with object detection in such instances. SR is the range within which an image is considered in satisfactory contrast for subsequent processing and do not require the histogram normalization. Therefore, an image is deemed appropriate for histogram equalization if its histogram mean and variance fall within the FR and extend beyond the predefined SR. The determination of these ranges is contingent on the average mean and variances derived from histograms of 100 challenging images. This dual-range strategy ensures the selection of images with the potential for

successful histogram equalization, effectively excluding those with insufficient information or problematic exposure.

The significance of exposure correction extends beyond immediate improvements in object detection. Enhanced image quality contributes to the overall robustness of analytical processes, supporting downstream tasks such as image segmentation, feature extraction, and classification. Moreover, in applications where real-time decision-making is crucial, the prompt correction of exposure issues ensures that the data used for analysis is consistently reliable, reducing the risk of inaccuracies in time-sensitive scenarios.

## Object Detection:

Detecting objects in post-flood environments is a challenging task because of the unpredictable nature of these objects. These objects vary in size, shape, and position and can either float on the water's surface or be submerged underwater, including fallen trees, broken bridges, debris, and floating leaves. Some of these objects, such as floating leaves, may not be critical for the safe navigation of surface water vehicles, while others, like fallen trees and hard debris, can pose a serious threat. Traditional vision-based detection methods are not ideal for identifying these objects due to their random nature. Accurately detecting these floating objects is difficult because of the dynamic and often turbulent water conditions and the presence of occlusions, such as shadows and reflections, which further obscure the objects. Therefore, to address these complexities in such challenging context, this thesis has integrated deep learning models that have demonstrated remarkable capabilities in detecting complex patterns and features, making them particularly well-suited for the complex task of object detection over water surfaces.

In this section, three distinct deep learning architectures are employed and rigorously compared to present their respective strengths and weaknesses, providing a comprehensive understanding of the detection trade-offs inherent in each technique.

The selected deep learning architectures fall into three categories: *single-stage detectors*, *two-stage detectors*, and *transformer-based detectors*. Each category represents a distinct paradigm in object detection, leveraging various architectural principles and methodologies. The single-stage detectors, such as *YOLO* (You Only Look Once), aim for efficiency by performing object localization and classification in a single forward pass. Two-stage detectors, including *Faster R-CNN* (Region-based Convolutional Neural Network) and Mask R-CNN, follow a more elaborate process, first proposing regions of interest and subsequently refining these proposals for accurate object detection. Transformer-based detectors, representing a newer paradigm, leverage attention mechanisms to capture contextual information effectively, as seen in models like *DETR* (DEtection Transformer). Each of these architectures are explained in Appendix E.

The evaluation of these architectures is conducted on two distinct datasets. The first is the widely used Marine Obstacle Detection Dataset (MODD/MODD2) dataset explained in section 4.1.1, providing a benchmark for assessing the performance of the detectors in a standardized setting. Additionally, a novel dataset, denoted as the Multi-modal Shallow waTER (MASTER) dataset (discussed in section 4.1.2), is introduced. The MASTER dataset is precisely captured across various waterbodies, including lakes, riversides, and ports, ensuring diversity in environmental conditions. This dataset aims to enhance

the robustness of the evaluation by encompassing scenarios that may not be adequately represented in existing datasets.

The training process involves exposing the selected deep learning architectures to both the MODD/MODD2 and MASTER datasets, facilitating the adaptation of the models to the specific challenges posed by post-flood environments. The comparative analysis encompasses standard metrics, including average precision, and recall, to comprehensively evaluate the performance of each detector. Precision reflects the accuracy of positive predictions, and recall measures the ability to capture all positive instances. The results are presented, considering the strengths and weaknesses of each architecture under various conditions, shedding light on their applicability in post-flood environments.

### Data augmentation:

In the context of this study, where multiple frames in the dataset come from the same video sequences, a process of computing data dissimilarities among images within the dataset becomes imperative. The objective is to anticipate the intrinsic variations between images and subsequently define a threshold to identify dissimilar frames. This involves the computation of structural similarity index (SSI), peak signal to noise ratio (PSNR) and mean square noise (MSE), with the determined threshold values presented in Table 6.1. The chosen threshold values are involved in the systematic selection of dissimilar images, a critical step in enhancing the diversity and robustness of the dataset.

**Table 6.1:** Threshold Metrics for Image Selection in Video Sequences. This table presents threshold values for SSI, PSNR, and MSE to filter out similar images, ensuring that only dissimilar images are used for efficient training.

| SSI  | PSNR | MSE |
|------|------|-----|
| 0.63 | 25.4 | 178 |

To further augment the dataset and amplify its variability, ten distinct augmentation techniques are systematically employed. These augmentation techniques play a pivotal role in introducing additional diversity to the dataset, thereby enhancing the model’s capacity to generalize across a multiple real-world scenarios. The augmentation techniques employed are pivotal not only in increasing dataset variability but also in fortifying the model’s resilience to noise during the training phase.

Each augmentation technique is characterized by a unique set of transformations applied to the original images. These transformations include, but are not limited to, rotations, flips, zooms, and changes in brightness and contrast. By subjecting the images to these transformations, the dataset is effectively expanded, presenting the model with a more extensive range of visual variations. These augmentations are especially valuable in scenarios where the availability of labeled data is limited, as they simulate diverse conditions that the model might encounter in real-world applications.

### Comparative Analysis

Table 6.2 defines the key features and hyperparameter configurations of each object detection model under consideration. Each model undergoes training across 500 epochs,

with the image dimensions standardized at 640x640 pixels. Table 6.3 offers a detailed quantitative comparison of the performance of each algorithm, trained independently. A notable observation is the superior average precision (AP) of the single-stage YOLOX model over both Faster R-CNN and DETR. When comparing the YOLOX small and large models, the small model excels in AP, while the large model demonstrates a higher average recall.

**Table 6.2:** Object Detection Model Hyper-parameter Specifications. This table enumerates the hyper-parameters utilized across all object detection models discussed in the thesis. These hyper-parameters are optimized to ensure a uniform training, facilitating fair comparisons between the performance outcomes of the various models.

| <i>Image Scale</i> | <i>Max Epochs</i> | <i>Interval</i> | <i>Batch Size</i> |
|--------------------|-------------------|-----------------|-------------------|
| 640                | 500               | 10              | 8                 |

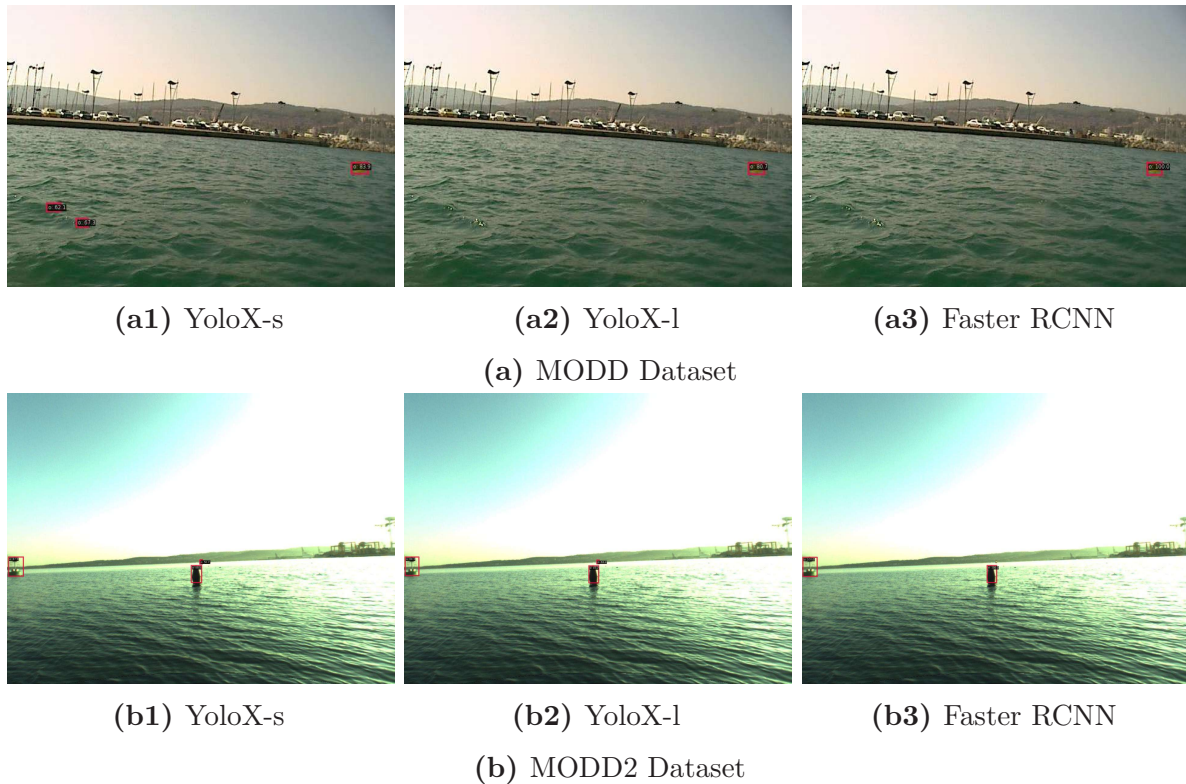
**Table 6.3:** Comparative analysis of object detection techniques across various datasets. This table indicates the Mean Average Precision (mAP) and Recall at Intersection over Union (IoU) thresholds of 50% and 50% – 95% respectively.

| <b>Datasets</b> | <b>YoloX S</b> |               | <b>YoloX L</b> |               | <b>Faster RCNN</b> |               | <b>Detr</b> |               |
|-----------------|----------------|---------------|----------------|---------------|--------------------|---------------|-------------|---------------|
|                 | <i>mAP</i>     | <i>Recall</i> | <i>mAP</i>     | <i>Recall</i> | <i>mAP</i>         | <i>Recall</i> | <i>mAP</i>  | <i>Recall</i> |
| MODD            | 0.98           | 0.66          | 0.98           | 0.66          | 0.92               | 0.60          | 0.056       | 0.012         |
| MODD2           | 0.96           | 0.6           | 0.94           | 0.62          | 0.40               | 0.35          | 0.102       | 0.033         |
| MASTER          | 0.88           | 0.72          | 0.83           | 0.68          | 0.86               | 0.61          | 0.11        | 0.08          |
| MIX             | 0.72           | 0.61          | 0.75           | 0.62          | 0.79               | 0.58          | 0.19        | 0.15          |

The comparative study between the YOLOX small and large models reveals a subtle distinction: while the small model achieves a higher AP, indicative of its precision in correctly identifying positive cases, the large model excels in average recall, demonstrating its proficiency in minimizing missed detections. In the context of ensuring safe navigation and mapping in post-flood environments, the priority shifts towards minimizing false negatives rather than false positives. In practical scenarios, detecting an additional, non-existent obstacle (false positive) is typically less detrimental than the failure to detect an actual obstacle (false negative). The latter can lead to hazardous situations, particularly in environments where obstacles can be unpredictable and varied due to the aftermath of flooding.

Given these considerations, an algorithm that exhibits higher sensitivity to false negatives – one with a higher recall – is deemed more suitable for the task at hand. Such an algorithm’s ability to reliably identify potential obstacle is crucial in ensuring safe navigation and effective mapping in post-flood scenarios. This criterion leads to the selection of the YOLOX large model as the preferred object detection architecture for this application. The large model’s robust recall makes it adept at detecting a wide range of obstacles, which is critical in environments where the terrain and obstacles can be significantly altered by flooding.

This decision is supported by the understanding that in post-flood scenarios, the environment is often drastically changed, with new obstacles emerging that were not present before. The ability of the YOLOX large model to identify these obstacles accurately,



**Figure 6.8:** Qualitative comparison of object detection algorithms on the MODD and MODD2 dataset. (a1) shows results using YoloX-s, which achieves 80% precision but with a higher number of true negatives. (a2) depicts the YoloX-l algorithm, which also reaches 80% precision but with no true negatives. (a3) illustrates the performance of the Faster RCNN algorithm, which has a lower precision of 70% and no true negatives. (b1) shows results using YoloX-s, which achieves 76.7% precision but with no true negatives. (b2) depicts the YoloX-l algorithm, which also reaches 75% precision but with no true negatives. (b3) illustrates the performance of the Faster RCNN algorithm, which has a lower precision of 66.7% and no true negatives. Each algorithm’s detection capability is visually represented by red bounding boxes around the detected objects.

despite the challenges posed by such environments, is a key factor in its selection. Its higher recall ensures that fewer obstacles are missed, a factor that could be crucial in rescue and recovery operations, as well as in ensuring the safety of navigation in these areas.

### Semantic Segmentation:

In the context of a post-flood environment, the inherent unpredictability and randomness of objects present challenges for their effective detection on the water’s surface. The concept of semantic segmentation in this domain revolves around distinguishing essential entities common to both normal and post-flood environments, namely sky, water, and unknown obstacles. This section delves into the semantic segmentation of the water surface, focusing on these three entities. Rather than explicitly detecting individual objects, the objective is to identify the water area and integrate this information with detected objects, thereby enhancing the understanding of traversable and non-traversable regions.

Three distinct segmentation architectures have been implemented and evaluated across various datasets, including MaStr1325 (Section 4.1.1), ROSEBUD (Section 4.1.1), USVInland (Section 4.1.1), and MASTER (Section 4.1.2). The subsequent section provides insights into the applied data augmentation techniques, followed by a comprehensive exploration of the segmentation architectures and their corresponding results. This multifaceted approach aims to enhance the robustness of semantic segmentation in water surface environments, offering valuable insights into the classification of essential entities and the overall delineation of traversable and non-traversable areas within the context of post-flood scenarios. Table 6.4 provides the salient features and hyper-parameters of each model.

**Table 6.4:** Semantic Segmentation Detection Model Hyper-parameter Specifications. This table enumerates the hyper-parameters utilized across all segmentation models discussed in the thesis. These hyper-parameters are optimized to ensure a uniform training, facilitating fair comparisons between the performance outcomes of the various models.

| <i>Image Scale</i> | <i>Backbone</i> | <i>Max Epochs</i> | <i>Interval</i> | <i>Batch Size</i> |
|--------------------|-----------------|-------------------|-----------------|-------------------|
| 640                | Resnet          | 500               | 10              | 8                 |

### Comparative Analysis

The performance metrics in Tables 6.5, through 6.8 provide a rich dataset for comparing the efficacy of three semantic segmentation models – DeepLabV3+, UNet, and Segformer – across different datasets and classes. These tables highlight how each model performs in terms of Intersection over Union (IoU) and Accuracy (Acc) across various environmental conditions and classes, which are crucial for applications in environmental monitoring, autonomous navigation, and land cover classification. Below is a detailed analysis of the results from each table.

**Aerial Fluvial Dataset (Table 6.5)** - The DeepLabV3+ model demonstrates strong performance on the Aerial Fluvial Dataset with an IoU of 94.43 for the Environment class and 91.39 for Water, suggesting it is capable of distinguishing between these two classes with high precision. The Accuracy metric is also high, but it is notable that there is a slight decrease in performance for the Water class compared to the Environment, which might be due to the similar textural features present in both classes that can sometimes be challenging to differentiate. Besides, UNet shows a competitive performance but slightly lower than DeepLabV3+ in terms of IoU for both classes. However, it maintains high Accuracy, particularly for the Water class. This could indicate that while UNet may misclassify some pixels between the classes, it remains reliable overall. Furthermore, Segformer exhibits a similar trend with IoU and Accuracy, performing slightly lower than DeepLabV3+ but still maintaining high metrics. Its relatively balanced performance in both IoU and Accuracy suggests that it is a robust model for segmenting complex scenes in aerial imagery where the distinction between land and water can be subtle.

**Mastr1325 Dataset (Table 6.6)** DeepLabV3+ on the MASTR1325 Dataset shows a considerable variation in IoU across different classes. It performs exceptionally well for Water and Sky but struggles with the Unknown class, indicating that it may have difficulties with classes that are not well-represented or defined in the training data. Contrarily, UNet displays a more balanced IoU across the classes than DeepLabV3+, though it also shows a significant drop in performance for the Unknown class. Its Accuracy metrics are



**Table 6.5:** Performance comparison of Semantic Segmentation models over Aerial Fluvial Dataset, based on two metrics *Intersection over Union (IoU)* and *Accuracy (Acc)* for two classes: Environment and Water.

| Classes     | DeepLabV3+ |            | Unet       |            | Segformer  |            |
|-------------|------------|------------|------------|------------|------------|------------|
|             | <i>IoU</i> | <i>Acc</i> | <i>IoU</i> | <i>Acc</i> | <i>IoU</i> | <i>Acc</i> |
| Environment | 94.43      | 96.28      | 93.82      | 95.21      | 96.14      | 97.76      |
| Water       | 91.39      | 96.84      | 90.65      | 97.63      | 93.91      | 97.3       |

relatively high for Environment, Water, and Sky, confirming its efficacy for common classes but highlighting potential improvement areas for more ambiguous categories. Notably, Segformer’s performance is consistent with UNet’s but tends to have slightly better IoU and Accuracy in most classes except for the Unknown. This model’s architecture may provide it with an edge in generalizing across various classes, even those that are less distinct.

**Table 6.6:** Performance comparison of Semantic Segmentation models over Mastr1325 Dataset, based on two metrics *Intersection over Union (IoU)* and *Accuracy (Acc)* for four classes: Environment, Water, Sky, and Unknown.

| Classes     | DeepLabV3+ |            | Unet       |            | Segformer  |            |
|-------------|------------|------------|------------|------------|------------|------------|
|             | <i>IoU</i> | <i>Acc</i> | <i>IoU</i> | <i>Acc</i> | <i>IoU</i> | <i>Acc</i> |
| Environment | 91.45      | 96.52      | 90.25      | 96.62      | 91.64      | 96.7       |
| Water       | 98.84      | 99.19      | 98.54      | 98.97      | 99.07      | 99.46      |
| Sky         | 99.03      | 99.35      | 98.98      | 99.41      | 99.33      | 99.59      |
| Unknown     | 49.59      | 68.22      | 46.24      | 64.81      | 47.75      | 64.42      |

Rosebud Dataset (Table 6.7) DeepLabV3+ achieves impressive IoU and Accuracy scores on the Rosebud Dataset, with over 99% for the Environment class and nearly as high for Water. This indicates a strong capability for segmenting well-defined classes with distinct features. UNet matches DeepLabV3+ in IoU for both classes and surpasses it in Accuracy for the Water class, indicating its strength in segmenting images with high textural contrast. Segformer shows a slight decrease in IoU for the Environment class compared to DeepLabV3+ and UNet but a marginal improvement in the Water class. Its high Accuracy for both classes suggests it has a strong ability to correctly classify pixels.

**Table 6.7:** Performance comparison of Semantic Segmentation models over Rosebud Dataset, based on two metrics *Intersection over Union (IoU)* and *Accuracy (Acc)* for two classes: Environment and Water.

| Classes     | DeepLabV3+ |            | Unet       |            | Segformer  |            |
|-------------|------------|------------|------------|------------|------------|------------|
|             | <i>IoU</i> | <i>Acc</i> | <i>IoU</i> | <i>Acc</i> | <i>IoU</i> | <i>Acc</i> |
| Environment | 99.21      | 99.53      | 99.21      | 99.57      | 99.33      | 99.66      |
| Water       | 98.86      | 99.53      | 98.87      | 99.48      | 99.04      | 99.53      |

UVInland Dataset (Table 6.8) DeepLabV3+ presents an excellent IoU for both Environment and Water classes, indicating a high degree of overlap between the predicted segmentations and the ground truth. Its Accuracy is also high, showcasing its strength in segmenting

images with clarity and distinct features. UNet has a slightly lower IoU compared to DeepLabV3+ but maintains a high Accuracy score. This could imply that while its segmentation may not be as precise as DeepLabV3+'s, it is still accurate in its overall classification. Segformer's performance is comparable to UNet's but with a slightly better IoU for the Water class. Its high Accuracy scores for both classes affirm its capability as a reliable segmentation model.

**Table 6.8:** Performance comparison of Semantic Segmentation models over UvInland Dataset, based on two metrics *Intersection over Union (IoU)* and *Accuracy (Acc)* for two classes: Environment and Water.

| Classes     | DeepLabV3+ |            | Unet       |            | Segformer  |            |
|-------------|------------|------------|------------|------------|------------|------------|
|             | <i>IoU</i> | <i>Acc</i> | <i>IoU</i> | <i>Acc</i> | <i>IoU</i> | <i>Acc</i> |
| Environment | 97.5       | 98.81      | 94.62      | 95.73      | 98.47      | 99.19      |
| Water       | 96.88      | 98.32      | 93.52      | 98.53      | 98.08      | 99.08      |

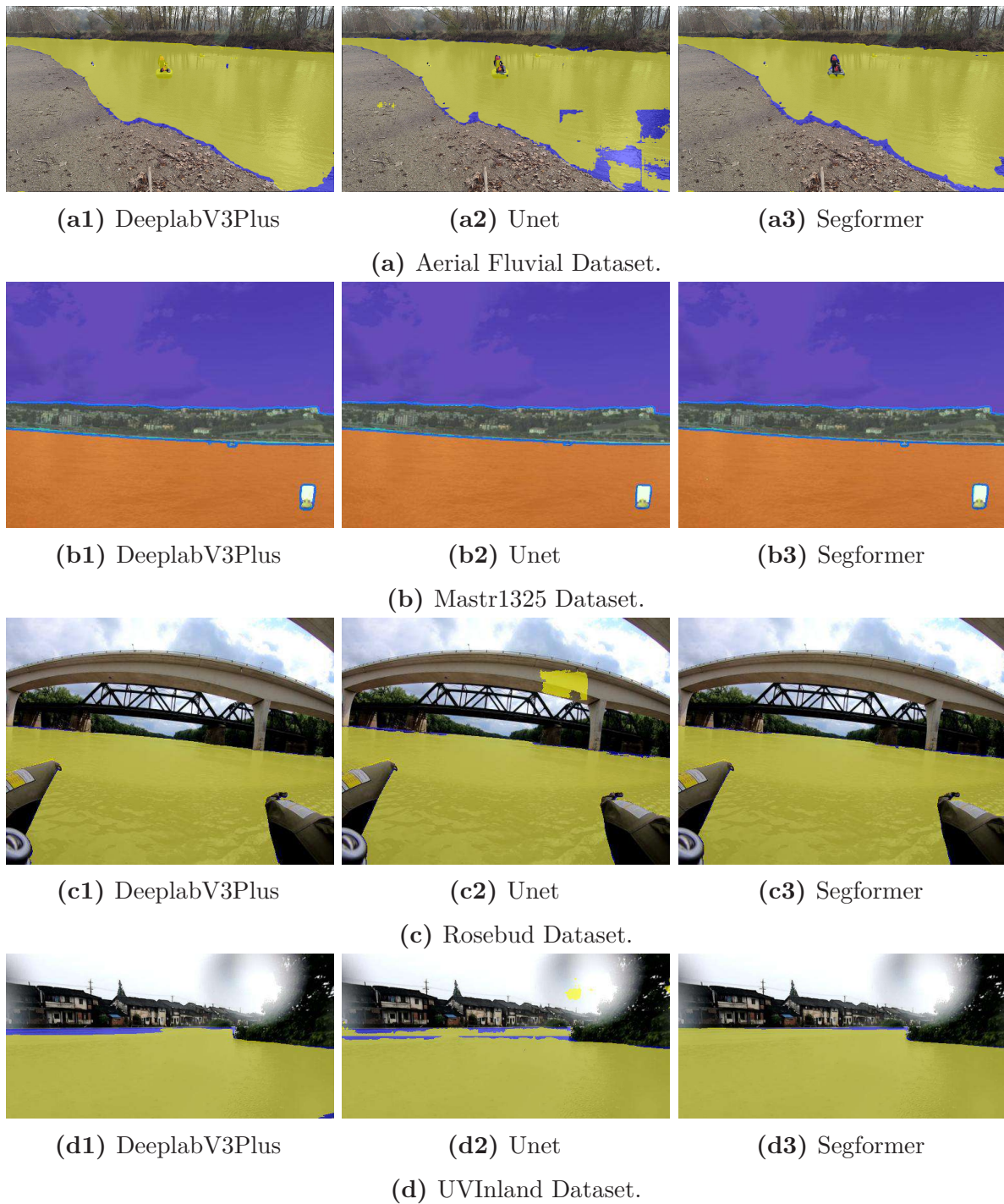
Figure 6.9 provides a visual representation of the semantic segmentation results from three different models – DeepLabV3+, UNet, and Segformer – applied to four different datasets: Aerial Fluvial, MASTR1325, Rosebud, and UvInland. The segmentation quality is color-coded, with yellow representing the water as segmented by the models and blue indicating the ground truth for water. In the MASTR1325 dataset, the ground truth is overlaid with multicolored segmented images. Across all datasets, DeepLabV3+ consistently provides high-fidelity segmentation with minimal misclassification. UNet, while generally accurate, shows a tendency toward under-segmentation across several datasets. Segformer demonstrates strong segmentation capabilities, often rivaling or even surpassing DeepLabV3+ in terms of alignment with the ground truth.

The visual analysis suggests that while all models perform well, the choice between them may come down to specific application needs. DeepLabV3+ would be suitable for scenarios where precision is paramount. UNet could be favored in situations where model simplicity and efficiency are more critical, despite some potential under-segmentation. Segformer's strong performance suggests its suitability for tasks requiring a balance between precision and generalization to various features within an image. Furthermore, this visual analysis underscores the importance of examining both quantitative metrics and qualitative visual outputs when evaluating semantic segmentation models, ensuring that the selected model is well-suited to the specific characteristics and requirements of the task at hand.

## 6.2.2 Point cloud Processing

The 3D LiDAR system is instrumental in generating a point cloud, a collection of points that collectively forms a comprehensive 3D representation of the surrounding environment. The quantity of points generated by a LiDAR system is contingent upon the number of channels it possesses. In essence, the greater the number of channels, the higher the resolution of the point cloud, at the cost of increased storage requirements. From a safety perspective, this work prioritizes the use of a high-resolution 128-channel 3D LiDAR, affording a more intricate and precise understanding of the environment.

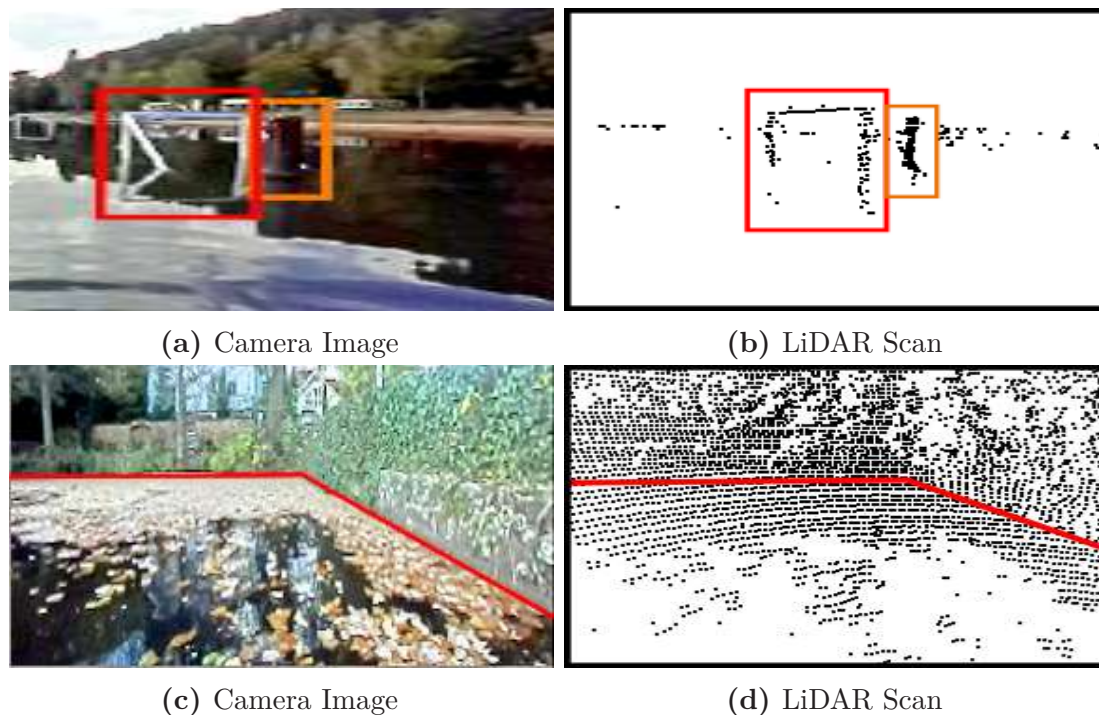
While the richness of information in a 3D point cloud is invaluable, it poses challenges, primarily due to the massive storage requirements for concatenating individual frames into



**Figure 6.9:** Visual Comparison of Semantic Segmentation Models. The Figure showcases the semantic segmentation capabilities of three models—DeepLabV3+, UNet, and Segformer—across four diverse datasets. Segmentation accuracy is visualized with yellow marking the model’s water detection and blue representing the ground truth.

a comprehensive map. Additionally, for tasks related to traversability analysis, the full 3D information may be excessive. Traversability analysis typically requires 2.5D information, where intensity information in a 2D map suffices to categorize obstacles—distinguishing whether an obstacle is floating or sinking. To address this, there is a need to compress the 3D information in each frame systematically, enabling the creation of a lightweight traversability map.

To tackle the storage and computational challenges associated with processing 3D LiDAR data, various techniques are employed to compress the massive amount of information while preserving the accuracy of traversability analysis. These techniques aim to distill the essential information needed for navigation and obstacle avoidance. Common approaches include voxelization, where the 3D point cloud is converted into a regular 3D grid of voxels, and octree-based methods, which hierarchically organize the point cloud into a tree structure to reduce redundancy and storage requirements. Additionally, techniques like image-to-pointcloud mapping are utilized to identify key elements within the point cloud, facilitating efficient representation without compromising critical information.



**Figure 6.10:** Effect of Surface water environment on LiDAR pointcloud. (a,b) LiDAR scan shows negligible amount of reflections from the surface of water. (c,d) LiDAR scan shows many points against the leaves floating over the surface of water. *The bounding boxes and lines are only for visualization.*

It is essential to highlight that a significant impedance mismatch between water and air impedes the transmission of LiDAR rays from water to air. Consequently, when LiDAR rays enter the water, they are unable to exit, resulting in the complete absence of water within the LiDAR point cloud. This phenomenon is accurately illustrated in Figure 6.10, depicting point clouds acquired in two distinct scenarios. This characteristic proves advantageous for the architecture, as the absence of reflections from water implies the

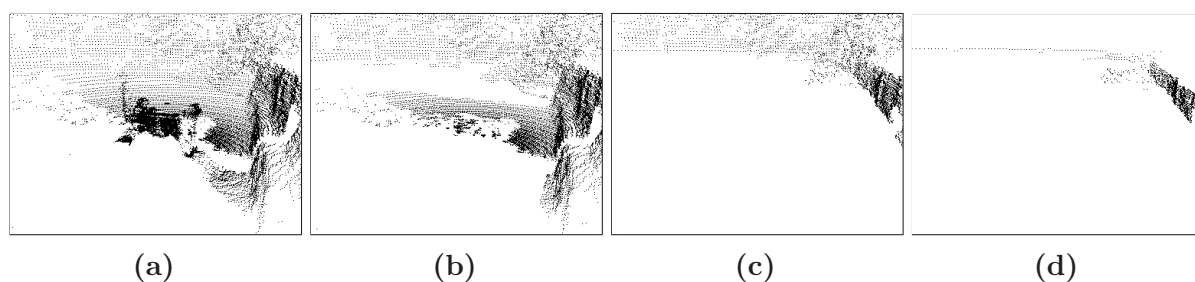
exclusion of floating objects from the LiDAR data, thereby enhancing the precision of the perception system.

### Multi-step Filtration

The 3D LiDAR sensor plays a crucial role in providing a comprehensive dataset of the surrounding environment, capturing a multitude of points as depicted in Figure 6.11a. However, not all points contribute to the generation of a useful traversability map. These points may include the points generated by colliding the surface vehicle's (SV) body, overhanging non-threat points, and random noisy points in LiDAR scan. A multi-step (three steps more precisely) filtration technique is implemented to clear out those redundant or unwanted points from pointcloud.

Initially, those points originating from the surface vehicle (SV) itself, are considered redundant and are subsequently filtered out from the LiDAR point cloud to enhance map accuracy. This filtering process ensures that SV's points do not interfere with the mapping process. In addition to SV's points, the LiDAR data may include points corresponding to overhanging tree branches. To address this, a predefined range is established, set at 1.5 times the height of the boat. Points falling within this overhanging range are deemed non-threatening and are retained in the dataset. This range is dynamically calculated based on the specific configuration of the vehicle, ensuring adaptability to different scenarios.

Furthermore, the LiDAR data may contain random points representing noise or extremely small (with respect to surface area) obstacle over the surface of water, each possessing a threat level. To simplify the representation of obstacles in the traversability map, a decision is made to treat obstacles less than 20cm in surface area as not a threat. To perform this filtration nearest neighbor algorithm is implemented to filter out points exceeding the threshold. This surface area threshold is configurable, allowing for adjustments based on the characteristics of the environment and the vehicle. Figure 6.11 provides a visual representation of the each step considered within the presented filtration technique, showing the dynamic nature of the filtering process.



**Figure 6.11:** Results of Three-step Pointcloud Filtering. a) Raw LiDAR scan b) Box filter for filtering SV points. c) Filtration of LiDAR points exceeding SV kinematics d) Points filtration using Algorithm 6.1.

Table 6.9 provides a quantitative breakdown of the reduction in the number of LiDAR points at three distinct scenarios, corresponding to the techniques discussed above shown in Figure 6.11. The table clarifies the progressive decrease in the number of points at each stage of the filtering process, offering insights into the efficiency and adaptability of the LiDAR data preprocessing methodology. The ability to configure and fine-tune these

filtering parameters ensures that the traversability map accurately reflects the relevant features of the environment while mitigating the influence of extraneous or non-threatening elements.

**Table 6.9:** Reduction in Point Cloud Data Across Processing Stages. The table shows the progressive reduction of data points in point clouds for three scenarios through three stages of processing, illustrating the filtering and refining effects of these stages.

| Scenarios  | <i>Stage 1</i> | <i>Stage 2</i> | <i>Stage 3</i> |
|------------|----------------|----------------|----------------|
| Scenario 1 | 120k+          | 105k           | 43k            |
| Scenario 2 | 111k+          | 98k            | 38k            |
| Scenario 3 | 50k+           | 46k            | 28k            |

---

**Algorithm 6.1:** Point Reduction Process for Generating an Obstacle Map. This algorithm outlines a method for filtering a point cloud scan by reducing points based on their probability values to create a more efficient obstacle map.

---

```

1 Input:  $S_i \Rightarrow$  Incoming scan,  $P_{ptx} \Rightarrow$  Probability Vector of points in scan.
   Require:  $S_f \Rightarrow$  Filtered scan with lesser points.
   while  $k \leftarrow S_i$  do
     end
      $P_p \leftarrow P_k$ 
    $k \leftarrow p$  else
     end
     Delete  $p$  in  $S_i$ 

```

---

### Image to point cloud mapping

Object detection within point clouds presents a unique set of challenges compared to image-based detection due to the absence of color and texture information. Consequently, the object detection carried out in the preceding sections on images is deemed more robust and accurate. To leverage the detailed detection information obtained from images, it is essential to map this information onto the point cloud using a technique known as inverse-perspective mapping.

The process begins by preserving the objects detected in the image and subsequently translating this information onto the point cloud. The remaining points in the point cloud are then filtered out, resulting in an obstacle-specific point cloud. The accuracy of this obstacle point cloud is intricately tied to the precision of the calibration between the camera and LiDAR systems. The calibration process necessitates the existence of the same environment in both modalities, enabling the identification of corresponding features.

In our prior publication [Keen 21], a similar methodology was introduced, emphasizing the importance of robust calibration for accurate mapping between images and point clouds. The calibration process is visualized in Figure 6.12, highlighting the mapping of the point cloud onto the images. To achieve accurate calibration, a checkerboard pattern



**Figure 6.12:** Point cloud Mapping over Camera Imagery. The figure illustrates the overlay of a LiDAR point cloud on an image from the TUK campus dataset, with red points indicating the precise mapping of the point cloud data onto the corresponding features within the image.

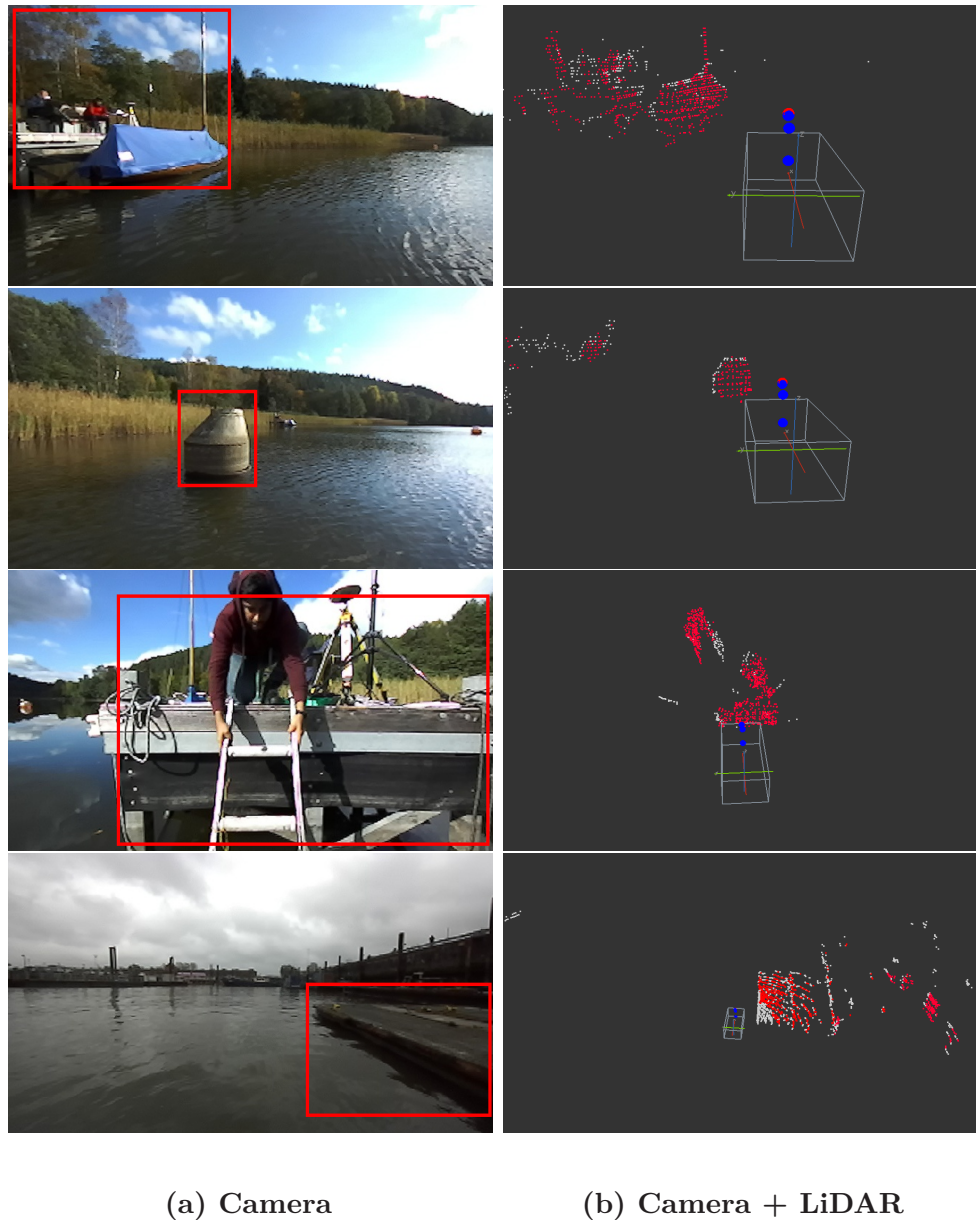
is employed, serving as a reference for identifying common features between the camera and LiDAR modalities.

The effectiveness of inverse-perspective mapping in translating object detection information from images to point clouds lies in its ability to maintain the spatial relationships and positions of objects within the environment. By aligning the detection results from images with the corresponding points in the LiDAR-generated point cloud, a unified and enriched representation of the environment is created. This integration of information is vital for downstream tasks such as navigation, scene understanding, and obstacle avoidance. Figure 6.13 shows the clear object after fusing information received from object detection and semantic segmentation step discussed in sections 6.2.1 and 6.2.1.

### Obstacle Motion Map

The motion mapping module within the context of underwater obstacle detection serves as a pivotal element for assessing potential threats posed by objects in a marine environment. This module extrapolates the dynamic characteristics of obstacles, providing an initial estimation of movement based on physical dimensions such as surface area and volume. Through empirical analysis, it has been inferred that larger obstacles tend to exhibit less mobility compared to their smaller counterparts. Such observations stem from the natural buoyancy and drag forces acting on objects immersed in a fluid medium. The module operates by meticulously analyzing the volumetric presence of an object on the water's surface, utilizing point cloud data to ascertain its three-dimensional structure. The point cloud, a set of data points in space, is typically generated by sonar or LiDAR systems that capture the nuances of the underwater landscape. By processing this data, the module can derive a probabilistic model that predicts the likelihood of an object's movement, taking into account factors such as current, wind, and the object's inherent buoyancy.

Subsequent chapters will delve into the integration of this motion mapping with threat level assessment algorithms. The threat level of an obstacle is not solely determined by its potential for movement but also by its underwater footprint — the area it encompasses on the seabed. By correlating motion map values with the spatial extent of the object's contact with the seabed, a more nuanced understanding of the obstacle's behavior and potential risk can be achieved. This multidimensional analysis is crucial for autonomous underwater vehicles (AUVs) and remotely operated vehicles (ROVs) that require real-time data to navigate safely and effectively in complex marine environments. This threat



(a) Camera

(b) Camera + LiDAR

**Figure 6.13:** Combining the Visual and Spatial Sensory Modalities. Figure (a) presents camera images with detected obstacles highlighted within red bounding boxes, while figure (b) displays the fusion of LiDAR point clouds with the detected obstacles from the camera imagery. The white points in the point cloud represent the raw spatial data, and the red points specifically denote the identified obstacles. *The boxes in the right images show the position and orientation of boat.*

assessment is not a static computation but a dynamic one that continuously evolves with the changing conditions and movements of the obstacles. It involves a complex algorithm that takes into account not only the physical parameters but also the historical data and predictive models of object behavior. For example, a stationary object that has a large underwater footprint might be considered a high-threat level due to its potential to hinder navigation routes or disrupt marine ecosystems. Moreover, the threat assessment is contextualized within the broader scope of marine operations and objectives. For certain



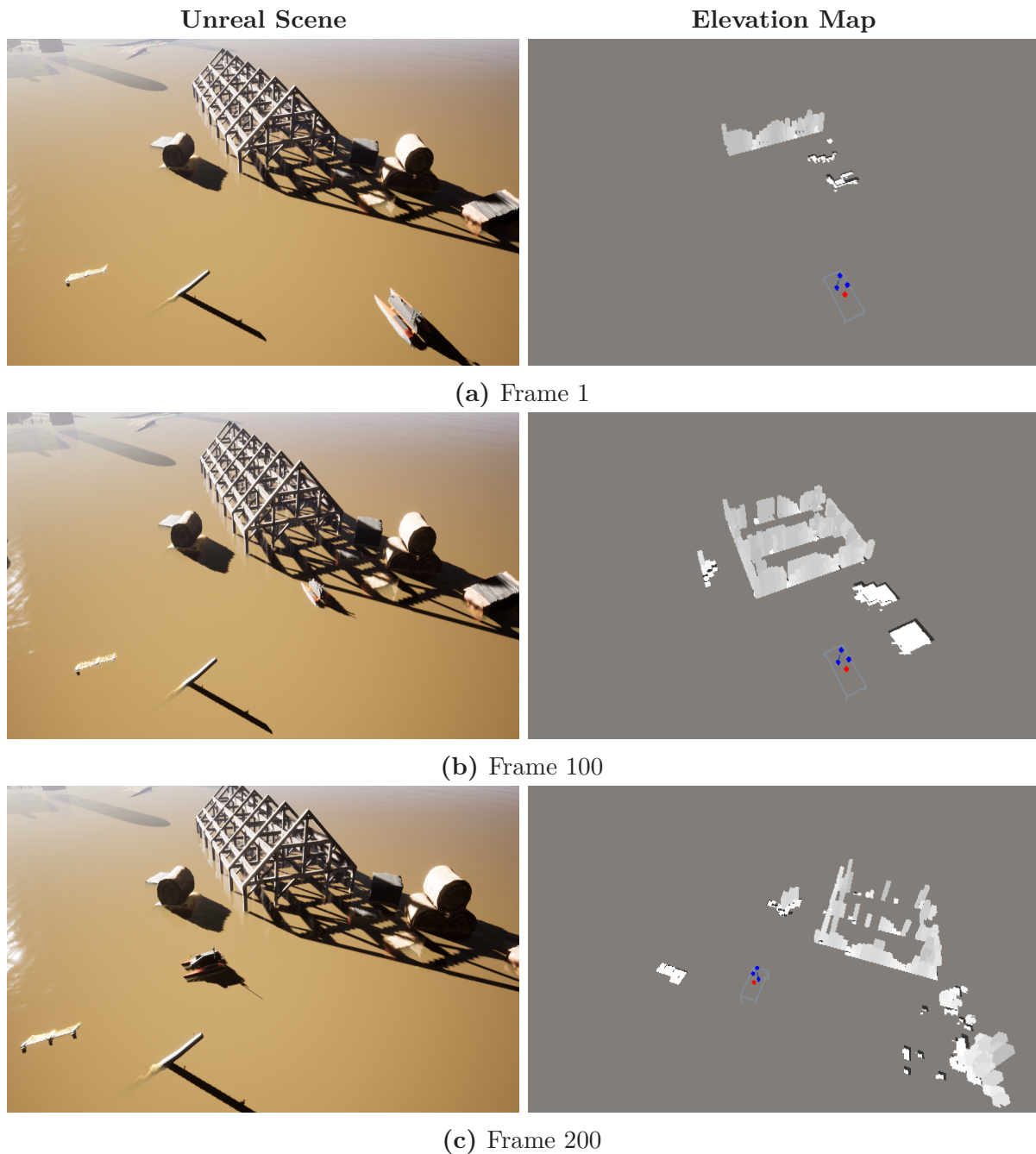
applications, such as underwater construction or salvage operations, a large immobile object may present a significant operational challenge, whereas for ecological surveys, the same object might be of high interest but low threat.

### 6.2.3 Obstacle Map

For traversability mapping in surface vehicles, a 2.5-dimensional occupancy grid map stands as a fundamental objective. This map helps identify and categorize obstacles that could pose a risk to safe navigation. The map's occupancy values indicate the level of danger presented by detected objects, making it a critical metric for assessing navigational risks. However, this mapping technique mainly focuses on surface-level obstacles and doesn't account for objects submerged below the water's surface, which could also pose significant navigational issues. Assessing the surface water challenges mainly relies on the vertical elevation of objects, meaning that taller structures present a more significant hazard. Furthermore, To create a global map from individual LiDAR (Light Detection and Ranging) scans, the system incorporates GPS/GNSS (Global Positioning System/Global Navigation Satellite System) data. This integration is essential for accurately aligning scans to form a universally referenced map. The onboard LiDAR system operates at dual frequencies of 10Hz and 20Hz, with the 10Hz frequency chosen for system evaluation purposes. This frequency selection, despite resulting in a large volume of data, is crucial for maintaining the resolution and effectiveness of the mapping process.

The challenge then becomes managing this extensive information without overwhelming storage capacities. A novel solution, as detailed in [Keen 20], involves updating the map only when the vehicle moves beyond the reach of the previously acquired scan. This method ensures that the map remains updated without becoming extremely dense. To ensure each new scan is accurately positioned relative to the existing map data, the Iterative Closest Point algorithm is utilized. This algorithm fine-tunes the alignment of each scan, ensuring that the global map remains a precise and reliable tool for navigation, providing a lightweight yet detailed representation of the surrounding environment, optimized for real-time application in surface vehicle navigation.

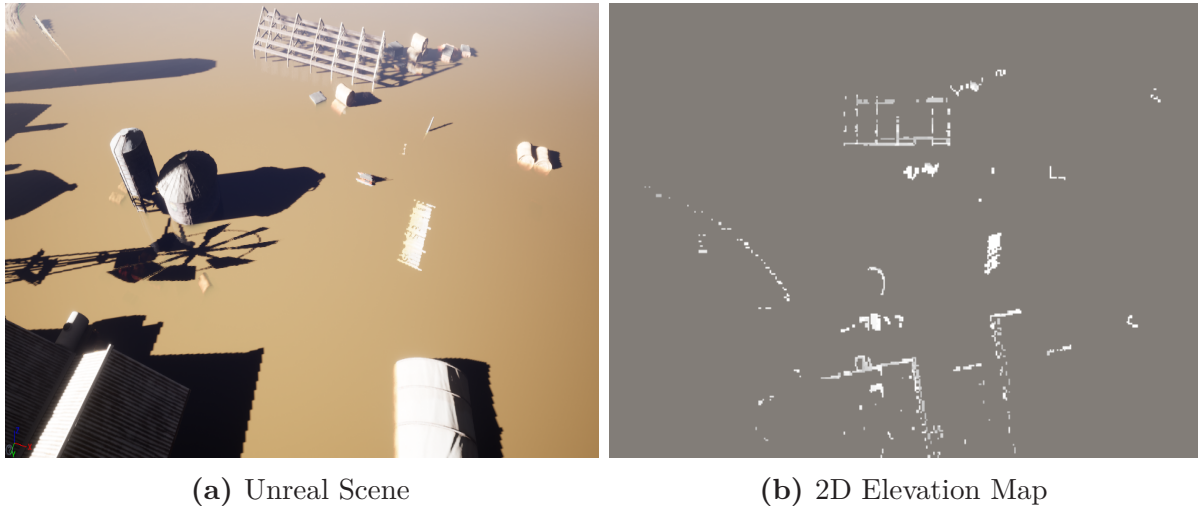
The Figure 6.14 illustrates a sophisticated simulation framework designed to enhance the capabilities of autonomous surface vehicles in terms of traversability and obstacle detection. It captures a series of frames from an Unreal Engine simulation, where the vehicle, equipped with both camera and LiDAR sensors, navigates a complex barnyard environment. The left side of each subfigure displays the simulated scene, depicting the vehicle's interaction with various static and dynamic obstacles within the virtual setting. On the right side, the elevation map can be seen, which is computed using data fusion of camera imagery and LiDAR point cloud. This elevation map is a key component of the surface traversability module, offering a visual representation of obstacles on the water's surface. Notably, the map's accuracy is within  $\pm 19$  cm, a proof to the precision of the simulation in mimicking real-world conditions. As the boat progresses through the environment, the frames at intervals of 100 demonstrate the iterative refinement of the map, showcasing the evolving understanding of the vehicle as it gathers more data. This level of detailed mapping is crucial for autonomous vehicles tasked with safely navigating aquatic environments, as it allows for the anticipation of potential hazards and the planning of efficient routes. The simulation's ability to provide ground truth eases



**Figure 6.14:** Simulation Result - Elevation Map at each frame. The image demonstrates the generation of map after consecutive frames by driving a surface vehicle within the simulated environment.

the validation process of the mapping accuracy, ensuring that the algorithms performing these operations are reliable and effective. The final map after traversing the complete simulated environment is shown in Figure 6.15.

Moreover, Figure 6.16 demonstrate the methodology tested over the natural water bodies. The images on the left depict views of various natural water bodies as captured by satellite, providing a top-down perspective similar to what one might find on a platform such

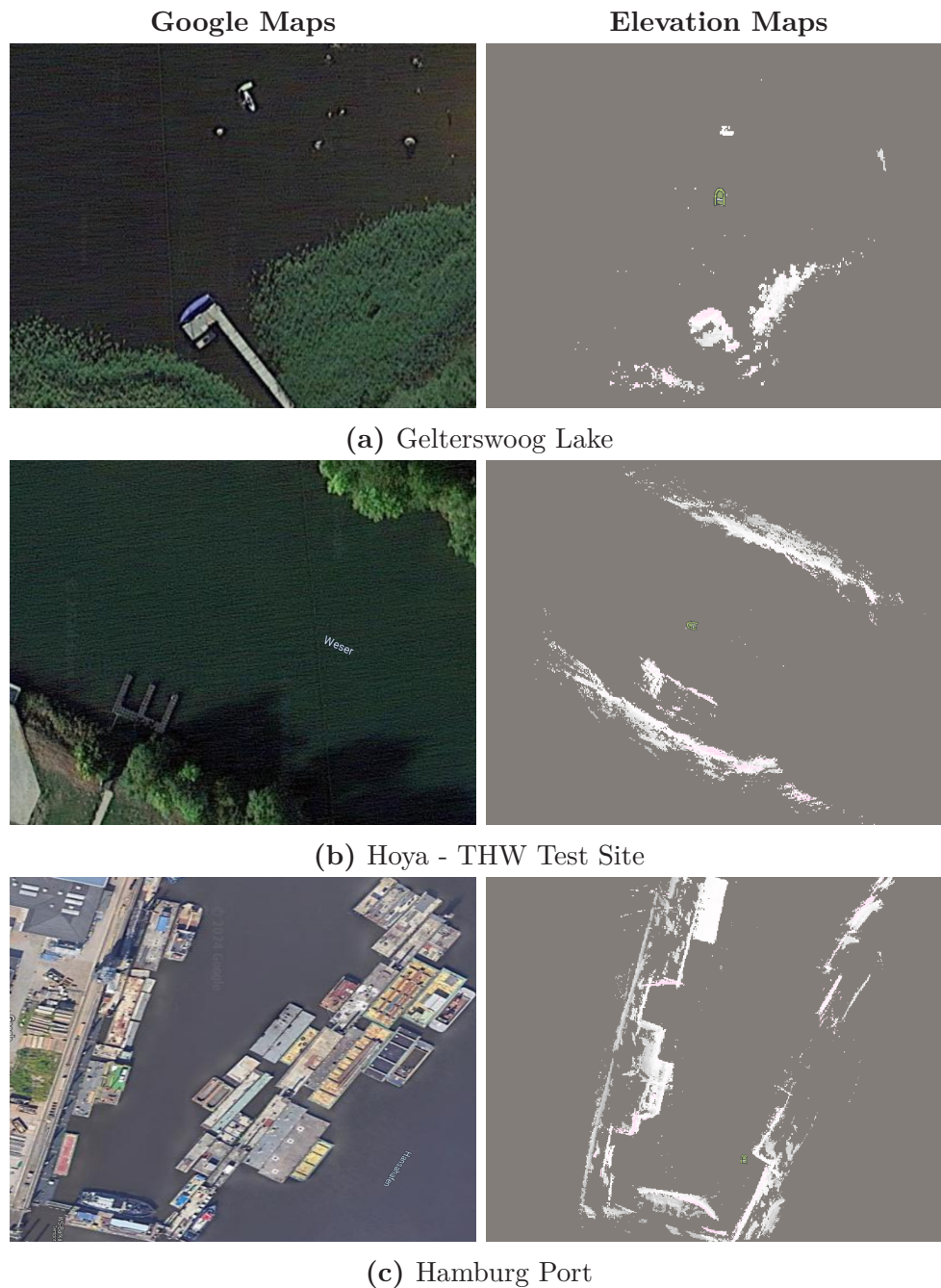


**Figure 6.15:** Simulation Result - Complete Elevation Map. Figure (a) shows a simulated barnyard scene with obstacles in the Unreal Engine, while Figure (b) illustrates the corresponding 2D elevation map generated by driving a simulated surface vehicle with multi-sensor system within the simulation environment. The map outlines the environment and height of those obstacles.

as Google Maps. These scenes offer a point of reference, showcasing the environment in its natural state, complete with surface features and surrounding landscapes. The corresponding images on the right, however, represent the elevation maps that reveal the potential navigational hazards above the water's surface.

In the simulated environment, the methodology deployed had previously demonstrated an accuracy of  $\pm 19$  cm, a metric that was thoroughly verified via rigorous testing. Drawing on these validated simulation results, a parallel can be drawn that the deviations in the natural water body maps would align closely with those observed in the controlled simulation conditions. While the true accuracy in natural settings cannot be explicitly quantified due to the absence of a definitive ground truth, the consistency and reliability of the sensor technologies employed provide a strong inference that the generated maps are indeed precise. In this regard, one of the prominent sensor would be RTK GPS system that provides an accuracy of approximately  $\pm 2$  cm. By leveraging such precise positional data, the resultant maps reflect a high degree of accuracy, reliably mirroring the actual layout and distribution of obstacles within the water body.

Moreover, the generated obstacle maps, when seen with their Google Maps counterparts, exhibit a striking similarity in the layout and presence of features, further underscoring the accuracy of the mapping process. This visual comparison serves as qualitative validation of the maps' precision, complementing the quantitative data provided by the GPS sensors. The implications of such precise mapping are manifold, extending beyond navigation to encompass areas like ecological conservation, disaster management, and the development of autonomous watercraft, where the understanding of surface and subsurface conditions is paramount.



**Figure 6.16:** Comparative presentation of natural water bodies with their respective Google Maps and Elevation Maps. The figure demonstrates satellite imagery from Google Maps with corresponding elevation maps generated by driving surface vehicle over three different water bodies, illustrating the results of obstacle mapping in Gelterswoog Lake, the THW test site in Hoya, and the Hamburg Port.

### 6.3 Discussion

This chapter provides a comprehensive comparison of various sensory modalities used in understanding a surface water environment, including cameras, LiDAR, and radars. Following this examination, a novel framework for Surface Water Traversability Mapping

(SWTM) is presented, which utilizes both camera and LiDAR sensors to map the surface water scenario.

In an aquatic environment, the camera images suffer from several challenges, including high reflections of sunlight from water surface and the low light conditions in cloudy environment that effect the overall exposure of camera imagery. Therefore, the camera images are enhanced by normalizing their exposure level using histograms. This exposure correction technique has contributed in the increase in robustness of obstacle detection within SWTM framework.

Additionally, several deep learning-based obstacle detection and semantic segmentation models are implemented in this chapter that has provided a detailed comparison of each model over surface water environment. Due to the limited relevant dataset, the obstacle detection suffer from false-positive detections that affect the precision of final traversability map. Therefore, the obstacles are further filtered based on their presence in critical region of the surface vehicle using the semantic information.

The 3D aspect of detected obstacle is computed by mapping it over LiDAR-generated point cloud, which provides detailed range information for each identified object. This fusion of data from cameras and LiDAR significantly elevates the precision of both detection and mapping within the SWTM framework. The obstacles identified through this process are subsequently represented on a 2D grid map, offering a comprehensive overview of the surface water environment. The accuracy of this map (within  $\pm 19$  cm variance) is dependent upon the precision of the GPS and the accuracy of object detection derived from camera imagery. This generated map lay foundation for the development of a fused traversability map in subsequent chapter, demonstrating the critical role of accurate surface water mapping in enhancing our understanding and navigation of complex environments.



## 7. Fused Traversability

Many marine applications, such as dock surveillance, marine resource exploration, and underwater terrain mapping, concentrate on perceiving either the surface or the underwater domain based on their specific requirements. This singular focus is generally enough for these domain-specific applications. For instance, dock surveillance primarily involves monitoring the water's surface, while underwater terrain mapping involves studying the ocean floor's topography. However, to map post-flood environments, it is crucial to consider both the surface and underwater domains, as floods can displace debris and unsafe obstacles, posing navigation risks. These obstacles can be found on the surface, partially submerged, or underwater, requiring a clear view of both domains for safe navigation.

Navigating the post-flood environment presents a formidable challenge due to the presence of many random and unstructured objects ranging from damaged infrastructure like broken bridges to natural debris such as floating leaves, submerged trees, or even vehicles, along with various forms of their underwater footprints. The inherent unpredictability and variability of these objects create a complex task that is extremely difficult to map accurately. Detecting and classifying these objects is a complex task because each one is unique and has its own characteristics like size, shape, material, position, and whether it is partially or fully submerged or floating on the surface.

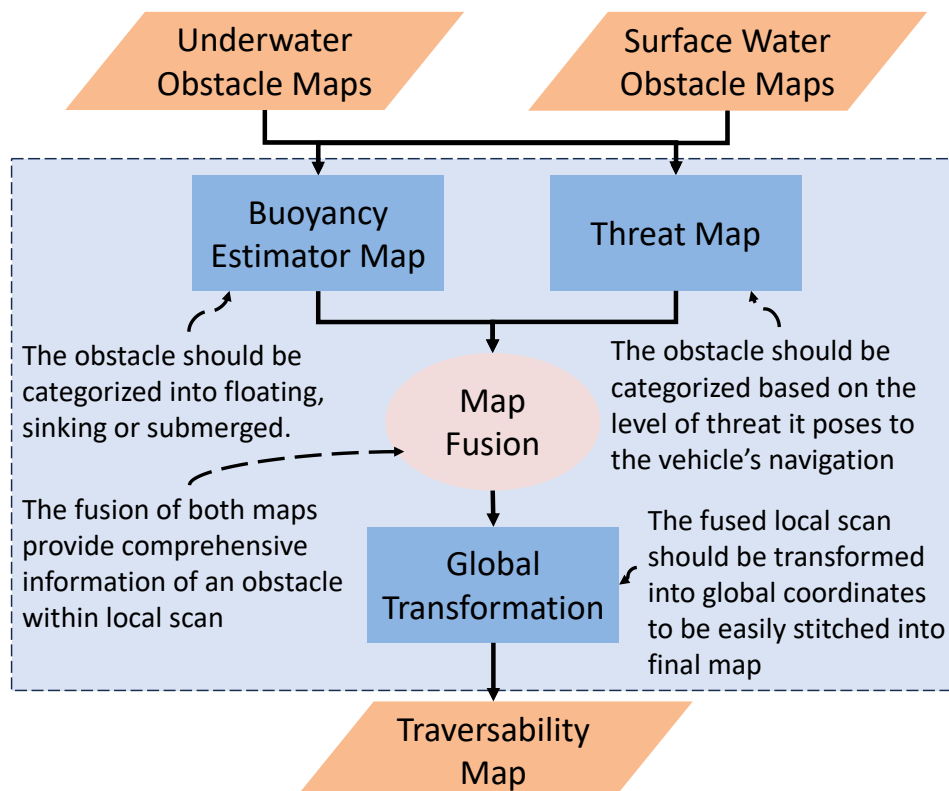
Furthermore, integrating data from various domains presents a challenge due to the differences in sensory modalities. For example, LiDAR provides detailed and accurate maps for surface water environments using three-dimensional range data. In contrast, sonar captures 2.5-dimensional sound reflection information to understand underwater terrains and identify submerged objects. However, the nature of the data obtained from these sensors is fundamentally different, making their integration a complex process. The varying types and formats of the data collected by these sensors require a sophisticated approach to data fusion.

To address these challenges and create an effective mapping system for post-flood environments, it is essential to transform the data from both domains into a comparable format. This transformation involves not only the alignment and integration of the data but also its standardization to a uniform scale and reference. Once this is achieved, the next crucial step is the categorization of detected obstacles into a limited number of basic classes. This

classification simplifies the complex array of objects into manageable groups, facilitating easier identification and analysis. By categorizing obstacles into fundamental classes, such as floating debris, submerged objects, or structural remnants, the mapping system can more efficiently process and interpret the data, leading to more accurate and actionable insights. This approach not only enhances the capability of the mapping system to deal with the diverse challenges of a post-flood environment but also significantly improves the reliability and utility of the data for rescue and recovery operations.

## 7.1 Fused Traversability Mapping Framework

The integration of surface and underwater obstacle maps is a crucial step in gaining a complete understanding of complex environments created by floods. This process is necessary to provide essential insights and characteristics of the post-flood situation, which is vital for creating maps that enable rescue boats to navigate safely. The Fused Traversability Mapping (FTM) framework is a critical component in integrating surface and underwater maps to produce a comprehensive and informative map that accurately represents the traversability of the post-flood scenario. To fully comprehend the workings of this module, it is necessary to discuss the specific functionalities and methodologies employed by both the surface and underwater traversability modules.



**Figure 7.1:** Schematic of the Fused Traversability Mapping (FTM) Framework. This figure illustrates the process of generating a Traversability Map by combining underwater and surface water obstacle maps with buoyancy data and threat levels, merging them, and then transforming the fused data into global coordinates for comprehensive navigation guidance.



The *Underwater Traversability Mapping (UTM)* framework, as elaborated in Chapter 5, employs a multifaceted approach to address the challenges posed by sonar imagery, which typically suffers from low signal-to-noise ratio (SNR). The UTM incorporates image enhancement and depth estimation techniques to refine this imagery, making it more conducive for analysis. Additionally, object detection techniques are utilized to identify underwater obstacles and integrate them into a localized map. A significant challenge with forward-looking sonar imagery is its 2.5-dimensional nature, which does not provide depth information on the waterbed. To overcome this, the UTM implements a depth estimation technique that computes the depth of the waterbed by detecting the intensity variation in pixel values. In its final stage, the UTM generates a local obstacle map that provides 2.5-dimensional information about the underwater footprint of objects and a depth map, providing a detailed view of the submerged environment.

In parallel, the *Surface Water Traversability Mapping (SWTM)* framework, presented in Chapter 6, focuses on detecting surface water obstacles. This is achieved through advanced deep learning algorithms specializing in semantic segmentation and object detection. Once these obstacles are detected, they are transformed into the LiDAR frame of reference, allowing for the creation of a three-dimensional point cloud representation of each obstacle. This transformation is crucial as it aligns the surface-based obstacle data with the LiDAR's spatial framework, thereby enabling a more accurate and comprehensive mapping of the surface environment. The SWTM's sophisticated use of LiDAR sensor not only enhances the detection process but also ensures that the surface map is computed within the LiDAR's frame of reference, thereby maintaining consistency and accuracy in the data.

The fusion of the data from these two modules – UTM and SWTM – by the FTM is a complex process that requires not only the integration of the data but also its alignment. This alignment ensures that the diverse data types, formats, and frames of reference from the underwater and surface modules are seamlessly added up to produce a unified, multi-class traversability map of the post-flood environment. Such a map is an invaluable asset in post-flood scenarios, providing rescue teams with crucial information that enhances their navigational capabilities and ensures safer operations. The detailed and dynamic nature of this map, encompassing both surface and underwater obstacles, reflects the multifaceted challenges of post-flood environments and represents a significant advancement in the field of environmental mapping and disaster response.

### **Categorizing Post-Flood Environment**

The diverse nature of obstacles, each varying in nature, size, and location, coupled with the unpredictability of the scenarios encountered, makes the task of detecting and classifying every obstacle a formidable challenge. Consequently, a novel approach to managing this complexity is the categorization of surface and underwater obstacles into simpler, more manageable groups. This categorization not only simplifies the detection and analysis process but also ensures a more efficient and targeted response in post-flood situations.

**Obstacle Buoyancy** - A practical method for categorizing obstacles in a post-flood environment is to classify them based on their interaction with water: floating, submerged, or completely sunken objects. Each category represents a distinct type of hazard and requires a different approach for detection and navigation. For example, underwater debris, such as sunken vehicles or construction materials, may not be

visible on the surface but can create a highly unsafe environment for navigation. This is particularly true in post-flood scenarios where the water is often shallow, and such submerged objects can pose significant risks to boats and other watercraft. In contrast, floating obstacles like leaves, branches, or even displaced household items present a different type of challenge. While these objects do not have an underwater footprint, they can still impede navigation and require detection and avoidance strategies that focus on the water's surface.

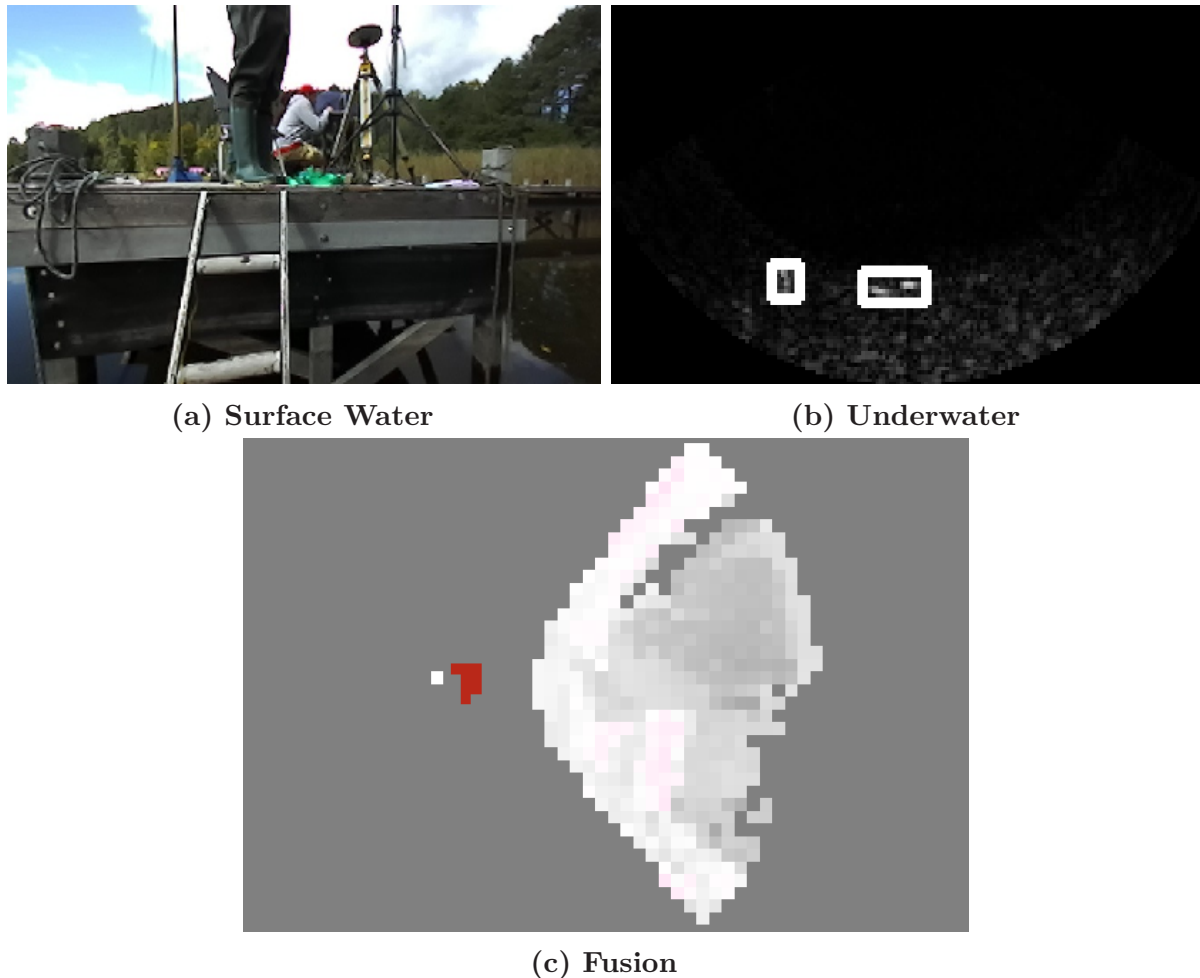
**Obstacle Threat Level** - In a post-flood scenario, not all obstacles present the same level of risk or impediment to navigation. This variance in the threat posed by different obstacles necessitates a more subtle approach to categorization, one that goes beyond mere detection and extends into assessing the degree of navigational risk each obstacle poses. Thus, it becomes essential to further categorize obstacles based on their threat level, creating a spectrum that ranges from easily navigable to completely impassable. For instance, when considering surface water obstacles, the nature and position of each object can significantly influence its impact on navigability. A standing tree or a floating boat represents a substantial obstacle, creating physical barriers that are not traversable and hence pose a high risk to navigation. In contrast, smaller obstacles like floating leaves, although present on the water's surface, generally do not pose a significant threat to navigation. These can often be traversed with minimal risk, allowing rescue boats to maintain their route with little to no deviation.

To understand the environment with respect to categories presented above, FTM presents two important modules i.e. Obstacle Buoyancy Estimator and Fused Local Threat Map.

### 7.1.1 Buoyancy Estimator

The complex task of categorizing obstacles in post-flood scenarios is a critical component of navigational safety and efficiency. This categorization, which hinges on the concept of buoyancy, is essential in differentiating between floating, submerged, and sunken obstacles. The distinction among these categories is primarily based on the presence or absence of surface and underwater footprints, characteristics that can be accurately identified through the data provided by underwater and surface water mapping modules. The integration of this data is a complex process, requiring a specialized Buoyancy Estimator module that plays a pivotal role in fusing the footprints from both surface and underwater environments to categorize the obstacles accurately.

The categorization process begins with an analysis of the footprints. Sunken obstacles are identified by their exclusive presence in the underwater domain without any corresponding surface footprint. These typically include objects that have completely settled at the bottom of the water body and are fully submerged. In contrast, floating objects are those with visible footprints on the surface water but no underwater presence. These objects, often lighter and less dense, remain buoyant and pose a different set of challenges for navigation. The third category encompasses submerged obstacles, which are characterized by their presence in both domains. These objects, partially submerged, are particularly challenging as they possess properties of both floating and sunken objects, requiring careful analysis for accurate categorization. Figure 7.2, it is illustrated that how surface water

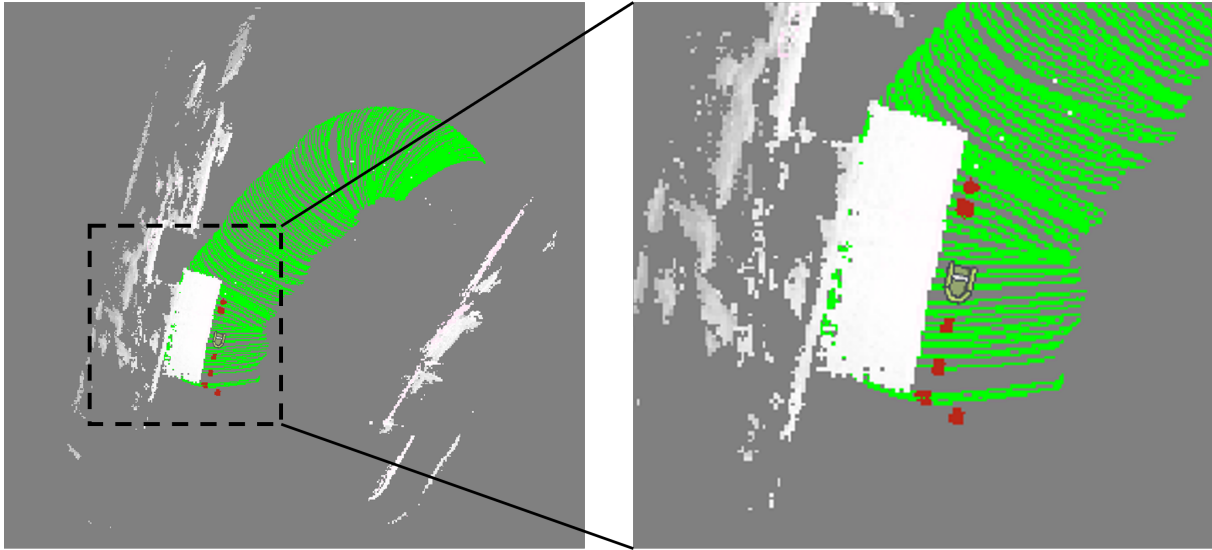


**Figure 7.2:** Surface and Underwater Detection in Gelterswoog Lake. This image presents the detection of (a) ladder in sonar due to its high reflective properties. (b) The docking area is not visible to limited sonar field of view. (c) The red shows its submerged area whereas gradient from white to gray shows the surface water elevation of docking area.

and underwater detections are integrated. The camera imagery captures a docking point complete with a steel ladder, while the sonar, leveraging its sensitivity to the reflective properties of objects, detects the same ladder. This dual detection allows for the ladder to be categorized as a submerged object, indicated by red points in Figure 7.2c. Conversely, portions of the dock not detected underwater are classified as floating, with a gradient from white to gray indicating their elevation above the water's surface.

### 7.1.2 Fused Local Threat Map

While categorizing obstacles based on buoyancy – floating, submerged, or sunken – is a fundamental step in understanding the post-flood environment, this classification alone is not sufficient to ascertain the navigational safety of these obstacles. It is crucial to recognize that not all obstacles detected in the maps pose a significant threat to navigation. For example, leaves floating on the surface of the water, though classified as 'Floating' by the buoyancy estimator, are typically safe to traverse and do not constitute significant navigational hazards. Similarly, sunken obstacles might be deemed safe if they are located



**Figure 7.3:** Fused Threat Map of Hamburg Port site. This color-coded traversability map of the Hamburg Port site illustrates the sonar field of view in green, submerged obstacles in red, and the elevation of surface water obstacles transitioning from white to gray, offering a detailed representation of both underwater and surface features.

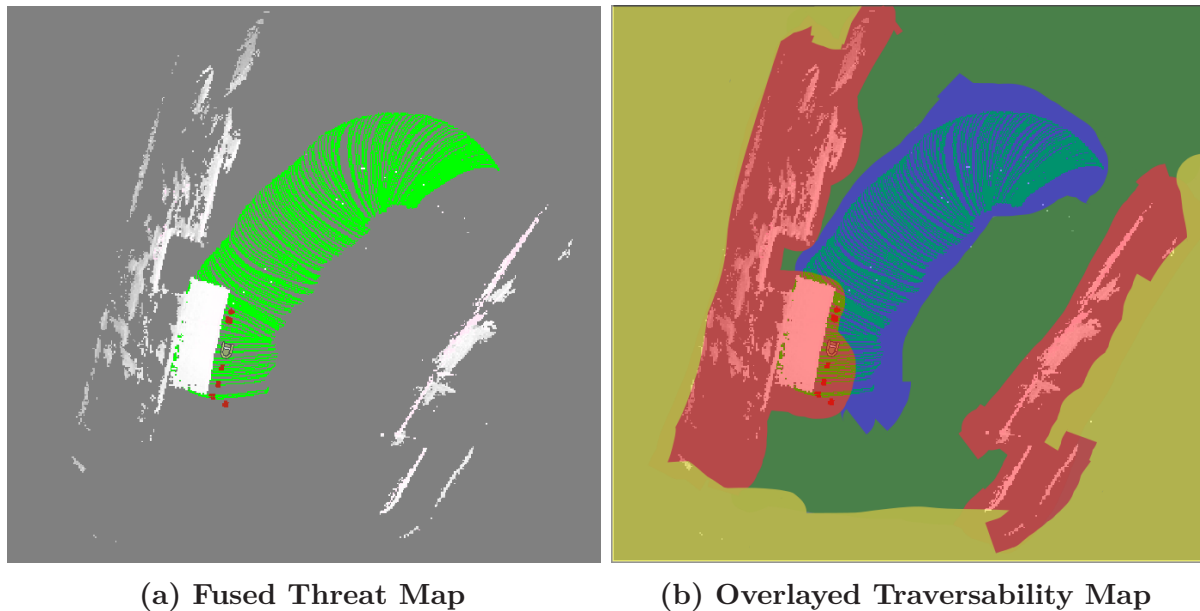
in regions that do not impede safe traversability. Therefore, a more simpler approach is required, one that assesses the threat level of each obstacle based on additional criteria, such as its volume and protrusion above and below the water's surface.

This approach involves computing a threat level for each obstacle, taking into account its physical dimensions and location relative to the water's surface. The concept is straightforward: the greater the height or volume of a floating obstacle above the water, the higher its potential threat level. This is because larger obstacles protruding above the water's surface can pose significant barriers to navigation, potentially causing damage to vessels or impeding their passage. Conversely, for sunken obstacles, the focus shifts to their volume under the water. Large sunken obstacles that occupy significant space beneath the surface can be hazardous, particularly in shallower waters where they may not be immediately visible but still pose a risk of collision. Figure 7.3, displays a fused threat map of the Hamburg Port site. This map highlights the underwater footprint of the docking point, providing vital information on submerged obstacles that could hinder navigation. The areas beyond the sonar's field of view, shown in green, remain uncaptured, emphasizing the limitations and challenges inherent in creating fully comprehensive maps. Nonetheless, the ability of these maps to fuse data from different sources into a coherent visual representation of navigability emphasizes their value.

### 7.1.3 Traversability Map

Traversability maps have become an essential tool for environmental analysis, particularly in assessing areas affected by natural disasters such as floods. These maps are designed to depict not only the physical landscape but also the navigability of an area, taking into account various obstacles and terrain features that might impede movement. The basis of traversability maps can be traced back to the fields of robotics and autonomous vehicle navigation, where understanding the environment is crucial for safe and efficient operation.

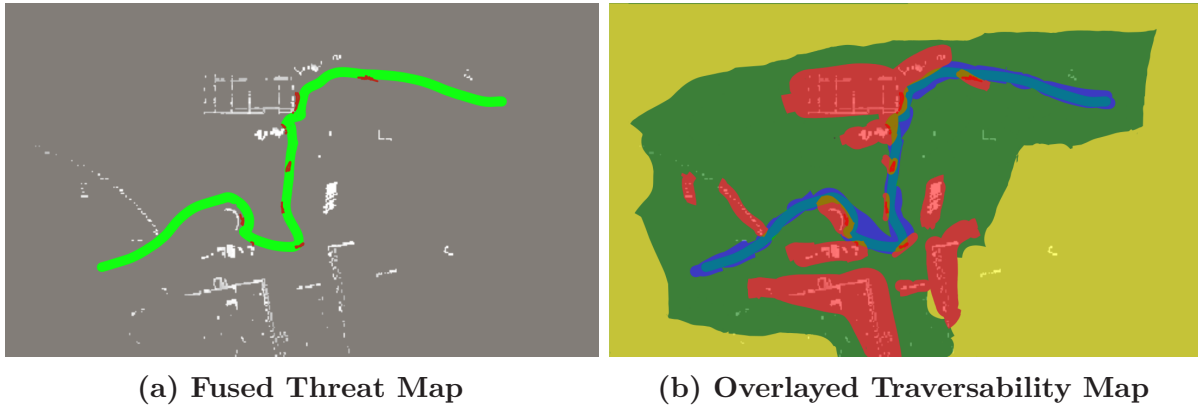
In these contexts, traversability maps are used to represent the environment in a way that machines can interpret, indicating areas that are safe to traverse versus those that are not. This concept has been adapted and expanded for disaster response and environmental management, where the ability to quickly assess and navigate through affected areas can significantly impact rescue and recovery efforts.



**Figure 7.4:** Traversability map of Hamburg Port. This image displays the traversability map, where the *blue* area indicates the traversable region with respect to both the underwater and surface water maps. The *red* area depicts the non-traversable region, while the *green* area shows the surface traversable region that is unknown underwater. The *yellow* area represents the completely unknown region.

In the final experiments of this thesis, the utility of updated traversability maps is demonstrated that uses fused threat map to provide a comprehensive environmental analysis. These maps are generated through the fusion of data from multiple sensors, including cameras and sonar, which capture distinct aspects of the environment. Figure 7.4, shows a fused threat map of Hamburg Port and a generated traversability map that indicates the regions where navigation is likely to be obstructed. The map in Figure 7.4b shows a traversability map overlaid upon the fused threat map, where different colors are used to communicate various levels of navigability. The *blue* areas signify regions that are traversable, meaning they are free of obstacles on the surface and underwater, making them safe for navigation. *red* areas denote non-traversable regions where obstacles are present, posing a danger to vehicles or rescue operations. *green* areas represent regions where the surface is traversable but the status underwater is unknown; hence, there might be submerged obstacles that are not accounted for. Lastly, *yellow* areas indicate completely unknown regions, where there is no information available regarding potential obstacles.

Similarly, in the Figure 7.5, a similar pair of maps are shown for a simulated barnyard environment. The traversability map (shown in Figure 7.5b) uses the same color coding as the previous image. It provides a visual assessment of traversability, with *blue* indicating clear paths for navigation, *red* showing obstructed areas, *green* suggesting potentially clear



**Figure 7.5:** Traversability map of Simulated Barnyard Environment. This image displays the traversability map, where the *blue* area indicates the traversable region with respect to both the underwater and surface water maps. The *red* area depicts the non-traversable region, while the *green* area shows the surface traversable region that is unknown underwater. The *yellow* area represents the completely unknown region.

surface areas with unknown underwater conditions, and *yellow* highlighting areas where no data is available.

These maps are crucial for autonomous vehicle navigation and disaster response. In robotics and autonomous vehicle systems, traversability maps guide path planning and decision-making processes by finding safe paths and highlighting areas to avoid. When adapted for disaster response, such maps are essential for planning rescue and recovery operations, allowing responders to quickly identify safe routes, understand the extent of damage, and navigate through affected areas with higher efficiency and safety. The integration of underwater and surface data provides a more comprehensive environmental analysis, aiding in the deployment of resources and efforts during critical response activities.

## 7.2 Discussion

The chapter presents a novel approach to mapping post-flood environments, addressing the unique challenges posed by the integration of surface and underwater domains. It introduces the Fused Traversability Mapping (FTM) framework, which combines data from the Underwater Traversability Mapping (UTM) and the Surface Water Traversability Mapping (SWTM) frameworks to create comprehensive maps that enhance the safety and efficiency of rescue operations. By categorizing obstacles based on buoyancy and threat level, the framework simplifies the complex array of post-flood challenges, facilitating more targeted and efficient navigational strategies. The Buoyancy Estimator and the Fused Local Threat Map play crucial roles in this process, enabling the accurate assessment of obstacles' impact on navigability. This integrated approach significantly better from traditional mapping techniques, offering a robust tool for disaster response and environmental monitoring. The FTM framework not only improves understanding of post-flood environments but also contributes valuable insights into environmental conservation and disaster management, showcasing the potential of advanced mapping technologies in enhancing navigational safety and operational efficiency in disastrous scenarios.

## 8. Conclusion

The thesis presented here provides a comprehensive examination of the challenges involved in mapping environments that have been affected by floods. Traditional mapping tools, designed for stable and predictable surroundings, are not adequate for natural disasters like floods. After a flood, there are many unpredictable factors that make navigation and mapping difficult, such as the presence of unstructured objects like debris and destroyed structures that are scattered across the landscape. Additionally, the water in these post-flood scenarios is often opaque due to a high concentration of suspended particulates like sand and silt, which can obscure the submerged topography and make mapping underwater even more difficult. This opacity, coupled with the variable water depth and the presence of submerged obstacles, create significant navigational hazards. Therefore, it is essential to create a mapping system specifically designed to address these challenges. This system should not only navigate the complexities introduced by floods but also generate precise and comprehensive maps of the affected environment, enabling more effective response efforts in the aftermath of such events.

This thesis makes several contributions to the field of environment mapping for surface water vehicles, with the introduction of the novel Shallow Water Traversability Mapping (SWiM) architecture standing out as a significant achievement. The SWiM architecture has undergone rigorous testing across various surface water vehicles, each characterized by distinct kinematic and dynamic properties. This testing is carried out in a range of water bodies, including lakes, riversides, and ports. However, the challenge of obtaining sensor data from post-flood environments, which present complex and heavily destructed scenario, limits the direct evaluation of the architecture in such contexts. To address this, a highly realistic post-flood simulation is developed using Unreal Engine, featuring diverse environments from remote barnyards to urban cities. Additionally, a novel simulated sonar plugin, utilizing Generative Adversarial Networks (GANs), is created to generate artificial sonar imagery from the simulated environments.

A key element of the SWiM architecture is the Underwater Traversability Mapping (UTM) framework, designed to detect underwater obstacles potentially unsafe to vehicles. This framework introduces a novel three-step filtration technique for enhancing the quality of raw sonar imagery. It also incorporates an innovative depth estimation method that

estimates the water bed's depth by analyzing pixel-based intensity variations in sonar acoustic data between no-return and echo-return zones. As a result, the UTM framework provides comprehensive underwater obstacle and depth maps in the form of 2.5D occupancy grid maps.

For mapping the surface water environment, the SWiM architecture presents the Surface Water Traversability Mapping (SWTM) framework, which integrates camera and LiDAR data to identify surface water obstacles. This involves the application of various object detection and semantic segmentation algorithms to camera images, enabling the robust identification and mapping of obstacles onto a 3D point cloud. Similar to UTM framework, the SWTM framework also delivers an obstacle map for surface navigation.

Furthermore, this thesis contributes to the field through the integration of surface and underwater obstacle maps with the depth map to produce a multi-class traversability map. This map distinguishes itself by being both lightweight and less complex than traditional obstacle maps, thereby offering a novel approach to navigating surface water vehicles in challenging environments.

## 8.1 Key Findings

This thesis uncovers several key findings relevant to the challenges of mapping post-flood environments, where traditional mapping solutions like Google Maps or OpenStreetMaps become impractical due to the extensive destruction of road networks. In this context, the comprehensive assessment of surface and underwater environments is crucial for executing effective rescue operations. While aerial or satellite mapping provides some utility in these scenarios, it fails to map the underwater environment, which is essential for the safe navigation of rescue boats. Consequently, deploying surface water vehicles has emerged as an essential component for thoroughly understanding underwater and surface water conditions.

In underwater sensing, the thesis identifies two principal challenges posed by post-flood environments: the extreme shallowness and the variability of the waterbed. These factors contribute to increased levels of random speckle and reverberation noise in sonar imagery. The image enhancement technique developed in this research significantly improves sonar image signal-to-noise ratio (SNR) and contrast. Moreover, the depth estimation derived from sonar imagery exhibits remarkable precision, proving critical for navigation in exceptionally shallow waters.

The lack of datasets representing post-flood conditions is addressed by creating a realistic simulation encompassing complex scenarios similar to those encountered in flooded areas. Given the challenges associated with the physics modeling of sonar acoustics—principally due to random speckle and reverberation noise—the use of Pix2Pix Conditional Generative Adversarial Networks (GANs) for generating sonar images has offered a more accurate and realistic alternative. This approach effectively narrows the gap between simulated environments and real-world post-flood scenarios.

Moreover, the thesis introduces a traversability map that is both less complex and lightweight, facilitating rapid and efficient data transmission to other rescue boats. Lastly, the integration of Real-Time Kinematic (RTK) GPS technology, with its 2cm positional accuracy, significantly enhances the overall accuracy of the traversability map. These



findings collectively contribute to advancing navigation and mapping technologies for rescue operations in post-flood environments, offering new avenues for safe and effective disaster response strategies.

## 8.2 Future Work and Outlook

The research presented in this thesis lays a solid foundation for the advancement of post-flood environment mapping, by the integrating highly accurate GPS/GNSS systems, the fusion of multiple sensors including cameras and LiDAR, and the rigorous evaluation of methodologies through both simulated and natural water bodies. These factors collectively contribute to a better accuracy of up-to-date traversability maps. However, the journey towards refining and enhancing the utility of these maps does not end here. Future research initiatives should prioritize the advancement of underwater object recognition techniques. This should explicitly focus on using material properties, such as acoustic reflectance, to significantly enhance the capability to classify obstacles based on their safety levels for navigation. Additionally, the exploration of texture-based classification methods appears promising for surface water object detection. This is particularly relevant in flooded environments that often contain a variety of traversable obstacles, including floating leaves and soft debris, that traditional detection methods may overlook.

Another significant area for future exploration involves addressing the diversity in sonar imagery resulting from distinct operating frequencies and hardware specifications of various sonar devices. It is crucial to develop more generalized deep-learning models that can adapt to this diversity. Achieving this goal necessitates collecting a broader range of underwater data utilizing various sonar equipment, which could significantly improve sonar-based mapping and navigation techniques.

Enhancing the robustness of the proposed architecture could also benefit from integration of additional sensing modalities. For instance, advanced synthetic aperture sonars (SAS) and multispectral imaging, could unveil new dimensions of environmental insight. These sensors hold the potential to offer detailed information about underwater terrains and surface water vegetation, respectively. Moreover, the accuracy of underwater obstacle detection could be further improved through the fusion of multiple sonars specifically having different frequencies. Similarly, the precision of the traversability map may be enhanced by fusing multi-perspective maps generated by various surface water vehicles.

In the long term, conducting data collection and architectural evaluations in actual flooded environments and their digital twins could substantially increase the efficiency and applicability of the proposed architecture. Additionally, integrating maps derived from other sources, including satellite, aerial, and the SWiM architecture, holds the potential to significantly reduce the time required for comprehensive environmental mapping. Such a multifaceted approach to data integration and analysis would not only refine the accuracy of environmental mapping but also contribute to the development of more efficient and effective navigation and rescue operations in post-flood scenarios.

Looking ahead, the outlook for post-flood environment mapping is promising, with the potential to significantly impact disaster management and recovery processes. The integration of real-time data collection and mapping technologies offers the prospect of developing dynamic maps that can provide immediate guidance to rescue and recovery teams in the

aftermath of a flood. This real-time capability, combined with the advancements in sensor technologies and data processing algorithms, could revolutionize the presented approach to disaster response, making it more agile and effective. Moreover, the ongoing collaboration between academic researchers, technology developers, and disaster management agencies will be critical in driving these innovations forward, ensuring that the methodologies developed are not only technologically advanced but also practically applicable in real-world scenarios. As we continue to witness the increasing frequency and severity of flooding events due to climate change, the importance of advancing the mapping capabilities cannot be overstated. The work carried out in this thesis represents a step towards a future where communities are better prepared to respond to and recover from flood disasters, underscoring the critical role of technology and innovation in enhancing our resilience to environmental challenges.

# A. System and Specification

The preceding chapters 5 - 7 present a novel SWiM (Shallow Water Traversability Mapping) architecture that is specifically designed to map a post-flood environment, where the up-to-date standard maps are unavailable due to the abrupt changes in the environment, significantly impeding the safe navigation of rescue teams. This architecture seamlessly fuses several sensory modalities to compute a robust traversability map encompassing both surface and underwater domains of a post-flood environment. As the choice of a sensor is essential for robustness and better accuracy of this architecture, this appendix discusses the sensory modalities that are used to evaluate the functionality of SWiM architecture. Furthermore, the detailed specifications of each sensor are provided to offer the in-depth technical insight.

## Camera

A post-flood environment is often highly cluttered with random and unstructured objects such as fallen trees, broken bridges and floating debris. These objects create several navigational challenges for rescue teams specifically the detection of critical and damaging obstacles over and under the water's surface. To overcome the detection problem over the surface of water, the SWiM architecture has integrated several deep learning models (discussed in Sections 6.2.1 and 6.2.1) that solely rely on camera imagery to compute the visual features. Therefore, the camera imagery stands as an essential sensory modality, and its significance has increased by the growing field of deep learning. Moreover, stereo cameras increase this utility by providing not only high-resolution imagery but also critical depth information about the environment. This depth perception is essential for creating a three-dimensional representation of the surroundings, typically in the form of a point cloud. Such 3D environmental data, obtained from stereo camera systems, is integral in enhancing the perception, navigation, and decision-making capabilities of autonomous systems, thereby solidifying the role of camera imagery in the domain of advanced autonomous technologies.

The surface vehicles presented in this thesis utilize Stereolabs' ZED2i stereo camera, as depicted in Figure A.1. The ZED2i camera is versatile, offering a range of resolutions from



**Figure A.1:** ZED stereo camera attached on ASTRIDER.

VGA to 2K and supporting frame rates varying from 15fps to 100fps, catering to diverse imaging requirements. A significant enhancement in this model is the integrated Inertial Measurement Unit (IMU), which significantly improves visual odometry, surpassing the capabilities of its predecessors. Additionally, the ZED2i is precisely engineered for outdoor applications, boasting an IP66 rating that signifies its resilience against environmental elements. Table A.1 provides an extensive overview of the ZED2i's specifications, highlighting its adaptability, durability, and advanced imaging features, making it an ideal choice for robust outdoor autonomous systems. The camera's combination of high-resolution imaging, variable frame rates, and enhanced odometry, all within a durable outdoor-friendly design, underscores its suitability for sophisticated applications in autonomous navigation and environmental mapping.

**Table A.1:** Specifications of ZED2i camera (source: <https://www.stereolabs.com/zed-2i/>).

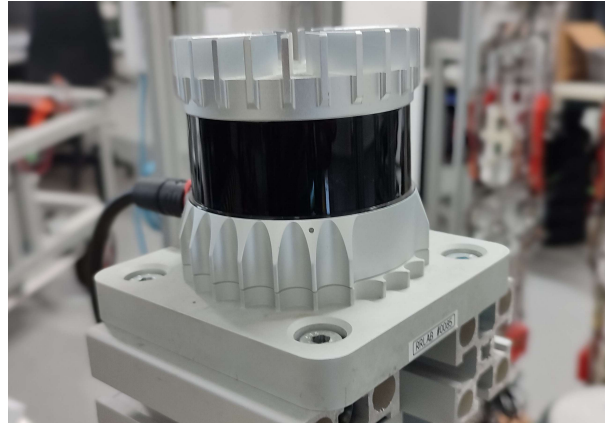
| Property                  | Value  |
|---------------------------|--|
| Serial Number             |  |
|                           | <b>Imaging Sensor</b>  |
| Sensor Type               | Dual 1/3" 4MP CMOS   |
| Resolution                | 2688 x 1520 px   |
| Pixel Size                | 2 $\mu$ m x 2 $\mu$ m  |
| Shutter                   | Electronic synchronized rolling shutter  |
| Output Resolution         | Side by Side<br>2x (2208x1242) @15fps - cropping mode<br>2x(1920x1080) @15/30fps - cropping mode<br>2x (1280x720) @15/30/60fps - binning 2x2 mode<br>2x(662x376) @15/30/60/100fps - binning 4x4 mode |
| Output Format             | YUV 4:2:2 - UYV(8bits)   |
| Max S/N Ration            | 38.3 dB  |
| Dynamic Range             | 64.6 dB  |
| Sensitivity               | 1900 mV/Lux-sec  |
| Chroma                    | Color  |
| Baseline                  | 12 cm (4.72 in)  |
|                           | <b>Temperature Sensor</b>  |
| Temperature Range         | -40 to 125 °C  |
| Abs. Temperature Accuracy | +/-0.5 °C  |
| Output Data Rate          | 25 Hz  |
|                           | <b>Inertial Measurement Unit</b>   |
| Accelerometer Range       | +/- 8G   |
| Accelerometer Resolution  | 0.244 mg   |

|                             |  |
|-----------------------------|--|
| Accelerometer Noise Density | 3.2 mg   |
| Gyroscope Range             | +/- 1000 dps                                   |
| Gyroscope Resolution        | 0.03 dps                                       |
| Gyroscope Noise Density     | 0.16 dps                                       |
| Sensitivity Error           | +/- 0.4%                                       |
| Output Data Rate            | 400 Hz   |
| <b>Magnetometer</b>         |  |
| Magnetic Field Range        | +/- 2500 $\mu$ T (z)<br>+/- 1300 $\mu$ T (x,y) |
| Magnetic Field Resolution   | 0.3 $\mu$ T                                    |
| Output Data Rate            | 50 Hz  |
| <b>Barometer</b>            |  |
| Pressure Range              | 300 to 1100 hPa                                |
| Pressure Resolution         | 0.18 Pa  |
| Relative Pressure Accuracy  | 0.12 hPa                                       |
| RMS Noise                   | 0.2 Pa   |
| Output Data Rate            | 25 Hz  |
| <b>Physical</b>             |  |
| Interface                   | USB Type C - External cable (up to 10m)        |
| Dimensions (w, h, d)        | 175.25 x 30.25 x 43.1 mm                       |
| Weight                      | 166g   |
| Power Consumption           | 380mA / 5V USB powered                         |
| Operating temperature       | -10° to +45°C                                  |

## LiDAR

The mapping of surface water environment in a post-flood scenario suffers from several perception challenges including variable light conditions and reflections from water surface. These issues significantly effect the exposure of camera imagery and hence degrade the object detection. In such scenarios, relying solely over visual data may be problematic and hence different sensor modality should be integrated including LiDAR. It is a frequently-used sensor in autonomous systems for environmental sensing. Utilizing light beams, primarily in the infrared spectrum, LiDAR measures the distance to obstacles based on the time of flight and the angle at which the beam reflects off an object. This technology can be categorized into two-dimensional or three-dimensional variants, with 3D LiDAR offering a detailed point cloud that represents the surrounding environment comprehensively. As an active sensor, LiDAR's effectiveness is independent of ambient light conditions, making it exceptionally reliable in low-light or even no-light scenarios. Additionally, LiDAR systems often provide intensity images, which capture the reflective properties of surfaces, adding another layer of environmental data. This feature enhances object detection and terrain analysis, further supporting LiDAR's integral role in the sensory array of autonomous systems.

The Shallow Water Traversability Mapping (SWiM) architecture uses 3D LiDAR, specifically the Ouster OS0 model with 128 lines, for mapping obstacles as detected by cameras. This particular LiDAR model, illustrated in Figure A.2, has a 90-degree vertical field of view, enabling it to capture high-resolution environmental data. Suited for close-range applications, the Ouster OS0 provides enhanced detail in mapping immediate surroundings. Table A.2 details the comprehensive specifications of the Ouster OS0. Additionally,



**Figure A.2:** Ouster OS0-128 3D LiDAR attached on ASTRIDER.

this LiDAR is equipped with a user-friendly web interface, facilitating data logging and basic parameter configurations, such as setting IP addresses and firmware updates. This interface significantly streamlines the process of data acquisition and system maintenance, making the Ouster OS0 a valuable asset in the SWiM architecture for precise and efficient environmental mapping.

**Table A.2:** Specifications of Ouster OS0 (source: <https://ouster.com/downloads/>).

| Property  | Value  |
|---|--|
| Serial Number   |  |
| Inventory Number  |  |
| <b>Optical Performance</b>  |  |
| Range<br>(80% Lambertian reflect.,<br>1024 @ 10 Hz mode)            | 45 m @ 100 klx sunlight, >90% detection probability<br>50 m @ 100 klx sunlight, >50% detection probability |
| Range<br>(10% Lambertian reflect.,<br>1024 @ 10 Hz mode)            | 15 m @ 100 klx sunlight, >90% detection probability<br>20 m @ 100 klx sunlight, >50% detection probability |
| Minimum Range   | 0.3 m (10% reflective target)  |
| Range Accuracy  | $\pm 3$ cm for lambertian targets, $\pm 10$ cm for retroreflectors   |
| Precision<br>(10% Lambertian reflect.,<br>1024 @ 10 Hz mode, 1 std) | 0.3 - 1 m: $\pm 2$ cm<br>1 - 10 m: $\pm 1$ cm<br>10 - 15 m: $\pm 1.5$ cm<br>>15 m: $\pm 5$ cm              |
| Range Resolution  | 0.1 cm<br>: For the Range Resolution = 0.8cm   |
| Vertical Resolution   | 32, 64, or 128 channels  |
| Horizontal Resolution   | 512, 1024, or 2048 (configurable)  |
| Field of View Vertical  | 90° (+45° to -45°)   |
| Horizontal  | 360°   |
| Ang. Sampling Accuracy  | Vertical: $\pm 0.01^\circ$ / Horizontal: $\pm 0.01^\circ$  |
| False Positive Rate   | 1/10,000   |
| Rotation Rate   | 10 or 20 Hz (configurable)   |
| <b>Laser</b>  |  |
| Laser Product Class   | Class 1 eye-safe per IEC/EN 60825-1: 2014  |
| Laser Wavelength  | 865 nm   |

|  |  |
|--|--|
| Beam Diameter Exiting  | 5 mm   |
| Beam Divergence  | 0.35° (FWHM)   |
| <b>Lidar Output</b>  |  |
| Connection   | UDP over gigabit Ethernet  |
| Points Per Second  | up to 655,360 (32 channel)<br>up to 1,310,720 (64 channel)<br>up to 2,621,440 (128 channel)  |
| Data Rate (Mbps)<br>(Legacy Mode)  | up to 66.2 Mbps (32 channel)<br>up to 129.1 Mbps (64 channel)<br>up to 255 Mbps (128 channel)  |
| Data Rate (Mbps)<br>(Dual Return Profile)<br>Not applicable for 1024x20<br>2048x10 Modes | up to 43.6 Mbps (32 channel)<br>up to 85.6 Mbps (64 channel)<br>up to 169.4 Mbps (128 channel)   |
| Data Rate (Mbps)<br>(Low Data Rate Profile)  | up to 23.6 Mbps (32 channel)<br>up to 44.6 Mbps (64 channel)<br>up to 86.6 Mbps (128 channel)  |
| Data Rate (Mbps)<br>(Single Return Profile)  | up to 65.6 Mbps (32 channel)<br>up to 128.5 Mbps (64 channel)<br>up to 254.3 Mbps (128 channel)  |
| Data Per Point   | Range, signal, reflectivity, near-infrared, channel, azimuth angle, timestamp  |
| Timestamp Resolution   | < 1 $\mu$ s  |
| Data Latency   | < 10 ms  |
| <b>IMU Output</b>  |  |
| Connection   | UDP over gigabit Ethernet  |
| Samples Per Second   | 100  |
| Data Per Sample  | 3 axis gyro, 3 axis accelerometer  |
| Timestamp Resolution   | < 1 $\mu$ s  |
| Data Latency   | < 10 ms  |
| Additional Details   | InvenSense ICM-20948; datasheet for more details:<br><a href="https://www.invensense.com/products/motion-tracking/9-axis/icm-20948/">https://www.invensense.com/products/motion-tracking/9-axis/icm-20948/</a> |

## Forward-Looking Imaging Sonar

The standard perception sensors including cameras and LiDAR are generally useful in surface water obstacle detection but are not workable solution in underwater obstacle detection systems. This is due to the higher attenuation of light waves within the water compared to air. Contrarily, the acoustic impedance of water is significantly higher as compared to air, therefore, sound is lesser attenuated in water. Sonar technology, which operates by emitting sound waves and measuring their reflection time to detect obstacles, is a pivotal sensor in underwater environments.

Forward-looking sonar is specifically utilized for identifying obstacles to ensure safe navigation in aquatic settings. The resolution of sonar imagery is closely linked to the frequency of the emitted sound waves; higher frequencies are crucial for detecting smaller objects, as lower frequencies typically yield lower resolution and may miss finer details. In post-flood environments, where detecting submerged obstacles is essential for safe navigation, high-frequency sonar is preferred due to its ability to discern smaller

objects in shallow waters. Additionally, imaging sonar systems can convert sound wave data into visual images, providing a more intuitive and detailed representation of the underwater environment. This capability is particularly beneficial in complex post-flood scenarios, where visualizing the nature and distribution of obstacles can significantly enhance navigational safety and efficiency. Thus, in the context of post-flood traversability, the deployment of high-frequency imaging sonar is instrumental in creating accurate and detailed maps of submerged landscapes.



**Figure A.3:** Tritech’s Gemini 720ik attached on ASTRIDER.

Within the Shallow Water Traversability Mapping (SWiM) architecture, the Underwater Traversability Module (UTM) employs the advanced Tritech Gemini 720ik multibeam imaging sonar, a forward-looking sonar system shown in Figure A.3, for effective underwater navigation and mapping. This high-resolution sonar is adept at producing detailed acoustic images, an essential feature in turbid and low-visibility water conditions typical of post-flood environments. The optical systems’ limitations in such settings are effectively counteracted by the Gemini 720ik’s proficiency in penetrating murky waters. Its capability to function reliably under these adverse conditions is critical for the precise mapping of submerged terrains and the identification of navigational hazards. The sonar’s detailed acoustic imaging facilitates the detection of obstacles that could potentially hinder navigation and impede rescue operations. Consequently, the Gemini 720ik sonar stands as a key component in the UTM of the SWiM architecture, enhancing the safety and efficacy of operations in challenging post-flood scenarios by ensuring thorough and accurate underwater terrain analysis. Table A.3 presents the detailed specification of Gemini sonar.

**Table A.3:** Specifications of Tritech’s Gemini 720ik sonar (source: <https://www.tritech.co.uk/products/gemini-720ik>).

| Property      | Value |
|---------------|-------|
| Serial Number |       |



|                                |   |
|--------------------------------|---|
|                                | <b>Imaging Sensor</b>   |
| Operating Frequency            | 720kHz  |
| Angular resolution             | 1.0° acoustic, 0.25° effective  |
| Supply voltage                 | 19V to 74V DC   |
| Power requirement              | 16W - 27W (range dependent)   |
| Main port protocol             | Ethernet  |
| Depth rating                   | 350m / 1148ft   |
| Weight in air                  | 1.40kg / 3.08lbs  |
| Weight in water                | 0.43kg / 0.96lbs  |
| Range                          | 0.1m - 120m / 4in - 394ft   |
| Number of beams                | 512   |
| Horizontal beam width          | 120°  |
| Vertical beam width            | 20°   |
| Range resolution               | 4mm / 0.2in   |
| Update rate                    | 5 - 97Hz (range dependent)  |
| Mode of operation              | CHIRP or CW   |
| Speed of Sound                 | Integrated VoS sensor   |
|                                | <b>Interface</b>  |
| Supply voltage                 | 19V to 74V DC   |
| Power requirement              | 16W - 27W (range dependent)   |
| Main port protocol             | Ethernet  |
| Auxiliary port protocol        | RS232, TTL in, pass-through power (2.5A max)                          |
| Connector type                 | MAIN port: Impulse MKS(W)-307-FCR<br>AUX port: Impulse MKS(W)-307-FCR |
|                                | <b>Physical specification</b>   |
| Depth rating                   | 350m / 1148ft   |
| Weight in air                  | 1.40kg / 3.08lbs  |
| Weight in water                | 0.43kg / 0.96lbs  |
| Temperature rating (operating) | -10°C to 35°C / 14°F to 95°F  |
| Temperature rating (storage)   | -20°C to 50°C / 4°F to 122°F  |

## Global Navigation Satellite System

In a post-flood rescue operations, the rescue teams often require the up-to-date environment map to navigate and plan accordingly. This map needs to be globally referenced with gps position for easy understanding and usage of the map. The incorporation of Global Navigation Satellite System (GNSS) technology in outdoor robotic mapping represents a significant enhancement in the field, providing critical geospatial data essential for various applications. GNSS, encompassing systems like GPS, GLONASS, Galileo, and BeiDou, offers comprehensive global coverage and enhanced accuracy in positional data, crucial for the effective functioning of outdoor robots in diverse environments. Moreover, GNSS technology enhances the autonomy of outdoor robots, enabling real-time decision-making and adaptive path planning in dynamic and unpredictable settings. This autonomy is crucial for tasks that require navigation over large areas or in conditions where human intervention is limited. Additionally, GNSS facilitates coordinated operations in multi-robot systems, allowing for synchronized task execution and collaborative mapping efforts, further enhancing efficiency and data quality.

The integration of GNSS with other sensory technologies, such as LiDAR, cameras, and inertial measurement units, enriches the spatial data acquired, enabling the creation of



(a) Trimble BX992 GNSS receiver enclosure



(b) Right side antenna from dual-antenna system.

**Figure A.4:** A dual-antenna Trimble GNSS system with integrated inertial navigation system powered by the BD992-INS.

comprehensive, multi-dimensional maps. This integration not only improves the accuracy of the mapping process but also contributes to the development of sophisticated algorithms for spatial analysis, obstacle detection, and environmental interaction.

The Trimble Dual Antenna BX992 GNSS (shown in Figure A.4a) is a high-end, state-of-the-art receiver integrates dual-antenna inputs with a GNSS engine, facilitating superior heading and position data crucial for applications requiring the utmost spatial accuracy and reliability. The BX992's dual-antenna (Figure A.4b) configuration is particularly advantageous in scenarios where the orientation of the device is as critical as its geographic location, such as in autonomous vehicle navigation, marine surveying, and aerial mapping. One of the key features of the BX992 is its ability to deliver Real-Time Kinematic (RTK) positioning, providing centimeter-level accuracy essential for precision-dependent tasks like land surveying, agricultural automation, and construction. This high level of accuracy is achieved through advanced signal processing algorithms and error correction techniques, which mitigate issues such as ionospheric and tropospheric interferences. Additionally, the robust build and design of the BX992 make it suitable for harsh and demanding outdoor environments, ensuring consistent performance even in challenging conditions.

**Table A.4:** Specifications of Trimbles BX992 Receiver (source: <https://info.intech.trimble.com/bx992-datasheet>).

| Property                         | Value       |
|----------------------------------|-------------|
| Serial Number                    |             |
| <b>Performance Specification</b> |             |
| <b>Time to First Fix</b>         |             |
| Cold Start                       | <45 seconds |
| Warm Start                       | <30 seconds |
| Signal Re-acquisition            | <2 seconds  |
| <b>Velocity Accuracy</b>         |             |
| Horizontal                       | 0.007 m/sec |
| Vertical                         | 0.020 m/sec |

---

Maximum acceleration GNSS Tracking  $\pm 11g$

**Inertial Sensors**

Maximum acceleration  $\pm 6g$   
Maximum angular rate  $\pm 350$  deg/sec

**Maximum Operating Limits**

Velocity 515 m/sec  
Altitude 18,000 m

---

**Interface**

Supply voltage 9V to 30V DC  
Power requirement Typically 3W  
IO Connector D-sub DE9 and DA26  
Antenna Connector 2× TNC (Female)  
Minimum required LNA Gain 32.0 dB

---

**Environmental Characteristics**

Operating Temperature  $-40$  to  $+75^{\circ}C$   
Storage Temperature  $-55$  to  $+85^{\circ}C$   
Weight in water 0.43kg / 0.96lbs  
IP Rating IP67

---



## B. Surface Water Vehicles

Disaster management teams globally heavily depend on surface vehicles (SVs) for executing efficient post-flood rescue operations. However, SVs exhibit diverse sizes and distinctive kinematics when navigating through water. In certain scenarios, the deployment of a larger SV might prove impractical due to limited mobility. Therefore, an adaptable robotic system that can operate efficiently on any vehicle type is crucial in disaster management. To address this challenge, it is necessary to test and evaluate the robotic system's performance on different SVs with varying sizes, kinematics, and sensor system positions. Chapter 4 presents a dataset gathered on three distinct SVs to assess the robotic system's capability. The dataset provides a comprehensive evaluation of the robotic system's performance under different conditions.

The PONTONBOOT project, as referenced in [BMW 21], provides the framework for analyzing specialized surface vehicles designed for post-flood relief and rescue operations. This project, initiated in December 2018 under the DIVERS<sup>1</sup> program and funded by the BMWI<sup>2</sup>, had the objective of developing a highly versatile and modular surface water robotic platform. The PONTONBOOT project was conceived with the intention of fostering diversification in defense companies, steering them towards the creation of technologies for civilian security purposes. The project's focus was to innovate platforms that could adapt to the dynamic and challenging environments typically encountered in post-flood scenarios, thereby enhancing the efficacy and scope of relief efforts. This initiative represents a significant step in the integration of advanced robotic technology in disaster management, particularly in the development of specialized equipment capable of operating effectively in the aftermath of floods, where conventional rescue methods may be inadequate or infeasible.

In this research thesis, an in-house built novel prototypical raft called ASTRIDER is presented that is designed specifically for this study. In addition, a novel dataset called MASTER is introduced in Section 4.1.2. This dataset is mostly collected using ASTRIDER,

---

<sup>1</sup>Innovationsprogramm zur Unterstützung von Diversifizierungsstrategien von Verteidigungsunternehmen in zivile Sicherheitstechnologien

<sup>2</sup>Bundesministerium für Wirtschaft und Energie

but some parts of the dataset and results are also obtained using other commercial drone boats such as BASILISK and EINZELBOOT, which are also described below.

## Astrider

The development of ASTRIDER (Autonomous Shallow waTeR Inspection and Depth Exploration Robot) is a crucial phase in this research. It serves as a lightweight experimentation platform and a carrier for sensors. This raft is designed to be easily transported and flexible for experimentation, embodying the principles of adaptability and practicality. It can be easily disassembled into several main components for convenient transport in a commercial vehicle. These components include a body made from versatile Minitect profiles, a sensor array, a reconfigurable autonomy box, custom-made holders for buoyancy units, and a propulsion unit. The design's modularity allows for efficient transport and assembly on-site, enabling swift deployment by a team of two individuals.



**Figure B.1:** The prototypical ASTRIDER equipped with the autonomy box.

The RRLAB astrider, as shown in Figure B.1, is equipped with advanced technical specifications and features that enable it to navigate autonomously and map its surroundings. It is designed to be reconfigurable and waterproof, which makes it suitable for challenging aquatic environments with not too strong water stream. One of its unique features is the adjustable sonar tilt and depth capabilities, which allows the vehicle to adapt to changing underwater conditions and improve its obstacle detection and avoidance abilities. The vehicle is powered by an industrial-grade processor and 16GB graphical processing unit, which enhances its computational power and allows for real-time data processing and decision-making.

The ASTRIDER has undergone rigorous testing in various water bodies, proving its operational excellence in different aquatic scenarios. The vehicle has a compact design, measuring 1.4 meters in length, 1.1 meters in width, and 0.75 meters in height, without compromising on versatility. Despite its small size, it is strong enough to carry a maximum payload of 32 kilograms, weighing up to 25 kilograms. Furthermore, the kinematics of the ASTRIDER are powered by a single motor drive, which enables efficient and streamlined movement through water. This propulsion system allows the vehicle to reach a maximum speed of 1.5 meters per second, ensuring swift and agile navigation. The power source of

the vehicle comprises of a 15.75 ampere-hour battery with a voltage of 25.9 volts, providing sustained operational capabilities for up to 2 hours.

The technical specifications of ASTRIDER make it an ideal solution for underwater exploration and surveillance, equipped with cutting-edge technology. The platform's autonomous navigation capabilities, combined with a customizable chassis and a robust waterproof design, make it an adaptable tool for a wide range of applications. Whether it's navigating through complex surface and underwater environments or conducting surveys in challenging water bodies, the ASTRIDER is a fusion of advanced technology and practical engineering.

## Basilisk

The BASILISK is a remote-controlled Unmanned Surface Vehicle (USV) (Figure B.2) used in the PONTONBOOT project. It has several technical features, including autonomous navigation and a waterproof system, which make it reliable and efficient for use in different water bodies. Furthermore, BASILISK is a compact and maneuverable surface vehicle, measuring 1.4m x 1.1m x 0.75m, with a weight of up to 270 kg. Its twin-hull kinematics design, with an individual motor in each hull, allows for precise maneuvers, including rotation on the spot, making it easy to navigate through various water conditions. The vehicle can carry a maximum load of up to 32 kg, making it suitable for a variety of applications. It can also move quickly and efficiently through the water, with a maximum speed of 1.5m/s.



**Figure B.2:** The BASILISK equipped with autonomy kit and the sensor array of ASTRIDER. [Meckel 22]

The vehicle has a battery capacity of 15.75 Ah and a voltage of 25.9V. It is equipped with an autonomy kit that includes sensors and hardware (detailed in Appendix A), which draw power efficiently from the host vehicle, optimizing the overall energy management system. The autonomy kit has a working time of 5 hours and seamlessly integrates with the SWiM architecture based on FINROC, through an embedded ROS node that acts as a mediator. This hardware abstraction (discussed in Appendix C) ensures a seamless connection between the BASILISK and SWiM architecture, allowing for harmonization of the diverse components within our robotic system.

## Pontonboot

The PONTONBOOT project has developed novel pontoon boats with modular designs and configurable building blocks. This design emphasizes standardization and modularization, allowing for quick and easy on-site reconfiguration of boat shapes directly in water. This adaptability contrasts with the compact vessels explored in [Kelly 19], where reconfiguration is achieved through automated means. In the case of the pontoon boats, the reconfiguration process is manual, making it faster to adjust to the operation's requirements.



(a) A PONTONBOOT core module with autonomy kit and coupled stern module.



(b) One of the PONTONBOOT work platform configurations with payload.

**Figure B.3:** The PONTONBOOT hardware kit is equipped with an additional touchscreen. [Meckel 22]

The primary focus of the project is on the single boat configuration, known as EINZELBOOT (shown in Figure B.3a), with dimensions of  $12.0, m \times 2.5, m \times 2.0, m$ . This configuration represents a robust and versatile platform that accommodates the autonomy kit seamlessly. However, it is essential to note that the autonomy kit's applicability extends beyond the single-boat setup. It has been successfully deployed on working platforms and floating bridge configurations (shown in Figure B.3b), highlighting its adaptability and universality across various configurations.

It is worth mentioning that General Dynamics European Land Systems (GDELS) crafted the prototypes, which underwent public demonstrations in November 2021 at the Elbphilharmonie in Hamburg, Germany. The successful showcase demonstrated the boats' technical capabilities and their applicability in disaster response and control operations, achieving the project's objectives.



## C. Framework for Intelligent Robot Control

The methodology of SWiM architecture provided in this thesis is implemented and evaluated over several surface water vehicles including ASTRIDER, BASILISK, and EINZELBOOT (discussed in Appendix B) using Framework for Intelligent Robot Control, commonly known as FINROC [Reichardt 12], [Reichardt 13]. This framework is a middleware capable of handling various robotic platforms, sensors, and tasks. The core architecture of FINROC is a sophisticated framework designed to facilitate real-time computation and seamless data transfer. Key components within this architecture include serialization, threading, runtime construction, and the utilization of distinct plugins for loading dataports. At its essence, FINROC is structured around the concepts of modules, groups, and parts, offering a hierarchical approach to development <sup>1</sup>. The framework operates with a high level of abstraction to simplify the development process, providing a more intuitive way for developers to interact with the system. This abstraction is particularly beneficial for users who want to focus on task-specific functionalities without delving into the intricacies of low-level control.

The hierarchical structure of FINROC is organized into two pivotal components: RRLIBS and LIBRARIES. These components are pre-existing modules and groups, strategically crafted to be directly integrated into any project. Leveraging RRLIBS and LIBRARIES significantly expedites the development process, allowing developers to build upon existing functionalities rather than starting from scratch. This modularity and reusability are crucial aspects of FINROC's design philosophy, promoting efficiency and reducing development time. Additionally, serialization and threading are integral aspects of FINROC's real-time computation capabilities. Serialization facilitates the conversion of complex data structures into a format that can be easily transmitted and reconstructed. Threading, on the other hand, enables parallel execution of tasks, enhancing the framework's performance and responsiveness. These features are essential for applications where timely and synchronized data processing is critical, such as robotics and automation.

---

<sup>1</sup><https://www.finroc.org/>

Furthermore, the runtime construction of modules and the loading of dataports through plugins contribute to the adaptability of FINROC. The framework can dynamically incorporate new functionalities and interface with various sensors and actuators, making it well-suited for applications with evolving requirements or diverse hardware configurations.

## Key Features

**Modularity and Extensibility:** A key strength of FINROC lies in its modular architecture. It is designed to accommodate various modules that encapsulate specific functionalities. This modularity enhances extensibility, allowing developers to integrate new sensors, actuators, or algorithms seamlessly. This adaptability is crucial for deploying robots in different scenarios with distinct requirements.

**Sensor Integration:** FINROC places a strong emphasis on sensor integration, acknowledging the significance of perception in robotic systems. The framework supports a wide array of sensors, including cameras, LiDAR, IMUs, and more. This diversity enables robots equipped with FINROC to perceive and interpret their surroundings effectively.

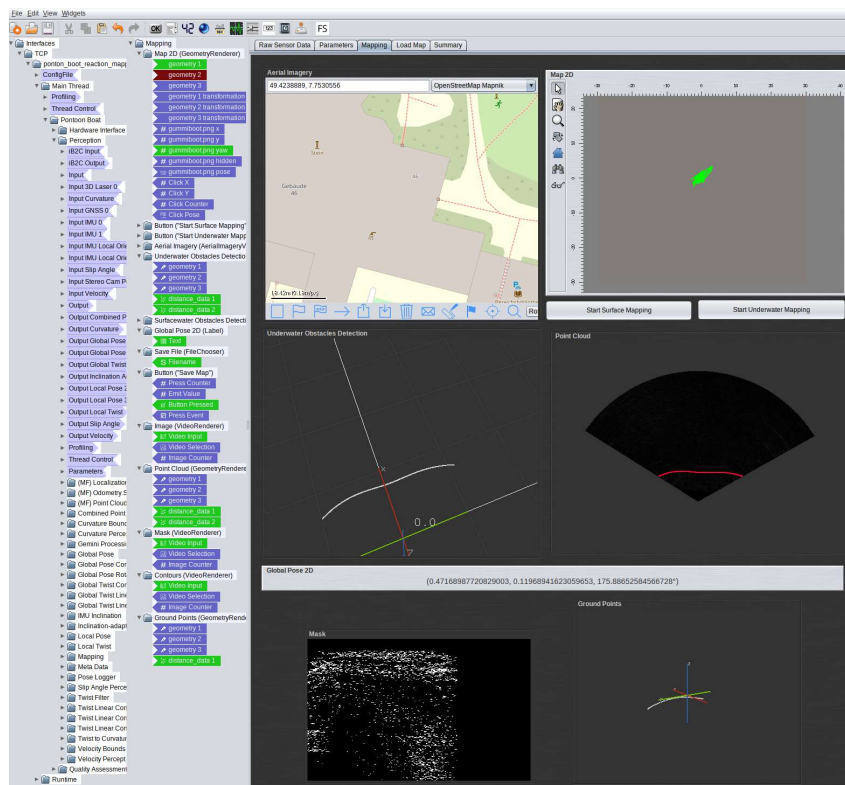
**Task Abstraction and Planning:** Task abstraction is a critical aspect of intelligent robot control. FINROC provides a high-level abstraction layer for defining tasks, allowing developers to focus on task planning rather than low-level control intricacies. This feature streamlines the development of complex robotic applications, from autonomous navigation to manipulation tasks.

**Real-time Control** The framework is optimized for real-time control, a prerequisite for many robotic applications. Whether it's navigating through dynamic environments or executing precise manipulation tasks, FINROC's real-time capabilities ensure the responsiveness and agility of robotic systems.

## Tools

The FINROC framework stands as a comprehensive suite offering indispensable tools to strengthen the development of robots. This support extends across critical functionalities such as data recording, playback, real-time visualization of data and software states, and the provision of a user interface for seamless interaction with robotic systems. At the core of these capabilities is the FINGUI (FINROC GUI), a powerful component that enables users to assess visualization outputs and craft tailored interfaces to meet specific robotic requirements. This tool serves as a pivotal component facilitating user interaction and visualization. One of its distinctive features is the ability to offer a highly customizable interface, allowing users to shape the visual representation and interaction paradigms according to the unique needs of the robotic system or specific control aspects. This level of customization ensures that the interface aligns precisely with the intricacies of the robotic application, enhancing user experience and system usability.

Furthermore, FINGUI is equipped with an array of widgets (shown in Figure C.1), each serving a specific purpose in the visualization and control processes. These widgets encompass a diverse set of tools, including joysticks, labels, LED fields, 2D and 3D rendering canvases, camera systems, maps, oscilloscopes, and an extensive collection of



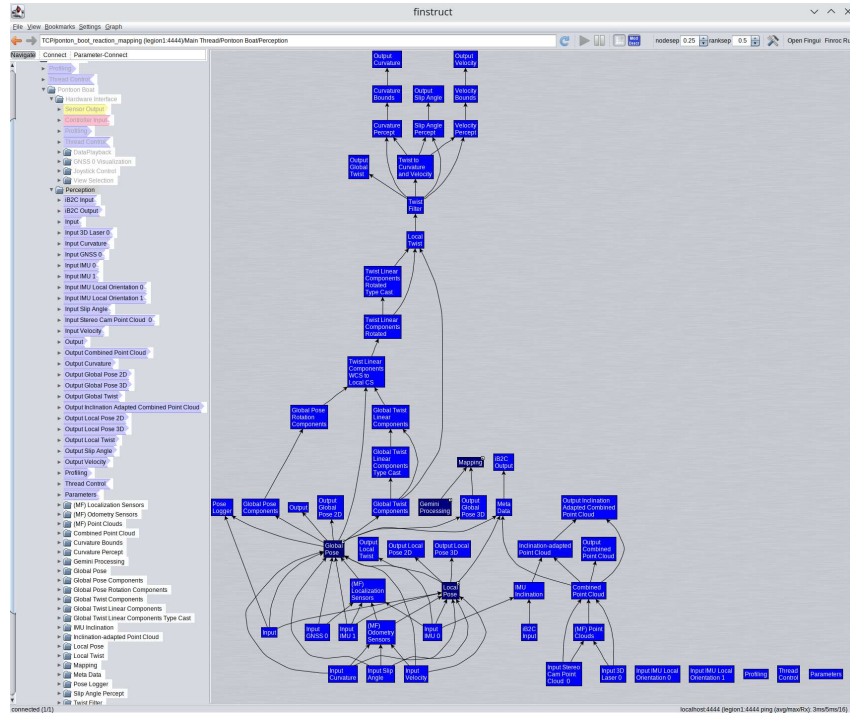
**Figure C.1:** FINGUI data processing visualization of PONTONBOOT Project

additional elements. The provision of such a rich variety of widgets empowers users to construct highly specialized and context-aware interfaces. Whether it involves real-time monitoring of sensor data, controlling actuators, or orchestrating complex robot behaviors, this gui offers the flexibility needed to adapt to diverse robotic applications.

Additionally, FINROC introduces a complementary tool known as FINSTRUCT, designed to provide insight into the intricate structural elements of the control software. This tool serves a crucial role in visualizing the component graph structure and port connections within the control software, offering a profound understanding of the system's architecture. Given the inherent complexity of control software in robotic applications, FINSTRUCT becomes an invaluable asset, allowing developers to navigate and comprehend the complex relationships between different components.

A key functionality of FINSTRUCT is its ability to create custom views as shown in figure C.2, offering users the flexibility to focus on specific program regions. These program regions typically comprise groups and modules, which are fundamental units in the Finroc framework. This feature is particularly beneficial in managing the complexities of control software, providing developers with the means to tailor their views and concentrate on relevant portions of the system. As a result, developers can streamline their workflow and efficiently navigate through the complexities of the control software.

Within FINSTRUCT, the visualization of port connections is a key aspect that facilitates the comprehension of data flow and interaction between different components in the control software. The accessibility of port data and timestamps further enriches the visualization, offering a comprehensive understanding of the temporal aspects of data exchange. Furthermore, FINSTRUCT provides the feature of the iB2C view, which offers a



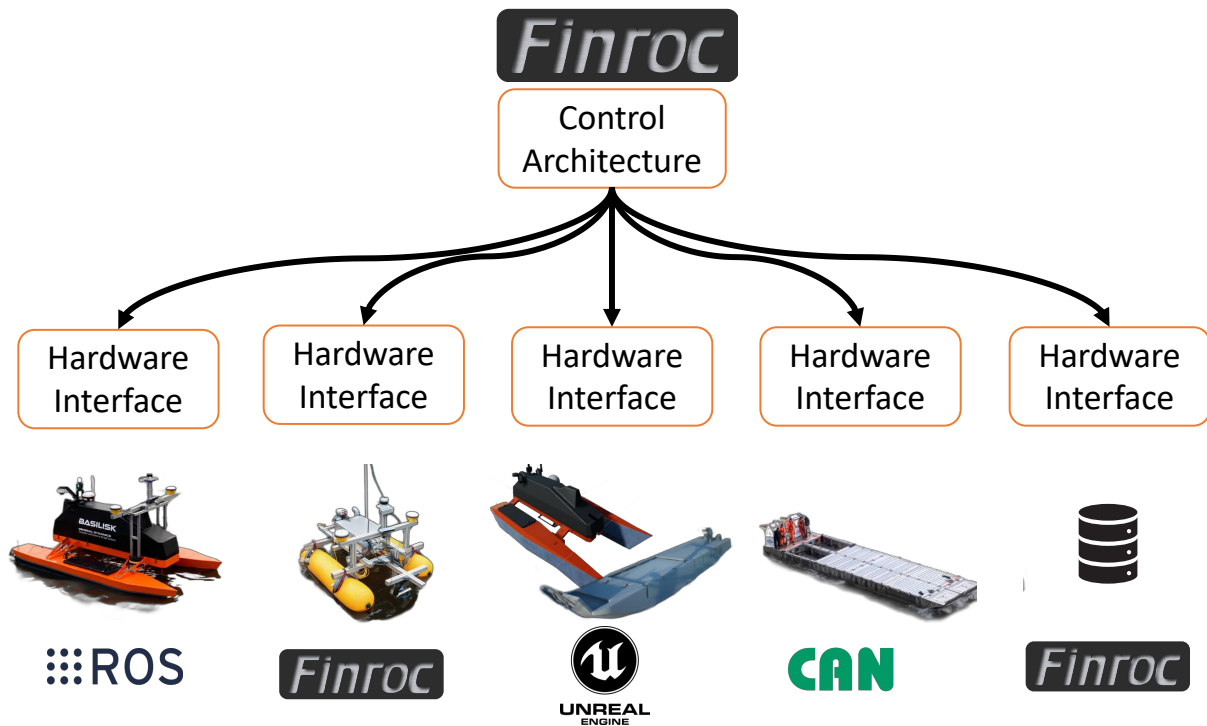
**Figure C.2:** FINSTRUCT component graph visualization of PONTONBOOT Project

dynamic representation of the network’s interface status. This visualization employs color-coded indicators to depict different aspects of activity. Green indicates active elements, red signifies elements with target ratings, and yellow represents elements with potential. Moreover, the colors of connections convey crucial information about the nature of the connection, distinguishing between stimulation and inhibition. This feature enhances the situational awareness of developers, providing real-time feedback on the status and activity levels of different components in the network.

## Hardware Interface

The *Hardware Interface* in the REACTION framework [Wolf 22] is implemented in this thesis through a variety of backend types, encompassing an array of hardware, simulations, and playback systems, as depicted in Figure C.3. This interface serves as a gateway to both actuators and sensors, facilitating interaction with the Unmanned Surface Vehicle’s (USV’s) physical components. It additionally provides essential details regarding the USV, such as its dimensional specifications and information about the integrated sensors. Leveraging the versatility of the behavior-based paradigm and the *iB2C* framework’s fusion capabilities, the control architecture dynamically configures its perception structure based on the information received from the hardware description.

The REACTION system is designed to support various *Hardware Interface* implementations and the corresponding hardware descriptions, ensuring broad applicability and adaptability. In an enhancement from the original interface, this implementation includes added support for sonar sensors, recognizing their importance in marine environments. Conversely, support for wheel-based odometry, more relevant to terrestrial vehicles, has been omitted. Moreover, modifications have been made to the descriptions of vehicle



**Figure C.3:** Common hardware interface developed to communicate with control architecture. The figure shows that the control architecture which is implemented in Finroc can connect to different backends through a common hardware interface. [Meckel 22]

properties to better suit the specific requirements and characteristics of USVs. These adaptations underscore the framework's commitment to providing a comprehensive and flexible interface that caters to the needs of different vehicle types, particularly in marine applications.

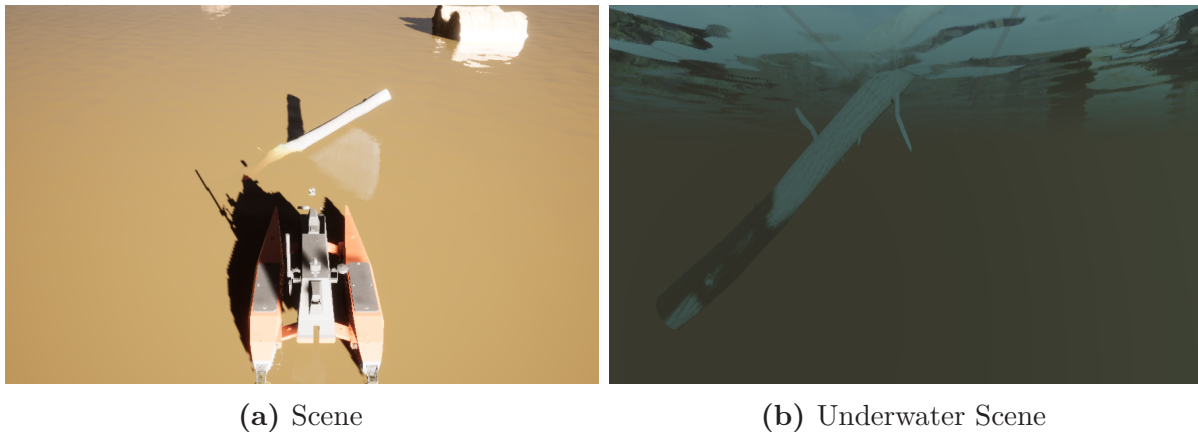


## D. Unreal Engine and Flood Simulation

Due to the unavailability of standard post-flood datasets, the rigorous evaluation of SWiM architecture is not possible. The simpler scenarios for detecting surface and underwater obstacles can be performed on shallow water lakes, and rivers but the complex scenarios cannot be performed and tested. In such cases, a realistic simulation may help in evaluating the complex scenarios. Consequently, this thesis presents a novel flood simulation that includes the design of a complex post-flood environment and generation of raw sensory data. This simulation is developed within Unreal Engine, a state-of-the-art real-time 3D creation tool, originally designed for the gaming industry, has found significant applications in the field of robotics, offering a robust platform for simulation and visualization. The use of Unreal Engine in robotics extends beyond mere visual fidelity, providing a comprehensive suite for the development, testing, and deployment of robotic systems.

The ability of Unreal Engine to render complex scenarios in real-time is a vital asset in robotics. It allows for dynamic simulation environments where parameters can change instantaneously, closely mimicking the unpredictability of real-world conditions. For instance, in autonomous vehicle development, the engine can simulate diverse weather conditions, varying traffic scenarios, and different times of the day, providing a comprehensive test bed for algorithms before actual field deployment. This reduces the need for extensive physical prototyping, thereby saving time and resources while enhancing safety. The Unreal Engine's advanced physics engine offers accurate modeling of physical interactions, which is crucial in robotics. This includes the simulation of various materials, collision detection, and the response of robots to different forces and terrains. This realistic physical simulation (shown in Figure D.1) helps in the fine-tuning of robotic control systems, ensuring that they are robust and reliable in real-world operations. Moreover, the engine's support for high-quality lighting and shadow effects adds to the realism, which is critical for vision-based robotic systems that rely on visual data for navigation and task execution.

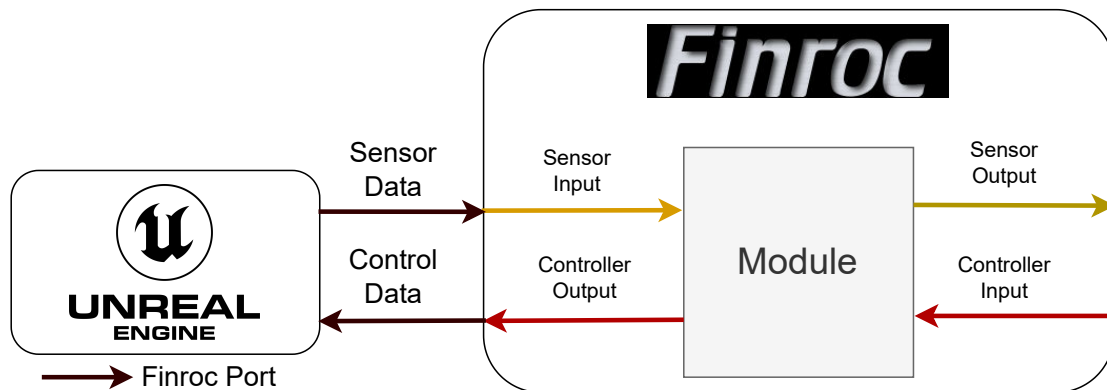
The upcoming sections explain the developed protocols of Unreal Engine to communicate with FINROC, and the sensor plugins that provide realistic sensory data. The novel surface water vehicles models and sonar plugin are already discussed in Sections 4.2.3 and 4.2.4.



**Figure D.1:** Simulation Scene with Underwater Scenario. The figure shows a submerged tree trunk in a simulated barnyard environment with underwater post-processed environment.

## Communication with Finroc

The integration of Unreal Engine with Finroc, a robust software framework for robotics and automation, through a dedicated plugin, represents a significant advancement in the field of robotics simulation and development. Unreal Engine, renowned for its high-fidelity real-time 3D rendering capabilities, provides an ideal platform for creating immersive and realistic simulation environments. When combined with Finroc's modular and flexible architecture, it enables developers to design, test, and refine complex robotic systems with unprecedented efficiency and accuracy.



**Figure D.2:** Unreal Engine to FINROC Interface.

The Unreal Engine plugin for Finroc leverages the engine's powerful graphics and physics simulation capabilities to bring realistic environmental dynamics into robotic simulations. This integration is especially crucial for applications requiring high levels of environmental interaction, such as autonomous vehicles, unmanned aerial vehicles (UAVs), and service robots. The plugin allows for the creation of detailed virtual worlds where robots can navigate and interact, providing valuable data for system development and testing. These environments can range from urban landscapes for testing autonomous cars to indoor settings for domestic robots, offering a versatile platform for a wide array of robotic applications.



A significant advantage of the Unreal Engine plugin for Finroc is its support for real-time data exchange between the simulated environment and the Finroc framework. This feature allows for seamless integration of control algorithms developed within Finroc with the simulated robotic platforms in Unreal Engine. Developers can test their control strategies, monitor their performance, and make iterative improvements in a highly realistic virtual setting, thereby reducing development cycles and improving the robustness of the robotic systems.

Furthermore, the plugin facilitates advanced scenario testing, including dynamic interactions with the environment and other agents. This is essential for developing adaptive and intelligent robotic systems capable of operating in unpredictable real-world situations. Scenarios such as urban streets flooded with sandy water or complex floating debris can be accurately modeled, providing valuable insights into the behavior and performance of robotic systems in challenging conditions.

## Sensors

One of the key features of the plugin is its ability to simulate a variety of sensory inputs that are critical in robotics, such as LIDAR, cameras, and localization sensors. This is particularly important for developing perception algorithms essential for tasks like obstacle avoidance, path planning, and environmental mapping. By accurately simulating these sensors within the Unreal Engine environment, the plugin enables the validation and fine-tuning of sensor fusion algorithms and perception systems under diverse and controlled conditions, which would be difficult or impractical to replicate in the real world.

### Localization Sensor

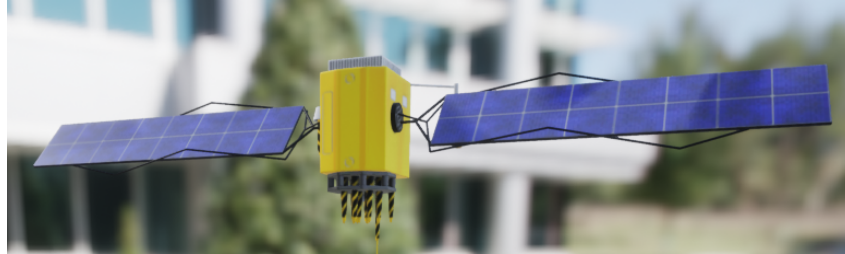
#### Inertial Measurement Unit

IMUs offer acceleration readings for each axis of a robot, along with 3D orientation information. Nevertheless, IMU data is susceptible to errors that accumulate over time and may be influenced by magnetic objects in close proximity. Within UE4, the simulated IMU records the linear velocity vector  $\vec{v}$  and angular accelerations, calculated using tick duration  $\Delta t$ . Standard gravity is applied to replicate earth gravity, and the robot's rotation can be accessed directly. Preprocessing can introduce drifts or slips over time to mimic the physical sensor's quality.

#### Global Navigation Satellite Systems

The Global Navigation Satellite System (GNSS) is crucial for global localization in robotics but is sensitive to environmental factors like shadowing, affecting its accuracy. This impact is quantified using dilution of precision metrics (HDOP, VDOP, PDOP), based on satellite constellations' geometry. In simulations, these metrics are approximated using satellite count to save computational resources, based on real-world data from a robot equipped with GNSS receivers. GNSS error in latitude, longitude, and altitude is determined by visible satellite configuration and random noise. The number of visible satellites is categorized directionally, influencing the GNSS fix category, which ranges from 'No Fix' to '3D Fix'. Unreal Engine 4 (UE 4) simulates GNSS using ray tracing and a satellite manager

to realistically model satellite visibility and environmental effects on GNSS performance. This approach offers a sophisticated and realistic method for understanding GNSS behavior in various conditions. The Figure D.3 shows the simulated mesh of satellite in unreal simulation.



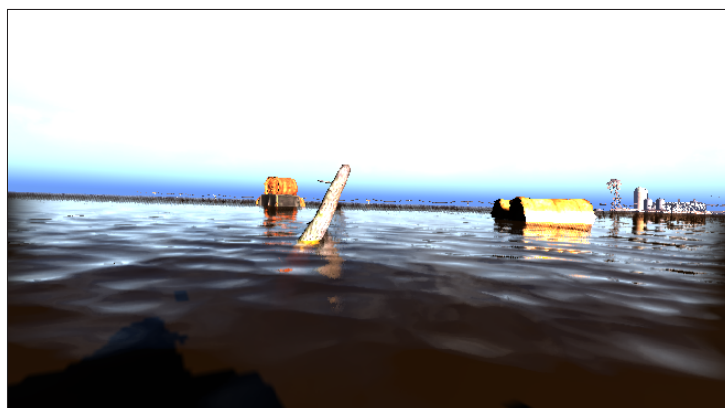
**Figure D.3:** Simulated Satellite Mesh in Unreal Engine.

## Vision and Range Sensors

Vision sensors are categorized into several types, including standard cameras, depth cameras, and laser scanners, each with distinct capabilities and applications. The simulation of these sensors in a virtual environment, such as that provided by Unreal Engine 4 (UE4), involves the emulation of real-world effects such as exposure variations and the presence of particulates like dust. These effects are replicated through the application of post-processing techniques and particle systems within the engine.

### Cameras

Standard cameras within UE4 operate using the engine's rendering pipeline, with camera parameters such as field of view and resolution configured to match the specifications of real sensor data sheets. Simulated cameras (an image shown in Figure D.4) issue render commands that are processed asynchronously, allowing the output to be published once rendering is complete. This non-blocking approach enables the simultaneous use of multiple cameras within complex scenes without impeding the system's update rate.



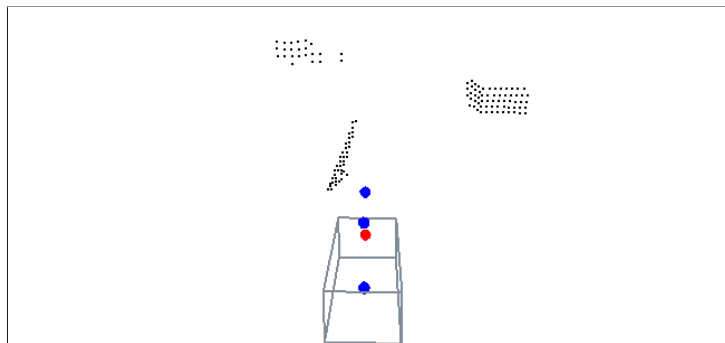
**Figure D.4:** Simulated Camera Image from Flooded Barnyard Environment. The figure demonstrates the simulated camera imagery captured in a realistic barnyard environment by surface water vehicle modeled within Unreal Engine.

Additionally, specialized camera types can be developed by extending a base camera class and customizing properties and post-processing effects. These customizations can be dynamically adjusted via control software, allowing for updates triggered by the robot's perception algorithms.

For the simulation of stereo and depth cameras, which provide additional distance information alongside visual data, UE4 employs techniques such as reading from the depth buffer of the rendering engine. Nonetheless, simulating stereo cameras is more complex due to the requirement of matching a calibrated pair of images. This matching process can create distortions in the resulting point cloud, a byproduct of image disparity. Issues such as reflections, insufficient texture details, and lens anomalies in the simulated device often lead to inaccuracies in the point data, such as spurious obstacles or missing information. To achieve a more realistic sensor simulation, computationally intensive methods like semi-global matching are used to process two simultaneously rendered images, as proposed in [Hirschmuller 08]. This process involves the use of two virtual camera actors with parallel lenses to capture the scene, after which the acquired image pair is rectified and matched to simulate depth perception accurately.

## LiDAR

In parallel, the simulation of Light Detection and Ranging (LiDAR) sensors diverges from that of stereo cameras by utilizing depth rendering techniques rather than image matching. LiDAR sensors, known for their superior precision and range, do not require the simulation of distortions that are commonly associated with stereo vision. The accuracy of LiDAR simulation (shown in Figure D.5) is further enhanced by introducing particle effects like dust into UE4's depth buffer, mimicking the behavior of real-world sensors that are susceptible to environmental particulates, resulting in potentially spurious readings.



**Figure D.5:** Simulated LiDAR Point Cloud from Flooded Barnyard Environment. The figure demonstrates the simulated LiDAR point cloud captured in a realistic barnyard environment by surface water vehicle modeled within Unreal Engine. *The cube in the image is surface vehicle only for visualization.*

The vertical resolution of a simulated LiDAR sensor is often less than its horizontal resolution due to the common use of a limited number of laser lines (e.g., 16 to 128 lines). This discrepancy in resolution is accounted for by omitting certain lines in the depth image, which inherently has a higher resolution due to the square pixel arrangement. The process of line skipping, taking into account the horizontal offset, results in a linearly rendered image. This approach is suitable for simulating single-line laser scanners. However, when

simulating rotating laser scanners, a hyperbolic distortion in the point cloud is often observed due to the polar coordinate mapping inherent to the rotation of the laser beam. To counter this, a hyperbolic mapping equation D.1 is applied to the depth image to approximate the distortion, resulting in a pixel mapping that closely resembles real LiDAR data.

$$h = \sqrt{a^2 + \frac{x^2 \cdot a^2}{b^2}} [\text{Wolf 20}] \quad (\text{D.1})$$

An example of such a simulation can be compared with the Ouster OS0-128 laser scanner, which uses 128 laser lines. The simulated point cloud demonstrates a high degree of similarity to actual point cloud data. Nevertheless, it is noted that data points proximal to the scanner exhibit a triangular pattern, an artifact of rendering through multiple camera frustums. This pattern can be mitigated, if necessary, by employing ray-tracing techniques or by vertically oversampling the depth images to produce a more uniformly distributed point cloud.

## E. Deep Learning Architectures

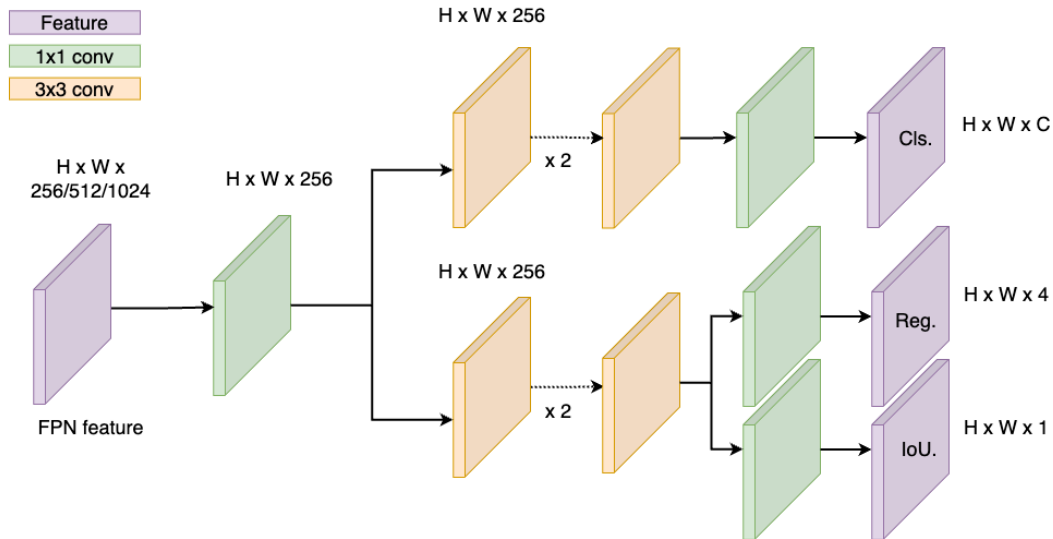
Throughout this thesis, the SWiM architecture has thoroughly examined and integrated several deep-learning based networks for object detection (Section 6.2.1), and semantic segmentation (Section 6.2.1). This appendix explains each in detail to help readers understand these architectures, drawing on relevant literature to clarify their theoretical foundations and practical applications. The discussion covers the subtle architectural differences, intricate algorithms, and contextual use of these deep learning paradigms. These algorithms have been implemented and rigorously tested using surface and underwater sensory datasets (explained in Section 4.1.1). The thesis aims to bridge the gap between complex computational models and their real-world applications by providing in-depth explanations and contextual evaluations. It allows readers to comprehend technological advancements and their implications within the broader context of computer vision and marine exploration.

### Object Detection

To detect the surface and underwater obstacles, deep learning based object detection networks may prove helpful. It is worth noting that multi-class object recognition and classification is not possible in post-flood environment due to random and unstructured obstacles. Therefore, SWiM architecture assumes only one class i.e. *obstacle*. Furthermore, the deep-learning based object detection algorithms can be categorized into three types: single-stage detectors, two-stage detectors, and transformer-based detectors. Single-stage detectors are optimized for speed and efficiency and perform object detection and classification in a single pass through the network. Two-stage detectors generate region proposals first and then classify those regions, resulting in higher accuracy but slower computational speed. Transformer-based detectors can handle varying object sizes and relationships without requiring region proposal mechanisms. Each category has its strengths and trade-offs. Following are the three algorithms (one from each type) implemented in this thesis for object detection in SWiM architecture.

## YOLOX

Zheng Ge et al. [Ge 21] introduced YOLOX, an improved version of the YOLO series, offering advancements in speed and accuracy. YOLOX adopts an anchor-free framework, eliminating the need for preset anchor boxes for object localization. It employs a center-based detection mechanism, predicting the central point, width, and height of objects directly as shown in Figure E.1. This simplifies detection and reduces hyperparameters, streamlining training. YOLOX features a CSPDarknet backbone with cross-stage partial connections for effective feature propagation and a spatial pyramid pooling module for feature capture across scales. It includes practical enhancements like dynamic anchor assignment, adapting to object scales and aspect ratios, and supports multi-scale and multi-input testing for robust detection in varied environments. Moreover, the YOLOX offers a variety of models, such as tiny, small, large, and X. These models differ in their image scale and architecture size. This thesis uses the small and large models to compare and analyze the trade-offs presented by each model. Table E.1 provides essential analytics of both models<sup>1</sup>.



**Figure E.1:** Architectural Comparison between YOLOX and YOLO Versions 3 to 5 [Ge 21]. This figure contrasts the YOLOX architecture with earlier YOLO versions, highlighting YOLOX’s decoupled head design which separates classification and regression tasks for improved learning. Unlike YOLOv3-v5’s coupled head that integrates all outputs, YOLOX refines predictions with distinct pathways and an added IoU branch, enhancing detection across different scales without relying on predefined anchors.

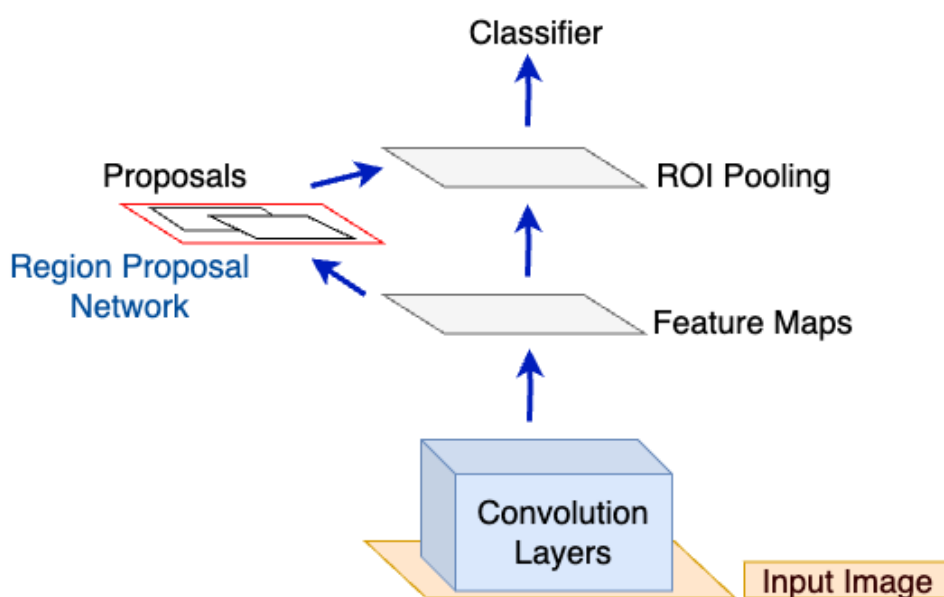
**Table E.1:** Essential analytics of implemented YOLOX architecture.

| Models  | Size | Mem (Gb) | box AP |
|---------|------|----------|--------|
| YOLOX-s | 640  | 7.6      | 40.5   |
| YOLOX-l | 640  | 19.9     | 49.4   |

<sup>1</sup><https://github.com/open-mmlab/mmdetection/tree/main/configs/yolox>

## Faster RCNN

Faster R-CNN, introduced by Ren et al. [Ren 15], enhances its predecessor, R-CNN, by integrating a Region Proposal Network (RPN). This network shares convolutional features with the detection network, streamlining proposal generation. The RPN identifies potential object locations within images, refining them into proposals passed to the detection network, typically Fast R-CNN. Fast R-CNN refines bounding boxes and classifies objects using a Region of Interest (ROI) pooling layer. Faster R-CNN optimizes computational resources by generating proposals directly from CNN features as shown in Figure E.2, reducing time and memory requirements. Although complex, it offers highly accurate object detection, minimizing false positives. However, effective training and extensive datasets are needed. Table E.2 provides key analytics of the architecture<sup>2</sup>.



**Figure E.2:** Architectural Overview of the Faster R-CNN Framework [Ren 15]. The Faster R-CNN model processes images through convolutional layers to create feature maps, uses a Region Proposal Network to identify potential object regions, and then applies ROI pooling and classification to detect and categorize objects within the image. This hierarchical process facilitates a streamlined flow from raw image data to precise object localization and classification.

**Table E.2:** Essential analytics of implemented Faster RCNN architecture.

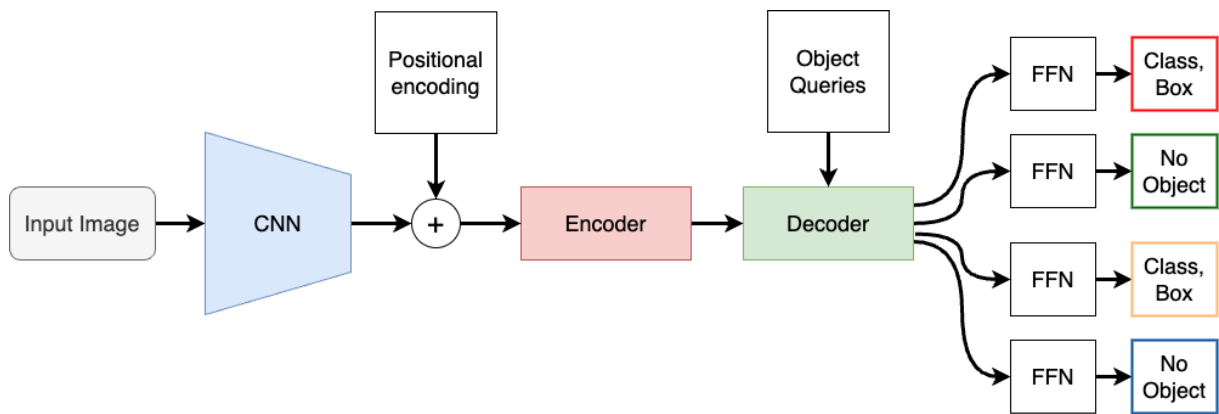
| Backbone | Lr schd | Mem (Gb) | Int time (fps) | box AP |
|----------|---------|----------|----------------|--------|
| R-50-FPN | 1x      | 4        | 21.4           | 37.4   |

## Detr

The DETR (DEtection TRansformer) algorithm, introduced by Carion et al. [Carion 20], revolutionizes object detection by leveraging the transformer architecture, renowned for its success in natural language processing (NLP). Unlike traditional methods, DETR processes

<sup>2</sup>[https://github.com/open-mmlab/mmdetection/tree/main/configs/faster\\_cnn](https://github.com/open-mmlab/mmdetection/tree/main/configs/faster_cnn)

entire images using self-attention mechanisms, enabling it to capture dependencies and interactions across the image. By predicting object locations and categories simultaneously in a single forward pass, DETR eliminates the need for region proposal networks (RPNs), streamlining the detection process. Figure E.3 shows the detailed illustration of DETR algorithm. Its innovative use of positional encodings and transformer-based backbone provides spatial awareness and enables direct modeling of object interactions, making it proficient in complex scenarios like occlusion. However, DETR's richness requires substantial training data and entails complexities in implementation and training. Nonetheless, it represents a significant advancement in object detection, offering an efficient alternative to traditional methods. Table E.3 provides essential analytics<sup>3</sup> for the implemented architecture.



**Figure E.3:** DETR Architecture for Object Detection [Carion 20]. The DETR algorithm employs a CNN to extract image features and positional encodings, which are then processed by a transformer encoder and decoder, with object queries to directly predict object classifications and bounding boxes. This end-to-end architecture simplifies the object detection workflow, as demonstrated by the example output where objects are accurately identified and localized.

**Table E.3:** Essential analytics of implemented DETR architecture.

| Backbone | Lr schd | Mem (Gb) | box AP |
|----------|---------|----------|--------|
| R-50     | 150e    | 7.9      | 39.9   |

## Semantic Segmentation

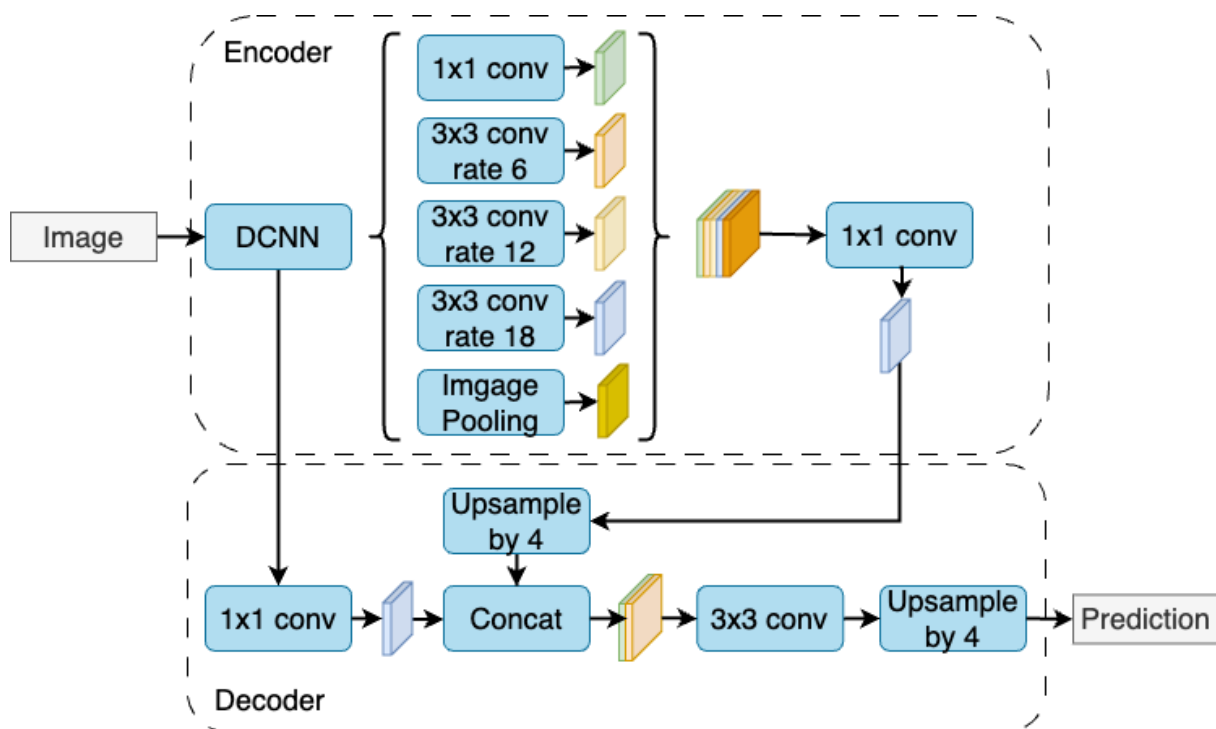
For accurate traversability mapping, the SWiM architecture implements semantic segmentation for surface water environment to segment water from remaining environment. The architecture assumes that the obstacle floating over the water is more critical as compared to the boundaries of water body. The semantic segmentation aims to assign a categorical label to every pixel in an image, effectively enabling a comprehensive understanding of the scene at a granular level. This complex task not only detect the presence of objects within an image but also defines their precise boundaries, offering a pixel-wise classification. This thesis implements three leading networks including DeepLabv3+, UNet, and SegFormer, each representing a unique approach to tackling the challenges inherent in semantic segmentation.

<sup>3</sup><https://github.com/open-mmlab/mmdetection/tree/main/configs/detr>



## DeepLabv3+

The DeepLabv3+ architecture [Chen 18] represents a significant advancement in semantic segmentation, leveraging atrous convolutions to enhance context and scale awareness. By systematically controlling the resolution of feature responses, DeepLabv3+ (shown in Figure E.4) effectively enlarges the field of view of filters without increasing computational burden. This allows for precise segmentation of objects within an image, particularly at boundaries where spatial accuracy is critical. The atrous spatial pyramid pooling (ASPP) module further enhances multi-scale information capture, enabling robust segmentation across diverse object sizes and resolutions. Additionally, the encoder-decoder structure refines segmentation results by leveraging deep semantic features from the encoder and low-level features from earlier layers. The depthwise separable convolution operations within the encoder module contribute to a rich representation of the input image, while the decoder module utilizes this representation to predict pixel-level segmentation. This combination ensures sharp object delineation and high performance across varied image contexts. Table E.4 provides essential analytics regarding the DeepLabv3+ architecture's implementation<sup>4</sup> in this thesis.



**Figure E.4:** DeepLabv3+ Architecture for Semantic Segmentation [Chen 18]. The framework processes an input image through an encoder with atrous convolutions and an image pooling layer to extract multi-scale features, which are then refined and combined with low-level features in the decoder. The result is an upsampled prediction with precise object segmentation, as seen with the example of a segmented cat.

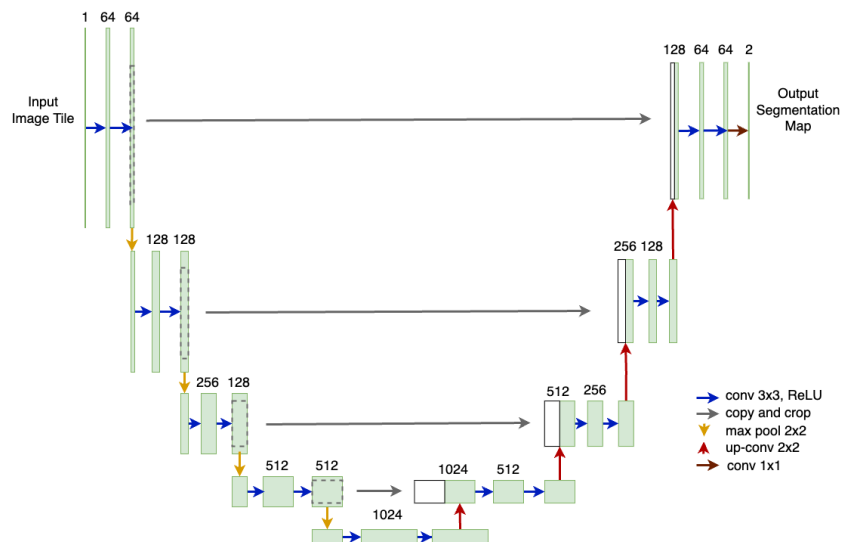
<sup>4</sup><https://github.com/open-mmlab/msegmentation/tree/main/configs/deeplabv3plus>

**Table E.4:** Essential analytics of implemented DeepLabv3+ architecture.

| Backbone | Crop Size        | Lr schd |
|----------|------------------|---------|
| R-50-D8  | $512 \times 512$ | 160k    |

## U-Net

U-Net [Ronneberger 15], named for its U-shaped architecture, is a powerful tool in medical imaging, providing pixel-wise classification for precise localization. Its design incorporates two main paths: a contracting path for capturing contextual information and an expanding path for mapping high-level features back onto the pixel space. The contracting path employs convolutional layers and max pooling for down-sampling, while the expanding path uses up-convolutions and concatenates features from the contracting path to preserve spatial information. Crucially, U-Net includes skip connections between these paths shown in Figure E.5, ensuring that high-resolution features are passed to the up-sampling path. This design choice allows U-Net to retain important spatial details lost during pooling operations. Moreover, U-Net’s ability to generalize well with minimal training samples makes it particularly suited for medical imaging tasks where annotated data is scarce. Beyond medical imaging, U-Net’s adaptable architecture has found applications in diverse fields such as satellite image analysis and autonomous vehicle navigation, where precise segmentation is crucial. Its efficient use of computational resources and robust performance in data-limited scenarios have contributed to its widespread adoption across various domains. Table E.5 provides essential analytics<sup>5</sup> for U-Net’s implementation in this thesis.



**Figure E.5:** U-Net Architecture for Semantic Segmentation [Ronneberger 15]. The U-Net architecture depicted processes an input image through a series of convolutional and pooling layers, followed by up-convolutions and concatenations that integrate fine-grained and abstract features for precise pixel-wise image segmentation.

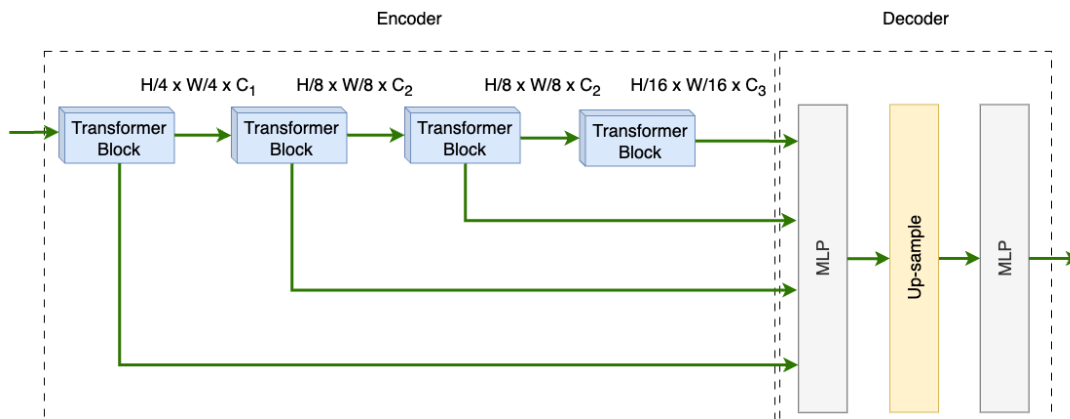
<sup>5</sup><https://github.com/open-mmlab/mmssegmentation/tree/main/configs/unet>

**Table E.5:** Essential analytics of implemented Unet architecture.

| Method      | Backbone    | Loss          | Crop Size        | Lr schd |
|-------------|-------------|---------------|------------------|---------|
| U-Net + FCN | UNet-S5-D16 | Cross Entropy | $256 \times 256$ | 160k    |

## Segformer

SegFormer is a technique for semantic segmentation that leverages transformers to tackle complex image segmentation tasks. Unlike traditional convolutional neural networks (CNNs), SegFormer excels at capturing fine details and global context crucial for accurate pixel-level image understanding. At its core lies a transformer encoder employing self-attention mechanisms to model long-range dependencies across the entire image, allowing for a holistic scene understanding as illustrated in Figure E.6. The hierarchical architecture efficiently integrates features across different scales, generating a multi-scale feature representation progressively refined by the decoder through upsampling and merging. This approach preserves high-resolution details while effectively combining them with broader context, making it particularly effective in scenarios with varying object sizes and fine-grained segmentation tasks. SegFormer also introduces a novel positional encoding scheme and a lightweight all-MLP decoder, providing crucial spatial information and precise object boundaries. Its design facilitates end-to-end training, simplifying the training and deployment process compared to other semantic segmentation frameworks. SegFormer’s scalability and adaptability enable seamless integration into various image resolutions and segmentation challenges. Table E.6 offers essential analytics<sup>6</sup> for implementing this architecture in this thesis.



**Figure E.6:** Segformer Architecture for Semantic Segmentation [Xie 21]. The architecture integrates transformer blocks with overlap patch embeddings in the encoder, and a decoder that merges multi-scale features for detailed semantic segmentation, computes in class-specific high-resolution prediction maps.

**Table E.6:** Essential analytics of implemented Segformer architecture.

| Backbone | Crop Size | Lr schd |
|----------|-----------|---------|
| MIT-B0   | 512       | 160k    |

<sup>6</sup><https://github.com/open-mmlab/mmssegmentation/tree/main/configs/segformer>



# List of Figures

|      |  |    |
|------|--|----|
| 1.1  | Post-flood Images from a flood in Pakistan in 2022 . . . . .   | 3  |
| 2.1  | KRISO Surface Water Vehicle [Jung 22] . . . . .  | 11 |
| 2.2  | Collision Avoidance Framework in KRISO [Choi 20b] . . . . .  | 12 |
| 2.3  | Water Strider during the competition final in Kalamata, Greece . . . . .   | 13 |
| 2.4  | Dataflow and Software Architecture of Water Strider. [Zube 22] . . . . .   | 14 |
| 2.5  | ROBOAT - Quarter Scale Prototype [Wang 19a] . . . . .  | 14 |
| 2.6  | ROBOAT - Half Scale Prototype [Wang 20b] . . . . .   | 15 |
| 2.7  | ROBOAT - Full Scale Prototype [Wang 20b] . . . . .   | 16 |
| 2.8  | Autonomy framework of Roboat II [Wang 20b] . . . . .   | 16 |
| 2.9  | JingHai-I surface water vehicle [Li 22] . . . . .  | 17 |
| 2.10 | System Architecture of JingHai-I [Peng 17] . . . . .   | 18 |
| 2.11 | Charlie testing in Genova harbor. [Caccia 07] . . . . .  | 19 |
| 2.12 | Hierarchical control architecture of Charlie [Caccia 09] . . . . .   | 20 |
| 3.1  | Block Diagram of the Shallow Water Traversability Mapping (SWiM) Architecture for Traversability Analysis. . . . . | 27 |
| 4.1  | Example Data from MODD Dataset [Kristan 15]. . . . .   | 35 |
| 4.2  | Example Data from Mastr1325 Dataset [Bovcon 19]. . . . .   | 36 |
| 4.3  | Example Data from Rosebud Dataset [Lambert 22]. . . . .  | 37 |
| 4.4  | Example Data from USVInland Dataset [Cheng 21]. . . . .  | 38 |
| 4.5  | Example Data from Marine Debris Dataset [Singh 21]. . . . .  | 39 |
| 4.6  | Example Data from UATD [Xie 22]. . . . .   | 40 |
| 4.7  | Example Data from MASTER. . . . .  | 42 |
| 4.8  | Example Segmentation Data from MASTER. . . . .   | 43 |
| 4.9  | Water Plugin - Unreal Engine . . . . .   | 44 |
| 4.10 | Representation of Visibility in a Simulated Underwater Environment. . . . .  | 45 |
| 4.11 | Flooded environments simulated in Unreal Engine . . . . .  | 46 |

|      |  |    |
|------|--|----|
| 4.12 | Surface vehicle models designed in UNREAL Engine . . . . .   | 47 |
| 4.13 | Comparative visualization between real and simulated sonar imagery. . . . .  | 49 |
| 4.14 | Architecture of a Generative Adversarial Network (GAN) . . . . .   | 50 |
| 4.15 | Architecture of a Conditional Generative Adversarial Network (cGAN) . . . . .  | 51 |
| 4.16 | Architecture of a Pix2Pix Conditional Generative Adversarial Network (cGAN) . . . . .                                  | 51 |
| 4.17 | Effect of Augmentation Techniques over Synthetic Images generated using Pix2Pix cGAN . . . . .                         | 53 |
| 4.18 | Implementation of Selective Augmentation Techniques for Enhanced Pix2Pix cGAN Performance. . . . .                     | 54 |
| 4.19 | Generated Imagery from Marine Debris Dataset at different field sizes . . . . .  | 56 |
| 4.20 | Generated Imagery from Marine Debris Dataset at different field sizes . . . . .  | 57 |
| 4.21 | Comparison of Different GANs using YoloX-s, YoloX-l, and Faster RCNN Detection Model on Marine Debris Dataset. . . . . | 61 |
| 4.22 | Comparative Object Detection on UATD and UATD+MASTER Datasets. . . . .   | 64 |
| 4.23 | Comparison of Different GANs using YoloX-s, YoloX-l, and Faster RCNN Detection Model on UATD Dataset. . . . .          | 68 |
| 4.24 | Simulation Data - Scenario 1 . . . . .   | 70 |
| 4.25 | Simulation Data - Scenario 2 . . . . .   | 71 |
| 4.26 | Simulation Data - Scenario 3 . . . . .   | 72 |
| 5.1  | Detection of cargo vessel using Simrad SX90 . . . . .  | 77 |
| 5.2  | Dual Frequency SeaKing-Bottom profiler . . . . .   | 77 |
| 5.3  | Multibeam Imaging sonars with different operating frequencies and beamwidths . . . . .                                 | 78 |
| 5.4  | Speckle Noise in real sonar imagery . . . . .  | 79 |
| 5.5  | Comparison of sonar imagery at different waterbed depth level . . . . .  | 80 |
| 5.6  | Comparison of sonar imagery at different gain level . . . . .  | 81 |
| 5.7  | Physical principle of sonar 2D imagery. . . . .  | 82 |
| 5.8  | Workflow of the Underwater Traversability Mapping (UTM) Framework. . . . .   | 83 |
| 5.9  | Sequential Procedure for Enhancing Underwater Imagery. . . . .   | 87 |
| 5.10 | Background Estimation Technique Applied to Sonar Imagery. . . . .  | 87 |
| 5.11 | Discrete wavelength decomposition of sonar imagery . . . . .   | 88 |
| 5.12 | Comparative Analysis of Filtering Techniques on Sonar Imagery for Enhancement (1). . . . .                             | 89 |
| 5.13 | Comparative Analysis of Filtering Techniques on Sonar Imagery for Enhancement (2). . . . .                             | 90 |
| 5.14 | Sonar Imaging for Seabed Analysis. . . . .   | 92 |

|      |  |     |
|------|--|-----|
| 5.15 | Obstacle Detection Flow Diagram . . . . .  | 94  |
| 5.16 | Three-Step Morphological Processing for Obstacle Detection in Sonar Imagery.                                       | 95  |
| 5.17 | Two-Step Canny Edge and Contour-Based Object Detection in Sonar Imagery.   | 96  |
| 5.18 | Illustration of Field of View and Distance Impact on Sonar Resolution. . .   | 97  |
| 5.19 | Application of Object Tracking in Sonar Image Processing for Obstacle<br>Detection. . . . .                        | 100 |
| 5.20 | Underwater Mapping Results - Scenario 1. . . . .   | 101 |
| 5.21 | Underwater Mapping Results - Scenario 2. . . . .   | 102 |
| 5.22 | Underwater Mapping Results - Scenario 3. . . . .   | 102 |
|      |  |     |
| 6.1  | Post-flood destruction captured in Satellite and Aerial Imagery . . . . .  | 106 |
| 6.2  | Debris due to post-flood environment. . . . .  | 107 |
| 6.3  | Example data from TUK Campus Dataset. [Keen 21] . . . . .  | 109 |
| 6.4  | Surface Water Traversability Mapping (SWTM) Framework - Block Diagram  | 111 |
| 6.5  | Example data from MASTER Dataset. . . . .  | 112 |
| 6.6  | Exposure Correction in Underexposed Image. . . . .   | 114 |
| 6.7  | Exposure Correction in Overexposed Image. . . . .  | 115 |
| 6.8  | Qualitative comparison of object detection algorithms on the MODD and<br>MODD2 dataset. . . . .                    | 119 |
| 6.9  | Visual Comparison of Semantic Segmentation Models . . . . .  | 123 |
| 6.10 | Effect of Surface water environment on LiDAR pointcloud . . . . .  | 124 |
| 6.11 | Results of Three-step Pointcloud Filtering. . . . .  | 125 |
| 6.12 | Point cloud Mapping over Camera Imagery. . . . .   | 127 |
| 6.13 | Combining the Visual and Spatial Sensory Modalities. . . . .   | 128 |
| 6.14 | Simulation Result - Elevation Map at each frame . . . . .  | 130 |
| 6.15 | Simulation Result - Complete Elevation Map . . . . .   | 131 |
| 6.16 | Comparative presentation of natural water bodies with their respective<br>Google Maps and Elevation Maps . . . . . | 132 |
|      |  |     |
| 7.1  | Schematic of the Fused Traversability Mapping (FTM) Framework. . . . .   | 136 |
| 7.2  | Surface and Underwater Detection in Gelterswoog Lake. . . . .  | 139 |
| 7.3  | Fused Threat Map of Hamburg Port site. . . . .   | 140 |
| 7.4  | Traversability map of Hamburg Port. . . . .  | 141 |
| 7.5  | Traversability map of Simulated Barnyard Environment. . . . .  | 142 |
|      |  |     |
| A.1  | ZED stereo camera attached on ASTRIDER. . . . .  | 148 |
| A.2  | Ouster OS0-128 3D LiDAR attached on ASTRIDER. . . . .  | 150 |

---

|     |   |     |
|-----|---|-----|
| A.3 | Tritech’s Gemini 720ik attached on ASTRIDER. . . . .  | 152 |
| A.4 | A dual-antenna Trimble GNSS system with integrated inertial navigation system powered by the BD992-INS. . . . . | 154 |
| B.1 | The prototypical ASTRIDER equipped with the autonomy box. . . . .   | 158 |
| B.2 | The BASILISK equipped with autonomy kit and the sensor array of ASTRIDER. [Meckel 22] . . . . .                 | 159 |
| B.3 | The PONTONBOOT hardware kit is equipped with an additional touchscreen. [Meckel 22] . . . . .                   | 160 |
| C.1 | FIGUI data processing visualization of PONTONBOOT Project . . . . .   | 163 |
| C.2 | FINSTRUCT component graph visualization of PONTONBOOT Project . . . . .   | 164 |
| C.3 | Common hardware interface developed to communicate with control architecture. . . . .                           | 165 |
| D.1 | Simulation Scene with Underwater Scenario . . . . .   | 168 |
| D.2 | Unreal Engine to FINROC Interface . . . . .   | 168 |
| D.3 | Simulated Satellite Mesh in Unreal Engine. . . . .  | 170 |
| D.4 | Simulated Camera Image from Flooded Barnyard Environment. . . . .   | 170 |
| D.5 | Simulated LiDAR Point Cloud from Flooded Barnyard Environment. . . . .  | 171 |
| E.1 | Architectural Comparison between YOLOX and YOLO Versions 3 to 5 [Ge 21]. . . . .                                | 174 |
| E.2 | Architectural Overview of the Faster R-CNN Framework [Ren 15]. . . . .  | 175 |
| E.3 | DETR Architecture for Object Detection [Carion 20]. . . . .   | 176 |
| E.4 | DeepLabv3+ Architecture for Semantic Segmentation [Chen 18]. . . . .  | 177 |
| E.5 | U-Net Architecture for Semantic Segmentation [Ronneberger 15]. . . . .  | 178 |
| E.6 | Segformer Architecture for Semantic Segmentation [Xie 21]. . . . .  | 179 |



# List of Tables

|     |  |     |
|-----|--|-----|
| 2.1 | Existing surface water vehicles and their essential characteristics. . . . .   | 21  |
| 4.1 | Threshold Metrics for Image Selection in Video Sequences. . . . .  | 52  |
| 4.2 | Evaluation Metrics for Generated Images Across Different Networks and Datasets. . . . .  | 55  |
| 4.3 | Indirect Evaluation of Generated Images in Marine Debris Dataset. . . . .  | 59  |
| 4.4 | Performance Metrics of Object Detection Algorithms on Different Datasets.  | 63  |
| 4.5 | Indirect Evaluation of Generated Images in UATD Dataset. . . . .   | 66  |
| 5.1 | Range vs Height Uncertainty of obstacle. . . . .   | 82  |
| 5.2 | Performance metrics of object detection using various filtering techniques on sonar imagery. . . . .                                   | 91  |
| 5.3 | Range vs Detection Accuracy . . . . .  | 98  |
| 6.1 | Threshold Metrics for Image Selection in Video Sequences. . . . .  | 117 |
| 6.2 | Object Detection Model Hyper-parameter Specifications. . . . .   | 118 |
| 6.3 | Comparative analysis of object detection techniques across various datasets  | 118 |
| 6.4 | Semantic Segmentation Model Hyper-parameter Specifications. . . . .  | 120 |
| 6.5 | Performance comparison of Semantic Segmentation models over Aerial Fluvial Dataset . . . . .   | 121 |
| 6.6 | Performance comparison of Semantic Segmentation models over Mastr1325 Dataset . . . . .  | 121 |
| 6.7 | Performance comparison of Semantic Segmentation models over Rosebud Dataset . . . . .  | 121 |
| 6.8 | Performance comparison of Semantic Segmentation models over UvInland Dataset . . . . .   | 122 |
| 6.9 | Reduction in Point Cloud Data Across Processing Stages. . . . .  | 126 |
| A.1 | Specifications of ZED2i camera (source: <a href="https://www.stereolabs.com/zed-2i/">https://www.stereolabs.com/zed-2i/</a> ). . . . . | 148 |
| A.2 | Specifications of Ouster OS0 (source: <a href="https://ouster.com/downloads/">https://ouster.com/downloads/</a> ). . . . .             | 150 |

---

|     |  |     |
|-----|--|-----|
| A.3 | Specifications of Tritech’s Gemini 720ik sonar (source: <a href="https://www.tritech.co.uk/products/gemini-720ik">https://www.tritech.co.uk/products/gemini-720ik</a> ). . . . . | 152 |
| A.4 | Specifications of Trimble’s BX992 Receiver (source: <a href="https://info.intech.trimble.com/bx992-datasheet">https://info.intech.trimble.com/bx992-datasheet</a> ). . . . .     | 154 |
| E.1 | Essential analytics of implemented YOLOX architecture. . . . .   | 174 |
| E.2 | Essential analytics of implemented Faster RCNN architecture. . . . .   | 175 |
| E.3 | Essential analytics of implemented DETR architecture. . . . .  | 176 |
| E.4 | Essential analytics of implemented DeepLabv3+ architecture. . . . .  | 178 |
| E.5 | Essential analytics of implemented Unet architecture. . . . .  | 179 |
| E.6 | Essential analytics of implemented Segformer architecture. . . . .   | 179 |

# Bibliography

- [BMW 21] “Neues Ponton-Boot-System für den Katastropheneinsatz vor dem Einsatz – gefördert durch das Bundeswirtschaftsministerium”, <https://www.bmwi.de/Redaktion/DE/Pressemitteilungen/2021/11/20211130-neues-ponton-boot-system-fuer-katastropheneinsatz.html>, 11 2021. last accessed 2022/02/01.
- [Bovcon 18] B. Bovcon, R. Mandeljc, J. Pers, M. Kristan, “Stereo obstacle detection for unmanned surface vehicles by IMU-assisted semantic segmentation”, *Robotics Auton. Syst.*, vol. 104, pp. 1–13, 2018.
- [Bovcon 19] B. Bovcon, J. Muhovi, J. Pers, M. Kristan, “The MaSTr1325 dataset for training deep USV obstacle detection models”, *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3431–3438, 2019.
- [Buades 05] A. Buades, B. Coll, J. M. Morel, “A non-local algorithm for image denoising”, in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, vol. 2. 2005, pp. 60–65 vol. 2.
- [Caccia 07] M. Caccia, M. Bibuli, R. Bono, G. Bruzzone, G. Bruzzone, E. Spirandelli, “Unmanned surface vehicle for coastal and protected waters applications: The Charlie project”, *Marine Technology Society Journal*, vol. 41, no. 2, pp. 62–71, 2007.
- [Caccia 09] M. Caccia, M. Bibuli, G. Bruzzone, G. Bruzzone, R. Bono, E. Spirandelli, “Charlie, a testbed for usv research”, *IFAC Proceedings Volumes*, vol. 42, no. 18, pp. 97–102, 2009.
- [Caimi 08] F. M. Caimi, D. M. Kocak, F. Dalglish, J. Watson, “Underwater imaging and optics: Recent advances”, in *OCEANS 2008*, vol. 2008-Supplement. 2008, pp. 1–9.
- [Canny 86] J. Canny, “A Computational Approach to Edge Detection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.
- [Carion 20] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, “End-to-end object detection with transformers”, in *European conference on computer vision*, Springer. 2020, pp. 213–229.
- [Cerqueira 17] R. Cerqueira, T. Trocoli, G. Neves, S. Joyeux, J. Albiez, L. Oliveira, “A novel GPU-based sonar simulator for real-time applications”, *Computers Graphics*, vol. 68, pp. 66–76, 2017.
- [Chaillan 07] F. Chaillan, C. Fraschini, P. Courmontagne, “Speckle noise reduction in SAS imagery”, *Signal Processing*, vol. 87, no. 4, pp. 762–781, 2007.

- [Chen 18] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation”, in *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 801–818.
- [Chen 19] W. Chen, K. Gu, W. Lin, Z. Xia, P. Le Callet, E. Cheng, “Reference-Free Quality Assessment of Sonar Images via Contour Degradation Measurement”, *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5336–5351, 2019.
- [Chen 21] W. Chen, K. Gu, T. Zhao, G. Jiang, P. L. Callet, “Semi-Reference Sonar Image Quality Assessment Based on Task and Visual Perception”, *IEEE Transactions on Multimedia*, vol. 23, pp. 1008–1020, 2021.
- [Cheng 21] Y. Cheng, M. Jiang, J. Zhu, Y. Liu, “Are We Ready for Unmanned Surface Vehicles in Inland Waterways? The USVInland Multisensor Dataset and Benchmark”, *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3964–3970, 2021.
- [Choi 20a] H. Choi, J. Jeong, “Despeckling Algorithm for Removing Speckle Noise from Ultrasound Images”, *Symmetry*, vol. 12, no. 6, 2020.
- [Choi 20b] J. Choi, J. Park, J. Jung, Y. Lee, H.-T. Choi, “Development of an autonomous surface vehicle and performance evaluation of autonomous navigation technologies”, *International Journal of Control, Automation and Systems*, vol. 18, pp. 535–545, 2020.
- [Choi 21] W.-S. Choi, D. R. Olson, D. Davis, M. Zhang, A. Racson, B. Bingham, M. McCarrin, C. Vogt, J. Herman, “Physics-based modelling and simulation of multibeam echosounder perception for autonomous underwater manipulation”, *Frontiers in Robotics and AI*, vol. 8, p. 706646, 2021.
- [DeMarco 15] K. J. DeMarco, M. E. West, A. M. Howard, “A computationally-efficient 2D imaging sonar model for underwater robotics simulations in Gazebo”, in *OCEANS 2015 - MTS/IEEE Washington*. 2015, pp. 1–7.
- [din 17] *Underwater image dehaze using scene depth estimation with adaptive color correction*. 2017.
- [Dube 16] R. Dube, A. Gawel, C. Cadena, R. Siegwart, L. Freda, M. Gianni, “3D localization, mapping and path planning for search and rescue operations”, in *2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. 10 2016, pp. 272–273.
- [Dudgeon 84] D. E. Dudgeon, R. M. Mersereau, *Multidimensional digital signal processing*, Prentice-hall, 1984.
- [Frost 82] V. S. Frost, J. A. Stiles, K. S. Shanmugan, J. C. Holtzman, “A model for radar images and its application to adaptive digital filtering of multiplicative noise”, *IEEE Transactions on pattern analysis and machine intelligence*, no. 2, pp. 157–166, 1982.
- [Galceran 12] E. Galceran, V. Djapic, M. Carreras, D. P. Williams, “A Real-time Underwater Object Detection Algorithm for Multi-beam Forward Looking Sonar”, *IFAC Proceedings Volumes*, vol. 45, no. 5, pp. 306–311, 2012. 3rd IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles.

- [Ge 21] Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun, “YOLOX: Exceeding YOLO Series in 2021”, *ArXiv*, vol. abs/2107.08430, 2021.
- [Goodfellow 20] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, “Generative adversarial networks”, *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [Guha-Sapir 04] D. Guha-Sapir, D. Hargitt, P. Hoyois, *Thirty years of natural disasters 1974-2003: The numbers*, Presses univ. de Louvain, 2004.
- [Haddad 91] R. A. Haddad, A. N. Akansu et al, “A class of fast Gaussian binomial filters for speech and image processing”, *IEEE Transactions on Signal Processing*, vol. 39, no. 3, pp. 723–727, 1991.
- [Han 19] J. Han, Y. Cho, J. Kim, “Coastal SLAM With Marine Radar for USV Operation in GPS-Restricted Situations”, *IEEE Journal of Oceanic Engineering*, vol. 44, no. 2, pp. 300–309, 2019.
- [Henson 19] B. T. Henson, Y. V. Zakharov, “Attitude-Trajectory Estimation for Forward-Looking Multibeam Sonar Based on Acoustic Image Registration”, *IEEE Journal of Oceanic Engineering*, vol. 44, no. 3, pp. 753–766, 2019.
- [Hirschmuller 08] H. Hirschmuller, “Stereo Processing by Semiglobal Matching and Mutual Information”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328–341, 2008.
- [Huang 79] T. Huang, G. Yang, G. Tang, “A fast two-dimensional median filtering algorithm”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 1, pp. 13–18, 1979.
- [Huntsberger 11] T. Huntsberger, H. Aghazarian, A. Howard, D. C. Trotz, “Stereo vision-based navigation for autonomous surface vessels”, *Journal of Field Robotics*, vol. 28, no. 1, pp. 3–18, 2011.
- [Hurtós Vilarnau 14] N. Hurtós Vilarnau et al, “Forward-looking sonar mosaicing for underwater environments”, *Doctoral Dissertation*, 2014.
- [Jain 15] S. Jain, S. Nuske, A. Chambers, L. Yoder, H. Cover, L. Chamberlain, S. Scherer, S. Singh, *Autonomous River Exploration*, Cham: Springer International Publishing, 2015, pp. 93–106.
- [Jia 19] Y. Jia, X. Ye, S. Guo, H. Yang, “A Piecewise Nonlinear Enhancement Method of Side Scan Sonar Images”, in *OCEANS 2019 - Marseille*. 2019, pp. 1–6.
- [Jung 19] J. Jung, J. Park, J. Choi, H.-T. Choi, “Terrain Based Navigation for an Autonomous Surface Vehicle with a Multibeam Sonar”, in *OCEANS 2019 - Marseille*. 2019, pp. 1–4.
- [Jung 22] J. Jung, J. Park, J. Choi, H.-T. Choi, “Bathymetric Pose Graph Optimization With Regularized Submap Matching”, *IEEE Access*, vol. 10, pp. 1–1, 01 2022.

- [Karoui 15] I. Karoui, I. Quidu, M. Legris, “Automatic Sea-Surface Obstacle Detection and Tracking in Forward-Looking Sonar Image Sequences”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 8, pp. 4661–4669, 2015.
- [Keen 20] H. E. Keen, K. Berns, “Generation of elevation maps for planning and navigation of vehicles in rough natural terrain”, in *Advances in Service and Industrial Robotics: Proceedings of the 28th International Conference on Robotics in Alpe-Adria-Danube Region (RAAD) 28*, Springer. 2020, pp. 488–495.
- [Keen 21] H. Keen, Q. Hamza Jan, K. Berns, “Drive on Pedestrian Walk. TUK Campus Dataset”, in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2021, pp. 3822–3828.
- [Keen 23] H. E. Keen, K. Berns, “Probabilistic Fusion of Surface and Underwater Maps in a Shallow Water Environment”, in *Advances in Service and Industrial Robotics: Proceedings of the 28th International Conference on Robotics in Alpe-Adria-Danube Region (RAAD)*, T. Petrič, A. Ude, L. Žlajpah, Eds. Cham: Springer Nature Switzerland, 2023, pp. 195–202.
- [Kelly 19] R. H. Kelly, *Algorithms for planning and executing multi-robot shapeshifting*, 2019.
- [Kristan 15] M. Kristan, V. S. Kenk, S. Kovacic, J. Pers, “Fast Image-Based Obstacle Detection From Unmanned Surface Vehicles”, *IEEE Transactions on Cybernetics*, vol. 46, pp. 641–654, 2015.
- [Kuan 85] D. T. Kuan, A. A. Sawchuk, T. C. Strand, P. Chavel, “Adaptive noise smoothing filter for images with signal-dependent noise”, *IEEE transactions on pattern analysis and machine intelligence*, no. 2, pp. 165–177, 1985.
- [Lambert 22] R. Lambert, J. Chavez-Galaviz, J. Li, N. Mahmoudian, “ROSEBUD: A Deep Fluvial Segmentation Dataset for Monocular Vision-Based River Navigation and Obstacle Avoidance”, *Sensors*, vol. 22, no. 13, 2022.
- [Lee 81] J.-S. Lee, “Refined filtering of image noise using local statistics”, *Computer graphics and image processing*, vol. 15, no. 4, pp. 380–389, 1981.
- [Li 22] A. Li, C. Gong, X. Huang, X. Sun, G. Liu, “Overview of Key Technologies for Water-based Automatic Security Marking Platform”, *Electrical Science Engineering*, vol. 4, p.30, 06 2022.
- [Liu 16] Z. Liu, Y. Zhang, X. Yu, C. Yuan, “Unmanned surface vehicles: An overview of developments and challenges”, *Annual Reviews in Control*, vol. 41, pp. 71–93, 2016.
- [Long 23] H. Long, L. Shen, Z. Wang, J. Chen, “Underwater Forward-Looking Sonar Images Target Detection via Speckle Reduction and Scene Prior”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–13, 2023.
- [Lv 22] H. Lv, D. Kong, G. Pang, B. Wang, Z. Yu, Z. Pang, G. Yang, “GuLiM: A Hybrid Motion Mapping Technique for Teleoperation of Medical Assistive Robot in Combating the COVID-19 Pandemic”, *IEEE Transactions on Medical Robotics and Bionics*, vol. 4, no. 1, pp. 106–117, 2022.

- [Masson-Delmotte 18] V. P. Z. H.-O. P. D. R. J. S. P. R. S. A. P. e. a. G. w. o. . . C. Masson-Delmotte, “An IPCC Special Report on the impacts of global warming of 1 no”, 5, 2018.
- [Meckel 22] D. Meckel, H. Keen, C. Heupel, K. Berns, “Transferring off-road control concepts to watercraft used in flooded areas”, in *Commercial Vehicle Technology 2022*, K. Berns, K. Dressler, R. Kalmar, N. Stephan, R. Teutsch, M. Thul, Eds. Wiesbaden: Springer Fachmedien Wiesbaden, 2022, pp. 121–136.
- [Mogstad 19] A. A. Mogstad, G. Johnsen, M. Ludvigsen, “Shallow-Water Habitat Mapping using Underwater Hyperspectral Imaging from an Unmanned Surface Vehicle: A Pilot Study”, *Remote Sensing*, vol. 11, no. 6, 2019.
- [Odetti 20] A. Odetti, G. Bruzzone, M. Altosole, M. Viviani, M. Caccia, “SWAMP, an Autonomous Surface Vehicle expressly designed for extremely shallow waters”, *Ocean Engineering*, vol. 216, p. 108205, 2020.
- [Oliveira 21] A. J. Oliveira, B. M. Ferreira, N. A. Cruz, “A Performance Analysis of Feature Extraction Algorithms for Acoustic Image-Based Underwater Navigation”, *Journal of Marine Science and Engineering*, vol. 9, no. 4, 2021.
- [Oskard 89] D. N. Oskard, T.-H. Hong, C. A. Shaffer, “Spatial mapping system for autonomous underwater vehicles”, in *Sensor Fusion: Spatial Reasoning and Scene Interpretation*, vol. 1003, SPIE. 1989, pp. 439–450.
- [pen 21] *Unveiling of How Image Restoration Contributes to Underwater Object Detection*. 2021.
- [Peng 17] Y. Peng, Y. Yang, J. Cui, X. Li, H. Pu, J. Gu, S. Xie, J. Luo, “Development of the USV ‘JingHai-I’ and sea trials in the Southern Yellow Sea”, *Ocean Engineering*, vol. 131, pp. 186–196, 2017.
- [Raber 19] G. T. Raber, S. R. Schill, “Reef Rover: A Low-Cost Small Autonomous Unmanned Surface Vehicle (USV) for Mapping and Monitoring Coral Reefs”, *Drones*, vol. 3, no. 2, 2019.
- [Reichardt 12] M. Reichardt, T. Föhst, K. Berns, “Introducing FINROC: A Convenient Real-time Framework for Robotics based on a Systematic Design Approach”. 2012.
- [Reichardt 13] M. Reichardt, T. Föhst, K. Berns, “Quality and Maintainability-Bridging the gap between end user expectations , vendors ’ business prospects , and software engineers ’ requirements on the ground ( SQM 2013 ) On Software Quality-motivated Design of a Real-time Framework for Complex Robot Control Systems”. 2013.
- [Ren 15] S. Ren, K. He, R. Girshick, J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks”, *Advances in neural information processing systems*, vol. 28, 2015.
- [Ronneberger 15] O. Ronneberger, P. Fischer, T. Brox, “U-net: Convolutional networks for biomedical image segmentation”, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, Springer. 2015, pp. 234–241.

- [Ropertz 17] T. Ropertz, P. Wolf, K. Berns, “Quality-Based Behavior-Based Control for Autonomous Robots in Rough Environments.”, in *14th International Conference on Informatics in Control, Automation and Robotics*. 2017.
- [Ruetz 22] F. Ruetz, P. Borges, N. Suenderhauf, E. Hernández, T. Peynot, “Forest Traversability Mapping (FTM): Traversability estimation using 3D voxel-based Normal Distributed Transform to enable forest navigation”, in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2022, pp. 8714–8721.
- [Salhi 12] A. Salhi, A. Y. Jammaoussi, “Object tracking system using Camshift, Mean-shift and Kalman filter”, *International Journal of Electronics and Communication Engineering*, vol. 6, no. 4, pp. 421–426, 2012.
- [Seraji 03] H. Seraji, “New Traversability Indices and Traversability Grid for Integrated Sensor/Map-Based Navigation”, *Journal of Robotic Systems*, vol. 20, no. 3, p. 121–134, Feb 2003.
- [Shafi 20] U. Shafi, R. Mumtaz, N. Iqbal, S. M. H. Zaidi, S. A. R. Zaidi, I. Hussain, Z. Mahmood, “A Multi-Modal Approach for Crop Health Mapping Using Low Altitude Remote Sensing, Internet of Things (IoT) and Machine Learning”, *IEEE Access*, vol. 8, pp. 112708–112724, 2020.
- [Singh 21] D. Singh, M. Valdenegro-Toro, “The Marine Debris Dataset for Forward-Looking Sonar Semantic Segmentation”, in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. 2021, pp. 3734–3742.
- [Sinisterra 17] A. Sinisterra, M. Dhanak, N. Kouvaras, “A USV platform for surface autonomy”, in *OCEANS 2017 - Anchorage*. 2017, pp. 1–8.
- [Teixeira 16] P. V. Teixeira, M. Kaess, F. S. Hover, J. J. Leonard, “Underwater inspection using sonar-based volumetric submaps”, in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2016, pp. 4288–4295.
- [Tomasi 98] C. Tomasi, R. Manduchi, “Bilateral filtering for gray and color images”, pp. 839–846, 1998.
- [tor 16] *Submerged marine debris detection with autonomous underwater vehicles*. 2016.
- [Trucco 06] E. Trucco, A. T. Olmos-Antillon, “Self-Tuning Underwater Image Restoration”, *IEEE Journal of Oceanic Engineering*, vol. 31, no. 2, pp. 511–519, 2006.
- [Viola 01] P. Viola, M. Jones, “Rapid object detection using a boosted cascade of simple features”, vol. 1, pp. I–I, 2001.
- [Wang 19a] W. Wang, B. Gheneti, L. A. Mateos, F. Duarte, C. Ratti, D. Rus, “Roboat: An Autonomous Surface Vehicle for Urban Waterways”, in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2019, pp. 6340–6347.
- [Wang 19b] Y. Wang, S. James, E. K. Stathopoulou, C. Beltrán-González, Y. Konishi, A. Del Bue, “Autonomous 3-D Reconstruction, Mapping, and Exploration of Indoor Environments With a Robotic Arm”, *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3340–3347, 2019.



- [Wang 20a] W. Wang, T. Shan, P. Leoni, D. Fernández-Gutiérrez, D. Meyers, C. Ratti, D. Rus, “Roboat ii: A novel autonomous surface vessel for urban environments”, in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2020, pp. 1740–1747.
- [Wang 20b] W. Wang, T. Shan, P. Leoni, D. Fernández-Gutiérrez, D. Meyers, C. Ratti, D. Rus, “Roboat II: A Novel Autonomous Surface Vessel for Urban Environments”, in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 1740–1747.
- [Wang 92] X. Wang, “On the gradient inverse weighted filter (image processing)”, *IEEE Transactions on Signal Processing*, vol. 40, no. 2, pp. 482–484, 1992.
- [Wei 21] M. Wei, H. Bian, S. Li, F. Zhang, “Mutual Information Re-Registration of Sensitive Region in Forward-Looking Sonar Images Combined With Particle Swarm Optimization Algorithm”, *IEEE Access*, vol. 9, pp. 7064–7075, 2021.
- [Wolf 20] P. Wolf, T. Groll, S. Hemer, K. Berns, “Evolution of Robotic Simulators: Using UE 4 to Enable Real-World Quality Testing of Complex Autonomous Robots in Unstructured Environments”. 07 2020.
- [Wolf 22] P. Wolf, *Cognitive Processing in Behavior-Based Perception of Autonomous Off-Road Vehicles*, 10 2022.
- [Xie 21] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, P. Luo, “SegFormer: Simple and efficient design for semantic segmentation with transformers”, *Advances in Neural Information Processing Systems*, vol. 34, pp. 12077–12090, 2021.
- [Xie 22] K. Xie, J. Yang, K. Qiu, “A Dataset with Multibeam Forward-Looking Sonar for Underwater Object Detection”, *Scientific Data*, vol. 9, no. 1, p. 739, 2022.
- [Yang 18] K. Yang, L. M. Bergasa, E. Romera, J. Wang, K. Wang, E. López, “Perception Framework of Water Hazards Beyond Traversability for Real-World Navigation Assistance Systems”, in *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2018, pp. 186–191.
- [Ye 07] C. Ye, “Navigating a Mobile Robot by a Traversability Field Histogram”, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 2, pp. 361–372, 2007.
- [Yuan 21] F. Yuan, F. Xiao, K. Zhang, Y. Huang, E. Cheng, “Noise reduction for sonar images by statistical analysis and fields of experts”, *Journal of Visual Communication and Image Representation*, vol. 74, p. 102995, 2021.
- [Zaheer 21] M. H. Zaheer, S. A. Mehdi, H. E. Keen, K. Berns, “Detection of Fungus in Gladiolus Fields Using a Quadcopter”, in *Advances in Service and Industrial Robotics: Proceedings of the 28th International Conference on Robotics in Alpe-Adria-Danube Region (RAAD)*, S. Zeghloul, M. A. Laribi, J. Sandoval, Eds. Cham: Springer International Publishing, 2021, pp. 127–134.

- [Zhang 22] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, C. Li, “Underwater Image Enhancement via Minimal Color Loss and Locally Adaptive Contrast Enhancement”, *IEEE Transactions on Image Processing*, vol. 31, pp. 3997–4010, 2022.
- [Zhenpeng 09] H. Zhenpeng, M. Dinesh, Z. Liangjun, “Ttm Terrain traversability mapping for autonomous excavator navigation in unstructured environments”, *ArXiv*, 2109.
- [Zube 22] A. Zube, D. Kleiser, A. Albrecht, P. Woock, T. Emter, B. Ruf, I. Tchouchenkov, A. Buller, B. Wagner, G. Baatar, J. Petereit, “Autonomously mapping shallow water environments under and above the water surface”, *at - Automatisierungstechnik*, vol. 70, no. 5, pp. 482–495, 2022.

# Curriculum vitae

## Education

- 09.2018 – 07.2024 Ph. D. program, Robotics Research Lab, Dep. of Computer Science, University of Kaiserslautern-Landau, Germany.  
Thesis: *Traversability Mapping in Post-Flood Environment*.  
Degree: **Doctor of Engineering (Dr.-Ing.)**
- 09.2014 – 06.2016 Electrical Engineering (Master), Department of Electrical Engineering, Syed Babar Ali School of Science and Engineering, Lahore University of Management Sciences (LUMS) Lahore, Pakistan.  
Thesis: *Conflict Avoidance among multiple Unmanned Aircrafts using Reinforcement Learning*.  
Degree: **Master of Science (M. S.)**  
Specialization: **Control and Automation.**
- 09.2009 – 10.2013 Electrical Engineering (Bachelor), Department of Electrical Engineering, University of Engineering and Technology (UET) Lahore, Pakistan.  
Senior Year Project: *Design and Development of Intelligent Dust Cleaner with Wall Following Technology*.  
Degree: **Bachelor of Science (B. Sc.)**  
Specialization: **Electronics and Communication.**

## Work Experience

- 03.2024 – today Senior Researcher, Research and Innovation, XITASO GmbH, Germany.
- 09.2022 – 02.2024 Research Associate, Robotics Research Lab, Department of Computer Science, University of Kaiserslautern-Landau, Germany.
- 02.2018 – 08.2022 Student Research Assistant, Robotics Research Lab, Department of Computer Science, University of Kaiserslautern-Landau, Germany.
- 06.2016 – 01.2018 Lecturer, Department of Computer Science, University of Central Punjab Lahore, Pakistan.
- 02.2014 – 01.2016 Lab Instructor, Department of Computer Science, University of Central Punjab Lahore, Pakistan.