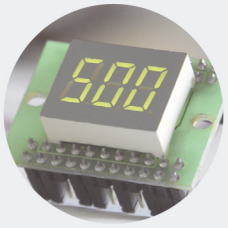
INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO
MESTRADO EM ENGENHARIA ELECTROTÉCNICA E DE COMPUTADORES






# Acoustic caracterization and dection of lost fishing gear

**RUI FILIPE MARTINS DE OLIVEIRA**
outubro de 2024

POLITÉCNICO
DO PORTO

Politécnico do Porto
Instituto Superior de Engenharia do Porto

# Acoustic characterization and detection of lost fishing gear

**Rui Filipe Martins Morais de Oliveira**

Master in Electrical and Computer Engineering
Specialization Area of Automation and Systems


Departamento de Engenharia Eletrotécnica
Instituto Superior de Engenharia do Porto

October 28, 2024

*This dissertation partially satisfies the requirements of the Thesis/Dissertation course of the program Master in Electrical and Computer Engineering, Specialization Area of Automation and Systems.*

**Candidate:** Rui Filipe Martins Morais de Oliveira, No. 1210180, 12101807@isep.ipp.pt

**Scientific Guidance:** Eng. Eduardo Pereira da Silva , eps@isep.ipp.pt

**Scientific Co-Guidance:** Eng. Alfredo Oliveira Martins, aom@isep.ipp.pt


Departamento de Engenharia Eletrotécnica
Instituto Superior de Engenharia do Porto
Rua Dr. António Bernardino de Almeida, 431, 4200–072 Porto

October 28, 2024

*Dedico este trabalho a todas as pessoas que se cruzaram comigo e que ajudaram a torná-lo possivel. À minha família, por ser meu alicerce incondicional, e por encorajarem-me necessários para que eu pudesse seguir em frente, mesmo nos momentos mais desafiadores e dificis. Aos meus amigos, que foram uma fonte de inspiração e apoio. Aos meus professores e orientadores, por sua dedicação, orientação e pelo conhecimento compartilhado, que foi fundamental para a realização deste projeto.*

# Acknowledgements

I dedicate this work to my family, whose unwavering support throughout my academic journey has always been a source of belief and encouragement in my pursuit of knowledge. In particular, to my wife, who has always supported and helped me along this path, whose presence and understanding brought joy during the difficult moments. This work is also dedicated to my teachers and mentors, whose wisdom and guidance have shaped my thinking and led me through the intricate paths of research. I am grateful for their patience and continuous inspiration. Special thanks to Eng. Alfredo, who, from the beginning of this project, provided support and assistance, making its realization possible. To the friendships cultivated throughout this academic journey, which lightened the days and made the challenges easier to overcome. Every conversation, laugh, and shared support contributed to the richness of this journey...

# Abstract

The exploration and use of the oceans have led to an increase in the number of objects and equipment left behind on the seafloor, including abandoned fishing nets, known as "ghost nets," and other lost items. These marine debris pose significant environmental threats, affecting underwater ecosystems, navigation, and even human life. Ghost Nets, in particular, continue to trap and kill marine life indiscriminately, contributing to the degradation of marine habitats and a decrease in biodiversity. Moreover, they can also hinder underwater exploration efforts by entangling and damaging submersible drones.

This thesis investigates methods and technologies for detecting and identifying abandoned fishing nets and other lost objects on the seafloor. The study evaluates different sonar systems and other underwater detection methods, as well as image processing techniques, particularly in their application to unmanned underwater vehicles (AUVs) and unmanned surface vehicles (USVs). By combining multiple systems and sensors, the aim is to improve detection accuracy and object identification.

Laboratory and real-world tests in open sea environments were conducted to validate the effectiveness of the proposed solutions. The research also explores the challenges associated with sonar detection in shallow waters, including issues related to the acoustic Backscatter from the seafloor and the limitations of current sonar technologies in accurately identifying mid-water targets.

Through this investigation, the thesis aims to contribute to developing advanced technological solutions that can be used in ocean cleanup operations and preserving marine ecosystems, promoting the sustainability of marine resources for future generations.

**Keywords**: Marine Debris, Ghost Nets, Sonar Detection, Underwater Imaging, AUVs (Autonomous Underwater Vehicles), USVs (Unmanned Surface Vehicles), Environmental Impact, Marine Ecosystems, Submersible Drones, Acoustic Backscatter, Shallow Water Detection, Marine Habitat Preservation, Sustainability, Image Processing, Seafloor Mapping.

# Contents

# List of Figures

# Listings

# Glossary

**ALDFG**

(Abandoned, Lost or Otherwise Discarded Fishing Gear) Fishing gear, such as nets and traps, that is abandoned, lost, or discarded, continuing to cause environmental harm and impact marine life after being lost.

**AUV**

(Autonomous Underwater Vehicle) An autonomous, unmanned underwater vehicle used for underwater exploration, monitoring, or research missions.

**Backscatter**

Acoustic backscatter refers to the reflection of the sonar signal back to the transmitter, used to detect submerged objects or structures.

**BlueView**

A type of sonar manufactured by Teledyne Marine, used for underwater detection, characterized by being more affordable but providing lower image quality compared to other sonars, such as the M3.

**Ghost Nets**

Abandoned or lost fishing nets that continue to capture and kill marine life indiscriminately, impacting underwater ecosystems.

**Greyscale Image**

An image rendered in shades of gray. In the context of underwater image processing, greyscale images are used to simplify data processing by neural networks

**M3 Sonar**

A high-quality sonar system manufactured by Kongsberg, used for underwater imaging and bathymetry, providing higher resolution images compared to other technologies.

**Neural Network**

A computational model inspired by the human brain, used for image and data processing to automatically identify and classify underwater objects.

**ROS Bag**

A file format used in the Robot Operating System (ROS) to store data, including sensor readings and images captured during underwater testing.

**Sonar**

A system that uses sound waves to detect and locate submerged objects. It is widely used in maritime and underwater applications.

**Target**

Refers to the submerged objects that are to be detected and classified during research or inspection missions. Examples include tires, fishing nets, and traps.

**TVG**

(Time-Variable Gain) A dynamic adjustment applied to the acoustic signals captured by sonar to compensate for signal attenuation due to the distance from the target.

**USV**

(Unmanned Surface Vehicle) An unmanned surface vehicle used for surface operations, such as hydrographic surveying and environmental monitoring

**VGG16**

A convolutional neural network architecture widely used for image recognition tasks, including object classification and detection.

# Acronyms

**ALDFG** *Abandoned, lost or otherwise discarded fishing gear* 8

**ANN** *Artificial neural networks* 18

**AUV** *Autonomous underwater Vehicle* 2, 6

**CLAHE** *Contrast-limiting adaptive histogram equalization* 16

**CNN** *Convolutional neural network* 13

**DFO** *Department of Fisheries and Oceans* 8

**DGRM** *Direção Geral Recursos Naturais, Segurança e Serviços Marítimos* 24

**DVL** *Doppler velocity log* 13

**EIQ** *Enhanced Image Quality* 49

**FLS** *Forward-Looking Sonar* 6, 9, 20

**FPF** *False Positive Fraction* 19

**FPS** *Frames per second* 49

**GTMUR** *Georgia Tech Miniature Underwater Robot* 12

**INESC TEC** *Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciencia* 14, 39

**ISEP** Instituto Superior de Engenharia do Porto 14, 39

**LIBS** *Laser-induced break-down spectroscopy* 15

**LSA** *Autonomous Systems Laboratory* 14

**LSTM** *Long short-term memory* 37

**MRF-Net** *Multiple Receptive Field Network* 9

**PIL** *Python Imaging Library* 52

**RNN** *Recurrent neural network* 36

**ROS** *Robot Operating System* 39

**SAS** *Synthetic Aperture Sonar* 20

**SBC** *Single board computer* 17

**SNR** *Signal-to-noise Ratio* 12

**SPPnets** *Spatial Pyramid Pooling Networks* 21

**TNF** *True Negative Fraction* 19

**TPF** *True Positive Fraction* 19

**TPI** *Topographic Position Index* 16

**TRI** *Terrain Ruggedness Index* 16

**TVG** *Time Variable Gain* 8

**USBL** *Ultra-short baselin* 13

**USV** *Unmanned Surface Vehicle* 2

**YOLO** *You only look once* 17

# Chapter 1

# Introduction

The exploration and use of the oceans have grown significantly in recent decades, leading to an increase in the amount of equipment and materials left on the seafloor, including abandoned fishing nets, also known as 'ghost nets,' and other lost objects. These marine debris represent a serious threat to the environment, affecting underwater ecosystems, navigation, and, in some cases, even human life. Ghost nets, in particular, continue to capture and kill marine life indiscriminately, contributing to the degradation of underwater habitats and the decrease in biodiversity. Furthermore, during exploration, underwater drones can become entangled in these types of debris, potentially leading to damage and halting the exploration mission until the drone is rescued. The effective detection and removal of these objects have become priorities for various environmental and governmental organizations. However, the precise identification and location of these submerged materials present significant technical challenges due to the complexity and diversity of oceanic conditions, as well as the limitations of traditional monitoring systems. In this context, advances in sonar technologies and image processing have offered new opportunities to tackle this problem more effectively. Modern sonars, combined with data processing algorithms, enable not only the detection of submerged objects but also their automatic identification and classification, differentiating, for example, between fishing nets, natural structures, and other debris. There have also been many advancements in underwater camera technology and parallel use methods to assist, such as light calibration algorithms to more optimized formats for capturing images.

In the light of this, the objective of this thesis is to explore methods and technologies for the detection and identification of abandoned fishing nets and other lost objects on the seafloor. Different types of sonars, among other target detection methods on the seabed, and image processing techniques will be analyzed, with a focus on their application in *Autonomous underwater Vehicle* (AUV) and *Unmanned Surface Vehicle* (USV). The feasibility of combining multiple systems and sensors to improve the accuracy of detection and identification of submerged objects will be evaluated through laboratory tests and real-world scenarios, culminating in a series of experiments conducted in open water. This investigation is hoped to contribute to the development of advanced technological solutions that can be used in ocean cleanup operations and the preservation of marine ecosystems, thereby promoting the sustainability of maritime resources for future generations.

## 1.1 Contextualization

With the increasing pollution in the oceans, it has become clear that there is a growing need to preserve maritime ecosystems. One of the existing problems is the presence of numerous fishing nets and other materials associated with fishing gear that are abandoned or lost on the seafloor or in the water column, causing issues for marine life and the ecosystem as a whole. This can result in fish becoming entangled, destruction of natural habitats, harm to larger species, and various other examples. Hence, the removal of these objects from the sea becomes crucial. Initiatives like the NetTag project have emerged to study ways to detect fishing nets and lost fishing gear objects.

## 1.2 Problem Formation

Abandoned or lost of fishing net has surfaced as a significant conservation issue that continues to compromise the economic, social, and ecological aspects of the marine environment. Even for AUV, underwater fishing nets represent a danger. To avoid irreparable damage to the AUV, they need to be capable of identifying and locating fishing nets successfully, and avoid them. The AUV ability to successfully detect depends on the accuracy and efficiency of the detection system.

As this intention is necessary, it is crucial to develop better means to detect and determine fishing nets and other objects, as well as the ability to determine what is being detected. With this goal in mind, it was essential to study various types of sonar to find one that would allow the determination and characterization of what was being seen through it, facilitating its removal.

### 1.2.1 Objective

The main objective is to determine how a system can be implemented to find and identify fishing nets and other abandoned objects on the seabed. To do this, it was necessary to find the best method and system for detection, considering possible real-case scenarios such as underwater drones and boats on the water's surface. It was also important to verify if combining several systems would positively impact the detection solution.

These methods required to be validated through laboratory tests and then on the high seas so they can be approved in a real-world situation. Alongside this, it was necessary to find methods for processing the collected data in order to check whether or not it was possible to detect the various targets that might appear and finally, whether these systems could still identify them.

### 1.2.2 Expected results

As mentioned previously, the aim was to develop a suitable solution that would be better to find targets on the seabed or in the water column so that drones can both avoid them and provide information about the target so that it can be removed. To this end, it is hoped to find a system that allows targets to be visualized with sufficient quality so that another system can then be used to identify and classify them automatically.

## 1.3 Work Plan

Initially, it was decided that it would be necessary to analyze other case studies to understand what had already been done and what remained unexplored, and through these studies, define a promising approach and establish the next steps. An approach was determined that could be the best way to investigate and carry out the study. To test and prove this investigation, a test plan was developed to validate whether or not it would be a feasible option to pursue, and similarly, these tests served to validate sonar options. Tests were conducted in a controlled environment to validate the solution, and after analyzing the initial data, it became possible to move on to real-world tests. After the tests were conducted, the results were analyzed, and then we moved on to data processing so that, through this process, it would be possible to determine and classify the target. Finally, acquiring as much data as possible became important, as the more data available, the better the training of the neural network, which would lead to more accurate results and allow for more effective target characterization.

## 1.4 Organization of the dissertation

Throughout this dissertation, various topics related to the theme will be addressed. Following this introduction, a review and research on related works are conducted, where we first explore studies that cover different methods for detecting underwater objects, allowing us to define an approach to follow. Subsequently, based on these same studies, data processing methods and ways to optimize the processing are also analyzed. After that, a discussion is held about everything analyzed and studied up to this point. In the third chapter, a detailed description of the problem is provided, where the targets to be visualized are discussed, particularly those that are most common along the Portuguese coast. Next, a theoretical study is conducted on underwater detection methods and data processing techniques, especially image processing. This leads to the implementation section, where the tests carried out are discussed, along with the planning and results. In that same chapter, the data processing methodology is presented, validating the theoretical part.

# Chapter 2

# Related work

Considering the addressed topic, it was necessary to determine the possible ways to detect objects underwater. For this purpose, various research works related to this topic were analyzed. This part did not specify the type of component to be detected, but rather ways to find multiple targets with different shapes, different materials. It was also verified for various scenarios, namely whether they are in the water column or on the seabed. Subsequently, solutions combining more than one detection system were also examined in an attempt to achieve better detection results. Finally, data processing methods were examined to automatically determine what was being observed. Although the only real-time system found relied on human detection, and in cases of systems combining multiple detection systems, there were also post-data collection analysis systems that, using artificial intelligence and component analysis methods, were able to train systems to determine with some accuracy and precision what was being observed.

## 2.1 Detection methods.

For target detection, several studies were analyzed that aim to find other objects on the seabed or in the water column, and afterwards, focus shifted to specific scenarios with Sound Navigation And Ranging (SONAR) and video cameras. Further solutions were also examined, but priority was effectively given to SONARs or combined solutions that included SONARs and other systems.

### 2.1.1 Sound Navigation And Ranging.

Several studies were analyzed in an attempt to determine the most commonly used and effective sonars for detecting objects underwater, both on the seabed and in the water column. Following this, a survey of the most important and widely used sources of information was conducted, enabling a discussion based on this topic at the end. Just for introduction can be see some topics of tips of sonar that were found during the research.

1. Side-Scan Sonars: These sonars emit acoustic beams to the sides of the vessel, allowing for detailed imaging of the seafloor and submerged objects alongside the vessel's path.

2. Depth Sounders: Designed to measure water depth, these sonars are common on vessels to prevent grounding.

3. Acoustic Imaging Sonars: These sonars provide detailed images of the seafloor and submerged objects, similar to an underwater camera but using sound waves instead of light.

4. Search and Rescue Sonars: Designed to locate submerged objects or individuals, such as sunken vessels or drowning victims.

5. Seabed Mapping Sonars: Used to map the seafloor and identify geological features such as underwater mountains, valleys, and channels.

6. Fish Finders: Especially used in commercial fishing, these sonars help locate schools of fish.

7. Underwater Navigation Sonars: Assist in underwater navigation, allowing divers and Autonomous Underwater Vehicles (AUVs) to locate themselves and avoid obstacles.

One of the studies used [1]for reference was conducted by the School of Engineering & Physical Sciences at Heriot-Watt University in Edinburgh, UK. In this study, a *Forward-Looking Sonar* (FLS) imagery, this solution was for AUV, and with the aim of locating debris and litter in the ocean, sea, lakes, or rivers. In this paper, they reflect on various studies and conclude that the small size of marine debris presents a problem, because most of the systems don't have a resolution for that and sonar sensors with big fields of view, such as Synthetic Aperture and Sidescan, have issues detecting small objects as their spatial resolution is close to 3 cm. Objects up to 1 meter in length can be identified with such sensors.

So they use ARIS Explorer 3000, at 1.8MHZ setting with a range of 1 to 3 meters, and this setting was used in a water tank and in a real case on the ocean. Over 2000 images were captured inside a water tank containing different types of debris

objects. After processing the data, all with the same resolution and image size, they were passed through a neural network to distinguish between seafloor debris and the seabed itself, and even to differentiate fish from other objects.

Across the ocean at Dalhousie University [2], at Halifax, Nova Scotia they conducted another study that target methods of gear detection, to increase the chances of retrieving the most possible of ALDFG, they targeted in Canada's most productive American lobster (Homarus americanus) fishing area, 27 side scan sonar (SSS) transects were conducted in Lobster Fishing Area (LFA) 34 over a 12-day survey. On their studies exist various approaches for lost gear detection and monitoring, ranging from manual efforts, such as SCUBA diving missions to AUV. According to their studies of the detection methods used in published research, acoustic instruments, such as side scan sonar, are increasingly being used in underwater detection as they yield distinctive results about objects that are visible on the seafloor. Side scan sonar systems capture acoustic data from a swath of the seafloor to generate images depicting the surface characteristics of the seabed. These sonar instruments can be mounted on a tethered towfish, towed behind a vessel, or remotely operated vehicles. When towed, side scan sonar can be challenging to use in environments with highly variable or steep bottoms. Adjusting operating frequencies and transducer configurations allows for the assessment of different seafloor substrate types and can impact the effectiveness of object detection. Dual-frequency side-scan sonar systems are increasingly utilized in marine monitoring and exploration missions. These sensors typically function at two distinct frequencies, offering varying coverage and resolution. The pulse frequency emitted by the system varies based on the desired range and system specifications. Successful identification of marine debris on the seafloor using side scan sonar depends on factors such as weather conditions, noise levels, equipment quality, and post-processing software.

To optimize practical tests and the final data collection, they conducted a study with coastal authorities and fishermen to collect data in areas with a higher probability of finding more ALDFG. With this information, they were able to map the areas of greatest interest, using a color gradient representation where warmer colors indicated a higher number of reported targets. The reported lost fishing gear locations representing lost gear up until June 2021 were obtained from Fisheries and Oceans Canada and were represented spatially by displaying their latitudinal (Y) and longitudinal (X) coordinates in Esri mapping software, ArcPro 2.7. Prio.

Side scan sonar surveys were carried out using an EdgeTech 4205 MPMT Dual Frequency Sonar, which operated at both low and high frequencies: 230kHz for low frequency and 850kHz for high frequency. The low-frequency signal scanned a 150m wide swath of seafloor on each side of the side scan sonar towfish, while the high-frequency signal covered a 50m wide swath on each side. Post-acquisition corrections to the side-scan data were made using Chesapeake Technology's SonarWiz

7. Corrections applied to the raw side-scan sonar data included automatic TVG, which adjusts the acoustic signal strength relative to the distance from the towfish to compensate for signal attenuation through the water column. Environmental parameters and fishing pressure data were collected from multiple sources and organized in ArcGIS Pro 2.7. All geospatial data layers were managed and recorded within an Esri geodatabase. Potential ALDFG objects were visually identified from the processed side-scan data. Coordinates of these potential targets were extracted and distributed for retrieval. Initial target identifications were based on raw side-scan data and processed sonar transects, leading to the generation of final identification targets. Initial targets refer to those found in surveys before post-processing, while final identification targets are confirmed after a second review to ensure accuracy.

Two years later, more tests were conducted at the same site and a new study [3] was carried out with the same objective using a side scan sonar, and in this case, the approach was to attempt area mapping, generating a grid with a resolution of 3 km x 3 km, primarily aimed at detecting fishing traps on the seabed. They mapped 34 lobster fishing areas, which took approximately 12 days to obtain the data. With this, it was possible to detect around 114 potential *Abandoned, lost or otherwise discarded fishing gear* (ALDFG), which were visually confirmed. These 34 areas were considered because of the information collected from the data generated because of since 2020, the *Department of Fisheries and Oceans* (DFO) has mandated reporting lost and retrieved fishing gear in all Canadian commercial fisheries. The data was collected by using an EdgeTech 4205 which is a dual-frequency sonar, was operated at a low frequency of 230 kHz and a high frequency of 850 kHz. The low-frequency signal scanned a strip of seafloor that was 150 meters wide on each side of the side scan sonar towfish, whereas the high-frequency sonar scanned a strip of seafloor that was 50 meters wide on each side. After acquiring the side-scan data, corrections were made using Chesapeake Technology SonarWiz 7. A single sonar analyst was responsible for conducting the post-processing and identifying ALDFG targets. Was includes *Time Variable Gain* (TVG) and layback positioning. The TVG correction balanced the backscatter and modified the acoustic signal strength relative to the distance from the towfish. This adjustment compensated for signal weakening through the water column and adjusted for the perceived size of the target. For the position is cross information from SonarWiz7 and GPS vessel position was allowed to achieve accurately the position.

This article revealed that it becomes challenging to cover such a large area. According to the reported analyses, the best approach would be to reduce the intervention and detection area to achieve greater reading accuracy and ensure more certainty about the targets being observed. Another issue encountered was that due to the terrain and depth of some capture zones, it was challenging to obtain accurate results. This resulted in some cases where possible targets were not identified,

and some of the things that were initially thought to be targets turned out to be nothing more than seabed features. Adverse weather conditions also hindered data collection, delaying the mission and affecting measurements on certain days because it did not allow the sonar to be 100% stable. In some cases, it was suspected that the targets may have shifted position between data collection and retrieval time due to unstable conditions.At the end was processed during 15 days allocated to do 109 tows, but in 72% of the tows wasn't retrieved any targets.

There is also some articles from MDPI jornal [4] that talk about this subject, one of them is *Multiple Receptive Field Network* (MRF-Net) for Autonomous Underwater Vehicle Fishing Net Detection Using Forward-Looking Sonar Images" on this article they approach the importance of avoiding the underwater fishing nets, because they represent a danger faced by the AUV, for the test they used to recognize FLSimages, with the proposed architecture is a center-point-base detector, which uses a novel encoder-decoder structure to extract features and predict the center points and bounding box size, with an interference of reverberation and speckle noises in the FLS image, they use a series of prepossessing operations to reduce noises. For the data collection on their tests was also used a Gemini 720i.

In this paper, they also discuss the complex and variable characteristics of seawater and the reverberation at the bottom and surface of the sea. FLS images typically exhibit three properties: first, low resolution; second, a lack of information due to the low level of accuracy and precision of the targets when viewed by FLS on a large scale; and third, the readings can be affected by noise present in the measurements, which can originate from reflections on the seabed or from the targets. Various researchers have proposed several preprocessing methods to suppress noise interference in images and object detection algorithms based on object proposals. Although these algorithms have achieved great success, the process of selecting object proposals is complex and time-consuming. Moreover, they also noted that there is still considerable room for improvement in the accuracy of object and target detection.

In their initial research, the appearance characteristics (area, centroid, perimeter, etc.) of the object through threshold segmentation of FLS images were estimated, or complex manual features to extract information from the images were proposed. However, as of the latest publication, the object detection algorithms used for FLS images are primarily based on the actual task, and the appropriate algorithm is selected according to the task requirements.They also encountered some problems while collecting the dataset because the environmental inconsistency generated noise that disrupted and distorted some images, affecting the accuracy of the AUV's position and location. The precision of the ground truth of the images is very important for supervised object detection in FLS images. They marked and recorded all intersections with possible targets, and using tools, it was possible for the model

to distinguish between fishing nets and other obstacles. They added cloth and plastic bags to the dataset. Our dataset includes three types of obstacles at different distances (0-5 and 5-10 m): fishing net, cloth, and plastic bag.

On Macedonian part of Lake Ohrid has conducted a pilot study [5] by Hydro-biological Institute Ohrid and with another country's held, they first studied the problem of that area, and they concluded approximately the quantity of existing ALDFG and their types to get a better understanding of the target they needed to find. They also analyzed whether they would find them closer to the sea bottom or more present in the water column.

In the next step they gathered from fishermen about places where they have lost parts of their gear or have encountered ghost nets in the past. Then they defined four techniques to detect or remove the fishing nets, sonar's was use to detect Lowrance Elite-9 Ti" device of 433/800 KHz dual frequency (DownScan ® HD / StructureScan Imaging™). Was also use two different types of creeping and was also diving on the places that the locals warned existing ALDFG. In this study was also on the sonar it was equipped with CHIRP sonar with a display resolution of 800 × 480 pixels. During fieldwork, the echo sounder was mounted on a 12-meter-long research vessel. It was used consistently, either independently or simultaneously with divers' surveys and creeping exercises, to evaluate its reliability for ghost-net detection. At all times, the echo sound reached the bottom of the screened lake areas, were recorded the position of the nets using GPS of the echo sounder. The location and depth of the recovered nets were recorded, along with the methods used for detection and removal. Additionally, the length, height, mesh sizes, and condition of each net were documented.

From their result its possible to see that in the 18 days of field work, a total distance of 512km, was possible to encounter nets in 38 locations, they retrieved 243 nets. The majority of the recovered targets were just small parts of nets, but there were cases where they removed entire nets that had either gotten stuck on the bottom or been abandoned in the water column. But the detectability of nets by the sonar was limited, and only free-floating nets became visible on the echo sounder's screen. The deepest nets detected with the sonar were at 60 m, and surveys conducted at depths ranging from 60 to 250 m did not yield any positive signals. Was also tested raised the nets that were in the bottom of the lake by 2 m and that way was already possible to identify the nets.

Sonar imaging systems were also utilized by the Centre for Marine Science and Technology at Curtin University, Australia,[6] where they aimed to develop a method to enhance the detection and identification of marine fauna using sonar imaging. In this case, they employed the Tritech Gemini imaging sonar intending to observe sharks. The targets ranged between 1.4m and 2.7m in length and were positioned at distances between 1 and 50 meters from the sonar, at depths up to 15 meters.

At close range, it was easy to determine the presence of a shark based on its shape. However, beyond 5 meters, identifying the shark based on shape alone became difficult, and from this distance onward, the analysis relied on recognizing movement patterns to distinguish a "shark-like" target before it became merely an acoustic target at greater distances. Another important factor they observed was that the seafloor reflected the acoustic beams, and this reflection at shallower depths could dominate the image and mask potential targets. This highlighted the need to identify noise and establish the necessary ranges and depths to extract the best possible data, thus enabling more accurate target identification. The Tritech Gemini 720i 300M sonar system they use operates at 720 kHz, featuring a 120° horizontal and 20° vertical beamwidth, with a -10° elevation angle. The system consists of 256 dynamically focused beams across the 120° horizontal field of view, providing an effective azimuthal beam resolution of 0.5°. The along-beam resolution varies depending on the range setting, but it can achieve a resolution as high as 8 mm.

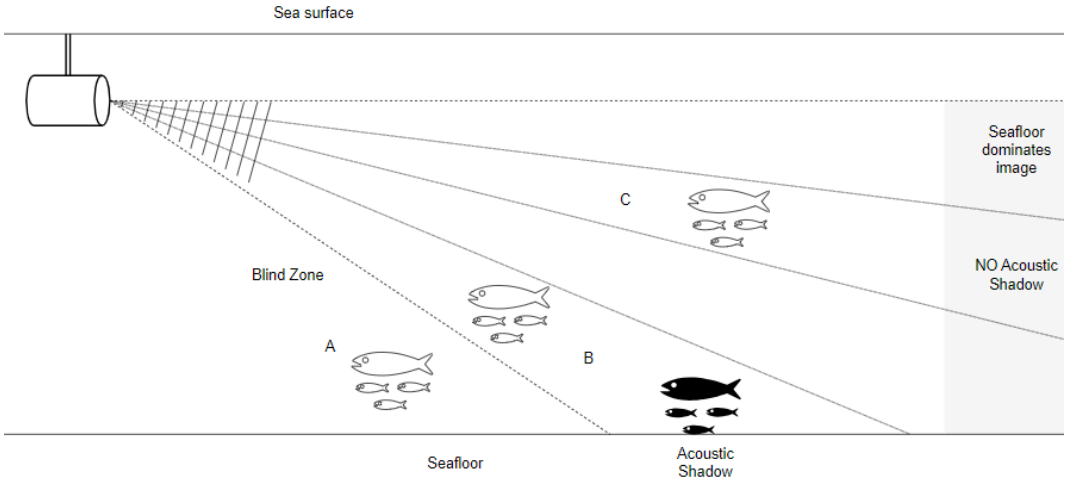



Figure 2.1: Schematic of Gemini imaging sonar beam.
[6]

A diagram of the Gemini imaging sonar beam is shown alongside examples of sharks positioned at similar distances from the sonar head but at different depths, resulting in varied acoustic images. These include: a) being outside the acoustic beam and in the "blind zone," b) being in the lower part of the acoustic beam, which creates a weaker image of the shark but produces a strong acoustic shadow, and c) being high enough in the water column to generate a strong image of the shark from back-scatter, with any shadow falling outside the sonar's designated range. The diagram also indicates the horizontal and slant range, with the "blind zone" and the area where the "seafloor dominates the image" marked by mottled and grey regions, respectively.

According to the information found, it can be stated that the range and position of the target in the water column relative to its position within the sonar beam

have a significant impact. As the horizontal range to the target decreases, if the target is not visible, it is likely located below the sonar beams, which is why it is not being detected. Once within the sonar's field of view, the target will reflect acoustic backscatter to the receiver, and targets at a shorter slant range than the seafloor will provide the highest *Signal-to-noise Ratio* (SNR). Suppose the target is within the acoustic beam and at a shorter slant range than the first contact of the acoustic beam with the seafloor. In that case, one can expect not only some reflected backscatter but also an acoustic shadow of the target, as seen in point B.

On the other hand, targets in the upper part of the acoustic beam create an acoustic shadow that is outside the sonar's configured range. According to the same study, the return from the seafloor is high compared to that of a target in mid-water; once part of the acoustic beam insonifies the seafloor, the sonar image is dominated by backscatter from the seafloor, and any target at the same slant range will be difficult to distinguish from the reverberation. Therefore, the system's optimal performance occurs when the target's range is shorter than that of the seafloor.

### 2.1.2 Video Camera.

During the analysis of various projects, it was observed that several solutions utilized cameras both for detecting nets and even for checking their condition.[7] For example, the Georgia Institute of Technology used a net analysis system with the use of a forward-facing camera capable of recording videos while the *Georgia Tech Miniature Underwater Robot* (GTMUR) passed through them, allowing for later analysis of the fishing nets' condition.

Although initially, they had considered using unprocessed images to subsequently use in the neural network, this did not yield the results they intended because, in a clean environment where only the net is present in the image, it is possible to assess its condition. However, in a real-life scenario where the net is dirty or fish may appear in the images, this makes it difficult for the data to be processed. As a result, they ended up using depth images instead of RGB to make it easier for the neural network to process the data.

Stereo vision, which involves two cameras, was then utilized. In this system, since the same point is visible in both the right and left cameras, it is possible to calculate the depth of these points, thereby obtaining more data through the images. Although stereo vision greatly reduces background noise in the images, real-world underwater images typically have poor quality due to low light, which greatly complicates the quality and precision of what is in each image. However, by using various algorithms, it was possible to filter and correct some of these aspects, but this required the system to rely on much more computational resources, resulting

in an average processing time of 4 seconds per image, thus not allowing for real-time analysis.

Exist also another study [8] to approach turbid underwater environments, because the visibility is significantly degraded by floating particles that cause light attenuation, which is one of the main problems for accurate underwater inspection by optical cameras. To obtain clear images for net inspection they use a AUV V pose control strategy for fish farming net inspection in turbid water, they also use a *Convolutional neural network* (CNN) that train offline in the way of determine what was the best method, for that they use experimental results both in swimming pool and in real cases with fish farm environments and than they were able to demonstrate the effectiveness of the proposed methods. For effective AUV net inspection, it is crucial to maintain the correct relative pose between the AUV and the net to acquire clear images. Various sound wave-based sensors, such as *Ultra-short baselin* (USBL) and *Doppler velocity log* (DVL), can determine the relative pose, but these sensors are too costly and heavy to be installed on small-sized AUVs. However floating particles significantly degrade visibility in turbid underwater environments due to light attenuation, which is one of the main factors impeding accurate underwater inspection. To obtain clear images in turbid water, the AUV should be close to the net from a far distance. Because the closer the camera is to the net, the gradient of an image increases, and that changes the sensitivity of the AUV to the movements.

The average gradient value of an image can indicate its clarity since it reflects the average rate of pixel intensity change. An image of a net taken from a far distance displays low gradient values, making the net appear unclear. Conversely, an image of a net taken from a close distance exhibits very high gradient values, making the net visible. Based on this observation, the mean gradient information of an image can serve as a highly distinctive feature for a vision-based approach to controlling the relative pose of an AUV concerning a net in turbid water. To increase the sensitivity of the mean gradient of an image to changes in the object's gradients, they propose using only those gradients with values exceeding the optimal threshold as determined by Otsu's method. To obtain clear images of fishnets, the AUV must maintain the correct distance and heading direction relative to the net. They determine the degrees-of-freedom (DOF) motion and achieve the dynamic model of the AUV, connecting it to the pose to enable better image capture.

They tested on a pool environment first and then they use a real fish farm in Wando-gun Jeollanam-do, Korea, to create a turbid underwater environment, a certain amount of soil was added to clean water. Also was use BlueROV2 with open-source autopilot hardware and software, and applied to 2 different AUV with communication to a surface computer, to process the data in real-time and control the entire process. With QGroundControl software, the operator can obtain the current image frame transmitted from the AUV to check the current state of the

AUV. To see the light interference was also test during tree different times. At the morning the pool is exposed to sunlight, with slightly noisy lighting conditions around the water surface, during the daytime the pool is shaded, with less noisy lighting conditions. The dataset used in training is generated at this time. And for the last the test was also made at night no natural light, requiring an extra light source from the AUV.

Experimental results in swimming pools and real fish farms demonstrated the effectiveness of the proposed methods. A real fish farm experiment showed the practical feasibility of the approach. A CNN-based controller for distance target selection allows the AUV to move to an area where the net is clearly visible, working effectively in different environments, including nighttime and real fish farms, and demonstrating robustness against various lighting conditions and turbidity.

### 2.1.3 Others systems.

Exist another project [9] like the IRIS AUV that was developed under the NetTag project, NetTag consists in tagging fishing gears and enhancing onboard best practices to promote waste-free fisheries, aims to reduce and prevent marine litter derived from fisheries, bringing together scientists, engineers, and the fisheries industry.

This solution was developed by the *Autonomous Systems Laboratory* (LSA), which is part of Instituto Superior de Engenharia do Porto (ISEP) and *Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciencia* (INESC TEC). NetTag, at this moment developed a solution composed by two technological devices: low cost, miniature, and environmental-friendly acoustic tags (transponders) and acoustic transceivers for unique localization (with fisher's personal ID) of lost gear. These acoustic tags and automated short-range recovery system according to the

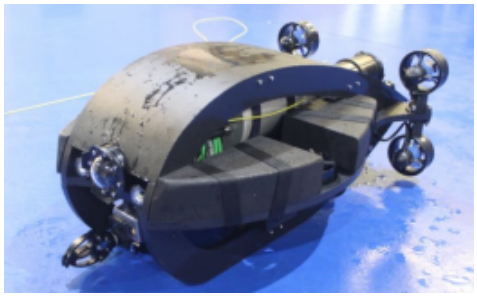


Figure 2.2: IRIS AUV.
[9]

project and is study will be innovative in terms of cost (low-cost), location accuracy (able to uniquely identify an item of fishing gear, precisely determine its position and facilitate its recovery), life cycle (have the sufficient battery life to allow location even after being lost for several months), personal and user-friendly (with a simple,

intuitive and reliable configuration interface customized for fishers to set up the devices with their personal ID numbers).

The IRIS AUV also uses a DVL sonar (Water Linked A50 DVL) to obtain velocity over the sea bottom. Through the use of tags, it was possible to locate the nets and other targets. The tags were low-cost to ensure the viability of this solution, as increasing the cost of the nets could hinder the ability of fishermen to access them. However, it would still be necessary to use systems to detect these tags, which many people working in this field do not have, so that fishermen who lose their nets can recover them. This AUV already enables both functionalities as it is equipped with suitable equipment to detect the tags and has also been developed with a robotic arm capable of retrieving them.

### 2.1.4 Solutions with two or more systems.

There are also solutions where studies [10] concluded that a single system would not be sufficient, so they opted to combine solutions, as seen in the GEOMAR paper. In this paper published by them, the solution involves equipping an AUV with a sonar to detect the object from a distance. The AUV then approaches the object until it is close enough for the onboard camera to process the image and determine if the detected object is the intended target. The solution also includes GPS location so that the position of the target can be marked in case the onboard system cannot recover it. Additionally, the AUV has the capability to remove and capture the target. First is done an acoustic mapping using the AUV following a predefined lawn mower trajectory to obtain a first map of a large area covering square kilometers. Then using that map was extract the data in order to identify the zones with high probability of containing Mn-nodules(manganese nodules). After having identified the working area with high Mn-nodules probability within the acoustic maps, the AUV navigate closer to the seafloor in order to optically detect the potential nodules using a downward-looking camera. In this phase, the robot detects and identifies any object on the sediment that resembles a manganese nodule. Upon identifying a potential nodule, the system calculates its 3D position and size. Following that the AUV performs an operational inspection on the nodules using *Laser-induced break-down spectroscopy* (LIBS). This will provide an estimate of the percentage of relevant minerals per nodule. For the acoustic detection of the Mn-nodules a multibeam echo sounder was used, in bathymetry and backscatter provided was given by the device valuable information. It was obtained by first gridding the multi beam data and smoothing it with a Gaussian filter whose kernel size depends on the noise characteristics of the device.

The depth of an AUV is often limited, the areas beyond a maximum depth rating must be avoided. This information is important because of two factors to avoid collisions of the AUV with cliffs, steep flanks of depressions, or large obstacles

during its mission that are steeper than the maximum rise rate of the AUV, and the Mn-nodule density tends to vary with slope. Another feature that they used was the *Topographic Position Index* (TPI). TPI can be used to classify geological regions for example ridges, upper slopes, lower slopes, and valleys.

Besides the TPI, they also use the *Terrain Ruggedness Index* (TRI) identified as valuable for nodule density estimation. Low TRI values indicate a flat or smooth terrain, while high values indicate roughness. For the visual detection of manganese nodules using a deep learning approach. The initial step involves proper image prepossessing on the acquired images, which is essential due to uneven illumination and other underwater effects such as backscatter and attenuation. They divide this in four parts first they define a lighting correction in order to try to compensate for the acquired images absence of precise information on the placement and nature of the light sources, the distance between the camera and the seabed, and the 3D structure of the scene. This is called lighting inhomogeneity corrections. To correct the inhomogeneity and vignetting-like artifacts in a single step a 2D "inverse illumination distribution" was applied to the original input images. Then was apply noise reduction, because in underwater images the reduced quantity of the light or mist in the water and other factors could reduce visibility and create a lot of noise, for they apply bilateral filter that soothes the images while preserving edges using a nonlinear combination of nearby image values. The third part is image dehazing which aims at removing the degradation of images affected by haze and mist, or also underwater the expected is lost of contrast for poor visibility conditions, then they applied *Contrast-limiting adaptive histogram equalization* (CLAHE) that was a well-known enhancing technique that restores the contrast of the images. For the last, they applied Convolutional Neural Networks Detection Methods.

## 2.2 Data processing methods.

One of the studies [1] used for reference was conducted by the School of Engineering & Physical Sciences at Heriot-Watt University in Edinburgh, UK, was describe the approach for the debris detection in FLS imagery, that consists in two main components Training data pre-processing and a CNN architecture. With that, it was trained two convolutional networks, one that makes binary decision(debris or background) and another that makes a multi-class decision. To improve the process, the image was prepared to have 96 X 96 sliding window with a stride of 8 over the input image, and all the image was processing in gray scale. For the training, it was obtain a dataset of 22446 images. In the results they show that they have for the FLS images in the binary classifier around 80.8% correct detection's for the samples of the debris, that was a good result considering that this was in real time, the AUV was recording online the data for the model to process. But this also shows that

in this mod exist less detection and for that exist less false positives. But in the binary problem is easier than the multi-class problem, as discriminating between object and non-object is easier than the discrimination of non-object versus several object classes. Having additional training data will mitigate this issue and lead to the development of a significantly more precise multi-class detector. Furthermore, class balancing plays a crucial role, especially given that the multi-class problem tends to exhibit less balance compared to the binary scenario.

Another study [7] that references this topic is the Georgia Institute of Technology used a net analysis system with the use of a forward-facing camera capable of recording videos, two methods were addressed, first was the Single Shot Multibox detector, in this method a deep neural network was generated to allow real-time detection, as such for this to happen it needed to be a fast system, for this purpose the input had to be as light as possible, allowing it to operate within the AUV, so each image is passed only once, producing an output immediately. Predictions are based on default bounding boxes that are estimated at the beginning of detection and are refined at the end of the detection pipeline. The second method is *You only look once* (YOLO), this is another single-shot deep learning network, and both perform data augmentation to assist with generalizing the neural network acceleration. for the training were used two dataset one with 1671 examples for testing and later another we 3345 images, after the test were achieved the parameters for the calibration of the network processor. Was trained on 10000 iterations for the nets, and runs at approximately 20Hz on singles GPU. Later, 1000 interactions for the holes detection and the hole is detected in over 95% of the video's images, but all this data was from evaluated on a controlled environment, they do not test on a real case.

Is also a reference on some of MDPI jornal articles, [4] that were used together with the data recovered an object detection multiple receptive field network (MRF-Net), and they used to mitigate the impact of reverberation and speckle noises in the FLS image, a series of pre-process operations is employed to reduce these noises. They also carried out the experiment of detecting and avoiding fishing nets in real-time in the sea with the embedded *Single board computer* (SBC) module and the e NVIDIA Jetson AGX Xavier embedded system of the AUV platform. During this process, they reach the conclusion that denoising algorithms usually inevitably cause images to blur, which affects the detection results. So they prepossessed the images with threshold segmentation to reduce the interference of noises on the objects. Because the gray levels of the target and noise are close to similar, they used gray stretching operation to improve contrast between the target and the noise before threshold segmentation. The structure of this solution is mainly divided into two parts, one is the feature extraction network, which is mainly the feature extraction network that combines various levels of features extracted from FLS images, while

the prediction module is tasked with determining the bounding box of the object.

There is also data processing for the nettag project[9], which is not so similar to the others, because this system uses acoustic receivers to locate the tags and identify them, as each one has its own ID, and it also uses the trilateration method to determine where the tags are and thus know where the nets are. Although it locates the nets, it doesn't determine exactly what state they are in or whether or not they are stuck on the seabed, or even if there are obstacles in the way. In the normal case is made 3 reading around the first good reading of the target at the end is possible to see in the figure below that is found the target in the interception of the three readings, in the image is marked in red.

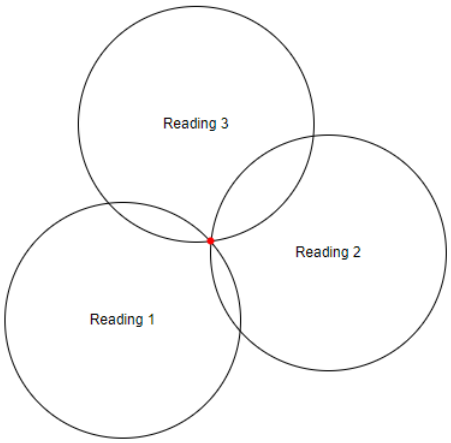



Figure 2.3: trilateration method.

As demonstrated in the GEOMAR paper[10], Convolutional Neural Network detection methods were utilized. Specifically, *Artificial neural networks* (ANN) were employed, which are computational models inspired by biological neural networks and are capable of outperforming earlier forms of artificial intelligence in typical machine learning tasks. CNN is a specific kind of ANN, and regarding that was used in their work a comparison between the performance of Fast-RCNN with DarkNet Yolo. Both algorithms are general purpose object detection schemes with fast and accurate implementations. YOLO trains on full images and directly optimizes detection performance. That proves to be fast no need to have a complex pipeline, simply running the neural network on a new image at test time to predict detection and YOLO makes predictions by considering the entire image as a whole. This approach differs from sliding window and region proposal-based techniques. In the last part of the YOLO application learns generalize representations of objects.

Due to YOLO's high generalizability, it is less prone to failure when encountering new domains or unexpected inputs. On the other hand, Regions with CNN features (R-CNN) proposes to apply a high-capacity convolutional neural network

to bottom-up region proposals in order to localize and segment objects, but this method was more slow so they to find a way to speed-up this method that was to substitute a CNN forward pass for each object proposal for Spatial pyramid pooling networks (SPPnets) able to share computations. The SPPnet approach generates a convolutional feature map for the whole input image and then classifies each object proposal using a feature vector derived from this shared feature map.

They were able to generate a map highlighting the most significant areas, even identifying regions with large, flat zones suitable for more effective seabed mining studies. The network was trained with 6000 images for both methods through visual inspection for each algorithm the Dice Similarity Coefficient, Area of Overlap, True Positive Fraction, True Negative Fraction, and False Positive Fraction were computed. So at the end they reach the follows conclusions they were able to generate a map highlighting the most significant areas, even identifying regions with large, flat zones suitable for more effective seabed mining studies. The network was trained with 6000 images for both methods through visual inspection for each algorithm the Dice Similarity Coefficient, Area of Overlap, True Positive Fraction, True Negative Fraction, and False Positive Fraction were computed. So at the end they reach the follows conclusions Since the images primarily contain sediment and nodules, lower thresholds are preferable as the likelihood of any detection being a nodule is high.

The high dice values and low overlap area values indicate that almost all of the large nodules in the ground truth are detected. The discrepancy arises from smaller nodules, which are over-detected due to missing nodules in the ground truth. The *True Positive Fraction* (TPF)values of 69% for YOLO and 94% for RCNN suggest that almost all pixels containing nodules are detected. The *True Negative Fraction* (TNF) shows that almost all zones without nodules are correctly identified, which is reasonable given that there are more pixels without nodules than with nodules. The *False Positive Fraction* (FPF)is the inverse of the TNF. The RTNF, RTPF, and RFPF metrics are useful indicators of the system's ability to detect a nodule without requiring a perfect match between the ground truth bounding box and the detected bounding box. For instance, if the detected bounding box overlaps with a ground truth nodule, it counts as a positive example. The RTPF shows that 96% (YOLO) and 99% (RCNN) of nodule detection's overlap with a labeled nodule. RCNN better fits the labeled nodules than YOLO, which only detects nodules in the ground truth data.

## 2.3 Discusions.

Detecting submerged objects is a technical challenge that involves accurately identifying targets of different shapes, sizes, and materials in the water column and

seabed. Various methods have been analyzed to address this issue, with sonar being one of the main detection systems used. This investigation explored different types of sonars and their characteristics. One such type is Side-Scan Sonar, which is widely used for detailed seabed mapping and object detection along the vessel's path. This sonar emits acoustic beams to the sides, creating detailed images that are useful for identifying structures and debris on the seafloor. Another type is Acoustic Imaging Sonar, which provides detailed images of the seafloor and submerged objects, functioning similarly to an underwater camera but using sound waves instead of light, making it particularly advantageous in low-visibility environments. FLS is ideal for autonomous underwater vehicles (AUVs) and is used for forward detection, crucial for avoiding collisions and detecting objects such as marine debris.

The investigation also identified several challenges in sonar detection. One issue is the limited resolution of sonars with large fields of view, such as *Synthetic Aperture Sonar* (SAS) and Side-Scan Sonar, which struggle to detect small objects due to their spatial resolution of approximately 3cm. Additionally, the accuracy of the results can be affected by environmental conditions, such as terrain variability and depth. Studies suggest combining different detection systems to improve detection effectiveness. For example, using sonars together with underwater cameras allows for the collection of complementary data, which can then be processed to provide more accurate target identification. Data processing methods, such as artificial neural networks, have been employed to automate the detection and classification of objects, distinguishing between marine debris and the seabed, or even differentiating fish from other objects.

In specific cases, such as detecting lost fishing nets, vision-based systems like stereoscopic cameras and pose control techniques in AUVs have been explored to overcome visibility limitations in murky underwater environments. While effective, these systems require intensive computational processing, which may limit their real-time application. A notable example in this field is the NetTag project, which develops solutions for locating and recovering lost fishing nets using low-cost acoustic tags that can be detected by specific sonars.

In summary, detecting submerged objects, especially when it involves identifying small targets or operating in complex environments, significantly benefits from the combination of different detection technologies, with an emphasis on the use of sonars. The integration of advanced data processing systems, such as artificial intelligence, is also crucial for improving the accuracy and efficiency of these operations. However, environmental and technical challenges remain barriers that require innovative solutions and continuous technological advancements.

Also the detection of submerged objects is a constantly evolving area, driven by the development of advanced technologies such as sonar and convolutional neural

networks. Although sonar-based methods, such as Side-Scan Sonar and Forward-Looking Sonar, already provide detailed and effective images in various underwater scenarios, integration with advanced data processing systems, such as artificial intelligence, has significantly increased accuracy and efficiency in target identification. Recent studies have demonstrated the potential of deep neural networks, such as YOLO and RCNN, to improve real-time detection, especially when combined with data pre-processing techniques to mitigate underwater noise. However, challenges such as the variability of environmental conditions and the need for greater accuracy in multi-class classification problems still persist. The combination of different detection systems and the use of intelligent data processing techniques are helping to overcome these limitations, although more testing in real environments is needed. Therefore, underwater detection continues to benefit from advances in artificial intelligence and combined methodologies, but there is still a long way to go to achieve completely accurate and applicable real-time solutions in complex and variable scenarios. Based on the analysis, it was found that YOLO is the best approach for real-time detection with high generalization. Its speed makes it the optimal choice for future real-world applications, such as AUVs or other systems that need to avoid targets. However, for cases where precision is more important than speed, RCNN offers better results, particularly when it comes to identifying and distinguishing between multiple object classes. So, the system that best fits the application should be chosen accordingly, if you don't mind.

A balance between speed and precision can be achieved by using networks like Fast-RCNN combined with techniques such as *Spatial Pyramid Pooling Networks* (SPPnets). This approach allows for the reuse of computations to optimize processing time while maintaining a high detection rate. This method is particularly useful when detection needs to be performed near-real time without compromising classification accuracy. To further enhance speed, preprocessing of the information to reduce noise can be implemented using threshold segmentation and gray-scale stretching techniques. However, care should be taken about the blurring effect these methods may cause in the images, as it could negatively affect the network's performance.

During the analysis of these studies, it was observed and confirmed that there is few documented content specifically related to the search and detection of fishing nets. This further highlighted the need for this study. Considering this, the study was approached in a way that focused primarily on lost and abandoned fishing nets, as well as other ALDFG, since there are already more studies available for certain types of objects and targets. For example, fishing traps in Canada, and tires in Brazil, among others.

# Chapter 3

# Description of the problem and path followed

## 3.1 Description of the problem

As mentioned before, there are many abandoned fishing nets and objects derived from fishing gear[11]. In short, anything that is ALDFG is a danger to the underwater environment and even to people who do research or rely on the sea, as it affects the entire ecosystem. In the case of the ecosystem, these objects, also known as 'ghost nets', continue to capture and kill marine life indiscriminately. Species such as fish, marine mammals, turtles, and birds are often trapped in these artifacts, resulting in injury, exhaustion, or death. In addition, ALDFG contributes to the degradation of marine habitats, including coral reefs and algae beds, which are essential for biodiversity. The slow decomposition of the plastic materials used in many of these artifacts also leads to the proliferation of microplastics, which enter the food chain and affect a wide range of organisms, from plankton to large predators, and eventually humans. Finally, ALDFG represents a significant economic problem, causing vessel damage, loss of commercially valuable species, and high clean-up and removal costs. Therefore, effective management and mitigation of ALDFG is crucial for the health and sustainability of marine ecosystems.

ALDFG poses a significant risk to autonomous underwater vehicles (AUVs), unmanned surface vehicles (USVs), and diving operations. [12][13] For AUVs and USVs, ghost nets and other debris can cause entanglement, damage sensors, and

thrusters, and disrupt research and monitoring missions. This interference can result in high repair costs and loss of equipment. In diving operations, ALDFGs present a considerable risk to diver safety. Divers can become trapped in the ghost nets, leading to underwater emergencies, including running out of air and drowning. In addition, the presence of ALDFG can hinder visibility and navigation, increasing the danger during diving operations. Therefore, the presence of ALDFG in the marine environment not only negatively impacts marine life and habitats, but also poses a danger to the technology and human safety involved in underwater and surface operations.

The objectives are to find methods of detecting targets, so the state-of-the-art analyzed various studies in the area of finding objects underwater, which allowed us to define what sonar would be the best approach, they could require approaching the same target from various angles so that it would be easier to verify what the Target was, but in terms of detection, it is the most viable option for finding targets.

The next objective was to define the type of sonar to be used. Various types of sonar were analyzed and a set of tests were conducted in a controlled environment to confirm the decision on the type of sonar. Once the feasibility of the situation had been verified, the next step was real-life testing. The next objective is to analyze the results and then define the best data processing methods. As we saw in the state-of-the-art, neural networks were the best option, so it was decided that it would be necessary to run the acquired data through various options of neural networks, thus possibly defining which were the most viable to use.

Ultimately, the final result would enable USVs and AUVs to gather information in real-time so that they can avoid obstacles and also determine the type of obstacle, allowing for its retrieval if necessary.

## 3.2 Possible targets to be found

### 3.2.1 Art of fishing in Portugal

[14]Taking into consideration from PORDATA, this information is from 2022 in terms of "Fish caught, and total from main species". 121000 tons of fish were caught in 2022, more than 50% of this catch was sardines 24 thousand tons, mackerel with 21 thousand tons, and horse mackerel with 18 thousand tons. From this data, we can infer that the mackerel and sardine fisheries have the largest volume of catches in Portugal and through [15] *Direção Geral Recursos Naturais, Segurança e Serviços Marítimos* (DGRM), that means Directorate General for Natural Resources, Safety and Maritime Services, decree no. 218/2023 we can see that the gear used in this case is purse seine fishing.

Types of fishing nets used in Portugal are, for example, Gillnets that can be either on the surface or the sea floor. In both cases, buoys are used to signal. The

main difference is that the one on the sea floor has anchors to hold it to the sea floor, and the other only has enough weight to keep it upright.

Trammel nets are bottom-set gillnets made up of three overlapping mesh panels, the outer two identical and with large meshes, and the inner one, higher up, has a smaller mesh. These nets are forbidden on the high seas.

In the case of bottom-set gillnets and trammel nets, the maximum cumulative length of each is 5000m. Drift nets can be up to 500m long and 10m high. Purse seines are used in the open sea to target dense volumes of fish, is a vertical net curtain that surrounds the school of fish, the bottom of which is then drawn together to enclose the fish, rather like tightening the cords of a drawstring purse. In open water is generally considered to be an efficient form of fishing, and has low levels of bycatch (accidental catch of unwanted species). But to ensure that they leave enough fish in the ocean to reproduce, for that is use a mesh size large enough to allow smaller fish to swim free. Purse seines are between 300 and 800m long and 60 to 150m deep. Exists also the trawling nets, these have a cone-shaped body and closed cod end that holds their catch. Trawl nets are designed to be towed by a boat through the water column (midwater trawl) or along the sea floor (bottom trawl). These can be up to 10m long and 2m high, and each fisherman can use up to 8 nets.

Depending on the species to be caught, different mesh sizes can be used:

1. 50-59mm;
2. 60-79mm;
3. 80-99mm;
4. $>=$ 100 mm.

Drift nets can be used with mesh sizes between 35 and 40 mm. The minimum mesh size for gillnets is 110 mm. Purse seines have a minimum mesh size of 16mm.

Trap fishing in Portugal is mainly done using these 3 types of traps.

1. Ceramic and sometimes plastic buckets;
2. Metal frames with mesh all around;
3. Net with hoops to make a flexible trap;

In this case also exists different mesh sizes depending on the species that are pretend to be caught.

1. 8 mm a 29 mm;
2. 30 mm a 50 mm;

3. > 50 mm;

Cage traps 8 to 29 mm from the shoreline, for vessels up to 9 m, or 30 to 50 mm and > 50 mm, from the shoreline, for boats up to 9 m, and out to 1 mile from the shoreline, for other boats.

## 3.3 Detection method

### 3.3.1 SONAR

SONAR, which stands for SOund NAvigation and Ranging, is an instrument that uses sound waves to investigate the ocean.[16] Scientists primarily use sonar to create nautical charts, detect underwater hazards to navigation, find and identify objects in the water column and on the seafloor, such as shipwrecks, and map the seafloor itself. It is preferred in oceanography because sound waves travel farther in water than radar and light waves. The sonar system includes physical sound sensors known as transducers. Scientists can use a single transducer or a group called a transducer array, which can be mounted on various platforms like remotely operated vehicles, autonomous underwater vehicles, ships, or platforms such as a towfish or glider.

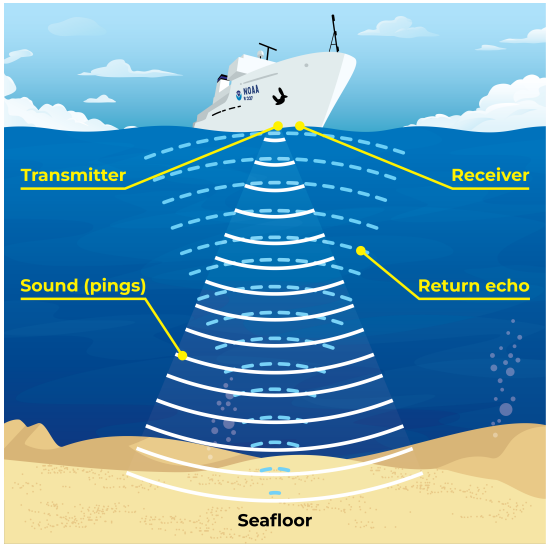



Figure 3.1: Sonar Pings illustration.
[17]

#### Types of Sonars

The Sonars are divided into three particular groups[18], one of them is the searchlight sonar provides a 360-degree view of underwater surroundings around the vessel. The sonar displays information such as fish schools and sea currents by continuously rotating its sensor. It's similar to using a flashlight to search the area near the

vessel. This type of sonar is commonly installed on small fishing boats and has become increasingly popular on recreational boats in recent years. The Sonar is usually represented on the screen as a point at the center (the boat), surrounded by echoes in a circle around the ship.

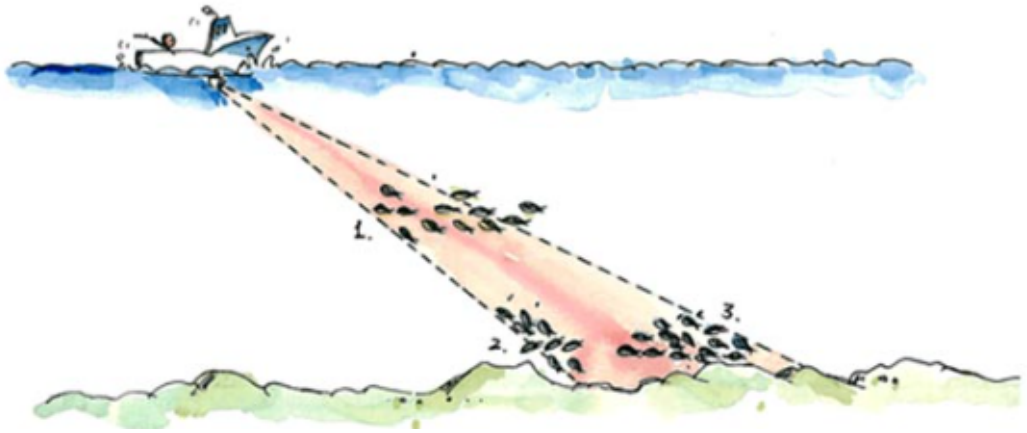



Figure 3.2: Searchlight Sonar.
[18]

The scanning sonar emits ultrasonic waves in a full 360-degree sweep around the ship, detecting and displaying the returning echoes immediately. This method is much quicker than traditional scanning sonar, allowing instant detection of the surrounding area. It enables real-time tracking and assessment of fast-moving fish like bonito and tuna.

1. Full-circle scanning Sonar: The system sends out ultrasonic waves in every direction around the ship in a single pulse, enabling it to detect and display all surrounding objects immediately.

2. Half-circle scanning Sonar: Ultrasonic sensors mounted on the ship's underside instantaneously scan a 180-degree area beneath the vessel.

The last one is Sector scanning Sonar, similar to the searchlight sonar, but the sector scanning sonar operates in 45-degree increments. This allows it to be 4 to 7 times faster than the searchlight sonar.

[19]During the analysis of various studies and more theoretical components, it was verified that SONARs have some aspects to be take in concerned. When choosing a sonar, several important aspects must be considered to ensure it meets the specific needs of your project or application. Here are some key factors to consider, like frequency is an important consideration. High frequency provides higher resolution and detail, making it ideal for accurately identifying small objects and detailed structures on the seabed. However, it has a lower penetration capability and range in water. Low frequency offers greater range and better penetration, making it useful for covering larger areas and deeper waters, all though the resolution is lower.

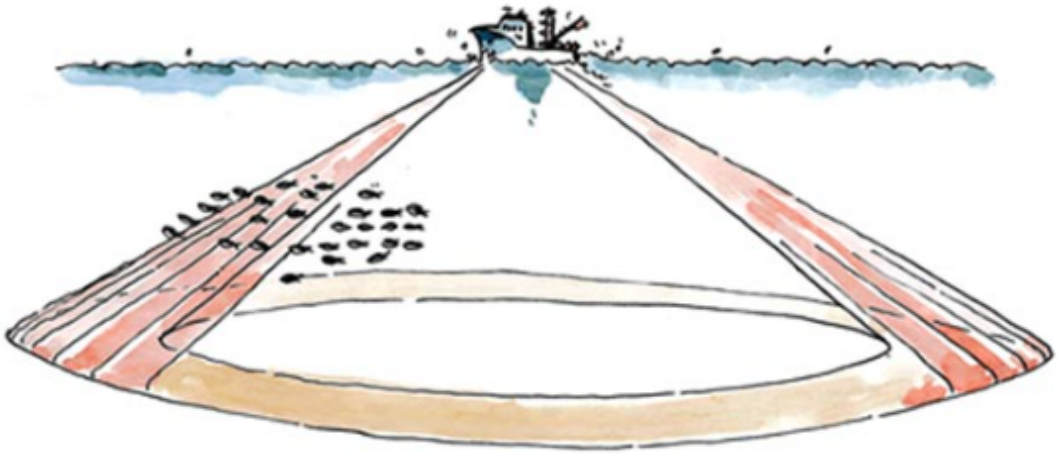


Figure 3.3: Scanning Sonar.
[18]

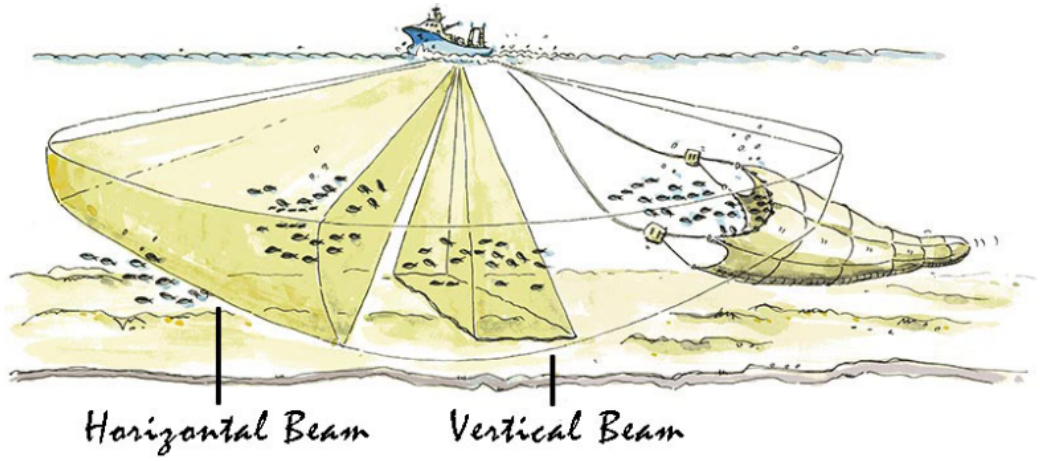



Figure 3.4: Sector Scanning Sonar.
[18]

There are also sub-types of sonar to consider. Side-scan sonar is ideal for mapping large areas of the seafloor and is useful for detecting objects and characterizing marine habitats. Multibeam sonar provides a three-dimensional view of the seafloor, which is beneficial for detailed topographic and bathymetric surveys. Single-beam sonar is simpler and more economical, making it suitable for basic depth measurements. Range and beam coverage are also crucial. A wider swath can cover more area in less time, but it may compromise resolution and accuracy. The distance at which objects can be detected and captured is important, so finding the right balance between these two aspects is essential to achieve high detection rates with the necessary quality to accurately identify what is being observed.

Resolution is the sonar's ability to distinguish between objects that are close to each other, and higher resolution is essential for tasks that require precise details.

Portability and ease of use are also factors, particularly in field operations. More compact and easy-to-handle systems are preferable, and in most cases, a viable solution for AUVs is a compact sonar that can be applied without affecting the dynamics of its movement. Environmental conditions must also be considered, including the sonar's ability to operate in different conditions, such as varying levels of salinity, water temperature, and the presence of suspended particles. In some cases, it might be important to have systems that help increase visibility. Cost is another determining factor, and it's important to balance the cost of the sonar with the required functionalities and performance. Integration with other systems is essential as well, particularly with navigation and positioning systems like GPS, which are critical for accurate mapping and object detection. Lastly, maintenance and support should be taken into account. The availability of technical support and the ease of maintenance of the equipment are important, as sonars that require less maintenance and have good technical support can be more reliable in the long term. These aspects will help you choose the most suitable sonar for your operations.

**Acoustic imaging sonars.**

Acoustic imaging sonars, also known as imaging sonars or high-definition sonars, use sound waves to create detailed images of the seabed, submerged objects, and the water column.[20] Here is an explanation of how they work and their main advantages. Sonar works by emitting high-frequency sound pulses, typically in the kHz to MHz range, from a transducer. These pulses travel through the water and are reflected back to the sonar when they encounter an object or the seabed. The time it takes for the pulse to return is used to calculate the distance to the object. Using a set of transducers, or an array, the sonar captures echoes from different directions, allowing it to create detailed two or three-dimensional images of the underwater environment. The data received is then processed to create a visual image, similar to that of a camera, but generated through sound waves. The software associated with the sonar interprets these signals and builds the image, highlighting different materials and structures based on their acoustic properties. Acoustic imaging sonars offer several advantages. They are capable of generating detailed, high-resolution images of the seabed and submerged objects, allowing for precise identification of small details and structures. These sonars perform well even in adverse conditions, such as turbid waters or areas with poor visibility, where optical methods like cameras would fail due to a lack of light or the presence of suspended particles. They are excellent for detailed mapping of submerged areas, including the identification of marine habitats, detecting wrecks, and inspecting underwater infrastructure. Acoustic imaging sonars are also useful for detecting specific objects, such as lost fishing nets, mines, or other types of debris, which is crucial for rescue operations, underwater archaeological research, and maritime security. Their versatility allows them

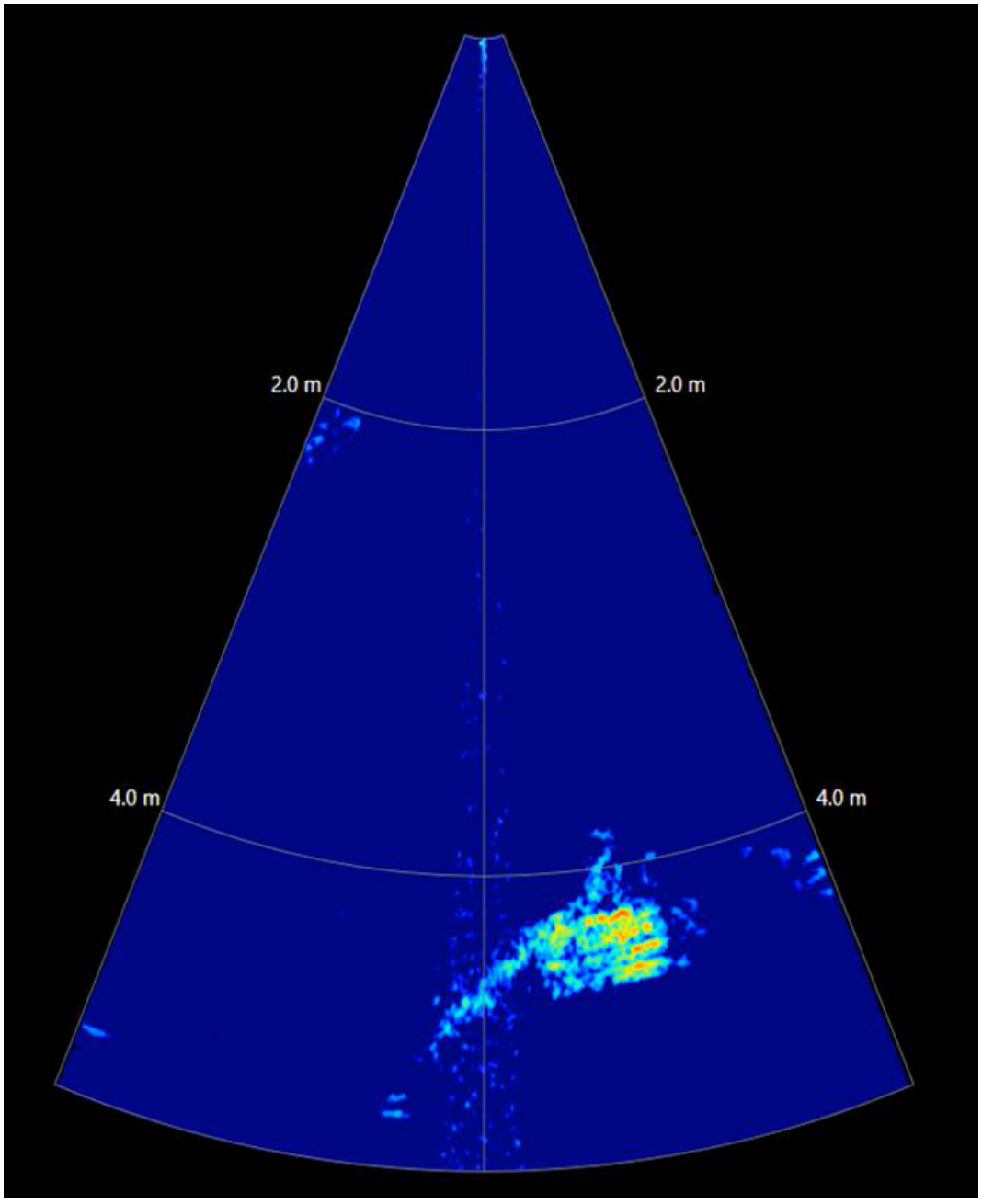



Figure 3.5: Palette saw for blueview from water tank using in tests.

to be used on various platforms, including autonomous underwater vehicles (AUVs), research vessels, and diving equipment. Additionally, these sonars can cover large areas quickly, making them efficient for wide-area surveys and saving time and resources. They are capable of operating at various depths, from shallow waters to deeper areas of the ocean, offering flexibility for different types of missions. Some acoustic imaging sonar systems can even generate three-dimensional data, providing a volumetric view of the underwater environment, which is valuable for detailed

analysis and intervention planning. Acoustic imaging sonars are powerful and versatile tools that provide a detailed view of the underwater environment, facilitating a wide range of commercial, scientific, and security applications.

**Analyzed sonars.**

To carry out the tests and validate the practical content, an analysis was made of the sonars present in the laboratory that could meet the project's needs. A market analysis was also carried out to find a possible equivalent sonar within the range of possible values for the project. First was analyzed M3 multi-frequency from Kongsberg that works with the following settings and specifications[21]:

1. Frequency: 700kHz to 1400kHZ
2. Minimum detectable: 0,2m to 140m
3. Resolution: 1cm
4. Field of View: 120º
5. Frame Rate: 40 fps

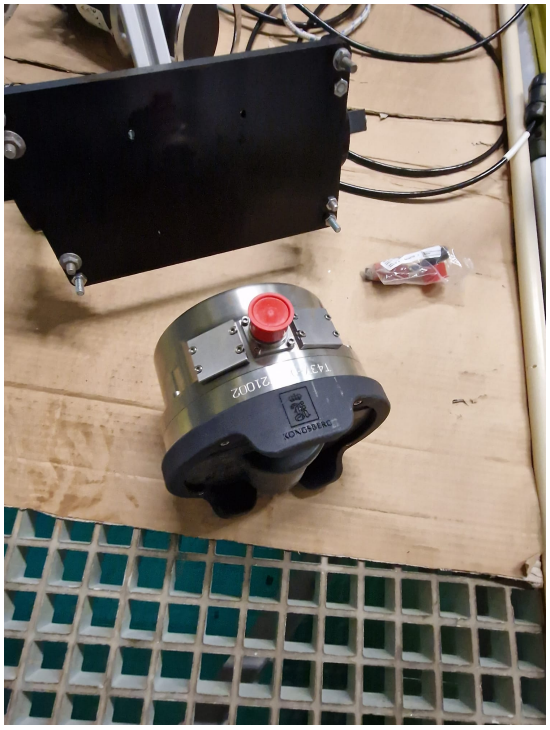


Figure 3.6: M3 multi-frequency for Kongsberg.

[22]Second was analyzed BlueView MB2200 - FLS that works with the following settings and specifications:

1. Frequency: 900k to 2250kHz
2. Minimum detectable: 2 a 60m e 0.5 a 7m
3. Resolution: 1,3cm e 0,5cm
4. Field of View: 130º
5. Frame Rate: 25 fps

Third was analyzed Tritech Gemini 1200iK that works with the following settings and specifications:

1. Frequency: 720k to 1200kHz
2. Minimum detectable: 0,1 a 120m e 0.1 a 5m
3. Resolution: 4mm e 2,4mm
4. Field of View: 130º
5. Frame Rate: 65 fps

This tritech sonar[23] was analyzed as a possible solution, considering it also works in imaging, and was capable of being applied to an AUV like the others and even came with support for this aspect, it is within the same range of values as the other options. The previous version of the sonar in question has even been used in various projects, including some of those mentioned in the state of the art. There is even a public database of data collected by the 720i version sonar for use in studies of the same order and for training neural networks. From the results analyzed, this sonar did not provide significantly better results to compensate for the laboratory's investment, considering that there were already two other options.

# Chapter 4

# Data processing methods

Based on the research conducted in the previous chapters, a specific architecture and structure for the implementation were decided. First, the selection of sonars to be used was made, considering only the sonars available at the LSA Laboratory: the M3 Sonar from Kongsberg and the Blueview from Teledyne Marine. Then, the targets were defined to be used in a controlled environment to test which of the two sonars would be the better option for the solution. After that, the data was collected in a controlled environment, specifically in the laboratory tank, which measures 10 meters in length, 6 meters in width, and 5 meters in depth. Once the sonar was selected, it was necessary to move on to testing in a real environment. For this, the selected sonar was applied to an AUV or USV, and the same targets were used again to extract new data, allowing for further analysis of the collected data.

The data had to be extracted from ROS Bag format. Initially, the data was extracted in color and then converted to Greyscale Image. Afterward, all the data would be used to train the Neural Network.

## 4.1 Neural Networks

Deep learning and neural networks are important topics in computer science and the IT sector because they currently offer the best solutions for data processing. Neural networks are used in machine learning, which refers to a category of computer programs that learn without definite instructions. Specifically, neural networks are used in deep learning an advanced type of machine learning that can conclude unlabeled

data without human intervention. The studies discussed above have led to the conclusion that there are various methods for processing information, but well-trained neural networks can provide a high degree of certainty and thus determine targets.

A neural network, or more precisely an ANN,[24] is a subset of machine learning that mimics the network of neurons in a brain to process complex data inputs. These networks can adapt, learn, and improve over time, forming the backbone of what we know as Artificial Intelligence. These neurons work together to learn patterns from the data and make predictions or classifications.

The ability of a neural network to solve or rather "think" has revolutionized and evolved computer science. This intelligent solution is capable of interpreting data and improving context. There are four critical steps that neural networks need to work effectively. It is training that allows neural networks to 'remember' patterns, in other words, if an unknown pattern appears to the computer, it will associate the pattern with the closest match present in its memory, then it will classify or organize the data or patterns into predefined classes, after which it groups and identifies a unique aspect of each instance of the data to classify it even without any other context present, thus generating the prediction capacity that will then produce the expected results using a relevant input, even when all the context is not provided beforehand. These neural networks usually require a high level of processing power to perform these functions accurately, especially in near-real time. This can be accomplished by using several processors with good capacity operating in parallel and organized in layers.

The first part of the neural network process begins with the first layer, which receives the raw input data, in the same way that a human's optic nerves receive visual input, then each successive layer receives the results of the previous one, after which it continues until the final layer has processed the information and thus generates output results. Each individual processing node contains its own database, including all the data that has ever passed through it and the rules with which it was originally programmed or developed and improved as more data passed through over time. These nodes and layers are all highly interconnected. The training process begins as soon as a neural network is structured for a specific application, this training can follow a supervised or unsupervised approach. In the first part of training, the network receives data that has already been classified as correct in order to train more efficiently.Unsupervised training occurs when the network interprets inputs and generates results without instruction or external support. Adaptability is one of the essential qualities of a neural network. This characteristic allows machine learning algorithms to be automated as they learn from their training and subsequent operations. Learning models are fundamentally centered on the weight of inputs, where each node assigns a weight to the input data it receives from previous nodes. Inputs that prove instrumental in deriving the correct answers are given greater

weight in subsequent processes.

Having adaptability, neural networks take advantage of various principles to define their operating rules and make determinations; fuzzy logic, gradient-based training, Bayesian methods and genetic algorithms all play a role in the decision-making process at the node level. This helps individual nodes decide what should be sent to the next layer based on the inputs received from the previous layer. Basic rules about object relationships can also help ensure higher quality data modeling, the more correct data is given at the start and the more rules and indications the network is given, the better the results will be and the faster and more efficient the processing will be.

Adding rules is not always beneficial to the system, as these rules can lead to incorrect assumptions when the algorithm tries to solve problems unrelated to the rules. Loading the wrong set of rules can lead to the creation of neural networks that provide irrelevant, incorrect, useless or counterproductive results.

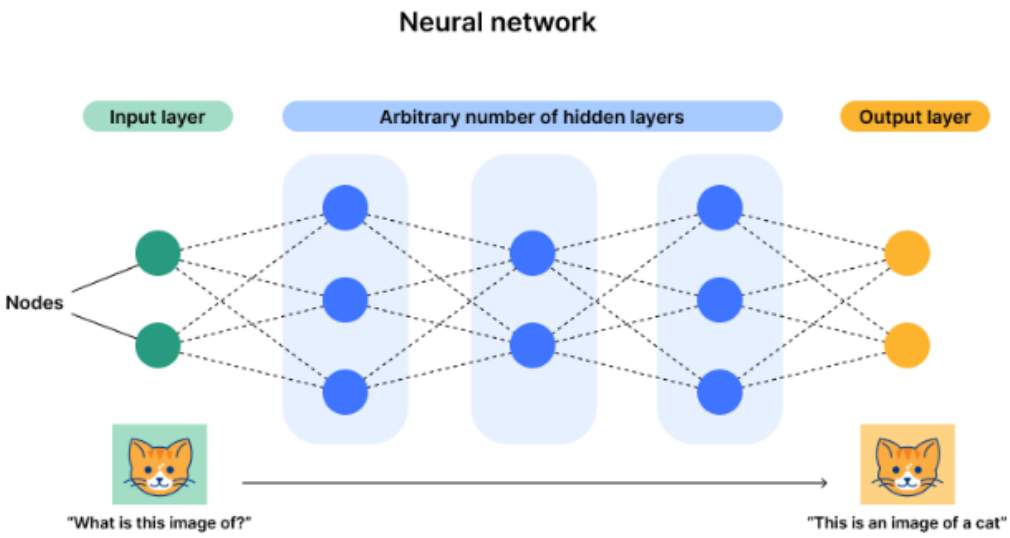



Figure 4.1: Example neural networks.
[25]

### 4.1.1 Types of Neural Networks.

Neural networks can typically be divided into a few types[26], but in one way or another, the rest end up being derived from these more standard types. Standard artificial neural network, pattern networks are made up of multiple hidden layers, with several interconnected processing nodes, or neurons, which communicate with each other through the process of synapses. Each neuron (node) receives inputs from several other neurons and produces an output which is passed on to other neurons in the network. The number of connections, and the robustness of nodes of the rules, generate a weight for the features and thus influence the information

throughout the system, which is usually trained using a technique called backpropagation. Backpropagation involves adjusting the weights of the connections between the nodes based on the network's performance on a set of training data. As a result, it generates the ability to perform complex tasks such as pattern recognition and predictions.

CNN these are a type of neural network well suited to image classification tasks, internally they have a composition of convolutional layers, which extract features and pooling layers, to reduce the size and quantity of the feature maps. A convolutional layer applies a set of filters to an image that is placed at its input, resulting in a set of feature maps that represent different patterns in the image, and also has a pooling layer that is used to reduce the size of the convolutional layer's output by sampling the feature maps. This type of algorithm on the images placed in its input as a series of filters designed to detect the various specific characteristics of the input. Its output is fed into a fully connected layer, which thus produces the final result, which will be the ability to classify the image. The first layer consists of a convolutional layer that uses a set of filters learnable from the input image to extract features. The feature maps fed by pooling layers thus reduce the width and height of the maps, while maintaining the depth, which is the number of channels. The output of the convolutional layers is flattened into a one-dimensional vector, which is passed on to the connected layers.CNNs learn to detect complex patterns in images and are therefore one of the most widely used systems for object detection.

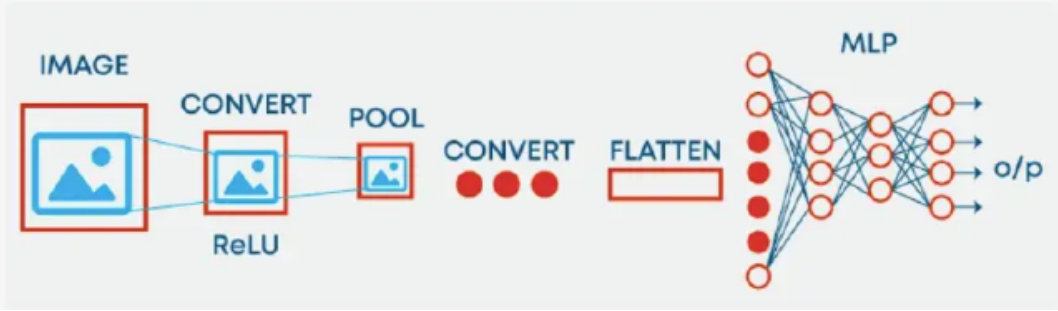



Figure 4.2: Most common struct of Convolutional neural network.
[26]

*Recurrent neural network* (RNN), are deep neural networks that can store information from previous calculations and then process this information sequentially. It is a type of neural network where the output of the previous time step is used as input for the current time step, together with the new input, thus creating a sequence element, allowing the RNN to learn the dependencies between them. This allows RNNs to be very adaptable for tasks such as speech recognition and machine translation, where understanding the context is crucial. RNNs can also be used for predicting time series, such as stock market movements or weather patterns. However, training RNNs can be difficult due to the vanishing gradient problem, which

can cause the error to propagate and increase as more data is accumulated.

The *Long short-term memory* (LSTM) is a type of network with the mechanism and capacity to store information for long periods, thus giving it the ability to learn from the experiences that have been passed on to it over time, so it is possible to say that the LSTM is a type of network used to process data sequentially. As they can remember long-term dependencies, this makes them suitable for tasks such as machine translation and speech recognition. One of the main benefits of LSTMs is that they can be trained on relatively small data sets. This is because LSTMs can learn from context, which allows them to generalize better than other types of neural networks. LSTM networks are similar to recurrent neural networks (RNN), but they have a forgetting gate that allows them to forget information that is no longer needed.

Stacked Autoencoders,[27] autoencoders are a type commonly used to learn efficient representations for data. A stacked autoencoder is a type of autoencoder where a neural network encodes the input, and then the output of that network is passed as input to another encoding network. By stacking several of these autoencoders on top of each other, it is possible to gain knowledge so that the network can process increasingly large and complex representations of the data, which can be beneficial for tasks such as image recognition, where a deep understanding of the data is required to make accurate predictions. This process can be used to start other types of networks.

Variational autoencoders This type of neural network is widely used for dimensionality reduction and generative modeling. The main idea behind variational autoencoders is to learn a latent representation of the data that is of a lower dimension than the input. The difference between variational autoencoders and stacked autoencoders is that stacked autoencoders learn a compressed version of the input data, while variational autoencoders learn a probability distribution. Here is a quick diagram representing the most common type of neural networks including ANN, CNN and RNN.

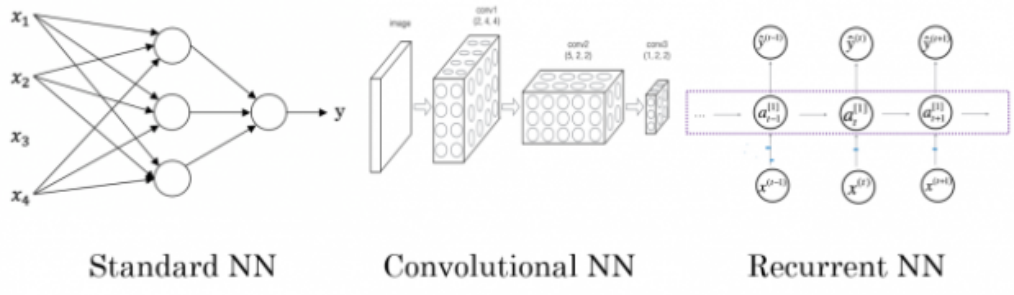



Figure 4.3: Most common type of neural networks.
[27]

The fact that neural networks are "black box" in nature is a disadvantage, as you have no idea how the network generated that result. Although there are libraries that simplify the creation of neural networks, there are situations where you need more control over the details of the method, such as when using machine learning to solve a challenging problem that no one has ever faced before. According to their structure, artificial neural networks require processors with parallel processing power.

[28]There is a model called VGG16 neural network model[29], this model is a deep convolutional neural network model that was developed by Karen Simonyan and Andrew Zisserman from the Visual Geometry Group (VGG) at the University of Oxford. This model was public in a 2014 paper, "Very Deep Convolutional Networks for Large-Scale Image Recognition." In particular, this speaks about a VGG model that has 16 layers that have weights, including convolutional layers, connected layers, and pooling layers. The most important feature of VGG16 is a model consists of several convolutional layers stacked on top of each other. Each layer typically uses a 3x3 filter with a stride of 1 pixel. This small receptive field allows the model to learn the finer details of the image effectively, other important aspect is that have 16 weight layers, VGG16 is considerably deeper than many previous network architectures. This depth is crucial for learning a wide hierarchy of features, from simple to complex. Have a ReLU (Rectified Linear Unit) activation function is used. This non-linear activation function helps introduce non-linear properties into the network, enabling it to learn more complex patterns. This network is using max pooling layers intermittently between convolutional layers. The pooling layers perform spatial downsampling, reducing the dimensionality of the feature maps, thus reducing the number of parameters and computations in the network. Towards the end of the network, there are three fully connected layers. The last layer in VGG16 uses a softmax activation function. This function is used to turn the logits (raw prediction scores) from the last fully connected layer into probabilities by comparing the exponential growth of each output in comparison to the sum of all exponentials of the outputs.

This model is known for simplicity and good performance and sets goals in image classification and tasks, because have a god straightforward architecture that can be easily replicated and adapted for a variety of image recognition tasks. Its architecture has become a standard for many CNN-related tasks and research.

# Chapter 5

# Implementation

Based on the research conducted in the previous chapters, a specific architecture and structure for the implementation were decided. First, the selection of sonars to be used was made, considering only the sonars available at the LSA Laboratory the M3 from Kongsberg and the Blueview from Teledyne Marine. Then, the targets were defined to be used in a controlled environment to test which of the two sonars would be the better option for the solution. After that, the data was collected in a controlled environment, specifically in the laboratory tank, which measures 10 meters in length, 6 meters in width, and 5 meters in depth. Once the sonar was selected, it was necessary to move on to testing in a real environment. For this, the selected sonar was applied to an AUV or USV, and the same targets were used again to extract new data, allowing for further analysis of the collected data. The data had to be extracted from *Robot Operating System* (ROS) bag format. Initially, the data was extracted in color and then converted to greyscale. Afterward, all the data would be used to train one neural network.

## 5.1 Test planning

For this solution, tests were conducted to compare the two options. These tests took place, first in a controlled environment within a tank located in the LSA, affiliated with ISEP and INESC TEC. The second part of the tests was conducted on the sea near the port of Leixoes, Matosinhos.

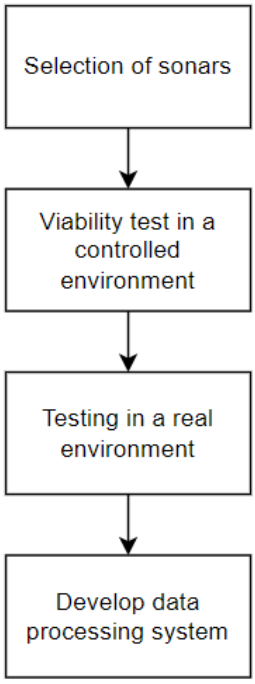



Figure 5.1: Architecture for project implementation.

In the controlled environment, it was decided that the following topics should be tested and analyzed. The first resolution test involves using a known target to analyze the system's ability to capture and perceive to determine the target. The second was the sensitivity tests that evaluate the sonar's sensitivity by adjusting it to detect targets of different sizes and depths. The third range test analyzes and determines the minimum and maximum recommended range to detect the target. Next was the discrimination test determining the sonar's ability to distinguish between various targets, and the last was the performance test at different frequencies which consisted of testing the sonar at different frequencies to determine which is the most effective for specific water conditions and targets.

The second scenario was to run the same tests with the same targets in a real environment, all laboratory tests need to be repeated in an offshore context as well, to compare information and data possibly. It also considered extra tests like noise rejection tests that test the sonar's ability to filter or eliminate unwanted noise. The second was the depth test to verify the depth at which the sonar can operate accurately. The last was the robustness test which exposed the sonar to adverse conditions such as murky water, extreme temperatures, or turbulent waters to assess its ability to operate reliably. It was decided that the tests in the real case would be carried out on an AUV and, if possible, with the two sonar options present in the laboratory. It was also planned to carry out the tests with a USV, in both cases keeping scenarios in which the target was seen from various perspectives so that it

would be possible to check the contour of the whole object.

## 5.2 Laboratory tests

The tests were carried out in the LSA laboratory, where the tank was about 5 meters deep with 6m wide and 10m long dimensions. It had a movable platform at one end, to which 2 sonars were placed in the middle of the platform, making the sonar approximately 1.25m from the end, leaving a length of 8.75m. The sonars were placed at a depth of 0.5m and with a $30^{o}$ inclination towards the bottom.

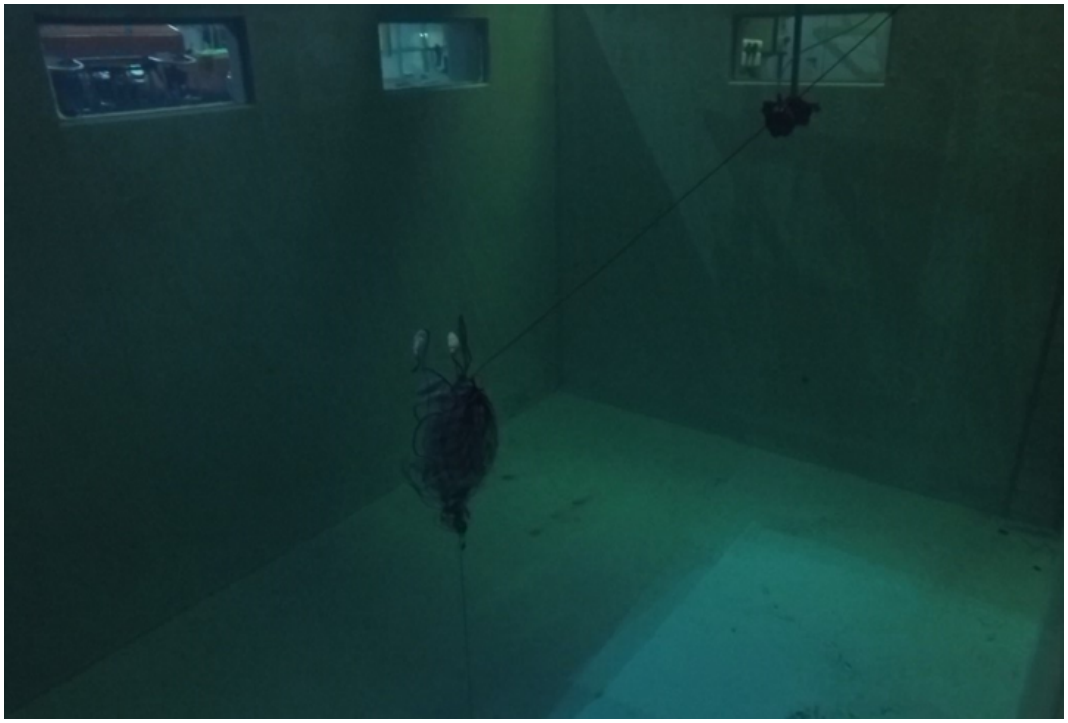


Figure 5.2: Ball of netting.

The first test was carried out with a ball of netting(Figure 5.2), followed by tests with a wooden stick, a plastic water bottle, a tire, 2 different nets with different configurations, 2 fishing traps, and finally octopus hunting traps. The targets suspended in the water column were 2.5m deep, centered in the tank and facing the sonar. The fishing traps were placed at the bottom of the tank centered on it, and the octopus traps were placed together in the same way, then in the water column, and after that they were placed like a curtain, individually attached to the same rope.

As you can see in the picture, we had to put weight on some of the targets to anchor them so that they would stay in place. Parameters used in the M3 Kongsberg frequency at 1400kHz, with a distance of 5m in high-frequency imaging mode, for the blueview, tests were carried out at a frequency of 2250kHz at a distance of 6m

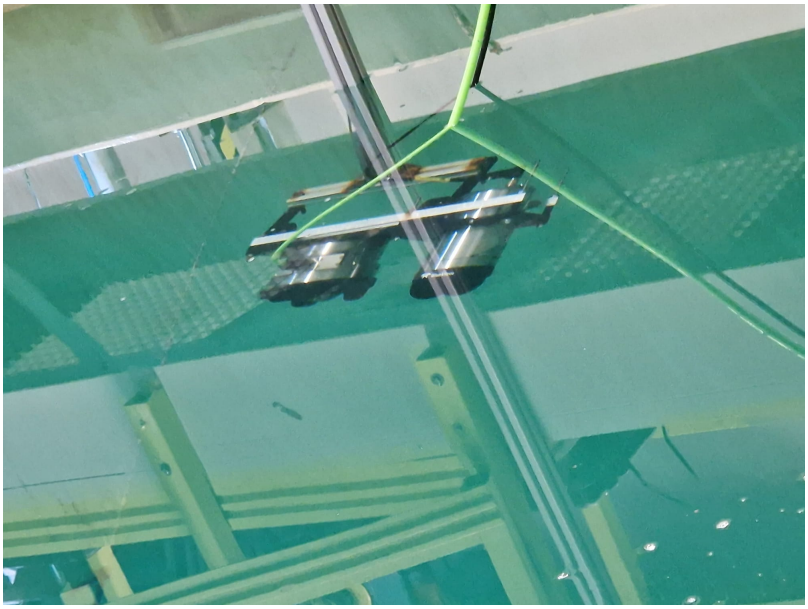


Figure 5.3: Sonars applied to the support at 30º.

with the field of view at 45º in imaging mode. This test with the grid allowed us to analyze whether or not the grid became more visible the closer we got to it, to try and check whether the inner squares were visible.

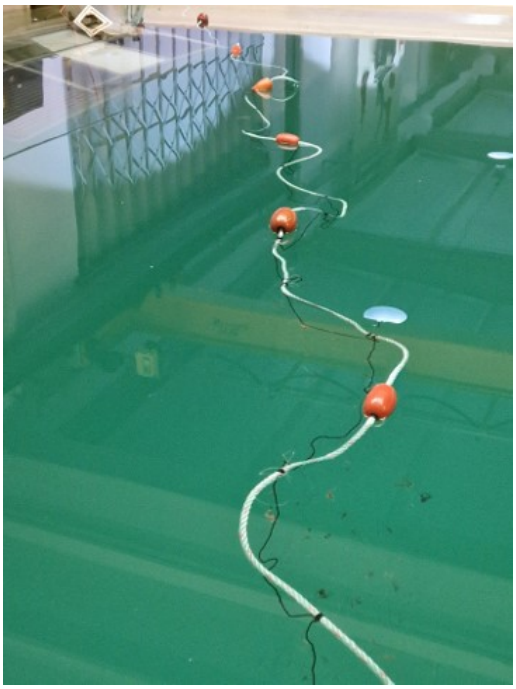


Figure 5.4: Blue net with horizontal application.

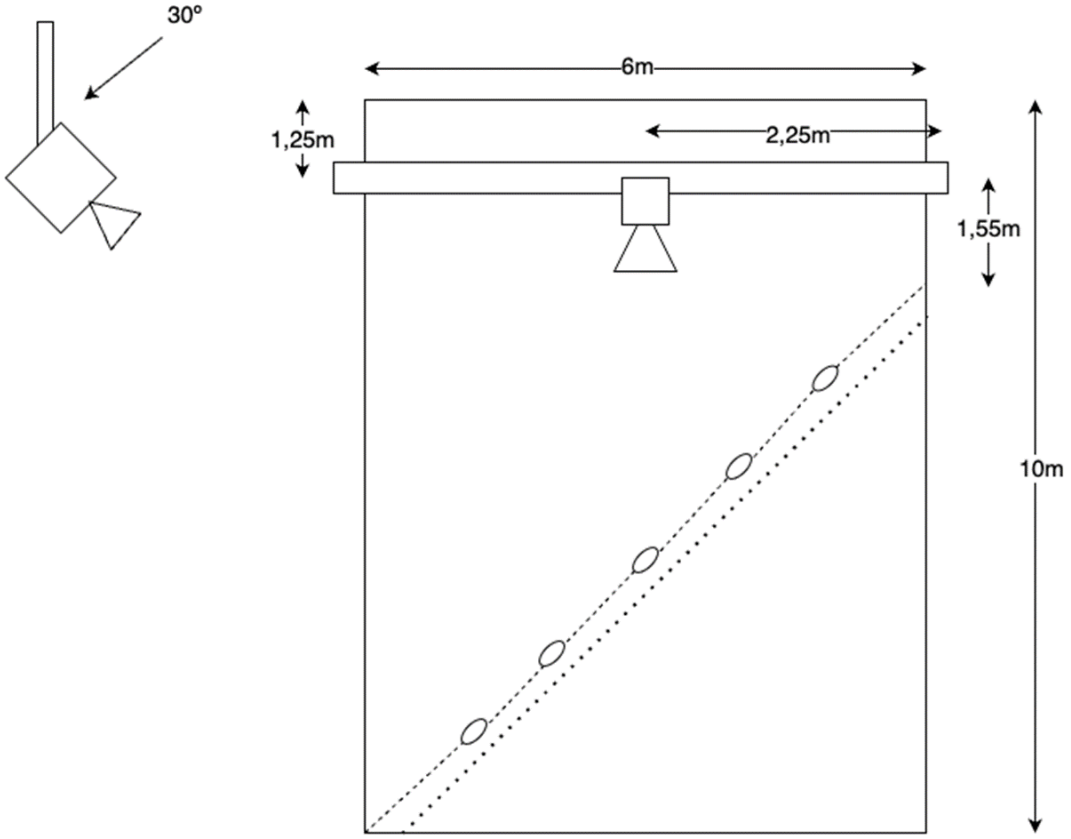



Figure 5.5: Blue net application scenario.

During the practical tests in the laboratory, there were sandbags at the bottom of the tank from other projects, which can also be seen in the images. During the tests, the nets were tested with various configurations and were even moved to allow the net's movement in the water column to be analyzed. While testing the targets, images were taken of each sonar and then compared to check which of the two would be more viable for testing on the high seas if it wasn't possible to go ahead with a practical test with the two solutions and to try to draw some conclusions about the next approaches.

Conclusions related to the M3, all the objects were detected at a frequency of 1400KHZ at a distance of approximately 4m, from a distance of 5m we lost a little resolution and we were able to visualize the sandbags and the cable at the bottom of the tank better than our objects. Both nets were detected. The black net with its denser center wire was more detectable than the blue one. The traps were not detected, either in the water column or at the bottom of the tank. The octopus traps were not detected either all together or in a chain arrangement.

Conclusions related to Blueview, all the objects were detected at a frequency of 2250KHZ at a distance of 4m where from 5m away we lost the intensity of the

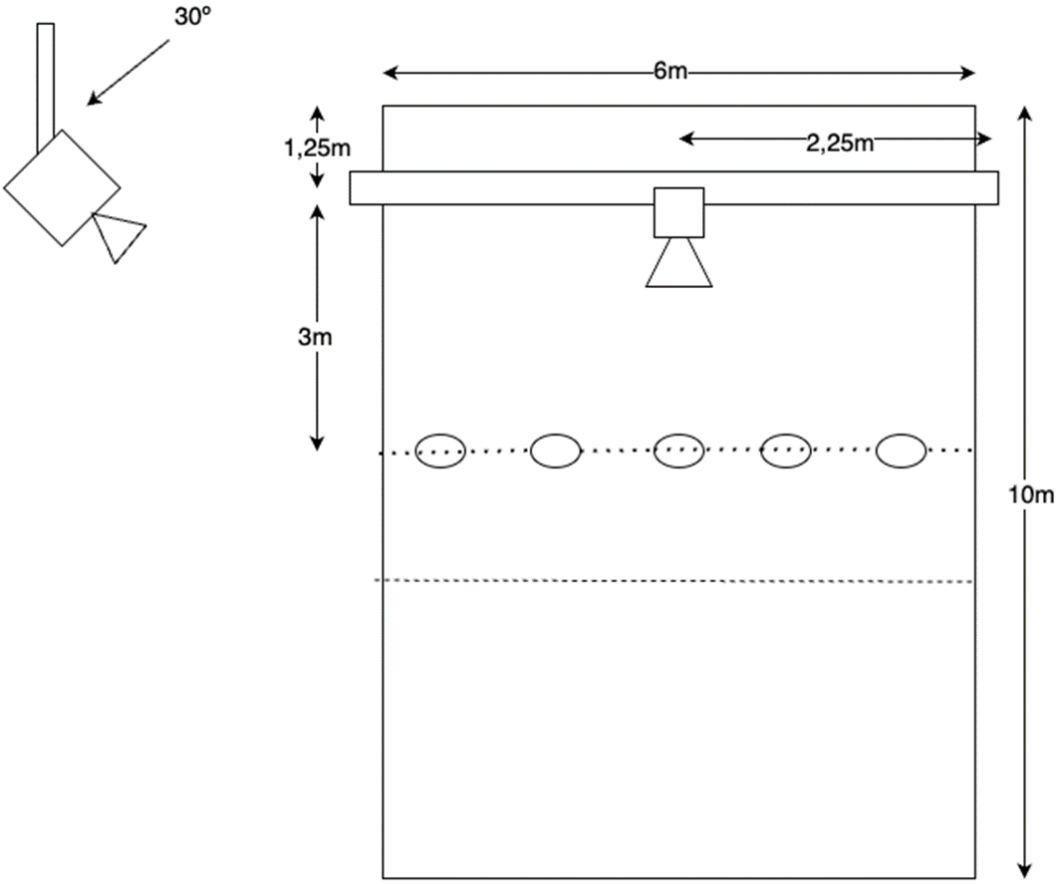



Figure 5.6: Black net application scenario.

objects, but it was still possible to detect them. Both nets were detected. The traps were not detected in the water column, but obstacles at the bottom of the tank were 'detected'. The octopus traps were 'detected' in the water column. There's a big difference between the two sonars, on the one hand, we have the Blueview, which can pick up the intensity of obstacles much better, but with worse visual acuity, and on the other hand we have the M3, where we can visualize objects better but have a big loss in detection.

Both options are valid, always depending on the end goal we want to achieve: if it's related to obstacle detection, the Blueview will be the better option, if it's more related to visual object detection, the M3 will be the better option.

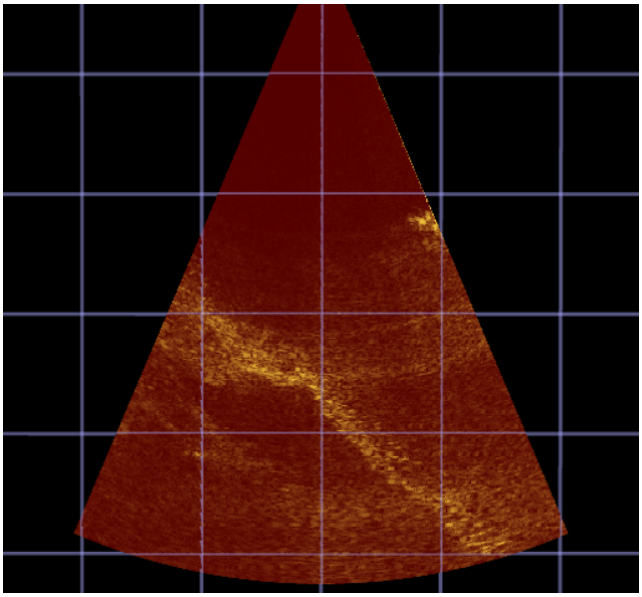


Figure 5.7: Result from M3 diagonal fishing net.

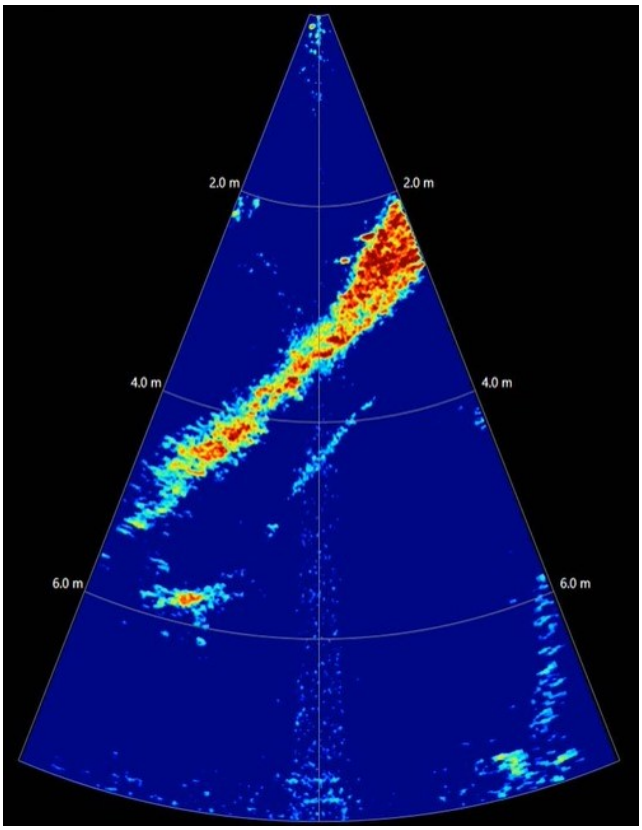



Figure 5.8: Result from Blueview diagonal fishing net.

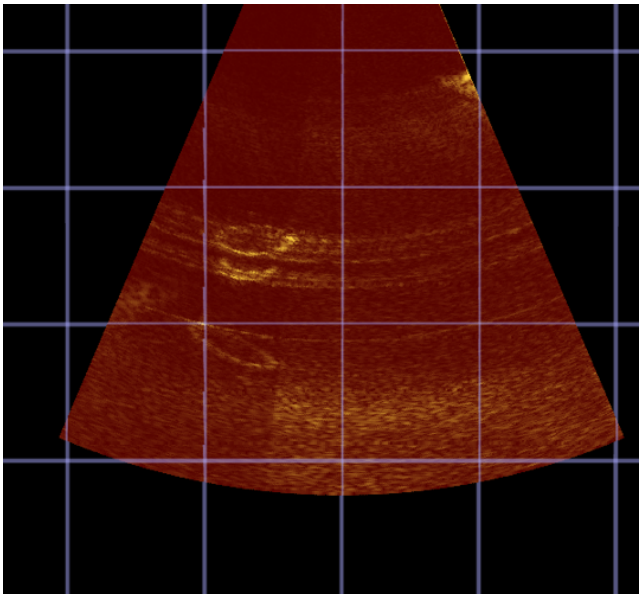


Figure 5.9: Result from M3 for the tyre.

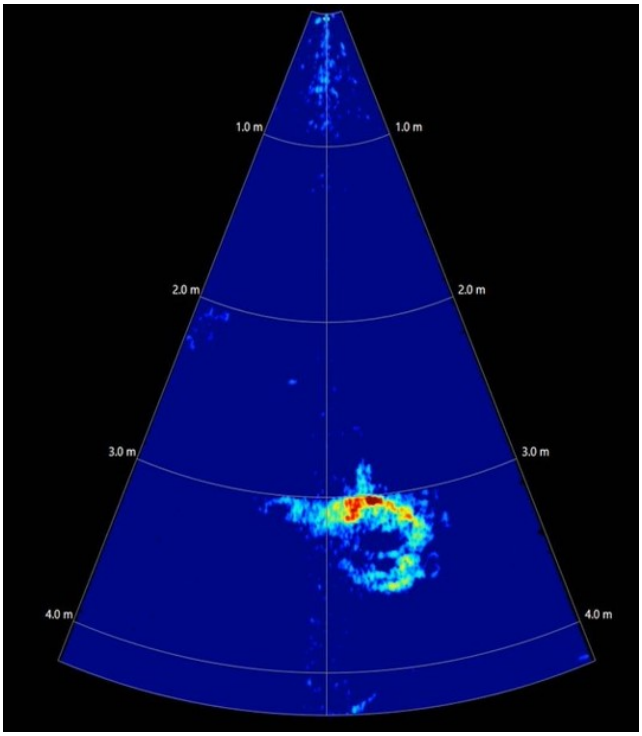



Figure 5.10: Result from Blueview for the tyre.

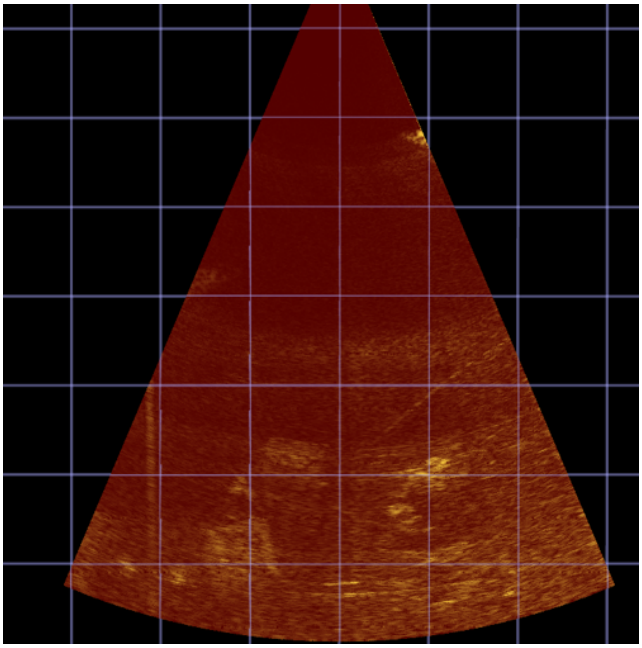


Figure 5.11: Result from M3 for fishing traps.

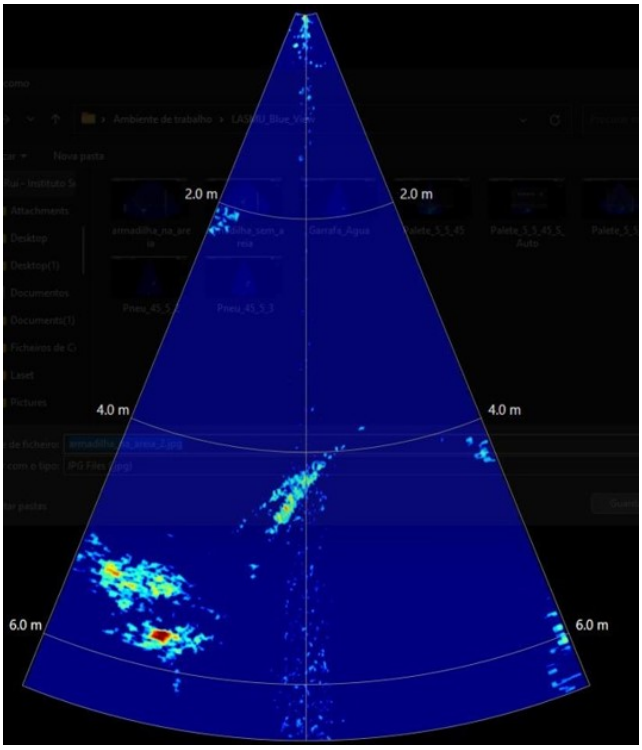



Figure 5.12: Result from Blueview for fishing traps.

## 5.3 Real-world tests

Continuing with the tests to validate the best approach and have access to more data for data processing, it was decided to do the tests in a real case, initially in the AUV, but eventually, the possibility of using the USV PORTUS present in the LSA laboratory arose, for the first phase of the tests in a real scenario, two outings were carried out, one on June 26, 2024, the second on July 25, 2024, under the supervision of Eng. Guilherme Silva, the laboratory team applied the M3 Kongsberg sonar to the USV, with an inclination of 30º as shown in the image, the sonar was located near the center of the vessel and applied to a piece that allowed it to vary its height, which benefited the test because lowering the piece as much as possible meant that the sonar support was 1 meter from the submerged part of the USV, thus minimizing the possibility of disturbing the data.

This USV was transported to the Leixões Marina, where, once in the water, it headed to the maneuvering basin where the tests were conducted. These tests were accompanied by a small boat from which it was possible to prepare the targets and control the conditions under which the tests were conducted. In the first day of tests, in this basin, as the tide was high, there were areas with depths of 16 meters and others less deep. However, due to some wind and to allow the USV to be more stable, the tests were carried out in an area where the depth ranged from 6 meters to 8 meters.


Figure 5.13: M3 Kongsberg sonar on the USV PORTUS.

In the first test, a fishing net anchored at an angle was read, appearing to descend in a zone with a depth of 16 meters. Subsequently, the test continued with a trap on the seabed, but this time the tests were moved to one end of the maneuvering basin to be conducted more stably. This trap was then placed in an area with a depth of

6 meters. After that, tests were carried out with a wooden pallet, a plastic cone, a cluster of nets, and two plates—one translucent and the other opaque. Lastly, not satisfied with how the net test was conducted, the test was repeated with the net at the surface, using buoys at both ends for signaling.

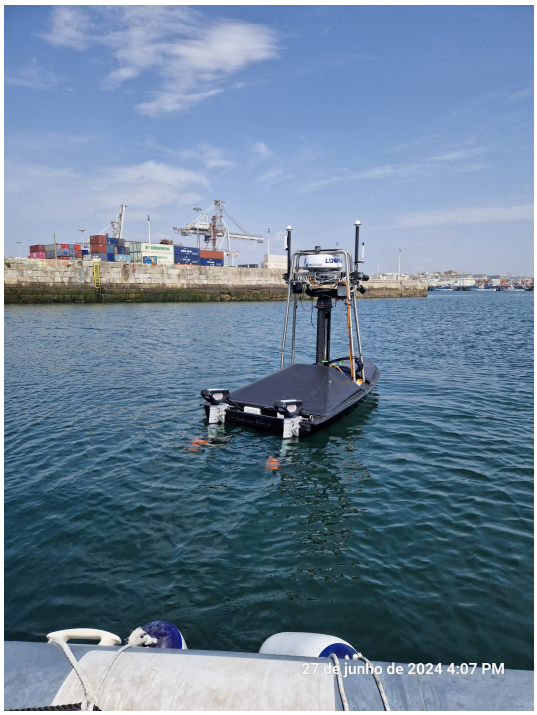



Figure 5.14: USV PORTUS on the first day of test localization.

Unfortunately, there were communication issues between the USV and the terminal that was controlling it. This terminal was located in containers at INESC TEC where a team from the LSA (Laboratory of Autonomous Systems) of researchers was controlling the USV. The team encountered the problem of insufficient communication speed to capture the sonar data and simultaneously visualize what was being seen on it. Therefore, the data were still acquired, but verifying what was being recorded was impossible. At the same time, the *Enhanced Image Quality* (EIQ) mode was activated, which reduces the *Frames per second* (FPS). This mode takes 12 pings and estimates an average, thus generating an optimized image. This makes this module suitable for slow tests, in this case where the USV was moving at a significant speed. In most logs, when the target starts to be noticed in one image, the USV has already passed the target in the next image, resulting in only the cluster of nets being visible in the readings made.


Figure 5.15: USV PORTUS net test.

## 5.4 Data processing

Regarding the data processing, a structure was developed based on the analysis of various studies. A diagram was created to outline the implementation of the data processing workflow. First, a script was developed to extract information from the data obtained during the tests, generating optimized images so that the subsequent process would be more precise and efficient. Then, the plan was to input the data into a neural network and use the acquired images to train it. Once the presence of the targets could be identified, it was also necessary to characterize them to determine what was found.

The data received from the M3 sonar, from which most of the test data were acquired, were extracted in ROS bag format (Robot Operating System). To visualize this data and record it frame by frame for second-by-second analysis, and at the same time to feed the data to the neural network as images, a Python script was developed. This script generated both color and gray-scale images to expedite the neural network's processing since certain targets are more perceptible in gray-scale and others in color.

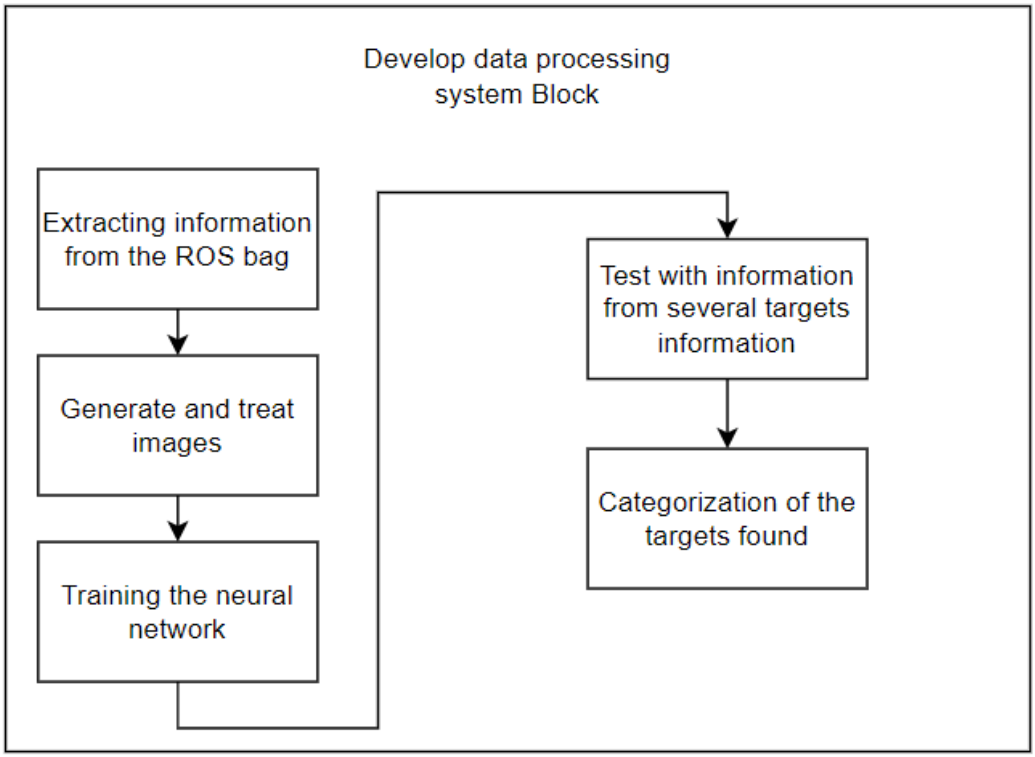



Figure 5.16: Architecture for project implementation.

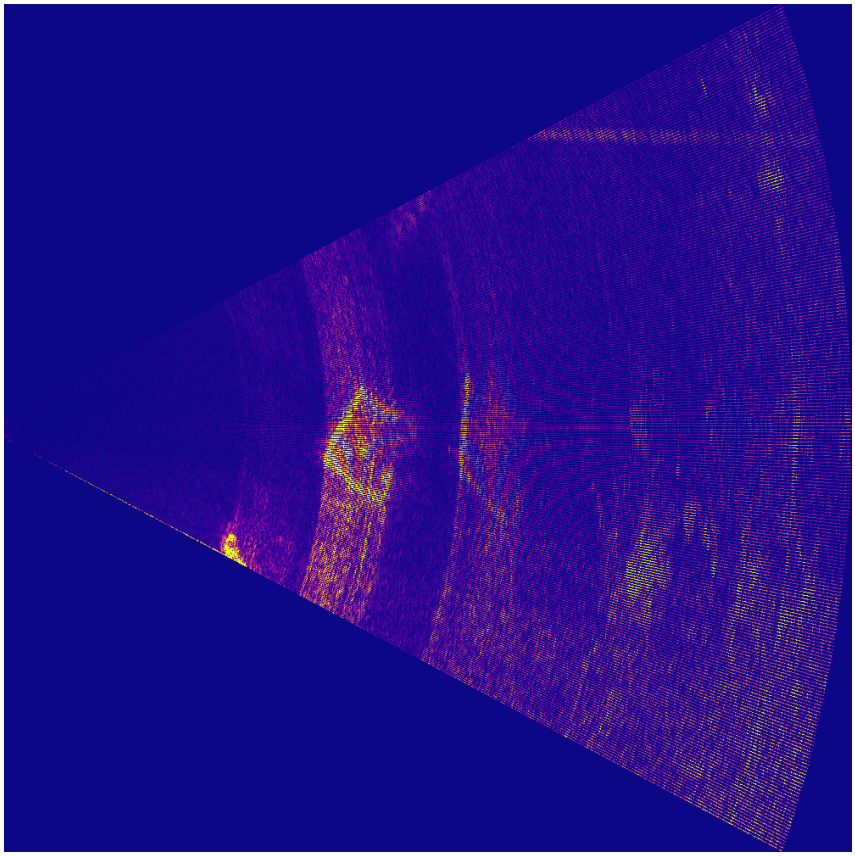


Figure 5.17: Locally generated images in colors of a fishing trap.

Using various libraries, data extraction from ROS bag files was achieved efficiently. The "os" library handled operating system tasks, like directory creation. "Numpy" was essential for numerical processing and array manipulation, while "open3d" enabled 3D data processing and visualization. The "*Python Imaging Library* (PIL)" library allowed for image file creation and manipulation. "rosbag" facilitated reading data from ROS bag files, working alongside "sensor_msgs.point_cloud2" to handle point cloud messages in ROS format. Finally, "matplotlib" was used to generate color-mapped images from color scales. Next, the data is extracted and processed, taking into account the x, y, and z coordinates and an intensity value, to then create and save an image derived from the point cloud generated with the obtained data. This image is then converted to grey-scale, saving both versions, and the point cloud is also stored. This script ultimately serves as a comprehensive tool for transforming raw point cloud data from sensors into visually accessible formats that are useful for subsequent analysis.

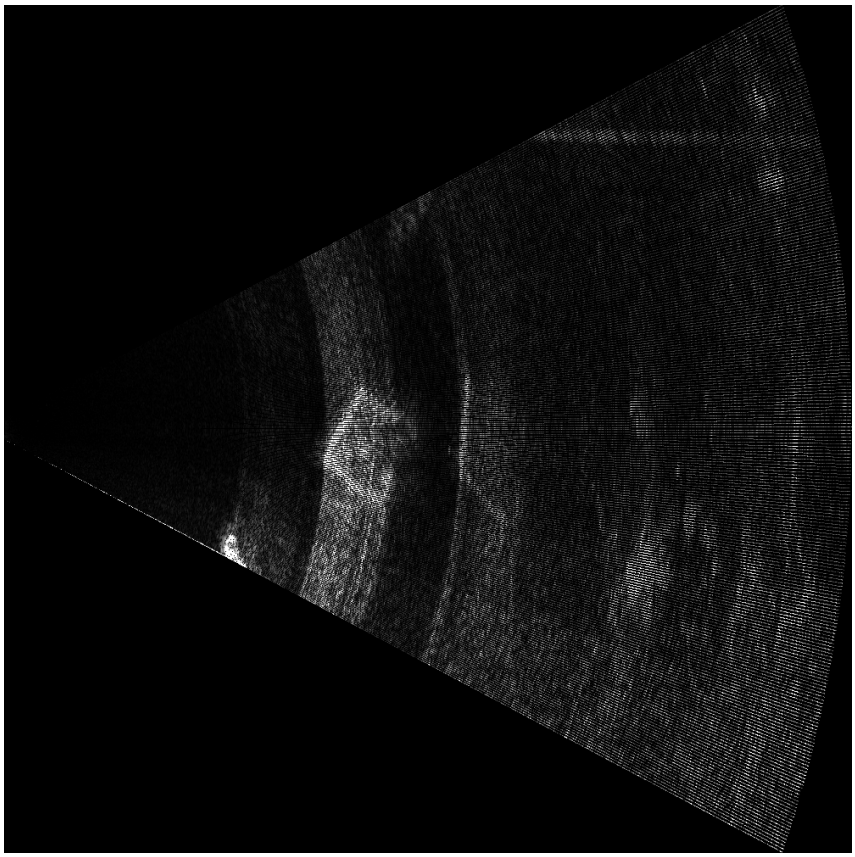


Figure 5.18: Locally generated images in grey scale of a fishing trap.

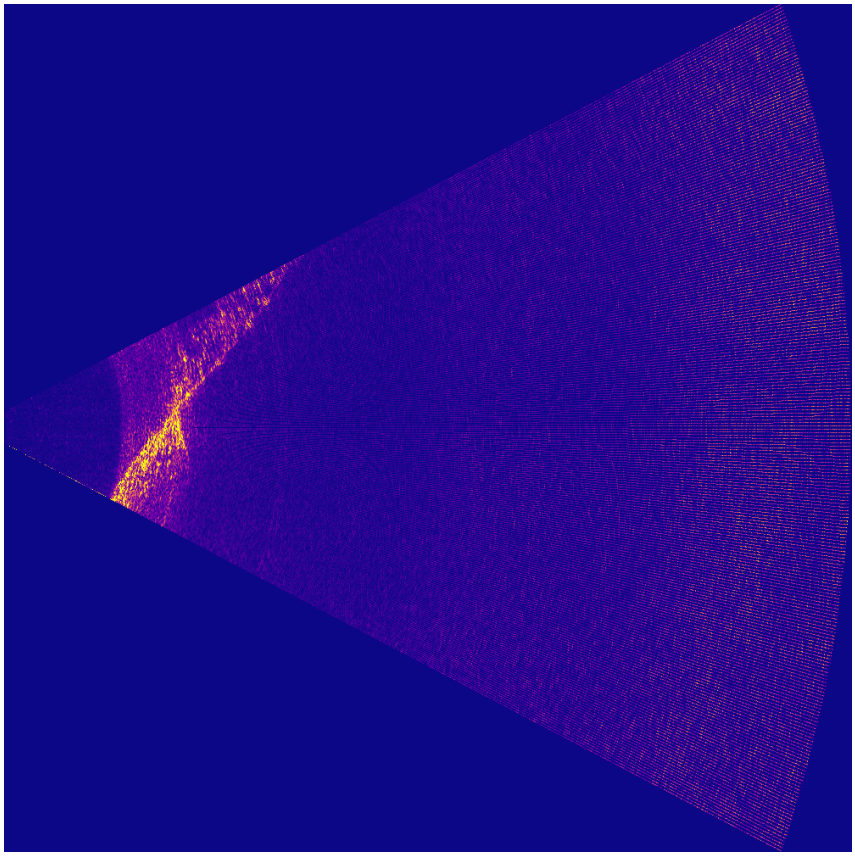


Figure 5.19: Locally generated images in colors of a fishing net.

Next, another code was developed to process the generated images. First, a structure for the neural network and its type was designed, where it was determined that it would be best to use CNNs. Several examples of existing practical projects were analyzed, and it was found that the best tool to use was Google Colaboratory, which offers a free Jupyter notebook environment with access to GPUs and TPUs. It also allowed information to be saved to Google Drive, freeing up space on the local machine. Based on the initial research, it was decided to create a system that would use Fast R-CNN. Fast R-CNN is an object detection model that improves the efficiency of the classic R-CNN by sharing the computation of feature maps across the entire image. Instead of extracting Regions of Interest (RoIs) after applying convolutional networks (which is computationally expensive), Fast R-CNN applies a convolutional network to the entire image once and then extracts features for each RoI using a technique called RoI Pooling.

```
csv_file_paths = [
    '/content/drive/My Drive/Colab Notebooks/armadilhapython/labels.csv',
    '/content/drive/My Drive/Colab Notebooks/rede/rlabels.csv',
    '/content/drive/My Drive/Colab Notebooks/pneu/pneulabels.csv',
    '/content/drive/My Drive/Colab Notebooks/garrafa/garrafalabels.csv',
    '/content/drive/My Drive/Colab Notebooks/palete/paletelabels.csv'
]
images_dir_paths = [
    '/content/drive/My Drive/Colab Notebooks/armadilhapython/',
    '/content/drive/My Drive/Colab Notebooks/rede/',
    '/content/drive/My Drive/Colab Notebooks/pneu/',
    '/content/drive/My Drive/Colab Notebooks/garrafa/',
    '/content/drive/My Drive/Colab Notebooks/palete/'
]
empty_image_dir = ['/content/drive/My Drive/Colab Notebooks/empty/']  # Caminho para imagens vazias
```

Figure 5.20: Google drive path for loafing the files.

It was also necessary to include SPP-Net, which introduces a Spatial Pyramid Pooling layer that allows the network to accept inputs of any size and produce a fixed output size, which is a requirement for some subsequent layers in CNNs. The spatial pyramid pooling can effectively handle variations in object size and scale by dividing the input into several scales and applying pooling at each level.

To integrate Fast R-CNN with SPP-Net, you can follow these steps: use a base CNN, such as VGG16, to extract feature maps from the entire image, and instead of RoI Pooling as in Fast R-CNN, you can use SPP to process Regions of Interest (RoIs) extracted from the image. This involves dividing each RoI into segments, applying pooling to each segment, and then combining the results to form a fixed-size feature vector, regardless of the RoI size. Finally, the resulting features are used to feed a fully connected layer that performs both object classification within the RoI and bounding box regression to obtain the precise location of the object.

Using the VGG16 architecture in TensorFlow/Keras, it can be adapted to handle multiple categories and empty images. First, the data is imported from the drive, and then the images and their respective labels are loaded. The data loading and network training process can be done as follows: first, load the data, then prepare it using LabelEncoder to transform textual labels into numerical ones and split the data into training and test sets. Next, we build the model using VGG16 as the base, adding layers for global pooling and classification. After that, it is necessary to compile and train the model by configuring the optimization process and training the model with the prepared data. Finally, the model is saved locally so that when new images appear, it is not necessary to retrain the model. For the new images, the script loads a set of new images, applies the model to make predictions, and prints the predicted labels for each image.

```
Mounted at /content/drive
Downloading data from https://storage.googleapis.com/tensorflow/keras-applications/vgg16/vgg16_weights_tf_dim_ordering_tf_kernels_notop.h5
58889256/58889256 ━━━━━━━━━━━━━━━━━━━━ 0s 0us/step
Epoch 1/100
3/3 ━━━━━━━━━━━━━━━━━━━━ 63s 21s/step - accuracy: 0.2048 - loss: 1.9683 - val_accuracy: 0.2727 - val_loss: 1.6272
Epoch 2/100
3/3 ━━━━━━━━━━━━━━━━━━━━ 83s 22s/step - accuracy: 0.2740 - loss: 1.6486 - val_accuracy: 0.6364 - val_loss: 1.4252
Epoch 3/100
3/3 ━━━━━━━━━━━━━━━━━━━━ 64s 23s/step - accuracy: 0.5376 - loss: 1.4776 - val_accuracy: 0.6364 - val_loss: 1.3154
```

Figure 5.21: Begin of data processing in Google Colaboratory.

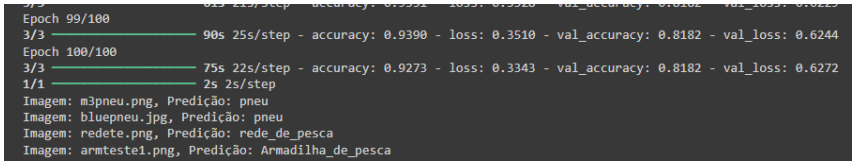



Figure 5.22: Ending of data processing in Google Colaboratory.

The VGG16 used is derived from a convolutional neural network architecture developed by Karen Simonyan and Andrew Zisserman from the University of Oxford in 2014. It is considered one of the most influential computer vision architectures due to its features: it has a depth of 16 convolutional layers, allowing it to capture a complex hierarchy of visual features. All these convolutional layers use very small filters (3x3), but many stacked layers increase the effective depth. It also uses max pooling (MaxPooling) to reduce spatial dimensions after certain convolutional blocks. After feature extraction by VGG16, layers were added to adapt the model to the specific task of classifying targets, a Global Average Pooling (GAP) layer follows the last convolutional layer. GAP helps reduce each feature map's spatial dimensions to a single value per channel. After the GAP, we implemented a dense (Fully Connected) layer to perform the learning of high-level features and the final decision. The last layer is a dense layer with a softmax activation function, which is used for multiclass classification. Softmax is used in the output layer to convert the logits (raw output values from the last dense layer) into class probabilities. For evaluation, we use cross-entropy loss for the classification task, which measures the model's performance by comparing the predicted probabilities with the true labels. The model is also trained over several epochs, allowing it to adjust the weights to minimize the loss function iteratively. Additionally, to aid processing and enhance the performance of the solution, an 'empty' class was added so that this system could recognize when there is nothing to process.

In conclusion, the combination of all these factors allowed the creation of a network with a robust and flexible system for the task of classifying images with possible targets, identifying and classifying them. The choice of VGG16 as a base provided a strong starting point due to its proven efficacy in computer vision tasks.

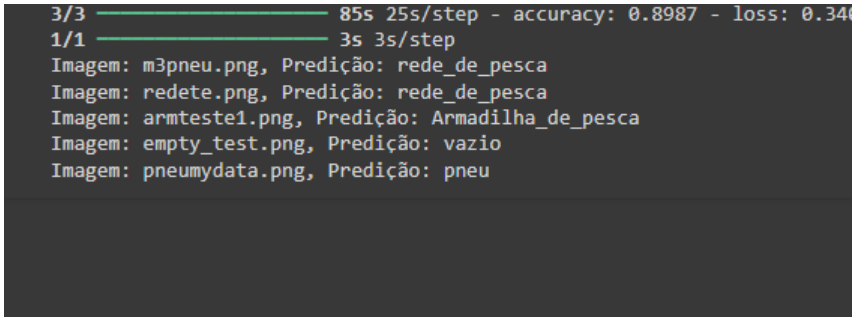



Figure 5.23: Ending of data processing 89% accuracy 1h20min of time training.

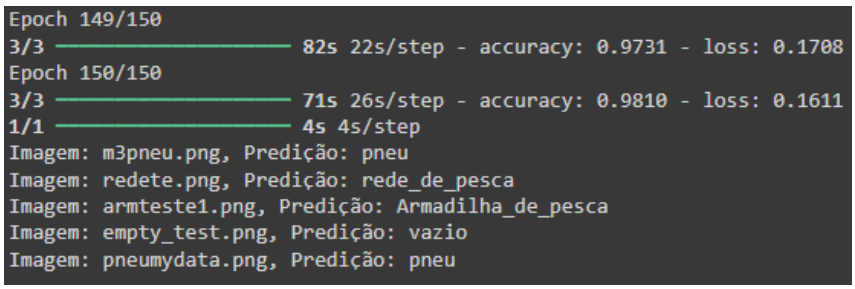



Figure 5.24: Ending of data processing 98% accuracy 2h of time training.

In the end, it was possible to generate the images successfully and even optimize them, as they were creating a linearization between the previous image and the next one to minimize the impact of noise. This allowed the extraction of the test data both locally from the lab results and from the open-sea tests. Regarding the image processing aimed at determining the existence of targets and classifying them, both locally generated images and cropped images provided in real-time by the software of the two sonars used in the tests were added as inputs. This allowed compatibility with both real-time images and those provided by the software developed in the previous section to be validated. It was necessary to adjust some settings and perform several iterations for the model to recognize the data being processed with greater accuracy. After the system did multiple readings to improve the precision of capturing new images, it became necessary to increase the processing to read the existing images 100 times to enhance the accuracy of the reading, but when the number of images for the learning was reduced the reading and learning process to 50.

With the obtained data, it was possible to test various targets and assess the response to these targets, gaining an understanding that the network was following the correct path. The main subject of this study was tested, and favorable results were obtained for solving the problem. Although more examples are necessary for more efficient and precise training of the neural network, it demonstrated that for

similar shapes and even different targets, it was able to efficiently determine what was requested and also identify which images were empty without a target. However, the results would have been more precise if more tests had been conducted in a real-world environment to gather more data and examples for further training the network. Due to some unforeseen events and difficulties, it was not possible to carry out more tests, but it was evident that the developed network functions effectively. At the same time, this study made another step forward in addressing the detection and search for ALDFG, as it was possible to detect and characterize fishing nets and other targets as intended, even with few real-world tests conducted.

## Chapter 6

# Conclusions and Future Work

During the development of this study, it was found that it is increasingly necessary to find effective methods to detect and identify ALDFG, there are various possible solutions, as mentioned in the state of the art, but the chosen solution was to use sonar systems, specifically the M3 multi-frequency sonar from Kongsberg. Although there are other possible solutions, such as those seen in the state of the art, like the Nettag project, where applying tags to the nets allows for triangulation of their position, it was a good method to locate the nets. However, the problem lies in the difficulty of convincing fishermen to use these tags, as it would involve spending money. In many cases, if their use is not mandatory by legislation, most of the fishing industry would not adopt them. Additionally, locating the nets would always require devices to find the tags, while with sonar, it would only be necessary to take readings with a specific sonar and then process the data, without obligating the industry to use tags on the nets. However, not just any sonar can be used, there are many types of sonar, and the best one to use may vary depending on the scenario in which it is applied.

For this study, it was decided that using an imaging sonar would be the best option for better image processing with the help of neural networks, allowing for the characterization of the extracted images. The availability of high-quality sonar, such as the M3, in the laboratory, which supports both imaging and bathymetry modes, allowed for testing its effectiveness and obtaining good test results. This validated its functionality and confirmed that this was indeed the best approach to follow during this study.

The tests conducted in the tank allowed for controlled testing and enabled the best conclusions to be drawn regarding the two sonars tested. It was observed that the M3 provided better quality and accuracy in detecting targets, while the Blueview could detect a target more easily but with significantly lower quality. Based on the results obtained, it was decided to proceed with real-world testing using only the M3 Sonar. Finding an AUV or USV drone to mount the Sonar and carry out the tests was necessary for real-world tests. Due to the lack of available AUVs, the tests were conducted with the Portus USV.

For image processing with the developed algorithms and the optimizations applied, such as converting the images to greyscale, it was possible to verify the results observed during the lab tests with greater accuracy and confirm what had been seen in the live stream through Rviz. Additionally, the real-world test results were analyzed. Through a few iterations, it was possible to optimize the results and acquire good examples for training the neural network.In several examples, such as the tire, fishing trap, and wooden pallet tested in the lab, as well as the fishing net in the maneuvering basin, the results were promising, proving that it is possible to identify the targets present in some of the images.

Based on everything studied and addressed, the planned architecture was followed for the neural processing part, allowing this study to achieve favorable results. As such, the approach of using an imaging sonar proved to be the best method, as it allowed for the generation of images that provided a precise understanding of the object's shape. Even in cases where only part of the object could be determined, it was observed that once a target was detected, a maneuver to circle the object and obtain more angles could be initiated. By combining all the collected data, a more accurate depiction of the target could be achieved, or even explicitly identify what was being detected. It is important to note that the surrounding aspects of the target can often affect its precision and identification. However, with the use of sonar, these problems are drastically reduced compared to using cameras, for example. The study confirmed that sonars are indeed the most viable option for collecting underwater data, and using various algorithms can facilitate target detection. In addition to the possibility of performing pre-defined maneuvers to circle the target, there is also the potential, as done in this study, of using image processing with the aid of a neural network. This network, trained with the data from the tests conducted so far, proved effective. The script developed for data extraction allowed the generation of images with intensity levels, enhancing target precision. Grayscale images were also generated to provide another level of precision for the extracted data. Both types of images demonstrated good target accuracy and were quite similar when compared. However, grayscale images showed that processing them through the neural network was faster.

After generating the data, training the neural network to automate target detection was possible. Initially, finding a model that met the needs of this solution was challenging. However, using the VGG16 architecture allowed the creation of a fast and accurate network. Despite the scale at which the neural network was built in this study, it proved to be a viable solution for the project, with significantly positive results, achieving an accuracy rate of 90% in classification. The success rate for detecting the presence of a target was 100%, proving that this is the right approach for underwater object detection. More data needs to be added to the network's training to improve object characterization, and having data from multiple angles for the same target would further enhance this process.

Due to the lack of existing content on the search for fishing nets, it was difficult to find viable data from other sonar systems and research that could be added to the obtained data to provide more images for data processing. However, since the images were from different types of sonar, they would not be a good option as they could compromise the results obtained with the initially used data. Nevertheless, the lack of existing data in other studies reinforced the certainty of the importance that the results of this study will allow progress and take another step forward in finding a solution to address this problem.

## 6.1 Future Work

Several aspects allow us to discuss future work hypotheses. While some may be a way to continue projects in this field, others are strictly related to the limitations and difficulties encountered during the process. To begin with, it would be important to conduct more tests in real-world environments to generate quality data that would enable more effective training of the neural network. However, despite being aware of this need, access to more tests was hindered by the lack of availability of an AUV or USV to carry them out. In addition, when access to the USV was obtained, new challenges still arose: days when the sea was too rough, bad weather, and malfunctions, among others. In an academic Master's context, where the time for research is limited, these problems present additional difficulties but also future opportunities, as these tests may still be conducted and their results could influence future studies. Another issue relates to the lack of real-world tests, that is, tests where the targets were not intentionally positioned. The tests conducted had authorization from the port authority to place nets in the maneuvering basin of the Leixões Terminal. The positioning of the targets was always done in a controlled manner. In the future, for a deeper understanding of the techniques' performance and for a different kind of validation, studies could be conducted in partnership with organizations or communities working offshore. This way, the tests could be evaluated in real situations not controlled solely by the researchers.

During the project, it was possible to understand the effectiveness of a good sonar, such as the M3 that was used. This idea aligns with other studies analyzed for comparison. However, this use also allowed for reflection on possible improvements for future studies. The use of these sonars with real-time characterization, as previously discussed, would be a viable option and a significant improvement to the work, as it would allow the sonar to avoid obstacles or alert the user controlling the USV or AUV of their presence. It became clear that in a target search mission, optimization, and efficiency could see significant improvements if there were a preplanned maneuver or trajectory planning. This would allow continuous coverage of the desired area, with an immediate notion of the target's position once found, usually obtained through data from other sensors whose function is to provide location. In addition to this maneuver, it would also be advantageous to explore data collection techniques for characterizing detected targets. Thus, when a target is found, it would be important to circle it and collect data from different angles, generating a more complex and accurate characterization of the target, and facilitating its removal, both in terms of efficiency and even safety.

Finally, there are still several possibilities to continue research on this topic regarding the impact of this solution on the preservation of the aquatic environment. As analyzed, this would be a solution that allows for the detection and identification of various harmful targets to the environment, whose removal would help reduce the impact on the degradation of aquatic ecosystems by marking nets and other objects that pose a risk to aquatic life. The removal of such objects, particularly ghost nets, would also be beneficial for research, as it reduces the risk of equipment loss, as observed in studies where AUVs became entangled in nets or collided with other objects, hindering the project's development.

Future studies can always start with the motivation behind this research: the ocean and aquatic environments, in general, remain a central stage for human life and the life of the planet. The constant and significant increase in ghost nets and other ALDFG components highlights the continued importance of studies like this and their relevance in the future, especially when considering that the aquatic environment, so crucial for biodiversity and sustaining life, is being severely harmed by human activity. With studies like this one, these problems may be mitigated or even solved.

# References


[1] M. Valdenegro-Toro *et al.*, "Submerged marine debris detection with autonomous underwater vehicles," *International Conference on Robotics and Automation for Humanitarian Applications (RAHA)*, 2016. [Cited on pages 6 and 16]

[2] L. M. Fulton, "Evaluating the use of side scan sonar for improved detection and targeted retrieval of abandoned, lost, or otherwise discarded fishing gear," Master's thesis, Memorial University of Newfoundland, Cambridge, 2020. [Cited on page 7]

[3] L. Fulton *et al.*, "Evaluating the use of side scan sonar for improved detection and targeted retrieval of abandoned, lost, or otherwise discarded fishing gear," *Continental Shelf Research*, p. 265, 2023. [Cited on page 8]

[4] R. Qin *et al.*, "Multiple receptive field network (mrf-net) for autonomous underwater vehicle fishing net detection using forward-looking sonar images," *MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliation*, p. 21, 2021. [Cited on pages 9 and 17]

[5] Z. Spirkovski *et al.*, "Ghost net removal in ancient lake ohrid: A pilot study," *Fisheries Research*, no. 211, pp. 46–50, 2019. [Cited on page 10]

[6] M. Parsons *et al.*, "Detection of sharks with the gemini imaging sonar.," *Acoustics Australia*, no. 42, pp. 101–108, 2014. [Cited on pages 10 and 11]

[7] J. Zheng *et al.*, "Underwater fishnet and hole detection using miniature underwater robot," *Report from Georgia Institute of Technology*, 2014. [Cited on pages 12 and 17]

[8] H. Lee *et al.*, "Autonomous underwater vehicle control for fishnet inspection in turbid water environments.," *International Journal of Control, Automation, and System*, no. 20, pp. 3383 – 3392, 2022. [Cited on page 13]

[9] A. Martins *et al.*, "A robotic solution for nettag lost fishing net problem," *2020 Global Oceans: Singapore*, 2020. [Cited on pages 14 and 18]

[10] J. Quintana *et al.*, "Towards automatic recognition of mining targets using an autonomous robot.," *2018 OCEANS*, pp. 1 – 8, 2018. [Cited on pages 15 and 18]

[11] G. Macfadyen *et al.*, "Abandoned, lost or otherwise discarded fishing gear.," *Food and Agriculture Organization of the United Nations*, no. 523, pp. 1 – 115, 2009. [Cited on page 23]

[12] E. Gilman *et al.*, "Abandoned, lost and discarded gillnets and trammel nets: Methods to estimate ghost fishing mortality, and the status of regional monitoring and management," *Food and Agriculture Organization of the United Nations*, no. 600, pp. 1 – 79, 2016. [Cited on page 23]

[13] K. Richardson *et al.*, "Understanding causes of gear loss provides a sound basis for fisheries management," *Marine Policy*, no. 101, pp. 30 – 39, 2019. [Cited on page 23]

[14] P. B. de Dados de Portugal Contemporâneo., "Peixe capturado, por principais espécies.." Available at `https://www.pordata.pt`, 2022. (Last accessed in 20/06/2024). [Cited on page 24]

[15] S. e. S. M. D. Direção-Geral de Recursos Naturais, "Decreto-lei n.º 218/2023, sobre as artes de pesca e a gestão dos recursos pesqueiros.." Available at Diário da República and `https://www.dgrm.mm.gov.pt`, 2023. (Last accessed in 20/06/2024). [Cited on page 24]

[16] N. Ocean Explorer, National Oceanic and Atmospheric Administration, "Mapping the ocean floor. ocean explorer, national oceanic and atmospheric administration." Available at `https://oceanexplorer.noaa.gov/technology/sonar/sonar.html`. (Last accessed in 12/07/2024). [Cited on page 26]

[17] Deep Ocean Education Project, "Sonar: Seeing with sound. deep ocean education project.." Available at `https://deepoceaneducation.org/resources/sonar-seeing-with-sound/`. (Last accessed in 20/07/2024). [Cited on page 26]

[18] Furuno, "Sonar basics: How sonar works. furuno.." Available at `https://www.furuno.com/en/technology/sonar/basic/`. (Last accessed in 10/06/2024). [Cited on pages 26, 27, and 28]

[19] Vedantu, "Sonar waves: Definition, uses, working, and application. vedantu.." Available at `https://www.vedantu.com/evs/sonar-wave`. (Last accessed in 15/07/2024). [Cited on page 27]

[20] Blue Robotics, "A smooth operator's guide to underwater sonars and acoustic devices.." Available at `https://bluerobotics.com/learn/a-smooth-operators-guide-to-underwater-sonars-and-acoustic-devices/`. (Last accessed in 20/06/2024). [Cited on page 29]

[21] Kongsberg, "M3 sonar: Multibeam imaging sonar.." Available at `https://www.kongsberg.com/discovery/seafloor-mapping/sonars/multibeam-sonar-m3-sonar/`. (Last accessed in 11/11/2023). [Cited on page 31]

[22] Teledyne Marine., "Blueview: Multibeam imaging sonar solutions.." Available at `https://www.teledynemarine.com/blueview`. (Last accessed in 05/12/2023). [Cited on page 31]

[23] Tritech., "Gemini 1200ik multibeam imaging sonar.." Available at `https://www.tritech.co.uk/products/gemini-1200ik`. (Last accessed in 08/02/2024). [Cited on page 32]

[24] Spiceworks., "What is a neural network?." Available at `https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-a-neural-network/`. (Last accessed in 08/07/2024). [Cited on page 34]

[25] Cloudflare., "What is a neural network?." Available at `https://www.cloudflare.com/pt-br/learning/ai/what-is-neural-network/`. (Last accessed in 15/07/2024). [Cited on page 35]

[26] KnowledgeHut., "Types of neural networks and their applications.." Available at `https://www.knowledgehut.com/blog/data-science/types-of-neural-networks`. (Last accessed in 10/07/2024). [Cited on pages 35 and 36]

[27] Rare Connections., "Types of neural networks.." Available at `https://rareconnections.io/types-of-neural-networks/`. (Last accessed in 12/07/2024). [Cited on page 37]

[28] Builtin., "Vgg16 - convolutional network for classification and object detection." Available at `https://builtin.com/machine-learning/vgg16`. (Last accessed in 20/07/2024). [Cited on page 38]

[29] Vitalflux., "Deep neural network examples from real life.." Available at `https://vitalflux.com/deep-neural-network-examples-from-real-life/`. (Last accessed in 12/07/2024). [Cited on page 38]

# Appendix A

# Appendix

## A.1 section 1 - Image generator

```
import os
import numpy as np
import open3d as o3d
from PIL import Image
import rosbag
import sensor_msgs.point_cloud2 as pc2
import matplotlib.pyplot as plt
import matplotlib.cm as cm

# Diret rios de entrada e sa da
input_bag_file = 'AIRSHIP_SENSORES_2024-06-27-14-11-24.bag'
#input_bag_file = 'AIRSHIP_SENSORES_2024-06-27-14-37-37.bag' vazio
#input_bag_file = 'AIRSHIP_SENSORES_2024-06-27-15-10-21.bag'
#input_bag_file = 'AIRSHIP_SENSORES_2024-06-27-15-20-07.bag'
#input_bag_file = 'AIRSHIP_SENSORES_2024-06-27-15-26-14.bag'
#input_bag_file = 'AIRSHIP_SENSORES_2024-06-27-15-28-16.bag'
#input_bag_file = 'AIRSHIP_SENSORES_2024-06-27-15-36-14.bag'
#input_bag_file = 'AIRSHIP_SENSORES_2024-06-27-15-42-48.bag'#vazio
#input_bag_file = 'AIRSHIP_SENSORES_2024-06-27-15-49-14.bag'
#input_bag_file = 'AIRSHIP_SENSORES_2024-06-27-16-04-02.bag'


output_dir_pcd = 'PointCloudFiles'
```

```
output_dir_images_color = 'PointCloudImagesColor'
output_dir_images_gray = 'PointCloudImagesGray'

# Create output directories if none exist
os.makedirs(output_dir_pcd, exist_ok=True)
os.makedirs(output_dir_images_color, exist_ok=True)
os.makedirs(output_dir_images_gray, exist_ok=True)

# Function to extract points and intensities from PointCloud2
def extract_points_and_intensities(msg):
    points = []
    intensities = []
    for p in pc2.read_points(msg, skip_nans=True, field_names=("x"
        , "y", "z", "intensity")):
        x, y, z, intensity = p
        points.append([x, y, z])
        intensities.append(intensity)
    return np.array(points), np.array(intensities)

# Function to create and save intensity image with colormap
def create_and_save_image(points, intensities, img_size,
    color_filename, gray_filename, intensity_scale=2):
    img_color = np.zeros((img_size, img_size, 3), dtype=np.uint8)
    img_gray = np.zeros((img_size, img_size), dtype=np.uint8)

    min_x, max_x = np.min(points[:, 0]), np.max(points[:, 0])
    min_y, max_y = np.min(points[:, 1]), np.max(points[:, 1])

    for idx, point in enumerate(points):
        x, y, z = point
        intensity = intensities[idx] * intensity_scale

        # Normalize coordinates to image range
        ix = int((x - min_x) / (max_x - min_x) * (img_size - 1))
        iy = int((y - min_y) / (max_y - min_y) * (img_size - 1))

        img_gray[iy, ix] = min(int(intensity), 255)

    # Use colormap to generate the color image
    colormap = cm.get_cmap('plasma')
    img_color = colormap(img_gray / 255.0)[:, :, :3] * 255

    img_color = Image.fromarray(img_color.astype(np.uint8))
    img_color.save(color_filename)

    img_gray = Image.fromarray(img_gray)
    img_gray.save(gray_filename)

# read arquive .bag using rosbag
```

```
bag = rosbag.Bag(input_bag_file)
point_cloud_topic = '/m3Image'  # Use the correct identified topic

frame_id = 0
img_size = 1000  # Increase resolution to improve sharpness

for topic, msg, t in bag.read_messages(topics=[point_cloud_topic]):
    points, intensities = extract_points_and_intensities(msg)

    print(f'Frame {frame_id}: {len(points)} points, Intensity range: {np.min(intensities)} - {np.max(intensities)}')

    # Save the point cloud
    pcd = o3d.geometry.PointCloud()
    pcd.points = o3d.utility.Vector3dVector(points)
    pcd.colors = o3d.utility.Vector3dVector(np.tile(intensities, (3, 1)).T / 255.0)  # Normalizar intensidades
    o3d.io.write_point_cloud(os.path.join(output_dir_pcd, f'frame_{frame_id:04d}.pcd'), pcd)

    # Create and save images
    color_filename = os.path.join(output_dir_images_color, f'frame_{frame_id:04d}.png')
    gray_filename = os.path.join(output_dir_images_gray, f'frame_{frame_id:04d}.png')
    create_and_save_image(points, intensities, img_size, color_filename, gray_filename, intensity_scale=2)

    frame_id += 1

bag.close()
```

## A.2 section 2 - Neural network

```


import os
import cv2
import numpy as np
import pandas as pd
import pickle
from tensorflow.keras import layers, models
from tensorflow.keras.applications import VGG16
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
```

```
from tensorflow.keras.models import load_model
from google.colab import drive

drive.mount('/content/drive')

# Defini  o de caminhos para arquivos e diret rios
model_path = '/content/drive/My Drive/Colab Notebooks/model_vgg16.keras'  # Mudan a para .keras
label_encoder_path = '/content/drive/My Drive/Colab Notebooks/label_encoder.pkl'
#csv_file_paths = [
#    '/content/drive/My Drive/Colab Notebooks/armadilhapython/labels.csv',
#    '/content/drive/My Drive/Colab Notebooks/rede/rlabels.csv'
#]
#images_dir_paths = [
#    '/content/drive/My Drive/Colab Notebooks/armadilhapython/',
#    '/content/drive/My Drive/Colab Notebooks/rede/'
#]
csv_file_paths = [
    '/content/drive/My Drive/Colab Notebooks/armadilhapython/labels.csv',
    '/content/drive/My Drive/Colab Notebooks/rede/rlabels.csv',
    '/content/drive/My Drive/Colab Notebooks/pneu/pneulabels.csv',
    '/content/drive/My Drive/Colab Notebooks/garrafa/garrafalabels.csv',
    '/content/drive/My Drive/Colab Notebooks/palete/paletelabels.csv'
]
images_dir_paths = [
    '/content/drive/My Drive/Colab Notebooks/armadilhapython/',
    '/content/drive/My Drive/Colab Notebooks/rede/',
    '/content/drive/My Drive/Colab Notebooks/pneu/',
    '/content/drive/My Drive/Colab Notebooks/garrafa/',
    '/content/drive/My Drive/Colab Notebooks/palete/'
]
empty_image_dir = ['/content/drive/My Drive/Colab Notebooks/empty/']  # Caminho para imagens vazias

def load_images_and_labels(csv_file_paths, images_dir_paths, empty_image_paths=None):
    images = []
    labels = []
    for csv_file_path, images_dir_path in zip(csv_file_paths, images_dir_paths):
        labels_df = pd.read_csv(csv_file_path)
        labels_df.set_index('filename', inplace=True)
        for img_name in os.listdir(images_dir_path):
            img_path = os.path.join(images_dir_path, img_name)
```

```
52              if img_name in labels_df.index:
53                  img = cv2.imread(img_path)
54                  if img is not None:
55                      img = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
56                      img = cv2.resize(img, (224, 224))
57                      images.append(img)
58                      labels.append(labels_df.loc[img_name]['label'
                            ])
59
60      # Carregar imagens vazias
61      if empty_image_paths:
62          for empty_img_path in empty_image_paths:
63              for img_name in os.listdir(empty_img_path):
64                  img_path = os.path.join(empty_img_path, img_name)
65                  img = cv2.imread(img_path)
66                  if img is not None:
67                      img = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
68                      img = cv2.resize(img, (224, 224))
69                      images.append(img)
70                      labels.append('vazio')  # Rotular como vazio
71
72      return np.array(images), np.array(labels)
73
74  if os.path.exists(model_path) and os.path.exists(
        label_encoder_path):
75      model = load_model(model_path)
76      with open(label_encoder_path, 'rb') as file:
77          label_encoder = pickle.load(file)
78  else:
79      images, labels = load_images_and_labels(csv_file_paths,
            images_dir_paths, empty_image_dir)
80      label_encoder = LabelEncoder()
81      labels_encoded = label_encoder.fit_transform(labels)
82      images = images.astype('float32') / 255.0
83      train_images, test_images, train_labels, test_labels =
            train_test_split(images, labels_encoded, test_size=0.2,
            random_state=42)
84      base_model = VGG16(include_top=False, input_shape=(224, 224,
            3))
85      base_model.trainable = False
86      model = models.Sequential([
87          base_model,
88          layers.GlobalAveragePooling2D(),
89          layers.Dense(64, activation='relu'),
90          layers.Dense(len(label_encoder.classes_), activation='
                softmax')
91      ])
92      model.compile(optimizer='adam', loss='
            sparse_categorical_crossentropy', metrics=['accuracy'])
```

```
93      model.fit(train_images, train_labels, epochs=100,
            validation_data=(test_images, test_labels))
94      model.save(model_path)
95      with open(label_encoder_path, 'wb') as file:
96          pickle.dump(label_encoder, file)
97
98  def load_new_images(images_dir_path):
99      images = []
100     image_files = []
101     for img_name in os.listdir(images_dir_path):
102         img_path = os.path.join(images_dir_path, img_name)
103         img = cv2.imread(img_path)
104         if img is not None:
105             img = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
106             img = cv2.resize(img, (224, 224))
107             images.append(img)
108             image_files.append(img_name)
109     return np.array(images), image_files
110
111 def predict_images(model, images):
112     images = images.astype('float32') / 255.0
113     predictions = model.predict(images)
114     predicted_classes = np.argmax(predictions, axis=1)
115     predicted_labels = [label_encoder.classes_[i] for i in
            predicted_classes]
116
117     return predicted_labels
118
119 new_images_dir_path = '/content/drive/My Drive/Colab Notebooks/
        new_images/'
120 new_images, image_files = load_new_images(new_images_dir_path)
121 predicted_labels = predict_images(model, new_images)
122
123 for img_name, label in zip(image_files, predicted_labels):
124     print(f"Imagem: {img_name}, P r e d i   o : {label}")
```