Journal of
Marine Science and Engineering

MDPI

Article

Acoustic Imaging Learning-Based Approaches for Marine Litter Detection and Classification

Pedro Alves Guedes [1,*], Hugo Miguel Silva [1], Sen Wang [2], Alfredo Martins [1,3], José Almeida [1,3] and Eduardo Silva [1,3]

[1] INESCTEC—Institute for Systems and Computer Engineering, Technology and Science, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal; hugo.m.silva@inesctec.pt (H.M.S.); alfredo.martins@inesctec.pt (A.M.); jose.m.almeida@inesctec.pt (J.A.); eduardo.silva@inesctec.pt (E.S.)
[2] Imperial College London, South Kensington Campus, London SW7 2AZ, UK; sen.wang@imperial.ac.uk
[3] ISEP—School of Engineering, Polytechnic Institute of Porto, Rua Dr. António Bernardino de Almeida 431, 4249-015 Porto, Portugal
[*] Correspondence: pedro.e.guedes@inesctec.pt

Abstract: This paper introduces an advanced acoustic imaging system leveraging multibeam water column data at various frequencies to detect and classify marine litter. This study encompasses (i) the acquisition of test tank data for diverse types of marine litter at multiple acoustic frequencies; (ii) the creation of a comprehensive acoustic image dataset with meticulous labelling and formatting; (iii) the implementation of sophisticated classification algorithms, namely support vector machine (SVM) and convolutional neural network (CNN), alongside cutting-edge detection algorithms based on transfer learning, including single-shot multibox detector (SSD) and You Only Look once (YOLO), specifically YOLOv8. The findings reveal discrimination between different classes of marine litter across the implemented algorithms for both detection and classification. Furthermore, cross-frequency studies were conducted to assess model generalisation, evaluating the performance of models trained on one acoustic frequency when tested with acoustic images based on different frequencies. This approach underscores the potential of multibeam data in the detection and classification of marine litter in the water column, paving the way for developing novel research methods in real-life environments.

Keywords: multibeam echosounder; water column data; macroplastics; classification; detection; convolutional neural networks; machine learning; marine litter; support vector machines; YOLOv8

check for updates

Citation: Guedes, P.A.; Silva, H.M.; Wang, S.; Martins, A.; Almeida, J.; Silva, E. Acoustic Imaging Learning-Based Approaches for Marine Litter Detection and Classification. *J. Mar. Sci. Eng.* **2024**, *12*, 1984. https://doi.org/10.3390/jmse12111984

Academic Editors: Tieshan Li and Dong-Sheng Jeng

Received: 29 July 2024
Revised: 26 September 2024
Accepted: 1 November 2024
Published: 3 November 2024

1. Introduction

Higher standards and improved living conditions have increased the demand for more materials, leading to greater waste generation [1]. Billions of metric tons are discarded annually, and without proper waste disposal, many environments are being damaged [2]. Recycling policies are failing to adequately address the disposal of plastic products, primarily composed of single-use plastics. Many companies engage in greenwashing, providing misinformation about their environmental actions to enhance their reputations and gain competitive advantages among environmentally conscious consumers [3,4].

The United Nations Environment Programme (UNEP) defines marine litter as any persistent, manufactured, or processed solid material discarded, disposed of, or abandoned in the marine and coastal environment [5–7], with plastic being the most prominent. The accumulation of plastic in the water is rapidly increasing [6]. It is estimated that 60% of all plastics ever made have been discarded in landfills or the natural environment. Positively buoyant plastic objects, commonly called floating plastic debris, illustrated in Figure 1, are influenced by a wide range of physical transport processes [7–10].

Submerged plastic debris can originate from low-density plastics combined with heavier materials, resulting in negatively buoyant masses that sink below the water's surface and biofouling.

Figure 1. Marine litter in the water column. Courtesy of Unsplash by Naja Jensen.

Microplastics can originate from fragmented parts of larger objects [6]. Detecting submerged macroplastics in the water column is necessary to capture them before they generate microplastics, reducing potential global damage. Most marine litter detection methods are related to debris on the surface of a body of water, with data essentially acquired by satellites [11]. Some solutions use drones or cameras in aircraft to acquire data to survey the amount of marine debris that covers the surface of rivers and seas. More recently, novel techniques are being applied to data extracted by hyperspectral cameras to better detect and classify marine litter [12].

Robotics can dramatically improve the detection and prediction of risks related to water pollution, providing new tools for the global management of water resources [13]. The multibeam echosounder (MBES) acquires acoustic data, and depending on the application, its raw data can be applied differently. MBES acoustic data, although composed of backscatter data, can be expressed as bathymetric data and as water column data (WCD) in the form of acoustic images.

While acoustic images can detect objects in the water column, it is crucial to test the ability to classify the footprint of a target in such images. Most methods use classical classifiers, such as support vector machine (SVM), Bayesian classifiers, and random forest classifiers. Other approaches explore deep neural networks for forward-looking sonar image classification [14]. There is limited research on classification and detection tasks performed with WCD, largely due to a low number of available datasets.

This paper contributes to the aforementioned topics and uses the imaging system setup and the foundations of the qualitative characterisation study that was made in [15] by leveraging multibeam water column data at various frequencies to detect and classify marine litter. The study encompasses:

- Using the previously acquired test tank data with different types of marine litter at multiple acoustic frequencies.
- Creation of an extensive acoustic image dataset with detailed labelling and formatting.
- Implementation of two novel classification algorithms: support vector machine (SVM) and convolutional neural network (CNN).
- Utilisation of two detection algorithms based on transfer learning: single-shot multibox detector (SSD), and You Only Look Once (YOLO), specifically YOLOv8 tuned for marine litter detection.

The results demonstrate the system's effectiveness in distinguishing between different classes of marine litter using varied algorithms for both detection and classification. Cross-frequency studies were conducted to evaluate model generalisation, assessing performance when models trained on one acoustic frequency were tested with images based on different frequencies. This innovative characterisation of multibeam data underscores its potential in enhancing the detection and classification of marine litter within the water column.

This paper is organised as follows: Section 2 has an overview of the related work in marine litter detection, focusing on water column litter detection. Section 3 describes the acoustic acquisition system setup and the solution's high-level architecture. Section 4 details the acoustic images dataset and the different acoustic image representations. Section 5 details the training procedures that were made for all the different models and a discussion of the results. Finally, Section 6 draws some conclusions on the work carried out and describes future work.

2. Related Work

This section reviews MBES acoustic imaging methods for marine litter detection and underwater image classification. Section 2.1 covers techniques like aerial surveys, hyperspectral imaging, and acoustic imaging. Section 2.2 explores classical and deep learning methods for classifying and detecting underwater acoustic data.

2.1. Marine Litter Detection

Marine litter detection mostly relies on remote sensing methods, often using unmanned aerial vehicles (UAV) [14,16]. Other methods include using satellite data, aircraft cameras, or remote hyperspectral imaging to assess debris on water surfaces [12]. Artificial intelligence (AI) is employed to detect and classify marine litter using surface imaging datasets (RGB and hyperspectral cameras), bathymetry studies with acoustic sensors, and in situ observations at the seabed. However, the latter often occurs without sonar imaging [14].

A literature review by Politikos et al. [14] summarises studies on marine litter research, focusing on detection-based algorithms for the automatic visual recognition and identification of macroplastics. Most marine debris data are acquired from the surface. Of the eighty scientific articles surveyed by Politikos et al. [14], only eight involved underwater data collection. This underwater data collection is primarily conducted using remotely operated vehicles (ROVs) and autonomous underwater vehicles (AUV) equipped with echosounders and cameras [14,17]. Only three of these publications addressed floating debris in the water column, with the remainder focusing on seafloor studies, as do most studies that rely on echosounder data. The author specified different types of data collected, including optical and acoustic (sonar) images.

Table 1 is based on a review on marine litter surveying and contains the eight publications on underwater marine debris studies [14]. These studies evaluated how the adopted algorithms recognised targets from the collected data.

Table 1. Submerged Marine Litter Surveying, based on [14,18].

References	Sampling System	Dataset Type	Litter Domain	Task
Aleem et al. [19]	Sonar	Sonar image	Floating, Seafloor	Classification, Detection
Bajaj et al. [20]	AUV/ROV	Optical image	Seafloor	Detection
Deng et al. [21]	AUV/ROV	Optical image	Seafloor	Classification, Detection
Fossum et al. [22]	AUV/ROV	Optical image	Seafloor	Classification, Detection
Fulton et al. [23]	AUV/ROV	Optical image	Seafloor	Classification, Detection

Table 1. *Cont.*

References	Sampling System	Dataset Type	Litter Domain	Task
Hong et al. [24]	AUV/ROV	Optical image	Floating, Seafloor	Classification
Politikos et al. [18]	AUV/ROV	Optical image	Seafloor	Classification, Detection, Segmentation
Valdenegro-Toro [25]	Sonar	Sonar image	Floating	Classification, Detection

From the perspective of underwater submerged litter detection, only two studies in the survey focused on this topic using sonar imaging [19,25]; however, these studies focused on a small-scale dataset collected from the bottom of a water tank. There is a lack of field measurements of underwater plastic debris, especially when concerned with submerged plastic [26]. Some studies used echosounders to detect macroplastics in the water column. Tests by Broere et al. [17] using a low-cost commercial fish finder identified plastic objects within the extracted acoustic images in controlled and uncontrolled environments. This study demonstrated that plastics can be detected in the water column with acoustic sensors and that different macroplastics have specific signatures related to their backscatter intensity, allowing for the estimation of litter size or even classification. It was confirmed that backscatter is affected by the orientation and deformation of an object. Water flow velocity significantly impacts detection, as increased velocity decreases the exposure time to the acoustic beams, reducing detection rates. High flow velocity can deter small macroplastics, which may not produce significant reflections. Additionally, the variability of reflections from the same object can impact object classification.

2.2. Underwater Acoustic Image Classification and Detection

Since most underwater marine litter detection applications do not rely on acoustic data, most of the work in this subsection, although varying in application, comprises solutions based on acoustic images and classification methods.

Object detection similar to that performed with multibeam echosounders is often facilitated by radars, which provide comparable information and share similar working principles. Deep neural networks were employed to recognise objects in 300 GHz radar images using returned power data, akin to multibeam backscatter intensity [27]. This study examined how power data varies with range, orientation, and different receivers in a laboratory environment. Due to the limited data available from this type of sensor, transfer learning was utilised. The study explored deep learning methods for detection and classification in scenes with multiple objects.

The detection and classification of marine litter using multibeam echosounders face similar challenges to underwater object recognition tasks performed with optical and sonar images. Deep convolutional neural networks have proven effective in these tasks, but they often require extensive datasets to generalise well to unseen examples. Acquiring and labelling such large volumes of data is costly and time-consuming, particularly for rare objects or in real-time operations. Few-shot learning (FSL) methods can be promising in addressing low data availability. Recent research has compared several supervised and semi-supervised FSL methods using underwater optical and side-scan sonar imagery, demonstrating that FSL significantly outperforms traditional transfer learning methods. These insights can be applied to improve the detection and classification of marine litter with multibeam echosounders, leveraging FSL to enhance model performance despite limited data [28].

A real-time underwater object detection algorithm using forward-looking imaging sonar [29] utilised Haar-like features as a weak classifier combined with AdaBoost to make a strong classifier, improving target detection performance in echosounder images. These features can reflect changes in the grey level, allowing the identification of edges, bars, and other simple image structures [29,30]. An algorithm for automatic detection and segmentation of gas plumes from water column images (WCIs) was proposed by Zhao et al. [31]. This author also combined Haar features with local binary patterns (LBP), which are highly discriminating and invariant to monotonic grey-level images, as are WCIs. Histogram of oriented gradients (HOG) was considered for feature extraction in underwater coloured acoustic images [32]. The HOG descriptor focuses on the structure or shape of an object, generating histograms using the magnitude and orientations of the gradient for each image region. Due to the large dimension of the HOG feature vector, a large amount of redundant information was reduced. An SVM classifier trained the extracted features and chose the radial basis function as the kernel function of the SVM, which defined the optimal classification surface.

A back propagation neural network (BPNN) was developed as a classifier for seabed sediments with MBES backscatter data in [33]. Of the thirty-four dimensions extracted that made up the initial feature vector, eight features with high classification were selected, reducing the workload of the classifier and improving classification efficiency and accuracy. Particle swarm optimisation was applied to increase the global optimisation ability of the model as well as to achieve an optimal initial weight for the parameters. Another optimisation was AdaBoost, which uses weak classifiers in the training set and assigns weights based on the weak classifier's error. The author proposed the PSO-BP-AdaBoost classification algorithm and compared it with a one-level decision tree, a PSO-Backpropagation algorithm, and an SVM, achieving better accuracy than the three.

According to Valdenegro-Toro [34], classic computer vision methods like AdaBoost only work well in objects that produce large echosounder shadows, as the Haar features previously mentioned. The author proposed to detect new objects without class information, which applies to detecting hard-to-model objects such as marine debris. That was achieved by using a CNN to estimate and rank the objectness maps from sonar images, being a data-driven approach. Using forward-looking sonar (FLS) images, different types of marine litter, such as plastic bottles, cans, glass bottles, tyres, and more, were successfully detected at the bottom of a test tank by multiple authors [34,35]. Later, an adapted faster region-based convolutional neural network (R-CNN) algorithm was developed [19]. Transfer learning was applied for feature extraction to address data scarcity, incorporating the Residual Network 50 (ResNet-50) into the algorithm.

Transfer learning was used by many authors, with trained networks on large datasets such as the COCO and ImageNet datasets. A multi-branch shuttle network embedded in You Only Look Once 5 (YOLO5) was proposed to detect fish with an FLS [36], and a real-time object detection was proposed with a YOLO5 model for an obstacle avoidance algorithm in the underwater environment [37]. Transfer learning has been used to create an automatic, multi-class object classifier using data from a side-scan sonar to detect sunken shipwrecks and drowning victims, among others [38–40]. A CNN model was developed based on transfer learning, where a visual geometry group (VGG) model was pre-trained with ImageNet data and all of its trained layers. This model's last fully connected layers were transferred to a new CNN, which was fine-tuned with a semi-synthetic sonar dataset [40].

The related work on underwater acoustic image classification and detection is summarised in Table 2. The table details the related work, considering the following items: the author, the year of publication, whether the acquired data are in the water column, developed algorithms, the application and the type of sensor used.

MBES data, previously acquired in [15], were necessary due to the scarcity of sonar images for object detection and classification in the scientific community compared to optical datasets. Existing sonar datasets typically focus on specific target objects and exhibit varied representations. Additionally, most sonar images do not capture the water column or are not for marine litter detection. The acquired test tank dataset, previously characterised qualitatively, was augmented with synthetic data, including multiple targets in a single acoustic image for numerous detections. This work extends earlier efforts by developing, training, and testing a multi-label SVM classifier and a shallow CNN multi-label classification model. Transfer learning was also applied to a single-shot detector (SSD) and the YOLO8 model to detect and label marine litter in the test tank data, validating the effectiveness of marine litter detection using MBES.

Table 2. MBES acoustic data classification and detection related work summary.

Reference	Year	WCD	Task/Technique	Sensor Type	Application
Gaida [41]	2020	Yes	Statistical Classification	MBES	Acoustic sediment classification
Janowski et al. [42]	2018	Yes	SVM Classification	MBES	Benthic habitat classification
Aleem et al. [19]	2022	No	Faster-RCNN classification model with pre-trained ResNet-50 Model	Adaptive Image Resolution Sonar (ARIS) in FLS configuration	Litter classification in test tank bottom
Yu et al. [37]	2021	No	Pre-trained YOLO5 detection model	Side-Scan Sonar	Shipwreck and submerged container detector
Ge et al. [40]	2021	No	Pre-trained CNN classification model	Side-Scan and synthetic data	Detection of acoustic targets within synthetic data
Fuchs et al. [39]	2018	No	Pre-trained CNN model	Forward-looking imaging sonar	Detection of targets for obstacle avoidance
Wang et al. [36]	2022	No	Pre-trained YOLO5 detection model and YOLO5 adaption	Forward-looking imaging sonar	Detection of weak and small litter
Valdenegro-Toro [34]	2019	No	Supervised CNN detection model	ARIS in FLS configuration	Detect litter without class information
Wang et al. [32]	2018	No	SVM detection with HOG features	Colourful imaging sonar	Wood stakes detection
Zhao et al. [30]	2020	No	AdaBoost cascade detector	MBES	Detection of gas plumes
Kim and Yu [29]	2017	No	AdaBoost cascade detector	Forward-looking imaging sonar	Object detection
Ji et al. [33]	2020	No	PSO-BP-AdaBoost classifier	MBES	Acoustic sediment classification
Ochal et al. [28]	2020	No	Few shot learning	Optical and Side-scan sonar	Underwater image classification

3. Multibeam Echosounder Test Tank Acquisition Setup

This section describes the setup of the acoustic imaging system, including the characteristics of the multibeam echosounder (MBES) used, and the test tank setup to perceive marine debris targets.

3.1. System Setup

Water column data were collected using the Kongsberg M3 high-frequency model, which provides high-resolution underwater acoustic imaging composed of beams and reflections. A sonar beam is defined by its shape, with horizontal and vertical angles that vary based on the type of sonar used. This array is divided into bins that associate intensity values with their respective distances and the reflections in the environment [43]. Each acoustic image from this MBES consists of 256 beams, each with 1573 reflections. A Windows processing unit manages the data acquisition process. The Kongsberg application programming interface (API) interfaces with the MBES, handling communication with the sonar head and processing raw data, specifically for MBES beamforming. This communication occurs over a standard ethernet cable using TCP/IP commands.

A custom software module was developed within the Robot Operating System 1 (ROS1) framework. This module acquires sensor message data packages. A dedicated subscriber was implemented to capture these messages and extract essential MBES data, including range, bearing, and normalised intensity.

The MBES operates at four distinct frequencies. The field of view (FOV) changes with each frequency, affecting the angular separation between the 256 beams. Higher frequencies enhance spatial resolution but may reduce detections because of a narrower FOV. The MBES uses maximum pulse power for imaging applications, with pulse duration adjustments to increase long-range detection. This results in a lower ping rate to extend the detectable range and manage power consumption. The sonar head was configured for image-enhanced mode, using a time delay fast Fourier transform beamformer for high-quality image generation. Table 3 depicts the main features of the Kongsberg M3 multibeam high-frequency echosounder model characteristics in image enhanced mode.

Table 3. Kongsberg M3 Multibeam High-Frequency Echosounder model characteristics in Image Enhanced mode.

	Operating Frequency			
Specifications	**700 kHz**	**950 kHz**	**1200 kHz**	**1400 kHz**
Angular Resolution	140° × 30°	140° × 27°	75° × 21°	45° × 18°
Range Interval (m)	0.2–150 (m)			
Beams	256			
Reflections per beam	1573			
Pulse Type	Continuous Wave and Linear Frequency Modulation (Chirp)			

Data acquisition was carried out in a test tank, as illustrated in Figure 2a. The data acquired in [15] lacked variety and quantity for each class of selected marine debris. Tests were conducted in a controlled tank environment, where limitations allowed data acquisition at only two distances and three directions of arrival relative to the sonar head. These constraints were caused by the tank's walls, bottom, and possible surface backscatter. Range limitations were addressed by setting a maximum range based on the target's position. The MBES auto-range feature was disabled, ensuring all 1573 reflections per beam were confined within the defined range to enhance image resolution.

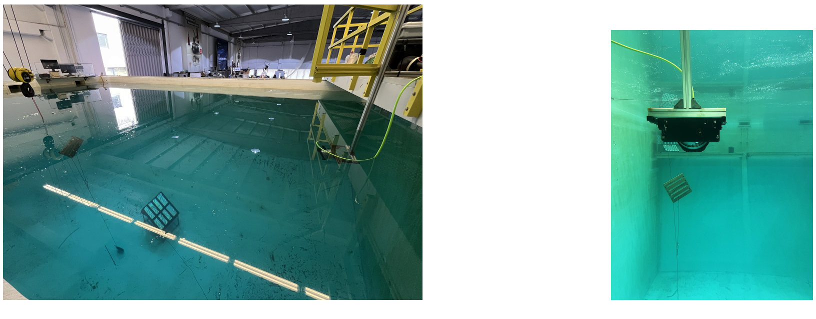
(**a**) (**b**)

Figure 2. Kongsberg M3 Multibeam High-Frequency Echosounder system setup in the test tank. (**a**) Test tank setup, (**b**) MBES capturing the Wooden deck in the water column.

During the data acquisition in [15], targets were placed approximately 2.8 m away, directly in front of the sonar head. In the latest tests, data were collected at approximately 1.66 m and 3.3 m from the sonar head at a depth of around 2.5 m. For each range, targets were positioned at angles of 0° (directly in front), −18° (left), and 18° (right). Targets were also attached to ropes in varying ways to alter their appearance in the water column.

The MBES parameters were adjusted during the tests, while time-variant and image gains remained constant across different frequencies and acquisition ranges.

The initial MBES mounting was designed for test tank conditions but later upgraded to a versatile mounting pole suitable for tank and surface vehicle applications, as illustrated in Figure 2b. This mount pole allows depth adjustment and pitch configuration for various mounting angles. Different angles were tested by adjusting the MBES pitch to evaluate their effects on the results, considering the MBES azimuth. The mounting point was designed for future compatibility with an AUV.

Targets were placed at depths that prevented the detection of the tank's bottom and walls, minimising high backscatter intensities from hard surfaces. This arrangement ensured that the highest backscatter intensities, normalised by the MBES, originated from the targets, simulating real-world conditions. The depth placements varied for each target, depending on its size, to maintain this setup. Data acquisition involved varying the frequency of the acoustic signals to generate acoustic images.

3.2. Marine Debris Selection

This study selected a collection of marine debris objects typically composed of plastic materials, such as a square made of polyvinyl chloride (PVC), for experimental testing. Various objects were chosen, some with the same squared shape, so it would be possible to study if algorithms could distinguish objects without relying solely on their shape.

The objects were selected for macroplastics detection and the NetTag+ project, with the primary objective of identifying concentrations of fish nets. Consequently, target sizes ranged from a PVC square measuring 0.5 m × 0.5 m to a smaller PVC square of 0.15 m × 0.15 m. Specific details of these objects are outlined in Table 4. Although these objects can occur naturally in the water column, their buoyancy was artificially modified for the experiments since the materials were new and had not been affected by wear or biofouling. Positively buoyant targets were weighted while negatively buoyant targets were tethered using buoys or a crane. The crane facilitated placing the targets at different fields of view (FOVs) and ranges and allowed for easy adjustment of the target's depth. These objects compose the current test tank dataset, with each object corresponding to a class.

Table 4. Marine Litter objects in the dataset.

Object	Description
PVC Square (1)	Window PVC square of 0.5 m × 0.5 m and small PVC square of 0.15 m × 0.15 m
PVC traffic cone (2)	Traffic cone lying vertically
Wooden deck (3)	Square wooden deck with slats
Vinyl sheet (4)	Thin squared shaped vinyl sheet
Fish net (5)	Agglomerate with fish nets and buoys

The PVC square class contains more occurrences due to the additional data collected with the smaller square of the same material. This enabled testing whether the material could still be detected at varying ranges and directions, as well as assessing the potential for false positives.

Each type of debris in Table 4 has a number that can be mapped in Figure 3.

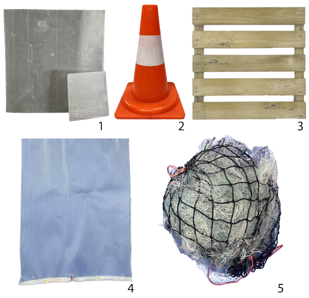

Figure 3. Marine debris used for the test tank dataset. PVC Squares (1); PVC traffic cone (2); Wooden deck (3); vinyl sheet (4); fish net (5).

3.3. High Level Architecture

The acoustic imaging for detection and classification problems of high-level architecture is summarised in Figure 4. After raw data acquisition, data are processed into three different acoustic imaging representations, which will be detailed in Section 4. Labelling and annotations were performed and tailored for the specific tasks and model architectures, enabling the training and validation of different types of networks.

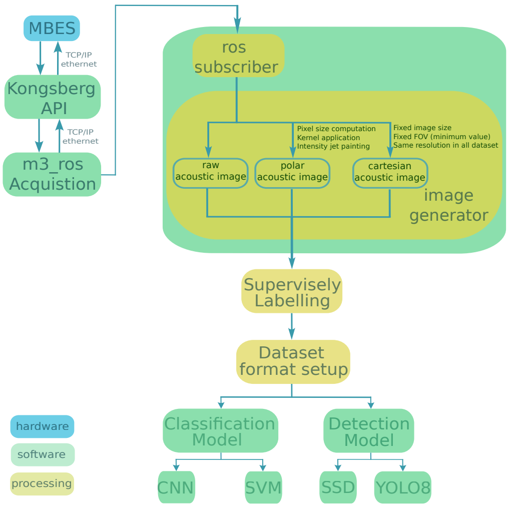

Figure 4. High-level architecture for the MBES sensor and acoustic imaging for detection and classification problems.

4. Acoustic Images Dataset

This section outlines water column image processing. Section 4.1 describes three representations of acoustic images. Section 4.2 details the labelling and formatting of datasets for different models, ensuring readiness for training and validation.

4.1. Acoustic Images Representation

The experimental setup described in Section 3 enabled the extraction of raw MBES data, represented in the `PointCloud2` format. This format includes the following fields:

- x, y: Acoustic image pixel coordinates.
- *height*: Acoustic image height representing 1573 backscatter points.
- *width*: Acoustic image width representing 256 beams.
- *intensity*: Normalised backscatter intensity.

Three types of acoustic image representations were generated: raw, polar, and Cartesian. The raw and polar representations were tested as inputs for multiple models to evaluate, which led to better results. The raw representation was designed for real-time visualisation of MBES acoustic images due to its low storage requirements and fast generation. The polar representation was created to provide a format that is more interpretable for human operators. The following subsections detail each type of representation.

4.1.1. Raw Acoustic Image

The raw acoustic image representation, depicted in Figure 5, is a direct mapping of the normalised intensity of each pixel to its corresponding x and y coordinates, being composed of a single channel (greyscale). This representation has a fixed resolution of (256 × 1573) and ensures consistent pixel spacing. However, the varying field of view (FOV) associated with different acoustic frequencies and the changing range of targets relative to the MBES induce distortions in target shapes, as shown in Figure 5. This illustration kept the range constant, yet the resulting acoustic images differ significantly due to the varying field of view (FOV). Although the range, angle, and intensity data are preserved, this representation distorts target shapes, making it more challenging to distinguish targets, especially with varying ranges. This deformation makes it harder to use the objects' shape as a feature for classification and detection models. Nonetheless, this method enables fast image generation due to the simplicity of its creation, which could be used for real-time visualisation.

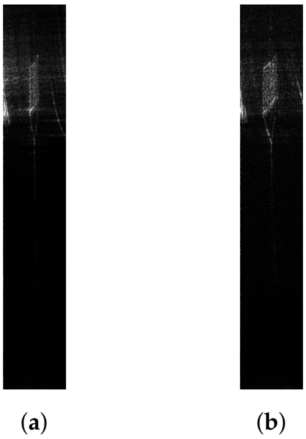

Figure 5. Raw acoustic images of a PVC square at the same range, with varying FOV due to the different acoustic frequencies. (**a**) Raw acoustic image of 1200 kHz, (**b**) Raw acoustic image of 1400 kHz.

4.1.2. Cartesian Acoustic Image

The Cartesian acoustic image shown in Figure 6 facilitates the identification of target locations, and sonar operators typically use it. In this representation, the range is associated with the $y-$axis, with lower y values indicating closer targets. The $x-$axis represents the bearing relative to the sonar head, with the centre corresponding to a target directly in front, having a direction of arrival (DoA) of 0°. The entire field of view is divided in half, each representing left and right bearings, providing spatial orientation information. This representation is based on variable pixel sizes influenced by the minimum distance between acoustic backscatter points, which affect the FOV and range parameters. Consequently, the image size varies due to these pixel size differences. To mitigate gaps in the data, a kernel is applied to the image to interpolate empty pixels. While this enhances visualisation clarity, many interpolated pixels lack precise intensity data. This Cartesian representation was developed as an alternative to the `rviz` ROS package, enhancing the visualisation of captured targets, particularly with adjustable kernel sizes. Additionally, post-processing was made, where a jet colour map was used on the normalised intensity data. This representation was not utilised for classification or detection models, and the image generation by this method is slower than the other two methods due to the kernel application.

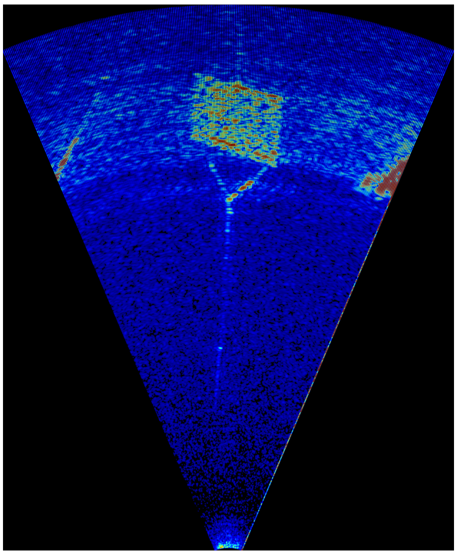

Figure 6. Cartesian acoustic image of a PVC square in the water column.

4.1.3. Polar Acoustic Image

The Polar acoustic image illustrated in Figure 7 addresses the issues present in the raw representation by offering a consistent and interpretable format.

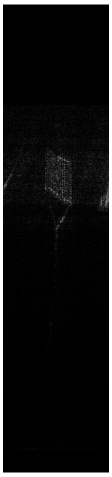

Figure 7. polar acoustic image of a PVC square in the water column.

The x and y coordinates of points, along with their intensity values, are used as input to generate a two-dimensional (2D) image representing these points in polar space. The maximum accepted field of view (FOV) and range are considered. A fixed FOV is applied for all frequencies to ensure all acoustic images have the same fixed width without pixels lacking intensity values. The algorithm computes the range (r) and angle (θ) for each point (i). The range is computed using the Euclidean distance formula:

$$range_i = \sqrt{x_i^2 + y_i^2} \tag{1}$$

The angle (θ) is computed using the arctangent function:

$$theta_i = arctan2(y_i, x_i) \tag{2}$$

The range resolution is defined as 0.002 m/pixel based on specifications, and the theta resolution is set to 0.1°, given by:

$$theta_{resolution} = \frac{minimum_{fov}}{beams_{number}} \tag{3}$$

The dimensions of the output image are determined based on these resolutions and the maximum values, ensuring consistent image resolution without compromising scale:

$$image_{height} = \frac{range_{max}}{range_{resolution}} + 1 \tag{4}$$

$$image_{width} = \frac{fov_{max}}{theta_{resolution}} + 1 \tag{5}$$

Two mapping operations convert polar coordinates to polar image indices. Equation (6) converts the angular position (θ) into a column index (x) within the polar image:

$$x = \left\lfloor \left(\frac{image_{width} - 1}{fov_{max}} \right) \theta + \frac{image_{width} - 1}{2} \right\rfloor \tag{6}$$

The current angle (θ) is scaled to fit within the image width, defined as the first member of Equation (6), and then centred to ensure symmetrical angular positions across the image width.

Equation (7) converts the radial distance (r) into a row index (y) within the polar image:

$$y = \left\lfloor \left(\frac{image_{height} - 1}{range_{max}} \right) r \right\rfloor \quad (7)$$

The range (r) is scaled to fit within the image height. The mapped indices assign each point's normalised intensity to its corresponding polar image position. This process iterates over all points represented in polar coordinates, effectively populating the polar image with intensity values based on their spatial distribution.

This method allows for generating acoustic images that facilitate the computation of a target's range and bearing relative to the sonar head. The varying FOV with different acoustic frequencies does not affect the resulting image dimensions or scale due to the fixed range and theta resolutions. Consistent image dimensions are beneficial for input into classification and detection models.

4.2. Dataset Labelling and Format

The dataset comprises five classes, as detailed in Table 4. Three acoustic frequencies were selected for the dataset: 950 kHz, 1200 kHz, and 1400 kHz. The 700 kHz frequency was excluded due to excessive noise in the MBES backscatter data, likely caused by the large field of view (FOV) and side-lobe interference from the test tank walls. The number of images per class at each frequency is presented in Table 5. The dataset is primarily balanced, except for the PVC square, which has higher occurrences.

Table 5. Test tank dataset with its classes and number of occurrences at each operating frequency.

		Operating Frequency	
Class	**950 kHz**	**1200 kHz**	**1400 kHz**
PVC Square	551	549	453
PVC traffic cone	317	356	305
Wooden deck	356	355	344
Vinyl sheet	356	355	425
Fish net	301	330	313

For classification tasks, images were stored in directories named after their respective classes, facilitating organised dataset management and straightforward class identification during training. Simple data augmentation techniques, including horizontal flips and brightness adjustments, were applied to enhance the dataset. Since the data consist of sonar images, horizontal flips were used to simulate different fields of view (FOVs). Brightness adjustments were made to mimic variations in backscatter intensity. These transformations aimed to improve generalisation and have been discussed by other authors [44,45]. Although synthetic data augmentation methods, such as generative adversarial networks (GANs), were considered, they were not implemented in the current scope of development [46].

Labelling and annotation were necessary for detection tasks due to the presence of bounding boxes. This was accomplished using the Supervisely platform, where the bounding boxes were annotated for the entire dataset, including augmented data. The augmented dataset included images with multiple objects and some without targets. Data augmentation techniques were also applied to this dataset, including horizontal and vertical flips, zoom, and brightness adjustments. The Common Objects in Context (COCO) format was used for detection tasks, and a specific COCO format adaptation was made for the You Only Look Once 8 (YOLOv8) model.

This structured approach to dataset labelling and annotation, combined with data augmentation, ensures that both classification and detection models are trained on diverse and comprehensive datasets, enhancing their performance and robustness.

5. Detection and Classification

This section details the detection and classification algorithms implemented and tested to detect marine debris objects in the water column acoustic images. Section 5.1 describes the chosen algorithms based on the state-of-the-art approaches. Section 5.2 details how the algorithms were implemented and their results.

5.1. Algorithms for Detection and Classification Problems

The complexity and variability of marine litter, along with the characteristics of sonar images, present significant challenges for developing accurate and efficient classification and detection methods. The selection of algorithms was based on the nature of the acoustic images, their footprints, and the review of the state-of-the-art techniques and where they were commonly applied. Machine learning and deep learning methods, typically used for this type of data, can offer promising solutions. Selecting appropriate algorithms is crucial for achieving high accuracy and efficiency. Testing these algorithms with the acquired data was necessary to determine if it was possible to accurately discriminate the selected marine debris in the test tank dataset.

SVM and CNN were chosen for the multi-class classification of sonar acoustic images. The decision to use SVM is based on its robustness in high-dimensional feature spaces, which is advantageous for the complex features in sonar images. Although SVMs can be slower to train with large datasets, efficient kernels, such as the Radial Basis Function (RBF), mitigate this issue [47].

CNNs were also selected due to their ability to capture spatial hierarchies in images through convolutional layers. This capability allows CNNs to automatically learn and extract relevant features from images, reducing the need for manual feature engineering. CNNs have demonstrated superior performance in image classification tasks due to their proficiency in modelling complex patterns. Despite requiring substantial computational resources and time for training, CNNs generally achieve higher accuracy than traditional methods like SVM [48].

YOLO (You Only Look Once) version 8 and the single-shot multibox detector (SSD) were chosen for object detection in sonar acoustic images. YOLOv8 is designed for real-time detection, treating detection as a single regression problem that directly predicts bounding boxes and class probabilities. This design facilitates efficient processing, maintaining a good balance between speed and accuracy, essential for real-time applications. While YOLOv8 may not achieve the highest accuracy among detection models, its efficiency makes it a practical choice for applications requiring real-time processing [49].

The SSD balances speed and accuracy, making it suitable for applications where both factors are essential. SSD uses feature maps at different scales to detect objects of various sizes, enhancing detection performance. Like YOLO, SSD requires only a single forward pass through the network to detect objects, ensuring efficiency. Although SSD is more complex due to its use of multiple feature maps and default boxes, it allows for greater flexibility in detecting objects of different sizes [50].

Three acquisitions were conducted, each corresponding to one of the three operating frequencies: 950 kHz, 1200 kHz, and 1400 kHz. Consequently, three models were trained for each frequency. The training was performed on an NVIDIA GeForce RTX 3060 Mobile GPU. The development environment primarily utilised TensorFlow 2, with specific APIs employed for different models. The SVM model was implemented using the `sklearn` library, and the CNN model was built with `keras`. The detector models were trained using transfer learning from initial weights derived from the COCO dataset, including the SSD and the YOLOv8. The SSD model was trained using the TensorFlow 2 Detection API, the YOLO8 model was trained using the Supervisely platform, and the inference was performed using the `ultralytics` library.

The machine and deep learning algorithms used in this study are widely applied in state-of-the-art research across various domains. The focus of this work was not on developing new algorithms but rather on applying proven learning-based methods to this specific application. The successful characterisation and detection of the proposed targets in a controlled environment serve as essential groundwork for future real-world experiments.

5.2. Training and Results

5.2.1. Support Vector Machine (SVM)

An algorithm was developed to optimise the performance of the support vector machine (SVM) model through grid search. The process involves systematically exploring and evaluating hyperparameter combinations to identify the best-performing configurations. The aim was to ensure that the SVM classifier is robust and generalises unseen data well.

The algorithm considers three critical hyperparameters for optimisation: the regularisation parameter (`C`), the kernel coefficient (`gamma`), and the kernel types (that were chosen taking into account that it is a multi-class classification problem):

- `C`—varies from 0.2 to 1.6 with a step of 0.2.
- `gamma`—varies from 0.25 to 2.0 with a step of 0.25.
- Kernel types—polynomial (poly), radial basis function (RBF), and sigmoid.

The combined dataset, made of the original and augmented data, was split into training, validation, and test sets. This step ensures that the model has sufficient data to learn from while also providing separate sets for validation and testing to evaluate model performance.

The grid search algorithm evaluated all possible combinations of the specified hyperparameters using 2-fold cross-validation (CV) for computational efficiency. The mean performance scores across the cross-validation folds are recorded for each combination. These scores are analysed to identify the best hyperparameters for each performance metric. The algorithm also evaluates the model performance based on accuracy, precision, and recall. Custom scoring functions are defined to assess precision and recall with weighted averaging, which accounts for class imbalances. The optimal hyperparameters for accuracy, precision, and recall are identified, defining an optimal classification surface. If different hyperparameters are optimal for each metric, multiple models are trained, each optimised for a specific metric. This ensures that the best possible model is available for different evaluation criteria. If the optimal parameters are the same across all metrics, a single model is trained with these parameters.

Inference was applied to the dataset corresponding to the frequency the model was trained on and the data of the other frequencies. This cross-frequency evaluation helps determine the models' robustness and adaptability when applied to data acquired under different acoustic conditions. The results of each model from the other acoustic frequencies data are available in Table 6. The models trained on a specific frequency demonstrated overfitting when tested with acoustic images of the same frequency. This overfitting can be attributed to the low variability of the images obtained in a controlled environment. In such settings, the images did not capture objects from different angles, resulting in limited variations in the object footprints, underlining the current dataset limitations. This is notorious across all trained models. This will be addressed in the near future, where MBES data will be acquired from an unmanned surface vehicle (USV) in a real-life controlled environment where targets will be placed in a harbour within known locations.

Table 6. Performance Metrics for SVM Models with different multibeam acoustic frequencies.

Class	Precision	Recall	F1-Score	Accuracy
950 kHz SVM model				
950 kHz data				
Fish net	1.00	1.00	1.00	1.00
Wooden deck	1.00	1.00	1.00	
PVC traffic cone	1.00	1.00	1.00	
Vinyl	1.00	1.00	1.00	
PVC square	1.00	1.00	1.00	

Table 6. *Cont.*

Class	Precision	Recall	F1-Score	Accuracy
1200 kHz data				
Fish net	0.00	0.00	0.00	0.19
Wooden deck	0.19	1.00	0.32	
PVC traffic cone	0.00	0.00	0.00	
Vinyl	0.00	0.00	0.00	
PVC square	0.00	0.00	0.00	
1400 kHz data				
Fish net	0.00	0.00	0.00	0.19
Wooden deck	0.19	1.00	0.32	
PVC traffic cone	0.00	0.00	0.00	
Vinyl	0.00	0.00	0.00	
PVC square	0.00	0.00	0.00	
1200 kHz SVM model				
950 kHz data				
Fish net	0.72	0.67	0.69	0.50
Wooden deck	0.46	0.74	0.56	
PVC traffic cone	0.31	0.35	0.33	
Vinyl	0.51	0.51	0.51	
PVC square	0.56	0.32	0.41	
1200 kHz data				
Fish net	1.00	1.00	1.00	1.00
Wooden deck	1.00	1.00	1.00	
PVC traffic cone	1.00	1.00	1.00	
Vinyl	1.00	1.00	1.00	
PVC square	1.00	1.00	1.00	
1400 kHz data				
Fish net	1.00	0.16	0.28	0.26
Wooden deck	0.23	1.00	0.38	
PVC traffic cone	0.54	0.11	0.19	
Vinyl	0.00	0.00	0.00	
PVC square	0.40	0.10	0.16	
1400 kHz SVM model				
950 kHz data				
Fish net	0.38	0.81	0.52	0.42
Wooden deck	0.66	0.36	0.46	
PVC traffic cone	0.26	0.40	0.32	
Vinyl	0.44	0.45	0.44	
PVC square	1.00	0.18	0.31	
1200 kHz data				
Fish net	0.97	0.18	0.30	0.35
Wooden deck	0.22	0.65	0.32	
PVC traffic cone	0.35	0.25	0.29	
Vinyl	0.50	0.12	0.20	
PVC square	0.56	0.54	0.55	
1400 kHz data				
Fish net	1.00	1.00	1.00	1.00
Wooden deck	1.00	1.00	1.00	
PVC traffic cone	1.00	1.00	1.00	
Vinyl	1.00	1.00	1.00	
PVC square	1.00	1.00	1.00	

5.2.2. Convolutional Neural Network (CNN)

A convolutional neural network (CNN) algorithm was developed, and class activation maps (CAM) were employed to identify the most informative regions in the images used by the model. The images and labels were split into training, validation, and test sets. The labels are one-hot encoded to facilitate multi-class classification.

The CNN architecture comprises several layers: an input layer and two convolutional layers with 32 and 64 filters, each with a kernel size of 3 × 3 and rectified linear unit (ReLU) activation functions. Max-pooling layers follow each convolutional layer to reduce the spatial dimensions of the feature maps. The output from these layers is then flattened and passed through a dense layer with 64 units and ReLU activation. The final output layer uses a softmax activation function to produce a probability distribution over the classes. The model is compiled using the Adam optimizer, categorical cross-entropy as the loss function, and metrics including accuracy, precision, and recall.

The model's performance is evaluated on the test set, generating key metrics such as test loss, accuracy, precision, and recall. Class activation maps (CAM) were generated to visualize the regions in the input images that CNN focused on when making its predictions. This helps in understanding which parts of the pictures are most influential in the model's decision-making process, providing insight into the model's interpretability.

Each trained model was evaluated on datasets corresponding to different acoustic frequencies to assess generalisation. The models were trained using both raw and polar acoustic image representations. The results for each acoustic image representation are summarised in Table 7 and Table 8, respectively. This cross-frequency evaluation, similar to that applied in the support vector machine (SVM) models, aimed to determine the robustness and adaptability of the CNN models when applied to data acquired under different acoustic conditions.

The models trained on polar acoustic images were less prone to overfitting compared to those trained on raw images, a trend observed across all frequencies. While these models still showed signs of overfitting, the effect was less pronounced than with the SVM model. The generation of CAMs provided insight into the models' focus areas, revealing whether they concentrated on pixels and features associated with backscatter data from marine debris targets. Upon generating the CAMs, it was confirmed that the marine debris objects and their backscatter significantly influenced the model's decision-making process, as illustrated in Figure 8. This finding underscored the importance of the marine debris and their backscatter in the model's predictions, affirming that the models primarily relied on these features despite the overfitting issue.

Table 7. Performance Metrics for CNN Models with different multibeam acoustic frequencies with the raw acoustic image representation.

Class	Precision	Recall	F1-Score	Accuracy
		950 kHz raw CNN model		
		950 kHz data		
Fish net	1.00	0.70	0.82	0.77
Wooden deck	0.45	1.00	0.62	
PVC traffic cone	1.00	0.51	0.68	
Vinyl sheet	1.00	0.96	0.98	
PVC Square	1.00	0.69	0.82	
		1200 kHz data		
Fish net	0.52	0.81	0.63	0.56
Wooden deck	0.70	0.77	0.74	
PVC traffic cone	0.48	0.33	0.39	
Vinyl sheet	0.41	0.54	0.47	
PVC Square	0.82	0.39	0.53	

Table 7. *Cont.*

Class	Precision	Recall	F1-Score	Accuracy
		1400 kHz data		
Fish net	0.36	0.87	0.51	0.46
Wooden deck	0.60	0.83	0.69	
PVC traffic cone	0.12	0.02	0.03	
Vinyl sheet	0.43	0.34	0.38	
PVC Square	0.62	0.29	0.39	
		1200 kHz raw CNN model		
		950 kHz data		
Fish net	1.00	0.47	0.64	0.46
Wooden deck	1.00	0.53	0.69	
PVC traffic cone	0.30	0.28	0.29	
Vinyl sheet	0.36	0.62	0.45	
PVC Square	0.34	0.41	0.37	
		1200 kHz data		
Fish net	0.88	1.00	0.94	0.94
Wooden deck	0.99	0.83	0.90	
PVC traffic cone	0.96	0.88	0.92	
Vinyl sheet	0.89	0.99	0.93	
PVC Square	1.00	1.00	1.00	
		1400 kHz data		
Fish net	0.65	0.94	0.77	0.51
Wooden deck	0.60	0.42	0.49	
PVC traffic cone	0.04	0.02	0.02	
Vinyl sheet	0.33	0.71	0.45	
PVC Square	1.00	0.45	0.62	
		1400 kHz raw CNN model		
		950 kHz data		
Fish net	1.00	0.19	0.32	0.33
Wooden deck	0.51	0.31	0.39	
PVC traffic cone	0.03	0.00	0.01	
Vinyl sheet	0.49	0.31	0.38	
PVC Square	0.25	0.74	0.38	
		1200 kHz data		
Fish net	1.00	0.21	0.34	0.55
Wooden deck	0.72	0.45	0.55	
PVC traffic cone	0.61	0.52	0.56	
Vinyl sheet	0.42	0.92	0.57	
PVC Square	0.58	0.61	0.60	
		1400 kHz data		
Fish net	1.00	0.96	0.98	0.96
Wooden deck	1.00	0.98	0.99	
PVC traffic cone	0.83	1.00	0.91	
Vinyl sheet	0.97	1.00	0.98	
PVC Square	1.00	0.88	0.94	

Table 8. Performance Metrics for CNN Models with different multibeam acoustic frequencies with the polar acoustic image representation.

Class	Precision	Recall	F1-Score	Accuracy
		950 kHz polar CNN model		
		950 kHz data		
Fish net	1.00	1.00	1.00	0.98
Wooden deck	1.00	0.99	0.99	
PVC traffic cone	0.93	1.00	0.96	
Vinyl sheet	0.97	1.00	0.99	
PVC Square	1.00	0.95	0.97	
		1200 kHz data		
Fish net	0.85	0.83	0.84	0.65
Wooden deck	0.92	0.76	0.83	
PVC traffic cone	1.00	0.42	0.59	
Vinyl sheet	0.36	1.00	0.53	
PVC Square	0.96	0.41	0.57	
		1400 kHz data		
Fish net	0.57	0.71	0.63	0.59
Wooden deck	0.67	0.62	0.64	
PVC traffic cone	0.55	0.47	0.50	
Vinyl sheet	0.49	0.83	0.61	
PVC Square	1.00	0.35	0.52	
		1200 kHz polar CNN model		
		950 kHz data		
Fish net	0.85	0.75	0.85	0.76
Wooden deck	0.95	0.66	0.80	
PVC traffic cone	0.68	0.84	0.75	
Vinyl sheet	0.86	0.55	0.71	
PVC Square	0.61	0.93	0.73	
		1200 kHz data		
Fish net	0.99	1.00	0.99	0.97
Wooden deck	1.00	0.83	0.91	
PVC traffic cone	0.90	1.00	0.95	
Vinyl sheet	0.97	1.00	1.00	
PVC Square	0.98	1.00	0.99	
		1400 kHz data		
Fish net	0.67	0.99	0.80	0.67
Wooden deck	0.70	0.37	0.48	
PVC traffic cone	0.48	0.98	0.64	
Vinyl sheet	1.00	0.47	0.64	
PVC Square	0.82	0.63	0.71	
		1400 kHz polar CNN model		
		950 kHz data		
Fish net	0.66	0.66	0.66	0.63
Wooden deck	0.77	0.55	0.64	
PVC traffic cone	0.53	0.82	0.64	
Vinyl sheet	0.64	0.79	0.71	
PVC Square	0.62	0.44	0.51	

Table 8. *Cont.*

Class	Precision	Recall	F1-Score	Accuracy
1200 kHz data				
Fish net	0.90	0.88	0.89	0.75
Wooden deck	0.66	0.18	0.28	
PVC traffic cone	0.80	0.75	0.78	
Vinyl sheet	0.51	1.00	0.68	
PVC Square	0.97	0.88	0.92	
1400 kHz data				
Fish net	0.95	1.00	0.98	0.90
Wooden deck	0.91	0.91	0.91	
PVC traffic cone	0.85	0.98	0.91	
Vinyl sheet	0.84	0.93	0.88	
PVC Square	0.91	0.74	0.85	

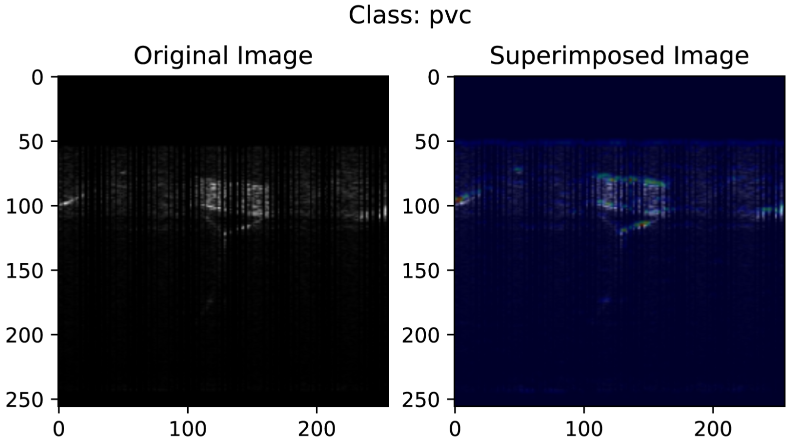

Figure 8. Class Activation Map applied to the CNN with a polar image of a PVC square as an input.

5.2.3. Single Shot Detector (SSD)

A single-shot multiBox detector (SSD) algorithm was implemented to detect objects in images across different acoustic frequencies. The workflow encompasses data preprocessing, model training, evaluation, and visualisation of the results using bounding boxes.

The data were loaded and pre-processed using annotations in COCO format. The data were divided into training and validation sets, ensuring the data distribution was maintained across sets. Annotations and images were processed to create TFRecord files to use the TensorFlow 2 Object Detection API. This step involved reading image files, encoding them, and normalising the bounding box coordinates. The annotations included bounding box coordinates and class labels.

The model architecture included a pre-trained backbone, RetinaNet50, followed by multiple convolutional layers to predict bounding boxes and class probabilities. The training involved optimising a multi-task loss function, a weighted sum of localisation loss and confidence loss. The Adam optimiser was used to minimise this loss function, facilitating efficient convergence.

The SSD model was loaded, and inference was performed on test images. The inference process involved passing images through the model and extracting bounding box predictions, class labels, and confidence scores. The COCO evaluation toolkit was integrated to compare the predicted bounding boxes with ground truth annotations, facilitating a detailed performance analysis. The results are summarised in Table 9.

It is noticeable from the results that the average precision (AP) for the intersection over union (IoU) for both thresholds is higher when the models that were trained with a specific acoustic frequency detect targets within images from the same frequency. The same goes for all the other metrics collected and showcased in Table 9. The results are visualised using bounding boxes drawn on the images, as illustrated in Figure 9, with scores above the 0.5 thresholds considered as valid detections. Cross-frequency detection seems unlikely for detection models.

Table 9. Performance Metrics for SSD Models with different multibeam acoustic frequencies with the polar acoustic image representation.

Class	Precision	Recall	mAP	AP for IoU Threshold at 0.6	AP for IoU Threshold at 0.75
950 kHz SSD model					
950 kHz data					
PVC traffic cone	0.8657	0.8712	0.8680	0.9693	0.8680
PVC Square	0.8440	0.8489	0.8494	0.5347	0.8494
Fish net	0.7104	0.7222	0.7186	0.7123	0.7186
Wooden deck	0.9704	0.9769	0.9780	0.9802	0.9780
Vinyl sheet	0.9831	0.9900	0.9872	0.958	0.9872
1200 kHz data					
PVC traffic cone	0.1725	0.3354	0.2843	0.3248	0.2843
PVC Square	0.0012	0.0039	0.0012	0.0	0.0012
Fish net	0.2353	0.3518	0.2631	0.0	0.2631
Wooden deck	0.0	0.0	0.0	0.0	0.0
Vinyl sheet	0.0245	0.0820	0.0295	0.6505	0.0295
1400 kHz data					
PVC traffic cone	0.0016	0.0047	0.0195	0.0297	0.0195
PVC Square	0.0	0.0	0.0	0.0	0.0
Fish net	0.0	0.0	0.0	0.0	0.0
Wooden deck	0.0049	0.0126	0.0172	0.0713	0.0172
Vinyl sheet	0.0209	0.0323	0.0295	0.0	0.0295
1200 kHz SSD model					
950 kHz data					
PVC traffic cone	0.2799	0.3509	0.2889	0.5049	0.2889
PVC Square	0.0042	0.0076	0.0270	0.0000	0.0270
Fish net	0.3153	0.4266	0.3528	0.0	0.3528
Wooden deck	0.0000	0.0000	0.0000	0.0000	0.0000
Vinyl sheet	0.0000	0.0000	0.0000	0.0000	0.0000
1200 kHz data					
Fish net	0.9501	0.9832	0.9720	-1.0	0.9720
PVC traffic cone	0.9716	0.9801	0.9789	0.9822	0.9789
PVC Square	0.7256	0.7423	0.7332	0.5743	0.7332
Wooden deck	0.8730	0.8810	0.8812	1.0000	0.8812
Vinyl sheet	0.9633	0.9900	0.9762	0.0000	0.9762
1400 kHz data					
Fish net	0.0	0.0	0.0	0.0	0.0
PVC traffic cone	0.0	0.0	0.0	0.0	0.0
PVC Square	0.0004	0.0008	0.0143	0.0218	0.0143
Wooden deck	0.0023	0.0029	0.0456	0.0000	0.0456
Vinyl sheet	0.1963	0.2766	0.2695	0.0000	0.2695

Table 9. *Cont.*

Class	Precision	Recall	mAP	AP for IoU Threshold at 0.6	AP for IoU Threshold at 0.75
			1400 kHz SSD model		
			950 kHz data		
PVC traffic cone	0.0009	0.0017	0.0138	0.0049	0.0138
PVC Square	0.2227	0.3113	0.3111	0.0000	0.3111
Wooden deck	0.0021	0.0064	0.0384	0.0000	0.0384
Fish net	0.0029	0.0048	0.0159	0.0178	0.0159
Vinyl sheet	0.0005	0.0014	0.0242	0.0000	0.0242
			1200 kHz data		
PVC traffic cone	0.0003	0.0003	0.0160	0.0198	0.0160
PVC Square	0.0129	0.0355	0.1301	0.3307	0.1301
Wooden deck	0.0067	0.0070	0.1225	0.0000	0.1225
Fish net	0.0391	0.0472	0.1714	0.1772	0.1714
Vinyl sheet	0.5067	0.6278	0.5426	0.8000	0.5426
			1400 kHz data		
Fish net	0.9847	0.9901	0.9930	0.9871	0.9930
PVC traffic cone	0.9791	0.9900	0.9862	0.9851	0.9862
PVC Square	0.9425	0.9501	0.9479	0.8752	0.9479
Wooden deck	0.8700	0.8713	0.8739	0.4950	0.8739
Vinyl sheet	0.8896	0.9092	0.9017	0.9000	0.9017

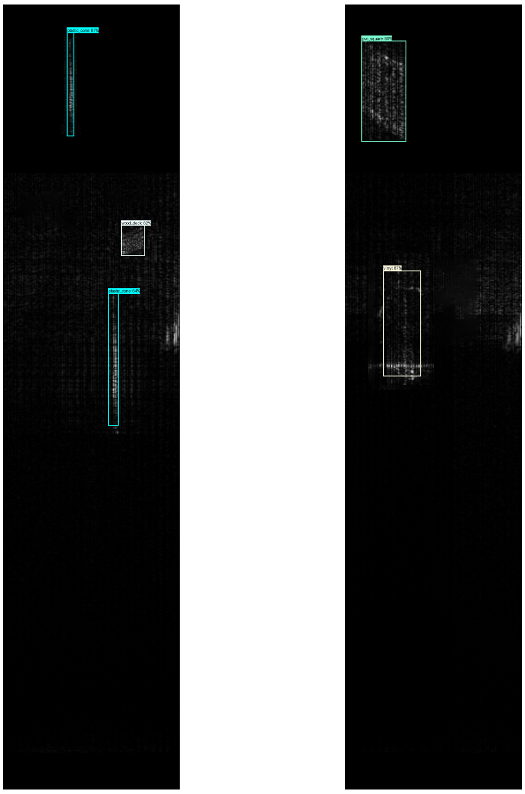

Figure 9. SSD model inference in two polar acoustic images with multiple targets with the target detection confidence.

5.2.4. You Only Look Once 8 (YOLOv8)

The YOLOv8 model, pre-trained using the Supervisely platform, is loaded from a checkpoint file containing weights for the previously labelled dataset. This model is configured to detect objects in images with a confidence threshold of 0.5 to filter predictions.

Inference begins by loading and iterating through a set of photos from a predefined directory. The YOLOv8 model processes each image to detect objects, producing bounding boxes and c for each detected object. The detection results are saved as images with highlighted objects, and the bounding boxes are extracted and stored for further analysis. The YOLOv8 model's performance is evaluated using metrics such as precision, recall, and F1 score metrics to assess accuracy and reliability.

During training, the YOLOv8 model was fine-tuned for 100 epochs with a patience of 50 epochs and a batch size of 16. The input image size was 640 pixels. The AdamW optimiser was used with an initial learning rate of 0.01, a final learning rate of 0.01, a momentum of 0.937, and a weight decay of 0.0005. The training utilised automatic mixed precision (AMP). Data augmentations included HSV-Hue augmentation (0.015), HSV-Saturation augmentation (0.7), and HSV-Value augmentation (0.4). Eight worker threads were used for data loading to ensure efficient training.

The results are summarised in Table 10.

Table 10. Performance Metrics for YOLOv8 Models with different multibeam acoustic frequencies with the polar acoustic image representation.

Class	Precision	Recall	F1	mAP	Fitness
950 kHz YOLOv8 model					
950 kHz data					
Fish net	1.00	1.00	1.00	0.9921	0.9603
PVC traffic cone	0.9303	0.9966	0.9623	0.9186	
Wooden deck	0.9925	1.00	0.9962	0.9757	
Vinyl sheet	0.9915	1.00	0.9957	0.9911	
PVC Square	0.9925	0.9213	0.9556	0.9100	
1200 kHz data					
Fish net	0.7490	0.6316	0.6853	0.2702	0.0695
PVC traffic cone	0.0157	0.0120	0.0136	0.0031	
Wooden deck	0.1016	0.1000	0.1008	0.0260	
Vinyl sheet	0.00	0.00	0.00	0.00	
PVC Square	0.00	0.00	0.00	0.00	
1400 kHz data					
Fish net	0.0134	0.9630	0.0264	0.0119	0.0031
PVC traffic cone	0.0025	0.50	0.0050	0.0012	
Wooden deck	0.00	0.00	0.00	0.00	
Vinyl sheet	0.00	0.00	0.00	0.00	
PVC Square	0.0015	0.0714	0.0030	0.0002	
1200 kHz YOLOv8 model					
950 kHz data					
Fish net	0.9333	1.00	0.9655	0.6900	0.1599
PVC traffic cone	0.0268	0.9545	0.0520	0.0488	
Wooden deck	0.0238	0.9565	0.0464	0.0133	
Vinyl sheet	0.0097	0.9259	0.0193	0.0010	
PVC Square	0.0	0.0	0.0	0.0	
1200 kHz data					
Fish net	0.9906	1.00	0.9953	0.9950	0.9789
PVC traffic cone	1.00	0.9958	0.9979	0.9279	
Wooden deck	0.9943	1.00	0.9971	0.9831	
Vinyl sheet	0.9980	1.00	0.9990	0.9905	
PVC Square	0.9945	1.00	0.9972	0.9887	

Table 10. *Cont.*

Class	Precision	Recall	F1	mAP	Fitness
1400 kHz data					
Fish net	1.00	0.00	0.00	0.00	0.1024
PVC traffic cone	1.00	0.00	0.00	0.00	
Wooden deck	0.00	0.00	0.00	0.0751	
Vinyl sheet	0.7445	0.9048	0.8169	0.3754	
PVC Square	0.00	0.00	0.00	0.00	
1400 kHz YOLOv8 model					
950 kHz data					
Fish net	0.00	0.00	0.00	0.0047	0.1109
PVC traffic cone	1.00	0.00	0.00	0.00	
Wooden deck	0.00	0.00	0.00	0.00	
Vinyl sheet	0.00	0.00	0.00	0.00	
PVC Square	0.9494	0.9130	0.9309	0.5055	
1200 kHz data					
Fish net	1.00	0.00	0.00	0.00	0.0891
PVC traffic cone	1.00	0.00	0.00	0.04	
Wooden deck	1.00	0.00	0.00	0.00	
Vinyl sheet	0.9108	0.9167	0.9137	0.3401	
PVC Square	1.00	0.00	0.00	0.00	
1400 kHz data					
Fish net	0.7756	1.00	0.8736	0.8568	0.7822
PVC traffic cone	0.8707	0.7921	0.8295	0.5482	
Wooden deck	0.9857	1.00	0.9928	0.8250	
Vinyl sheet	0.7938	0.9529	0.8661	0.7661	
PVC Square	1.00	0.8476	0.9175	0.8106	

Bounding boxes were drawn on the images, as illustrated in Figure 10, with scores above the 0.5 thresholds considered valid detections. As with the SSD models, it is noticeable that the mean average precision (mAP) and the fitness metric are higher when the models trained with a specific acoustic frequency detect targets within images from the same frequency, as with the other metrics. The results are visualised using bounding boxes drawn on the images, as illustrated in Figure 10, with scores above the 0.5 thresholds considered valid detections.

5.3. Results Discussion

The original dataset from [15] consisted of targets captured directly in front of the sonar head at multiple frequencies. However, the dataset lacked variation in fields of view and ranges relative to the sonar head. As a result, most models overfitted when tested on acoustic images using the same frequency as the training set, particularly in classification tasks, where results of 1.0 were observed. To mitigate overfitting and improve model generalisation, additional data were collected in the test tank using the same three acoustic frequencies, incorporating new data for all object classes, including the addition of a smaller PVC square. This process involved acquiring data at two additional distances and from three fields of view (FOV) relative to the sonar head, with targets arranged in different ways to create varied footprints in the sonar images. The ranges were selected to ensure an equal number of occurrences for each frequency, considering the minimum field of view (FOV) of the multibeam echosounder (MBES). Varying the range altered the appearance of targets across the water column, allowing assessment of whether the models could still detect and classify the targets accurately. Due to the test tank's physical constraits, such as its walls, bottom, and surface reflections, more distances could not be tested. While the expanded dataset reduced overfitting in most models, SVM models continued to struggle with generalisation and required significantly longer training times compared to other models, even if there were changes in the cross-frequency test results.

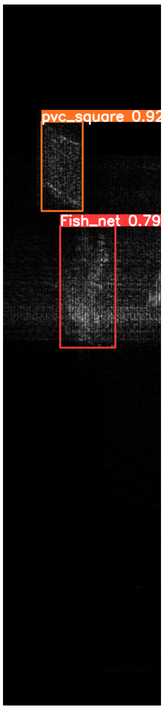

Figure 10. YOLO8 model inference in polar acoustic images with multiple targets with the target detection confidence.

CNN models demonstrated that the polar representation of the acoustic images outperformed the raw image representation, showing better generalisation. However, despite the mostly balanced test tank dataset, the lack of environmental and range variability still led to overfitting, even with simple data augmentation and early stopping techniques in place. These issues were primarily due to test tank limitations. Class activation maps revealed that the CNN models focused on pixels and features linked to the backscatter from marine debris targets. CNN models also trained faster than SVM models.

The SSD model produced good results with acoustic images at the same frequency as the training set, outperforming the YOLOv8 model in this respect. However, SSD models required more time and resources for training and were only effective for detecting objects at the same acoustic frequency they were trained on.

The YOLOv8 models, while not as effective as SSD models in same-frequency tests, performed better in cross-frequency tests. The training was faster and more optimised. YOLOv8 might yield better results with more exhaustive hyperparameter tuning.

Overall, the classification and detection of different types of marine debris were achieved, despite some selected marine debris sharing similar shapes. This suggests that features other than shape were key to distinguishing between the selected materials, allowing the material's characterisation with the features available within the dataset. Data were gathered at different acoustic frequencies to assess whether there was greater discriminative power for each type of object and material at specific frequencies and to verify if the objects could be reliably detected at those frequencies. Additionally, a cross-frequency study was conducted, confirming the expected outcome that models would underperform when trained and tested across different frequencies. This was a preliminary investigation into the potential of multi-frequency approaches for target detection.

Across the models, the cross-frequency tests generally yielded better results in those trained with 1400 kHz data. Higher acoustic frequencies tended to produce better metrics, such as accuracy, likely due to the improved spatial resolution, which helps distinguish between targets based on different types of acoustic data. This improvement is primarily linked to the reduced distance between beams at higher frequencies. However, these tests are preliminary and inconclusive, as they do not provide sufficient information to conduct a comprehensive multi-frequency study, which would require the use of different hardware. However, a multi-frequency MBES is not yet available for data acquisition, which could otherwise facilitate techniques similar to those used in hyperspectral imaging systems [12]. Another approach worth exploring involves using multiple single-frequency models to detect and classify the same target, potentially increasing its detectability. Although this method does not achieve the same results as a multi-frequency or adaptive-frequency multibeam echosounder, it presents a feasible line of investigation when the required hardware is unavailable. It is not possible to conclude if the models can generalise with the current test tank dataset. To achieve generalisation and enable the application of new algorithms, a more extensive dataset obtained using the MBES in real-world scenarios across all frequencies and various ranges and poses relative to placed targets is required.

6. Conclusions and Future Work

Addressing the growing threat of marine litter in the oceans is crucial. This study obtained acoustic images using a multibeam echosounder setup at various frequencies to facilitate classification and detection. Data were collected in a test tank scenario with multiple targets positioned in the water column, enabling their identification through acoustic imagery.

Supervised machine learning methods were employed for automated target classification. Two algorithms, CNN and SVM, were implemented, and a model was trained to categorise different object classes at different acoustic frequencies. Deep learning approaches based on transfer learning were also trained with new weights, specifically for SSD and YOLO8 detection models. These methods demonstrated the capability to classify and detect marine debris targets at different acoustic frequencies from water column imaging data, distinguishing different debris even if their shape was similar. However, while mostly balanced, the data lacked sufficient variety for the models to generalise across multiple scenarios. The test results revealed that models trained on data with the same acoustic frequency exhibited overfitting, revealing a dataset limitation. Cross-frequency inference yielded poorer results, particularly in models based on frequencies with better spatial resolution. It highlights their difficulty in detecting or classifying targets in images with lower spatial resolution, such as those based on 950 kHz frequency, being more notorious in the detection algorithms. Tests with different algorithms indicated that the polar representation of acoustic data led to better results than the raw image representation, as demonstrated in the CNN model study.

The successful characterisation and detection of the proposed targets in a controlled environment serve as essential groundwork for future real-world experiments that are already planned.

Future work will focus on acquiring more data in real-world scenarios and placing various marine debris targets at known locations to develop more generalised models. Plans include creating a dataset with multiple acoustic frequencies and varying MBES orientations relative to the targets at different ranges. A data acquisition campaign is being organised within the scope of the NetTag+ project to gather data from the same marine debris classes used in this study, along with additional debris types, to expand the dataset and improve the models' generalisation capabilities. This dataset will enable testing of current state-of-the-art networks and support the development of new techniques for detecting and classifying marine litter. The successful characterisation and detection proposed by this work laid the foundation for the planned dataset campaign.

Moreover, further experiments with multiple MBES systems will be necessary to advance marine litter detection and classification. However, this study has already demonstrated that acoustic data collected at different single frequencies using a MBES can differentiate between various objects in the water column in test tank conditions. Future tests will investigate whether detecting the same target at multiple frequencies, even if multi-band information is not available simultaneously, can enhance the discriminative power of marine debris detection algorithms.

Author Contributions: Conceptualisation, H.M.S.; methodology, H.M.S.; software, P.A.G.; validation, H.M.S. and S.W.; formal analysis, H.M.S. and S.W.; investigation, P.A.G.; data curation, P.A.G.; writing—original draft preparation, P.A.G.; writing—review and editing, H.M.S., S.W. and P.A.G.; supervision, H.M.S. and S.W.; project administration, H.M.S., A.M., J.A. and E.S.; funding acquisition, H.M.S., A.M., J.A. and E.S. All authors have read and agreed to the published version of the manuscript.

Funding: National Funds finance this work through the Portuguese funding agency, FCT—Fundação para a Ciência e a Tecnologia, within the project UIDB/50014/2020 with DOI (https://doi.org/10.54499/uidb/50014/2020), the Studentship for Doctoral Research Funding Programme, with the DOI (https://doi.org/10.54499/SFRH/BD/151460/2021) and by the European Union under the Horizon Europe Program, Grant No. 101112812 (NETTAG+).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset available at https://doi.org/10.5281/zenodo.13759505 [51].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fauziah, S.H.; Rizman-Idid, M.; Cheah, W.; Loh, K.H.; Sharma, S.; NoorMaiza, M.; Bordt, M.; Praphotjanaporn, T.; Samah, A.A.; bin Sabaruddin, J.S.; et al. Marine debris in Malaysia: A review on the pollution intensity and mitigating measures. *Mar. Pollut. Bull.* **2021**, *167*, 112258. [CrossRef]
2. Alizadeh, L.; Liscio, M.C.; Sospiro, P. The phenomenon of greenwashing in the fashion industry: A conceptual framework. *Sustain. Chem. Pharm.* **2024**, *37*, 101416. [CrossRef]
3. Li, M.; Trencher, G.; Asuka, J. The clean energy claims of BP, Chevron, ExxonMobil and Shell: A mismatch between discourse, actions and investments. *PLoS ONE* **2022**, *17*, e0263596. [CrossRef]
4. Yu, E.P.y.; Van Luu, B.; Chen, C.H. Greenwashing in environmental, social and governance disclosures. *Res. Int. Bus. Financ.* **2020**, *52*, 101192. [CrossRef]
5. Galgani, F.; Michela, A.; Gérigny, O.; Maes, T.; Tambutté, E.; Harris, P.T. Marine Litter, Plastic, and Microplastics on the Seafloor. *Plastics and the Ocean: Origin, Characterization, Fate, and Impacts*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2022; pp. 151–197.
6. Löhr, A.; Savelli, H.; Beunen, R.; Kalz, M.; Ragas, A.; Van Belleghem, F. Solutions for global marine litter pollution. *Curr. Opin. Environ. Sustain.* **2017**, *28*, 90–99. [CrossRef]
7. Sivadas, S.K.; Mishra, P.; Kaviarasan, T.; Sambandam, M.; Dhineka, K.; Murthy, M.R.; Nayak, S.; Sivyer, D.; Hoehn, D. Litter and plastic monitoring in the Indian marine environment: A review of current research, policies, waste management, and a roadmap for multidisciplinary action. *Mar. Pollut. Bull.* **2022**, *176*, 113424. [CrossRef]
8. Egger, M.; Quiros, L.; Leone, G.; Ferrari, F.; Boerger, C.M.; Tishler, M. Relative abundance of floating plastic debris and neuston in the eastern North Pacific Ocean. *Front. Mar. Sci.* **2021**, *8*, 626026. [CrossRef]
9. Corcoran, P.L. Benthic plastic debris in marine and fresh water environments. *Environ. Sci. Process. Impacts* **2015**, *17*, 1363–1369. [CrossRef]
10. Soto-Navarro, J.; Jordá, G.; Compa, M.; Alomar, C.; Fossi, M.; Deudero, S. Impact of the marine litter pollution on the Mediterranean biodiversity: A risk assessment study with focus on the marine protected areas. *Mar. Pollut. Bull.* **2021**, *165*, 112169. [CrossRef]
11. Topouzelis, K.; Papageorgiou, D.; Karagaitanakis, A.; Papakonstantinou, A.; Arias Ballesteros, M. Remote sensing of sea surface artificial floating plastic targets with Sentinel-2 and unmanned aerial systems (plastic litter project 2019). *Remote Sens.* **2020**, *12*, 2013. [CrossRef]
12. Freitas, S.; Silva, H.; Silva, E. Hyperspectral Imaging Zero-Shot Learning for Remote Marine Litter Detection and Classification. *Remote Sens.* **2022**, *14*, 5516. [CrossRef]
13. Ribotti, A.; Magni, P.; Mireno, B.; Schroeder, K.; Barton, J.; McCaul, M.; Diamond, D. New cost-effective, interoperable sensors tested on existing ocean observing platforms in application of European directives: The COMMON SENSE European project. In Proceedings of the OCEANS 2015-Genova, Genova, Italy, 18–21 May 2015; pp. 1–9.
14. Politikos, D.V.; Adamopoulou, A.; Petasis, G.; Galgani, F. Using artificial intelligence to support marine macrolitter research: A content analysis and an online database. *Ocean. Coast. Manag.* **2023**, *233*, 106466. [CrossRef]
15. Guedes, P.; Silva, H.; Wang, S.; Martins, A.; Almeida, J.; Silva, E. Multibeam Multi-Frequency Characterization of Water Column Litter. In Proceedings of the OCEANS 2024-Singapore, Singapore, 15–18 April 2024; pp. 1–6.

16. Garaba, S.P.; Park, Y.J. Riverine litter monitoring from multispectral fine pixel satellite images. *Environ. Adv.* **2024**, *15*, 100451. [CrossRef]
17. Broere, S.; van Emmerik, T.; González-Fernández, D.; Luxemburg, W.; de Schipper, M.; Cózar, A.; van de Giesen, N. Towards underwater macroplastic monitoring using echo sounding. *Front. Earth Sci.* **2021**, *9*, 628704. [CrossRef]
18. Politikos, D.V.; Fakiris, E.; Davvetas, A.; Klampanos, I.A.; Papatheodorou, G. Automatic detection of seafloor marine litter using towed camera images and deep learning. *Mar. Pollut. Bull.* **2021**, *164*, 111974. [CrossRef]
19. Aleem, A.; Tehsin, S.; Kausar, S.; Jameel, A. Target Classification of Marine Debris Using Deep Learning. *Intell. Autom. Soft Comput.* **2022**, *32*, 73–85. [CrossRef]
20. Bajaj, R.; Garg, S.; Kulkarni, N.; Raut, R. Sea debris detection using deep learning: Diving deep into the sea. In Proceedings of the 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), Kuala Lumpur, Malaysia, 24–26 September 2021; pp. 1–6.
21. Deng, H.; Ergu, D.; Liu, F.; Ma, B.; Cai, Y. An embeddable algorithm for automatic garbage detection based on complex marine environment. *Sensors* **2021**, *21*, 6391. [CrossRef]
22. Fossum, T.O.; Sture, Ø.; Norgren-Aamot, P.; Hansen, I.M.; Kvisvik, B.C.; Knag, A.C. Underwater autonomous mapping and characterization of marine debris in urban water bodies. *arXiv* **2022**, arXiv:2208.00802.
23. Fulton, M.; Hong, J.; Islam, M.J.; Sattar, J. Robotic detection of marine litter using deep visual detection models. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 5752–5758.
24. Hong, J.; Fulton, M.; Sattar, J. Trashcan: A semantically-segmented dataset towards visual detection of marine debris. *arXiv* **2020**, arXiv:2007.08097.
25. Valdenegro-Toro, M. Submerged marine debris detection with autonomous underwater vehicles. In Proceedings of the 2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA), Amritapuri, India, 18–20 December 2016; pp. 1–7.
26. Van Emmerik, T.; Kieu-Le, T.C.; Loozen, M.; Van Oeveren, K.; Strady, E.; Bui, X.T.; Egger, M.; Gasperi, J.; Lebreton, L.; Nguyen, P.D.; et al. A methodology to characterize riverine macroplastic emission into the ocean. *Front. Mar. Sci.* **2018**, *5*, 372. [CrossRef]
27. Sheeny, M.; Wallace, A.; Wang, S. 300 GHz radar object recognition based on deep neural networks and transfer learning. *IET Radar, Sonar Navig.* **2020**, *14*, 1483–1493. [CrossRef]
28. Ochal, M.; Vazquez, J.; Petillot, Y.; Wang, S. A comparison of few-shot learning methods for underwater optical and sonar image classification. In Proceedings of the Global Oceans 2020: Singapore–US Gulf Coast, Biloxi, MS, USA, 5–30 October 2020; pp. 1–10.
29. Kim, B.; Yu, S.C. Imaging sonar based real-time underwater object detection utilizing AdaBoost method. In Proceedings of the 2017 IEEE Underwater Technology (UT), Busan, Republic of Korea, 21–24 February 2017; pp. 1–5.
30. Zhao, J.; Mai, D.; Zhang, H.; Wang, S. Automatic Detection and Segmentation on Gas Plumes from Multibeam Water Column Images. *Remote Sens.* **2020**, *12*, 3085. [CrossRef]
31. Zhao, S.; Wang, C.; Bai, B.; Jin, H.; Wei, W. Study on the polystyrene plastic degradation in supercritical water/CO_2 mixed environment and carbon fixation of polystyrene plastic in CO_2 environment. *J. Hazard. Mater.* **2022**, *421*, 126763. [CrossRef]
32. Wang, X.; Wang, J.; Yang, F.; Zeng, G. Target detection in colorful imaging sonar based on HOG. In Proceedings of the 2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Qingdao, China, 14–16 September 2018; pp. 1–5.
33. Ji, X.; Yang, B.; Tang, Q. Acoustic seabed classification based on multibeam echosounder backscatter data using the PSO-BP-AdaBoost algorithm: A case study from jiaozhou bay, China. *IEEE J. Ocean. Eng.* **2020**, *46*, 509–519. [CrossRef]
34. Valdenegro-Toro, M. Learning objectness from sonar images for class-independent object detection. In Proceedings of the 2019 European Conference on Mobile Robots (ECMR), Prague, Czech Republic, 4–6 September 2019; pp. 1–6.
35. Singh, D.; Valdenegro-Toro, M. The marine debris dataset for forward-looking sonar semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 3741–3749.
36. Wang, J.; Feng, C.; Wang, L.; Li, G.; He, B. Detection of Weak and Small Targets in Forward-Looking Sonar Image Using Multi-Branch Shuttle Neural Network. *IEEE Sens. J.* **2022**, *22*, 6772–6783. [CrossRef]
37. Yu, Y.; Zhao, J.; Gong, Q.; Huang, C.; Zheng, G.; Ma, J. Real-time underwater maritime object detection in side-scan sonar images based on transformer-YOLOv5. *Remote Sens.* **2021**, *13*, 3555. [CrossRef]
38. Huo, G.; Wu, Z.; Li, J. Underwater object classification in sidescan sonar images using deep transfer learning and semisynthetic training data. *IEEE Access* **2020**, *8*, 47407–47418. [CrossRef]
39. Fuchs, L.R.; Gällström, A.; Folkesson, J. Object recognition in forward looking sonar images using transfer learning. In Proceedings of the 2018 IEEE/OES Autonomous Underwater Vehicle Workshop (AUV), Porto, Portugal, 6–9 November 2018; pp. 1–6.
40. Ge, Q.; Ruan, F.; Qiao, B.; Zhang, Q.; Zuo, X.; Dang, L. Side-scan sonar image classification based on style transfer and pre-trained convolutional neural networks. *Electronics* **2021**, *10*, 1823. [CrossRef]
41. Gaida, T. Acoustic Mapping and Monitoring of the Seabed: From Single-Frequency to Multispectral Multibeam Backscatter. Ph.D. Thesis, TU Delft, Delft, The Netherlands, 2020.
42. Janowski, L.; Trzcinska, K.; Tegowski, J.; Kruss, A.; Rucinska-Zjadacz, M.; Pocwiardowski, P. Nearshore benthic habitat mapping based on multi-frequency, multibeam echosounder data using a combined object-based approach: A case study from the Rowy Site in the Southern Baltic Sea. *Remote Sens.* **2018**, *10*, 1983. [CrossRef]

43. Gonçalves, P.M.; Ferreira, B.M.; Alves, J.C.; Cruz, N.A. Image segmentation and mapping in an underwater environment using an imaging sonar. In Proceedings of the OCEANS 2022, Hampton Roads, VA, USA, 17–20 October 2022; pp. 1–8.
44. Chandrashekar, G.; Raaza, A.; Rajendran, V.; Ravikumar, D. Side scan sonar image augmentation for sediment classification using deep learning based transfer learning approach. *Mater. Today Proc.* **2023**, *80*, 3263–3273. [CrossRef]
45. Qin, X.; Luo, X.; Wu, Z.; Shang, J. Optimizing the Sediment Classification of Small Side-Scan Sonar Images Based on Deep Learning. *IEEE Access* **2021**, *9*, 29416–29428. [CrossRef]
46. Zhu, J.; Li, H.; Qing, P.; Hou, J.; Peng, Y. Side-Scan Sonar Image Augmentation Method Based on CC-WGAN. *Appl. Sci.* **2024**, *14*, 8031. [CrossRef]
47. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
48. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 6999–7019. [CrossRef]
49. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
50. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
51. Guedes, P. MBES M3 Kongsberg HF Model Test Tank Marine Debris Dataset. 2024. Available online: https://zenodo.org/records/13759505 (accessed on 13 September 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.