

Review

A Review of Underwater Mine Detection and Classification in Sonar Imagery

Stanisław Hożyń 

Faculty of Mechanical and Electrical Engineering, Polish Naval Academy, 81-127 Gdynia, Poland; s.hozyn@amw.gdynia.pl; Tel.: +48-261-262-586

Abstract: Underwater mines pose extreme danger for ships and submarines. Therefore, navies around the world use mine countermeasure (MCM) units to protect against them. One of the measures used by MCM units is mine hunting, which requires searching for all the mines in a suspicious area. It is generally divided into four stages: detection, classification, identification and disposal. The detection and classification steps are usually performed using a sonar mounted on a ship's hull or on an underwater vehicle. After retrieving the sonar data, military personnel scan the seabed images to detect targets and classify them as mine-like objects (MLOs) or benign objects. To reduce the technical operator's workload and decrease post-mission analysis time, computer-aided detection (CAD), computer-aided classification (CAC) and automated target recognition (ATR) algorithms have been introduced. This paper reviews mine detection and classification techniques used in the aforementioned systems. The author considered current and previous generation methods starting with classical image processing, and then machine learning followed by deep learning. This review can facilitate future research to introduce improved mine detection and classification algorithms.

Keywords: mine detection; mine classification; sonar imagery; mine countermeasure; mine-like object



Citation: Hożyń, S. A Review of Underwater Mine Detection and Classification in Sonar Imagery. *Electronics* **2021**, *10*, 2943. <https://doi.org/10.3390/electronics10232943>

Academic Editor: Manohar Das

Received: 7 November 2021

Accepted: 25 November 2021

Published: 26 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Underwater mines are a strategic military tool to protect any country's naval borders. They constitute fully autonomous devices composed of an explosive charge, sensing device and fuse mechanism. Previous generation mines needed physical contact with the ship to trigger an explosion. The newly developed mines, on the contrary, are equipped with sophisticated sensors, usually detecting some combinations of acoustic and magnetic signals. Some of them are smart mines equipped with artificial intelligence to detect any false signals that attempt to release them. These mines need to pose extreme danger for ships and submarines. However, their small operational range makes their effectiveness minimal. To maximise effectiveness, a group of mines—a minefield—is deposited in a specific seabed area. In this formation, they pose a tactical threat for all types of ships.

Navies around the world use mine countermeasure (MCM) units to protect against mines that involve both passive and active tactics. Passive countermeasures modify the specific target vessel characteristics or signatures which are used to trigger mines. These include building vessels with fibreglass or altering a steel vessel's magnetic field through degaussing. Alternatively, active countermeasures aim to find mines using specially designed ships or underwater vehicles. The first active measure, minesweeping, utilises a contact sweep or wire drag to cut the mooring wires of floating mines. In other scenarios, it uses a distance sweep to mimic a ship. The second active measure, mine hunting, requires searching for all the mines in a suspicious area. It is generally divided into four stages:

- Detection—finding targets from different signals (acoustic, magnetic);
- Classification—determining if the target is a mine-like object (MLO) or a benign object;
- Identification—using additional information (diver, underwater vehicle equipped with a camera) to validate classification results;

- Disposal—neutralising a mine.

The detection and classification steps are usually performed using a sonar mounted on a ship's hull or on an underwater vehicle. After retrieving the sonar data, military personnel scan the seabed images to detect targets and classify them as MLOs or benign objects. This procedure is tedious and time consuming due to the difficult task of differentiating a multitude of received sonar images. Since a human factor is involved in scrutinising these images, fatigue and stress could lead to misclassification errors.

As mine technology becomes more advanced, MCM, as a result, has become more complex and sophisticated. Currently, navies employ a range of different mine countermeasures that include the human factor. Navies' primary target is to have less human lives involved, such as divers, in risky underwater operations of minefield detection. For this purpose, MCM units use either autonomous underwater vehicles (AUVs), remotely operated vehicles (ROVs), or unmanned surface vehicles (USVs). The vehicles often perform their missions autonomously, equipped with high-resolution sensors (sonars, magnetometers, optical cameras). After the mission, the data gathered by the vehicles are analysed by a sonar technical operator aboard the ship to detect, classify and identify objects in a verified area.

To reduce the technical operator's workload and decrease post-mission analysis time, computer-aided detection (CAD), computer-aided classification (CAC) and automated target recognition (ATR) algorithms have been introduced. They are based on the analysis of different types of image characteristics, which can be grouped into three categories [1]:

- Texture-based features—patterns and local variations of the image intensity;
- Geometrical features—e.g., length, area;
- Spectral features—e.g., colour, energy.

Two attempts for utilising image features in the aforementioned algorithms are depicted in Figure 1.

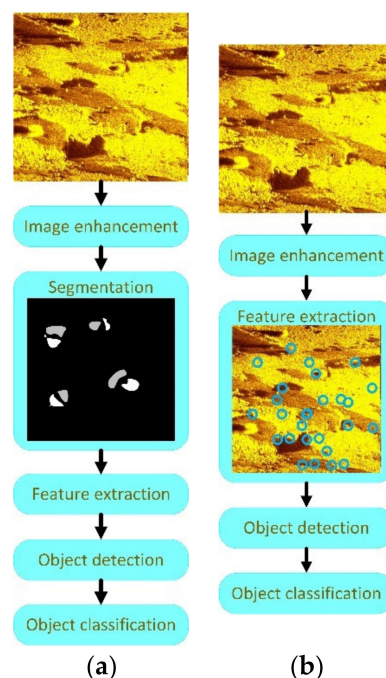


Figure 1. Mine detection and classification (a) based on segmentation, and (b) on texture feature extraction.

The first attempt employs an image segmentation method, which uses geometrical and spectral features to divide the image into homogeneous regions: seabed, highlights and shadows. In the first step, the image is enhanced to facilitate the retrieval of distinctive

geometrical features. Then, the image is segmented; all areas that differ from the seabed are marked as regions of interest (ROIs) and analysed, considering the geometrical features of the included regions. Finally, if the geometrical features of the ROI correspond to pre-defined mine features, the object is classified as an MLO. In the second attempt, ROI detection and classification are performed using texture-based features. In this case, the parameters of the features and their mutual localisation are analysed to detect ROIs and classify them as MLOs or benign objects. It also demands image enhancement to recognise the most distinctive features. Finally, the features are compared with pre-defined mine features in the classification step.

Even though extensive research has been devoted to developing CAD, CAC and ATR algorithms, they are still not efficient enough to perform their task autonomously. Their responsibilities mainly focus on notifying a technical operator about a potential target in the form of visual cues. Consequently, the operator is responsible for making the final decision regarding classification. This situation stems from the low quality of sonar images. Sonar images are complicated to analyse due to speckle noise and environmental conditions causing spurious shadows, sidelobe effects and multipath return [2]. This results in significant variability in targets, clutter and background signatures. Additionally, mines constitute small objects, often partially buried in the seabed, which makes them difficult to distinguish, even for a highly skilled technical operator. Therefore, the task of developing efficient detection and classification methods is very challenging.

The detection and classification methods presented in the literature can be divided into classical image processing, machine learning (ML) and deep learning (DP) techniques [3]. Classical image processing demands expert supervision to select the required features [4]. The features are mostly based on background, highlight and shadow combinations (Figure 2). Considering the depiction of MLOs in images, the object can be recognised as the following background (A–B), highlight (B–C), background (C–D) and shadow (D–E) regions. For this purpose, the techniques employing geometrical or spectral feature analyses are applied. Classical signal processing needs a lot of time and resources, as well as being expensive to implement. Therefore, researchers have attempted to replace it with intelligent techniques such as machine learning or deep learning [3].

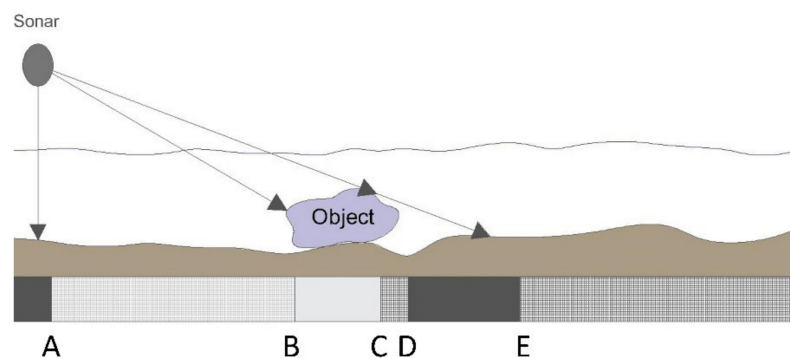


Figure 2. Background, shadow and highlight combination.

Machine learning is when a computer uses data to detect or predict hidden characteristics or patterns automatically. The detection or prediction needs to be preceded by a supervised or unsupervised learning process. In the supervised option, the input and output data pair is used to teach the system. In unsupervised learning, the system is taught utilising input data only. In both cases, the imaging data should be high quality, which is difficult to achieve using sonar images. ML also has other weaknesses. It needs to divide the problem statement into several parts and then combine the result, which is a time-consuming process. Additionally, it considers only the object's features and does not include all background information. Furthermore, the substantial amount of data reduces its reliability [3]. These problems can be solved using deep learning.

Deep learning is a subfield of machine learning, providing algorithms inspired by the structure of the brain's neuron connections called artificial neural networks (ANNs). It constitutes computational models composed of multiple layers to process data with different levels of abstraction. Deep learning works with unstructured and structured data and performs automatic feature extraction. Its algorithms work excellently with a vast amount of data, performing efficiently and overcoming machine learning's weaknesses described above, meaning it is more reliable [3]. Figure 3 presents the working process of machine learning and deep learning algorithms and the performance in relation to the quantity of data provided.

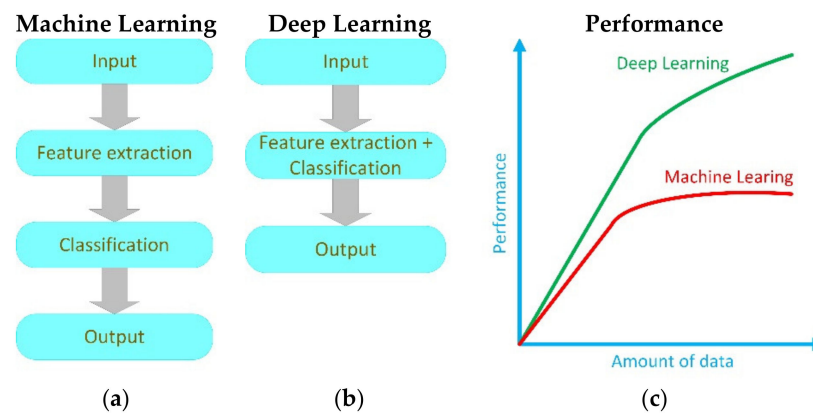


Figure 3. (a) Working process of ML; (b) working process of DL; (c) performance of ML and DP as a function of the amount of data.

Deep learning algorithms are classified into supervised, semi-supervised, unsupervised and reinforcement learning. A supervised learning approach trains a model with the categorised data or labels to predict the mapping function, while unsupervised learning identifies patterns in datasets containing data points that are neither classified nor labelled. As for semi-supervised learning, it combines a small amount of labelled data with a large amount of unlabelled data. Finally, reinforcement learning learns in an interactive environment by trial and error using feedback from its own actions and experiences.

Deep learning algorithms demand a vast quantity of data. However, due to the lack of publicly available datasets and confidentiality issues connected with the military character of mine detection tasks, the problem of having high-quality data necessary for training neural networks still exists. To deal with this problem, some techniques to collect the sonar data are deployed. One of them is sonar data simulation, which plays an essential role in tuning, detection and classification methods. Various simulators have been presented in the literature. Cerqueira et al. [5] developed a GPU-based sonar simulator for real-time applications. A different approach was presented in [6,7], where the tube ray-tracing method was utilised to generate realistic sonar images. The research in [8] proposed a simulation method called uSimActiveSonar to simulate an active sonar system with hydrophone data acquisition. In [9], the authors used a definite difference time-domain approach for pulse propagation simulation.

Data augmentation (DA) is also applied to create data for deep learning algorithms. It is a technique that artificially expands a training set's size by creating modified data from the existing data. Its main task is to prevent overfitting when an initial dataset is too small to train on. DA comprises image processing methods such as flipping, rotation, scaling, translation, cropping or generative adversarial networks. Other techniques used by DA constitute colour space transformation, kernel filtering, random erasing or mixing images.

Transfer learning is also used to deal with insufficient training data. It is based on gaining knowledge from solving one problem and applying it to a different but related task. For instance, a network trained to detect wrecks on the seabed transfers this training

to a network dedicated to mine detection. In this case, a smaller database can be applied in the training procedure.

Another concept to improve the reliability of mine detection and classification methods is algorithm fusion [10,11]. Algorithm fusion combines the detection and classification outputs from two or more algorithms. This is beneficial because the performance of the individual method depends on many factors, such as the image quality, environmental conditions or object's characteristics. By fusing two or more algorithms, the probability of mine detection and proper classification increases.

Another fusion technique is to combine classical image processing with deep learning [3]. A classical method distinguishes ROIs in the detection step while deep learning classifies ROIs as MLOs or benign objects in this technique. This combination can improve the performance of the classification step since the neural network analyses only the ROIs' areas in the image. It can also reduce the negative impact of unbalanced data, which appears while using sonar images for mine detection. Unbalanced data result from unequal representations of mine and seabed areas in the image. Mines are depicted as small objects on any seabed area. Therefore, their presence in the given data is hardly noticeable, while the seabed areas have a significant articulation.

A detailed overview of underwater mine detection and classification in sonar imagery is presented in this paper. In Section 2, a sonar device is described, taking into consideration the ability to generate underwater images. Section 3 briefly introduces the concept of highlights and shadows in sonar imagery. This concept is crucial since most classical image processing methods utilise highlights and shadows to detect and classify objects. Section 4 presents target detection methods, while Section 5 describes object classification techniques. In the review, the author focused on methods that were implemented in mine detection and classification algorithms. This is because mines constitute complicated objects to detect and classify due to their size and characteristics. Consequently, object detection algorithms that detect various seabed objects with sonar imagery are the least reliable for mine detection.

This review includes the most recent methods as well as previous generation ones. The latter, even though they very often achieved low efficiency, should still be considered in newly developed algorithms. This is because their past training was based on low-quality and low-resolution images, but they should perform better given the latest high-resolution images. Consequently, their implementation in up-to-date applications is a positive aspect. This is especially important because of the actual joining of classical image processing with deep learning algorithms.

2. Underwater Sonar

Sonar is an acronym for SOund Navigation And Ranging. Sonar's basic principle is to locate objects via transmitting acoustic waves. Acoustic waves constitute mechanical vibrations produced by longitudinal waves through an elastic medium at a certain speed. Their speed is measured, and a target distance can be estimated based on the time between echo transmission and the return. However, when in water, acoustics waves spread faster than when in air. The sound velocity connected to this speed differs when in water compared to when it is in air. The factors that contribute to this speeding up process are depth, temperature, pH and salinity. It is estimated that the speed of sound in underwater conditions varies from 1405 to 1550 m/s (in the air, it is approximately 340 m/s) [12].

Acoustic signals produce vibrations. They are characterised by frequency f (expressed in Hz) or by period T (linked to the frequency by the relation $T = 1/f$). In airborne conditions, the frequency used in communication can reach 300 GHz. The frequencies utilised in underwater applications vary between 10 Hz and 1 MHz [13]. The main constraints on the frequencies used for a particular application are as follows [12]:

- The sound wave attenuation in water, limiting the maximum usable range;
- The dimensions of the sound sources, which increase at lower frequencies, for a given transmission power;
- The spatial selectivity related to the directivity of acoustic sources and receivers;

- The target acoustic response, depending on frequency.

The wavelength of the acoustic signal is the spatial correspondent to time periodicity. It is described as the spacing between two points in the medium, undergoing the same successive vibration state. In other words, this is the distance travelled by the wave during one period of the signal T with velocity c :

$$\lambda = cT = \frac{c}{f} \quad (1)$$

This means that when a sound velocity is equal to 1500 m/s, the underwater acoustic wavelength amounts to 150 m at 10 Hz, then to 1.5 m at 10 kHz and finally to 0.0015 m at 1 MHz.

The sonar equation depicts the relationship amongst all elements of the sound travelling between a transducer and a target. It defines how much energy is demanded to return to a transducer from a target in order to be detected [13]:

$$DT \leq SL - 2 \times TL + TS - (NL - DI) \quad (2)$$

where DT is the detection threshold, which defines the sensitivity of the sonar to detect acoustic energy; SL is the source level, which defines the initial acoustic energy; TL is transmission loss caused by distortion and attenuation of water; TS is the target strength determined by the reflectivity of a target and defined by its shape, size and composition; NL is the noise level in the water column which may disturb the sonar's sound waves; DI is the directivity index, which indicates the narrowness of the sound beam that the sonar produces.

The main effect of wave propagation is the decreasing signal amplitude caused by absorption and geometrical spreading. Absorption is connected to the chemical properties of seawater and is a crucial factor in the propagation of underwater acoustic waves, which limits the reachable range at high frequencies. The approximation of propagation losses is crucial in the evaluation of sonar system performances. Since the aquatic propagation medium is restricted by two well-marked interfaces (the seabed and the sea surface), the propagation of a signal is often accompanied by a series of multiple paths generated by unwanted reflections at these two interfaces. Furthermore, the velocity of acoustic waves varies spatially in the depths of the ocean. Due to temperature and pressure, the paths of sound waves are thus refracted depending on the velocity variations encountered. This phenomenon complicates the modelling and interpretation of the sound field's spatial structure. Consequently, the most manageable and most efficient modelling technique is geometrical acoustics, based on the local values of the wave propagation direction and velocity according to the well-known Snell–Descartes law [13].

The acoustic wave is characterised by three factors [14]:

- The local motion amplitude of each particle in the propagation medium around its position of equilibrium;
- The fluid velocity corresponding to this motion;
- The acoustic pressure, which is the variation around the average hydrostatic pressure.

Practically, acoustic pressure is the most often used quantity in underwater acoustics, and it is measured using hydrophones, the marine equivalent of aerial microphones.

The range is defined as the radial distance between the sonar and the reflector. It can be estimated according to the following procedure:

- Transmitting a short pulse of duration T_p in the direction of the reflector;
- Recording the signal by the receiver until the echo from the reflector has arrived;
- Estimating the time delay τ from time series.

The range to the target is then calculated as

$$R = \frac{c\tau}{2} \quad (3)$$

The accuracy of the range is related to the pulse length T_p . Shorter pulses offer a better resolution but have less energy, which limits the propagation range. Another approach is based on modulated (phase-coded) pulses. In this case, the resolution is reduced by the bandwidth of the acoustic signal.

Underwater sonar is composed of hydrophones that transmit and receive acoustic waves, in this case, known as active sonar. Alternatively, it only receives echoes from underwater; the term passive sonar is used to describe this operation. When a hydrophone transmits an acoustic wave, it propagates and is reflected by any interface with a different characteristic impedance $Z = \rho c$, where ρ is the density of the interface (in Rayl), and c is the speed of sound in water.

A transmitting hydrophone emits acoustic waves to a scene containing acoustic reflecting material characterised by a reflectivity function $\gamma(x, y)$. The scattered acoustic field is received by one or more hydrophones (Figure 4). Sonar imaging poses the inverse problem of estimating the reflectivity function from the received data. It can be divided into angular and range processing, which is called data beamforming. Beamforming is a processing method that focuses the signal from several receivers in a specific direction. It can be employed in all types of sonars. Range processing includes signal processing that is applied to separate time events. The processing is a function of transmitting waveforms. Various types of signals are used to construct waveforms. Basic sonar often uses gated continuous-wave pulses, sometimes referred to as pings. The latest generation of sonar uses phase-coded transmit signals where the phase coding establishes the signal bandwidth. This is because phase-coded waveforms increase the transmit signal energy and maintain a large signal bandwidth. The range processing conducted on phase-coded signals is called pulse compression or matched filtering [4].

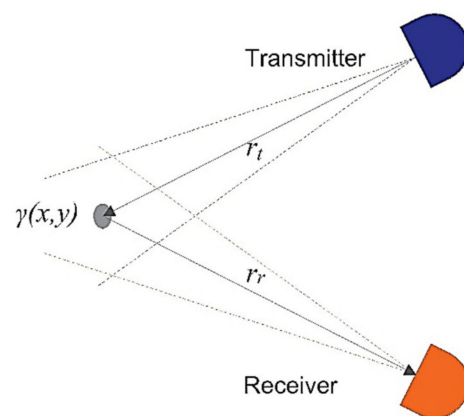


Figure 4. Basic imaging geometry.

The angular resolution defines the quality of sonar imaging. It is defined as the minimum angle at which two reflectors can be separated in the sonar image. Assuming that a phased array receiver of length L consists of some elements, as illustrated in Figure 5, the angular resolution is the angle difference at which the echo from two reflectors causes destructive interference in the receivers. For small angles, the angular resolution can be approximated as follows:

$$\beta = \frac{\lambda}{L} \quad (4)$$

where λ is the wavelength, and L is the receiver length.

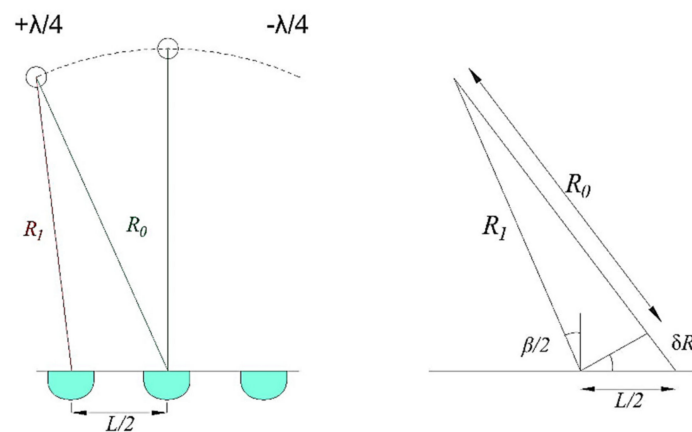


Figure 5. Angular resolution (L —receiver length; $\beta/2$ —the angle of the reflector; δR —the distance between two reflectors, R_0 and R_1).

For high-frequency sonar imaging, beamforming applications can be arranged into three different types [15]:

- Sectorscan sonar (Figure 6), which generates a two-dimensional image for each pulse. These images are usually displayed pulse by pulse. Sectorscan sonars are often mounted to a ship's hull for forward-looking or broad-swath imaging. In some applications, 360-degree views are produced by the arrangement of hydrophones into a cylindrical array.
- Sidescan sonar (Figure 7), which utilises the vehicle motion to cover the seabed area. It generates one or a few beams to create an image by using repeated pulses. This method demands a relatively simple hardware architecture; consequently, it is more affordable than other techniques.
- Synthetic-aperture sonar (Figure 8), which uses multiple pulses to create a sizeable synthetic array. An image of the seafloor is created from multiple pulses that form each pixel on the seafloor. This technique can be considered as the combination of sidescan and sectorscan sonars.

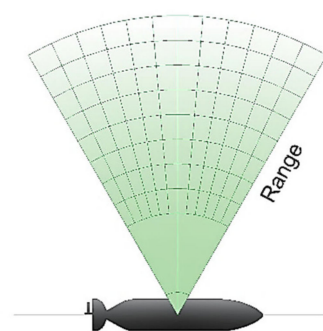


Figure 6. Sectorscan sonar.

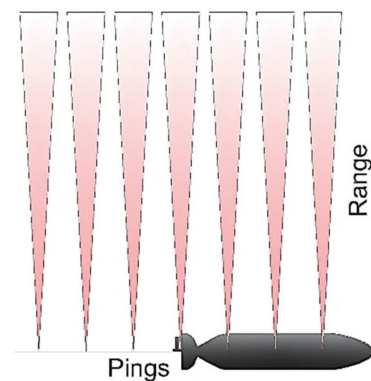


Figure 7. Sidescan sonar.

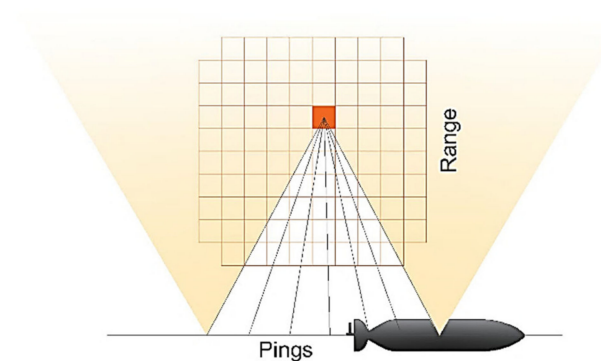


Figure 8. Synthetic-aperture sonar.

3. Highlights and Shadows

In sonar images, a highlight region is laterally followed by a shadow region for most targets of interest. The highlight's brightness and the shadow's size are determined by the target's strength (TS) of the sonar equation. The target's strength is related to the target's reflectivity, and it is built on a complex relationship between the target's dimension, shape, roughness and thickness. Additionally, the sonar's pulse duration, frequency and the incident angle of the waveform also influence the reflectivity [16].

The target structure is a direct factor; acoustically soft materials are less reflective than acoustically dense ones. The acoustic resistance relies on the speed of sound into the object in comparison to the aquatic elastic medium surrounding it. Additionally, the target's thickness is essential because the variations in the speed of sound continue for several wavelengths. The target's size and the acoustic beam's incident angle should guarantee that the entire beam reaches the surface. When an object is smaller than the dimensions of the beam width, or the incident angle does not hit the target surface entirely, the produced reflection is insufficient. The shape is also important because it decides the direction of the reflection. For example, a conic surface will return the waveforms in ripple formations in many directions [16].

Unlike reflectivity, the acoustic shadow does not depend on the object's structure and surface. The shadow is formed because the waveforms, when hitting the object, are obstructed by the side of the object and do not reach the seabed behind it, which is the required area. Consequently, no data are returned when the waveforms do not reach the specified area. This case is depicted in Figure 9, where the cliff blocks the waveforms from hitting the seabed.

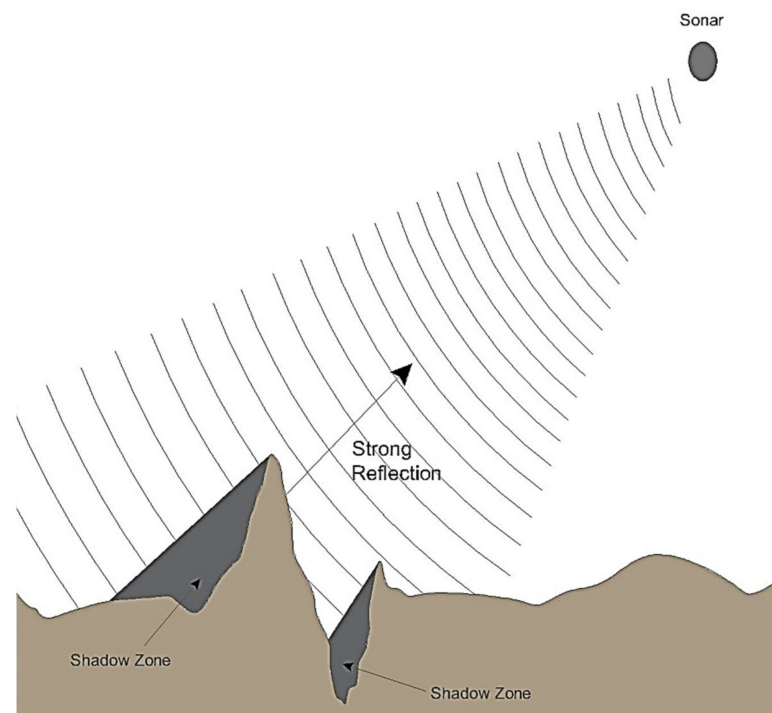


Figure 9. Acoustic reflection.

The shadow provides data about the vertical shape and characteristics of an object lying at the bottom. The lateral length of the shadow depends on the distance to the object, the vehicle's altitude and the height of the target. Acoustic shadows can also represent bottom depressions. This happens when the depression's depth is low enough to affect the seabed, causing it to prevent acoustic waveforms from reaching the depression. In this case, the vehicle's altitude and the size of the depression determine the shape of the shadow. However, this type of shadow is not preceded by a reflective target, meaning a highlight before the shadow is not created. This is why the combination of highlights and shadows is essential when searching for targets of interest.

4. Object Detection

Object detection in sonar imagery is a complicated task due to the nature of the environment, causing spurious shadows, multipath returns and sidelobe effects [2]. These obstacles cause significant variability in targets, resulting in clutter and various background signatures. However, the properties of underwater sound propagation assure that the mine area can be generally detected in the image. This is because a mine is usually manufactured using denser material than seabed objects and creates a highlight segment on the image. Additionally, a shadow segment, which is orthogonal to the sonar, is generated. The highlight and shadow segments create a spotted cluttered surrounding. As a result, the ROI has a background segment relating to the reverberation from the sea [17].

4.1. Image Enhancement

The first step of object detection and classification is constituted by the image enhancement technique [18–21]. It is mainly used to normalise images and reduce noise. In an image normalisation step, both highlights and shadows become distinct from the background. One method of image normalisation is histogram equalisation, which transforms an image, making its histogram nearly regular. All intensity bands have a similar number of pixels in a regular histogram [22]. Another method utilises a serpentine forward-backwards filter, which considers the neighbourhood of each pixel to normalise its value [18]. Both methods produce a consistent background and emphasise highlight and shadow pixel

intensities. Apart from normalisation, noise reduction is essential during the image enhancement step. For this purpose, a diamond- or square-shaped median filter is employed. This operation smooths the image and reduces noise [23].

Similarly, the Wiener filter analyses local mean and variance in order to reduce the changes in pixels' intensity [24]. The changes are also reduced by wavelet-based de-noising via Donoho's shrinkage, which implies choosing and shrinking a specific level in the wavelet [25]. Apart from the aforementioned methods, there are the Gaussian filter and the difference of Gaussians technique. Both are applied to reduce noise and smooth the image [26]. The image enhancement step is devoted to preparing the image for object detection and classification. Therefore, the choice of a particular technique should conform to the detection and classification schemes.

4.2. Image Segmentation

Image segmentation refers to the technique of grouping image pixels into several classes. The pixels belonging to the same homogeneous regions are assigned the same labels. Segmentation is a popular technique to separate the highlight and shadow regions correlated with mines in sonar imagery [27–34]. This is because the pixels representing mines have higher values than the average pixel intensity in the image, while the pixels representing shadows created by mines have lower values. The most straightforward approach to segment the image is determining thresholds in pixel intensities to distinguish background, highlight and shadow regions. Some improvements can be achieved by applying adaptive thresholding: for example, the threshold based on the local mean [18] or based on local histograms [35].

Acosta et al. [31] developed an algorithm to segment objects into two regions: acoustical highlight and seafloor reverberation areas. The proposed solution does not need any a priori assumption about the nature of sonar images. It constitutes an adaptation of the Cell Average–Constant False Alarm Rate (CA–CFAR) used in radar technology for detecting moving objects. The proposed algorithm provides similar results in image segmentation concerning other frequently used approaches but demands less computational resources and parameters to set. Due to its simplicity and accuracy, it can be used in real-time applications.

More complex techniques such as fuzzy functions [27] or Markov random field (MRF) [19] are also used for segmentation. The first technique, fuzzy functions, constitutes fuzzy k-means clustering, which utilises the mean and variance of luminance within small windows. This technique also can deploy c-means clustering, which additionally uses lightness. However, fuzzy clustering is responsive to speckle noise [24]. The other technique, MRF, enables reliable segmentation since it uses pixel dependencies. It considers the surrounding pixels' labels, assuming that a pixel surrounded by shadow pixels is more likely to belong to the shadow region [19].

In [36], a new unsupervised method of sonar image segmentation was introduced. This method is based on the amplitude dominant component analysis (ADCA) technique and comprises a multi-channel filtering and saliency map. The estimated saliency map is associated with the input from the sonar image, while multi-channel filtering uses the Gabor filter to reconstruct the input image in the narrowband components. The presented results indicate the usefulness of exploiting the saliency regions in the sonar images.

The authors in [37] established the efficient convolutional network (ECNet) architecture for semantic segmentation of SSS images, which utilises a novel encoder to learn rich hierarchical features. The architecture consists of an encoder network to capture context and a corresponding decoder network to restore full input-size resolution feature maps. Additionally, it employs a single stream deep neural network with multiple side outputs to optimise edge segmentation. Their solution performs image-to-image prediction by leveraging fully convolutional neural networks and deeply supervised nets. The model uses weighted loss to overcome the imbalanced classification problem, where the target pixels result in a larger weight in the loss function. According to the authors, ECNet allows

performing predictions much faster and more efficiently, making it possible to utilise the limited resources on embedded platforms effectively.

An unsupervised statistically based algorithm for image classification was presented in [38]. Unlike the other methods, it does not employ a statistical model of the shadow region. It merges highlight and shadow detection utilising a weighted likelihood ratio test taking into consideration the spatial distribution of the target. In the next step, the sonar elevation and scan angle are calculated. Using these acquired data and the statistical features of the pixels, a support vector machine (SVM) classifies shadow and background regions. The authors claimed that the algorithm is robust and does not require knowledge about the target's shape or size.

McKay et al. [39] used a sparse reconstruction-based classification (SRC), which shows resiliency to noise, blur and occlusion. Their method incorporates a novel interpretation of spike and slab probability distributions as Bayesian discrimination combined with a dictionary learning scheme for patch extractions. It also facilitates anomaly detection to avoid false identifications without additional training implementation. Accessing the database provided by the U.S. Naval Surface Warfare Center in this method proves robustness by classifying targets with diverse geometric arrangements, bothersome Rayleigh noise and background clutter [39].

Deep convolutional neural networks were used to classify underwater targets in synthetic-aperture sonar (SAS) imagery in [40]. The authors used a massive database of sonar data collected at sea during different expeditions in various geographical locations. The database consisted of dummy mine shapes, more realistic mine-like targets, other human-made objects and calibrated rocks. The authors developed a new training procedure to augment the training data and avoid overfitting. In this procedure, the deep networks performed several binary classification tasks in which different objects had to be discriminated. The results showed that deep networks can learn valuable differences between similar objects and outperform traditional feature-based classifiers.

4.3. MLO Detection

Segmentation defines ROIs by considering all segments, including highlights, shadows or combinations. These regions contain potential mine-like objects. To distinguish a mine, the detection step is needed. Some detection solutions utilise matched or template filters [17,28,41]. In simple algorithms, the templates are used for detecting highlight and shadow combinations. More sophisticated procedures take additional steps, including highlight clutter and shadow clutter region recognition. Some also convolve templates with distinct areas in the image and employ a threshold to establish if the analysed area constitutes an MLO [17]. In this attempt, the template covers all possible MLOs but is able to recognise the bottom clutters [28]. Another template matching method generates a subspace of shapes with six vertices representing the MLO. Subsequently, it compares the MLO's normalised shape with the subspace of defined shapes [42]. As a result, the most distinct regions are considered for further processing.

The main idea of the template matching technique is briefly described below. Assuming that the vehicle is moving at a known distance from the seabed and range to the MLO (see Figure 10), the shadow length can be solved as a function of the range using the following equation:

$$S(R) = \frac{O \cdot R}{A - O}, \quad (5)$$

where A is the vehicle's distance from the seabed, R is the range to the object, O is the object's height and S is the shadow length.

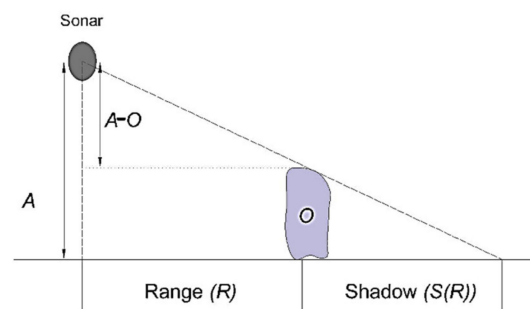


Figure 10. Template's geometry.

The obtained data can be compared with the characteristics of the known MLOs to determine similarity. Figure 11 presents the most distinctive regions utilised during the template matching test.

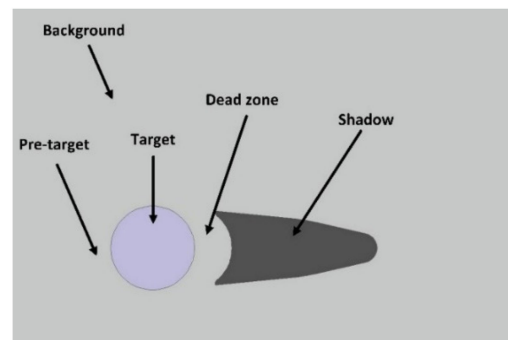


Figure 11. Template of an MLO.

In ref. [43], a Gabor-based deep neural network architecture was developed to detect MLOs. The steerable Gabor filtering modules are embedded within the cascaded layers to enhance images' scale and orientation. The proposed Gabor neural network is designed as a feature pyramid network with a small number of trainable weights. It merges strong and weak features to detect MLOs at multiple scales accurately. Feature extraction is performed using a parametrised Gabor layer to improve the generalisation ability and efficiency. Additionally, the steerable Gabor filter is implemented in the cascade layer to improve images' scale and orientation decomposition. The Gabor neural network (GNN) is taught utilising sonar images with labelled MLOs. The authors compared the detection performance of their method with others devoted to MLO detection such as the Haar-like detector [44], LBP cascade detector [4], VGG+GAP [45] and SNN+GAP [46]. The experimental results indicated that the proposed solution is an effective MLO detection method for AUVs in terms of accuracy and model size. The obtained results are presented in Table 1.

Table 1. The performance of MLO detection methods in comparison with a GNN detector [43].

Method	Correct Detections	Incorrect Detections	Ground Truth Not Detected	Frame/s
GNN detector	174	46	42	3.018
Haar-like detector	18	319	198	0.053
LBP cascade detector	23	267	193	0.051
VGG + GAP	46	107	170	0.004
CNN + GAP	9	134	207	0.007

They also performed a comparison with state-of-the-art object detectors which were not devoted to mine detection. They considered the following methods: R-CNN [44], Fast R-CNN [45], Faster R-CNN [46], SSD300 [47] and YOLOv3 [48]. However, it is worth mentioning that the YOLOv4 and YOLOv5 methods have been introduced recently. They constitute an improved version of the YOLOv3 algorithm; hence, their accuracy can probably be higher in mine detection tasks. This implies that the YOLOv4 and YOLOv5 methods should be taken into consideration for newly developed methods.

The obtained results (see Table 2) show that the proposed method outperforms the others in terms of accuracy. It also achieves a great reduction in size in comparison with other detectors. The speed performance is worse than that for the R-CNN, Fast R-CNN and Tiny YOLOv3 methods, but the authors underlined that their main concern was to improve the detection accuracy for a reliable MLO detection algorithm.

Table 2. The performance of generic object detection methods used for MLO detection [43].

Method	AP \pm SD (%)	Frame/Second	Total Trainable Weights	Model Size (KB)
GNN	79.93 \pm 7.66	3.01	2,084,880	8305
R-CNN	37.62 \pm 11.85	0.3	25,502,912	100,632
Fast R-CNN	18.26 \pm 3.9	1.29	25,436,865	100,371
Faster R-CNN	9.41 \pm 3.16	2.89	25,703,510	101,423
Tiny YOLOv3	70.54 \pm 8.91	28.41	8,669,876	3399
Full YOLOv3	72.76 \pm 9.53	15.07	61,523,734	241,082
SSD300	27.08 \pm 6.9	7.17	91,427	91,427

The research in [49] deployed a DNN to detect MLOs on the seafloor in sidescan sonar imagery. The authors analysed the impact of the DNN depth, memory requirements, calculation requirements and training data distribution on detection efficacy. Additionally, they incorporated visualisation techniques to facilitate a user's interpretation of the model's behaviour. According to them, more complex DNN models generate better accuracy (98%) than simple ones (93%) and yield a better performance than SVMs (78%). The most complicated DNN models achieved a 1% increase in efficacy at the cost of a seventeen-fold increase in the number of trainable parameters. The presented solution requires fewer computational resources than DNNs developed for multi-class classification tasks. As a result, it is suitable for autonomous underwater vehicles.

5. Object Classification

For object classification, a set of features needs to be extracted. Based on the extracted features, the classification can be performed by calculating distances between them [42]. Other approaches utilise parallel lines via the Hough transform [29] or shape and signal-to-noise ratio analyses [41]. For example, Dobeck et al. [50] proposed 25 features due to their uniqueness. The most significant ones are listed below:

- Maximum highlight pixel value in the ROI;
- Minimum pixel value in the ROI;
- Size of highlight region;
- Size of shadow region;
- Mean of highlight region;
- Mean of shadow region;
- Distance from shadow to highlight;
- Number of shadow pixels below threshold;
- Number of highlight pixels above the threshold.

Other techniques employ feature extraction algorithms such as a wavelet packet-based feature extraction [51] or canonical correlation analysis (CCA), which is able to segment the image simultaneously [18]. In ref. [52], the k-nearest neighbour (K-NN) technique for underwater target discrimination was described. It is based on the k-nearest neighbour

regions, defined by training data and used to classify the candidate region. The neighbour's distance to the candidate region determines the belief score. In this approach, smaller distances generate higher belief scores. Another technique, the cooperating statistical snake (CSS) model, was used in [19] to recognise the boundaries of highlights and shadows. In this attempt, two statistical snakes are applied to determine the target highlight and shadow separately. It is performed by reducing the movement of the snakes by applying the relationship between the highlight and shadow.

The boundaries of shadows and highlights are often indistinguishable due to noise in the sonar images. To remedy this problem, the method of fitting a superellipse was developed in [2]. Superellipses are defined as Lamé curves in analytical geometry. They form shapes such as rectangles, rhomboids and ellipses utilising changes in the squareness of the superellipse function. In this work, superellipses were used to classify the ROIs.

The change detection technique was applied for mine classification in [28]. This technique compares identified MLOs to establish any changes that can facilitate the classification step. Only highlights with matching shadows are taken into consideration. The mine classification is based on the pixel distribution analysis of each change. Human-made targets are characterised by a normal pixel distribution in comparison to the regular distribution of other underwater objects. Finally, the classification algorithm employs a finite state Markov machine, which uses four feature vectors: highlight, shadow, highlight clutter and shadow clutter. Then, these vectors are compared to four threshold values to perform the classification step.

Underwater object classification methods mainly focus on detecting all possible mine-like objects (MLOs) and classifying them as mine or not-mine [35]. In classical methods, model-based or data-driven approaches have been widely utilised. They extract a set of features from MLOs to a training dataset; however, the obtained results depend on the similarity between the test and training data [53]. In [19], the authors used the Hausdorff distance between synthetic shadows and a real object shadow as well as size information to produce a membership function. They also implemented the object classification step based on Dempster–Shafer information. The obtained classification efficiency was improved in [54] using multi-angle view mine simulation and template matching. Apart from model-based approaches, local feature descriptors without prior knowledge have also been deployed for mine classification. Among them, the most popular are: the Haar-like feature [55], the combination of Haar features and learned features from a human operator's brain electroencephalogram (EEG) [56] and Haar-like and local binary pattern (LBP) features [4]. The extracted features are usually analysed using machine learning techniques, such as boosting [55] and support vector machines (SVMs) [57].

Other feature-based methods use geometric visual descriptors, such as the scale-invariant feature transform (SIFT) [57–59] and local binary patterns (LBPs) [4,60]. In [57], the authors applied dense SIFT feature extraction with various window sizes for calculating orientation histograms. Barngrover et al. [4] considered the detection capabilities using synthetic data compared to real-world images. For this purpose, they developed an algorithm based on AdaBoost to distinguish features and an optimised cascade of features to classify objects. This training algorithm demands many training examples to facilitate the machine learning module selection of the most distinct features. The authors performed experiments with the local binary pattern (LBP) features and Haar-like features. The results showed that semisynthetic training and testing can improve the classification performance in case of scarcity in real data. Additionally, it can allow determining the most promising features.

The main disadvantage of feature-based methods is that feature extractors must be manually designed to generate a feature vector from the input image window. Therefore, in recent years, MLO detection and classification methods have used deep learning (DL) to process sonar images in their raw form without manual feature engineering [49,61,62]. In [49], Gebhardt et al. proposed different structures of convolutional neural networks (CNNs), where a global average pooling (GAP) layer is used before each fully connected

layer to produce a class activation map. A four-step pipeline of MLO detection, including synthetic data generation, one-class classification, background extraction and binary classification, was presented in [62]. The second and fourth steps are accomplished using an auto-encoder and a pre-trained network, VGG-19, respectively.

Transfer learning with pre-trained CNNs for mine detection and classification was used in [61], where the feature vectors train a support vector machine (SVM) on a small sonar dataset. The authors tested several pre-trained CNNs for SVM and modified CNN problems: VGG16, VGG19, VGG-f and Alexnet. The obtained results showed that the SVM using CNN features can improve detection in the case of a small training dataset. This is because, when using a small dataset, CNN is not able to tune its parameters correctly. Therefore, to obtain reliable results, highly discriminative features should be utilised. Performance can also be increased using deeper and more detailed networks.

In ref. [63], the authors presented another usage of DL for classifying MLOs in synthetic sonar aperture images. In this work, a fused anomaly detector is deployed to narrow down the pixels in SAS images and extract target-sized tiles. Then, the detector calculates the confidence map of the same size as the original image by estimating the target probability value for all pixels based on a neighbourhood around it. The confidence map considers only ROIs for the classification step.

The research in [59] compared feature detection algorithms utilising a synthetic sonar image dataset. The dataset contains MLOs located on grass, sand ripple and sand. The authors analysed the Shi–Tomasi, Harris, SURF, SIFT, ORB, STAR and FAST feature detectors and descriptors on each of these backgrounds. In the classification step, the SVM technique was utilised. The obtained results were assessed with the receiver operating curve (ROC) by comparing the number of correctly identified object features and the number of incorrectly identified ones. The authors observed that the SURF method is most suitable for the sandy seabed. The ORB method indicated the best performance for the rippled seabed, while the Harris, Shi–Tomasi, SIFT and STAR methods were not applicable for the used dataset.

Fei et al. [64] proposed a filter method for feature selection and an ensemble learning scheme in the Dempster–Shafer theory framework for underwater mine classification. The filter method adopts the composite relevance measure. This measure constitutes a weighted arithmetic average of mutual information, modified relief weight and Shannon entropy, or a weighted geometric average of these three factors. The features provided by the maximum composite relevance measure using the sufficient feature set scheme were tested for a wide range of classifiers. The results showed that the features selected by the proposed methods deliver better performance without the requirement of manual setting of parameters. The proposed filter methods are also much faster than the filter methods in the literature. The authors also concluded that the incorporation of a priori knowledge about the classifier's performance increases accuracy.

The reviewed MLO detection and classification methods are listed in Table 3.

Table 3. Summary of MLO detection and classification methods reviewed in this article.

Application	Authors	Year	Technique	Remarks
Detection	Attaf et al. [36]	2016	Classical image processing	<ul style="list-style-type: none"> • Uses the amplitude dominant component analysis • Exploits the saliency of sonar images
Detection Classification	Reed et al. [19]	2003	Classical image processing	<ul style="list-style-type: none"> • Uses unsupervised Markov random field model • Utilises a priori information about the spatial relationship between highlights and shadows
Detection	Acosta et al. [31]	2015	Classical image processing	<ul style="list-style-type: none"> • Uses Cell Average–Constant False Alarm Rate • Suitable for autonomous underwater vehicles
Detection Classification	Tucker et al. [18]	2007	Classical image processing	<ul style="list-style-type: none"> • Uses canonical correlation analysis • High accuracy (88%)
Detection	Rao et al. [17]	2009	ML	<ul style="list-style-type: none"> • Used for real-time application • Good results on the database provided by the Naval Surface Warfare Center
Detection	Barngrover et al. [56]	2016	ML	<ul style="list-style-type: none"> • Uses brain–computer interface, Haar-like feature classifier and SVM
Detection Classification	Saisan et al. [42]	2008	ML	<ul style="list-style-type: none"> • Treats mine detection as a two-dimensional object recognition and localisation problem • Shows good results on benchmarking data created from the mine dataset
Detection	Abu et al. [38]	2019	ML	<ul style="list-style-type: none"> • Uses a support vector machine over the statistical features • Does not require knowledge about the target’s shape or size
Detection	McKay et al. [39]	2017	ML	<ul style="list-style-type: none"> • Uses a sparse reconstruction-based classification • Resilient to noise, blur and occlusion
Detection	Thanh et al. [43]	2020	DL	<ul style="list-style-type: none"> • Uses Gabor-based detector • Achieves competitive performance compared to the existing approaches

Table 3. Cont.

Application	Authors	Year	Technique	Remarks
Detection Classification	Gebhardt et al. [49]	2017	DL	<ul style="list-style-type: none"> • High accuracy (93%) • Suitable for autonomous underwater vehicles
Detection	Wu et al. [37]	2019	DL	<ul style="list-style-type: none"> • Uses the efficient convolutional network architecture for semantic segmentation • Applicable for real-time processing tasks
Detection Classification	Williams et al. [40]	2016	DL	<ul style="list-style-type: none"> • Uses deep networks learned for several binary classification tasks
Detection Classification	Ciany et al. [41]	2003	Classical image processing	<ul style="list-style-type: none"> • Uses signal-to-noise ratio and shape of the objects • Suitable for autonomous underwater vehicles
Classification	Neumann et al. [29]	2008	Classical image processing	<ul style="list-style-type: none"> • Uses the Hough transformation • Significantly reduces the number of false detections
Classification	Dobeck et al. [50]	1997	DL	<ul style="list-style-type: none"> • Uses the k-nearest neighbour attractor-based neural network classifier • Classifies mine-size regions which are similar to a mine's signature
Classification	Yao et al. [51]	2002	DL	<ul style="list-style-type: none"> • Uses feature selection and neural network classifier • Introduces a sub-band fusion mechanism for wideband data
Detection Classification	Dura et al. [2]	2008	Classical image processing	<ul style="list-style-type: none"> • Uses a superellipse fitting approach • The classification rate is higher than 80%
Classification	Wei et al. [28]	2009	Classical image processing	<ul style="list-style-type: none"> • Uses change detection techniques • Satisfactory results for two sets of bi-temporal sidescan sonar images
Classification	McKay et al. [61]	2017	DL	<ul style="list-style-type: none"> • Uses convolutional neural networks and transfer learning

Table 3. Cont.

Application	Authors	Year	Technique	Remarks
Classification	Galusha et al. [63]	2019	DL	<ul style="list-style-type: none"> • Uses convolutional neural networks for location based on cross-validation
Classification	Fei et al. [64]	2015	DL	<ul style="list-style-type: none"> • Uses an ensemble learning scheme in the Dempster–Shafer theory framework • Manual setting of parameters is not required

6. Conclusions

Underwater mine detection and classification in sonar imagery pose a challenging task due to the low quality of sonar images. Sonar images are complicated to analyse due to speckle noise and environmental conditions causing spurious shadows, sidelobe effects and multipath return. This results in significant variability in targets, clutter and background signatures.

To deal with these obstacles, numerous image processing techniques have been developed. They can be divided into classical image processing, machine learning and deep learning techniques.

The most conventional approach to MLO detection and classification is classical image processing. It incorporates highlight and shadow regions to spot suspicious objects on the seafloor. This process involves the segmentation step, which divides the image into highlights, shadows and background regions. Then, the templates are used for detecting highlight and shadow combinations. The flaw in classical image processing is that it demands more workforce and training efforts. It requires experts to design image features necessary for object detection and classification. Additionally, its performance is image quality dependent. Therefore, an extra enhancement step is needed to ensure its accuracy. However, a feature that compensates for its flaws is that it does not need an extensive database.

More advanced image processing techniques employ machine learning, where image features are used to detect and classify objects automatically. Detection and classification need a supervised or unsupervised learning process to work. In both modes, an expert needs to design the image feature specifications. ML detects and classifies objects based on their characteristics without utilising any of the seabed characteristics. In the ML technique, the imaging data stipulate a high quality, which is challenging to achieve using sonar images. Therefore, image enhancement is implemented in this technique as well. Since ML necessitates the design of image features, it is time demanding and a tiring operation. However, when trained on small datasets, it provides an appropriate level of accuracy. Thus, ML is considered one of the more reliable techniques in image processing.

A step further in image processing is deep learning, a subfield of ML that provides algorithms inspired by the brain's neuron connection composition. It employs structured and unstructured data and extracts features automatically. Since DL algorithms demand a vast quantity of data, a severe problem in its operation path is the lack of publicly available datasets, alongside military non-disclosure of mine detection tasks. To resolve this, some techniques including sonar simulators, data augmentation and transfer learning can be employed. However, these techniques do not compensate for the lack of actual representative datasets. As a result, the performance of DL algorithms for mine detection and classification is lower compared to other computer vision applications.

An added disadvantage of deep learning is that images demand manual labelling, which requires an additional workload and extended working hours. Furthermore, deep learning is also susceptible to low-quality data. Therefore, even though the image enhancement step is unnecessary for DL as it is in most of the most popular computer vision applications, researchers often implement it for noise reduction and contrast improvement of low-quality sonar images.

The upcoming trend is to combine classical image processing with DL. A classical method distinguishes ROIs in the detection step while DL classifies ROIs as MLOs or benign objects in this combined technique. This combination can improve the performance of the classification step since the neural network analyses only the ROIs' areas in the image. As a result, the computationally expensive DL algorithms are functional in analysing just small parts of sonar images.

This paper examined the current and previous generations of classical image processing, machine learning and deep learning methods. It can help upcoming research in developing new detection and classification algorithms, thus laying the groundwork for the next generation of advanced techniques.

Funding: This paper was funded by the Polish Ministry of Defence Research Grant entitled "Vision system for object detection and tracking in the marine environment".

Conflicts of Interest: The author declares no conflict of interest.

References

1. Del Rio Vera, J.; Coiras, E.; Groen, J.; Evans, B. Automatic Target Recognition in Synthetic Aperture Sonar Images Based on Geometrical Feature Extraction. *EURASIP J. Adv. Signal Process.* **2009**, *2009*, 109438. [[CrossRef](#)]
2. Dura, E.; Bell, J.; Lane, D. Superellipse Fitting for the Recovery and Classification of Mine-Like Shapes in Sidescan Sonar Images. *IEEE J. Ocean. Eng.* **2008**, *33*, 434–444. [[CrossRef](#)]
3. Neupane, D.; Seok, J. A review on deep learning-based approaches for automatic sonar target recognition. *Electronics* **2020**, *9*, 1972. [[CrossRef](#)]
4. Barngrover, C.; Kastner, R.; Belongie, S. Semisynthetic versus real-world sonar training data for the classification of mine-like objects. *IEEE J. Ocean. Eng.* **2015**, *40*, 48–56. [[CrossRef](#)]
5. Cerqueira, R.; Trocoli, T.; Neves, G.; Joyeux, S.; Albiez, J.; Oliveira, L. A novel GPU-based sonar simulator for real-time applications. *Comput. Graph.* **2017**, *68*, 66–76. [[CrossRef](#)]
6. Borawski, M.; Forczmański, P. Sonar Image Simulation by Means of Ray Tracing and Image Processing. In *Enhanced Methods in Computer Security, Biometric and Artificial Intelligence Systems*; Kluwer Academic Publishers: Boston, MA, USA, 2005; pp. 209–214.
7. Saeidi, C.; Hodjatkashani, F.; Fard, A. New tube-based shooting and bouncing ray tracing method. In Proceedings of the 2009 International Conference on Advanced Technologies for Communications, Hai Phong, Vietnam, 12–14 October 2009; pp. 269–273.
8. Danesh, S.A. *Real Time Active Sonar Simulation in a Deep Ocean Environment*; Massachusetts Institute of Technology: Cambridge, MA, USA, 2013.
9. Saito, H.; Naoi, J.; Kikuchi, T. Finite Difference Time Domain Analysis for a Sound Field Including a Plate in Water. *Jpn. J. Appl. Phys.* **2004**, *43*, 3176–3179. [[CrossRef](#)]
10. Maussang, F.; Rombaut, M.; Chanussot, J.; Hétet, A.; Amate, M. Fusion of local statistical parameters for buried underwater mine detection in sonar imaging. *EURASIP J. Adv. Signal Process.* **2008**, *2008*, 1–19. [[CrossRef](#)]
11. Maussang, F.; Chanussot, J.; Rombaut, M.; Amate, M. From Statistical Detection to Decision Fusion: Detection of Underwater Mines in High Resolution SAS Images. In *Advances in Sonar Technology*; IntechOpen: London, UK, 2009; ISBN 9783902613486.
12. Lurton, X. *An Introduction to Underwater Acoustics*; Springer: Berlin/Heidelberg, Germany, 2010; ISBN 978-3-540-78480-7.
13. Barngrover, C.M. *Automated Detection of Mine-Like Objects in Side Scan Sonar Imagery*; University of California: San Diego, CA, USA, 2014.
14. Tellez, O.L.L.; Borghgraef, A.; Mersch, E. The Special Case of Sea Mines. In *Mine Action—The Research Experience of the Royal Military Academy of Belgium*; IntechOpen: London, UK, 2017; pp. 267–322.
15. Doerry, A. Introduction to Synthetic Aperture Sonar. In Proceedings of the 2019 IEEE Radar Conference (RadarConf), Boston, MA, USA, 22–26 April 2019; pp. 1–90.
16. Atherton, M. *Echoes and Images, The Encyclopedia of Side-Scan and Scanning Sonar Operations*; OysterInk Publications: Vancouver, BC, Canada, 2011; ISBN 098690340X.
17. Rao, C.; Mukherjee, K.; Gupta, S.; Ray, A.; Phoha, S. Underwater mine detection using symbolic pattern analysis of sidescan sonar images. In Proceedings of the 2009 American Control Conference, St. Louis, MO, USA, 10–12 June 2009; pp. 5416–5421. [[CrossRef](#)]

18. Tucker, J.D.; Azimi-Sadjadi, M.R.; Dobeck, G.J. Canonical Coordinates for Detection and Classification of Underwater Objects From Sonar Imagery. In Proceedings of the OCEANS 2007—Europe, Aberdeen, UK, 18–21 June 2007; pp. 1–6.
19. Reed, S.; Petillot, Y.; Bell, J. An automatic approach to the detection and extraction of mine features in sidescan sonar. *IEEE J. Ocean. Eng.* **2003**, *28*, 90–105. [[CrossRef](#)]
20. Klausner, N.; Azimi-Sadjadi, M.R.; Tucker, J.D. Underwater target detection from multi-platform sonar imagery using multi-channel coherence analysis. In Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics, San Antonio, TX, USA, 11–14 October 2009; pp. 2728–2733.
21. Langner, F.; Knauer, C.; Jans, W.; Ebert, A. Side scan sonar image resolution and automatic object detection, classification and identification. In Proceedings of the OCEANS '09 IEEE Bremen: Balancing Technology with Future Needs, Bremen, Germany, 11–14 May 2009; pp. 1–8.
22. Hożyń, S.; Zalewski, J. Shoreline Detection and Land Segmentation for Autonomous Surface Vehicle Navigation with the Use of an Optical System. *Sensors* **2020**, *20*, 2799. [[CrossRef](#)] [[PubMed](#)]
23. Hożyń, S.; Żak, B. Local image features matching for real-time seabed tracking applications. *J. Mar. Eng. Technol.* **2017**, *16*, 273–282. [[CrossRef](#)]
24. Wang, X.; Wang, H.; Ye, X.; Zhao, L.; Wang, K. A novel segmentation algorithm for side-scan sonar imagery with multi-object. In Proceedings of the 2007 IEEE International Conference on Robotics and Biomimetics (ROBIO), Sanya, China, 15–18 December 2007; pp. 2110–2114.
25. Om, H.; Biswas, M. An Improved Image Denoising Method Based on Wavelet Thresholding. *J. Signal Inf. Process.* **2012**, *3*, 109–116. [[CrossRef](#)]
26. Hożyń, S.; Żak, B. Segmentation Algorithm Using Method of Edge Detection. *Solid State Phenom.* **2013**, *196*, 206–211. [[CrossRef](#)]
27. Celik, T.; Tjahjadi, T. A Novel Method for Sidescan Sonar Image Segmentation. *IEEE J. Ocean. Eng.* **2011**, *36*, 186–194. [[CrossRef](#)]
28. Wei, S.; Leung, H.; Myers, V. An automated change detection approach for mine recognition using sidescan sonar data. In Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics, San Antonio, TX, USA, 11–14 October 2009; pp. 553–558.
29. Neumann, M.; Knauer, C.; Nolte, B.; Brecht, D.; Jans, W.; Ebert, A. Target detection of man made objects in side scan sonar images segmentation based false alarm reduction. *J. Acoust. Soc. Am.* **2008**, *123*, 3949. [[CrossRef](#)]
30. Huo, G.; Yang, S.X.; Li, Q.; Zhou, Y. A Robust and Fast Method for Sidescan Sonar Image Segmentation Using Nonlocal Despeckling and Active Contour Model. *IEEE Trans. Cybern.* **2017**, *47*, 855–872. [[CrossRef](#)]
31. Acosta, G.G.; Villar, S.A. Accumulated CA–CFAR Process in 2-D for Online Object Detection From Sidescan Sonar Data. *IEEE J. Ocean. Eng.* **2015**, *40*, 558–569. [[CrossRef](#)]
32. Ye, X.-F.; Zhang, Z.-H.; Liu, P.X.; Guan, H.-L. Sonar image segmentation based on GMRF and level-set models. *Ocean Eng.* **2010**, *37*, 891–901. [[CrossRef](#)]
33. Fei, T.; Kraus, D. An expectation-maximisation approach assisted by Dempster-Shafer theory and its application to sonar image segmentation. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012; pp. 1161–1164. [[CrossRef](#)]
34. Szymak, P.; Piskur, P.; Naus, K. The Effectiveness of Using a Pretrained Deep Learning Neural Networks for Object Classification in Underwater Video. *Remote Sens.* **2020**, *12*, 3020. [[CrossRef](#)]
35. 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](#)]
36. Attaf, Y.; Boudraa, A.O.; Ray, C. Amplitude-based dominant component analysis for underwater mines extraction in side scans sonar. In Proceedings of the Oceans 2016—Shanghai, Shanghai, China, 10–13 April 2016. [[CrossRef](#)]
37. Wu, M.; Wang, Q.; Rigall, E.; Li, K.; Zhu, W.; He, B.; Yan, T. ECNet: Efficient Convolutional Networks for Side Scan Sonar Image Segmentation. *Sensors* **2019**, *19*, 2009. [[CrossRef](#)] [[PubMed](#)]
38. Abu, A.; Diamant, R. A Statistically-Based Method for the Detection of Underwater Objects in Sonar Imagery. *IEEE Sens. J.* **2019**, *19*, 6858–6871. [[CrossRef](#)]
39. McKay, J.; Monga, V.; Raj, R.G. Robust Sonar ATR Through Bayesian Pose-Corrected Sparse Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5563–5576. [[CrossRef](#)]
40. Williams, D.P. Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 2497–2502. [[CrossRef](#)]
41. Ciany, C.M.; Zurawski, W.C.; Dobeck, G.J.; Weilert, D.R. Real-time performance of fusion algorithms for computer aided detection and classification of bottom mines in the littoral environment. In Proceedings of the Oceans 2003. Celebrating the Past ... Teaming Toward the Future (IEEE Cat. No.03CH37492), San Diego, CA, USA, 22–26 September 2003; Volume 2, pp. 1119–1125.
42. Saisan, P.; Kadambe, S. Shape normalised subspace analysis for underwater mine detection. In Proceedings of the 2008 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. 1892–1895.
43. Thanh Le, H.; Phung, S.L.; Chapple, P.B.; Bouzerdoum, A.; Ritz, C.H.; Tran, L.C. Deep gabor neural network for automatic detection of mine-like objects in sonar imagery. *IEEE Access* **2020**, *8*, 94126–94139. [[CrossRef](#)]

44. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; Volume 1, pp. 580–587.
45. Girshick, R. Fast R-CNN. In Proceedings of the ICCV 2015, Santiago, Chile, 13–16 December 2015; pp. 1440–1448. [[CrossRef](#)]
46. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
47. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. *SSD: Single Shot MultiBox Detector*; Springer International Publishing: New York, NY, USA, 2016; pp. 21–37.
48. Redmon, J.; Farhadi, A. *YOLO v.3: An Incremental Improvement*; In Computer Vision and Pattern Recognition; Springer: Berlin/Heidelberg, Germany, 2018; pp. 1804–2767.
49. Gebhardt, D.; Parikh, K.; Dzieciuch, I.; Walton, M.; Vo Hoang, N.A. Hunting for naval mines with deep neural networks. In Proceedings of the OCEANS 2017—Anchorage, Anchorage, AK, USA, 18–21 September 2017; pp. 1–5.
50. Dobeck, G.J.; Hyland, J.C.; Smedley, L. *Automated Detection and Classification of Sea Mines in Sonar Imagery*; Dubey, A.C., Barnard, R.L., Eds.; SPIE: Bellingham, WA, USA, 1997; pp. 90–110.
51. Yao, D.; Azimi-Sadjadi, M.R.; Jamshidi, A.A.; Dobeck, G.J. A study of effects of sonar bandwidth for underwater target classification. *IEEE J. Ocean. Eng.* **2002**, *27*, 619–627. [[CrossRef](#)]
52. Li, D.; Azimi-Sadjadi, M.R.; Robinson, M. Comparison of different classification algorithms for underwater target discrimination. *IEEE Trans. Neural Netw.* **2004**, *15*, 189–194. [[CrossRef](#)] [[PubMed](#)]
53. Myers, V.; Fawcett, J. A Template Matching Procedure for Automatic Target Recognition in Synthetic Aperture Sonar Imagery. *IEEE Signal Process. Lett.* **2010**, *17*, 683–686. [[CrossRef](#)]
54. Cho, H.; Gu, J.; Yu, S.-C. Robust Sonar-Based Underwater Object Recognition Against Angle-of-View Variation. *IEEE Sens. J.* **2016**, *16*, 1013–1025. [[CrossRef](#)]
55. Sawas, J.; Petillot, Y.; Pailhas, Y. Cascade of boosted classifiers for rapid detection of underwater objects. In Proceedings of the European Conference on Underwater Acoustics, Istanbul, Turkey, 5–9 July 2010; pp. 1–8.
56. Barngrover, C.; Althoff, A.; DeGuzman, P.; Kastner, R. A Brain–Computer Interface (BCI) for the Detection of Mine-Like Objects in Sidescan Sonar Imagery. *IEEE J. Ocean. Eng.* **2016**, *41*, 123–138. [[CrossRef](#)]
57. Hollesen, P.; Connors, W.A.; Trappenberg, T. Comparison of Learned versus Engineered Features for Classification of Mine Like Objects from Raw Sonar Images. In *Lecture Notes in Computer Science, Proceedings of the Advances in Artificial Intelligence, St. John's, NL, Canada, 25–27 May 2011*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6657 LNAI, pp. 174–185, ISBN 9783642210426.
58. Zhu, Z.; Xu, X.; Yang, L.; Yan, H.; Peng, S.; Xu, J. A model-based Sonar image ATR method based on SIFT features. In Proceedings of the OCEANS 2014—TAIPEI, Taipei, Taiwan, 7–10 April 2014; pp. 1–4.
59. Tueller, P.; Kastner, R.; Diamant, R. A Comparison of Feature Detectors for Underwater Sonar Imagery. In Proceedings of the OCEANS 2018 MTS/IEEE Charleston, Charleston, SC, USA, 22–25 October 2018; pp. 1–6.
60. Dale, J.; Galusha, A.P.; Keller, J.; Zare, A. Evaluation of image features for discriminating targets from false positives in synthetic aperture sonar imagery. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXIV*; Isaacs, J.C., Bishop, S.S., Eds.; SPIE: Bellingham, WA, USA, 2019; p. 10.
61. McKay, J.; Gerg, I.; Monga, V.; Raj, R.G. What's mine is yours: Pretrained CNNs for limited training sonar ATR. In Proceedings of the OCEANS 2017—Anchorage, Anchorage, AK, USA, 18–21 September 2017; pp. 1–7.
62. Denos, K.; Ravaut, M.; Fagette, A.; Lim, H.-S. Deep learning applied to underwater mine warfare. In Proceedings of the OCEANS 2017—Aberdeen, Aberdeen, UK, 19–22 June 2017; pp. 1–7.
63. Galusha, A.P.; Dale, J.; Keller, J.; Zare, A. Deep convolutional neural network target classification for underwater synthetic aperture sonar imagery. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXIV*; Isaacs, J.C., Bishop, S.S., Eds.; SPIE: Bellingham, WA, USA, 2019; p. 5.
64. Fei, T.; Kraus, D.; Zoubir, A.M. Contributions to automatic target recognition systems for underwater mine classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 505–518. [[CrossRef](#)]