# Automatic Target Recognition for Mine Countermeasure Missions using Forward Looking Sonar data

Narcís Palomeras, Thomas Furfaro, David P. Williams, Marc Carreras, Samantha Dugelay

*Abstract*—The detection of objects on the sea floor is a complex task. The domain of the detection and classification of naval mines is additionally complicated by the high risk nature of the task. Autonomous underwater vehicles have been used in naval mine countermeasures (MCM) operations to search large areas using sensors such as side-scan or synthetic aperture sonars. These sensors generally have a high coverage rate, while sacrificing spatial resolution. Conversely, sensors with higher resolution but lower coverage (such as forward-looking sonars and electro-optical cameras) are employed for the later classification and identification stages of the MCM mission. However, in order to autonomously execute a target reacquisition mission, it is important to be able to collect and process data automatically and in near real-time, onboard an Autonomous Underwater Vehicle (AUV). For this purpose, an Automatic Target Recognition (ATR) system is required. This article proposes an ATR, which can be used on board an autonomous vehicle, capable of detecting mine-like objects in forward-looking sonar data. The ATR combines a detector and a classifier, based on convolutional neural network models, with a probabilistic grid map that filters out false positives and combines reported detections at nearby locations. A strategy, combining a survey pattern with target-mapping maneuvers automatically activated by the ATR, has been designed to maximize the performance of this ATR. The whole system has been tested in simulation as well as using data from previous mine countermeasure exercises, the results of which are presented here.

*Index Terms*—Unmanned underwater vehicles, sonar detection, machine learning, automatic control

## I. INTRODUCTION

IN 1854, the United States president John Quincy Adams defined mine warfare as "not fair and honest" [1]. However, these deadly devices, considered to be the most cost-effective form of naval warfare [2], continue to be employed and enhanced to the present day.

The manufacture and deployment of mines is relatively cheap, while detecting and removing them is a costly and dangerous effort. More than 50 countries currently have the capacity to deploy naval mines [3], with this number continuing to increase in recent decades [4]. Such devices might also be built and deployed by terrorist groups, as a form of underwater improvised explosive devices (IED). Therefore, it

N. Palomeras, T. Furfaro, D. P. Williams, and S. Dugelay were with NATO Science and Technology Organization Centre for Maritime Research and Experimentation (CMRE) Viale San Bartolomeo 400, 19126 La Spezia (SP), Italy e-mail: (npalomer@eia.udg.edu).

M. Carreras was with Underwater Robotics Research Center (CIRS), Computer Vision and Robotics Institute (VICOROB), Universitat de Girona, 17003 Girona, Spain

Manuscript received April 30, 2021.

is paramount to develop mechanisms to detect and counteract them.

Mine sweepers, mine hunters and MCM vessels, are the most employed systems to neutralize these threats. Mines sitting on the seabed (bottom mines) can be hard to detect, as features of the sea floor can obscure these devices from sonar. These mines use to carry a much larger warhead than moored ones and can have very sophisticated sensors which make them much harder to sweep [5]. To defeat these bottom mines, autonomous underwater vehicles (AUV) are increasingly being employed. The standard procedure is to carry out survey missions with a side scan sonar (SSS) or a synthetic aperture sonar (SAS) capable of detecting Mine-Like Contacts (MILCO). A classic approach for such a survey missions is to have a single vehicle following a "lawnmower" or paired-track pattern to acquire data that may be automatically processed or manually inspected by an operator once the vehicle is recovered. Advanced methods use automatic target recognition (ATR) systems on board that are able to provide the position of a contact, with some uncertainty, while the mission is still being executed [6]. Due to the limited resolution of these long range sonar systems it is not always easy to classify detections among actual mines or objects with a similar shape. To overcome this problem, work has begun on systems composed of multiple heterogeneous AUVs. The basic concept is that a long-range vehicle covers a large area searching for MILCOs while one or more AUVs, more maneuverable and equipped with lower-range but higher-resolution sensors, reacquire each of the contacts found during the wide area survey. This approach provides higher resolution images and additional viewpoints for each contact, thus simplifying the classification process. The work presented in this article envisions this cooperative methodology, and develops the necessary capabilities for an AUV, equipped with a high-resolution forward looking sonar (FLS), to reacquire additional data of a contact given only a rough initial estimate.

Forward-looking sonar are becoming increasingly popular for many applications such as mapping [7], obstacle avoidance [8] or automatic target recognition [9]. However, though high frequency FLS can provide higher resolution images than other sonars, the interpretation and analysis of their data is still challenging.

There are several methods in the literature to perform automatic object detection on sonar data. These methods are based on different techniques such as local contrast, template matching or machine learning to name some of the most

used. Among these techniques, convolutional neural networks (CNNs) have proven to be a method that generalizes well through different scenarios [10]. However, CNNs require large amounts of data to be properly trained, which can be a problem in the MCM domain where data is scarce. Our proposal consists of designing CNN models small enough to require a limited amount of data to be trained, as well as making use of techniques such as transfer learning when possible and necessary. Regardless of the techniques used, when dealing with noisy sonar data, an automatic detector will never reach an accuracy of 100%. For this reason, we propose to use a probabilistic grid map after the detector that will help both to reduce the impact of false positives, and to aggregate the weight of nearby detections who have a high probability of belonging to the same target. This probabilistic framework uses the spatial information of each sonar frame, taking advantage of the FLS high frame rate.

The multiple vehicle cooperative strategy proposed in this article was explored in the past in the NATO exercises: Trident Juncture MCM EXperiment 2015 (TJMEX'15) and Olives Noires MCM EXperiment 2016 (ONMEX'16). In these two campaigns, after surveying a large area with an AUV equipped with a SAS sonar, a small AUV was sent to execute a reacquisition mission on the contacts detected by the former. In those exercises, the data gathered in the reacquisition mission were then manually analyzed, once the vehicle was recovered, to detect and classify all the targets present in them. In this article we propose a FLS-based ATR that can be used in real time on board an AUV to automate this process, as well as a strategy to maximize the overall ATR performance. The datasets from TJMEX'15 an ONMEX'16 are the basis of the work presented here.

The paper is organized as follows. First, several works on sonar-based ATR approaches are reviewed. Then, Section III presents the proposed methodology. First the ATR and all the modules that compose it are presented in detail. Then, the strategy to be followed to maximize the ATR performance is defined. The document continues with the results obtained by applying the designed ATR to different data sets obtained during previously mentioned exercises, as well as by simulating the proposed strategy for target reacquisition. The article closes with a discussion of the results and the conclusions drawn.

## II. RELATED WORK

There is not a large body of published work regarding ATR systems for FLS. Thankfully, some of the solutions used in SAS or SSS can be adapted to FLS as the images generated by these systems share some similarities. Most sonar-based ATR makes use of one of these techniques: exploiting local contrast, template matching, or machine learning. Some of the most relevant proposals are reviewed next.

Several sonar-based object detection algorithms use the local contrast between regions to identify man-made objects within an image. To quickly calculate this local contrast, these algorithms take advantage of the integral-image representation described by [11]. The cascade of boosted classifiers have been

used to localize known objects both in SSS [12] and SAS [13]. This method looks for custom-made Haar-like features relying on previous knowledge of the target. If the geometry of the target, as well as the spatial relation between the target and the sonar are known, it is possible to estimate how the shadows will be projected and where high reflection spots will be localized. Using this information, several authors have proposed *ad hoc* filters, using the integral image approach, to segment local contrast regions. This technique has been applied to SSS and SAS data [6], but also to low resolution FLS [14] with considerable success. To further improve these results, [15] proposed to create a mosaic from the area of interest to increase the target contrast while reducing the ATR failure rate. Also based on the contrast between acoustic shadows and reflections [16] proposed a method that uses saliency techniques to detect and localize a diver in FLS imagery. The method applies frequency analysis to segment the acoustic highlight region from its surroundings and from this region extracting the acoustic shadow, improving overall the localization accuracy. In general, local contrast approaches have shown their ability to detect objects when the geometry is known. However, the kinds of features that can be extracted by these techniques are not always sufficient to distinguish between similar objects.

In cases where images of the to-be-recognized target are available, template matching techniques have also been used. Hurtós et al. [17] presented a system able to follow a chain using FLS data, where a template is created for each part of a chain link (e.g., lower left corner, upper right corner, straight segment, etc.) and detected in the FLS images using a standard template matching algorithm. In [18] a target signature generated from a simple acoustic model is compared with the actual SAS image of an object for classification purposes. Template matching is a feasible approach when there are few objects to detect or classify, those object classes are known a priori, and their geometries do not deform or otherwise change. However, methods based on template matching tend to be *ad hoc*, built-to-purpose solutions that do not generalize very well among similar but not exactly equal objects.

The most popular approach today, primarily in the optical imaging domain, but also increasingly in the sonar domain, is the use of machine learning. A method to semantically map an underwater environment using a low resolution FLS is presented in [19]. The method segments the structures in the scene and provides a descriptor combining geometric information and the acoustic intensities reflected by the objects. Then a support vector machine (SVM) is used to classify them. SVM algorithms do not require as many data as a CNN to be trained but need hand-crafted features. In contrast, CNNs are able to automatically detect important features without any human supervision, leading to generally better results at the cost of requiring more data to be trained [20]. However, it seems more clear every day that if enough data is available, the methods that deliver the best results are those based on CNNs. Several authors have designed CNNs to differentiate between 2 or more classes using FLS imagery. A CNN able to localize a small remotely operated vehicle (ROV) was proposed in [21]. The system had to be trained with 2 sets of images: one in

which only the small-ROV was visible and another in which only the typical background was visible. The CNN architecture used was a modification of the seminal CNN [22] to recognize handwritten digits.

Valdenegro-Toro proposes a coupled detector/classifier system also based on a CNN model [9]. The architecture has 2 convolutional layers followed by a pooling layer and 1 fully connected layer that are common for both the detector and the classifier. Additionally, both the detector and the classifier have 2 more fully connected independent layers. D. P. Williams presented several ATRs to work with SAS data. One of the first works in which he proposes a CNN to classify SAS data is [23]. In this work, a relatively small model (i.e., $\sim$30k parameters) was used to classify between 2 classes: targets and clutter (i.e. a mine-like objects or anything else). In an extension of this work [24], up to 8 even smaller CNN models were used in parallel, for a similar classification task, and their output was averaged. Williams claims that to determine the success of a CNN it is not enough to take into account the amount of data available for training but the relationship between the amount of data and the capacity of the model. While CNNs for classifying optical images have between $10^6$ to $10^8$ parameters to train, he proposes CNNs with less than $10^4$ parameters. Valdenegro-Toro also explored the concept of tiny networks in [25]. In this paper he argues not only the capability of these models with few parameters but also the possibility of being run in real time by computers with low computational performance.

The coupling of CNNs to the underwater sonar domain is complicated by the relative lack of training data. Data collected at sea is typically more expensive than the electro-optical analogue. Defense applications, such as MCM, suffer additionally as training data that does exist is not necessarily available for widespread use.

Three common solutions to mitigate the lack of training data for CNN are data augmentation, transfer learning [26] and synthetic data generation. Data augmentation techniques include data transformations such as scaling, cropping, flipping, padding, rotating, translating or changing contrast, which can artificially increase the training set size. When dealing with sonar images most of these conventional procedures should not be applied, because of the unique geometry and physics of target-sensor relationship. For example, to respect the properties of shadows generated by the sonar, images should not be rotated and only flipped around the range axis. Scaling images will also change the signal to noise ratio (SNR) and, obviously, any changes that affect color cannot be applied. Regarding transfer learning, several authors have exploited it by using CNNs, trained with very large optical image data sets, and then re-trained some or all their layers with sonar imagery. Published articles are not conclusive about the benefits of this method. Some examples, applied to SAS data, use pre-trained AlexNet and/or VGG-Net models [27], [28] with promising results. Others demonstrate that a custom model trained from scratch can outperform a fine-tuned pre-trained VGG-16 CNN in the context of seafloor classification [29]. Finally, new data can be generated synthetically. A classic approach to synthesizing underwater sonar data is to use simulation techniques such as ray tracing [30] or tube ray tracing [31]. Some of these methods can even work in real time [32]. However, simulations do not present all the artifacts that real sonar images have. Other solutions apply style transfer techniques to the simulation-derived images to mitigate these problems [32]. However, nowadays, one of the most promising approaches is the use of Generative Adversarial Networks (GANs) [33]. GANs are used for realistic data generation and have been applied in the underwater sonar domain mainly to generate SSS and SAS images [34] [35] but also for FLS [36].

There are few frameworks in the literature that address the whole problem of target reacquisition in the context of MCM missions. One of them is [37], in which the process of reacquiring a target while gathering FLS images with an AUV is described. The target position is identified using a SSS-based ATR and the gathered FLS images are analyzed once the mission finalizes. Another ATR working with FLS data was presented in [38]. It shows a modification of the ATR presented in [14] running on a FLS mosaic, built in real time, instead of using the raw data coming directly from the FLS.

## III. METHODOLOGY

This article proposes both an automatic target recognition (ATR) system and a data fusion step to be employed in local MCM search problems (called reacquisition). Both the ATR and the data fusion element have been designed to carry out target reacquisition missions in the context of mine counter-measure (MCM) operations. The ATR uses high resolution forward looking sonar (FLS) images and is composed of several modules that can run online onboard an autonomous underwater vehicle (AUV). The envisioned strategy commands the vehicle to gather data with the FLS, combining a survey pattern with automatically activated target-mapping maneuvers.

### A. Automatic Target Recognition

In the reacquisition strategy we envisage, the proposed ATR is the key component. The ATR ingests the FLS imagery and the vehicle odometry and returns the position and, if possible, the class of all detected targets. The ATR is composed by three main modules: a detector, one or more classifiers and a probabilistic grid map. Following the current trend in both optical and sonar object detection applications, we have decided to develop both the detection and classification modules using convolutional neural networks (CNNs). The function of the probabilistic grid map is three-fold: to filter out false positives; to combine multiple detections into a single one; and to spatially average the classifier results. Standard probabilistic tools have been used in the grid map implementation. The overall architecture of detector, classifier and the grid map is depicted in Fig. 1.

The detector and classifier of the proposed ATR have been tailored to the ARIS 3000 FLS model [39]. However, the main principles could be easily adapted to other high resolution FLS devices. The ARIS 3000 has up to 128 beams spaced 0.25° apart, with a total field of view of 30° horizontal and 15°
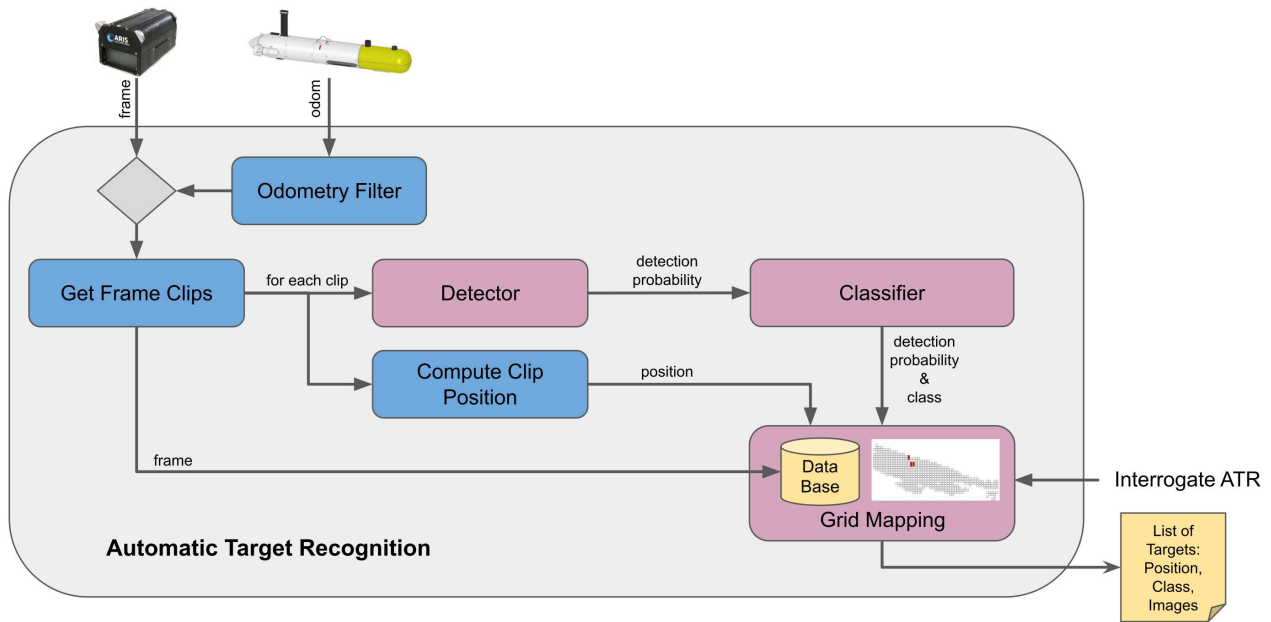
Fig. 1.  Automatic Target Recognition framework.

vertical with an along-beam resolution that ranges from 3 mm to 10 cm (see Fig. 2). Up to 1024 samples per beam can be obtained producing 128 × 1024 px polar images at 4 to 15Hz. The sonar can operate on two frequencies, 1.8MHz and 3MHz, with a maximum range of 15 m when using the lower.

Using the FLS at its maximum range presents some issues. Since all images are discrete at 128 × 1024 px maximum, the higher the range, the lower the resolution of the generated images. In addition, small variations in the AUV orientation may cause large changes in the ensonified areas (in particular at the furthest ranges), resulting in distorted images that are not useful for the ATR. In our tests we have seen that using half of its maximum range (i.e., 7.5 m) is a good compromise between coverage and resolution. To keep the physical proportions of the objects mapped by the sonar, first, the images gathered by the sonar, in polar coordinates, are transformed to Cartesian coordinates. Then, because the mine-like targets to be detected are between 0.75 and 1.5 m long, each sonar image is divided into overlapping clips to form regions of interest. A procedure has been implemented to create squared clips with a desired size and minimum overlap. Fig. 3 shows the clips obtained from an ARIS frame, already transformed to Cartesian coordinates, using a clip size of 1.5 × 1.5 m and a minimum overlap between clips of 40%. In this example 9 clips were obtained. From them, 3 contain only background, 3 are mainly covered by the target and the other 3 contain a mix of both (i.e., background with a small portion of the target). About 15% of the image is not being used as a result of forcing the clips to be square. The number of clips per frame may vary depending on: the sonar range, the size of
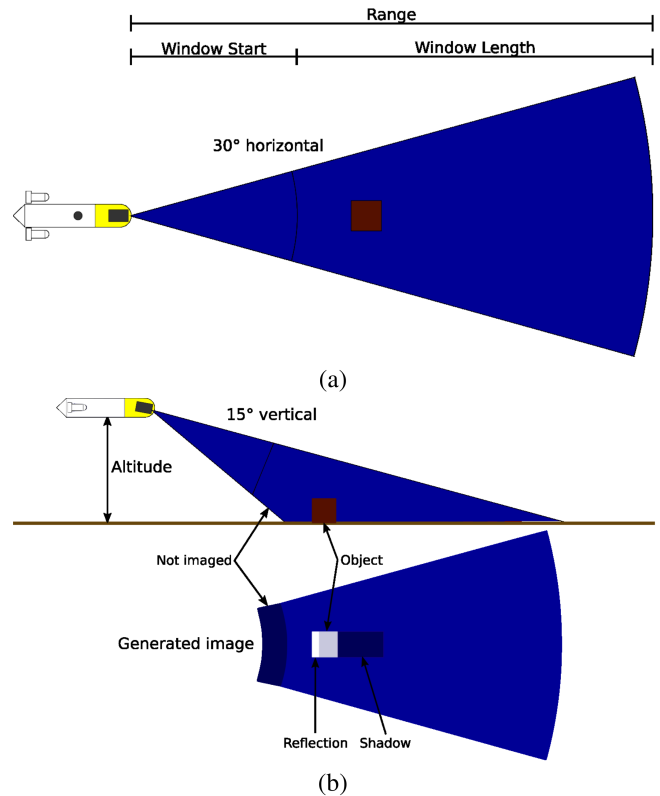


Fig. 2.  Forward Looking Sonar geometry: (a) top view and (b) side view with and schematic of a typical image (represented in Cartesian coordinates).
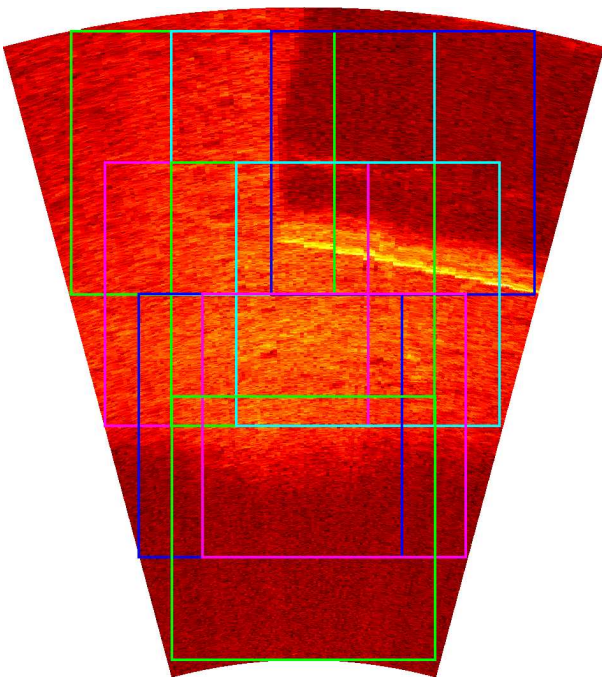
Fig. 3. Example of a FLS frame with 9 square clips where a cylindrical-shaped target can be seen.

the clips and the minimum overlap required. It is worth noting that we want the ATR to work online. Therefore, if each frame is divided into ∼10 clips and the sonar is working at ∼5 Hz, the ATR must be able to process ∼50 clips per second.

When the vehicle is far from the bottom or the roll or pitch are not adequate, the frames taken by the FLS contain little or no information. Also, if the surge or the angular velocity are too high, FLS images are distorted and useless for detection or classification purposes. Therefore, it is preferable to discard these frames to avoid introducing erroneous information into the system. When the FLS gathers a frame, a filter based on the AUV odometry determines whether the vehicle attitude, velocity (both linear and angular) and altitude are between certain user-defined thresholds that allow the FLS to correctly generate an image, and thus whether the image is a candidate to be ingested into the ATR. If a frame passes this filter, a set of clips are subsampled from it. For each clip, 2 processes are carried out. First, the position that the clip is mapping, with respect to the global system, is computed. Basic geometry operations involving the position and attitude of the AUV carrying the FLS, the transformation between the vehicle and the sonar and the sonar configuration parameters (i.e., window start, window length, sonar field of view and resolution) are used assuming an orthographic projection. We consider the sea floor to be planar and horizontal for this purpose, such that any pixel from FLS imagery corresponds to a geographic point on the flat horizontal sea floor. On the other hand, each clip is analyzed by the detector module. If the detector triggers a positive detection (i.e., an object of interest is detected within the clip), the clip is also passed to the classifier module. This classifier considers detection to be of one of two classes: target or clutter. If the clip belongs to a target class, the contact

with its resulting classifier information are stored. Otherwise, the detection is not considered. The position of the clip, the detection probability and, if available, the estimated target class are sent to the grid map system that combines all this information into a grid structure.

*1) Detection Module:* To implement the detection module, 4 of the CNN models presented in the related work [9], [22]–[24] were tested. A summary of each one of these models is presented in Table I. All these models use the ReLU activation function in all the layers, except for [23] where a Sigmoid function was used. From the two models proposed by Valdenegro-Toro and reviewed in the Related Work, [25] would possibly be more appropriate considering the small size of the available data set. However, since the article focuses on the overall framework and not on the individual performance of each component, we have used here the model presented in [9] to test a wider range of models instead of using an algorithm, which despite offering better performance with limited data, presents a similar approach to [23] and [24]. Also, of the 8 models proposed in [24] only the one that produced the best results with our data set was included in this comparison (called "Net B", in that reference).

The data used to train these models were collected in two campaigns. One was carried out in the St. Feliu de Guixols (Spain) harbour with the only purpose of gathering data to train these models. It includes 5 different objects, 3 of them deployed for this purpose, as well as images from the sea bottom. This data was taken by a team from Universitat de Girona (UdG), in collaboration with the NATO Centre for Maritime Research and Experimentation (CMRE), with a Sparus II AUV using an ARIS 3000 FLS. The second campaign was carried out by a CMRE team at the Balearic Islands (Spain) using the BlackCAT vehicle, a Bluefin 21 AUV modified to perform MCM missions, equipped with the same sonar model. Although in the second campaign only one target was detected and it appears in very few frames (i.e., only 72 frames), all the data was acquired at open sea and, therefore, it presents interesting phenomena such as sand ripples. The data collected in both scenarios have been combined to create a training and a validation data set where the targets used in the validation have never been seen during training (see Table II). The original frames were manually labelled and all clips, generated automatically, containing more than 50% of a target label were considered as a target clip, while those containing less than 10% of a target were considered seabed.

Most of the frames in this data set contained only seafloor images. In total, only around 2000 images contained an object of interest. Therefore, data augmentation operations were applied to those frames in which a target was mapped to balance both classes. Specifically, all clips containing a target were translated in various directions and flipped around the range axis. The data augmentation process has been executed automatically. This has resulted in a few more images containing targets than without.

All models in Table I were trained using a root mean square propagation optimizer and a binary cross entropy loss criterion for 100 epochs. The equivalent size of the clips was set at ∼1.4×1.4 m, which in this data set corresponds to clips that

TABLE I
CNN MODELS FOR ATR USING SONAR DATA IN LITERATURE

| Cite | Capacity | Input size w×h px @channels | Architecture | Accuracy avg. last 50 epochs |
|---|---|---|---|---|
| [22] | 2578 | 28×28 @ 1 | Conv(5, 5) @ 4<br>MaxPool(2, 2) @ 4<br>Conv(5, 5) @ 12<br>MaxPool(2, 2) @ 12<br>FullConnected(192, 1) | 87.74% $\pm$ 2.66% |
| [23] | 30201 | 167×167 @ 1 | Conv(32, 32) @ 6<br>AvgPool(2, 2) @ 6<br>Conv(17, 17) @ 8<br>AvgPool(2, 2) @ 8<br>Conv(7, 7) @ 10<br>AvgPool(2, 2) @ 10<br>Conv(5, 5) @ 24<br>AvgPool(2, 2) @ 24<br>FullConnected(216, 1) | 82.20% $\pm$ 4.01% |
| [9]<br>objectness | 1845409 | 96×96 @ 1 | Conv(5, 5) @ 32<br>MaxPool(2, 2) @ 32<br>Conv(5, 5) @ 32<br>MaxPool(2, 2) @ 32<br>FullConnected(14112, 128)<br>FullConnected(128, 96)<br>FullConnected(128, 1) | 80.62% $\pm$ 2.25% |
| [24]<br>Net B | 1509 | 267×267 @ 1 | Conv(8, 8) @ 4<br>AvgPool(4, 4) @ 4<br>Conv(6, 6) @ 4<br>AvgPool(4, 4) @ 4<br>Conv(4, 4) @ 4<br>AvgPool(2, 2) @ 4<br>Conv(5, 5) @ 4<br>AvgPool(2, 2) @ 4<br>FullConnected(4, 1) | 79.55% $\pm$ 1.58% |

TABLE II
COMPOSITION OF THE DATA SETS USED TO TRAIN AND EVALUATE THE TARGET DETECTION MODELS

| Class | Train | Validation |
|---|---|---|
| Sea bottom (background) | Harbour around objects 1, 2 and 3 Open-sea days 1 and 3 (6816 images) | Harbour around objects 4 and 5 Open-sea day 2 (1132 images) |
| Object of interest (targets) | Harbour objects 1, 2 and 3 (8470 images) | Harbour objects 4 and 5 Open-sea objects day 2 (1141 images) |

range from 300×300 to 400×400 px depending on the sonar configuration used. Each clip was scaled down to the network input size and normalized with $\mu = 0.5$ and $\sigma = 0.5$ before being processed.

The average accuracy obtained for each model, with our datasets, for the last 50 epochs is shown in Table I. All the values are around 80% with the exception of Y. Le Cun network. This model achieves a better performance at expenses of reducing the input size to only $28 \times 28$ px. This significantly improves the SNR but makes small objects almost impossible to detect.

According to the results obtained, several new models were designed, taking into account the following considerations:

- *Relation between the input size of the model and its first filter.* Unlike optical architectures, where the use of small filters is dominant, when dealing with FLS images, we have experimented that it is important for the first filter to be relatively large. This could be explained by two

reasons: the lack of detail and the high presence of noise in the sonar images.

- *Use of average pooling instead of max pooling.* This has been pointed in the literature as a technique to deal with noisy images [23]. In our tests we get small differences using the average or the maximum pooling, but we consistently obtained slightly better results with the average pooling.

- *Reduced capacity.* Overfitting is a well known problem for high capacity models when trained with small data sets [24]. According to the limited size of our training data set and the fact that the detector has to be executed on board an AUV with limited computational budget, keeping a low number of parameters is of paramount importance.

- Input size should be large enough to allow the detection of medium size objects but small enough to allow scale down the frame clips thus improving its SNR.
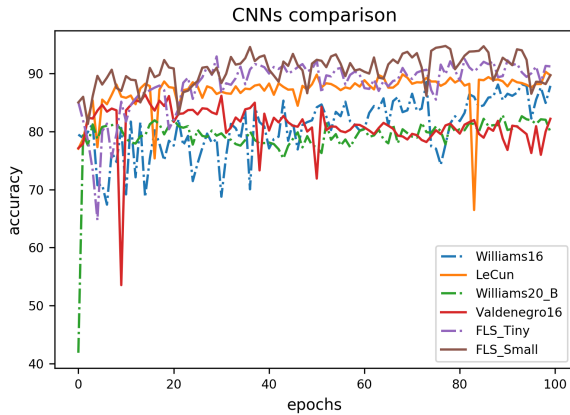
Fig. 4. Accuracy over epochs for each CNN model using the validation dataset.
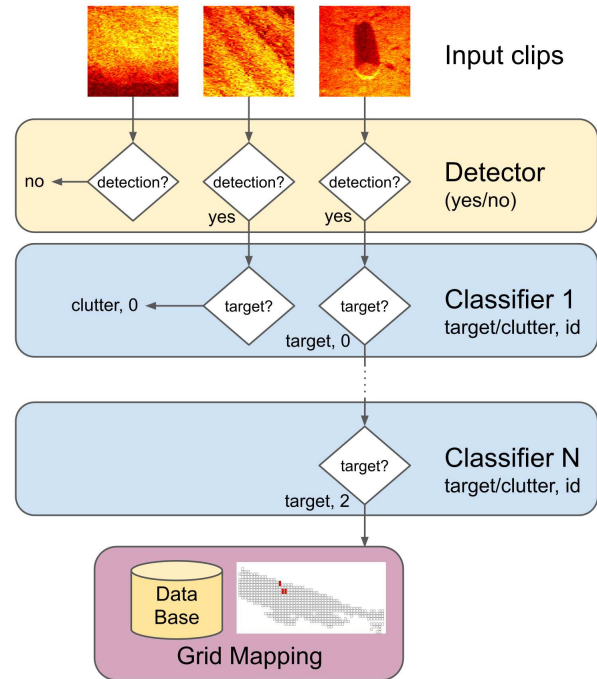


Fig. 5. Example of detection/classification with 3 different clips: clip 1 does not trigger any detection and therefore is not further processed. Clips 2 and 3 trigger detection and are analyzed by Classifier 1 which classifies clip 2 as a clutter class, and no further process is carried out, and clip 3 as a target class. Clip 3 is further processed until Classifier N classifies it as target class 2. This information is sent to the probabilistic grid map system.

The two most successful models that we have designed are described in Table III. These models significantly improve over the previous results. The main characteristics are: a significantly large initial filter (i.e., $16 \times 16$ and $32 \times 32$), use of average pooling, small capacity (i.e., $\sim 10^4$ parameters) and an input size of approximately 50% of the original image clip size (i.e., $167 \times 167$). A batch normalization layer have also been added after each convolution layer to accelerate the training process [40]. Fig. 4 shows a comparison of the accuracy for all the models in Table I and Table III.

*2) Classification Module:* The classification module is used as a second step, only if the detector has indicated that the clip contains an object of interest. The classifier is multi-class, but each class must be labelled as *target* or *clutter*. Only clips classified as a target classes, are sent to the next classification module or the probabilistic grid map system (see Fig. 5).

Training a multi-class classifier from scratch requires a large data set. However, we only have available a few hundred images of each of the classes that we would like to recognize. This amount of images is insufficient to properly train a classifier from zero. For this reason, we have used a technique known as transfer learning [26]. The idea behind transfer learning is to reuse part of a model already trained for a similar purpose. In our case, we have used the convolutional layers from the detector model in the classifier. Therefore, only the last fully connected layers of the classifier have been trained using the limited data set available. The convolutional layers perform the feature extraction. An example of the features extracted by these layers can be seen in Fig. 6 where the intermediate responses generated by the FLS_tiny model (see Table III) are shown. Despite it not being easy to interpret this responses, looking at the output after the first average pooling, it seems that channels 1 and 4 have segmented the main shadows in the input image, channel 2 the highlights and channel 3 the background. After the second average pooling, spatial information is almost completely lost and it is difficult to see which channels have segmented one element or another.

Two classifiers were implemented using the same methodology. The first one was designed to recognize 4 different objects of interest, all of them labeled as targets: truncated cone (see

Fig. 7(a) and (b)), horizontal cylinder (see Fig. 7(c)), box with rounded corners (see Fig. 7(d)) and two blocks linked (see Fig. 7(e)).

A total of 1,685 images (i.e., ~400 clips per class) were available in our data sets. These images were divided into a training and a validation data sets, with a ratio of 80%-20%. To train the classification model using transfer learning, the detector weights for the convolutional layers were fixed and the last fully connected layers were modified and trained using the new data set. Both the FLS_tiny and the FLS_small models were tested, obtaining the best results with the FLS_small. The two fully connected layers at the end of this model, were replaced by a $FullConnected(288, 128)$ with ReLU activation followed by a $FullConnected(128, 4)$ layer with a Sigmoid activation. To train the classifier, the cross-entropy loss criterion and the Adam optimizer were used for 200 epochs. Fig. 8 shows the accuracy and loss for the training and validation data over epochs. The averaged accuracy in the last 50 epochs reaches $91.92\% \pm 0.90\%$ for the training data set and it drops to $82.43\% \pm 2.85\%$ for the validation one. The confusion matrix in Table IV depicts how all the available clips are classified by the model. The most difficult objects to differentiate are the horizontal cylinder and the box with rounded corners. These two classes, due to their geometries and the nature of sonar imagery, are imaged quite differently depending on the aspect angle. Moreover, when these two classes appear only partially on a clip, the in-view component is very similar and the classifier easily confuses them.

TABLE III
PROPOSED CNN TO DETECT OBJECTS OF INTEREST ON SQUARED CLIPS FROM CARTESIAN SONAR IMAGES

| Name | Capacity | Input size | Architecture | Accuracy |
|------|----------|-----------|--------------|----------|
| FLS_tiny | 9661 | $167 \times 167$ @ 1 | Conv(16, 16) @ 4<br>AvgPool(4, 4) @ 4<br>Conv(11, 11) @ 8<br>AvgPool(4, 4) @ 8<br>Conv(6, 6) @ 16<br>AvgPool(2, 2) @ 16<br>FullConnected(16, 4)<br>FullConnected(4, 1) | $90.00\% \pm 1.56\%$ |
| FLS_small | 41749 | $167 \times 167$ @ 1 | Conv(32, 32) @ 4<br>AvgPool(2, 2) @ 4<br>Conv(15, 15) @ 8<br>AvgPool(2, 2) @ 8<br>Conv(8, 8) @ 16<br>AvgPool(2, 2) @ 16<br>Conv(5, 5) @ 32<br>AvgPool(2, 2) @ 32<br>FullConnected(288, 32)<br>FullConnected(32, 1) | $91.53\% \pm 2.03\%$ |

TABLE IV
TARGET CLASSIFICATION MATRIX

| Input Vs. Classifier Output | linked blocks | cylinder horizontal | truncated cone | box with rounded corners |
|---|---|---|---|---|
| linked blocks | 431 | 11 | 4 | 17 |
| cylinder horizontal | 26 | 354 | 8 | 15 |
| truncated cone | 1 | 23 | 321 | 7 |
| box with rounded corners | 13 | 60 | 4 | 386 |

The second classifier was trained to eliminate false detections triggered by sand ripples. The presence of sand ripples is one of the most frequent elements that may generate false positives in the detector module. Looking at the sand ripples in Fig. 7(i) and (j) and the cylindrical object in Fig. 7(c) it is easy to see the similarity between the two. Medium sized rocks also trigger the detector module producing false positives. Unfortunately, we were unable to collect enough data from them to train a classifier. The same process detailed before was used changing only the last layer for a $FullConnected(128, 2)$. The training data set contained: 1717 mine-like object clips, labeled as target, and 1998 sand ripple clips, labeled as clutter. The accuracy obtained was $88.83\% \pm 2.14\%$.

It should be noted that when sufficient computational resources are available, it is possible to run both classifiers one after the other. In this case, for instance, we would run the sand ripples classifier first and, if the clip is classified as a target, we could run the target classifier next as shown in Fig. 5.

*3) Probabilistic Grid Map System:* Using the CNN models presented before, the accuracy to predict if an image contains a target is ∼90% and the one to correctly classify a detected target is even lower. However, these results can be improved overall if the FLS frames are not considered independent observations but a representation of a particular space. Mapping each frame clip to a reference coordinate system will allow a higher-accuracy estimation of which zones are more likely to contain a target, and to better classify each detection.

As shown in Fig. 1, the probabilistic grid map system receives the position of the clip in the reference coordinate system, the output of the detection module and, if available,

the probability of belonging to a target class. The output of the detector is a value between 0 and 1 where 0 means 100% sure that the input is background and 1 means 100% sure that the input contains an object of interest. In order to map the detectors output to a binary category, a decision threshold ($\theta$) must be defined. This threshold is problem-dependent and must be tuned for each application. Usually, the best accuracy is achieved when both false positive and false negative are in equilibrium. However, for some applications can be more important to avoid false negatives or false positives. If the number of false negatives has to be reduced, $\theta$ has to be lowered, while if the number of false positives has to be reduced, $\theta$ has to be increased.

In the target reacquisition missions that we have analyzed, the frames containing a target account for less than 3% of the total. Due to the FLS high frame rate, when a target appears in the sonar field of view, it will appear in several consecutive frames. Combining these two ideas, we consider that it is better to incorrectly classify one of these frames containing a target as background than to get a 4 or 5% more false positives in the remaining 97% of data.

To see the performance of a classification model at all classification thresholds a receiver operating characteristic curve (ROC), as the one shown in Fig. 9, is used. This curve plots the true positive (TP) rate against the false positive (FP) rate for the FLS_tiny model. Measuring the area under the ROC curve (AUC) provides aggregate measure of performance across all possible classification thresholds. The AUC ranges from 0 to 1 being 0 a model whose predictions are 100% wrong and 1 a model whose predictions are 100% correct.
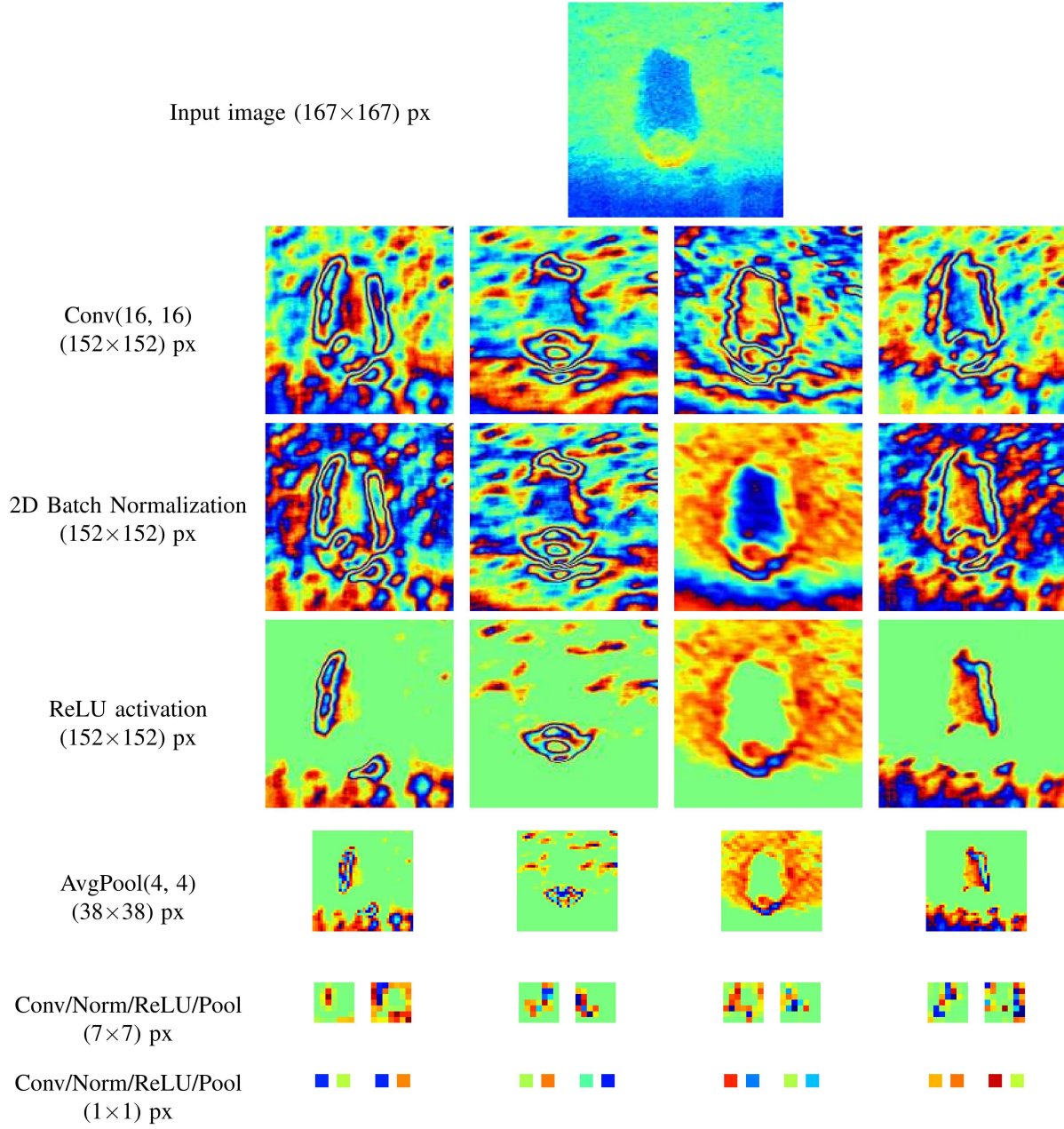
Input image $(167{\times}167)$ px

Conv(16, 16) $(152{\times}152)$ px

2D Batch Normalization $(152{\times}152)$ px

ReLU activation $(152{\times}152)$ px

AvgPool(4, 4) $(38{\times}38)$ px

Conv/Norm/ReLU/Pool $(7{\times}7)$ px

Conv/Norm/ReLU/Pool $(1{\times}1)$ px

Fig. 6. Intermediate responses for an input image (i.e., a truncated cone) using the FLS_tiny model. The output for the first 4 layers is shown individually. For the others, convolution, batch normalization, ReLU activation and average pooling are grouped. Note that images size do not match its real size to improve their visualization.

The FLS_tiny classifier has an AUC of 0.9768.

An important difference of our system with respect to a binary classifier (i.e. a detector), is that we use a probabilistic module after it, and therefore, we do not want the output of the detector to be true or false but a probability.

To obtain a corrected probability, $z_t$, according to the detector output ($p_t$) and the decision threshold ($\theta$), equation (1) is used.

$$z_t = \begin{cases} \dfrac{p_t}{2\theta}, & \text{if } p_t < \theta \\ \dfrac{1}{2} + \dfrac{p_t - \theta}{2(1-\theta)}, & \text{if } p_t \geq \theta \end{cases} \quad (1)$$

To accumulate all the probabilities obtained in a nearby place, a standard grid map is used [41]. The cell update policy for this grid map is defined by (2), where $P(n|z_t)$ is the probability to have a target in cell $n$ given a measurement $z_t$, where $z_t$ is the detector output $p_t$ corrected by (1). $P(n)$ is a prior probability and the previous estimate is defined by $P(n|z_{1:t-1})$.

$$P(n|z_{1:t}) = \left[ 1 + \frac{1 - P(n|z_t)}{P(n|z_t)} \frac{1 - P(n|z_{1:t-1})}{P(n|z_{1:t-1})} \frac{P(n)}{1 - P(n)} \right]^{-1} \quad (2)$$

Assuming that the distribution of targets in the map is initially unknown, all grid cells are initialized with $P(n) =$
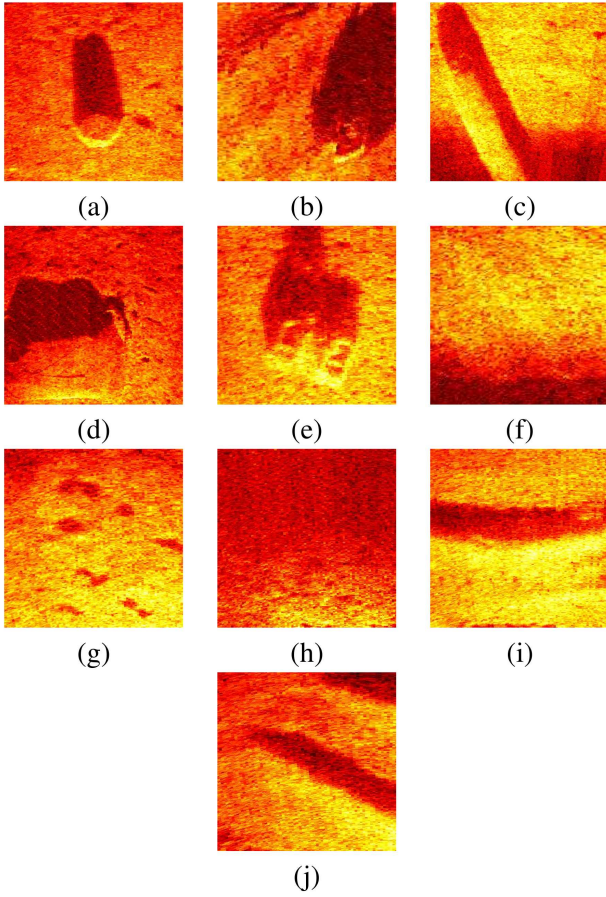
Fig. 7. Examples of frame clips. (a) and (b) truncated cone; (c) horizontal cylinder; (d) box with rounded corners; (e) two blocks linked; (f) to (h) show background clips; (i) and (j) sand ripples;



(a)



(b)

Fig. 8. 4-classes Multi-class classifier (a) accuracy and (b) loss over epochs for both training and validation data sets.

$0.5 \ \forall \ n \ \in$ grid. To optimize the update process, log-odds are used as in (3).

$$L(n|z_{1:t}) = L(n|z_{1:t-1}) + L(n|z_t)$$
$$L(n) = log \left[ \frac{P(n)}{1 - P(n)} \right] \quad (3)$$

A certain degree of adaptability in the grid is desired because not all the objects in the map are static (e.g., fish swimming within the FLS field of view) . To achieve this adaptability, a clamping update policy that defines an upper and lower bound on the cell estimate is applied using (4) per [42].

$$L(n|z_{1:t}) = \max(\min(L(n|z_{1:t-1}) + L(n|z_t), l_{max}), l_{min}), \quad (4)$$

with $l_{max} = 29.93$ and $l_{min} = -29.93$. Relatively high clamping values have been chosen as the probabilities obtained after applying (1) are always very close to 0 or 1.

The grid map is updated as the FLS frames are processed. At any moment (e.g., at the end of each survey track), the grid map can be checked to see if any cell has a probability of containing a target above a certain threshold. Only cells
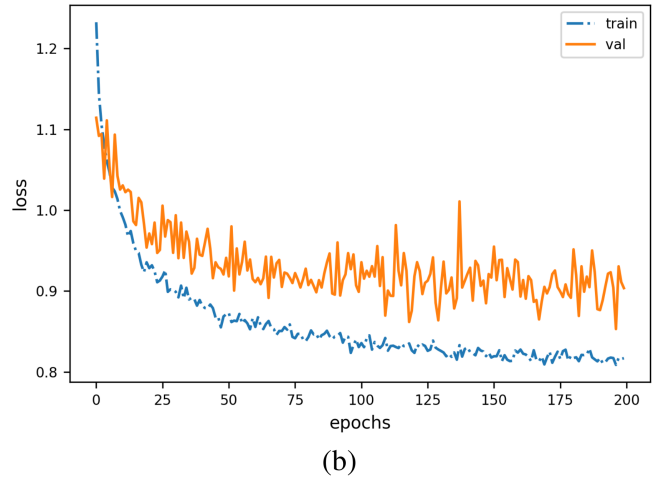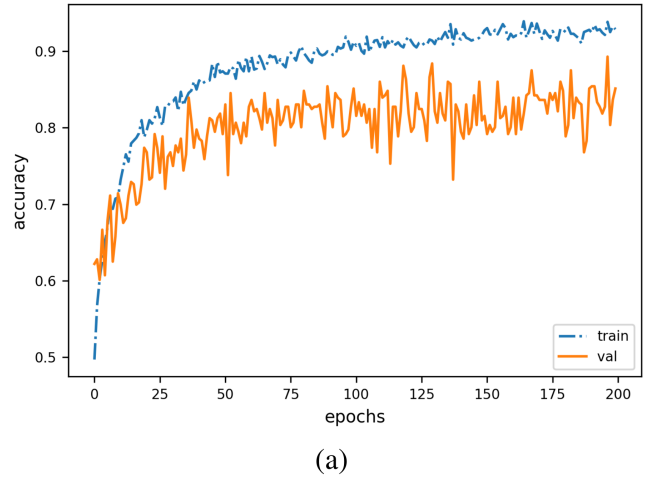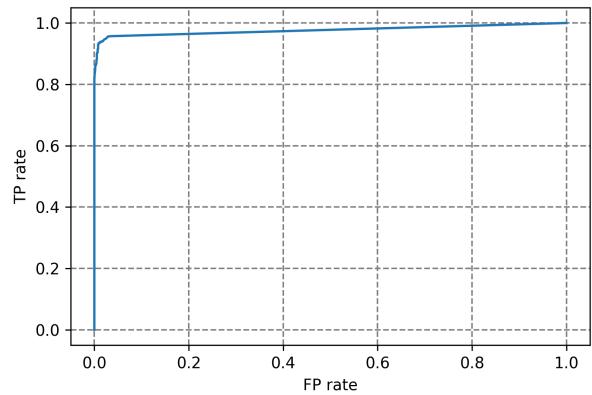


Fig. 9. ROC for the FLS_tiny model with an AUC of 0.9768.

accumulating a minimum number of detections are considered valid in order to ensure a consistent result. Since a single target can be detected in several nearby cells, a non-maximum suppression (NMS) filter is applied to avoid multiple detections for the same target.

Every time the grid is updated with a positive detection, both the frame from which the clip came and the estimated probability of belonging to a target class are stored in a database linked to the grid (see Fig. 1). When the ATR is interrogated, it can return any of the following elements: the resulting grid map, a list with the position of all detected targets, the averaged estimated class of each target and/or a set of images that triggered each detection. The size of this information represents a small fraction of the total size of the data gathered by the AUV during the survey. Depending on the number of images per target included it can range from some kilobyte to a few dozen megabytes. Such data can be transmitted using a low bandwidth channel like an acoustic modem. This allows for faster operations avoiding recovery of the vehicle or transmission the raw data which is usually in the order of some gigabytes.

### B. Target Reacquisition Strategy

The BIONDo AUV (see Fig. 10), in which the ATR framework will be first deployed, is a Sparus II class AUV built by Iqua Robotics [43] and modified by the CMRE to be used in MCM missions. The AUV has three thrusters (one vertical and two horizontal) that provide it with great maneuverability as well as hovering capability. It has a complete navigation sensor suite including a Doppler velocity logger (DVL) and a high-performance inertial navigation system (INS), several communication interfaces (i.e., WiFi, acoustic modem and radio) and an ARIS 3000 FLS [39] as its main perception sensor.

BIONDo uses the COLA2 software architecture [44] and is compatible with the distributed framework for embedded collaborative autonomy $D^2CAF$ [45]. Within $D^2CAF$, BIONDo can be considered as an autonomous agent with target reacquisition capabilities and therefore can be tasked by any other agent that requires this service. When a reacquisition task is created, if $D^2CAF$ allocates it to BIONDo, it receives the position (latitude, longitude), the area (width, height) and the orientation in which one or several MILCOs have been obtained. This zone should be close to BIONDo's current location and relatively small: BIONDo, and in particular the designated FLS, is not intended to survey large areas, but to revisit MILCOs already detected by other means. An area is given instead of a single point because it is necessary to survey around a given MILCO position in order to allow its re-localization. An error of several meters may exist between the MILCO given position and the one in which it is reacquired due to the differences that may exist between the navigation solutions of the different vehicles in the detection-reacquisition chain. Targets detected after a reacquisition mission can be reported back to $D^2CAF$.

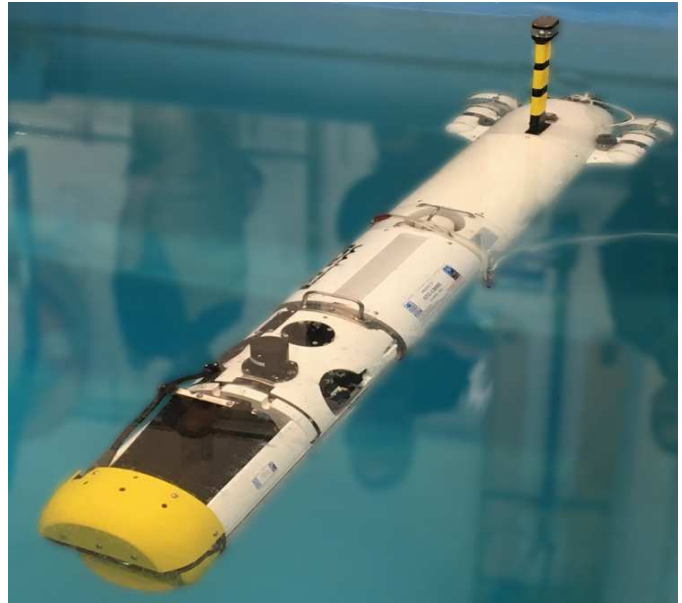The ARIS 3000 FLS installed in the BIONDo AUV has a maximum range of 15 m. However, to ensure the necessary



Fig. 10. BIONDo AUV, a Sparus II class vehicle prepared for MCM reacquision missions. It includes a high-performance INS and an ARIS 3000 FLS. BIONDo AUV is rated for depths up to 200 m.

resolution and the quality of the images taken, an average range of 7.5 m is used. With this configuration, if the vehicle navigates between 2 to 3 m over the bottom, the FLS frames start at around 3 m from the sonar and extends $\sim$4.5 m covering less than 3.5 m across track (see Fig. 2). Therefore, to ensure full coverage, the tracks of the survey mission have to be placed very close (i.e., between 2 and 3 m away). A standard lawnmower pattern (see Fig. 11(a)) can be used for this purpose. However, because the vehicle has to turn 180° in less than 3 m, resulting trajectories can be difficult to follow. To avoid unstable maneuvers, a sliding box trajectory (see Fig. 11(b)), where the AUV only has to execute 90° turns, can be used instead. While executing this survey patterns, if a target is reported by the ATR, a target-mapping maneuver can be triggered. This maneuvers consists on mapping a specific position (i.e., the position in which the target has been detected) from different points of view as seen in Fig. 11(c) and (d). It is important to highlight that these target-mapping maneuvers can only be executed if the position of the target to map is known with a high accuracy and the AUV navigation solution has not drifted much since the detection was made. A hovering capable vehicle like BIONDo AUV is well-suited for this kind of trajectory. Different target-mapping shapes can be defined. They range from a simple cross (see Fig. 11(c)) to a more complex back and forth double cross shape (see Fig. 11(d)). While the first will be faster to execute, the second will provide more viewpoints of the target and a larger area with multiple views.

Fig. 12 presents the architecture implemented to leverage the ATR results during mission execution. When the AUV is targeted by the $D^2CAF$ to reacquire a MILCO, it receives information about its approximate position. A sliding box survey mission is automatically generated around the general MILCO position. The vehicle navigates to the survey area on surface,
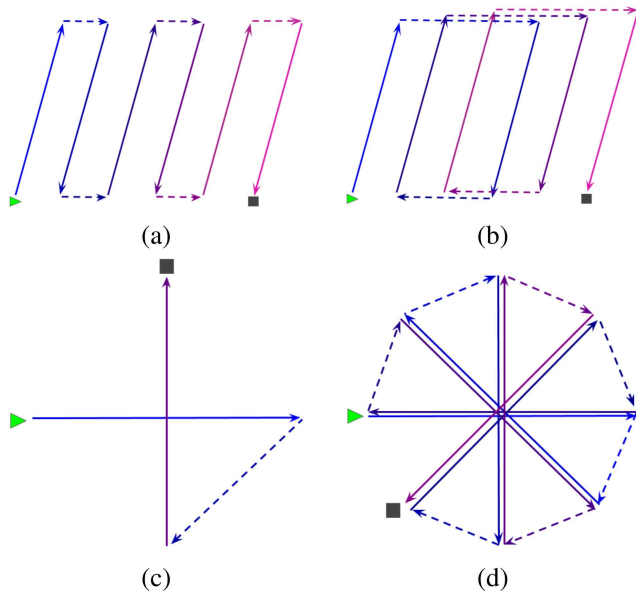
Fig. 11. Target reacquisition trajectories. (a) Lawnmower and (b) sliding box pattern are both useful for searching an area. (c) simple cross and (d) multi-view cross are useful for collecting data on a single position. Starting and ending position are indicated by a triangle and a square respectively.

to have access to GPS data, and once there it is submerged to reach the initial point of the survey mission. The ATR is then enabled and the survey starts. At the end of each survey track the ATR is interrogated. It returns the position of each new target candidate, and the survey mission is paused and a target-mapping maneuver is executed over each candidate position. If the navigation allows it, this interrogation can be done every 2 or 4 survey tracks instead. As soon as the survey is completed, the information calculated by the ATR (i.e., the position, class, and a set of images for each detected target) can be transmitted. The gigabytes of information collected by the vehicle during the mission are summarized in a few megabytes of data that can be sent even using a low-bandwidth communication channel.

## IV. RESULTS

In this section the ATR proposed before is tested using data sets collected during previous MCM exercises at sea. Both the detection and classification models as well as the benefits of using the introduced probabilistic grid map are evaluated. Next, an overall performance assessment of the reacquisition strategy is shown.

### A. ATR Results

Specific results for the detection and classification modules have already been reported in Table III and Table IV. According to these results, the FLS_small model behaves slightly better than FLS_tiny both for detection and classification tasks. However, FLS_small capacity is more than 4 times larger and, therefore, requires more computational power to execute. When running FLS_tiny in a computer with an i7 CPU and without GPU acceleration, it takes $\sim$4 ms to analyse a clip

while FLS_small takes almost twice the time (i.e., $\sim$7 ms). Considering that the ATR evaluation must be done for each clip in each frame (i.e., $\sim$50Hz), we propose to use the FLS_tiny architecture for the detection model, and use the FLS_small for the classification module activated only when a positive detection is triggered.

To analyze the ATR performance in the context of a real mission, data from previous campaigns were used. In particular, data gathered during the NATO exercises TJMEX 2015 in Cartagena (Spain) and ONMEX 2016 in Hyéres (France), as well as data gathered in St. Feliu de Guixols (Spain) in preparation for these two trials. In all of these trials, a Sparus II AUV equipped with an ARIS 3000 FLS executed a predefined survey trajectory over an area where a MILCO was previously detected. At the end of each inspection, the vehicle was recovered and the acquired data was manually processed. The AUV navigation suite used in all these exercises was based on a DVL and a low cost MEMS-based INS.

One of the goals of this paper is to show that with the proposed ATR it is possible to detect all the targets present in a surveyed area automatically online. It is worth noting that all the missions in which the vehicle flew over a target in the these campaigns have been used. Moreover, none of the data sets used to train and validate the detection model contain images from any of these 3 campaigns, which shows the system's generality.

The detector threshold was set at 0.999 to avoid false detections and the minimum number of hits per cell to be considered valid was established at 5. These values were used for all tests reported in this article. The clip size for each mission varied from 1.2 to 1.4 m, depending on the sonar parameters with which the experiment was performed. The size of the grid cells were always set to half the size of the clips.

Fig. 13 presents three missions performed during TJMEX campaign. Figures 13 (a), (c) and (e) show the vehicle's estimated trajectory in blue and the positive detections reported by the ATR detector as red stars. A dashed green circle has been added as a ground truth at the position where the target was manually identified.

Figures 13(b), (d) and (f) show the resulting grid maps for each run. White squares represent cells with a probability between 0 to 0.5 to contain a target while light pink to red squares have a probability between 0.5 to 1. We consider that a cell contains a target only if its probability is over 0.95. Moreover, only cells with a minimum number of updates are shown, indicating the actual sensor coverage.

In the first mission (see Figures 13(a) and (b)) the detector triggers 112 positive detections corresponding to 52 different frames. From these, only 2 were false positives and were filtered out by the grid map. Up to 5 cells were marked as containing a target but because they were all together, the NMS filter had no problems getting the ATR to report a single detection. The clip size was set at 1.25 m for this particular experiment.

In the second mission (see Figures 13(c) and (d)) the detector generated only one false positive. However, the target was seen twice during the survey and, due to the large drift of
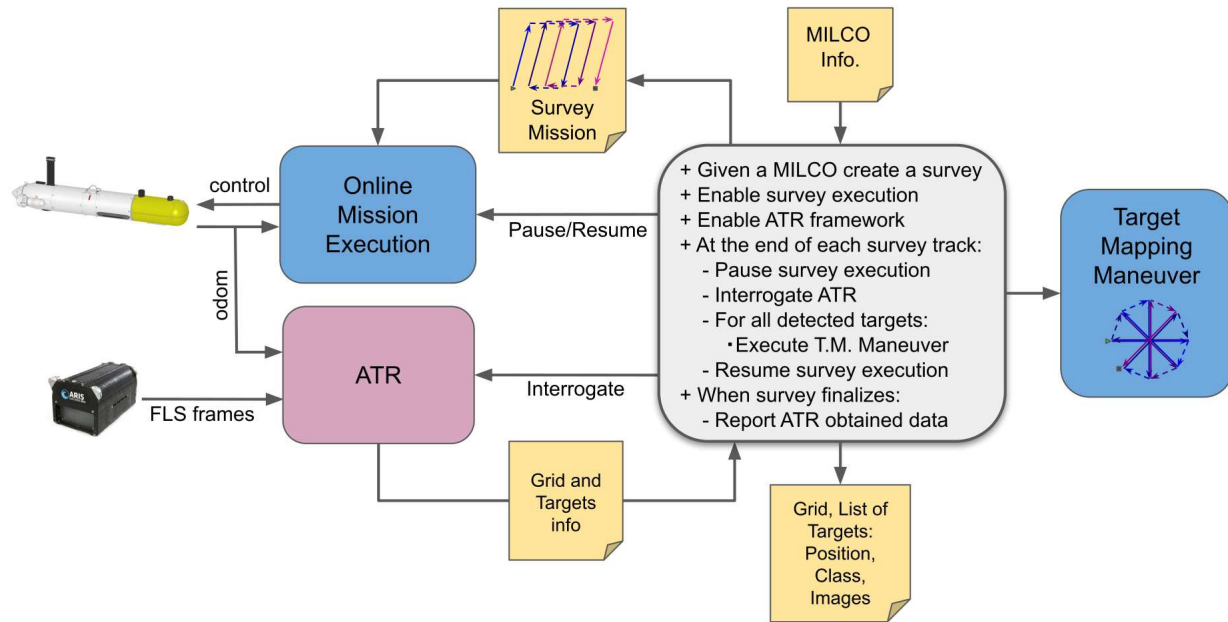
Fig. 12.  MCM target reacquisition architecture.

the AUV, it appears in two different locations (see Fig. 13(c)). The false detection disappeared after accumulating all cell updates but the two regions in which the target was detected also appeared in the grid map. In total, 282 detections were obtained scattered in an area of $3.5 \times 4$ m around the target and 6 cells were marked in the grid map with a probability over 0.95. Depending on the size of the NMS filter used, one (i.e., NMS $> 4.0$) or two detections (i.e., $1.5 <$ NMS $< 4.0$) were reported by the ATR. It is worth mentioning that the new navigation suite in the BIONDo AUV, which incorporates the same DVL but a high performance INS, should greatly mitigate this drift.

The third mission (see Fig. 13 (e) and (f)) was again over the first target. During the inspection, the vehicle crossed a school of fish near the southern edge of the trajectory area causing multiple detections (see Fig. 14(a)). However, these detections were not consistent over time and the probabilistic grid map filtered them out. At the end of the survey, only the real target position was reported by the ATR.

In the first and third experiments, the deployed target was completely different than the objects used to train the 4-classes classifier (see Fig. 14(c)). Therefore, the classifier was unable to identify the target among any of the classes learned. In the second run, the deployed target was a horizontal cylinder (see Fig. 14(b)). Despite in some frames it was misclassified, averaging all the classifier estimations it was correctly identified.

In the TJMEX trials it was observed that the AUV had some difficulties following the survey trajectory. For this reason, in ONMEX the lawnmower pattern was replaced by a sliding box pattern. Fig. 15 presents two missions performed during ONMEX. In the trials area, the soil was smooth but with rocks of different sizes scattered around. Most of these rocks were small enough that they did not trigger the detection module

systematically. Only one of these contacts (see Fig. 14(d)) was detected consistently enough to appear on the grid map (see detection in Fig. 15(a) around 9 m South and 16 m West). The only target present in the area was also detected in the West edge of the surveyed trajectory. Due to many frames in which only half of the target appears, it was misclassified as a box with rounded corners instead of a horizontal cylinder (see Fig. 14(e)). The mission was repeated but changing the center of the survey area for the position in which the target was located in the previous survey. Fig. 15(c) shows the result of this second run. Up to 7 areas triggered the target detection model but only the one located around 13 m North and 22 m West, corresponding to the mine-like target, was persistent enough to not be filtered by the probabilistic grid. All other detections, corresponding to small- or medium-sized clutter, were filtered out. In this second run the target was correctly classified as a horizontal cylinder. The second survey was aborted before completion due to a sonar malfunction. It is worth noting that, despite using the same navigation suite than in TJMEX, vehicle drift was much lower during these trials.

The last environment was more challenging because the bottom had several areas covered by sand ripples. To filter them, we used the classification model able to differentiate between targets and sand ripples. In the first trial a lawnmower pattern was used. By chance, the path of the vehicle was almost parallel to the sand ripples. With the classifier disabled, the detector found 463 possible detections (see Fig. 16(a)). The grid map was able to group all these detections into only 3 candidates, one of which was the desired target (see Fig. 16(b)). Enabling the sand-ripple classifier, only 83 of these detections remained (see Fig. 16(c)) and the grid map had no problems to filter out the scattered false positives returning only a target candidate corresponding to the real mine-like object (see Fig. 16(d)).
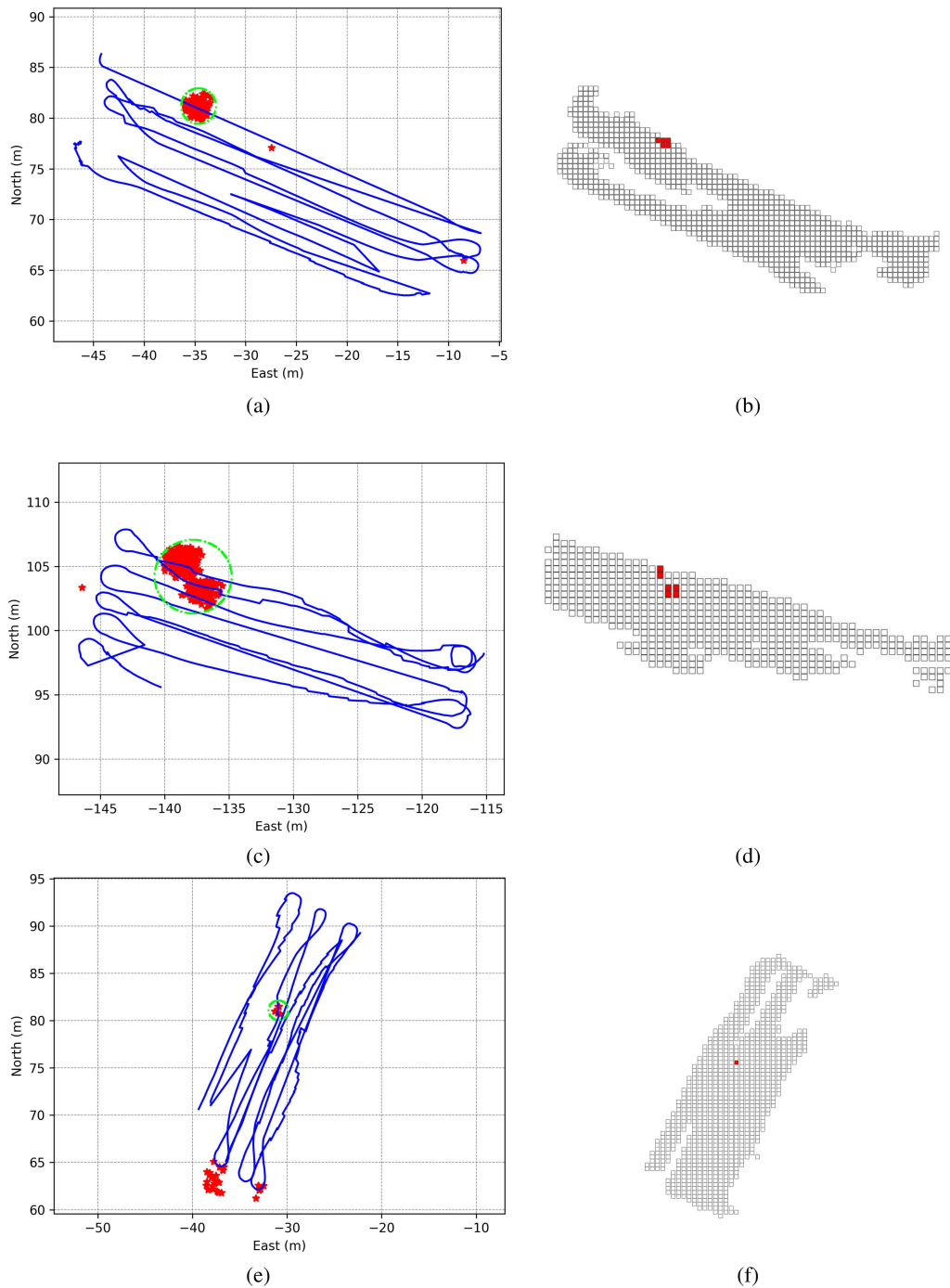
(a)



(b)



(c)



(d)



(e)



(f)

Fig. 13. TJMEX MCM reacquisition trials. Figures (a), (c) and (e) show the vehicle's trajectory in blue and the detections reported by the detection module as red stars. A dashed green circle has been added as a ground truth at the position where the target was manually identified. Figures (b), (d) and (f) show the resulting grid map. White squares represent cells with a probability between 0 to 0.5 to contain a target while light pink to red squares have a probability between 0.5 to 1. Only squares in which at least 5 updates have been accumulated are shown.
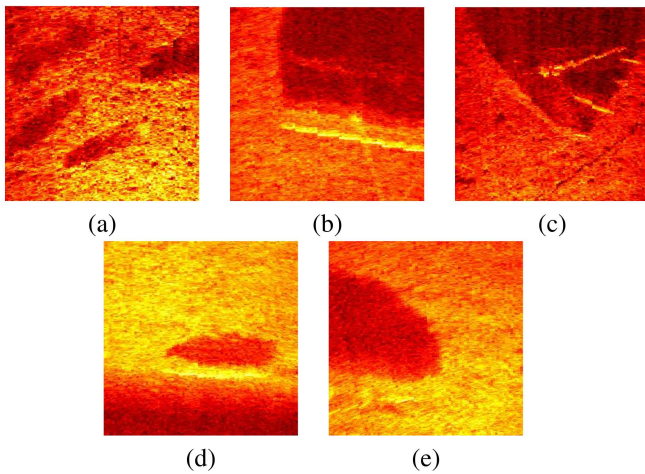
Fig. 14. Examples of frame clips gathered during TJMEX and ONMEX trials. Clip (a) shows one of the frames mapping a shoal of fish that caused several false positives. Clips (b) and (c) show the deployed targets in TJMEX. Clip (d) corresponds to a stone present in ONMEX survey area that triggered systematic false positives. Clip (e) contains part of a horizontal cylinder target deployed in ONMEX that was misclassified as a box with rounded corners by the system.

The same area was inspected again using a sliding box pattern and centering the target better. The same orientation was used, however, because a different survey pattern was used, some tracks mapped the sand ripples perpendicularly, producing even more false detections than in the previous execution. Without using the sand ripples classifier, the detector generated 2739 positive detections (see Fig. 16(e)) and the grid map system returned 5 candidate positions from which one was the real target (see Fig. 16(f)). Enabling the sand-ripple classifier, only 387 detections were triggered (see Fig. 16(g)) and, in the grid map, only the cell containing the target has a probability over 95% (see Fig. 16(h)). Executing the 4-class classifier in chain after the target-clutter classifier the deployed target was correctly identified in both trials.

With these two experiments it can be seen that the classifier is very effective in filtering out a particular type of false positives. In this particular scenario, about 90% of the detections produced by sand ripples were eliminated, while less than 10% of the detections produced by the real target were marked as clutter.

### B. Target Reacquisition Strategy Results

The previous section has shown the ability of the proposed ATR to detect and classify targets in FLS data sets collected during previous MCM exercises in different locations and conditions. Next, the performance of the proposed target reacquisition strategy has been evaluated based on 3 factors: the coverage, the navigation drift and the detection accuracy.

It is obvious that the system can only detect the presence of targets in regions that have been covered by the sonar. According to the sonar geometry, at 7.5 m range the sonar should cover around 3.5 m across track. However, several factors reduce the effective coverage. Some of these are: square clips cause part of each frame to be discarded; some frames

are discarded due to the odometry filter; and small variations in the AUV heading cause a considerable displacement in the FLS footprint because the sonar is always mapping several meters in front of its own position. In the previous trials it was seen that the AUV had difficulties following a lawnmower pattern where the tracks were very close. However, the sliding box trajectory can also result in unexplored gaps as shown in Fig. 16. To evaluate the coverage, taking into account the trajectory pattern and the separation between tracks, several simulations were conducted. In these simulations, the dynamic model of the Sparus II AUV [46], and the same trajectory controllers used in the actual vehicle, were used. A grid map with 0.5 $m^2$ cells was set in all the simulations and a cell was considered visited if at least 5 updates were accumulated in it. Fig. 17 shows the result of executing a lawnmower and a sliding box pattern, at 3.5 m and 2 m distance between tracks. The figure shows that the sliding box pattern gives, on average, a 10% better coverage than the lawnmower. It can also be seen that, to ensure a coverage of more than 90%, the tracks have to be at around 2 m distance with this particular sonar configuration.

The coverage analysis was done assuming an almost perfect localization. However, it was shown in the previous subsection that the navigation drift is one of the major problems to tackle. If the vehicle has a large drift, it is difficult to ensure coverage as well as to associate detections. To improve this, the MEMS-based INS installed in standard Sparus II AUVs has been replaced in the BIONDo AUV by a high performance INS that allows to reduce the localization drift over time.

To assess the performance of the target reacquisition strategy, several mission were simulated over an area with 2 targets, placed at (7.5, -5) and (-1, 1). Regarding the ATR simulation, since generating realistic FLS images synthetically is complex and beyond the scope of this work, instead of generating sonar clips and running them through the detector module, the ATR performance was emulated, bypassing the simulation of the sonar data at the sensor level. To set the detection probability of the detector module, the values obtained in real tests were used. According to the results in Section IV-A, the accuracy of the ATR is between 95% and 99.99% when dealing with data from previous MCM campaigns. When dealing with the data set that was created to validate the different CNN models, the accuracy decreases to ∼90%. It is worth noting that this validation dataset contains more extreme cases than what is typically found in the real missions. In the simulations, the ATR accuracy was set first to 99.9% and then to 90% and the localization drift was tuned to resemble both the high-accuracy INS first and the MEMS-based INS later.

Fig. 18 shows the obtained results when the accuracy of the ATR was set at 99.9% and the navigational drift commensurate with that of the high-grade INS available in the BIONDo. The two existing targets were discovered and only a simple target-mapping maneuver was performed on each of them. The few false positives produced by the ATR were filtered out by the probabilistic grid system. When the ATR accuracy was reduced to 90% (see Fig. 19) hundreds of false positives were reported by the ATR. These resulted in 5 target-mapping maneuvers being triggered. The additional data provided by
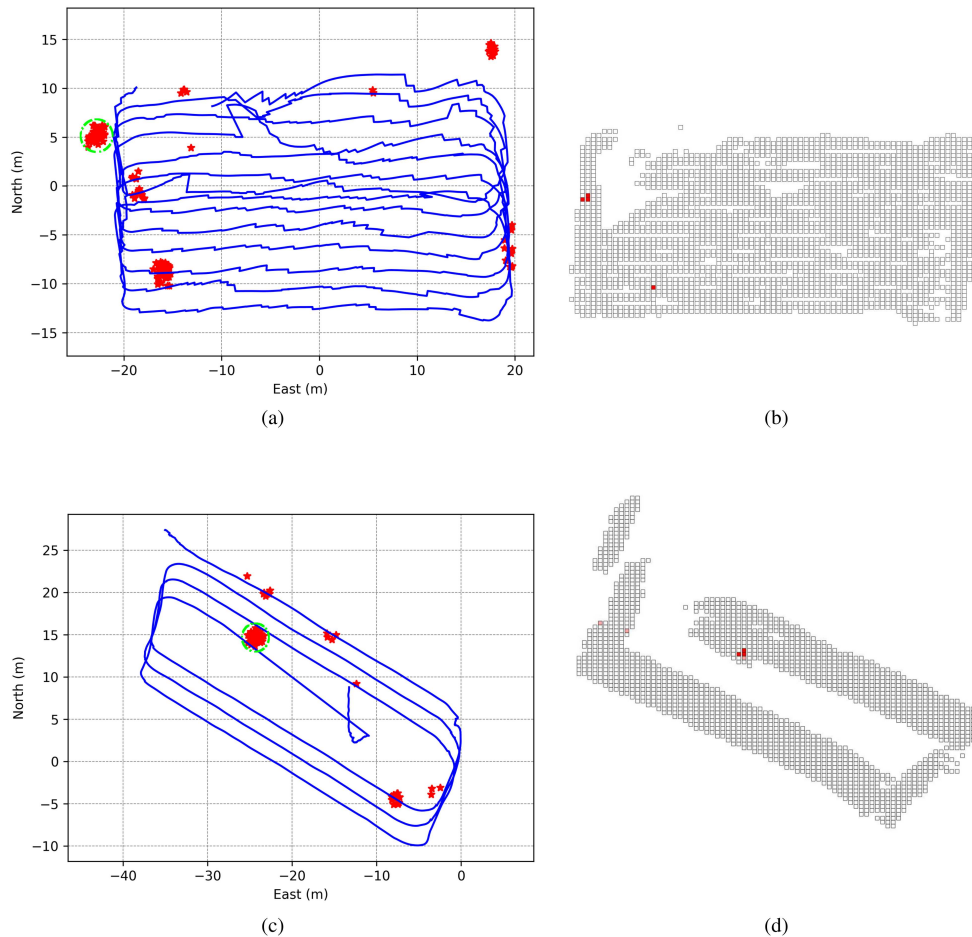
(a)



(b)



(c)



(d)

Fig. 15. ONMEX MCM reacquisition trials. Figures (a) and (c) show the AUV trajectory and the detection output while figures (b) and (d) present the resulting grid map.

these supplementary tracks allowed the probabilistic grid to filter out the false positives and only the position of the two targets was returned at the end of the mission.

When the same analysis was done using a less precise localization system, the main difference observed was that the coverage of the sliding box was significantly reduced. In our tests, this reduction went from 94% to 79.8% considering only the area covered by the sliding box pattern and not the additional tracks added by the target-mapping maneuvers. Though the two targets are correctly detected, as shown in Fig. 20 and Fig. 21, in a third of the simulations executed, one of the two targets was not detected because the area in which it was located was not covered. Of course, these results depend on how poorly the vehicle is localized. In Fig. 21, the lower accuracy of the ATR was not a major issue for the probabilistic grid system which marked only 2 cells with a probability greater than 0.95 of containing a target. Using the additional information provided by the target-mapping maneuvers, the system filtered out the false detections and kept only the valid target positions. Because the mapping maneuver was executed at the end of each survey track, the relative drift was small and the detection positions thus didn't change significantly.

## V. DISCUSSION

Results reported in Section III and IV have assessed the target reacquisition strategy and demonstrated the accuracy of the proposed ATR.

The ATR module has demonstrated an accuracy over 95% when dealing with real data from missions never encountered during its training phase. The classifier model has shown a lower accuracy (i.e., ~80% with 4 classes and ~90% with only two classes), but was overall sufficient to correctly classify most of the targets or to filter out many clutter elements. The probabilistic grid map system has proven to be a key module within the ATR. Most of the false positives generated by the detector were due to rocks, swimming fish, noise in the sonar data or sand ripples. Detections produced by sonar noise or fish occur sporadically and are spatially inconsistent. Therefore, when more data from the same region is observed, those detections are filtered out. The same applies to small clutter, mainly rocks, on the background: if they are observed several times or from different points of view, most of them disappear from the probabilistic grid map because their detection is not consistent. It is normal for medium to large objects, including rocks, to be detected as objects of interest by the detector module. Differentiating which of these objects should

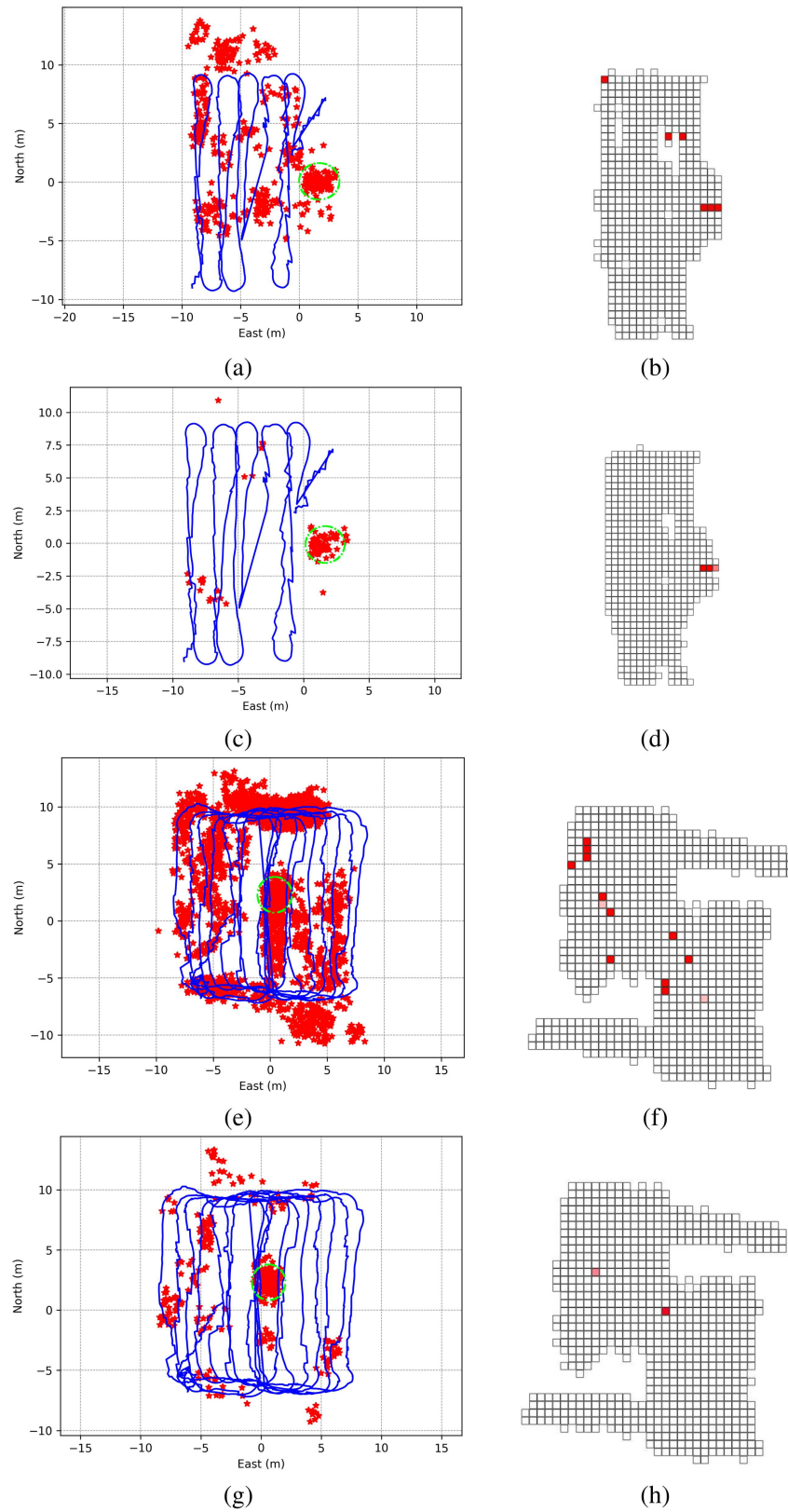(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Fig. 16. St. Feliu de Guixols shoreline and first reacquisition trials. Figures (a) and (c) show the same AUV trajectory and the target detection output while figures (b) and (d) present the resulting grid map. Figures (e) and (f) show the AUV same trajectory and the target detection output while figures (g) and (h) present the resulting grid map. The difference between (a) and (b) with respect to (c) and (d) and (e) and (f) woth respect to (g) and (h) is that the target-clutter classifier was disabled in the former pairs and enabled in the latter.
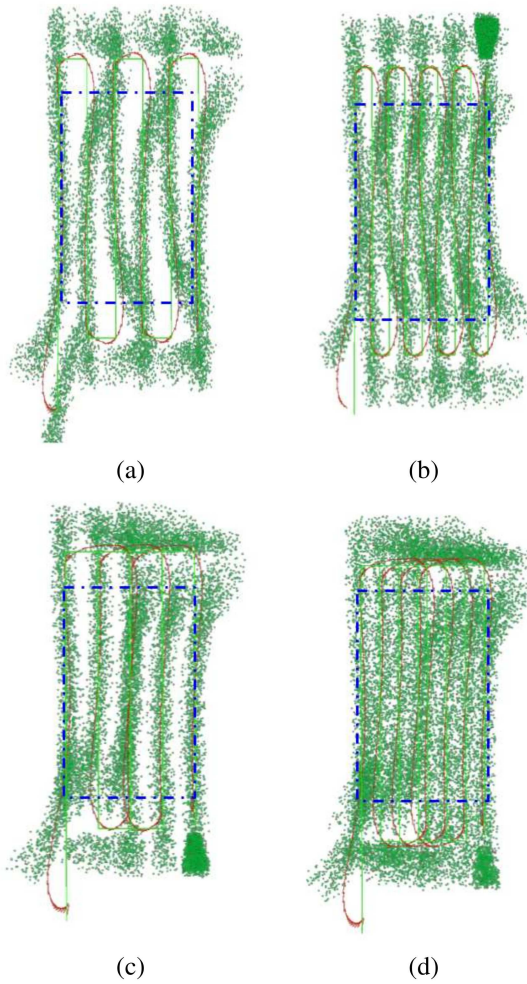
(a)



(b)



(c)



(d)

Fig. 17. Coverage comparison executing a lawnmower and a sliding box pattern in a $30 \times 15$ m area at a constant altitude of 2 m and a velocity of 0.5m/s with tracks at different separation. The green dots represent clip centers. (a) using a lawnmower with tracks at 3.5 m the coverage is 62.8%, (b) with a lawnmower with tracks at 2 m, the coverage is 84.5% (c) using a sliding box pattern with tracks at 3.5 m the coverage is 69.4%, and (d) with a sliding box survey with tracks at 2 m, the coverage reaches 94%. Reducing the tracking space from 3.5 m to 2 m, the execution time is increased a $\sim$34%.

be considered a target or clutter is the responsibility of the classifier.

Sand ripples have also proven to be very problematic. Unlike rocks or sporadic elements that cause detections at a particular position or moment, sand ripples may trigger consistent detections over a large area. In this situation it was key to have a classifier trained to differentiate between objects of interest and specific clutter elements. During the experiments, we also observed that the best way to deal with sand ripples was to orient the survey pattern parallel to them and avoid imaging the ripples perpendicular to the sonar.

In summary, the ATR was able to estimate the position of the deployed target in 7 out of 7 target reacquisition missions. Only in one, a medium-sized rock was also identified as a possible target. For the missions carried out in an area with sand ripples it was necessary to use a specific classifier, able to filter out the ripples, to obtain only one target. Otherwise,
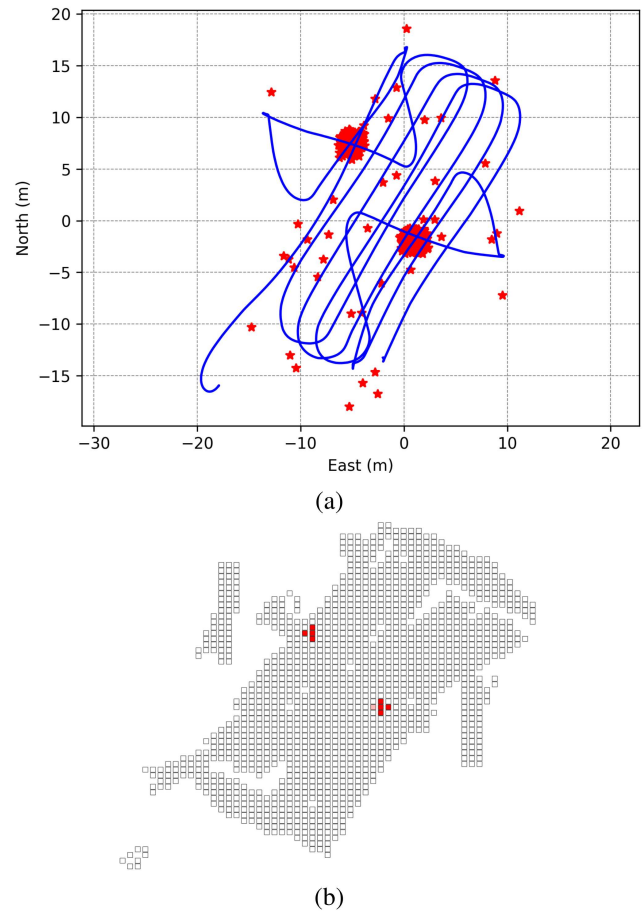


(a)



(b)

Fig. 18. Simulation of a reacquisition MCM mission using the proposed strategy. A high-precision navigation system and an ATR with an accuracy of 99.9% are simulated.

the ATR reported up to four more false candidates. However, even without using this specific classifier, the position of the deployed target was correctly estimated. The classifier was also used to identify the targets in 5 trials. In 4 of them, the target was correctly identified, while in one it was misclassified, mainly because only partial images of it were obtained.

Using simulation tools the ATR coverage was also evaluated under different conditions. The sliding box pattern was shown to provide slightly more coverage than the lawnmower pattern (i.e., around a 10%).

The proposed strategy, that combines a survey with automatically triggered target-mapping maneuvers, was also tested. Several conclusions can be drawn from the tests carried out. Concerning the target-mapping maneuvers at the end of each survey track, they allow the system to eliminate most of the false positives introduced by the ATR. Furthermore, these additional maneuvers provide multiple views for each target, which gives added confidence to the results.

Finally, it has been observed that it is essential to reduce the vehicle drift. Otherwise, it is not possible to ensure complete coverage of the survey area, and it is more difficult to associate detections to targets.
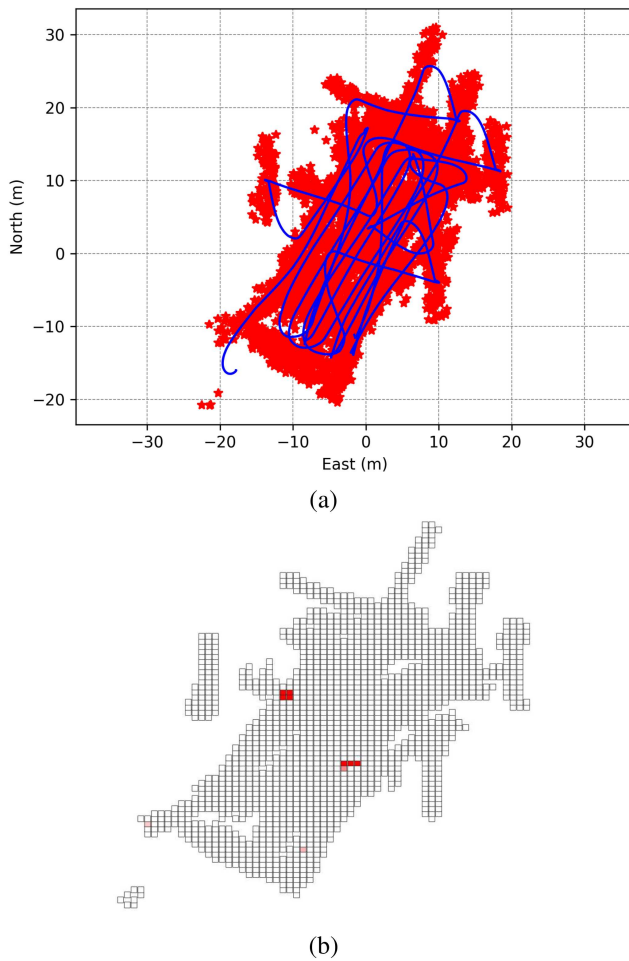
(a)



(b)

Fig. 19. Simulation of a reacquisition MCM mission using the proposed strategy. A high-precision navigation system and an ATR with an accuracy of 90% are simulated.
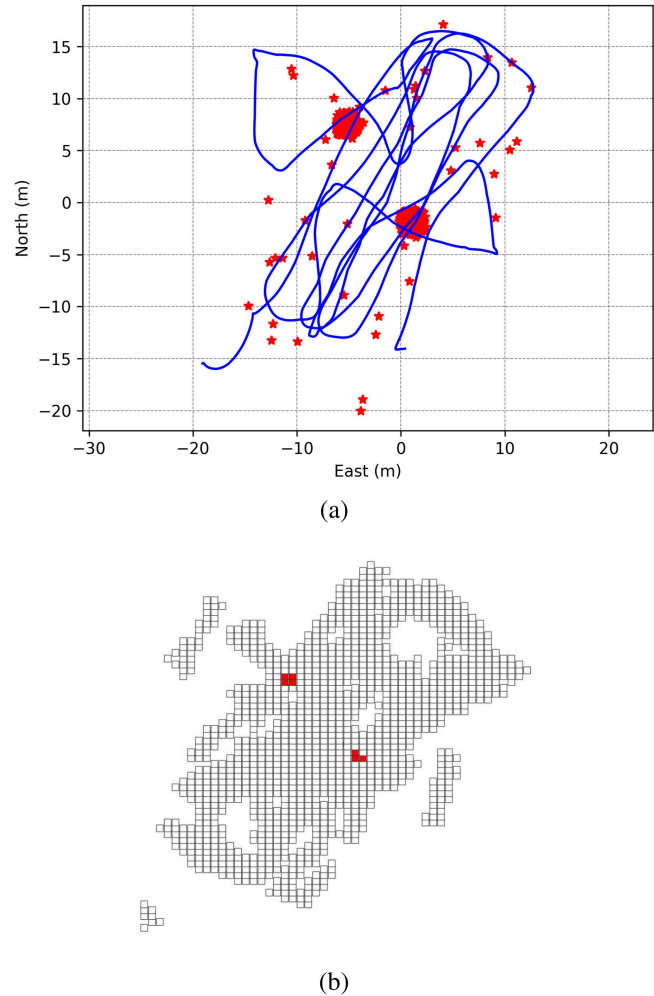


(a)



(b)

Fig. 20. Simulation of a reacquisition MCM mission using the proposed strategy. A MEMS-based INS navigation system and an ATR with an accuracy of 99.9% are simulated.

## VI. CONCLUSIONS

The purpose of this article was twofold. Firstly, an ATR for forward looking sonar, designed to run online on board an AUV, and capable of detecting and classifying mine-like targets, was developed. Secondly, a strategy was defined to exploit the performance of this ATR, as well as to obtain multiple views of each of the detected targets.

To develop the ATR, several models based on CNNs, already available in the literature, were tested using FLS data. Insights from those previous works led to the design of two new models, with improved performance when dealing with this particular sonar data. These models can be used to detect objects of interest in FLS imagery but they can be also retrained, using transfer learning, as multi-class classifiers to identify specific targets or clutter elements. Next, a probabilistic mechanism based on a grid map was implemented, acting as a spatial filter, to improve the overall ATR results. This system is capable of filtering out inconsistent detections as well as combining different detections belonging to the same target. Although CNNs and probabilistic grids are well-known techniques, the main novelty of this work lies in the unique way in which they have been combined and adapted within the

presented framework. The proposed ATR has been tested with data acquired in previous MCM missions. It has been shown to be able to re-locate all targets in a given area producing very few, if any, false positives when thresholds are set aggressively.

It is well understood that there is a trade off between false alarm rate and the aggressiveness with which low-scoring contacts are excluded. Operationally, in the context of defence and especially MCM related domains, this has important implications regarding acceptable risk. The best mitigation is to improve classifier performance while decreasing the exclusion mechanisms (such as the threshold, in our case).

The envisioned architecture, combining autonomous reacquisition trigger with ATR outputs, has also been tested through multiple simulations.

The whole system may run in real-time on board a vehicle and is now being deployed in the BIONDo AUV to be tested in future MCM campaigns. As new data from these campaigns becomes available, both the detector and the classifier models can be retrained to improve their performance. We believe that the classifier is the module that might benefit the most from this additional data, especially to differentiate
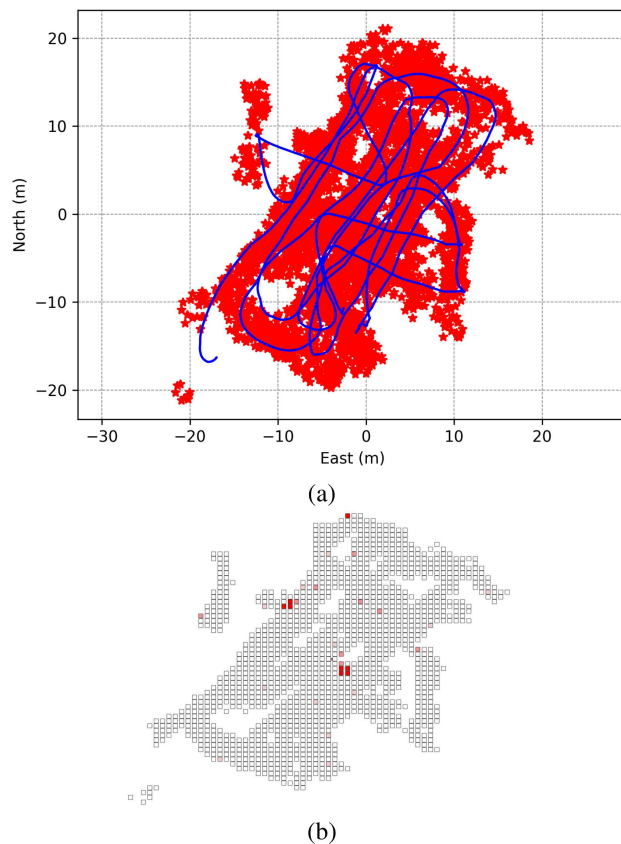
(a)



(b)

Fig. 21. Simulation of a reacquisition MCM mission using the proposed strategy. A MEMS-based INS navigation system and an ATR with an accuracy of 90% are simulated.

between clutter classes (e.g., rocks, seaweed, small objects) and target classes. This would reduce the number of false positives making the system even more robust. The framework presented is modular enough to be extended. For example, a simultaneous localization and mapping module, that uses the detections as landmarks, could be added to improve the overall navigation. Likewise, a terrain classification module linked to a live trajectory optimization system could be included to select the classification models to be used as well as to optimize the shape and orientation of the target reacquisition trajectories to be executed. A multi-aspect classifier could even be used to perform target-mapping maneuvers in a smarter way. Moreover, the new INS installed in the BIONDo AUV should reduce the navigation drift significantly, producing more consistent coverage.

## ACKNOWLEDGMENT

## REFERENCES

[1] M. Schiffer, J. Bedi, A. Daemmrich, and A. Molella, *Power Struggles: Scientific Authority and the Creation of Practical Electricity Before Edison*, ser. Lemelson Center Studies in Invention and Innovation. MIT Press, 2008. [Online]. Available: https://books.google.it/books?id=8yCle7H0Q_QC

[2] RussianNavy.com, "Russian navy weapons," http://milit.ru/mines.htm, 2020, accessed: 2020-03-09.

[3] MilitaryFactory.com, "Mine warfare fleet strength by country," https://www.globalfirepower.com/navy-mine-warfare-craft.asp, 2020, accessed: 2020-03-09.

[4] N. R. Council, *Oceanography and Mine Warfare*. NATIONAL ACADEMY PRESS Washington, 2000.

[5] T. Garrold, *Mine Warfare Introduction*. Officers School Command, U.S. Navy, 1998, ch. The Threat. Surface Warfare, pp. 17–81.

[6] D. P. Williams, "Fast target detection in synthetic aperture sonar imagery: A new algorithm and large-scale performance analysis," *IEEE Journal of Oceanic Engineering*, vol. 40, no. 1, pp. 71–92, Jan 2015.

[7] N. Hurtós, D. Ribas, X. Cufí, Y. Petillot, and J. Salvi, "Fourier-based registration for robust forward-looking sonar mosaicing in low-visibility underwater environments," *Journal of Field Robotics*, vol. 32, no. 1, pp. 123–151, 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21516

[8] Y. Petillot, I. Tena Ruiz, and D. M. Lane, "Underwater vehicle obstacle avoidance and path planning using a multi-beam forward looking sonar," *IEEE Journal of Oceanic Engineering*, vol. 26, no. 2, pp. 240–251, April 2001.

[9] M. Valdenegro-Toro, "End-to-end object detection and recognition in forward-looking sonar images with convolutional neural networks," in *2016 IEEE/OES Autonomous Underwater Vehicles (AUV)*, Nov 2016, pp. 144–150.

[10] Z. Zhao, P. Zheng, S. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, Nov 2019.

[11] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, Dec 2001, pp. I–I.

[12] J. Sawas, Y. Petillot, and Y. Pailhas, "Cascade of boosted classifiers for rapid detection of underwater objects," in *Proceedings of the 10th European Conference on Underwater Acoustics, ECUA 2010 : Istanbul, Turkey*, vol. 3, 7 2010, pp. 1507–1516.

[13] J. Sawas and Y. Petillot, "Cascade of boosted classifiers for automatic target recognition in synthetic aperture sonar imagery," *Proceedings of Meetings on Acoustics*, vol. 17, no. 1, p. 070074, 2012. [Online]. Available: https://asa.scitation.org/doi/abs/10.1121/1.4788639

[14] E. Galceran, V. Djapic, M. Carreras, and D. P. Williams, "A real-time underwater object detection algorithm for multi-beam forward looking sonar," *IFAC Proceedings Volumes*, vol. 45, no. 5, pp. 306 – 311, 2012, 3rd IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474667016306206

[15] F. Ferreira, V. Djapic, M. Micheli, and M. Caccia, "Forward looking sonar mosaicing for mine countermeasures," *Annual Reviews in Control*, vol. 40, pp. 212 – 226, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S136757881500053X

[16] J. Zhu, Y. Siquan, L. Gao, Z. Han, and Y. Tang, "Saliency-based diver target detection and localization method," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–14, 02 2020.

[17] N. Hurtós, N. Palomeras, A. Carrera, and M. Carreras, "Autonomous detection, following and mapping of an underwater chain using sonar," *Ocean Engineering*, vol. 130, pp. 336 – 350, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0029801816305765

[18] V. Myers and J. Fawcett, "A template matching procedure for automatic target recognition in synthetic aperture sonar imagery," *IEEE Signal Processing Letters*, vol. 17, no. 7, pp. 683–686, July 2010.

[19] M. Dos Santos, P. O. Ribeiro, P. Núñez, P. Drews-Jr, and S. Botelho, "Object classification in semi structured enviroment using forward-looking sonar," *Sensors*, vol. 17, no. 10, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/10/2235

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 436, no. 7553, pp. 436–444, May 2015.

[21] J. Kim and S. Yu, "Convolutional neural network-based real-time rov detection using forward-looking sonar image," in *2016 IEEE/OES Autonomous Underwater Vehicles (AUV)*, Nov 2016, pp. 396–400.

[22] Y. Le Cun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Proceedings of the 2nd International Conference on Neural Information Processing Systems*, ser. NIPS'89. Cambridge, MA, USA: MIT Press, 1989, p. 396–404.

[23] D. P. Williams and S. Dugelay, "Multi-view SAS image classification using deep learning," in *OCEANS 2016 MTS/IEEE Monterey*, Sep. 2016, pp. 1–9.

[24] D. P. Williams, "On the use of tiny convolutional neural networks for human-expert-level classification performance in sonar imagery," *IEEE Journal of Oceanic Engineering*, pp. 1–25, 2020.

[25] M. Valdenegro-Toro, "Real-time convolutional networks for sonar image classification in low-power embedded systems," *ESANN 2017 - Proceedings, 25th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pp. 577–582, 2017.

[26] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[27] P. Zhu, J. Isaacs, B. Fu, and S. Ferrari, "Deep learning feature extraction for target recognition and classification in underwater sonar images," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, Dec 2017, pp. 2724–2731.

[28] N. Warakagoda and O. Midtgaard, "Transfer-learning with deep neural networks for mine recognition in sonar images," in *4th International Conference on Synthetic Aperture Sonar and Synthetic Aperture Radar*, vol. 40, 2018, pp. 115–112.

[29] J. Chen, S. J., and T. J., "Interpretable semi-supervised deep learning with synthetic aperture sonar for automatic target recognition," in *4th International Conference on Synthetic Aperture Sonar and Synthetic Aperture Radar*, vol. 40, 2018, pp. 132–139.

[30] J. Bell and L. Linnett, "Simulation and analysis of synthetic sidescan sonar images," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 144, no. 4, pp. 219–226, 1997.

[31] D. GUERIOT, C. Sintes, and R. Garello, "Sonar data simulation based on tube tracing," in *OCEANS 2007 Europe*, Aberdeen, United Kingdom, Jun. 2007, pp. 1 – 6. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02165212

[32] K. J. DeMarco, M. E. West, and A. M. Howard, "A computationally-efficient 2d imaging sonar model for underwater robotics simulations in gazebo," in *OCEANS 2015 - MTS/IEEE Washington*, 2015, pp. 1–7.

[33] A. Makhzani, J. Shlens, N. Jaitly, and I. Goodfellow, "Adversarial autoencoders," in *International Conference on Learning Representations*, 2016. [Online]. Available: http://arxiv.org/abs/1511.05644

[34] A. I. Karjalainen, R. Mitchell, and J. Vazquez, "Training and validation of automatic target recognition systems using generative adversarial networks," *2019 Sensor Signal Processing for Defence Conference, SSPD 2019*, 2019.

[35] M. Jegorova, A. I. Karjalainen, J. Vazquez, and T. Hospedales, "Full-Scale Continuous Synthetic Sonar Data Generation with Markov Conditional Generative Adversarial Networks," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 3168–3174, 2020.

[36] D. Liu, Y. Wang, Y. Ji, H. Tsuchiya, A. Yamashita, and H. Asama, "Cyclegan-based realistic image dataset generation for forward-looking sonar," *Advanced Robotics*, vol. 35, no. 3-4, pp. 242–254, 2021. [Online]. Available: https://doi.org/10.1080/01691864.2021.1873845

[37] G. E. Packard, T. Austin, R. Christenson, J. Kaeli, R. Littlefield, M. Purcell, and R. Stokey, "Servo controlled rapid mcm target reacquisition using remus-100," in *2014 Oceans - St. John's*, Sep. 2014, pp. 1–3.

[38] F. Ferreira, V. Djapic, M. Micheli, and M. Caccia, "Improving automatic target recognition with forward looking sonar mosaics," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 3382–3387, 2014, 19th IFAC World Congress. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474667016421282

[39] soundmetrics.com, "Soundmetrics," http://www.soundmetrics.com/, 2020, accessed: 2020-03-16.

[40] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ser. ICML'15. JMLR.org, 2015, p. 448–456.

[41] H. Moravec and A. Elfes, "High resolution maps from wide angle sonar," in *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, vol. 2, March 1985, pp. 116–121.

[42] M. Yguel, O. Aycard, and C. Laugier, "Update policy of dense maps: Efficient algorithms and sparse representation," *Springer Tracts in Advanced Robotics*, vol. 42, 07 2007.

[43] iquarobotics.com, "Iqua robotics," http://iquarobotics.com/, 2020, accessed: 2020-03-16.

[44] N. Palomeras, A. El-Fakdi, M. Carreras, and P. Ridao, "COLA2: A control architecture for AUVs," *IEEE Journal of Oceanic Engineering*, vol. 37, no. 4, pp. 695–716, 2012.

[45] T. C. Furfaro, "A distributed framework for embedded collaborative autonomy," *OCEANS 2018 MTS/IEEE Charleston*, pp. 1–6, 2018.

[46] M. Carreras, J. D. Hernández, E. Vidal, N. Palomeras, D. Ribas, and P. Ridao, "Sparus II AUV — A hovering vehicle for seabed inspection," *IEEE Journal of Oceanic Engineering*, vol. 43, no. 2, pp. 344–355, 2018.

**Narcis Palomeras** (MSc 2004, PhD 2011) is a postdoctoral researcher at the University of Girona (UdG), and Member of the Computer Vision and Robotics Group (VICOROB). He has participated in several research projects, all related to underwater robotics, both national and European (TRIDENT, PANDORA, MORPH, MERBOTS, LOONDOCK, TWINBOT, 3DAUV, ATLANTIS...) as well as in different European competitions for AUVs such as SAUC-E and ERL. His research activity focuses mainly on underwater robotics in topics like planning, exploration, intelligent control architectures, mission control and localization. Currently is the coordinator of a Joint Degree Erasmus Mundus about Intelligent Field Robotic Systems.



**Thomas Furfaro** received his Bachelor and Master's of Ocean Engineering degrees from Florida Atlantic University in 2008 and 2012. In 2012 he joined the NATO Undersea Research Center to work on the EU FP7 project MORPH, on which he worked on relative navigation solutions using acoustic networks. Afterwards, at which point NURC was renamed the NATO STO Centre for Maritime Research and Experimentation, he worked on the implementation of software defined communications architectures for use in the maritime domain. Since 2016, he has been working on the topic of autonomy for naval mine countermeasures, with a focus on software architectures, interoperability, collaborative autonomy and autonomous behaviours.



**David P. Williams** received the B.S.E. (magna cum laude), M.S., and Ph.D. degrees in electrical and computer engineering from Duke University, Durham, NC, USA, in 2002, 2003, and 2006, respectively. He was the recipient of a James B. Duke Graduate Fellowship and a National Defense Science and Engineering Graduate Fellowship. He has been recognized with the Excellence in Review by the IEEE Journal of Oceanic Engineering twice, in 2013-2014 and 2017. From 2007 to 2021, he was with the NATO Science and Technology Organization, Centre for Maritime Research and Experimentation (formerly NATO Undersea Research Centre), in La Spezia, Italy. His research interests are in the fields of machine learning, pattern recognition, and mine countermeasures.

**Marc Carreras** received the B.Sc. degree in industrial engineering in 1998 and the Ph.D. degree in computer engineering in 2003 from the University of Girona, Girona, Spain. He is currently an Associate Professor with the Computer Vision and Robotics Institute, University of Girona. From 1999 to 2021, he has participated in 28 research projects (European and national projects). He is an author of about 150 publications. He has supervised five Ph.D. dissertations and participated in several European autonomous underwater vehicle (AUV) competitions (five times winner). His work in this paper includes defining the research problem, providing the hardware for the experimental setup, and planning and revising the experiments. His research activity mainly focuses on underwater robotics in research topics, such as intelligent control architectures, robot learning, path planning, and AUV design, modeling, and identification.

**Samantha Dugelay** obtained her PhD in mathematics from the University Paris-Sud, France in 1997. She was sponsored by the French Navy Hydrographic Service (SHOM) and served at IFREMER to exploit deep sea multibeam echo-sounder data for automatic seafloor characterisation winning best student paper at both UACE 96 and Oceans'96. She has since then remained in the maritime domain specialising in high resolution synthetic aperture sonar processing and statistical analysis of high resolution data at University College London and the University of Bath (UK) respectively. She then joined the UK Ministry of Defence Dstl as technical lead where she worked on projects ranging from anti-submarine warfare, autonomy, to sonar simulation for mine countermeasures. She moved to the NATO Science and Technology Centre for Maritime Research and Experimentation (CMRE) in 2013 as a project leader and in 2017 as the programme manager for autonomous mine countermeasures. Since 2021, she is a senior analyst at Thales UK, looking at making research an operational reality. Her main interests lie in synthetic aperture sonar processing, machine learning, autonomy and experimentation at sea with a strong emphasis on operational relevance.