

EXPECT THE UNEXPECTED: A MAN-MADE OBJECT DETECTION ALGORITHM FOR UNDERWATER OPERATIONS IN UNKNOWN ENVIRONMENTS.

Thomas SeeByte Ltd, Edinburgh, United Kingdom
Guerneve
PierreYves SeeByte Ltd, Edinburgh, United Kingdom
Mignotte

1 INTRODUCTION

When operating in underwater environments, AUVs can be expected to encounter a wide range of previously-unseen Man-Made Objects (MMO) as well as unfamiliar types of seafloors. In the context of Mine Countermeasure Missions (MCM) in particular, ATR algorithms are typically trained using large amounts of data to capture prior knowledge on specific environments. In order to achieve high detection rates while minimising the need for operator input, the algorithms are typically optimised for a specific sensor payloads and a particular set of objects of interest. This results in high-performance but also highly-specialised algorithms. In practise, the ever-changing environments of deployment of autonomous systems often require rapid adaptability to new types of man-made objects. While traditional machine learning algorithms can take between a few hours up to a few weeks to retrain on new datasets, it is not always possible to retrain an algorithm in-situ. In order to address this problem, we present a sensor-agnostic approach to the development of generic man-made object detection algorithms. Using a combination of simulation techniques and open-source datasets, we present a method to build a generic dataset, suitable for training DNN-based object detection algorithms. Using both qualitative and quantitative experimental results, we demonstrate the suitability of the approach for detecting previously-unseen man-made objects in previously-unseen sidescan sonar sensor data. We introduce an operator workflow to aid in-situ review of the data. We finally discuss the performance and practicality of such algorithms in comparison to traditional highly-specialised MCM algorithms. Finally we discuss the potential field applications of this approach such as performing change detection for monitoring of underwater man-made infrastructures.

2 LITTERATURE REVIEW

Early work on the detection of man-made objects¹ was based on separate classification and detection steps. The detection stage was typically based on segmentation, clustering or markov random fields² approaches. The classification was often achieved by performing matching geometric models to the segmented shape. In situations where further information on the shape of the object is available, a CAD model library can be

generated and employed as prior knowledge for detection and matching in sonar data³. All these methods require prior knowledge in the form of parameters that are typically tuned to the objects of interest. The potential of algorithms achieving better performance than expert human operators at the task of detecting man-made objects was also demonstrated in⁴. Over the last few years, Deep Neural Networks (DNN) have superseded model-based approaches in most computer vision research fields. These models effectively trade the learning of explicit models for large sets of multi-scale 2D filters that are learned from the data through backpropagation⁵. In particular in the domains of object detection and classification, DNNs achieved state-of-the-art performance^{6,7} on benchmark photographic datasets. Likewise in the underwater domain, a Resnet-based Retinanet model was employed to detect boulders in sidescan data⁸, showing the ability to detect small patterns with a similar level of performance as a human operator. A TR-YOLOv5s model was employed in⁹ to demonstrate real-time object detection in sidescan data. A yolo model with a transfer learning approach was used in¹⁰. The performance of detection algorithms being notoriously sensitive to the complexity of the environment^{11,12}, the reliability of algorithms when deployed on the field can be obtained either through online domain adaptability or the training of generic models. Through data-mining and sparse supervision, domain-specialised sidescan ATR algorithms have been shown to benefit from fine-tuning to operational conditions¹³.

3 METHOD

In order to mitigate the uncertainty on the environment of deployment of our algorithms, we create a large dataset by combining simulated and real datasets of various sources.

3.1 Sidescan sonar simulation

We generate simulated sidescan imagery using two different sidescan sonar simulation frameworks. The first framework¹⁴ provides realistic seafloor simulation by modelling seafloor patterns with wavelets. As illustrated in figure 1, the simulator can generate seafloors of various types and material including sand ripples, marine growth and rocky clutter. These types can be combined by adding multiple layers on top of each which allows for generating seabeds of various levels of complexity. The intensities are then rendered using a raytracing process with a Lambert model as reflection model and a Rayleigh model as environment noise model. The sound speed is assumed to be constant and set to 1480 m/s. The transmission loss is computed using the sonar equation, on the basis of a sensor operating at 600 kHz.

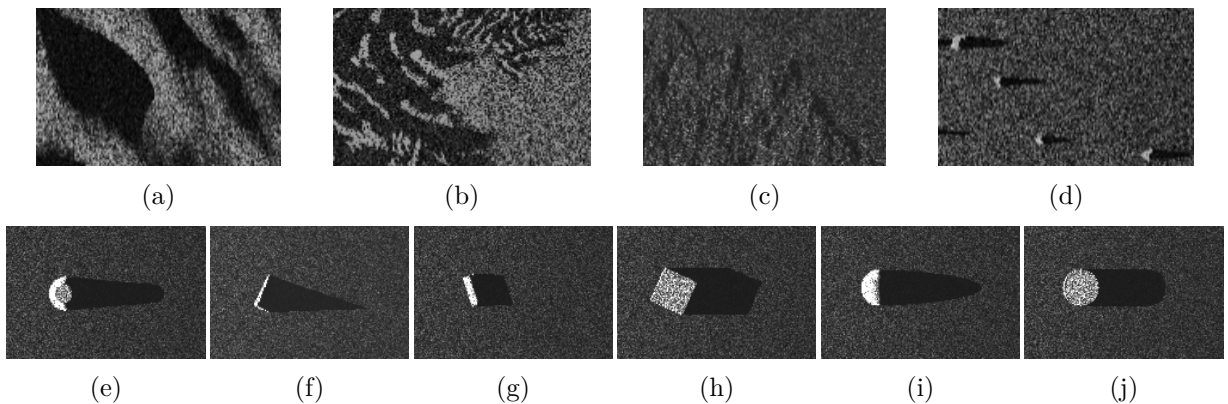


Figure 1: Illustration of different types of seafloors generated with our sidescan sonar simulator: low-frequency (a) and high-frequency (b) ripples, marine growth (c) and rocky clutter (d). 6 basic shapes can be inserted: cones (e), wedges (f), cylinders (g), parallelepipeds (h), hemispheres (i) and upright cylinders (j).

Our second simulation framework enables the insertion of CAD models on top of real sidescan sonar data¹⁵. The 3D models are rendered by first estimating the local bathymetry around the point of insertion. Following this a raytracing step is employed to generate sidescan-looking intensities using a similar process as in the first simulation framework. A sensor noise model is then added and a refinement step based on a Generative Adversarial Model (GAN) can then be optionally applied to match the appearance of specific sidescan sensors. In order to build a dataset with a diverse set of man-made objects, we collected around 100 different open-source CAD models available online¹⁶. As can be seen in figure 2, the models represent a diverse set of man-made objects, in particular: some mine shapes, UxOs (Unexploded Ordnance), lobster traps, AUV (Autonomous Underwater Vehicle), ROV (Remotely-Operated Vehicle), tyres, anchors, various containers, buoys, divers, baskets, ship and plane wrecks, pier items. In order to represent difficult seafloors, a few CAD models of rocks and corals were also employed.

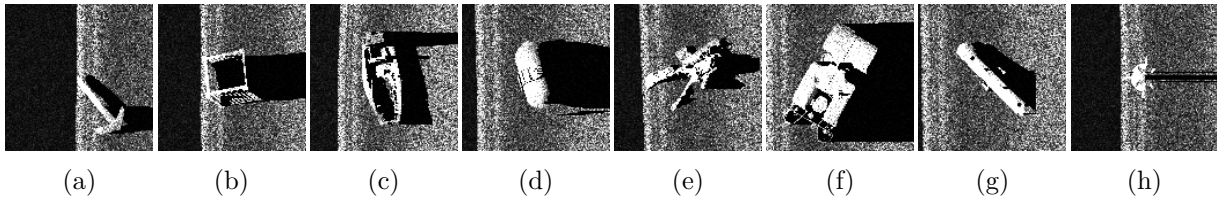


Figure 2: Simulated sidescan images of various CAD models of man-made objects: anchor (a), crate box (b), boat (c), gas tank (d), diver (e), ROV (f), AUV (g), floating mine (h).

Using these two simulation frameworks, we generated 200GB of sidescan sonar data following different navigation patterns and inserting objects in random locations and orientations to obtain observations at different ranges and angles. The data was generated at 2cm x 2cm resolution.

3.2 Open-source real datasets

The simulators described in the previous section generate data at arbitrary resolutions and ranges without taking into account all of the physical limitations encountered when using real sensors. As opposed to this, real sidescan sensors operate at a specific frequency and range. The speed of vehicle is then set based on the characteristics of the sensor and the desired range and resolution. The imagery delivered by real sensors is then characterised by frequency-specific noise patterns and resolutions. In addition to these characteristics, our simulated frameworks do not model for artefacts due to the presence of other acoustic devices (such as other sonar sensors or acoustic communications). Real datasets often exhibit the presence of fish in the watercolumn or so-called surface-returns created by the sea surface when operating in shallow waters. In order to include these elements in our training set, we employed open-source datasets found using Google dataset search tool¹⁷. These datasets^{18,19,20,21,22,23} contain sensors from several sonar manufacturers (Klein, Edgetech, Marine Sonics) and feature various resolutions (3cm to 15cm across-track resolution) and ranges (30m to 100m) with a total of 260GB. The vast majority of these datasets is composed of seafloor-only data with a few images of shipwrecks²¹ and mines²².

3.3 Training of a DNN-based MMO detector

As shown in figure 3, we combine simulated, hybrid (real seafloor with raytraced CAD models inserted) and real datasets together to form the training dataset to our MMO detector, with a total of 4000 observations. The combination is done at training time, randomly sampling images with an equal balance of images containing man-made objects and seafloor-only images.

Using this large training dataset composed of simulated and real data, we train a DNN model with a

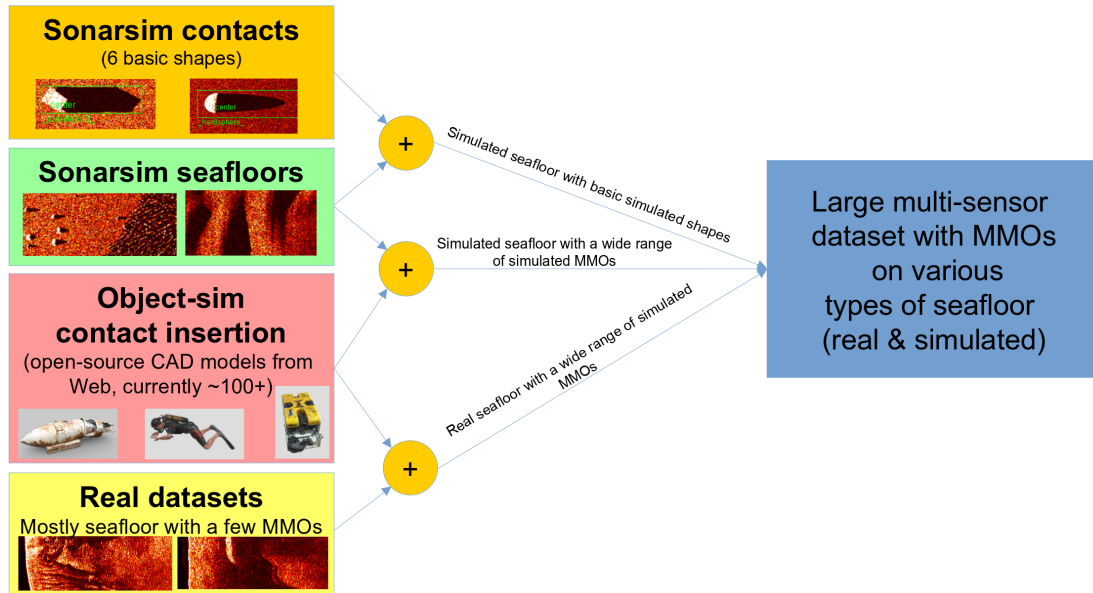


Figure 3: A large training dataset is obtained by combining simulated, hybrid and real datasets with a wide range of sensors, seafloors and man-made objects.

Resnet18 backbone using a small amount of augmentation in the form of geometric (affine) transforms, noise and dynamic range augmentation. The model is trained at a resolution of 4x4cm to allow for detecting objects with sizes in the range of 20cm up to a few meters.

4 EXPERIMENTS

In this section we provide qualitative and quantitative results obtained when evaluating the MMO detector on real datasets. Importantly, these datasets feature objects that were not represented in the training set to evaluate the capacity of the detector to generalise to other types of MMOs.

4.1 MMO dataset

The first dataset is a dataset composed of 40 real sidescan images containing MMO of different kinds. These images come from publicly-available sources^{24,25} and feature various sensors coming from Edgetech and Klein sonar manufacturers. Figure 4 shows the detections returned by the trained detector on different types of man-made objects.

Most of the MMOs present in these images are detected with a high level of confidence, including the objects in images 4-d,e,f,g which are types of objects that were not included in the training dataset. Image 4-a shows a missed detection (green box in the nadir area with very faint features) as well as a false alarm (white box in the top right corner) on a dredging mark on the seabed.

4.2 Human body dataset

The second dataset was provided by HEART²⁶ (Hutterian Emergency Aquatic Response Team), a charitable organization specialised in search and recovery of drowning victims. The dataset is composed of 3

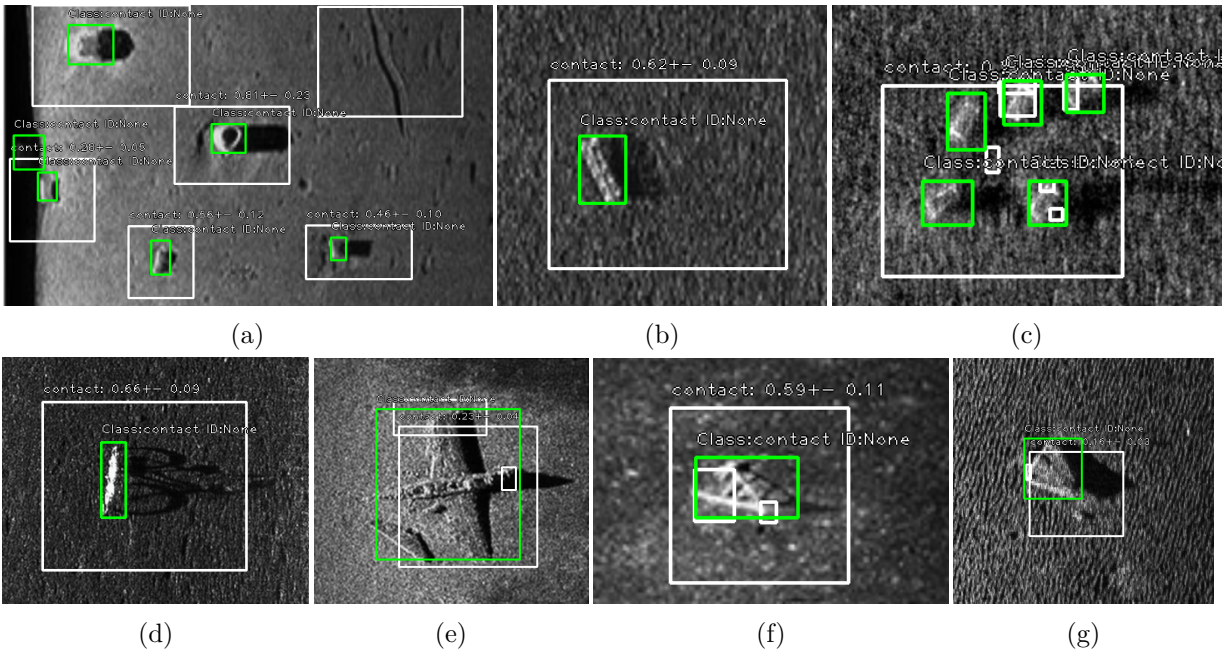


Figure 4: Detections (white boxes) obtained on real sidescan images featuring different types of MMOs (ground truth labels as green boxes). The first row of images (a,b,c) features containers of shapes that are similar in appearance to the MMOs simulated in the training set. The second row shows MMOs that were not present in the training set: a bicycle (d), a plane (e), an antenna (f) and a triangular mine (g).

missions acquired with a Marine Sonic sea scan HDS towed system and a total number of 51 observations on human bodies. The towed system was employed to survey 3 lakes in Saskatchewan, Canada. As visible in figure 5, the model successfully detects the human bodies present in the dataset despite the lack of highlight in their appearance.

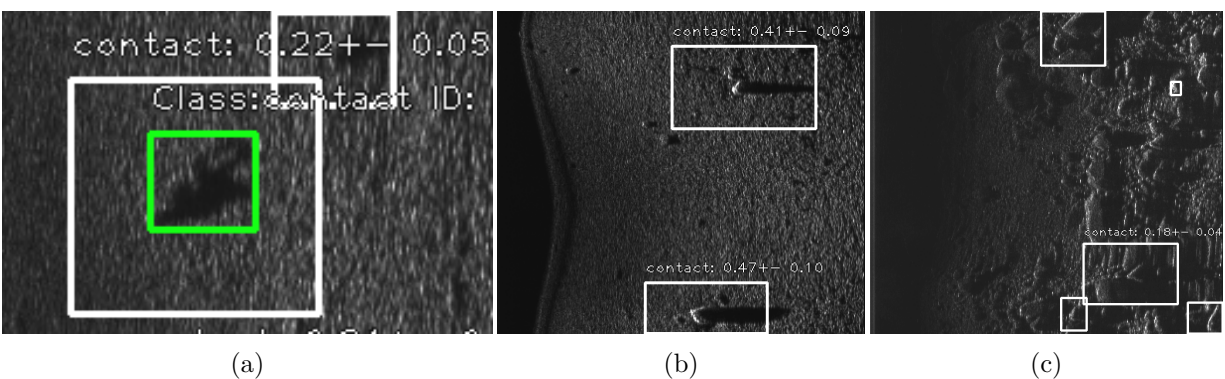


Figure 5: Detections (white boxes) and ground truth labels (green boxes) obtained on real Marine Sonic sidescan sonar images featuring: a human body (a), man-made-object-looking clutter (b) and rocks (c).

Although some CAD models of divers were included in the simulated training set (see figure 2-e), it should be noted that their appearance was significantly different as these were rendered with a much sharper highlight in comparison to figure 5-a where only the shadow is visible. As visible in figure 5-b,c, the MMO model returns detections on unmarked seabed items such as rocks with angular or mine-like features.

4.3 MMO detection performance and comparison to a traditional ATR

In addition to the qualitative results shown in sections 4.1 and 4.2, we perform a quantitative evaluation using a set of ground truth labels for each dataset. Based on these ground truth labels and the detections returned by the MMO detector, we derive the PD (Probability of Detection) and number of FA (False Alarms) per square kilometers. These two metrics are computed at various levels of confidence to generate a ROC (Receiver Operating Characteristic) curve on each dataset. In order to provide a baseline, we first plot the performance of the MMO detector on the large simulated dataset used for training the model as represented by the blue curve in Figure 6. The performance on the real MMO dataset described in section 4.1 is represented by the green curve and the performance on the human body dataset described in section 4.2 is represented by the red curve. In order to show the difference in confidence of the model on each dataset, asterisk markers are placed to represent the performance at a given level of confidence (0.17). In order to compare the results obtained by our MMO detector with a traditional mine-hunting ATR, we trained the same DNN on a dataset composed of simulated backgrounds and real mines²² with no other man-made objects. This ATR model was evaluated on the real MMO dataset and is represented by the dashed green curve, showing a PD that is 30% lower on average than the MMO detector.

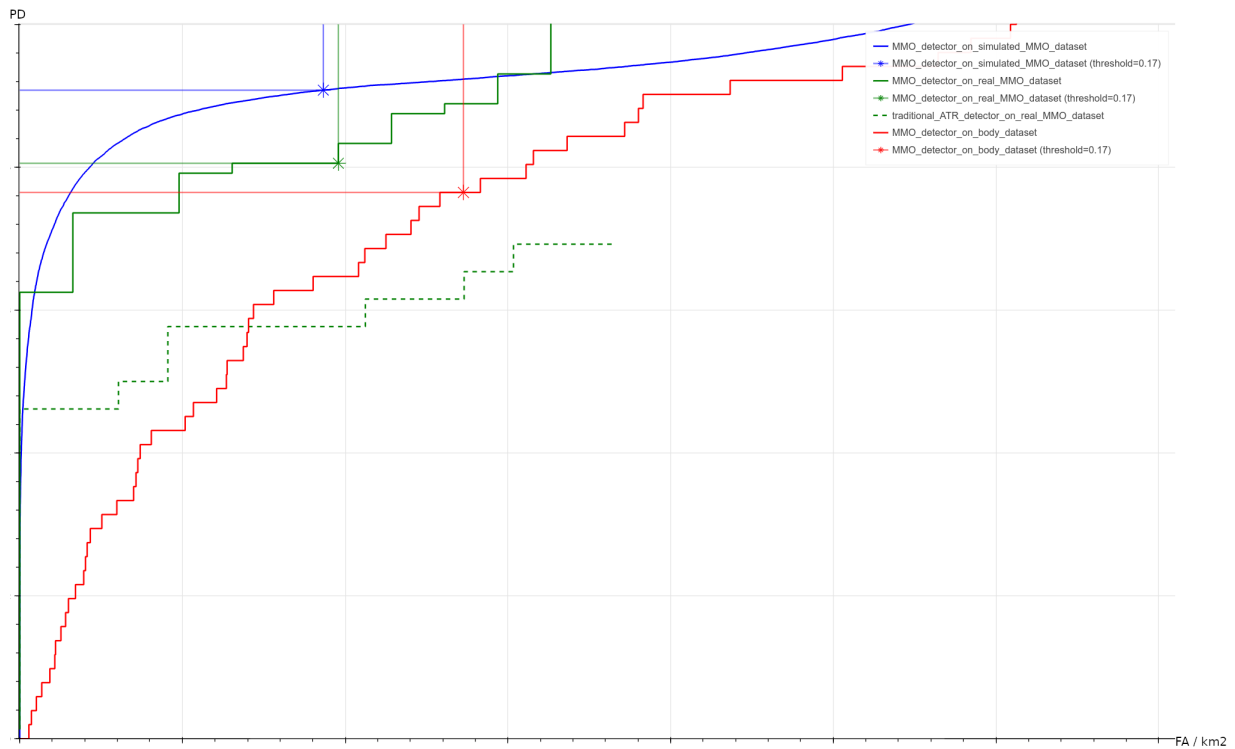


Figure 6: ROC curves achieved by the MMO detector on three datasets: training set with simulated MMOs (blue), real MMO dataset (green) and human body dataset (red). On each curve an asterisk marks the point corresponding to a confidence level of 0.17 to show the difference in performance for a given confidence level. The performance of a traditional (mine-hunting) ATR algorithm on the real MMO dataset is represented by the dashed green curve, showing a lower PD than the MMO detector.

5 DISCUSSION / ANALYSIS

The results presented in section 4 demonstrate the ability to detect previously-unseen MMOs in real data. For a given level of confidence (asterisks on figure 6), the model achieves higher PD on the simulated set

than on real data. This can be expected due to the differences in appearance between real and simulated targets which are typically sharper and feature brighter highlights than real targets. When processing real data, high PD is then obtained by operating a lower levels of confidence which results in higher number of false alarms. These false alarms can be mitigated by performing post-mission analysis with a trained operator. Practical MCM missions follow rigorous ConOps (Concept of Operations) that rely on multi-pass surveys. The fusion of multiple views can be operated based on contact type, size and location to reduce the number of false alarms. Using clustering techniques¹¹, frequent artefacts due to the presence of acoustic communications, local fauna or sand ripples could then be processed jointly to speed up the review process by avoiding the need for reviewing each detection independently.

As opposed to the approach presented in this paper, traditional ATR algorithms are typically trained on a small and specific set of mines⁹. This leads to high PD / FA rates but comes at the cost of not being able to detect other types of man-made objects, as shown on figure 6 when comparing the two green curves on the real MMO dataset. The ever-changing scenarios that can be encountered on a field of MCM operations require adaptability to previously-unseen environments. While this can be achieved through acquiring data on the new domain and retraining a specialised ATR algorithm, this is an expensive operation and is likely to require a significant amount of time. In this context, the experimental results show the benefit of employing models trained on larger sets of objects to detect previously-unseen data. These models could then be employed by an operator to compensate for the lack of adaptability of traditional ATRs.

6 CONCLUSION

The approach demonstrated in this paper shows that a DNN-based model can be trained on a generic dataset using a combination of simulation tools and open-source datasets to detect previously-unseen man-made objects in real sidescan sonar data. The experiments demonstrated the suitability of simulation tools to train models that perform well on real datasets. We also showed that this method can be used as an alternative to the expensive retraining of specialised ATR algorithms on new domain data. This approach provides a higher level of adaptability than traditional ATR algorithms at the cost of some false alarms detections on complex seabeds such as rocky clutter. The capacity to detect a larger range of objects than traditional ATRs benefits to landmark-based applications such as relocalisation and autonomous decision making on board of AUVs. Future work will leverage the ability to detect generic man-made objects by investigating change detection in the context of monitoring subsea infrastructures.

ACKNOWLEDGMENTS

The authors would like to thank Manuel Maendel as well as the HEART (Hutterian Emergency Aquatic Response Team)²⁶ organisation for sharing their dataset.

REFERENCES

1. Dura, Esther, et al. "Image processing techniques for the detection and classification of man made objects in side-scan sonar images." Sonar Systems. Makati, Philippines: InTech, 2011.
2. Reed, Scott, Yvan Petillot, and Judith Bell. "An automatic approach to the detection and extraction of mine features in sidescan sonar." IEEE journal of oceanic engineering 28.1 (2003): 90-105.
3. Guerneve, Thomas, Kartic Subr, and Yvan Petillot. "Underwater 3D structures as semantic landmarks in SONAR mapping." 2017 IEEE/RSJ International Conference on Intelligent Robots and

- Systems (IROS). IEEE, 2017.
4. Kessel, Ronald T., and Vincent L. Myers. "Discriminating man-made and natural objects in sidescan sonar imagery: Human versus computer recognition performance." *Automatic Target Recognition XV*. Vol. 5807. SPIE, 2005.
5. Hecht-Nielsen, Robert. "Theory of the backpropagation neural network." *Neural networks for perception*. Academic Press, 1992. 65-93.
6. Lin, Tsung-Yi, et al. "Focal loss for dense object detection." *Proceedings of the IEEE international conference on computer vision*. 2017.
7. Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
8. Feldens, Peter, et al. "Detection of boulders in side scan sonar mosaics by a neural network." *Geosciences* 9.4 (2019): 159.
9. Yu, Yongcan, et al. "Real-time underwater maritime object detection in side-scan sonar images based on transformer-YOLOv5." *Remote Sensing* 13.18 (2021): 3555.
10. Einsidler, Dylan, Manhar Dhanak, and Pierre-Philippe Beaujean. "A deep learning approach to target recognition in side-scan sonar imagery." *Oceans 2018 Mts/Ieee Charleston*. IEEE, 2018.
11. Terrain characterisation for online adaptability of automated sonar processing: Lessons learnt from operationally applying ATR to sidescan sonar in MCM applications, UACE 2023.
12. Daniell, Oliver, and Jose Vazquez. "Estimation of detection/classification performance in sonar imagery using textural features and self-organising maps." *OCEANS 2017-Anchorage*. IEEE, 2017.
13. de Bodinat, Jean, et al. "Similarity-based data mining for online domain adaptation of a sonar ATR system." *Global Oceans 2020: Singapore-US Gulf Coast*. IEEE, 2020.
14. Pailhas, Yan, et al. "Real-time sidescan simulator and applications." *OCEANS 2009-EUROPE*. IEEE, 2009.
15. Karjalainen, Antti Ilari, Roshenac Mitchell, and Jose Vazquez: "Training and Validation of Automatic Recognition Systems using Generative Adversarial Networks" in *Sensor Signal Processing for Defence Conference (SSPD)*, IEEE (2019).
16. <https://sketchfab.com/>
17. <https://datasetsearch.research.google.com/>
18. <https://www.data.gov.uk/dataset/6240585f-a9d3-42d8-9f07-3ef19a84baa1/raw-sidescan-sonar-data-from-dogger-bank-sci-cend-07-08>
19. <https://www.data.gov.uk/dataset/5773e015-c014-416c-aa43-a0397a8fd9e9/raw-sidescan-sonar-data-from-solan-bank-area-of-search>
20. Marco Bernardi, Brett Hosking, Chiara Petrioli, Brian J. Bett, Daniel Jones, Veerle A.I. Huvenne, Rachel Marlow, Maaten Furlong, Steve McPhail, Andrea Munafo. (2020). "AURORA, A multi sensor dataset for robotic ocean exploration." Web.
21. Sethuraman, Advait V., et al. "Machine Learning for Shipwreck Segmentation from Side Scan Sonar Imagery: Dataset and Benchmark." *arXiv preprint arXiv:2401.14546* (2024).
22. Santos, Nuno Pessanha, et al. "Side-scan sonar imaging data of underwater vehicles for mine detection." *Data in Brief* 53 (2024): 110132.
23. Commonwealth of Australia (Geoscience Australia) (2018): *The Search for Flight MH370 - Phase 2 Raw and Processed on the National Computational Infrastructure*. .dataset. <https://pid.geoscience.gov.au/dataset/ga/120962>
24. <https://www.edgetech.com/underwater-technology-gallery/>
25. Chosid, D. F. "Development of Side Scan Sonar Methodology to Survey Derelict Lobster Pots in Simple and Complex Habitats in Massachusetts." *Government of Massachusetts* (2017).
26. <https://www.hearteam.ca/>