



# Perception for Underwater Robots

John McConnell<sup>1</sup> · Ivana Collado-Gonzalez<sup>1</sup> · Brendan Englot<sup>1</sup>

Accepted: 4 October 2022 / Published online: 24 October 2022  
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

## Abstract

**Purpose of Review** The purpose of this review is to survey recent work in the area of perception for underwater robots. We consider problems such as object-level identification and classification, Simultaneous Localization and Mapping (SLAM), and 3D reconstruction. Our goal is to understand the current state of the art, how old shortcomings have been addressed, and the new issues introduced by recent work.

**Recent Findings** We consider findings in several key areas. In vision, lower-cost consumer-grade cameras have been employed to perform mapping of large-scale areas. In sonar imaging, we find major steps forward in terms of reconstruction quality, with datasets for semantic labeling still incomplete. In bathymetry, we find many mature and robust systems. Lastly, in side-scan sonar we find a budding research area, with much promise ahead and some impressive initial work.

**Summary** We break our survey down by sensor modality and consider three key areas; object-level identification and classification, SLAM, and 3D reconstruction.

**Keywords** Marine robotics · Sonar · Computer vision · Localization and mapping · Machine learning

## Introduction

Perception is a fundamental task for robots. We consider perception as the problem of processing observations of the surrounding environment to achieve situational awareness. Perceptual tasks can range from high-level mapping, down to object-level understanding and robot state estimation. In our view, perception is the first in several critical steps toward autonomy. Downstream processes such as planning and control require reliable, robust, and consistent perceptual systems to function.

In contrast to other environments, underwater perception is inhibited by the environment itself: plagued by low light, varying water conditions, the attenuation of light and other

signals in water, and environmental disturbances far beyond the control of human monitors. These factors limit the tools for underwater perception. Low-cost structured light sensors and infrared LiDAR cannot be used over long distances underwater because water absorbs most or all of the infrared laser energy directed through it, resulting in a very weak or absent return. Radio signals are similarly attenuated, making inter-platform communication a challenge. The Global Positioning System (GPS) provides key pose information for above-surface localization and navigation tasks. However, the electromagnetic signals from orbiting satellites are heavily attenuated by the water, thus, GPS is not adequate for underwater applications. This leads to a more challenging localization procedure, with subsea systems relying on inertial and acoustic sensors for dead reckoning, with numerical integration errors that worsen over time. A common strategy to rectify drift is self-correction by using environmental features; this can be achieved by employing Simultaneous Localization and Mapping (SLAM).

Optical underwater navigation is challenging due to its susceptibility to poor visibility caused by suspended particles, inadequate lighting, and unwanted distortions. Additionally, visual odometry feature extraction relies on texture-rich images, which can be sparse in underwater settings. Nevertheless, in contrast to acoustic sensing, visual

---

This article is part of the Topical Collection on *Underwater Robotics*

✉ Brendan Englot  
benglot@stevens.edu

John McConnell  
jmconn1@stevens.edu

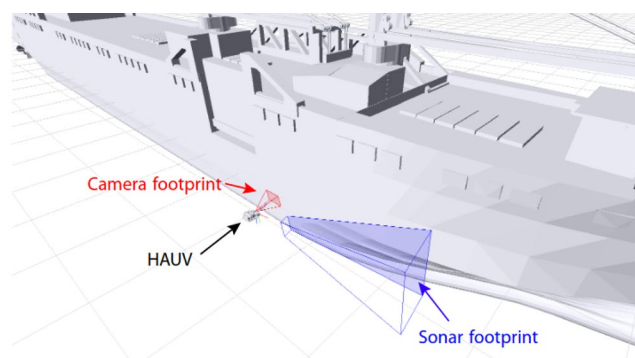
Ivana Collado-Gonzalez  
icollado@stevens.edu

<sup>1</sup> Stevens Institute of Technology, Hoboken, NJ, USA

navigation can be well suited for precise mapping of feature-rich scenarios such as ship hull and shipwreck inspections. Furthermore, cameras provide colorful and detailed information helpful in object detection and classification tasks. Moreover, the use of cameras in this domain has spawned a research area addressing the issues caused by the water medium.

The use of sonar sensors is motivated by the challenges for cameras imposed by the underwater environment. Sonar is immune to lighting conditions, has a long range due to the properties of sound in water, and is robust to water turbidity. Nonetheless, sonar is not a plug-and-play solution and has specific drawbacks. Several varieties of sonar are available, and here we will consider imaging, side-scan, and profiling. Profiling sonar uses an array of narrow “pencil” beams to recover accurate distance measurements at known bearing angles. However, to build an understanding of the environment, a dead reckoning system is required to stitch line scans together. Imaging sonar, on the other hand, is lower in cost and returns a panoramic view of the environment across a wide horizontal aperture. Imaging sonar, however, does not measure the location of observations in the vertical aperture, resulting in high ambiguity about the location of observed points in 3D space. Lastly, side-scan sonar has two transducer arrays that send and receive acoustic pulses; it is an excellent tool when performing seafloor searches, especially when covering large areas at low image resolution.

Underwater-specific obstacles have resulted in an impressive body of work containing all sensors employed in this domain. In this survey paper, we consider recent developments in perception for underwater robots, categorized by sensor and their respective subfields.



(a) Example of a sensor footprint in a ship hull inspection modality [1].

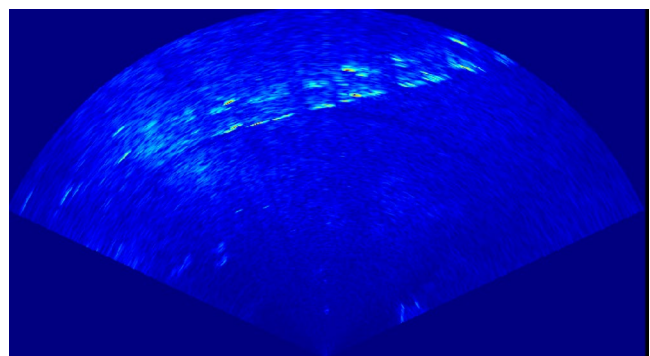
**Fig. 1** (a) shows an example of a Bluefin HAUV conducting a ship hull inspection mission; note the large volume of the imaging sonar footprint shown in blue (From: Hover F.S., et al. *The International Journal of Robotics Research* (Volume 31 Issue 12). Pages 1445-

## Imaging Sonar

In this section, we will discuss imaging sonar and its recent perceptual algorithms. It is known by several names that are often used interchangeably: imaging sonar, wide aperture multi-beam sonar, and forward-looking sonar. However, when we discuss imaging sonar, we mean a multi-beam sonar with a non-trivial vertical aperture. This sensor provides an expansive field of view and is effective for imaging large volumes of water, making it an excellent tool for situational awareness, allowing a robot to perceive much of the environment around it. Further, sonar is robust to lighting conditions and water quality, making it the sensor of choice when operating in turbid environments. However, imaging sonar does not report information in 3D. While the sensor can measure range and bearing, the elevation angle of an observed point is lost in the image formation process. Moreover, these sensors have a low signal-to-noise ratio as well as multi-path and other characteristics that make it challenging for automated perceptual algorithms. Examples of an imaging sonar’s field of view and imagery are given in Fig. 1.

## Object Identification

Identifying, classifying, and understanding objects in the environment is a fundamental task for underwater robots. Firstly when considering object identification, or automated target recognition, many classification methods have been applied. However recently, in settings such as vision and LiDAR, deep learning methods have far eclipsed classical approaches [2]. Critically though, when considering imaging sonar there are few datasets and even fewer with labels to support object classification or similar.



(b) Sample sonar image.

1464. ©2012 by SAGE Publications, Reprinted by Permission of SAGE Publications) [1]. (b) shows an example sonar image with a floating dock in view, and a range of 30m

The focus of recent work has been learning labels while only utilizing a few labeled samples. However, unlike in other settings, this is motivated by necessity and not the desire to reduce data procurement overhead when training. Transfer learning from other image domains to initialize the weights of a Convolutional Neural Network (CNN) is sometimes used, reducing the number of training samples required in the sonar data [3, 4]. A Generative Adversarial Network (GAN), in this case, CycleGAN, is used by Liu et al. [5] to generate synthetic training data. Chen et al. [6] use the method of few-shot learning, which requires only a few training examples of each class. Rather than requiring explicit labels, this work trains using image similarity. Interestingly, Wang et al. [7] avoids the need for unknown object classification by introducing an engineered target. This target can be classified using more standard methods, similar to AprilTags [8]. Although promising, relying on engineered targets [7] requires human effort for building and placing markers on existing structures.

### 3D Reconstruction

Recall that imaging sonar does not contain 3D information, even though the sensor observes a 3D volume of water. This creates some open questions, about how to recover the 3D data and build a 3D understanding of the environment. This is critical, especially when considering robots using this sensing modality that may need to operate outside of a fixed plane.

Firstly we consider approaches that utilize multiple views to resolve the ambiguity in the vertical aperture of an imaging sonar. Akin and Negahdaripour [9] use a space carving approach, with Westman et al. [10] enhancing space carving. Wang et al. [11] extend Acoustic Structure-From-Motion (ASFM) [12] in tracking salient point features through space to minimize their ambiguity. DeBortoli et al. [13] use a CNN to detect salient imagery, removing the imagery that is unproductive when completing 3D reconstruction with ASFM as the backend. Westman et al. [14] introduce a volumetric framework to perform 3D reconstruction, testing sonars with a narrow and wide aperture.

While using multiple views to resolve sonar ambiguity is principled and draws on many methods from outside underwater robotics, it may be desirable to recover 3D information without multiple frames. Guerneve et al. [15] use a blind deconvolution and Westman and Kaess [16•] use an acoustic generative model to recover the missing data in a sonar image. Both of these methods are effective and draw on classical methods to recover the elevation angles of the observed points. DeBortoli et al. [17] employ a CNN and synthetic training data to predict the missing data in a sonar image, with Wang et al. [18] learning from each beam rather than image-to-image, making a comparison to [17]. All of these

methods recover the missing elevation angle to accompany the 2D range-bearing observations of a single sonar image, at a single viewpoint.

A new set of methods, introducing a second sonar in a stereo array have also been proposed [19–21]. These methods recover point clouds from a pair of concurrent sonar images at the same robot viewpoint using sonars with different perspectives, with McConnell et al. [19] focusing on an orthogonal array of sonars. McConnell and Englot [22] extend their prior work [19] by using object-level inference about simple objects, greatly increasing the coverage rate for a given trajectory.

### SLAM

When considering perception, SLAM is considered a critical task, and in the context of imaging sonar a non-trivial one. In this section, we will consider three subtopics of SLAM; sonar odometry, loop closure, and SLAM systems.

Generally speaking, SLAM solutions depend on two categories of measurement constraints; sequential constraints, typically derived from odometry or scan-matching, and non-sequential constraints, often referred to as loop closures. Loop closures are typically achieved in two steps, identifying loop closures and computing the relevant transformation between frames to be inserted into the SLAM backend. Franchi et al. [23] consider the problem of deriving linear speed measurements from sonar imagery, akin to sonar odometry. Henson and Zakharov [24] build sonar mosaics using optical flow as a basis for transform estimation. Almanza-Medina et al. [25] use a neural network to learn 3-DOF transformations between sonar images and Song et al. [26] consider sonar image registration; both of these works could be utilized for either odometry or loop closure. Santos et al. [27] consider the use of a graph structure to describe a scene for both loop closure identification and registration. Ribeiro et al. [28] use a CNN with triplet loss to train a network for sonar image-based place recognition, framed as an image retrieval problem.

While using objects as landmarks for SLAM is relevant, most of the recent approaches have focused on pose SLAM or landmark-based SLAM using point features as landmarks. Firstly, Westman et al. [29] use point features in sonar imagery as a basis for SLAM, taking special care to handle their vertical ambiguity. Ribeiro et al. [30] use a local ASFM as well as loop closure to build a 6-DOF state estimate. Wang et al. [31] estimate the robot state in 3-DOF while enforcing in-plane motion; this work uses Iterative Closest Point (ICP) based scan-matching to derive both sequential and non-sequential measurement constraints. Teixeira et al. [32] use a factor graph with odometry as well and a surfel-based map. Hinduja et al. [33] propose a degeneracy-aware SLAM system to manage possible degenerate factors. Xu

et al. [34] use a sliding window to stay robust to front-end outliers. Recently a new area has been considered for underwater robots in littoral settings; using satellite images as an aid in the SLAM problem. Dos Santos et al. [35] use a neural network to compute the similarity between a possible location in the satellite image and a given sonar scan from the robot's current state. This similarity is used as the basis for particle filter localization in the survey area. McConnell et al. [36] utilize a similar concept, except instead of using a neural network to find the similarity it generates an image predicting the above-surface appearance of the observed underwater structures, to enable the registration of sonar images to the provided satellite image. McConnell et al. [36] then use this registration as another factor in a SLAM solution that includes odometry and standard loop closures.

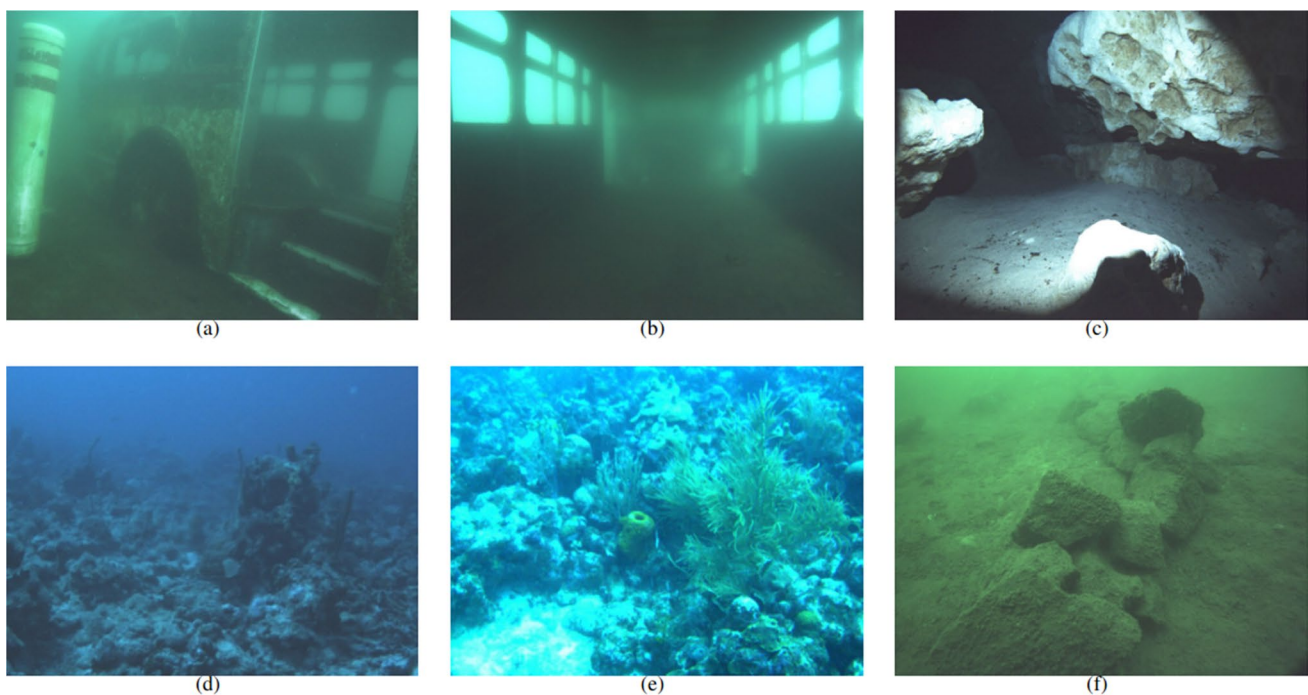
## Underwater Vision

In this section, we will consider camera-based systems in the underwater domain. Cameras provide dense, expansive information about the environment. Unlike sonars, they provide rich color and texture information and are easy to interpret by untrained human operators. Cameras can support a wide range of tasks from the object level to SLAM. Moreover, the broader vision field is large and quickly growing, providing a significant body of research to draw upon.

Cameras, like all sensors, have their drawbacks. These sensors are passive, and therefore require sufficient ambient lighting or lighting onboard the vehicle to make them useful. Further, camera data may be corrupted by water quality. If water turbidity is high and visibility low, cameras may simply not offer a useful path forward for underwater perception. Lastly, the properties of water create a loss of red channel as well as hazing in images, creating a loss of texture required for many automated perception methods. Examples of underwater imagery in different conditions are given in Fig. 2.

## Underwater Image Enhancement

The properties of water create a loss of red and texture in images that can be problematic for image processing systems. This has spurred an impressive body of research on color correction and image dehazing solutions. There have been many approaches to this problem but in general, they either use a physics-based model or employ machine learning to recover the lost information. Non-learning approaches are also used to recover some of the missing information [38–44]. These approaches are effective as they require little prior information about the environment to recover lost data. Machine learning in this case is challenging as labeled data is usually required, including examples of scenes with and without the effects of water immersion. For tank scenes this is feasible, but for scenes in the field, it may not be.



**Fig. 2** (a)–(f) show examples of camera images from a variety of environments, lighting conditions, and water qualities. (©[2019] IEEE. Reprinted, with permission, from [37].)

When learning image correction in underwater vision, a recent focus has been the use of GANs, to create synthetic data for supervised learning [45–50]. The use of a GAN is of great importance as it may not be possible to generate enough training data for image enhancement, even when using data-efficient methods. GANs, however, may create a performance gap due to training on synthetic data and testing on real-world underwater scenes.

## Object Identification

Object identification and classification are widely explored tasks in vision systems. Accurate and detailed analysis of subsea images is relevant for evaluating and monitoring ecosystem states [51], assessing environmental impact, robot orientation prediction [52], tracking life forms [53], and carrying out many other tasks. However when underwater, there are unique problems: mainly lower image quality and the lack of training data. Obtaining large datasets can be expensive and impractical in underwater environments because of high operational costs, time constraints, and scarcity of specific objects. Nonetheless, there are several examples of supervised learning for classification using deep learning for seafloor image classification [54], wildlife monitoring [55], and coral detection [51]. To deal with the lack of sufficient training data, a common technique in deep learning is to use transfer learning, pre-training on a large dataset then fine-tuning on the domain-specific data [56–58]. Another option to obtain sufficient data is to train using synthetic samples, shown by O’Byrne et al. [59]. To entirely avoid the need for a large dataset, algorithms capable of learning from only a handful of samples can be leveraged, Ochal et al. [60] analyze several “few-shot” learning techniques over underwater imagery. Furthermore, Yamada et al. [61] use environmental metadata, such as horizontal location and depth of the observed seafloor, to regularize learning and enhance the accuracy of the self-supervised classification process. Detection has also been used to enable “follow me autonomy”, where a vehicle follows along with a human diver [62, 63] or for multi-robot operations [64].

Extracting meaningful information over a broad range of environments and image quality conditions is challenging. As previously mentioned, producing large underwater datasets is costly and the intensive effort of manual labeling often makes supervised learning techniques impractical. Topic models are a family of Bayesian probabilistic models, suitable for unsupervised semantic clustering. Girdhar et al. [65] proposed a Realtime Online Spatiotemporal Topic (ROST) modeling framework, which attempts to model the semantics of the streaming observed visual data. Using ROST, a “surprise” score of incoming observations is computed. This score is used to determine the presence of high-level patterns in the scene, and differentiate between

previously observed or new data. Moreover, the topic labels computed using ROST are suitable for use by autonomous agents working with real-time constraints. Based on ROST, Kalmbach et al. [66] developed a full and robust feature extraction pipeline that can accommodate video recorded under less than ideal quality conditions.

## SLAM

Vision-based SLAM is a mature field with a large body of work to draw from. There are even some prominent open-source systems such as ORB-SLAM [67, 68] upon which underwater SLAM systems can be based. Underwater SLAM, however, is challenged by domain-specific environmental conditions: poor lighting conditions, water turbidity, and the reduction in image quality due to the properties of water.

The performance of open-source frameworks in underwater conditions have previously been evaluated [37, 69–71]. Zhang et al. [72] and Ferrera et al. [73] propose robust visual odometry systems for use in underwater environments. While the underwater domain imposes new challenges on cameras, it does offer some additional tools. These range from acoustic devices such as sonars and Doppler Velocity Logs (DVLs) to simple pressure sensors. Xu et al. [74] and Vargas et al. [75] enhance visual SLAM with DVL and inertial/DVL fusion, respectively. Sonar has been employed to augment a visual SLAM system with acoustic range measurements [76, 77]. Rahman [78] show stereo visual SLAM assisted by a pressure sensor and inertial measurement unit (IMU) and Hu et al. [79] use a pressure sensor for scale initialization with a monocular camera. Rahman et al. [80, 81] consider the artifacts of artificial light in an underwater cave, such as harsh shadows and contours. Most recently, however, a GoPro-9 camera with integrated IMU has been used for SLAM, potentially enabling complex missions while minimizing hardware overhead [82]. While Bosch et al. [83] does not consider the SLAM problem, it does open the door for the use of omni-directional cameras in an underwater setting. Xanthidis et al. [84] show an example of multi-robot visual SLAM around shipwrecks. Suresh et al. [85] avoid some of the issues in underwater vision, by looking at the ceiling outside of the water, and performing state estimation using those features. However, this approach is limited to indoor applications.

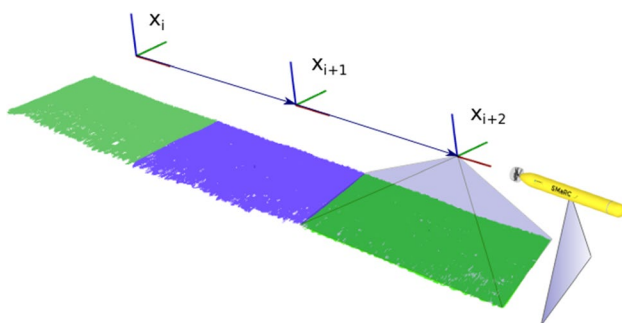
## Bathymetry and Profiling

Of all the underwater perception systems, profiling and bathymetric mapping are perhaps the most relevant in commercial and industrial settings. They are used in support of many offshore construction projects including oil and gas,

offshore wind, telecommunications cabling, and many more. Bathymetry is typically the process of using a narrow beam scanner to recover precise distance information between a robot and the seafloor/riverbed, while profiling can refer to scanning in any direction, including in a forward-looking orientation. Most commonly, these single scans are registered into submaps using a highly accurate inertial navigation system, often including a DVL. Once submaps are formed, tools such as point cloud registration can be applied to enhance inter-submap alignment and search for loop closures. In this section, we will consider recent developments in this area, including bathymetry and profiling performed with both narrow beam sonars and laser scanners. An example of submap formation during a bathymetry survey is shown in Fig. 3.

In bathymetric mapping and profiling, objects of interest in the survey area can greatly enhance SLAM performance. Guernev et al. [87] perform semantic mapping using prior CAD models of expected objects in the survey area. When these priors are available, they can be of great value, however, they may not always be an option, making the case for techniques such as submap registration. Torroba et al. [88] perform a comparison of modern methodologies for submap registration. Hitchcox and Forbes [89] use Gaussian process regression for point cloud registration, enabling bathymetric SLAM using a laser scanner. Jung et al. [90] enhance ICP performance by regularizing terrain height. ICP however can be degenerate, and it is critical to indicate to SLAM backends which measurements can be relied upon, often with a covariance matrix. Sprague et al. [91] learn ICP covariance rather than deriving it with variable initial guesses.

Improving submap registration may not be enough, and further work may be required. Campos and Garcia [92] filter point clouds acquired from noisy sensors by creating a surface mesh. Bore et al. [93] apply sparse Gaussian processes to reduce the storage overhead of bathymetric maps.



**Fig. 3** An example of profiling line scans being stitched into submaps shown in green and purple. Note the submap poses shown as  $x_i$ . (©[2019] IEEE. Reprinted, with permission, from [86])

While many backends have been applied to bathymetric SLAM, Zhang et al. [94] use a particle filter backend for SLAM, with Teng et al. [95] employing one specifically for detecting invalid loop closures. Torroba et al. [86] fuse dead reckoning information and maximize geometric consistency to produce accurate maps.

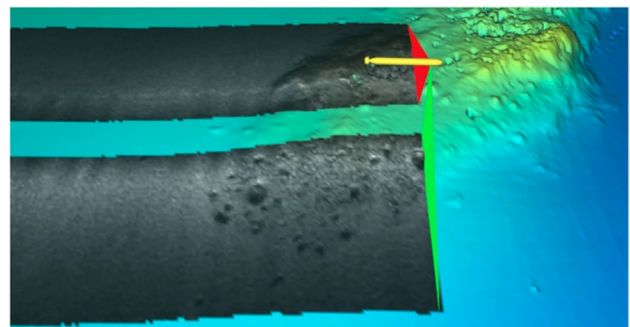
Lastly, profiling sensors can be applied for more than simply downward-looking bathymetric sweeps. Teixeira et al. [96] use profiling sonar submaps to perform 3D mapping of ship hulls. Palomer et al. [97] propose a calibration routine for a camera and laser scanner. Palomer et al. [98] use a laser scanner to perform SLAM about an object of interest. Norgren and Skjetne [99] propose SLAM around an iceberg, estimating robot state as well as iceberg drift rate, critical in ice monitoring.

## Side-Scan Sonar

In this section, we discuss the use of side-scan sonar (SSS). SSS provides an expansive 2D field of view, useful for large-scale searches of the seafloor, often-times with human operators monitoring incoming images. The field of view of a side-scan sonar is illustrated in Fig. 4.

Recently, deep learning methods have been applied to address the Automated Target Recognition (ATR) problem. However, much like other areas discussed in this paper, the lack of training data is a problem. Thus, the use of GANs is proposed [100, 101]. Furthermore, Yu et al. [102] employed the YOLO architecture applied to SSS images.

Place recognition has also been studied using SSS images. Larsson et al. [103] train a network to perform place recognition using triplet loss. Lastly, Xie et al. [104•] use a CNN to infer the missing dimension in SSS images.



**Fig. 4** An example of side-scan sonar imaging a survey area. Note the image mosaic in gray-scale, with terrain elevation shown in color. (©[2021] IEEE. Reprinted, with permission, from [101])

## Conclusions

Underwater perception has seen much work over the past five years. However, some areas stand out as requiring more work and new solutions. Firstly, when considering identifying and classifying objects in sensor data, there are few if any off-the-shelf solutions due to the lack of training data in this setting, especially for imaging sonar. While GANs and simulators have provided an outlet, evaluation is still challenging due to the lack of public benchmark data sets in a common setting (pool, reef, wreck, etc.). Critically, the lack of a common and robust object dataset has hampered efforts to incorporate semantic information into downstream processes.

When considering SLAM, while there are many innovative underwater SLAM solutions, many of them test on their own specially gathered data, and subsequently do not share this data in the public domain. Compared to other fields, such as ground robotics, where the KITTI dataset has proliferated, our community does not have an easy-to-use benchmark dataset. When considering the future in our community, especially in the sonar setting, more open-source code and public data are required to enable comparison, benchmarking, and replication of results.

While underwater state estimation has continued evolving over the past several years, many of these methods use sensor packages that often drastically increase the cost. A research avenue that is of large potential impact is low-cost perceptual and inertial sensor payloads running robust SLAM systems in 6-DOF. Such systems present the potential for larger-scale multi-robot deployments and an alternative to expensive inertial navigation systems that may be prone to failure in adverse conditions.

## Declarations

**Conflict of Interest** The authors declare no competing interests.

**Human and Animal Rights and Informed Consent** This article does not contain any studies with human or animal subjects performed by any of the authors.

## References

1. Hover FS, Eustice RM, Kim A, Englot B, Johannsson H, Kaess M, Leonard JJ. Advanced perception, navigation and planning for autonomous in-water ship hull inspection. *Int J Rob Res*. 2012;31(12):1445–64.
2. Milioto A, Vizzo I, Behley J, Stachniss C. Rangenet ++: Fast and accurate lidar semantic segmentation. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2019. p. 4213–20.
3. Fuchs LR, Gällström A, Folkesson J. Object recognition in forward looking sonar images using transfer learning. In: *IEEE/OES autonomous underwater vehicle workshop (AUV)*. 2018.
4. Valdenegro-Toro M, Preciado-Grijalva A, Wehbe B. Pre-trained models for sonar images. In: *OCEANS: San Diego – Porto*. 2021.
5. Liu D, Wang Y, Ji Y, Tsuchiya H, Yamashita A, Asama H. CycleGAN-based realistic image dataset generation for forward-looking sonar. *Adv Robot*. 2021;35(3–4):242–54.
6. Chen Y, Ma QM, Yu J, Chen T. Underwater acoustic object discrimination for few-shot learning. In: *4th international conference on mechanical, control and computer engineering (ICM-CCE)*, 2019. p. 430–4304.
7. Wang Y, Ji Y, Liu D, Tamura Y, Tsuchiya H, Yamashita A, Asama H. ACMarker: Acoustic camera-based fiducial marker system in underwater environment. *IEEE Robot Autom Lett*. 2020;5(4):5018–25.
8. Olson E. AprilTag: A robust and flexible visual fiducial system. In: *IEEE international conference on robotics and automation*, 2011. p. 3400–07.
9. Aykin MD, Negahdaripour S. Three-dimensional target reconstruction from multiple 2D forward-scan sonar views by space carving. *IEEE J Ocean Eng*. 2017;42(3):574–89.
10. Westman E, Gkioulekas I, Kaess M. A theory of fermat paths for 3D imaging sonar reconstruction. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2020. p. 5082–88.
11. Wang J, Shan T, Englot B. Underwater terrain reconstruction from forward-looking sonar imagery. In: *International conference on robotics and automation (ICRA)*. 2019. p. 3471–77.
12. Huang TA, Kaess M. Towards acoustic structure from motion for imaging sonar. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2015. p. 758–765.
13. DeBortoli R, Nicolai A, Li F, Hollinger GA. Real-time underwater 3D reconstruction using global context and active labeling. In: *IEEE international conference on robotics and automation (ICRA)*. 2018. p. 6204–11.
14. Westman E, Gkioulekas I, Kaess M. A volumetric albedo framework for 3D imaging sonar reconstruction. In: *IEEE international conference on robotics and automation (ICRA)*. 2020. p. 9645–51.
15. Guerneve T, Subr K, Petillot Y. Three-dimensional reconstruction of underwater objects using wide-aperture imaging sonar. *J Field Robot*. 2018;35(6):890–905.
16. Westman E, Kaess M. Wide aperture imaging sonar reconstruction using generative models. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2019. p. 8067–8074. **This paper represents a non-learning approach to the 3D reconstruction problem using a single imaging sonar. Importantly, the authors demonstrate results on real-world data with widely used hardware.**
17. DeBortoli R, Li F, Hollinger GA. ElevateNet: A convolutional neural network for estimating the missing dimension in 2D underwater sonar images. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2019. p. 8040–47.
18. Wang Y, Ji Y, Liu D, Tsuchiya H, Yamashita A, Asama H. Elevation angle estimation in 2D acoustic images using pseudo front view. *IEEE Robot Autom Lett*. 2021;6(2):1535–42.
19. McConnell J, Martin JD, Englot B. Fusing concurrent orthogonal wide-aperture sonar images for dense underwater 3D reconstruction. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2020. p. 1653–60.
20. Negahdaripour S. Analyzing epipolar geometry of 2D forward-scan sonar stereo for matching and 3D reconstruction. In: *OCEANS MTS/IEEE charleston*. 2018.
21. Negahdaripour S. Application of forward-scan sonar stereo for 3D scene reconstruction. *IEEE J Ocean Eng*. 2020;45(2):547–62.

22. McConnell J, Englot B. Predictive 3D sonar mapping of underwater environments via object-specific Bayesian inference. In: IEEE international conference on robotics and automation (ICRA). 2021. p 6761–67.
23. Franchi M, Ridolfi A, Allotta B. Underwater navigation with 2D forward looking sonar: An adaptive unscented Kalman filter-based strategy for AUVs. *J Field Robot*. 2021;38(3):355–85.
24. Henson BT, Zakharov YV. Attitude-trajectory estimation for forward-looking multibeam sonar based on acoustic image registration. *IEEE J Ocean Eng*. 2019;44(3):753–66.
25. Almanza-Medina JE, Henson B, Zakharov YV. Sonar FoV segmentation for motion estimation using DL networks. *IEEE Access*. 2022;10:25591–604.
26. Song S, Michael Herrmann J, Si B, Liu K, Feng X. Two-dimensional forward-looking sonar image registration by maximization of peripheral mutual information. *Int J Adv Robot Syst*. 14(6).
27. Santos MM, Zaffari GB, Ribeiro POCS, Drews-Jr PLJ, Botelho SSC. Underwater place recognition using forward-looking sonar images: A topological approach. *J Field Robot*. 2019;36(2):355–69.
28. Ribeiro POCS, dos Santos MM, Drews PLJ, Botelho SSC, Longaray LM, Giacomo GG, Pias MR. Underwater place recognition in unknown environments with triplet based acoustic image retrieval. In: IEEE international conference on machine learning and applications (ICMLA). 2018. p. 524–529.
29. Westman E, Hinduja A, Kaess M. Feature-based SLAM for imaging sonar with under-constrained landmarks. In: IEEE international conference on robotics and automation (ICRA). 2018. p. 3629–36.
30. Li J, Kaess M, Eustice RM, Johnson-Roberson M. Pose-graph SLAM using forward-looking sonar. *IEEE Robot Autom Lett*. 2018;3(3):2330–7.
31. Wang J, Chen F, Huang Y, McConnell J, Shan T, Englot B. Virtual maps for autonomous exploration of cluttered underwater environments. *IEEE J Ocean Eng*. 2022.
32. Teixeira PV, Fourie D, Kaess M, Leonard JJ. Dense, sonar-based reconstruction of underwater scenes. In: IEEE/RSJ international conference on intelligent robots and systems (IROS). 2019. p 8060–66.
33. Hinduja A, Ho B-J, Kaess M. Degeneracy-aware factors with applications to underwater SLAM. In: IEEE/RSJ international conference on intelligent robots and systems (IROS), 2019. p. 1293–99.
34. Xu Y, Zheng R, Zhang S, Liu M. Robust inertial-aided underwater localization based on imaging sonar keyframes. *IEEE Trans Instrum Meas*. 2022;71:1–12.
35. Dos Santos MM, De Giacomo GG, Drews-Jr PLJ, Botelho SSC. Cross-view and cross-domain underwater localization based on optical aerial and acoustic underwater images. *IEEE Robot Autom Lett*. 2022;7(2):4969–74.
36. McConnell J, Chen F, Englot B. Overhead image factors for underwater sonar-based SLAM. *IEEE Robot Autom Lett*. 2022;7(2):4901–8.
37. Joshi B, Rahman S, Kalaitzakis M, Cain B, Johnson J, Xanthidis M, Karapetyan N, Hernandez A, Li AQ, Vitzilaios N, Rekleitis I. Experimental comparison of open source visual-inertial-based state estimation algorithms in the underwater domain. In: IEEE/RSJ international conference on intelligent robots and systems (IROS). 2019. p. 7227–33.
38. Ancuti CO, Ancuti C, De Vleeschouwer C, Garcia R. Locally adaptive color correction for underwater image dehazing and matching. In: IEEE conference on computer vision and pattern recognition workshops (CVPRW). 2017. p. 997–1005.
39. Skinner KA, Johnson-Roberson M. Underwater image dehazing with a light field camera. In: IEEE conference on computer vision and pattern recognition workshops (CVPRW). 2017. p. 1775–82.
40. Skinner KA, Iscar E, Johnson-Roberson M. Automatic color correction for 3D reconstruction of underwater scenes. In: IEEE international conference on robotics and automation (ICRA). 2017. p. 5140–47.
41. Cho Y, Kim A. Visibility enhancement for underwater visual SLAM based on underwater light scattering model. In: IEEE international conference on robotics and automation (ICRA). 2017. p. 710–717.
42. Berman D, Levy D, Avidan S, Treibitz T. Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE Trans Pattern Anal Mach Intell*. 2021;43(8):2822–37.
43. Marques TP, Albu AB, L2uwe: A framework for the efficient enhancement of low-light underwater images using local contrast and multi-scale fusion. In: IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW). 2020. p. 2286–95.
44. Roznere M, Li AQ. Real-time model-based image color correction for underwater robots. In: IEEE/RSJ international conference on intelligent robots and systems (IROS). 2019. p. 7191–96.
45. Li J, Skinner KA, Eustice RM, Johnson-Roberson M. Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robot Autom Lett*. 2018;3(1):387–94.
46. Fabbri C, Islam MdJ, Sattar J. Enhancing underwater imagery using generative adversarial networks. In: IEEE international conference on robotics and automation (ICRA). 2018. p. 7159–65.
47. Islam MdJ, Xia Y, Sattar J. Fast underwater image enhancement for improved visual perception. *IEEE Robot Autom Lett*. 2020;5(2):3227–34.
48. Hu K, Zhang Y, Weng C, Wang P, Deng Z, Liu Y. An underwater image enhancement algorithm based on generative adversarial network and natural image quality evaluation index. *J Mar Sci Eng*. 2021;9(7).
49. Zhou Y, Yan K, Li X. Underwater image enhancement via physical-feedback adversarial transfer learning. *IEEE J Ocean Eng*. 2022;47(1):76–87.
50. Park J, Han DK, Ko H. Adaptive weighted multi-discriminator cycleGAN for underwater image enhancement. *J Mar Sci Eng*. 2019;7(7).
51. Modasshir Md, Rekleitis I. Enhancing coral reef monitoring utilizing a deep semi-supervised learning approach. In: IEEE international conference on robotics and automation (ICRA). 2020. p. 1874–80.
52. Joshi B, Modasshir Md, Manderson T, Damron H, Xanthidis M, Li AQ, Rekleitis I, Dudek G. DeepURL: Deep pose estimation framework for underwater relative localization. In: IEEE/RSJ international conference on intelligent robots and systems (IROS). 2020. p. 1777–84.
53. Dayoub F, Dunbabin M, Corke P. Robotic detection and tracking of crown-of-thorns starfish. In: IEEE/RSJ international conference on intelligent robots and systems (IROS). 2015. p. 1921–28.
54. Rimavicius T, Gelzinis A. A comparison of the deep learning methods for solving seafloor image classification task. In: Damaševičius R, Mikašytė V, editors. Information and software technologies. Cham; Springer International Publishing; 2017. p. 442–53.
55. Xu W, Matzner S. Underwater fish detection using deep learning for water power applications. In: International conference on computational science and computational intelligence (CSCI). 2018. p. 313–318.
56. Garcia R, Prados R, Quintana J, Tempelaar A, Gracias N, Rosen S, Vågstøl H, Løvfall K. Automatic segmentation of fish



- using deep learning with application to fish size measurement. *ICES J Mar Sci.* 2019;77(4):1354–66.
57. Chen Q, Beijbom O, Chan S, Bouwmeester J, Kriegman D. A new deep learning engine for CoralNet. In: *IEEE/CVF international conference on computer vision workshops (ICCVW)*. 2021. p. 3686–95.
  58. Levy D, Belfer Y, Osherov E, Bigal E, Scheinin AP, Nativ H, Tchernov D, Treibitz T. Automated analysis of marine video with limited data. In: *IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)*. 2018. p. 1466–68.
  59. O’Byrne M, Pakrashi V, Schoefs F, Ghosh B. Semantic segmentation of underwater imagery using deep networks trained on synthetic imagery. *J Mar Sci Eng.* 2018;6(3).
  60. Ochal M, Vazquez J, Petillot Y, Wang S. A comparison of few-shot learning methods for underwater optical and sonar image classification. In: *Global Oceans: Singapore – U.S. Gulf Coast*. 2020.
  61. Yamada T, Massot-Campos M, Prügel-Bennett A, Williams SB, Pizarro O, Thornton B. Leveraging metadata in representation learning with georeferenced seafloor imagery. *IEEE Robot Autom Lett.* 2021;6(4):7815–22.
  62. Islam MdJ, Sattar J. Mixed-domain biological motion tracking for underwater human-robot interaction. In: *IEEE international conference on robotics and automation (ICRA)*. 2017. p. 4457–64.
  63. Fulton M, Hong J, Sattar J. Using monocular vision and human body priors for AUVs to autonomously approach divers. In: *IEEE international conference on robotics and automation (ICRA)*. 2022. p. 1076–82.
  64. Shkurti F, Chang W-D, Henderson P, Islam MdJ, Higuera JCG, Li J, Manderson T, Xu A, Dudek G, Sattar J. Underwater multi-robot conveying using visual tracking by detection. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2017. p. 4189–96.
  65. Girdhar Y, Giguère P, Dudek G. Autonomous adaptive exploration using realtime online spatiotemporal topic modeling. *Int J Rob Res.* 2014;33(4):645–57.
  66. Kalmbach A, Hoeberechts M, Albu AB, Glotin H, Paris S, Girdhar Y. Learning deep-sea substrate types with visual topic models. In: *IEEE winter conference on applications of computer vision (WACV)*. 2016.
  67. Mur-Artal R, Montiel JMM, Tardós JD. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Trans Robot.* 2015;31(5):1147–63.
  68. Mur-Artal R, Tardós JD. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Trans Robot.* 2017;33(5):1255–62.
  69. Hidalgo F, Kahlefeldt C, Bräunl T. Monocular ORB-SLAM application in underwater scenarios. In: *OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO)*. 2018.
  70. Hidalgo F. ORBSLAM2 and point cloud processing towards autonomous underwater robot navigation. In: *Global Oceans: Singapore – U.S. Gulf Coast*. 2020.
  71. Li AQ, Coskun A, Doherty SM, Ghasemlou S, Jagtap AS, Modasshir MD, Rahman S, Singh A, Xanthidis M, O’Kane JM, Rekleitis I. Experimental comparison of open source vision based state estimation algorithms. In: *Proc international symposium on experimental robotics*. 2016.
  72. Zhang J, Ila V, Kneip L. Robust visual odometry in underwater environment. In: *OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO)*. 2018.
  73. Ferrera M, Moras J, Trouvé-Peloux P, Creuze V. Real-time monocular visual odometry for turbid and dynamic underwater environments. *Sensors* 2019;19(3).
  74. Xu S, Luczynski T, Willners JS, Hong Z, Zhang K, Petillot YR, Wang S. Underwater visual acoustic SLAM with extrinsic calibration. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2021. p. 7647–52.
  75. Vargas E, Scona R, Willners JS, Luczynski T, Cao Y, Wang S, Petillot YR. Robust underwater visual SLAM fusing acoustic sensing. In: *IEEE international conference on robotics and automation (ICRA)*. 2021. p. 2140–46.
  76. Rahman S, Li AQ, Rekleitis I. Sonar visual inertial SLAM of underwater structures. In: *IEEE international conference on robotics and automation (ICRA)*. 2018. p. 5190–96.
  77. Roznere M, Li AQ. Underwater monocular image depth estimation using single-beam echosounder. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2020. p. 1785–90.
  78. Rahman S, Li AQ, Rekleitis I. SVIn2: An underwater SLAM system using sonar, visual, inertial, and depth sensor. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2019. p. 1861–68.
  79. Hu C, Zhu S, Liang Y, Mu Z, Song W. Visual-pressure fusion for underwater robot localization with online initialization. *IEEE Robot Autom Lett.* 2021;6(4):8426–33.
  80. Rahman S, Li AQ, Rekleitis I. Contour based reconstruction of underwater structures using sonar, visual, inertial, and depth sensor. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2019. p. 8054–59.
  81. Weidner N, Rahman S, Li AQ, Rekleitis I. Underwater cave mapping using stereo vision. In: *IEEE international conference on robotics and automation (ICRA)*. 2017. p. 5709–15.
  82. • Joshi B, Xanthidis M, Rahman S, Rekleitis I. High definition, inexpensive, underwater mapping. *IEEE International Conference on Robotics and Automation (ICRA)*, 2022. pp 1113–1121. **This paper presents a large-scale underwater visual SLAM solution using minimal hardware, in this case a GoPro camera. The results demonstrate robustness across a wide variety of environmental conditions.**
  83. Bosch J, Istenič K, Gracias N, Garcia R, Ridao P. Omnidirectional multicamera video stitching using depth maps. *IEEE J Ocean Eng.* 2020;45(4):1337–52.
  84. Xanthidis M, Joshi B, Karapetyan N, Roznere M, Wang W, Johnson J, Li AQ, Casana J, Mordohai P, Nelakuditi S, Rekleitis I. Towards multi-robot shipwreck mapping. *Advanced Marine Robotics Technical Committee Workshop on Active Perception at IEEE International Conference on Robotics and Automation (ICRA)*. 2021.
  85. Suresh S, Westman E, Kaess M. Through-water stereo SLAM with refraction correction for auv localization. *IEEE Robot Autom Lett.* 2019;4(2):692–9.
  86. Torroba I, Bore N, Folkesson J. Towards autonomous industrial-scale bathymetric surveying. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2019. p. 6377–82.
  87. Guerneve T, Subr K, Petillot Y. Underwater 3D structures as semantic landmarks in sonar mapping. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2017. p. 614–619.
  88. Torroba I, Bore N, Folkesson J. A comparison of submap registration methods for multibeam bathymetric mapping. In: *IEEE/OES autonomous underwater vehicle workshop (AUV)*. 2018.
  89. Hitchcox T, Forbes JR. A point cloud registration pipeline using gaussian process regression for bathymetric SLAM. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2020. p. 4615–22.
  90. Jung J, Park J, Choi J, Choi H-T. Bathymetric pose graph optimization with regularized submap matching. *IEEE Access.* 2022;10:31155–64.
  91. Sprague C, Torroba I, Bore N, Folkesson J. PointNetKL: Deep inference for GICP covariance estimation in bathymetric SLAM. *IEEE Robot Autom Lett.* 2020;5(3):4078–85.

92. Campos R, Garcia R. Surface meshing of underwater maps from highly defective point sets. *J Field Robot*. 2018;35(4):491–515.
93. Bore N, Torroba I, Folkesson J. Sparse Gaussian process SLAM, storage and filtering for auv multibeam bathymetry. In: *IEEE/OES autonomous underwater vehicle workshop (AUV)*. 2018.
94. Zhang Q, Li Y, Ma T, Cong Z, Zhang W. Bathymetric particle filter SLAM based on mean trajectory map representation. *IEEE Access*. 2021;9:71725–36.
95. Teng M, Ye L, Yuxin Z, Yanqing J, Qianyi Z, Pascoal AM. Efficient bathymetric SLAM with invalid loop closure identification. *IEEE ASME Trans Mechatron*. 2021;26(5):2570–80.
96. Teixeira PV, Kaess M, Hover FS, Leonard JJ. Underwater inspection using sonar-based volumetric submaps. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. 2016. p. 4288–95.
97. Palomer A, Ridaio P, Forest J, Ribas D. Underwater laser scanner: Ray-based model and calibration. *IEEE ASME Trans Mechatron*. 2019;24(5):1986–97.
98. Palomer A, Ridaio P, Ribas D. Inspection of an underwater structure using point-cloud SLAM with an auv and a laser scanner. *J Field Robot*. 2019;36(8):1333–44.
99. Norgren P, Skjetne R. A multibeam-based SLAM algorithm for iceberg mapping using AUVs. *IEEE Access*. 2018;6:26318–37.
100. Karjalainen AI, Mitchell R, Vazquez J. Training and validation of automatic target recognition systems using generative adversarial networks. In: *Sensor signal processing for defence conference (SSPD)*. 2019.
101. Bore N, Folkesson J. Modeling and simulation of sidescan using conditional generative adversarial network. *IEEE J Ocean Eng*. 2021;46(1):195–205.
102. Yu Y, Zhao J, Gong Q, Huang C, Zheng G, Ma J. Real-time underwater maritime object detection in side-scan sonar images based on transformer-YOLOv5. *Remote Sens*. 2021;13(18).
103. Larsson M, Bore N, Folkesson J. Latent space metric learning for sidescan sonar place recognition. In: *IEEE/OES autonomous underwater vehicles symposium (AUV)*. 2020.
104. • Xie Y, Bore N, Folkesson J. Inferring depth contours from side-scan sonar using convolutional neural nets. *IET Radar, Sonar & Navigation* 14(2):328–334. **This paper uses side-scan sonar to infer the bathymetric contours of undersea terrain. Critically though, this work is enabled by convolutional neural nets, a method seldom employed in this setting due to lack of training data.**

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.