

Sonar-Based Deep Learning in Underwater Robotics: Overview, Robustness, and Challenges

Martin Aubard , Ana Madureira , *Senior Member, IEEE*, Luís Teixeira , and José Pinto 

Abstract—With the growing interest in underwater exploration and monitoring, autonomous underwater vehicles have become essential. The recent interest in onboard deep learning (DL) has advanced real-time environmental interaction capabilities relying on efficient and accurate vision-based DL models. However, the predominant use of sonar in underwater environments, characterized by limited training data and inherent noise, poses challenges to model robustness. This autonomy improvement raises safety concerns for deploying such models during underwater operations, potentially leading to hazardous situations. This article aims to provide the first comprehensive overview of sonar-based DL under the scope of robustness. It studies sonar-based DL perception task models, such as classification, object detection, segmentation, and simultaneous localization and mapping. Furthermore, this article systematizes sonar-based state-of-the-art data sets, simulators, and robustness methods, such as neural network verification, out-of-distribution, and adversarial attacks. This article highlights the lack of robustness in sonar-based DL research and suggests future research pathways, notably establishing a baseline sonar-based data set and bridging the simulation-to-reality gap.

Index Terms—Autonomous underwater vehicle (AUV), data sets, deep learning (DL), robustness, sonar-based.

I. INTRODUCTION

IN Recent decades, the world's oceans have become a focal point for numerous subjects of interest to scientists, industries, and military organizations, including underwater archaeology [1], maritime exploration [2], transportation logistics, renewable energy initiatives [3], and military applications [4]. These applications share a common challenge: conducting safe surveys, explorations, or data collection in the often unpredictable and hazardous underwater environment. Autonomous underwater vehicles (AUVs), used to collect data and perform operations underwater, have emerged as indispensable tools in addressing this challenge. AUVs come equipped with various sensors and instruments collecting underwater data, including

measuring temperature, salinity, currents, and cataloging marine species and seabed structures. However, the underwater environment presents challenges for navigation and environmental understanding due to its inherent physical uncertainties. Radio waves are rapidly absorbed by water, restricting the use of radar and global positioning system (GPS) sensors. Underwater vision is limited by reduced luminosity, turbidity, and the necessity for proximity to capture clear imagery. These challenges affect traditional vision sensors, such as red–green–blue (RGB) and grayscale cameras, and light detection and ranging (LiDAR) systems due to the significant scattering and absorption of light. On the other hand, sonar sensors, relying on sound waves that propagate efficiently and consistently underwater, provide a viable alternative for underwater sensing and navigation.

A. Sonars

Sonars, classified as active or passive, offer distinct underwater exploration and monitoring methodologies. Passive sonar captures the ambient underwater environment without emitting any signals, which is primordial for furtive surveillance. In contrast, active sonar emits sound pulses reflecting on underwater objects, marine life, or the seabed. By analyzing the return signals, we can determine object distances (time between emission and reception) and identify object types (variations in sound intensity), primordial for mapping and navigation (e.g., obstacle avoidance), providing comprehensive underwater environment information. In addition, the choice of operating frequency affects sonar performance since it directly influences the system's range and resolution as well as the level of penetration into soft bottoms, such as mud. Lower frequencies provide higher propagation distances but at the cost of detail, whereas high frequencies offer superior resolution but with reduced operating range. Thus, sonar data collection requires a tradeoff between range and resolution. Synthetic aperture sonar (SAS) has recently addressed this limitation by providing higher resolution without sacrificing ranges, making it a promising technology for underwater exploration.

Active sonars can also be categorized by configuration, such as single-beam and multibeam types. Single-beam offers a singular sound measurement for determining altitude, aiding navigation tasks (e.g., bottom tracking). In contrast, multibeam sonar, including side-scan sonar (SSS) and multibeam echo sonars (MBES), are designed to simultaneously capture a broader range of data points. MBES provides a bathymetric mapping from beneath the vehicle, whereas SSS focuses on

Received 10 April 2024; revised 3 December 2024; accepted 9 December 2024. Date of publication 1 April 2025; date of current version 16 July 2025. This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie under Grant 956200. (Corresponding author: Martin Aubard.)

Associate Editor: T. Sæbø.

Martin Aubard and José Pinto are with the OceanScan Marine Systems and Technology, 4450-718 Matosinhos, Portugal (e-mail: maubard@oceanscan-mst.com; zepinto@oceanscan-mst.com).

Ana Madureira is with the INESC INOV-Lab and ISRC (ISEP/P.PORTO), 4249-015 Porto, Portugal (e-mail: amd@isep.ipp.pt).

Luís Teixeira is with the Faculty of Engineering, University of Porto (FEUP), 4200-465 Porto, Portugal (e-mail: luisft@fe.up.pt).

For further information, please visit <https://remaro.eu>.

Digital Object Identifier 10.1109/JOE.2025.3531933

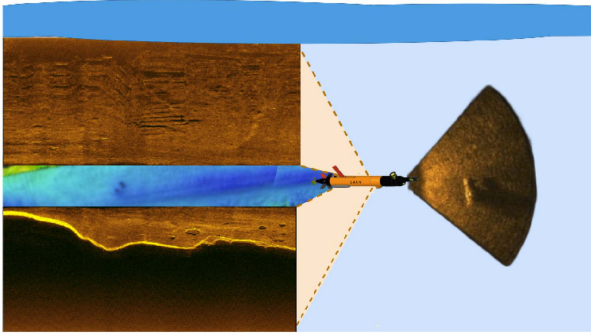


Fig. 1. Sonar perception with SSS [5] and FLS [6]. This figure represents the setup and visual information from the multibeam echo sonar (MBES), SSS, and FLS. It shows that while the SSS provides information on past data from the port to the starboard, the FLS gives current information from the front of the AUV and the MBES from beneath the vehicle.

generating high-resolution seafloor imagery on both sides of the vehicle (port and starboard). In addition, forward-looking sonar (FLS), whether single-beam or multibeam, maps the area directly ahead of the vehicle. A single-beam FLS might only provide distance measurements to upcoming obstacles, whereas a multibeam FLS offers a detailed representation of the area in front of the vehicle, enhancing obstacle avoidance and navigation. Fig. 1 represents three sonar sensors: an SSS on both port and starboard sides, an MBES beneath, and a multibeam FLS in front of the vehicle. Due to the position of the SSS, the usual SSS image lacks information between the two transducers, which is called the nadir gap; however, in this graphical representation, the MBES fills the nadir gap.

B. Deep Learning (DL) and Robustness

Artificial intelligence (AI) has emerged as a crucial solution to enhance the performance and capabilities of AUVs. AI's ability to process and interpret large data sets, identify patterns, and make informed predictions can significantly improve AUV navigation, collision avoidance, and real-time decision-making [7]. Over the past decade, machine learning (ML) and DL techniques have been applied to improve feature detection in data collected by AUVs, such as identifying mines [8] or shipwrecks [9]. Traditionally, ML and DL methods have been used in postprocessing vision-based data, where collected data are processed offline to detect features. This approach reduces the time needed for target identification, which typically requires domain expertise. However, due to the time-consuming nature of offline processing, AUVs cannot interact with potential underwater objects while surveying. Recently, the focus has shifted toward integrating DL algorithms directly onboard AUVs to enable real-time interaction with the underwater environment, enhancing responsiveness to environmental changes and improving the autonomy of AUVs [10]. This real-time approach enables AUVs to adapt their navigation dynamically in response to detected objects or obstacles, streamlining survey times and improving data quality [11]. However, relying on DL models for interpreting environmental data in real time raises essential questions about the reliability of sonar-based DL outputs.

The central question, “*How can we rely on real-time sonar-based DL models?*” raises several related issues regarding the robustness of DL models and the true meaning of robustness. Robustness refers to the ability of AI systems to handle errors or inconsistencies during their operation [12]. It is a subcategory of the broader concept of Safe AI, including Explainable AI (XAI), interpretability, privacy, and security. Safe AI aims to ensure the reliability of algorithms, vehicle safety, and the safety of surrounding environments [13]. Due to their opacity, DL algorithms are often considered opaque systems [14], as their reasoning process is unclear, requiring measures to ensure expected performance while mitigating unintended consequences and potential harm [15]. The challenge lies in balancing the benefits of AI with managing the risks of misuse or malfunction. XAI [16] provides insights into the training and implementation processes, helping users understand AI behaviors, identify issues, and improve system architecture. The EU’s General Data Protection Regulation [17] enforces a “right to explanation,” allowing individuals affected by AI decisions to request an explanation of how those decisions were made. Interpretability Gilpin et al. [18] further supported understanding the model structure, aiding in reliability assessments, diagnostics, and corrections. While Safe AI has seen significant progress in terrestrial applications, such as autonomous driving, there has been less focus on ensuring the safety of underwater perception systems, particularly sonar-based systems. Underwater environments pose unique challenges, including unreliable GPS positioning, poor visibility for optical sensors, noisy sonar data, strong currents, and constantly changing conditions. Hence, implementing DL onboard AUVs raises questions about vehicle safety, where for the AUV itself, safety means ensuring that it does not collide due to unreliable model behavior, does not get lost, and collects data accurately.

C. Motivations

DL models’ robustness examination aims to consistently ascertain their ability to make accurate predictions under all conditions. Furthermore, it is essential to understand the factors that could lead to erroneous sonar-based DL outputs. The data-driven nature of DL models requires relying on the integrity and relevance of training data, which brings up the following critical considerations: 1) Adequate data volume in the training set is essential for comprehensive learning and recognition abilities upon deployment. 2) The relevance of data to the operational environment is vital, prompting questions about what constitutes the resemblance of data to the deployment context regarding environmental and sensor data distributions. Those considerations are even more critical for sonar-based data sets due to their open-source accessibility limitation. Thus, if the available data prove insufficient, operators face the following alternatives: 1) collecting more data (time and cost-consuming); 2) utilizing simulators for data generation, requiring high fidelity to real-world conditions; or 3) leveraging generative models to increase the data set volume. Nevertheless, even with precise data, model-induced errors remain possible, underscoring the significance of neural network verification as a discipline

seeking to mathematically validate model reliability within defined noise thresholds. However, the current state of neural network verification tools reveals limitations in applicability across tasks and constraints related to model size or activation functions. Adversarial attack and out-of-distribution (OOD) detection emerge, focusing on identifying and mitigating noise that could deceive the DL model, emphasizing the importance of incorporating adversarial attack defenses as part of the pre-deployment model evaluation process. Sonar data is affected by various underwater noise sources, including self-noise (system-generated noise), multipath reflections (reflected signals from the ocean surface and bottom), interference from other sonars (external interference), ambient noise from marine life and oceanic activity (environmental noise), as well as speckle noise inherent in sonar imaging (imaging artifact) [19]. Furthermore, the disparate distribution between sonar brands can also mislead DL output. Thus, those pre-deployment steps remain essential for reducing unpredicted noises potentially encountered during a mission.

This document aims to guide practitioners in deploying DL models for sonar on AUVs, emphasizing model robustness considering the uncertainties inherent in sonar data. Such model presents unique challenges that are less documented than traditional optical data, requiring specific focused handling uncertainties and ensuring reliable model performance. The following sections explore the complexities of utilizing sonar-based DL in underwater robotics and provide a comprehensive overview of improving the robustness of these applications. The objective is to demonstrate how leveraging sonar-based DL can significantly enhance AUV capabilities while ensuring the safety and success of underwater missions. Section II reviews existing literature, positioning this work as a pioneering effort in addressing the robustness of sonar-based DL models. Section III discusses essential sonar-based tasks enabled by DL, including classification, object detection, segmentation, and simultaneous localization and mapping (SLAM). Section IV explores the current state of robustness in sonar-based applications, highlighting state-of-the-art open-source data sets, simulators for underwater sonar environments, and synthetic data generation techniques. It also emphasizes the emerging field of neural network verification in sonar contexts and discusses methodologies for OOD detection, adversarial attacks, and uncertainty quantification. Finally, Section V concludes this article.

II. SONAR-BASED DL-RELATED SURVEYS

The exploration of DL in sonar imagery for underwater applications is captured through the following insightful review papers, each delving into different aspects and methodologies in the field of sonar-based DL. First, D. Neupane and Seok [20] broaden the discussion by emphasizing the importance of sonars in underwater object detection and the challenges posed by the lack of accessible data sets. The article provides a structured analysis that spans sonar principles, the utility of DL over traditional AI and ML-based methods, and a detailed examination of data sets, preprocessing technologies, and DL architectures for automatic target recognition (ATR), highlighting the intrinsic

challenges of sonar data, such as nonhomogeneous resolution and acoustic shadowing. The review underscores the need for high-quality, shared data sets to advance the field. It recommends detailed documentation of data sets, simulators for data generation, and a more nuanced approach to data augmentation for enhancing ATR in sonar-based studies. Steiniger et al. [21] distinguishes itself by concentrating on SAS and SSS data, areas not extensively covered in prior reviews. Acknowledging the foundational work in [20], this article narrows its focus to SSS and SAS, excluding FLS and MBES from its analysis. It rigorously compares simple convolutional neural networks (CNN) algorithms across various tasks, such as feature extraction, classification, object detection, and segmentation, highlighting a comprehensive examination of over 60 publications related to SSS object detection. This article critically notes the absence of open-source SSS image data sets, which hampers the comparability of research outcomes, and suggests data augmentation and the generation of simulated data using generative adversarial network (GAN) models as potential remedies. However, it also highlights the scarcity of shared data sets and models that could facilitate broader research collaboration. Khan et al. [22] presented a selection of cutting-edge algorithms developed over the past seven years, filling gaps that have yet to be addressed by existing surveys and catalogs the applications for underwater object detection. It provides a succinct overview of architectures, including a comparative analysis of various you only look once (YOLO) versions and other CNN-based models, termed “ConVNNs.” This article calls for a more diversified and balanced data set, exploring deep transformer models and developing hybrid detection techniques, among other future directions. Teng and Zhao [23] conducted a comprehensive survey of DL-based detection methods for mines and manmade targets using underwater RGB and sonar imagery. They thoroughly examine the various types of noise encountered in sonar images and discuss methods to mitigate these noises during data processing. In addition, they evaluate several SOA models using identical data sets to benchmark performance for RGB image analysis. Tian et al. [24] work is the pioneering and sole survey paper on sonar segmentation. It meticulously reviews existing sonar segmentation literature, outlines the current challenges in the field, and proposes 12 research directions to advance sonar segmentation studies. Expanding the scope of our analysis, we study sonar applications in fish identification and shoreline surveillance. Yassir et al. [25] focused on fish classification and segmentation, comparing SOA DL models. In contrast, Chai et al. [26] provided an extensive survey covering fish classification, detection, segmentation, and denoising of sonar images. However, the comparison by Yassir et al. [25] across models trained on diverse data sets complicates definitive conclusions due to the potential variability in data set characteristics. Domingo et al. [27] explored DL methods for shoreline surveillance by classifying underwater vessels using passive sonar. This review highlights the adaptability of DL methods to several aspects of underwater exploration and monitoring, showcasing the breadth of potential applications for sonar technology. In synthesizing these reviews, it becomes evident that while significant strides have been made in applying

TABLE I
COMPARATIVE ANALYSIS BETWEEN THE SURVEYS ON SONAR-BASED DL

Survey	Year	Topics	Datasets	Simu.	Models	Sensors	Tasks	Robust.
Neupane et al. [20]	2020	ATR	✓	✓	✓	SSS, FLS	C, D, S	✗
Steiniger et al. [21]	2022	ATR	✗	✗	✓	SSS, SAS	C, D, S	UQ *
Khan et al. [22]	2024	ATR	✗	✗	✓	SSS, FLS	C, D	✗
Teng et al. [23]	2020	ATR	✗	✗	✓	SSS, FLS, RGB	C, D	✗
Tian et al. [24]	2020	Sonar Segmentation	✗	✗	✓	SSS, FLS	S	✗
Yassir et al. [25]	2023	Fish Identification	✗	✗	✓	MBES	C, S	✗
Chai et al. [26]	2023	Fish Identification	✓	✓*	✓	SSS, SAS, FLS	C, D, S	✗
Domingo et al. [27]	2022	Shoreline Surveillance	✓	✗	✓	Passive Sonar	C	✗
Ours	2024	ATR	✓	✓	✓	SSS, FLS	C,D,S,SI	OOD,AA,UQ

This table compares the surveys on the Sonar-based deep learning by analyzing the date of publication (year), the main topic of the survey (Topics), if they enumerate and compare SAO datasets (Datasets), if they enumerate and compare underwater simulator for Sonar-based image generation (Simu.), if they enumerate and compare SAO DL models (Models), the type of sensors used (Sensors), the type of tasks realized by the DL models (Tasks), where *C* stands for classification, *D* for detection, *S* for segmentation, and *SI* for SLAM, and finally the robustness method tackle in the surveys (Robust.), where *OOD* stands for out-of-distribution, *AA* for adversarial attack, and *UQ* for uncertainty quantization. The * symbol translates a brief mention of the topics in the paper.

DL to sonar imagery for underwater detection, the field still faces substantial limitations, such as the critical need for open-source data sets, a more granular understanding, and explanation of DL models tailored to sonar data. However, because of the need for a sonar baseline data set, comparing all the DL models does not give concrete insight into which one is the most suited for sonar images. Table I highlights a detailed comparative analysis of surveys on sonar-based DL models. This comparison underscores the singularity of our contribution relative to previous works in the scope of sonar-based DL. Our survey is the first to provide an in-depth comparison of current sonar open-source data sets and simulators and to address the robustnesses of sonar-based DL models, delving into critical areas such as OOD detection, adversarial attacks, and uncertainty quantification. This comprehensive approach aims to provide valuable insights into the current SOA of sonar-based DL in terms of models, data sets, simulators, and methods to improve the robustness of the DL prediction.

III. SONAR-BASED DL PERCEPTION

This section describes the principal tasks achievable by applying sonar-based DL onboard AUVs, explicitly focusing on classification, object detection, segmentation, and SLAM. This discussion includes a historical overview of the model developments and highlights the SOA models for these tasks. In this article, we consciously abstain from directly comparing the performance of published sonar-based DL models. This decision is twofold; first, as indicated in Section II, drawing direct comparisons across models is challenging due to the utilization of disparate data sets collected under varying conditions with different sonar equipment, most of which are not publicly accessible. Second, prior survey papers, including [20], [21], and [22], have already undertaken comprehensive comparisons of sonar-based DL models published respectively in 2020, 2022, and 2024. We direct interested readers to these surveys for in-depth comparisons. Conversely, our section's objective is to furnish readers with an encompassing perspective on the assortment of models deployable for sonar-based DL tasks, such

as classification, object detection, segmentation, and SLAM, offering insights into the evolving landscape of DL.

A. Classification and Object Detection

Classification and object detection integrated into underwater vehicles effectively enhance the situational awareness and navigation in a complex underwater environment. Classification algorithms allow for categorizing underwater objects or features into predefined classes, which is essential for tasks, such as marine life monitoring, habitat mapping, and underwater archaeology. Object detection, on the other hand, extends this capability by identifying these categories and locating and tracking objects within the sonar imagery, which is crucial for obstacle avoidance, target tracking, and detailed environmental assessment. The history of DL classification has seen significant evolution, especially with the advent of CNNs. Since the breakthrough achievement of AlexNet in 2012 [28], CNNs have become a staple in computer vision tasks, including object detection, segmentation, and classification. They leverage convolutional layers to reduce image size and enhance pattern recognition capabilities effectively. AlexNet, a successor to LeNet [29], introduced more filter layers and demonstrated remarkable classification abilities across over a thousand classes using RGB images. ResNet [30] further revolutionized DL by introducing residual networks with “skip connections” to combat the vanishing gradient problem, improving accuracy even as network depth increased. In the domain of object detection, the last decade has witnessed remarkable progress. It encompasses crucial tasks, such as object localization and classification within images, facilitated by both one- and two-stage detectors. Two-stage detectors excel in accuracy by separately addressing localization and classification, but often at the cost of efficiency. The inception of R-CNN in 2014 [31], followed by advancements, such as Fast [32] and Faster R-CNN [33], highlighted the efforts of optimizing the balance between accuracy and efficiency, mainly through innovations like Region of Interest pooling and separate networks for region proposal predictions. The emergence of one-stage detectors, notably through the YOLO [34] series,

underscored a pivotal shift toward real-time object detection by merging localization and classification tasks. Despite initial trade-offs in accuracy, subsequent iterations of YOLO and single shot detector (SSD) [35] have improved the efficiency and effectiveness of object detection models. These models, particularly YOLOv4 and its successors, have set new benchmarks in real-time detection capabilities, addressing challenges like small object detection through innovations like FPP necks and CSPDarkNet53 backbones. Initially introduced for natural language processing (NLP), transformers [36] have broadened their application to include computer vision tasks. Their architecture, centered around the self-attention mechanism, facilitates a deeper contextual understanding of input sequences and has proven particularly effective in overcoming the limitations of sequential models, such as recurrent neural networks [37] and long short-term memory [38], offering enhanced data handling and processing capabilities. Adapting transformer technology to vision tasks, as demonstrated by detection transformer [39] and visual transformer (ViT) [40], signifies a paradigm shift in object detection, achieving unparalleled accuracy and context awareness, although with higher data and computational requirements. Due to these requirements, their applicability in underwater scenarios is limited by significant data and computational demands. These models typically require extensive data sets to achieve performance comparable to conventional CNN-based SOA, a challenging underwater detection precondition due to the scarcity of large-scale data sets. Nevertheless, recent efforts have explored integrating transformer layers with CNNs to leverage the transformers' superior feature extraction capabilities while mitigating their data requirements, leading to the development of models, such as YOLOv5-TR [42], where the transformer layer is located between the neck and the backbone layers, demonstrating improved performance with the complexities of underwater environments in SSS images. Aubard et al. [43] compared the YOLOv5 with its transformer version YOLOv5-TR and a novel anchor-free object detection of the YOLOX [44] with an SSS wall data set. They conclude that YOLOv5-TR improved the classic version, whereas YOLOX gives the best result. Recently, the authors proposed the YOLOX-ViT model [45] in SSS images, improving its previous version implemented by a ViT layer, and proposed a lightweight version of their model called KD-YOLOX-ViT using knowledge distillation. Two object detection sonar samples are represented in Fig. 2. The left sample represents the RBoxNet rotation bounding end-to-end detector [41] detecting a shipwreck using an FLS. In contrast, the right sample shows the YOLOv5-TR [42] detecting shipwrecks on SSS images. Although both detections detect shipwrecks, those samples show the detection representation of shipwrecks on SSS and FLS images.

B. Segmentation

The goal of segmentation in underwater imagery is to categorize each pixel of an image into meaningful classes, which can significantly assist in tasks, such as habitat mapping, species identification, and monitoring underwater infrastructure, or ecological changes. The journey of DL-based segmentation began

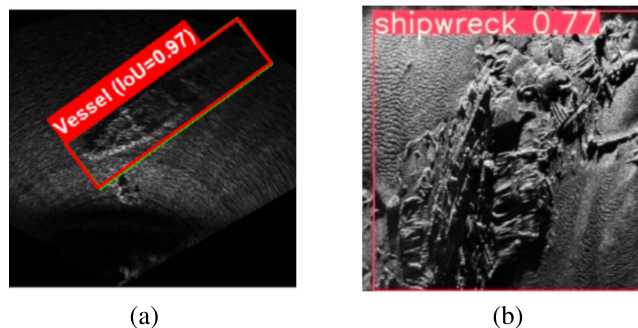


Fig. 2. Samples of object detection on sonar images. (a) Object detection model prediction for a vessel on a FLS image. (b) Object detection model prediction for a shipwreck on a SSS image.

with fully convolutional networks [46], which marked a departure from traditional patch-based classification methods by processing an entire image in a single forward pass and outputting a pixel-wise annotation map. U-Net's architecture [47], characterized by its symmetric expanding and contracting paths, was designed to capture context and localize features effectively. This model became a blueprint for many follow-up studies, including those focusing on underwater imagery, due to its efficiency in handling small data sets with high performance, a common scenario in underwater research. As the field progressed, models like DeepLab [48] and PSPNet [49] introduced approaches to capture broader context and achieve more precise segmentation boundaries. These models enhanced the segmentation of complex scenes, where the variability in scale and appearance of objects poses significant challenges. As for the object detection field, introducing attention mechanisms and transformers into segmentation models marked another improvement. Attention U-Net [52], for instance, adapted the U-Net architecture by incorporating attention gates, which help the model focus on relevant features while suppressing less important ones. This capability is particularly beneficial in underwater segmentation, where foreground-background contrast can be low, and objects of interest may be obscured. Models like SETR [53] and ViT-Seg [54], which leverage the transformer's ability to handle long-range dependencies, offer promising results for feature extraction of underwater segmentation into sonar images, such as MiTU-Net [55] a mix of a Transformer and U-Net used on FLS images. However, transformer-based models for feature extraction in sonars can be unstable due to unpredictable noise; He et al. [56] have proposed a hybrid CNN-Transformer-histogram of oriented gradient framework for FLS segmentation to address this challenge, outperforming the previous CNN and CNN-Transformer-based sonar segmentation. Most underwater segmentation studies predominantly utilize RGB images, reflecting a broader trend within computer vision [57], [58]. However, there is a notable gap in the availability of data sets derived from sonar imaging. This scarcity is compounded by the significant time and resources required for data set annotation, particularly for segmentation tasks. Segmentation, by its nature, demands detailed pixel-wise labeling, which is considerably more time-consuming and labor-intensive than the bounding box annotations used in object detection. Each image must be

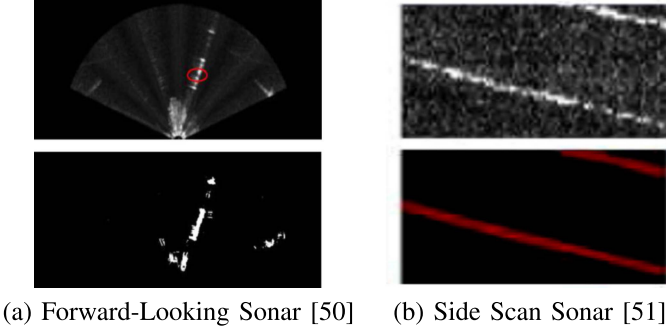


Fig. 3. Samples of segmentation on sonar images. (a) Segmentation model prediction on a FLS image. (b) Segmentation model prediction on a SSS image.

meticulously analyzed to ensure accurate classification of all pixels, which can be incredibly challenging in underwater environments where distinguishing between different elements can be complicated due to poor visibility, sound reflection/refraction, overlapping objects, and variable lighting conditions. The substantial annotation effort required for segmentation data sets sometimes aligns with the limited resources and time available to research teams. Fig. 3 represents two segmentation samples; the one on the left represents a saliency segmentation method for pipeline recognition on FLS [50], whereas the right sample is an SSS image where a combined residual and recurrent CNN called R^2 CNN is applied on a fishing net data set [51].

C. Simultaneous Localization and Mapping

The final subfield in computer vision addressed in this section is SLAM, commonly known as SLAM [59]. In contrast to detecting and localizing objects, SLAM focuses on mapping an unknown environment while simultaneously tracking the robot's position. SLAM algorithms are crucial in robotics and autonomous vehicle navigation, especially in scenarios where GPS is unreliable or unavailable, such as underwater. It usually employs probabilistic algorithms, such as Kalman filters (KF) [60] or Graph SLAM [61], to represent uncertainty in both the map and the robot's position estimate. These algorithms can handle several types of sensor data, including vision (optical and sonar), range, or odometry measurements, and fuse these measurements to create a consistent environment map while estimating the robot's pose. Deep SLAM [64] is an emerging research area that combines DL models with traditional SLAM. DL models enhance feature extraction, depth estimation, and image alignment in SLAM systems [65], [66], [67], [68]. However, implementing SLAM in underwater environments poses several challenges, such as limited visibility (due to turbidity or darkness), underwater currents (causing AUVs to drift, leading to errors in localization and mapping), sensor noise (which affects pose estimation accuracy), and computational complexity (due to limited power and processing capacity of AUVs). In addition, the lack of GPS underwater requires reliance on KF [69] to estimate vehicle position based on the last GPS fix and subsequent movements. Despite using KF or extended Kalman Filters (EKF), the ground truth is still an "estimated"

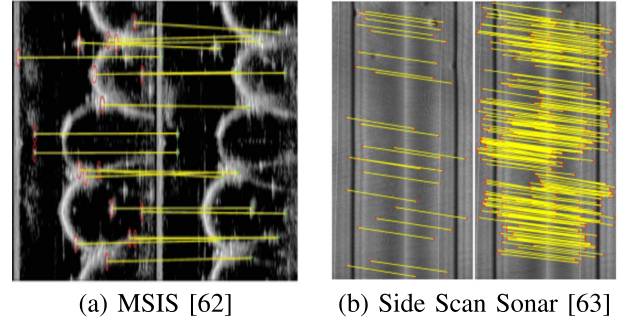


Fig. 4. Samples of SLAM on sonar images. (a) SLAM detected keypoint correspondence prediction on a MSIS image. (b) SLAM detected keypoint correspondence prediction on a SSS image. These points refer to identifying specific points of interest (key points) across multiple observations or images considered at the same physical location in the environment.

ground truth [70]. Recent methods combine MBES and optical data to leverage the strengths of both sensors while mitigating their limitations [71]. A detailed overview of the challenges and technologies for underwater SLAM using sonar and optical sensors can be found in [72]. However, sonar-specific DL SLAM remains underdeveloped, presenting opportunities for future research, such as using a CNN-based model for feature extraction by Yang et al. [73]. Fig. 4 showcases two sonar samples where non-DL SLAM algorithms are applied. On the left, an EKF SLAM algorithm using speeded-up robust features for feature extraction is applied to mechanical scanning imaging sonar (MSIS) images [62]. On the right, Graph SLAM uses the scale invariant feature transform for feature extraction in SSS images [63].

D. Challenges

The primary concern when applying DL for sonar vision is the availability and quality of data for training the DL models. Regrettably, many companies and laboratories opt to keep their data sets private due to the high costs associated with data collection. Furthermore, there is a lack of comparative studies on sonar data sets in the literature, which could help practitioners train their models more effectively. On occasions when open-source data sets are available, they often do not match the specific objects, environmental conditions, sensors, or noise characteristics, resulting in an unusable data set. Consequently, many underwater robotics researchers must collect and annotate their data, which is time-consuming and costly. The quality of the sonar data is paramount for practical training, such as interference from other sonar devices, marine life, or general ambient noise, which can distort sonar data [74]. Thus, through this article, we encourage researchers and practitioners to compare new models across data sets with varying characteristics when publishing to help readers understand the model's performance under different sonar conditions and environments, as demonstrated in [75]. To help researchers and practitioners Section IV-A aims to present in a single document the SOA data sets for sonar-based tasks, such as classification, object detection, and segmentation collected with SSS and FLS. Although

the lack of a data set is the main critical point for applying reliable sonar-based DL for underwater tasks, the difference between training and deployment data is another important one. Indeed, as previously explained, sonar sensors suffer from different potential noise sources, which can differ drastically from the images in training. This discrepancy can lead to DL model errors, such as misclassification or failure to detect objects. Much research focuses on denoising sonar images to mitigate this issue, aiming to preprocess data to remove noise [76], [77], [78]. While denoising is critical in sonar-based applications due to inherent noise issues, it faces three primary limitations. First, the unpredictability of underwater noise may lead to scenarios where the denoising process is ineffective, leaving residual noise that can still fool DL models. This limitation underscores the challenge of ensuring comprehensive noise removal without sacrificing crucial data features. Another limitation is the potential loss of critical information during the denoising process. While denoising aims to improve image clarity by removing noise, it can inadvertently eliminate important features for accurately detecting and classifying underwater objects. Finally, recent denoising approaches for sonar images rely on autoencoders [79] and deep autoencoder [80], which, to be effective in real-time applications, must operate swiftly to avoid delaying the vehicle's interaction with potential objects. This requirement for speed can compromise the denoising quality or the overall system's efficiency, posing a tradeoff between noise removal and operational effectiveness. Thus, even though denoising data increases the clarity of sonar images, more is needed to ensure the correctness of DL model prediction. Other alternative approaches, such as neural network verification, adversarial attack defenses, and OOD detection, need further study to enhance the robustness of DL models in sonar applications beyond denoising techniques. Thus, the four last sections of Section IV aim to provide the first robustness study under the scope of sonar neural network verification, adversarial attack, and OOD instead of denoising data.

IV. SONAR-BASED DL: TOWARD ROBUSTNESS

This section aims to comprehensively summarize available open-source sonar data sets mostly for FLS and SSS sensors. It also covers the synthesis of simulators for generating simulated sonar data and explores various methods for generating synthetic data. In addition, this section addresses the robustness of sonar-based DL models, detailing approaches like neural network verification, adversarial attacks, OOD detection, and uncertainty quantification. Conclusively, we propose a structured framework designed to enhance the robustness of sonar-based DL models, ensuring their reliability before deployment for real-world applications.

A. Data Set—State of the Art

In underwater exploration and research, sonar data sets are indispensable, offering a broad spectrum of applications from geological surveys to object detection. The previous section spotlighted the lack of open-source sonar data sets. Thus, this section is a detailed narrative synthesis of several notable data

sets, highlighting their characteristics, utilities, and the nuances of their collection methods, aiming to provide better knowledge for researchers of open-source sonar data sets: The UCI ML Repository—Connectionist Bench (Sonar, Mines versus Rocks) Data Set [81] specializes in sonar signal intensity data for classifying mines and rocks. This data set's strength lies in its focus on sonar signal characteristics, offering a resource for models designed to operate in environments where optical clarity is compromised. However, its specific focus on mines and rocks limit its direct application to broader image recognition or segmentation tasks. Sugiyama et al. [82] published SSS data from ice terraces in Glacier Grey, Patagonia [83] illuminating glacier formations. While offering ecological insights, this data set's utility could be more constrained by the need for specialized software to interpret the data. Ireland's Open Data Portal presents an eclectic collection of 14 sonar data sets [84], whose variety spans a broad spectrum of potential applications. Despite this diversity, the lack of detailed descriptions regarding the inclusion of SSS images or the nature of the data sets poses a challenge in identifying their applicability to specific research questions or projects. The U.S. Government's Open Data portal [85], with its range of data sets, including sonar data, promises a wealth of data for various applications. However, the expansive scope of the portal makes locating specific types of sonar data, such as SSS images, a daunting endeavor that demands considerable time and effort. The Marine_PULSE data set [86], introduced by Du et al. [87], features SSS images focusing on underwater objects, such as pipelines, cables, and engineering platforms. While the data set's grayscale, object-focused images provide specificity, they are limited by low resolution and a lack of broader contextual features, which could impede comprehensive model learning.

The Northern Adriatic Reefs data set [88] offers georeferenced SSS mosaics of biogenic reefs off Venice. Although these low-resolution, whole-mosaic images capture broader geological formations, their format may need to be more conducive to training models to detect specific objects or features within a diverse underwater environment. The Seafloor Sediments data set [91] boasts over 434 164 images derived from SSS waterfalls, showcasing a variety of seafloor types along the coast of Catalunya. The data set's large scale and seafloor-type diversity are invaluable. However, the lack of detailed information on the specific conditions of image collection (e.g., sonar range, altitude) might influence the generalizability of models trained on this data. The UXO data set [90] focuses on unexploded ordnance detection in underwater environments, containing 74 437 frames collected using the ARIS Explorer 3000 sonar in a controlled pool environment, with 48 462 corresponding Go-Pro optical frames. Annotations, including bounding boxes and object types, are available for camera frames, supporting object detection tasks. Similarly, SSS Imaging for Mine Detection [92], with 1170 annotated images for distinguishing mine-like and nonmine-like objects, directly facilitates training in underwater mine detection. Despite its utility, the data set would benefit from additional details on the original image sizes, sonar range, and AUV altitude during data collection, which is essential for understanding the detected object scale and appearance. Despite

its data augmentation potential, the focus on FLS images and the collection in a pool environment may only partially capture the complexities and variations of oceanic environments. The underwater acoustic target detection (UATD) data set [75] is an FLS data set for object detection, which provides 14 639 annotations among 9200 images. It provides a complete dual-frequencies data set (720 kHz, 1200 kHz) with 10 classes: cubes, balls, cylinders, human bodies, planes, circle cages, square cages, metal buckets, tires, and bluerovs. The data collection setup is described in detail in [75], benefiting potential users using the data set under the same conditions. Dual-frequency identification sonar (DIDSON) data set [93] provides a fish data set collected in the Rhode River, MD, USA, Indian River Lagoon, FL, USA, San Francisco Bay, CA, USA, and Bocas del Toro, Panama, resulting in 100 h of data where 1000 frames were extracted with eight fishes species labeled for segmentation model showcased by Perivoliot et al. [94]. The data set requires the DIDSON-V5 software provided by Sound Metrics to convert the DDF format files. The Marine Debris Turntable (MDT) data set [95] presented by Singh et al. [96] contains 2471 FLS images with 12 classes of objects, including bottles, pipes, platforms, and propellers annotated for the segmentation task. The Synthetic Aperture Sonar Seabed Environment Data Set (SASSED) [97] provides 129 complex-valued, high-frequency sonar snippets depicting various seafloor textures, such as hardpack sand, mud, sea grass, rock, and sand ripple. Each snippet includes a hand-segmented mask image, which groups similar textures without necessarily representing ground truth labels, providing a valuable resource for training and testing sonar-based ML models.

The AI4Shipwreck data set [89] contributes to underwater archaeology and research. This data set comprises 286 high-resolution SSS images collected from 24 distinct shipwreck sites within the Thunder Bay National Marine Sanctuary, utilizing the EdgeTech 2205 SSS technology. A notable aspect of this collection is its focus on shipwrecks, offering a lens through which the underwater past can be explored and studied. Each image within the data set has been annotated for segmentation tasks. The NKSID data set [98], introduced by Jiao et al. [99], provides a total of 2617 FLS images split into eight classes among big propellers, cylinders, fishing nets, floats, iron pipelines, small propellers, soft pipelines, tires, which makes it the most significant data set for classification task among our data set comparison. The data collection occurred in Bohai Bay with an ROV set-up with dual frequency (550 kHz, 1.2 MHz) Oculus M750 d as FLS. The SubPipe data set [57] encompasses a comprehensive underwater collection, uniquely combining grayscale and RGB camera imagery with SSS images to offer a holistic view of underwater pipeline environments. Alongside vision (optical and sonar) data, this data set enriches its offering with conductivity, temperature, and depth readings and navigational information, providing a multifaceted underwater exploration and analysis approach. Utilizing the Klein3000 sonar system at dual frequencies of 455 and 900 kHz, the data set captures high-quality imagery conducive to detailed study and model training. Annotations and benchmarks within the SubPipe data set are tailored to evaluate SOA models across various applications, including segmentation, SLAM utilizing RGB data, and object

detection using SSS images. The SSS data set section of the Subpipe data set counts 10 030 images, making it the biggest open-source SSS object detection data set in our comparison. Sonar wall detection data set (SWDD) [5] comprises 864 SSS images of walls, meticulously annotated following the COCO format, providing a resource for training and testing object detection models. The data set includes an SSS waterfall video spanning 6 min and 57 s, from which 6243 images have been extracted and annotated to support benchmark validation efforts further. Including YOLOX [44] and YOLOX-ViT [45] models in the data set's benchmarking process highlights the exploration of advanced object detection techniques in the context of SSS data. Utilizing the Klein3000 sonar at frequencies of 455 and 900 kHz, the data set offers high-resolution imagery conducive to detailed object detection tasks. It serves as a testing ground for innovative model architectures like YOLOX-ViT. In addition, an extended version of the data set has been recently introduced through [100], where three data sets were collected under different weather conditions to improve object detection model comparison. These additional data sets include a total of 797 SSS images, which complement the original data set. The Aurora data set [101] provides a multisensor collection for underwater exploration, integrating sonar, camera, and inertial sensors to facilitate the development of SLAM algorithms across different underwater environments. This comprehensive multisensor approach enriches SLAM research by enabling fusion between different types of sensory data, helping address the complexity of underwater localization and mapping. Krasnosky et al. [102] presented a data set that offers bathymetric surveys using MBES, complemented by high-precision GPS-based ground truth, which is particularly valuable for enhancing SLAM and bathymetric mapping methods. Similarly, Mallios et al. [103] capture sonar imaging of underwater caves, a unique and highly challenging environment for SLAM testing. Those SLAM data sets highlight the importance of comprehensive sensory data and ground-truth validation to advance SLAM capabilities in diverse and demanding underwater environments. Fig. 5 represents some samples of those data sets. Each data set's detailed account of data collection methods, including sonar range and environmental conditions, enriches underwater sonar research. Their diverse focuses—from geological formations to object detection—underscore the critical role of detailed documentation and the necessity for a broad range of data sets to address the multifaceted challenges of underwater exploration and monitoring.

With the advent of DL on sonar data, a significant demand exists for data sets. Unlike uncrewed aerial vehicles (UAVs) or autonomous ground vehicles, AUVs face unique data collection challenges due to the complexities of the underwater environment. This data set comparison aims to highlight current sonar-based data sets. Traditionally, open-source sonar data sets for DL applications were rare, hindering the ability to compare different models effectively due to the utilization of disparate data sets. However, recently, a promising trend towards the publication and availability of such data sets, from Aubard et al. [5], Tuñón et al. [57], Sethuraman et al. [89], Jiao et al. [98], Dahn et al. [90], and Santos and Moura [92] emerging in early 2024,

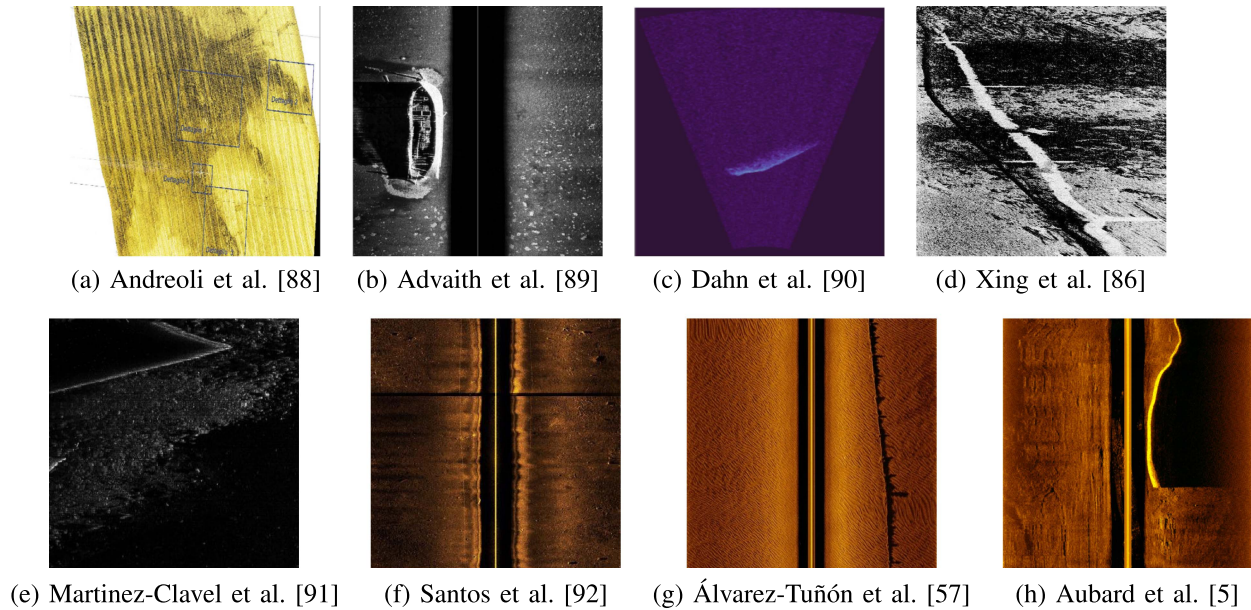


Fig. 5. Samples of underwater sonar data sets. (a)–(h) Illustrate various SOA sonar data sets, including GeoTiff [88], SSS [89], FLS [90] images with different objects, such as pipelines [57], walls [5], and shipwrecks [89].

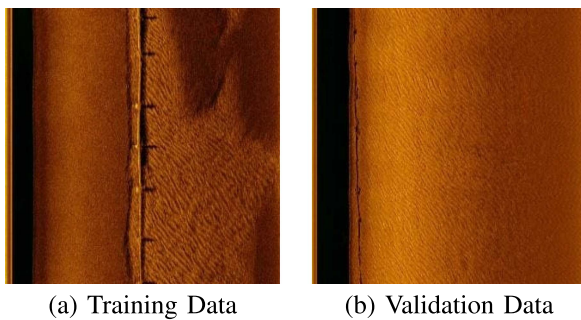


Fig. 6. Comparison between training and validation data sets [57]. (a) Training SSS image and (b) shows a validation SSS image from the same location, collected on different dates and at different altitudes. This comparison highlights the impact of variations, such as AUV altitude on SSS images, which can reduce the accuracy of DL models. Thus, several parameters, such as AUV's altitude, must be carefully considered when creating a training data set.

alongside Martinez-Clavel et al. [91], Ogburn et al. [93], and Xing et al. [86] from 2023. This growing repository, although still modest in volume, signals a shifting paradigm towards enhancing the reproducibility of experiments and incentivizing collaborative data collection efforts within the research community. Despite these advancements, sonar technology's inherent characteristics still require a delicate approach from training to inference data. Tuñón et al. [57] illustrated how model performance may vary between training and operating data sets. They compare data sets collected at different dates and vehicle altitudes, resulting in 98% in the average precision AP_{50-90} for training, whereas only 15% of AP_{50-90} for inference data. Fig. 6 represents two samples from the training and validation data set, where the pipeline size and shadow are different on both images despite the same environment. This variability underscores the

importance of considering environmental and operational factors for training and operating. Table II compares the SOA to the sonar data sets described. To the best of authors' knowledge, this is the most complete sonar data set comparison. It presents a comprehensive overview of sonar data sets, including 10 SSS and 6 FLS data sets, encompassing 4 for classification, 5 for detection, 5 for segmentation, and 3 for SLAM, alongside two data sets lacking annotations. It reveals that the NKSID [98] provides the most images in classification with 2617 images, while SubPipe [57] and UXO [90] in detection respectively with 10 030 and 74 437 images. In contrast, UATD [75] offers a broader scope with 10 distinct classes across 9200 images. The most complete data set for segmentation is Seafloor Sediments [91], boasting 434 164 images.

Moreover, unlike terrestrial and aerial optical images, which can be benchmarked with established data sets like COCO [105], or ImageNet [106] (for object detection), sonar data sets have yet to establish a universally recognized baseline data set. This absence complicates direct model comparisons, highlighting a current limitation for future development within the community. Establishing such a benchmark would greatly facilitate advancements in sonar-based DL, improving the field toward greater standardization and comparability. Thus, due to this limitation, we have created a dedicated GitHub repository centralizing current open-source underwater sonar data sets, providing a foundation for future benchmarking efforts. In addition, researchers and practitioners are encouraged to contribute by adding new data sets, ensuring the resource remains comprehensive and up to date. The repository is available online.¹

¹[Online]. Available: <https://github.com/remaro-network/OpenSonarDatasets>

TABLE II
COMPARISON OF THE OPEN SOURCE STATE-OF-THE-ART SONAR UNDERWATER DATA SETS

Dataset	Sonar	Data	No Data	Object labels	Annotation	Set-up	Year
Northern Adriatic Reefs [88]	SSS	GeoTIFF	7	Reefs	✗	✗	2010
Lago Grey [82]	SSS	Raw	✗	Glacier, Walls	✗	✓	2019
UCI ML [81]	✗	Raw	211	Mines, Rocks	Classification	✗	✗
SeabedObjects-KLSG [104]	SSS	Images	1190	Wrecks, Humans, Mines	Classification	✗	2020
Marine_PULSE [86]	SSS	Images	627	Pipes, Mounds, Platforms	Classification	✗	2023
NKSID [98]	FLS	Images	2617	Infrastructures, Propellers, Tires	Classification	✓	2024
UATD [75]	FLS	Images	9200	Tires, Mannequins, Boxes	Object Detection	✓	2022
SSS for Mine Detection [92]	SSS	Images	1170	Mines	Object Detection	✗	2024
SWDD [5]	SSS	Images	7904	Walls	Object Detection	✓	2024
SubPipe [57]	SSS *	Images	10 030	Pipelines	Object Detection	✓	2024
UXO [90]	FLS	Images/Raw	74 437	Unexploded Ordnances	Object Detection	✓	2024
MDT [95]	FLS	Images	2471	Infrastructures, Debris	Segmentation	✓	2021
SASSED [97]	SAS	Images	129	Muds, Sea Grass, Rocks, Sands	Segmentation	✗	2022
Seafloor Sediments [91]	SSS	Images	434 164	Rocks, Marine life	Segmentation	✓	2023
DIDSON [93]	FLS	Images	1000	Fishes Species	Segmentation	✓	2023
AI4Shipwreck [89]	SSS	Images	286	Shipwrecks	Segmentation	✓	2024
Cave Sonar [103]	MSIS *	Rosbag	500 meters	Cave Seabed	SLAM	✓	2017
Aurora [101]	MBES, SSS *	Raw	35h	Seabed, Marine habitats	SLAM	✓	2020
MBES-Slam [102]	MBES	Rosbag	4 missions	Seabed	SLAM	✓	2022

This table compares the state-of-the-art Sonar underwater dataset by analyzing the type of sonar (Sonar), type of data (Data), number of data (No data), objects labeled in the data (Object labels), if the data is annotated, for which DL tasks (annotation), if the data collection set up such as sonar frequency, altitude, etc. are described in the dataset or not (Set-up) and finally the year of the dataset publication (Year). The * is to mention that the dataset is not only limited to Sonar but extended to other sensors such as optical cameras.

B. Synthetic Data

Given the limited open-source sonar data set, a novel research trend relies on simulated data for real-world implementations, known as Simulation-to-Real [117]. The goal is to reduce the costly and time-consuming process of underwater data collection by relying on underwater simulators equipped with sensor payloads to generate simulated data, resulting in the development of multiple open-source simulators. UWSim [107] is the first underwater simulator designed for marine robotics applications, built on the OpenSceneGraph and osgOcean libraries to provide realistic underwater vision and physics that support various sensors and vehicles. Gwon et al. [108] improved the UWSim simulator by proposing an SSS plugin. DAVE [111] and UUV Simulator [109], anchored in the ROS [118] and Gazebo framework, primarily support conventional cameras and FLS to enhance acoustic visualization, making them ideal for users engaged with the ROS ecosystem. Furthermore, DAVE provides an underwater point cloud LIDAR sensor. Stonefish [112] uses ROS for publishing virtual sensor measurements and includes cameras, FLS, and SSS images, underscoring its proficiency in simulating custom acoustic data. However, the quality of simulated SSS images still needs to match actual SSS data. UNavSim [113], compatible with ROS and leverages high-detail rendering Unreal Engine 5 and AirSim [119], brings camera and underwater LIDAR technologies. MARUS [114], also compatible with ROS, offers a comprehensive sensor suite including underwater LIDAR, cameras, and FLS, highlighting

the simulator's commitment to providing diverse and realistic sensor data for underwater research. HoloOcean [115] also generates camera and FLS imagery to craft realistic underwater scenarios. Furthermore, Potokar et al. [116] recently improved their HoloOcean simulator by implementing multibeam imaging, multibeam profiling, SSS, and echo-sonar, which makes it one of the most complete simulators for simulated sonar images. Cerqueira et al. [110] presented ImagingSonarSimulator, a sonar simulator for FLS and MSIS. Using the Rock-Gazebo framework [120], it models physical forces in the underwater environment, providing real-time simulation for a virtual AUV. The simulator uses the OpenGL shading language [121] on a GPU to emulate sonar devices based on parameters, such as pulse distance, echo intensity, and field-of-view. Furthermore, the simulator is compatible with ROS 1. Ciuccoliet al. [122] provided a deeper review of sonar and nonsonar simulator analysis. This comparison underscores that while underwater simulators are progressively advancing as a research domain, offering increasingly lifelike representations of underwater environments through cameras and FLS, they still confront challenges in providing and accurately replicating the nuances of real SSS imagery. Table III systematizes the underwater simulators by highlighting their vision-based sensors, such as camera and underwater LIDAR, providing 3-D point cloud, FLS, SSS, and MSIS. This comparison shows that most simulators focus primarily on rendering camera images, while HoloOcean [116] and Stonefish [112] provide an extensive range of realistic sonar capabilities. The GitHub repository MASTODON [123] provides

TABLE III
COMPARATIVE ANALYSIS OF THE OPEN-SOURCE UNDERWATER SIMULATORS

Simulator	Year	ROS Support	Vision Sensors				
			Camera	Lidar	FLS	SSS	MSIS
UWSim [107]	2012	ROS 1	✓	✓	✗	✓ [108]	✗
UUV [109]	2016	ROS 1	✓	✓	✗	✗	✗
ImagingSonarSimulator [110]	2017	ROS 1	✗	✗	✓	✗	✓
DAVE [111]	2022	ROS 1	✓	✓	✓	✗	✗
Stonefish [112]	2019	ROS 1 & 2	✓	✗	✓	✓	✗
UNavSim [113]	2023	ROS 1 & 2	✓	✓	✗	✗	✗
MARUS [114]	2022	ROS 1 & 2	✓	✓	✓	✗	✗
HoloOcean [115]	2022	ROS 2	✓	✗	✓	✓ [116]	✗

This table compares the underwater simulator and provides insight into the vision sensors those simulators provide. The table analyzes the date of publication (year), the ROS support (ROS support), then if they provide vision sensors among camera (Camera), underwater LIDAR (LIDAR), forward looking sonar (FLS), side scan Sonar (SSS), and mechanical scanning imaging sonar (MSIS).

a novel sonar image simulator, as described by Woods [124]. However, this article is not open-access, limiting detailed information availability. This section analysis focuses exclusively on open-source sonar simulators to support reproducible research and open-access studies. Nevertheless, nonopen-source sonar simulators, such as Coiras et al. [125], are also available in the literature.

However, despite reducing data collection time, simulated data sets usually represent ideal conditions, while the real world contains many uncertainties. For instance, sonar images may require adding noises to align them with real-world sonar data. This difference between simulated and real data is known as the sim-to-real gap, an ongoing research topic providing promising results for object detection [126] and segmentation [127]. GANs [128], traditionally used in tasks like NLP [129] and image generation, have emerged as a promising avenue for underwater image generation. Current research explores the potential of GANs to merge simulated and real data, thereby producing expansive data sets [130]. Jaber et al. [131] used conditional GAN (cGANs) for increasing their FLS data set collecting with the Stonefish simulator and validating it in a pool environment. Similarly, Lee et al. [6] addressed the scarcity of sonar imagery by employing a Pix2Pix-based cGAN to generate synthetic sonar images. This method enhances segmentation model training by simulating diverse underwater conditions, leading to improved performance when real data is limited, highlighting that GANs could offer a viable solution to the prevailing challenge of insufficient underwater data. Diffusion models [132], another emerging class of generative models, employ a sophisticated process of progressively adding then removing noise from images, offering stable training and diverse outputs, and showing improvement in SOA object detection on sonar images [133]. In contrast, GANs create data through a competitive process between a generator and a discriminator, known for producing highly realistic images but with potential training instability. G. Huo et al. [104] proposed a semisynthetic data generation method to generate data from optical to sonar data of airplanes and drowning victims using image segmentation with intensity distribution. This semisynthetic method aims to crop the object on optical images and add specific shadows, sonar backgrounds, and sonar distributions, resulting in an image that looks like a sonar image. Following the same principle, Bai

et al. [134] proposed a global context external-attention network (GCEANet), which, from optical images, produces pseudo-SSS images corresponding to the absent categories for zero-shot SSS image classification. Furthermore, Data augmentation and transfer learning, two well-known methods for improving DL outputs, can also be implemented during the DL training process to improve the accuracy and robustness of models. Data augmentation [135] increases the size of the data set by filtering, rotating, and adding random noises in the original data set, improving the accuracy and model performance against noise. Transfer Learning enhances the model's accuracy [104]. Instead of starting the DL model training from scratch, pretrained DL weights with bigger data sets are transferred into the DL model before training with the underwater data set.

C. Neural Network Verification

Neural network verification [136] in the context of sonar aims to ensure that DL models can accurately interpret data under various noise conditions within defined boundaries. It validates the model's predictions against expected outcomes across all potential sonar inputs, ensuring reliability in diverse underwater environments. Applying DL to real-world scenarios, especially in challenging underwater settings, requires robust verification to ascertain reliability and robustness [137]. In image processing, DL models typically incorporate various layers—such as convolutional, pooling, and fully connected—and nonlinear activation functions, such as ReLU, softmax, and sigmoid. The challenge of verification is amplified by the high-dimensional input space and the complex nonlinearities introduced by activation functions. Several tools have been developed to address this challenge. Alpha-Beta-CROWN [138] uses optimized bounding methods to determine neuron activation bounds, efficiently reducing computational demands. ERAN [139] and DeepPoly [139] utilize abstract interpretation techniques to balance verification precision and scalability, while Reluplex [140] and Marabou [141] focus on verifying ReLU-based networks, with Marabou extending support to more architectures and activation functions. These tools are valuable for ensuring robustness in DL models used for image classification and object detection, where reliability is critical. However, despite the advancements in neural network verification, current tools are primarily designed for classification tasks and face limitations when applied to other areas, such as object detection, segmentation, and SLAM. In addition, adapting these methods to sonar data introduces new challenges due to the unique uncertainties of underwater environments, which require greater focus in future research.

D. Adversarial Attack

Neural network Verification methods often fail to keep pace with the growing complexity of SOA DL models, leading to significant demands on computational resources due to the nonlinearities and high dimensionality of DL models, making scalability a critical concern for researchers and practitioners [142], [143]. The fast gradient sign method (FGSM) [143] is the first white-box adversarial attack methodology that leverages the gradients of a neural network to craft adversarial examples

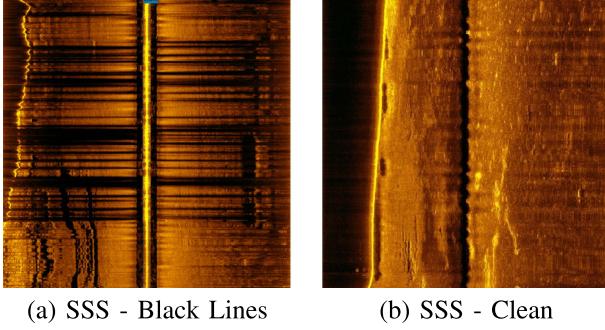


Fig. 7. Comparison between SSS images with and without loss of information from [100]. (a) SSS image with black lines caused by loss of information during a mission where the vehicle surfaced in windy weather, resulting in the transducer going out of the water. This noise caused wall nondetection, highlighting the lack of model robustness. (b) Clean SSS image with a nadir gap filtering applied.

efficiently and quickly. Building upon FGSM's foundation, Projected Gradient Descent (PGD) [144] introduces an iterative approach that enhances the basic concept. PGD refines FGSM's technique by applying multiple small-step adjustments to the input, ensuring the perturbation remains within a defined epsilon neighborhood of the original input, enhancing the adversarial example's effectiveness while maintaining its subtlety. DeepFool [145] diverges from PGD and instead focuses on the precision of perturbations; it seeks the minimal change required to alter a model's classification, providing a more nuanced estimation of model robustness. Carlini and Wagner's (C&W) Attack [142] identifies the minor possible perturbation that can still mislead a model, comparable to DeepFool's principle of minimal disruption. However, C&W designs their attack to be effective even against models fortified with defensive measures. In a targeted vector, the Jacobian-based saliency map attack (JSMA) [146] focuses on strategically modifying input features. By exploiting the output gradient concerning the input, JSMA identifies critical pixels whose alteration would specifically misguide the model into a chosen misclassification. As highlighted by Papernot et al. [147], transfer attacks exploit the phenomenon of perturbation transferability, highlighting the ability of adversarial examples to mislead different models. Query-based strategies, explored by Chen et al. [148] involved iterative input adjustments based on model feedback, aiming to find adequate adversarial inputs without accessing the model's gradients. Decision-based attacks, such as those by Brendel et al. [149], refine adversarial inputs using model outputs, score-based attacks, and Ilyas et al. [150] using confidence scores.

Most perception studies focus on grounded or aerial data sets for UAVs [151], predominantly working with RGB images [152]. However, underwater robots that rely mostly on sonar vision require a study of their specific uncertainties, which differ from RGB cameras. Fig. 7 shows an SSS sample with and without signal loss from [100], characterized by black lines in the image. These samples were collected at the surface during a storm, causing the sonar transducers to emerge from the water, resulting in missing data occasionally. Although this information loss is not adversarial noise in the traditional sense, it

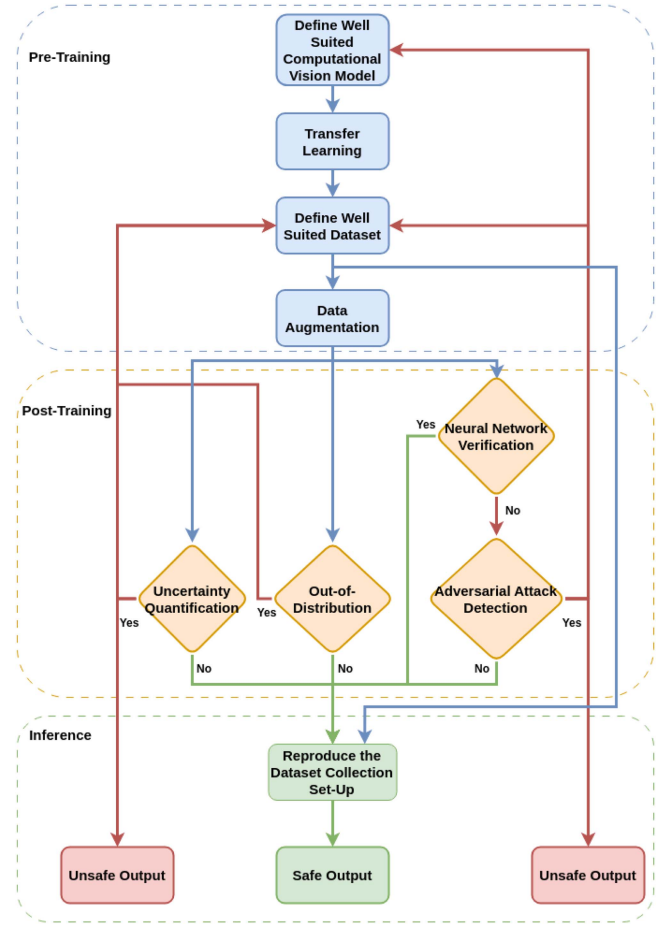


Fig. 8. Sonar-based DL—robustness workflow. This proposed workflow introduces two main steps: the pretraining step, represented by blue boxes (define well-suited computer vision model, transfer learning, define well-suited data set and data augmentation), and the posttraining steps before deploying a sonar-based DL model, represented by orange diamond, which should result in better model prediction under unexpected noises (neural network verification and adversarial attack detection) and underwater characteristics (OOD, epistemic uncertainty). Finally, suppose the output of the OOD, uncertainty quantification (epistemic and aleatory), and neural network verification (or adversarial attack detection) do not return any error; the sonar set-up from the data set collection should be reproduced (frequency, vehicle altitude) to ensure the correct model behavior during inference.

is a natural adversarial scenario, highlighting the unpredictable environmental conditions that can affect model performance. Such scenarios can be leveraged in adversarial retraining to enhance model robustness. Several laboratories are exploring adversarial methods for sonar images due to the growing interest in underwater tasks. ROSAR framework, proposed by Aubard et al. [100], utilizes PGD attacks to target specific safety properties of SSS, similar to the type of signal loss shown in Fig. 7. When these properties are compromised, the method generates counterexamples for adversarial retraining, resulting in up to 1.85% in detection model robustness. Ma et al. [153] proposed the noise adversarial network (NAN), which introduces noise into the data set and applies it to the Faster R-CNN object detection model, enhancing robustness by 8.9% mAP on a sonar data set. Furthermore, Ma et al. [154] presented the Lambertian Adversarial Sonar Attack (LASA), an adversarial attack for

SSS images based on the Lambertian reflection model. They compare LASA with FGSM, PGD, and Deepfool and conclude that LASA significantly improves the robustness of sonar-based classifiers. Feng et al. [155] examined the impact of adversarial noise on both CNNs and transformers when applied to sonar spectrograms.

E. Out-of-Distribution

Sonar images suffer from the disparity between training data and real-world sonar inputs, variations in sonar brands, operating frequencies, and environmental conditions, which presents a substantial challenge known as OOD [156]. Hendricks and Gimpel [157] introduced softmax probabilities to distinguish between correctly classified, misclassified, and OOD examples within neural networks. Building on this, Liang et al. [158] enhanced the model's ability to differentiate in-distribution (ID) from OOD data through temperature scaling and input preprocessing, introduced by their ODIN technique. Lee et al. [159] proposed the Mahalanobis distance for a sophisticated similarity measure between input features and class-conditional distributions, improving the efficacy of OOD detection. Generative models, such as GANs and variational autoencoders (VAEs) have enabled the identification of OOD inputs by learning latent representations of ID data. Kingma and Welling's [160] work, alongside Higgins et al.'s [161] introduction of β -VAEs, further enhance this by controlling differentiation in latent spaces, thereby improving OOD detection.

In applying these advancements to sonar imaging, Gerg and Cotner [162] directly addresses OOD detection by incorporating a perceptual metric prior (PMP) within the training classification loss. This approach notably improves the robustness of models, especially for sonar image classification tasks characterized by limited data and subtle distribution differences. It presents an innovative alternative to conventional methods like data augmentation and hyperparameter tuning, specifically tailored to overcome the unique challenges posed by sonar data analysis. Furthermore, Jiao and Zhang [163] introduces the balanced ensemble transfer learning (BETL) framework to address the compounded challenges of long-tail and few-shot classification in sonar images. This framework enhances classification accuracy while optimizing memory and inference time, indirectly aiding OOD detection by improving model performance on sparsely represented classes that often resemble OOD samples. As an extension of [163], Jiao et al. [99] presented the first comprehensive examination of open-set long-tail recognition specifically for sonar images, marking an improvement in sonar-specific OOD detection. This work navigates into the difficulties of classifying sonar data and setting new benchmarks by evaluating SOA algorithms and proposing the novel Push the Right Logit Up and the wrong Logit Down (PLUD) loss function. Cook et al. [164] used their null space analysis (NuSA) approach, which aims to detect outliers while testing for classification tasks, to detect unknown objects during ATR tasks in sonar data [165], and conclude that NuSA applies to sonar images outperform the OOD methods, such as unsupervised self-supervised outlier detection (SSD) [166] and rectified activation [167].

F. Uncertainty Quantification

Uncertainty quantification in DL models serves as a critical framework for assessing the confidence and reliability of predictions. It is categorized into two main types: aleatory uncertainty, coming from the inherent noise and variability in the data [168], and epistemic uncertainty, which stems from the model's lack of knowledge or uncertainty about the model itself [169]. Techniques for addressing these uncertainties have been extensively reviewed, highlighting the dual nature of predictive uncertainty within supervised learning frameworks [170], [171]. The methodologies range from Bayesian inference, which offers a deep-rooted framework for epistemic uncertainty, to heteroscedastic neural networks that effectively model aleatory uncertainty by allowing variance in predictions based on data noise [172]. Monte Carlo Dropout, introduced by Gal and Ghahramani [173], presents a more computationally feasible approximation of Bayesian inference, balancing practicality and theoretical rigor. Addressing aleatory uncertainty, heteroscedastic neural networks, as explored by Kendall and Gal [172], propose a model adjusting its confidence levels based on the inherent noise present in input data. This approach models uncertainty as a function of the data, allowing predictions to reflect the variability in the underlying data distribution. Ensemble methods, introduced by Lakshminarayanan et al. [174], emerge as a robust approach to encapsulate both types of uncertainty by aggregating predictions from a collection of models, which can capture model variance and reflect data variability. Despite their effectiveness, ensembles require multiple models to be trained and maintained, which may not be feasible in resource-constrained environments. When applied to sonar-based data, uncertainty quantification in DL models becomes crucial for enhancing the robustness and reliability of underwater object detection and classification tasks. Fuchs et al. [175] proposed a pipeline for generating simulated FLS data using cycleGAN and ensuring the quality of the simulated images for real deployment by analyzing the data uncertainty for detection and classification tasks. Tarling et al. [176] combined self-supervised learning with uncertainty quantification to improve training and measure prediction uncertainty for fish detection on FLS images. By estimating the noise variance of the data set images, they adjust the loss function to regulate the aleatoric uncertainty.

G. Limitations and Workflow

Section IV aims to help researchers and practitioners interested in leveraging sonar-based DL perception tasks by comparing and reviewing state-of-the-art sonar data sets and simulators into a consolidated document, facilitating access to simulated and field-collected data set. Surprisingly, our research revealed a lack of comprehensive surveys of these data sets, mainly because researchers were not used to revealing their data sets. However, this scenario is gradually changing with a novel trend toward data sharing, as evidenced by the recent availability of numerous open-source data sets from 2023 and 2024. This shift underscores a growing commitment within the research community towards openness and collaboration, significantly benefiting the field. The section also highlights the unique sonar uncertainties

TABLE IV
COMPARATIVE ANALYSIS OF THE PREVIOUS WORKS FOCUSED ON SONAR-BASED ROBUSTNESS SEPARATED INTO THREE MAIN SECTIONS: OOD, ADVERSARIAL ATTACK, AND UNCERTAINTY QUANTIFICATION

Paper	Year	Robustness	Method	Sonar	Task
Ma et al. [153]	2020	AA	NAN	SSS and SAS	D
Feng et al. [155]	2023	AA	FGSM [143] and PGD [144] on Spectrogram	Spectrogram	C
Shell et al. [154]	2021	AA	LASA	SSS	C
Aubard et al. [100]	2024	AA	ROSAR	SSS	D
Gerg et al. [162]	2023	OOD	PMP	SAS	C
Jiao et al. [163]	2022	OOD	BETL	FLS	C
Jiao et al. [99]	2024	OOD	PLUD	FLS	C
Cook et al. [165]	2020	OOD	NuSA [164] on Sonar	SAS	D
Fuchs et al. [175]	2022	UQ	cycleGAN	FLS	D, C
Tarling et al. [176]	2022	UQ	Prediction Uncertainty	FLS	D

This table analyses the date of publication (*Year*), the robustness method (*Robustness*) where *AA* stands for adversarial attack, *OOD* for OOD and *UQ* for uncertainty quantification, the method used in their papers (*Method*), the type of sonar use (*Sonar*), and the DL task realized (*Task*) where *C* stands for classification and *D* for detection.

encountered underwater, affecting DL models' performance. Discrepancies between training and operational data, including variations in environmental conditions, sonar setup, and seabed characteristics, can potentially mislead models. Presently, the focus within sonar methodologies leans heavily toward image denoising before model inference rather than supporting the intrinsic robustness of the models themselves. Consequently, this section presents strategies for addressing uncertainties inherent in sonar imaging and the first survey focusing on robustness for sonar-based DL models. However, this preliminary survey highlights a noticeable scarcity of research explicitly addressing neural network verification (0 papers), adversarial attacks (4 papers: [100], [153], [154], [155]), OOD detection (4 papers: [99], [162], [163], [165]), and uncertainty quantification (2 papers: [175], [176]), as detailed in Table IV. Through this article, we encourage further research into the robustness of sonar-based DL models, which would enhance the safety and autonomy of underwater robotic systems.

To ensure the robustness of sonar-based DL model implementation into AUVs, we support our document by proposing a framework describing the specific steps improving the models' robustness. Fig. 8 describes a comprehensive workflow by selecting a well-suited computer vision model and data set, applying transfer learning and data augmentation. If the unsafe output corresponds to OOD data or uncertainty quantification, the data set requires refinement to enhance detection reliability. Furthermore, the model must undergo either a neural network validation or adversarial attacks under specific underwater noises. If those two steps are invalid, the model is considered "unsafe," requiring a revision of the model and training data set. "Safe output" signifies the model's ability to perceive its environment, resulting in potential correct vehicle behavior. However, the training data set setup should match the inference setup, including sonar frequency, color map, and vehicle altitude. In the context of "unsafe," decisions by the autonomous vehicle driven by the DL model could engender unwanted outcomes, resulting in hazardous scenarios. For such AUVs relying on DL perception for navigation and decision-making, the goal remains

to reduce DL model output uncertainties, ensuring the vehicle's and its surroundings' safety.

V. CONCLUSION AND FUTURE RESEARCH TRENDS

The growing interest in underwater exploration, inspection, and monitoring has led to a specific need for underwater data collection and interaction, resulting in underwater robots like AUVs. While AUVs, mean autonomy under real-time human operators' supervision, they still seek fully autonomous actions and interactions in the deep sea. In contrast with other autonomous vehicles, such as autonomous cars and UAVs, AUVs suffer from a lack of communication, visibility, and available data, which results in an uncertain and dangerous environment for vehicles to perform tasks autonomously. Furthermore, autonomous behaviors rely on sensors that capture and understand the vehicle's surroundings to adapt to the vehicle's trajectory in real-time. Because of the often bad quality of underwater camera images, underwater vision mainly relies on sonars (e.g., SSS, FLS, MBES) to map and collect data from the underwater environment. Thus, improving AUVs' autonomy requires understanding the sonar data while surveying, which results in implementing computer vision DL models, such as classification, object detection, segmentation, and SLAM. However, implementing such a model onboard requires a safe DL model without real-time supervision and communication. Thus, the challenge is finding the tradeoff between autonomy and safety to improve the vehicle's autonomy without compromising its safety and the safety of its surroundings. In this article, we tackled the robustness of the sonar-based DL model with the following question: "How can we rely on a real-time sonar-based DL model?" aiming to spotlight the current research topics and method that can be applied to reduce the sonar-based DL models uncertainties. We compared previous surveys on sonar-based DL models, highlighting the need for robustness focus. Indeed, current surveys mostly compare sonar-based DL models and highlight the lack of open-source data sets but without referring to the robustness of the model itself, which is primordial to

ensure its good behavior. By presenting and critically analyzing 19 sonar data sets and comparing various underwater simulators, we have offered a comprehensive resource for accessing and generating open-source data crucial for advancing DL underwater applications. In addition, to support the need for accessible data sets, we provide a novel GitHub repository that clusters the sonar data sets in a single place, open for community contributions. The discussion on robustness, through the lenses of neural network verification, adversarial attacks, and OOD detection, underscores the need for the resilience of sonar-based DL models. The proposed workflow for enhancing model robustness aims to mitigate uncertainties, resulting in more reliable and safer underwater robotic operations. This increased implementation of onboard DL for underwater missions results in the need for robustness of sonar-based DL models, which will play a pivotal role in ensuring mission success and safety by bridging the gap between theoretical robustness and practical, real-world efficacy. Advancing this frontier will provide new potential in autonomous underwater navigation and data collection. This paper encourages future research to focus on sonar DL models' uncertainties to improve their robustness. Future work should reduce the gap between sim-to-real for sonar data collection and validation, improving the current SOA data sets by publishing collected data to create a baseline data set. Finally, the current method for DL robustness, such as neural network verification, adversarial attack, and OOD, in the scope of sonar images require a specific focus to ensure reliable DL predictions and AUV safety.

REFERENCES

- [1] J. Wu et al., "Multi-AUV motion planning for archeological site mapping and photogrammetric reconstruction," *J. Field Robot.*, vol. 36, no. 7, pp. 1250–1269, 2019, doi: [10.1002/rob.21905](https://doi.org/10.1002/rob.21905).
- [2] T. Nakatani et al., "AUV 'TUNA-SAND' and its exploration of hydrothermal vents at kagoshima bay," in *Proc. OCEANS 2008 - MTS/IEEE Kobe Techno-Ocean*, 2008, pp. 1–5.
- [3] M. Wright, W. Gorma, Y. Luo, M. Post, Q. Xiao, and A. Durrant, "Multi-actuated AUV body for windfarm inspection: Lessons from the bio-inspired robofish field trials," in *Proc. 2020 IEEE/OES Auton. Underwater Veh. Symp. (AUV)*, St Johns, NL, Canada, 2020, pp. 1–6.
- [4] P. E. Hagen, N. Storkersen, K. Vestgard, and P. Kartvedt, "The HUGIN 1000 autonomous underwater vehicle for military applications," in *Proc. Oceans 2003. Celebrating the Past... Teaming Toward the Future*, San Diego, CA, USA, 2003, pp. 1141–1145.
- [5] M. Aubard, L. Antal, M. Madureira, and E. Abraham, "SWDD: Sonar wall detection dataset," Jan. 2024, doi: [10.5281/zenodo.10528135](https://doi.org/10.5281/zenodo.10528135).
- [6] E.-h. Lee, B. Park, M.-H. Jeon, H. Jang, A. Kim, and S. Lee, "Data augmentation using image translation for underwater sonar image segmentation," *PLoS One*, vol. 17, Aug. 2022, Art. no. e0272602.
- [7] L. Christensen et al., "Recent advances in AI for navigation and control of underwater robots," *Curr. Robot. Rep.*, vol. 3, pp. 165–175, 2022.
- [8] T. R. Krogstad, M. Wiig, P. Cleophas, and O. Midtgaard, "Autonomous object identification in mine countermeasure missions," in *Proc. OCEANS '11 MTS/IEEE KONA*, Waikoloa, HI, USA, 2011, pp. 1–9.
- [9] L. Character, A.O. Jr, T. Beach, and S. Luzzadder-Beach, "Archaeologic machine learning for shipwreck detection using LiDAR and sonar," *Remote Sens.*, vol. 13, no. 9, 2021, Art. no. 1759.
- [10] M. Aubard et al., "LSTS toolchain framework for deep learning implementation into autonomous underwater vehicle," in *Proc. OCEANS 2023 - Limerick*, 2023, pp. 1–6.
- [11] M. Aubard, S. Quijano, O. Álvarez Tuñón, L. Antal, M. Costa, and Y. Brodskiy, "Mission planning and safety assessment for pipeline inspection using autonomous underwater vehicles: A framework based on behavior trees," in *Proc. OCEANS 2024 - Singap.*, 2024, pp. 1–6.
- [12] A. Tocchetti et al., "A.I. robustness: A human-centered perspective on technological challenges and opportunities," *ACM Comput. Surv.*, vol. 55, no. 12, pp. 1–38, 2023, doi: [10.1145/3665926](https://doi.org/10.1145/3665926).
- [13] N. Fulton and A. Platzer, "Safe AI for CPS (invited paper)," in *Proc. 2018 IEEE Int. Test Conf. (ITC)*, Phoenix, AZ, USA, 2018, pp. 1–7.
- [14] R. Shwartz-Ziv and N. Tishby, "Opening the black box of deep neural networks via information," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2390–2398, doi: [10.1109/ICCV.2017.259](https://doi.org/10.1109/ICCV.2017.259).
- [15] J. Whittlestone and S. Clarke, "AI challenges for society and ethics," in *The Oxford Handbook of AI Governance*, Oxford, U.K.: Oxford Univ. Press, Apr. 2024, doi: [10.1093/oxfordhb/9780197579329.013.3](https://doi.org/10.1093/oxfordhb/9780197579329.013.3).
- [16] D. Gohel and V. Singh, "Explainable AI: Current status and future directions," *Electronics*, vol. 9, no. 11, 2020, Art. no. 1972, doi: [10.3390/electronics9111972](https://doi.org/10.3390/electronics9111972).
- [17] B. Casey, A. Farhangi, and R. Vogl, "Rethinking explainable machines: The GDPRs 'right to explanation' debate and the rise of algorithmic audits in enterprise," *Berkeley Technol. Law J.*, vol. 34, p. 145, 2019.
- [18] L. H. Gilpin et al., "Explaining explanations: An overview of interpretability of machine learning," in *Proc. 2018 IEEE 5th Int. Conf. Data Sci. Adv. Anal.*, Turin, Italy, 2018, pp. 80–89, doi: [10.1109/DSAA.2018.00018](https://doi.org/10.1109/DSAA.2018.00018).
- [19] X. Lurton, "An introduction to underwater acoustics: Principles and applications," 2010. [Online]. Available: <https://api.semanticscholar.org/CorpusID:109354879>
- [20] D. Neupane and J. Seok, "A review on deep learning-based approaches for automatic sonar target recognition," *Electronics*, vol. 9, no. 11, 2020, Art. no. 1972. [Online]. Available: <https://www.mdpi.com/2079-9292/9/11/1972>
- [21] Y. Steiniger, D. Kraus, and T. Meisen, "Survey on deep learning based computer vision for sonar imagery," *Eng. Appl. Artif. Intell.*, vol. 114, 2022, Art. no. 105157. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0952197622002718>
- [22] A. Khan, M. Fouda, D.-T. Do, A. Almaleh, A. Alqahtani, and A. Rahman, "Underwater target detection using deep learning: Methodologies, challenges, applications, and future evolution," *IEEE Access*, vol. 12, pp. 12618–12635, 2024.
- [23] B. Teng and H. Zhao, "Underwater target recognition methods based on the framework of deep learning: A survey," *Int. J. Adv. Robot. Syst.*, vol. 17, no. 6, 2020, Art. no. 1729881420976307, doi: [10.1177/1729881420976307](https://doi.org/10.1177/1729881420976307).
- [24] Y. Tian, L. Lan, and H. Guo, "A review on the wavelet methods for sonar image segmentation," *Int. J. Adv. Robot. Syst.*, vol. 17, no. 4, 2020, Art. no. 1729881420936091, doi: [10.1177/1729881420936091](https://doi.org/10.1177/1729881420936091).
- [25] A. Yassir, S. Jai Andaloussi, O. Ouchetto, K. Mamza, and M. Serghini, "Acoustic fish species identification using deep learning and machine learning algorithms: A systematic review," *Fisheries Res.*, vol. 266, 2023, Art. no. 106790. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0165783623001832>
- [26] Y. Chai, H. Yu, L. Xu, D. Li, and Y. Chen, "Deep learning algorithms for sonar imagery analysis and its application in aquaculture: A review," *IEEE Sensors J.*, vol. 23, no. 23, pp. 28549–28563, Dec. 2023.
- [27] L. C. F. Domingos, P. E. Santos, P. S. M. Skelton, R. S. A. Brinkworth, and K. Sammut, "A survey of underwater acoustic data classification methods using deep learning for shoreline surveillance," *Sensors*, vol. 22, no. 6, 2022, Art. no. 2181. [Online]. Available: <https://www.mdpi.com/1424-8220/22/6/2181>
- [28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. in Neural Inf. Process. Syst.*, Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
- [29] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.
- [31] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. 2014 IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 580–587.
- [32] R. Girshick, "Fast R-CNN," in *Proc. 2015 IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 1440–1448.

- [33] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [34] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 779–788.
- [35] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland, 2016, pp. 21–37.
- [36] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA, 2017, pp. 6000–6010.
- [37] M. I. Jordan, "Serial order: A parallel distributed processing approach," *Adv. Psychol.*, vol. 121, pp. 471–495, 1997.
- [38] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [39] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.*, Glasgow, U.K., Berlin, Heidelberg: Springer-Verlag, 2020, pp. 213–229, doi: [10.1007/978-3-030-58452-8_13](https://doi.org/10.1007/978-3-030-58452-8_13).
- [40] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations*, 2021.
- [41] G. Neves, M. Ruiz, J. Fontinele, and L. Oliveira, "Rotated object detection with forward-looking sonar in underwater applications," *Expert Syst. Appl.*, vol. 140, 2020, Art. no. 112870. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417419305809>
- [42] Y. Yu, J. Zhao, Q. Gong, C. Huang, G. Zheng, and J. Ma, "Real-time underwater maritime object detection in side-scan sonar images based on transformer-YOLOV5," *Remote Sens.*, vol. 13, no. 18, 2021, Art. no. 3555.
- [43] M. Aubard, A. Madureira, L. Madureira, and J. Pinto, "Real-time automatic wall detection and localization based on side scan sonar images," in *Proc. 2022 IEEE/OES Auton. Underwater Veh. Symp. (AUV)*, 2022, pp. 1–6.
- [44] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.
- [45] M. Aubard, L. Antal, A. Madureira, and E. Ábrahám, "Knowledge distillation in YOLOX-ViT for side-scan sonar object detection," 2024, *arXiv:2403.09313*.
- [46] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [47] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, Cham, Switzerland, 2015, pp. 234–241.
- [48] L.-C. Chen, G. Papandreou, I. Kokkinos, K. P. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018. [Online]. Available: <https://api.semanticscholar.org/CorpusID:3429309>
- [49] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. 2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Los Alamitos, CA, USA, Jul. 2017, pp. 6230–6239, doi: [10.1109/CVPR.2017.660](https://doi.org/10.1109/CVPR.2017.660).
- [50] W. Chen, Z. Liu, H. Zhang, M. Chen, and Y. Zhang, "A submarine pipeline segmentation method for noisy forward-looking sonar images using global information and coarse segmentation," *Appl. Ocean Res.*, vol. 112, 2021, Art. no. 102691. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0141118721001681>
- [51] F. Yu et al., "Side-scan sonar images segmentation for AUV with recurrent residual convolutional neural network module and self-guidance module," *Appl. Ocean Res.*, vol. 113, 2021, Art. no. 102608. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0141118721000857>
- [52] O. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.
- [53] S. Zheng et al., "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 6877–6886.
- [54] R. Strudel, R. Garcia, I. Laptev, and C. Schmid, "Segmenter: Transformer for semantic segmentation," in *Proc. 2021 IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Los Alamitos, CA, USA, Oct. 2021, pp. 7242–7252, doi: [10.1109/ICCV48922.2021.00717](https://doi.org/10.1109/ICCV48922.2021.00717).
- [55] Y. Liang, X. Zhu, and J. Zhang, "MiTU-Net: An efficient mix transformer U-like network for forward-looking sonar image segmentation," in *Proc. 2022 IEEE 2nd Int. Conf. Comput. Commun. Artif. Intell. (CCAI)*, 2022, pp. 149–154.
- [56] J. He, J. Chen, H. Xu, and Y. Yu, "SonarNet: Hybrid CNN-transformer-HOG framework and multifeature fusion mechanism for forward-looking sonar image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4203217, doi: [10.1109/TGRS.2024.3368659](https://doi.org/10.1109/TGRS.2024.3368659).
- [57] O. Álvarez Tuñón, L. R. Marnet, M. Aubard, L. Antal, M. Costa, and Y. Brodskiy, "SubPipe: A submarine pipeline inspection dataset for segmentation and visual-inertial localization," in *Proc. OCEANS 2024 - Singap.*, 2024, pp. 1–7.
- [58] M. J. Islam et al., "Semantic segmentation of underwater imagery: Dataset and benchmark," in *Proc. 2020 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Las Vegas, NV, USA, 2020, pp. 1769–1776.
- [59] J. Leonard and H. Durrant-whyte, "Simultaneous map building and localization for an autonomous mobile robot," in *Proc. IEEE/RSJ Int. Workshop Intell. Robots Syst.*, 1991, pp. 1442–1447.
- [60] R. E. Kalman, "A new approach to linear filtering and prediction problems," *ASME. J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, Mar. 1960, doi: [10.1115/1.3662552](https://doi.org/10.1115/1.3662552).
- [61] G. Grisetti, R. Kümmerle, C. Stachniss, and W. Burgard, "A tutorial on graph-based SLAM," *IEEE Trans. Intell. Transp. Syst. Mag.*, vol. 2, no. 4, pp. 31–43, Winter 2010.
- [62] A. J. Oliveira, B. M. Ferreira, and N. A. Cruz, "Feature-based underwater localization using imaging sonar in confined environments," in *Proc. OCEANS 2021: San Diego - Porto*, 2021, pp. 1–7.
- [63] J. Zhang, Y. Xie, L. Ling, and J. Folkesson, "A fully-automatic side-scan sonar simultaneous localization and mapping framework," *IET Radar Sonar Navigation*, vol. 18, pp. 674–683, Nov. 2024, doi: [10.1049/rsn2.12500](https://doi.org/10.1049/rsn2.12500).
- [64] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Toward geometric deep SLAM," 2017, *arXiv:1707.07410*.
- [65] F. Hidalgo and T. Bräunl, "Review of underwater slam techniques," in *Proc. 6th Int. Conf. Automat. Robot. Appl. (ICARA)*, Queenstown, New Zealand, 2015, pp. 306–311.
- [66] S. Zhang et al., "Visual SLAM for underwater vehicles: A survey," *Comput. Sci. Rev.*, vol. 46, 2022, Art. no. 100510.
- [67] S. Rahman, A. Q. Li, and I. Rekleitis, "SVIn2: A multi-sensor fusion-based underwater slam system," *Int. J. Robot. Res.*, vol. 41, no. 11/12, pp. 1022–1042, 2022.
- [68] S. Rahman, A. Q. Li, and I. Rekleitis, "SVIn2: An underwater SLAM system using sonar, visual, inertial, and depth sensor," in *Proc. 2019 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2019, pp. 1861–1868.
- [69] D. Loeblis, R. Sutton, J. Chudley, and W. Naeem, "Adaptive tuning of a kalman filter via fuzzy logic for an intelligent AUV navigation system," *Control Eng. Pract.*, vol. 12, no. 12, pp. 1531–1539, 2004. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0967066103002582>
- [70] J. Zhang, Y. Xie, L. Ling, and J. Folkesson, "A dense subframe-based SLAM framework with side-scan sonar," *IEEE J. Ocean. Eng.*, vol. 49, no. 1, pp. 123–134, Jan. 2024, doi: [10.1109/JOE.2023.1234567](https://doi.org/10.1109/JOE.2023.1234567).
- [71] J. P. Coftelt, P. Kampmann, and B. Wehbe, "SAVOR: Sonar-aided visual odometry and reconstruction for autonomous underwater vehicles*," in *Proc. 2024 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2024, pp. 1236–1243.
- [72] X. Wang, X. Fan, P. Shi, J. Ni, and Z. Zhou, "An overview of key slam technologies for underwater scenes," *Remote Sens.*, vol. 15, no. 10, 2023, Art. no. 2496. [Online]. Available: <https://www.mdpi.com/2072-4292/15/10/2496>
- [73] W. Yang, S. Fan, S. Xu, P. King, B. Kang, and E. Kim, "Autonomous underwater vehicle navigation using sonar image matching based on convolutional neural network," *IFAC-PapersOnLine*, vol. 52, no. 21, pp. 156–162, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2405896319321858>
- [74] B. Mishachandar and S. Vairamuthu, "Diverse ocean noise classification using deep learning," *Appl. Acoust.*, vol. 181, 2021, Art. no. 108141.
- [75] K. Xie, J. Yang, and K. Qiu, "A dataset with multibeam forward-looking sonar for underwater object detection," *Sci. Data*, vol. 9, Dec. 2022, Art. no. 739.
- [76] X. Wang, Q. Li, J. Yin, X. Han, and W. Hao, "An adaptive denoising and detection approach for underwater sonar image," *Remote Sens.*, vol. 11, no. 4, 2019, Art. no. 396. [Online]. Available: <https://www.mdpi.com/2072-4292/11/4/396>

- [77] C. Wang, L. Shen, Y. Fan, T. Chen, and X. Tan, "Sonar image denoising based on anisotropic guided filtering," in *Proc. 5th Int. Conf. Intell. Auton. Syst. (ICoIAS)*, 2022, pp. 54–59.
- [78] X. Zhou, C. Yu, X. Yuan, and C. Luo, "Deep denoising method for side scan sonar images without high-quality reference data," in *Proc. 2nd Int. Conf. Comput. Control Robot. (ICCCR)*, 2022, pp. 241–245.
- [79] T. Liu, S. Yan, and G. Wang, "Remove and recover: Two stage convolutional autoencoder based sonar image enhancement algorithm," *Multimedia Tools Appl.*, vol. 83, pp. 55963–55979, 2023, doi: [10.1007/s11042-023-17673-z](https://doi.org/10.1007/s11042-023-17673-z).
- [80] F. Ji, G. Li, S. Lu, and J. Ni, "Research on a feature enhancement extraction method for underwater targets based on deep autoencoder networks," *Appl. Sci.*, vol. 14, no. 4, 2024, Art. no. 1341. [Online]. Available: <https://www.mdpi.com/2076-3417/14/4/1341>
- [81] T. Sejnowski and R. Gorman, "Connectionist bench (sonar, mines vs. rocks)," *UCI Mach. Learn. Reposit.*, 1988, doi: [10.24432/C5T01Q](https://doi.org/10.24432/C5T01Q).
- [82] S. Sugiyama, M. Minowa, and M. Schaefer, "Side-scan sonar and water profiler data from lagoon grey," 2019. [Online]. Available: <https://data.mendeley.com/datasets/jp252p9m9sc/1>
- [83] S. Sugiyama, M. Minowa, and M. Schaefer, "Underwater ice terrace observed at the front of glacier grey, a freshwater calving glacier in patagonia," *Geophysical Res. Lett.*, vol. 46, no. 5, pp. 2602–2609, 2019. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL081441>
- [84] Government of Ireland, "Ireland's open data portal," Accessed: Feb. 07, 2024. [Online]. Available: https://data.gov.ie/dataset?q=sonar&sort=title_stringdesc
- [85] Government of USA, "U.S. Government's open data," Accessed: Feb. 07, 2024. [Online]. Available: <https://data.gov/>
- [86] X. Du, "Side-scan sonar images of marine engineering geology (Marine_PULSE Dataset)," *Revealing Potential Deep Learn. Detect. Submarine Pipelines Side-Scan Sonar Images, Investigation Pre-Train. Datasets*, vol. 15, no. 19, 2023, Art. no. 4873, doi: [10.5281/zenodo.7922705](https://doi.org/10.5281/zenodo.7922705).
- [87] X. Du, Y. Sun, Y. Song, L. Dong, and X. Zhao, "Revealing the potential of deep learning for detecting submarine pipelines in side-scan sonar images: An investigation of pre-training datasets," *Remote Sens.*, vol. 15, no. 19:4873, 2023. [Online]. Available: <https://www.mdpi.com/2072-4292/15/19/4873>
- [88] E. Andreoli et al., "Side scan sonar mosaics of selected northern adriatic mesophotic biogenic reefs off venice (GeoTIFF image)," *Zenodo*, Mar. 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.4608083>
- [89] A. V. Sethuraman et al., "Machine learning for shipwreck segmentation from side scan sonar imagery: Dataset and benchmark," 2024, *arXiv:2401.14546*.
- [90] N. Dahn et al., "An acoustic and optical dataset for the perception of underwater unexploded ordnance (UXO)," in *Proc. OCEANS 2024 - Halifax*, 2024, doi: [10.1109/OCEANS55160.2024.10754316](https://doi.org/10.1109/OCEANS55160.2024.10754316).
- [91] F. D. B. Martinez-Clavel Valles, H. H. Rajani, N. Graciaset, and R. Garcia, "A large scale side-scan sonar dataset of seafloor sediments for self-supervised pretraining," *Zenodo*, Nov. 2023, doi: [10.5281/zenodo.10209445](https://doi.org/10.5281/zenodo.10209445).
- [92] N. P. Santos and R. Moura, "Side-scan sonar imaging for mine detection," Jan. 2024. [Online]. Available: https://figshare.com/articles/dataset/_i_Side-scan_sonar_imaging_for_Mine_detection_i_/24574879
- [93] M. Ogburn, J. Lefcheck, and J. Olson, "DIDSON fish data and code for analysis," Jan. 2023. [Online]. Available: https://smithsonian.figshare.com/articles/dataset/DIDSON_fish_data_and_code_for_analysis/19611510
- [94] T.-M. Perivolioti, M. Tušer, D. Terzopoulos, S. P. Sgardelis, and I. Antoniou, "Optimising the workflow for fish detection in didson (dual-frequency identification sonar) data with the use of optical flow and a genetic algorithm," *Water*, vol. 13, no. 9, 2021, Art. no. 1304. [Online]. Available: <https://www.mdpi.com/2073-4441/13/9/1304>
- [95] Marine debris turntable. 2024. Accessed: Feb. 27, 2024. [Online]. Available: <https://github.com/mvaldenegro/marine-debris-fls-datasets>
- [96] D. Singh and M. Valdenegro-Toro, "The marine debris dataset for forward-looking sonar semantic segmentation," in *Proc. 2021 IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, 2021, pp. 3734–3742. [Online]. Available: <https://api.semanticscholar.org/CorpusID:237091259>
- [97] J. T. Cobb, "Synthetic aperture sonar seabed environment dataset (SASSED)," *Mendeley Data*, vol. V4, Aug. 2022. [Online]. Available: <https://doi.org/10.17632/s5j5gzr2vc.4>
- [98] "NKSID," 2024. Accessed: Mar. 20, 2024. [Online]. Available: <https://github.com/Jorwnpay/Sonar-OLTR>
- [99] W. Jiao, J. Zhang, and C. Zhang, "Open-set recognition with long-tail sonar images," *Expert Syst. Appl.*, vol. 249, 2024, Art. no. 123495. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417424003609>
- [100] M. Aubard, L. Antal, A. Madureira, L. F. Teixeira, and E. Ábrahám, "ROSAR: An adversarial re-training framework for robust side-scan sonar object detection," 2024. [Online]. Available: <https://arxiv.org/abs/2410.10554>
- [101] M. Bernardi et al., "AURORA, a multi sensor dataset for robotic ocean exploration," in *IEEE Dataport*, Oct. 2020, doi: [10.21227/nmms-te61](https://doi.org/10.21227/nmms-te61).
- [102] K. Krasnosky, C. Roman, and D. Casagrande, "A bathymetric mapping and slam dataset with high-precision ground truth for marine robotics," *Int. J. Robot. Res.*, vol. 41, no. 1, pp. 12–19, 2022, doi: [10.1177/02783649211044749](https://doi.org/10.1177/02783649211044749).
- [103] A. Mallios, E. Vidal, R. Campos, and M. Carreras, "Underwater caves sonar dataset," *Int. J. Robot. Res.*, vol. 36, pp. 1247–1251, 2017. [Online]. Available: <https://api.semanticscholar.org/CorpusID:39662325>
- [104] G. Huo, Z. Wu, and J. Li, "Underwater object classification in sidescan sonar images using deep transfer learning and semisynthetic training data," *IEEE Access*, vol. 8, pp. 47407–47418, 2020.
- [105] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.
- [106] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. 2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.
- [107] M. Prats, J. Pérez, J. Soler Fernandez, and P. Sanz, "An open source tool for simulation and supervision of underwater intervention missions," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 2577–2582.
- [108] D.-H. Gwon, J. Kim, M. H. Kim, H. G. Park, T. Y. Kim, and A. Kim, "Development of a side scan sonar module for the underwater simulator," in *Proc. 14th Int. Conf. Ubiquitous Robots Ambient Intell. (URAI)*, 2017, pp. 662–665.
- [109] M. M. M. Manhães, S. A. Scherer, M. Voss, L. R. Douat, and T. Rauschenbach, "UUV simulator: A gazebo-based package for underwater intervention and multi-robot simulation," in *Proc. OCEANS 2016 MTS/IEEE Monterey*, 2016, pp. 1–8.
- [110] R. Cerqueira, T. Trocoli, G. Neves, S. Joyeux, J. Albiez, and L. Oliveira, "A novel GPU-based sonar simulator for real-time applications," *Comput. Graph.*, vol. 68, pp. 66–76, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0097849317301371>
- [111] M. M. Zhang et al., "Dave aquatic virtual environment: Toward a general underwater robotics simulator," in *Proc. 2022 IEEE/OES Auton. Underwater Veh. Symp. (AUV)*, 2022, pp. 1–8.
- [112] P. Cieślak, "Stonefish: An advanced open-source simulation tool designed for marine robotics, with a ROS interface," in *Proc. OCEANS 2019 - Marseille*, 2019, pp. 1–6.
- [113] A. Abdelhakim, O. Álvarez Tuñón, H. I. Ugurlu, J. le Fevre Sejersen, Y. Brodskiy, and E. Kayacan, "UNav-SIM: A visually realistic underwater robotics simulator and synthetic data-generation framework," in *Proc. 2023 Int. Conf. Adv. Robot. (ICAR)*, 2023, pp. 570–576.
- [114] I. Loncar et al., "MARUS - A marine robotics simulator," in *Proc. OCEANS, Hampton Roads*, 2022, pp. 1–7.
- [115] E. Potokar, S. Ashford, M. Kaess, and J. G. Mangelson, "HoloOcean: An underwater robotics simulator," in *Proc. 2022 Int. Conf. Robot. Automat. (ICRA)*, 2022, pp. 3040–3046.
- [116] E. Potokar et al., "HoloOcean: Realistic sonar simulation," in *Proc. 2022 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2022, pp. 8450–8456.
- [117] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *Proc. 2017 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2017, pp. 23–30.
- [118] M. Quigley et al., "ROS: An open-source robot operating system," in *Proc. ICRA Workshop Open Source Softw.*, 2009.
- [119] S. Shah, D. Dey, C. Lovett, and A. Kapoor, "AirSim: High-fidelity visual and physical simulation for autonomous vehicles," in *Field and Service Robotics*, Cham, Switzerland: Springer International Publishing, 2018, pp. 621–635.
- [120] T. Watanabe et al., "The rock-gazebo integration and a real-time AUV simulation," in *Proc. 12th Latin Amer. Robot. Symp. 2015 3rd Braz. Symp. Robot. (LARS-SBR)*, 2015, pp. 132–138.
- [121] R. J. Rost, J. M. Kessenich, B. Lichtenbelt, H. Malan, and M. Weiblen, *OpenGL Shading Language*, 2nd ed. Boston, MA, USA: Addison-Wesley Professional, 2006.

- [122] N. Ciuccoli, L. Screpanti, and D. Scaradozzi, "Underwater simulators analysis for digital twinning," *IEEE Access*, vol. 12, pp. 34306–34324, 2024.
- [123] MASTODON, 2024. Accessed: Nov. 25, 2024. [Online]. Available: <https://github.com/Sonar-Sim/MASTODON>
- [124] D. Woods, "An overview of the mastodon sonar simulator," *J. Acoustical Soc. America*, vol. 154, pp. A211–A211, Oct. 2023, doi: [10.1121/10.0023310](https://doi.org/10.1121/10.0023310).
- [125] E. Coiras, A. Ramirez-Montesinos, and J. Groen, "GPU-based simulation of side-looking sonar images," in *Proc. OCEANS 2009-EUROPE*, 2009, pp. 1–6.
- [126] M. Sung et al., "Realistic sonar image simulation using deep learning for underwater object detection," *Int. J. Control Automat. Syst.*, vol. 18, pp. 1–12, Feb. 2020.
- [127] A. V. Sethuraman and K. A. Skinner, "STARS: Zero-shot sim-to-real transfer for segmentation of shipwrecks in sonar imagery," in *Proc. Britis. Mach. Vis. Conf.*, 2023.
- [128] I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [129] T. Miyato, A. M. Dai, and I. Goodfellow, "Adversarial training methods for semi-supervised text classification," in *Proc. Int. Conf. Learn. Representations*, 2017.
- [130] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *Proc. 2018 IEEE Int. Conf. Robot. Automat. (ICRA)*, 2018, pp. 7159–7165.
- [131] N. Jaber, B. Wehbe, and F. Kirchner, "Sonar2Depth: Acoustic-based 3D reconstruction using cGANs," in *Proc. 2023 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2023, pp. 5828–5835.
- [132] J. Ho, A. Jain, and P. Abbeel, "Denosing diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, Curran Associates, Inc., 2020, pp. 6840–6851. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bfcfec8584af0d967f1ab10179ca4b-Paper.pdf
- [133] Y. Zhiwei, J. Zhao, H. Zhang, Y. Yu, and C. Huang, "A side-scan sonar image synthesis method based on a diffusion model," *J. Mar. Sci. Eng.*, vol. 11, May 2023, Art. no. 1103.
- [134] Z. Bai, H. Xu, Q. Ding, and X. Zhang, "Side-scan sonar image classification with zero-shot and style transfer," *IEEE Trans. Instrument. Measur.*, vol. 73, 2024, Art. no. 5009415, doi: [10.1109/TIM.2024.3352693](https://doi.org/10.1109/TIM.2024.3352693).
- [135] Y. Xu, Y. Zhang, H. Wang, and X. Liu, "Underwater image classification using deep convolutional neural networks and data augmentation," in *Proc. 2017 IEEE Int. Conf. Signal Process. Commun. Comput. (ICSPCC)*, 2017, pp. 1–5.
- [136] A. Albarghouthi, "Introduction to neural network verification," *Found. Trends Program. Lang.*, vol. 7, no. 1/2, pp. 1–157, 2021, doi: [10.1561/25000000051](https://doi.org/10.1561/25000000051).
- [137] X. Huang et al., "A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability," *Comput. Sci. Rev.*, vol. 37, 2020, Art. no. 100270. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1574013719302527>
- [138] S. Zhang, T.-W. Weng, P.-Y. Chen, C.-J. Hsieh, and L. Daniel, "Beta-CROWN: Efficient bound propagation with per-neuron split constraints for complete and incomplete neural network verification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, Art. no. 2289.
- [139] G. Singh, T. Gehr, M. Püschel, and M. Vechev, "An abstract domain for certifying neural networks," in *Proc. ACM Program. Lang.*, vol. 3, 2019, Art. no. 41.
- [140] G. Katz, C. Barrett, D. L. Dill, K. Julian, and M. J. Kochenderfer, "Reluplex: An efficient SMT solver for verifying deep neural networks," in *Proc. Int. Conf. Comput. Aided Verification*, 2017, pp. 97–117.
- [141] G. Katz et al., "The marabou framework for verification and analysis of deep neural networks," in *Proc. Int. Conf. Comput. Aided Verification*, Jul. 2019, pp. 443–452.
- [142] Y. Qin, J. Hu, and B. Wu, "Toward evaluating the robustness of deep learning based rain removal algorithm in autonomous driving," in *Proc. 2023 Secure Trustworthy Deep Learn. Syst. Workshop*, New York, NY, USA, 2023, Art. no. 7, doi: [10.1145/3591197.3591309](https://doi.org/10.1145/3591197.3591309).
- [143] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Proc. Int. Conf. Learn. Representations*, 2015.
- [144] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," in *Proc. Int. Conf. Learn. Representations*, 2018.
- [145] S.-M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "DeepFool: A simple and accurate method to fool deep neural networks," in *Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 2574–2582.
- [146] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, and A. Swami, "The limitations of deep learning in adversarial settings," in *Proc. 2016 IEEE Eur. Symp. Secur. Privacy*, 2015, pp. 372–387. [Online]. Available: <https://api.semanticscholar.org/CorpusID/7004303>
- [147] N. Papernot, P. McDaniel, and I. Goodfellow, "Transferability in machine learning: From phenomena to black-box attacks using adversarial samples," 2016, *arXiv:1605.07277*.
- [148] P.-Y. Chen, H. Zhang, Y. Sharma, J. Yi, and C.-J. Hsieh, "ZOO: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models," in *Proc. 10th ACM Workshop Artif. Intell. Secur.*, Nov. 2017, pp. 15–26, doi: [10.1145/3128572.3140448](https://doi.org/10.1145/3128572.3140448).
- [149] W. Brendel, J. Rauber, and M. Bethge, "Decision-based adversarial attacks: Reliable attacks against black-box machine learning models," in *Proc. Int. Conf. Learn. Representations*, 2018.
- [150] A. Ilyas, L. Engstrom, A. Athalye, and J. Lin, "Black-box adversarial attacks with limited queries and information," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2142–2151. [Online]. Available: <https://api.semanticscholar.org/CorpusID:5046541>
- [151] A. A. dos Santos et al., "Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVs," *Sensors*, vol. 19, no. 16, 2019, Art. no. 3595.
- [152] M. Yan, Z. Li, X. Yu, and C. Jin, "An end-to-end deep learning network for 3D object detection from RGB-D data based on hough voting," *IEEE Access*, vol. 8, pp. 138810–138822, 2020.
- [153] Q. Ma, L. Jiang, W. Yu, R. Jin, Z. Wu, and F. Xu, "Training with noise adversarial network: A generalization method for object detection on sonar image," in *Proc. 2020 IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Los Alamitos, CA, USA, Mar. 2020, pp. 718–727, doi: [10.1109/WACV45572.2020.9093467](https://doi.org/10.1109/WACV45572.2020.9093467).
- [154] Q. Ma, L. Jiang, and W. Yu, "Lambertian-based adversarial attacks on deep-learning-based underwater side-scan sonar image classification," *Pattern Recognit.*, vol. 138, 2023, Art. no. 109363. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S003132032300064X>
- [155] S. Feng, X. Zhu, S. Ma, and Q. Lan, "Adversarial attacks in underwater acoustic target recognition with deep learning models," *Remote Sens.*, vol. 15, no. 22, 2023, Art. no. 5386. [Online]. Available: <https://www.mdpi.com/2072-4292/15/22/5386>
- [156] F. Cai and X. Koutsoukos, "Real-time out-of-distribution detection in learning-enabled cyber-physical systems," in *Proc. 2020 ACM/IEEE 11th Int. Conf. Cyber-Phys. Syst.*, Sydney, NSW, Australia, 2020, pp. 174–183, doi: [10.1109/ICCPS48487.2020.00025](https://doi.org/10.1109/ICCPS48487.2020.00025).
- [157] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *Proc. Int. Conf. Learn. Representations*, 2017.
- [158] S. Liang, Y. Li, and R. Srikant, "Enhancing the reliability of out-of-distribution image detection in neural networks," in *Proc. Int. Conf. Learn. Representations*, 2018.
- [159] K. Lee, K. Lee, H. Lee, and J. Shin, "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," in *Proc. 32nd Conf. Neural Inf. Process. Syst.*, Montréal, Canada, 2018.
- [160] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proc. 2nd Int. Conf. Learn. Representations*, Banff, AB, Canada, Apr. 2014.
- [161] I. Higgins et al., " β -VAE: Learning basic visual concepts with a constrained variational framework," in *Proc. Int. Conf. Learn. Representations*, 2017.
- [162] I. D. Gerg and C. F. Cotner, "A perceptual metric prior on deep latent space improves out-of-distribution synthetic aperture sonar image classification," in *Proc. 2023 IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 6576–6579.
- [163] W. Jiao and J. Zhang, "Sonar images classification while facing long-tail and few-shot," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4210420, doi: [10.1109/TGRS.2022.3211847](https://doi.org/10.1109/TGRS.2022.3211847).
- [164] M. Cook, A. Zare, and P. Gader, "Outlier detection through null space analysis of neural networks," in *Proc. 37th Int. Conf. Mach. Learn. Workshop Uncertainty Robustness Mach. Learn.*, 2020.
- [165] M. Cook, A. Khoury, A. Zare, and P. Gader, "Null space analysis for detecting unknown objects during automatic target recognition tasks in sonar data," in *Proc. OCEANS 2023 - MTS/IEEE U.S. Gulf Coast*, Sep. 2023, pp. 1–5.

- [166] V. Schwag, M. Chiang, and P. Mittal, "SSD: A unified framework for self-supervised outlier detection," in *Proc. Int. Conf. Learn. Representations*, 2021.
- [167] Y. Sun, C. Guo, and Y. Li, "ReAct: Out-of-distribution detection with rectified activations," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2021, Art. no. 12. [Online]. Available: <https://api.semanticscholar.org/CorpusID:244709089>
- [168] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [169] C. E. Rasmussen, "in *Gaussian Processes in Machine Learning*. Berlin, Heidelberg: Springer, 2004, doi: [10.1007/978-3-540-28650-9_4](https://doi.org/10.1007/978-3-540-28650-9_4).
- [170] M. Abdar et al., "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Inf. Fusion*, vol. 76, pp. 243–297, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1566253521001081>
- [171] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods," *Mach. Learn.*, vol. 110, no. 3, pp. 457–506, Mar. 2021, doi: [10.1007/s10994-021-05946-3](https://doi.org/10.1007/s10994-021-05946-3).
- [172] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA, 2017, pp. 5580–5590.
- [173] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. 33rd Int. Conf. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.
- [174] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 6405–6416.
- [175] L. R. Fuchs, A. Norén, and P. Johansson, "GAN-enhanced simulated sonar images for deep learning based detection and classification," in *Proc. OCEANS 2022 - Chennai*, 2022, pp. 1–5.
- [176] P. Tarling, M. Cantor, A. Clapés, and S. Escalera, "Deep learning with self-supervision and uncertainty regularization to count fish in underwater images," *PLoS One*, vol. 17, no. 5, 2022, Art. no. e0267759. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85129439886&doi=10.1371%2Fjournal.pone.0267759&partnerID=40&md5=9eec9de7ad5bc349a15b5472538024fe>



Martin Aubard received the B.Sc. degree in electrical engineering and industrial computing from IUT de l'Indre, Châteauroux, France, in 2017, and the M.Sc. degree in control and industrial computing from the University of Technology of Troyes, Troyes, France, in 2021. He is currently working toward the Ph.D. degree in sonar-based deep learning for autonomous underwater vehicles at the Faculdade de Engenharia da Universidade do Porto (FEUP), Porto, Portugal, in collaboration with OceanScan Marine Systems and Technology, Matosinhos, Portugal. His research is

conducted within the European "Reliable AI for Marine Robotics" project. His research interests include sonar-based object detection, robust object detection, and real-time guidance for autonomous underwater vehicles.



Ana Madureira (Senior Member, IEEE) received the B.S. degree in computer science engineering from the Instituto Superior de Engenharia do Porto (ISEP), M.S. degree in electrical and computers engineering from the Faculdade de Engenharia da Universidade do Porto (FEUP), and the Ph.D. degree in production and systems from the University of Minho, Portugal, Habilitation in informatics, in 1993, 1996, 2003, and 2022, respectively. She is a Coordinator Professor with Habilitation with ISEP/P.PORTO and has significant experience in teaching Computer Science

undergraduate and graduate courses and in master and Ph.D. thesis supervision (since 1994). She is the Director of R&D Interdisciplinary Studies Research Center, Porto, Portugal. She has been involved in more than 20 R&D projects from which she was Principal Investigator of four R&D projects, mainly on the AI-based developing of intelligent scheduling systems. She has authored or coauthored three books and more than 150 papers in international scientific conference proceedings, journals, and book chapters. She has a patent authored or coauthored on the area of her research work which include "multi-agent system for distributed manufacturing scheduling with genetic algorithms and Tabu Search."



Luís Teixeira received the Licenciatura degree (5-year degree) in electrical and computer engineering, the M.Sc. degree in computer networks and services, and the Ph.D. degree in electrical and computer engineering from Universidade do Porto, Porto, Portugal, in 2001, 2004, and 2009, respectively.

He is an Associate Professor with the Department of Informatics Engineering, Faculdade de Engenharia da Universidade do Porto, Porto, and the Senior Researcher with the Center for Telecommunications and Multimedia of INESC TEC, Porto. Previously, he was a Researcher with INESC Porto (2001–2008), a Visiting Researcher with the University of Victoria (2006), Victoria, BC, Canada, a Senior Scientist with Fraunhofer AICOS (2008–2013), Porto, and the President of APRP, the Portuguese Association for Pattern Recognition (2019–2021). He has coauthored more than 60+ publications in international journals and conferences, having gathered 1100+ citations to his publications, and (co-) supervised or is (co-) supervising 100+ M.Sc. students and seven Ph.D. students. As a Researcher, he has participated in numerous projects and is currently the Principal Investigator of the CAGING project, a research project funded by FCT, the Portuguese Foundation for Science and Technology. His research interests include computer vision, machine learning, and human-centred computing.



José Carlos Queirós Pinto received the Licenciatura degree in computer science from the Faculty of Sciences, University of Porto, Porto, Portugal, in 2005, and the master degree in software engineering from the School of Engineering, University of Porto in 2007. He is a Robotics Engineer and Project Manager at OceanScan Marine Systems and Technology, Matosinhos, Portugal, where he serves as Software Team Leader. Previously, he was a Senior Researcher at the Underwater Systems and Technology Laboratory and a Teaching Assistant at FEUP (2006–2020).

Previously, he was a Senior Researcher with the Underwater Systems and Technology Laboratory and a Teaching Assistant with FEUP (2006–2020), Porto, Portugal. His expertise spans software infrastructures for networked autonomous vehicles, onboard AI-driven autonomy, and the application of robotics in oceanography and archaeology. He has coauthored more than 50 scientific publications, participated in more than 15 international R&D projects, and led numerous field trials involving autonomous underwater systems. His research interests include multivehicle coordination, AI in robotics, and innovative software tools for underwater exploration.