MMinE-SwEEPER

Deliverable Report

Deliverable No. 5.4: Landscape the existing AI approaches for munition detection

Funded by
the European Union

Authors | Denitsa Dimitrova, Josep Bosch, Sebastian Graber, Massimiliano Corsini

Contributing organisations:

DOCUMENT TRACKING DETAILS

Project Information

Project acronym	MMinE-SwEEPER
Project title	Marine Munition in Europe - Solutions with Economic and Ecological Profits for Efficient Remediation
Instrument	Horizon Europe
Grant Agreement	No. 101167839
Starting date	1 October 2024
Duration	31 March 2028
Project Coordinator	GEOMAR Helmholtz Centre for Ocean Research Kiel

Deliverable Information

Deliverable Name	Landscape the existing AI approaches for munition detection
Deliverable Number	5.4
Work Package (WP)	5
Related Task (s)	5.4
Deliverable Beneficiary	Royal Military Academy (RMA)
Author(s)	Denitsa Dimitrova,Josep Bosch, Sebastian Graber, Massimiliano Corsini
Due date	31.03.2026
Submission date	26.03.2026
Type of deliverable	Report
Dissemination level	PU

Acknowledgements

This report was produced by RMA, IQUA, north.io in the scope of the MMinE-SwEEPER project. This project has received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101167839.

Contact details

Questions, comments, and/or concerns about this document should be emailed to denitsa.dimitrova@mil.be subject line "MMinE-SwEEPER Deliverable 5.4 Landscape the existing AI approaches for munition detection".

Contents

Index of Tables

Executive summary

This deliverable provides the scientific foundation for the development of AI-approaches within Work Package 5 of the MMinE-SwEEPER project by presenting a compilation of the state-of-the-art approaches and methodological knowledge from the scientific literature. It primarily contributes to Objective B – Advance robotic applications, 3D imaging and AI-supported data analyses. In particular it lies the groundwork for the technical and operational activities under Objectives B2 and B4.

The report evaluates a wide range of hydroacoustic and optical detection approaches, showing that no single method performs best under all conditions. Modern one-stage detectors, particularly from the YOLO family, offer the best balance between accuracy, speed, and operational feasibility, while traditional methods remain suitable for low-resource systems and two-stage detectors for high-precision tasks. Optical imagery complements sonar by providing detailed visual information for classification and validation. The findings highlight that combining hydroacoustic and optical methods in a multimodal workflow is most effective, enabling robust, scalable, and efficient AI-based UXO detection with improved accuracy and reduced false positives across diverse environments.

Introduction

The detection, classification, and identification of munition- and mine-like objects in underwater environments represent a critical challenge for maritime safety, offshore infrastructure development, and environmental protection. Within the framework of the MMinE-SwEEPER project, Deliverable 5.4 provides a structured literature review of state-of-the-art approaches for analyzing hydroacoustic data as well as optical and acoustic camera imagery, with particular focus on their applicability to unexploded ordnance (UXO) detection. Modern underwater surveys rely increasingly on autonomous underwater vehicles (AUVs), remotely operated vehicles (ROVs), towed platforms, and dedicated sonar systems. These platforms generate vast amounts of heterogeneous data, primarily in two complementary sensing domains:

· Spatial hydroacoustic approaches, including multibeam echosounder (MBES), side-scan sonar (SSS), synthetic aperture sonar (SAS), and acoustic camera systems, and

· Optical imagery, acquired through underwater cameras.

The scale and complexity of these datasets make manual interpretation impractical and economically unsustainable. Automated and scalable data processing solutions are therefore indispensable. However, underwater environments impose severe sensing challenges. Optical data suffer from strong light attenuation, wavelength-dependent absorption, colour distortion, turbidity, low contrast, and highly complex seabed backgrounds. Objects of interest often occupy only a small fraction of the image and may be partially buried, covered by sediment, or obscured by marine growth. Hydroacoustic data, while not affected by optical degradation, introduce their own domain-specific complexities. Sonar imagery depends on backscatter physics, grazing angle, seabed composition, and shadow geometry. Noise, artifacts, and strong variability across survey conditions complicate robust interpretation. In both sensing domains, the detection problem is further aggravated by strong class imbalance: munition and mine-like objects are rare compared to the overwhelming background of benign seabed structures.

Over the past decade, artificial intelligence (AI) and in particular deep learning has emerged as the dominant paradigm for underwater data analysis. AI-based approaches enable scalable processing, improved generalization across environments, and the integration of contextual and geometric cues that are difficult to model using classical rule-based methods. This report synthesizes these approaches, compares domain-specific methodologies, and discusses their strengths, limitations, and integration potential. The objective is to provide a consolidated knowledge base that supports the development of robust, reliable, and scalable AI-driven detection systems within MMinE-SwEEPER, ultimately contributing to more efficient underwater survey operations.

Existing AI approaches for munition detection

Spatial hydroacustic approaches

1. Traditional and unsupervised signal processing

These approaches rely on handcrafted features, statistical models, and unsupervised learning instead of training on large neural datasets. They are designed for real time operations on vehicles with limited computational resources. Instead of "learning" how the targets look like, these methods exploit known geometric and physical properties of the sonar returns to separate man-made objects from natural seabed clutter.

Example of using these kinds of methods on sonar data is presented in the paper of G. F. Dobeck (Dobeck, 2001). In this paper a four-stage algorithm called AMDAC (Advanced Mine Detection and Classification) is used, which includes: Image enhancement, Detection, Stepwise feature selection and Classification. The data used includes 225 SAS and 120 SSS images from two different SSS systems. The algorithm shows performance equal or better than expert sonar operators, however it relies on background normalization which can fail in irregular seabed patterns. As a result, the AMDAC algorithm for the two SSS has a performance of 90% and 92% of Probability of detection and classification (PdPc) of mines with false alarm rates of 42% and 12%. For SAS, AMDAC can perform at 92% PdPc with 64% false alarm. Here it is important to add the performance of the operators which scored 80% PdPc for SSS with 72% false alarms and 62% PdPc for SAS with 23 % false alarms.

Another example of unsupervised method used on sonar images for object detection is presented in the paper of D. P. Williams (Williams, 2015). For that algorithm approximately 30,000 SAS images are used in different geographical locations. The method used therein is a unsupervised three stage cascade based on shadow, ripple and echo detection. The main advantage of the presented approach is the near real-time detection onboard of AUVs. However, the algorithm is stated to have difficulties in areas with heavy sand ripples and posidonia (seagrass). The evaluation of this method is given by Receiver Operating Characteristic (ROC) Curves, which shows how the probability of detection increases compared to the False Alarm Rate (FAR). As a final result for all seabed characteristics, the probability of detection is 88.8%.

2. One-stage detectors

One-stage detectors are deep learning–based approaches designed to perform object localization and classification in a single forward pass of the network, enabling high computational efficiency and real-time performance. These methods process the entire input image simultaneously and directly predict bounding box coordinates along with associated class probabilities. The detection framework is typically centred around a convolutional neural network (CNN) backbone responsible for hierarchical feature extraction, followed by detection heads that operate on multi-scale feature maps. One-stage detectors treat object detection as a direct prediction problem, where bounding box locations and class labels are estimated simultaneously from the input image. This approach removes the need for a separate step that first proposes candidate object regions. As a result, the overall detection process becomes faster and computationally more efficient. Among these approaches, YOLO (You Only Look Once) (Ultralytics, 2026a) is one of the most used due to its balance between detection accuracy and computational speed, and has been increasingly applied to sonar and hydroacoustic imagery in recent years, as presented in the next paragraphs.

A comprehensive analysis on the performance and the development of different YOLO versions until 2023, from YOLOv1 to YOLOv8 and YOLO-NAS is presented in (Terven et al., 2023). However, the development of YOLO versions continued and as of 15 January 2026 there are more versions developed by Ultralytics and freely usable (Ultralytics, 2026b) (released under the APGL 3,0 license).

Early versions of YOLO have been used for SSS images. YOLOv2 is proposed for object detection task on SSS images in (Einsidler et al., 2018). Already in this version improvements were introduces compared to the traditional YOLO such as batch normalization (improves the overall stability of the model), multi-scale training (learning and predicting objects from different dimensions), and anchor boxes (pre-definition of the bounding box).

Examples for YOLO3 on object detection using SSS data are presented in several publications (Li & Cao, 2022), (Wang et al., 2021), (Yulin et al., 2020). Specifically (Yulin et al., 2020) proposes and compares 3 models on shipwreck target recognition: traditional YOLOv3, transfer learning YOLOv3 (retraining by sharing learned model parameters with a new model) and Faster R-CNN model. The results show highest precision (89.76 %), recall (91.66%) in the transfer learning YOLOv3.

YOLOv5 is stated to increase accuracy and reduce the number of the parameters of the model (Ultralytics, 2026b), which also makes it attractive for object detection on underwater images. In (Yang et al., 2024) YOLOv5 has been employed in a study for generation of SSS images that have been trained on object detection. The results show very good performance, with recall of the model trained on SSS real and generated images shows results of 92.3 %.

YOLOv7 and improved versions of it have been used for underwater sonar image object detection as well. One example is shown in (Wen & Zhang, 2023) where YOLOv7 is used in combination with two improvements for better detection of objects: Swin-Transformer, a method used for dynamic attention that focuses on the target regions, and Convolutional Block Attention Module for improving the detection speed and increasing the accuracy. This improved YOLOv7 model is compared to other models such as YOLOv5, Sparse R-CNN and Cascade R-CNN, as well as the original version of YOLOv7, all trained on the same data, 400 SSS images separated into 70% for training, 15% for testing and 15% for validation. The results show highest, 85.32% Average Precision (AP) at an Intersection over Union (IoU) threshold of 50% results of the improved YOLOv7, which is around 5 % higher than all other tested models. Another paper using different improvements to YOLOv7 using sonar data for target recognition is comparing how different improvements affect the initial model (Chen et al., 2024). These improvements to the model include: module that enhances local and global feature capture; a module that enhances target recognition in complex scenes; a module that enhances feature learning, that helps handling clutter targets and targets with diverse shapes and sizes. The paper also presents an improved YOLOv7 that contains all the modules: DA-YOLOv7, with precision of 92.8% and recall of 89.7%. Compared to the YOLOv7 combined with just one of the modules, the combined DA-YOLOv7 scores highest performance on the same dataset.

The latest version of YOLO that is used in publications with sonar images, as of February 2026 is YOLOv11. In a study for naval mine detection using SSS data (Caricchio et al., 2025), the model performed with very high prediction of 93% and recall of 83%. The data used in the paper consists of 1200 images, 881 of which include mines. The database is split into 90% for training, 5% for testing, and 5% for validation.

3. Two-stage detectors

Two-stage detectors have a sequential detection strategy in which object localization and classification are performed in distinct processing stages, allowing for more precise discrimination in complex environments. In the first stage, candidate object regions are generated either through learned region proposal networks or through physics-informed preprocessing techniques (echo intensity thresholding, shadow–highlight extraction, seabed segmentation) (Du, L., Zhang et al., 2020). These proposals aim to identify regions that are likely to contain objects of interest. In the second stage, each candidate region is refined and classified using deep convolutional neural networks, producing final bounding boxes and object labels. Within spatial hydroacoustic applications, two-stage frameworks have demonstrated superior robustness in cluttered seabed conditions, significantly reducing false

alarm rates compared to single-stage detectors. Their ability to integrate acoustic propagation characteristics with data-driven classification makes them particularly well suited for munition detection in real-world environments.
In the study (Agrawal et al., 2025), the authors use Mask R-CNN as the baseline model for underwater mine-like object detection in side-scan sonar images. The paper's main contribution is a Syn2Real (Synthetic to Real) approach, which uses diffusion models to generate synthetic data. While the focus is on improving generalization through synthetic data, the study also evaluates the performance of Mask R-CNN as the detection framework. The model follows a two-stage process: a Region Proposal Network (RPN), which generates candidate object regions, and a mask head, which performs pixel-level segmentation and classification. The training dataset consists of 600 images: 200 real images and 400 synthetic images. The results show an Average Precision (AP) of 83.3% at an Intersection over Union (IoU) threshold of 50% for the model trained on the combined dataset and evaluated on 100 real SSS images.

Another used approach in side-scan sonar images is Faster R-CNN. The Faster R-CNN model consists of a RPN for generating object candidates and a classification/regression head for identifying and refining object locations. An example of its usage is found in a study on object detection in sonar images (Jiang et al., 2020), which employed a dataset of 5,400 sonar images across different categories, such as corpses, shipwrecks, and plane wreckage. The primary focus of the study is on active learning algorithms designed to reduce annotation costs by intelligently selecting the most informative images for labeling. It explores the model's performance under varying amounts of training data per class. While the study emphasizes active learning, it also evaluates the performance of the Faster R-CNN model, which achieved an average precision of slightly over 90% for plane wreckage, close to 90% for shipwrecks, and around 60% for corpses when trained using only 35% of the data (approx. 1,500 images).

A complete table comparing the spatial hydroacoustic approaches is shown on Table 1.

Table 1: Summarization of hydroacoustic approaches with used data and metrics showing the performance of each method

Method	Data	Metrics
Human Operators (baseline)	SSS and SAS datasets from Dobeck (Dobeck, 2001)	SSS: PdPc 80%, FAR 72%; SAS: PdPc 62%, FAR 23%
AMDAC – Traditional Unsupervised (Dobeck, 2001)	225 SAS + 120 SSS images (two SSS systems)	SSS: PdPc 90% & 92%, FAR 42% & 12%; SAS: PdPc 92%, FAR 64%
Shadow–Ripple–Echo Cascade – Unsupervised (Williams, 2015)	~30,000 SAS images from multiple seabeds	Probability of Detection = 88.8% (ROC-based)
YOLOv2 – One-stage (Einsidler et al., 2018)	SSS imagery	No explicit metrics reported
Transfer Learning YOLOv3 – One-stage (Yulin et al., 2020)	SSS shipwreck detection dataset	Precision 89.76%, Recall 91.66%
YOLOv5 – One-stage (Yang et al., 2024)	Real + generated SSS images	Recall 92.3%
Improved YOLOv7 (Swin-Transformer + CBAM) – One-stage (Wen & Zhang, 2023)	400 SSS images separated into 70% for training, 15% for testing and 15% for validation	85.32% mAP50
DA-YOLOv7 – One-stage (Chen et al., 2024)	Sonar dataset	Precision 92.8%, Recall 89.7%
YOLOv11 – One-stage (Caricchio et	1200 SSS images (881 mines),	Precision 93%, Recall 83%

al., 2025)	90/5/5 split	
Mask R-CNN – Two-stage (Agrawal et al., 2025)	600 SSS training images (200 real + 400 synthetic) + 100 real test	mAP50 = 83.3%
Faster R-CNN – Two-stage (Jiang et al., 2020)	5400 SSS images (plane wrecks, shipwrecks, corpses)	Average Precision (using 35% training data): Plane wrecks >90%; Shipwrecks ~90%; Corpses ~60%

Acoustic camera approaches

Traditionally, underwater unexploded ordnance (UXO) and mine detection has relied primarily on Side-Scan Sonar (SSS) and Synthetic Aperture Sonar (SAS) systems, which have historically constituted the dominant acoustic payloads on Autonomous Underwater Vehicles (AUVs). In recent years, however, there has been increasing interest in multibeam sonars with large vertical apertures, commonly referred to as acoustic cameras or Forward-Looking Sonars (FLS). Unlike traditional SSS and SAS systems, FLS generates real-time two-dimensional acoustic images that resemble optical camera outputs. This image-like representation offers a twofold advantage: first, improved human interpretability, as the visual similarity to optical imagery allows operators to more readily detect objects, protrusions, and anomalies compared to classical sonar waterfall displays; and second, transferability of computer vision methods, since the 2D intensity format enables the adaptation of algorithms originally developed for optical imagery, such as convolutional neural networks (CNNs) and modern object detection frameworks, often with minimal architectural modifications. Furthermore, in recent years several manufacturers have begun producing acoustic camera models with good image quality at competitive prices (Blueprint Subsea, 2026), (Tritech, 2026), (Einsidler et al., 2018). The growing adoption of FLS in both commercial and research AUV platforms further reinforces its operational relevance, as modern vehicles increasingly integrate acoustic cameras for navigation (Li et al., 2018), (Franchi et al., 2021), inspection (Zhang et al., 2022), (Tulsook et al., 2018) , obstacle avoidance (Mu et al., 2022), and target detection (Hurtós et al., 2017) ,(Parsons et al., 2017), particularly in turbid or low-visibility environments where optical cameras are ineffective.

With the rapid advancement of artificial intelligence (AI) and deep learning, researchers have begun investigating the use of machine learning methods for automated target detection in FLS imagery. IEEE Journal of Oceanic Engineering (Aubard et al., 2025) provides a comprehensive review of sonar-based deep learning in underwater robotics, analyzing robustness issues specific to acoustic imaging, including speckle noise statistics, reverberation, multipath interference, and domain shift across sonar configurations. The survey highlights that FLS imagery presents distinct challenges compared to optical data, particularly in terms of lower spatial resolution, anisotropic beam patterns, and geometry-dependent shadow formation, all of which must be considered in network design and training strategies.

For example, (Yang et al., 2024) proposed a lightweight deep neural network architecture specifically designed for underwater target detection in FLS images, emphasizing computational efficiency and robustness to sonar noise in real-time applications. Similarly, (Long et al., 2023) proposed a detection framework combining speckle reduction and scene-prior modeling to improve target detection performance in FLS images. Their approach integrates adaptive despeckling preprocessing with a CNN-based detector, demonstrating that modeling sonar-specific image degradation significantly improves detection accuracy compared to directly applying optical-domain architectures. In (Wang et al., 2025), the authors propose a modified YOLO-based framework tailored for forward-looking sonar imagery. Their contribution includes semantic–spatial feature enhancement modules designed to better capture elongated acoustic highlights and shadow regions characteristic of submerged objects. By incorporating sonar-aware feature fusion strategies, the method improves detection robustness under low signal-to-noise ratios and partial occlusions, which are common in FLS data.

Zacchini et al. (Zacchini et al., 2020) presented the development of an Automatic Target Recognition (ATR) methodology to identify and localize potential targets in FLS imagery. The core of the system is a Mask R-CNN architecture trained on real FLS data acquired in the CSSN test basin in La Spezia. Beyond bounding-box detection, the use of instance segmentation enables more precise target delineation, facilitating geometric feature extraction (e.g., size, shape descriptors) that is directly relevant for mine-like object discrimination and post-detection classification stages.

In (Cao et al., 2023), a deep learning-based obstacle candidate area detection algorithm is developed using YOLOv3. The work focuses on real-time obstacle detection for AUV navigation, but from an acoustic processing perspective it demonstrates the feasibility of deploying single-stage detectors directly on raw or minimally preprocessed FLS intensity images. The study evaluates detection performance under varying range and incidence angles, providing insight into how acoustic shadow length and highlight intensity influence network confidence—an aspect highly relevant to UXO-like object detection.

Despite these advances, most AI-based FLS research focuses on generic underwater target detection (Xie et al., 2022), not specifically UXO or mine detection. Scientific Data (Xie et al., 2022) introduces a multibeam forward-looking sonar dataset intended for general underwater object detection benchmarking. While not UXO-specific, it provides standardized annotated acoustic imagery that has enabled reproducible evaluation of detection architectures and comparative performance analysis across models.

At the time of writing this deliverable, the literature specifically addressing UXO detection using FLS imagery remains limited compared to the extensive body of work on SSS and SAS. However, recent efforts have focused on generating datasets suitable for training AI models in this domain. (Dahn et al., 2024) introduced a multimodal dataset combining acoustic and optical data for underwater UXO perception. The acoustic component consists of forward-looking sonar imagery acquired under controlled conditions with known target placements, including UXO-like objects. The dataset provides synchronized modalities, calibration parameters, and ground-truth pose annotations, enabling supervised detection, classification, and cross-modal fusion experiments. Importantly, it supports research into how acoustic shadow geometry and highlight intensity correlate with true object shape.

(Ściegienka & Blachnik, 2024) presented the development of a large forward-looking sonar acoustic image dataset specifically for UXO classification. The dataset was designed to evaluate transfer learning strategies by fine-tuning state-of-the-art CNN architectures pre-trained on optical datasets. The study systematically benchmarks multiple deep neural network models, analysing classification accuracy, generalization across acquisition scenarios, and sensitivity to acoustic noise. Their results demonstrate that, despite domain differences, transfer learning from optical imagery remains viable when combined with sonar-specific augmentation and normalization strategies.

Table 2: Summarization of acoustic camera approaches with used data.

Method	Data	Metrics
Lightweight deep neural network (FLSD-Net) (Yang et al., 2024)	UATD-2022 (Xie et al., 2022) (Non-UXO)	MAP 86.7% Recall 81.9%
UFIDNet Desplecking + CNN detector (Long et al., 2023)	Marine Debris dataset FLS (Ocean Systems Lab) (Non-UXO) UATD-2022 (Xie et al., 2022) (Non-UXO)	AP 70.5%

Yolo framework for FLS YOLO-SONAR (Wang et al., 2025)	Marine Debris dataset FLS (Ocean Systems Lab) (Non-UXO)	mAP 82.3%
Mask R-CNN ATR (Zacchini et al., 2020)	Custom Spezia CSSN basin Dataset (Non-UXO)	mAP 95%
Object avoidance based on YOLOv3 (Cao et al., 2023)	1000 custom images (800 training 200 validation) (Non-UXO)	mAP 97.92% (IoU 0.3) mAP 96.73% (IoU 0.5) mAP 93.36% (IoU 0.7)
Multimodal (FLS/Optic) dataset (Dahn et al., 2024)	ARISs 300 data in water tank with different UXO objects.	Not applicable
Dataset for UXO detection (Ściegienka & Blachnik, 2024)	Simulated data	Not applicable.

Optical camera approaches

Underwater optical cameras are extensively used for inspection, monitoring, and mapping tasks in environments where visibility allows, including habitat surveys (Armstrong et al., 2019), infrastructure inspection (Bobkov et al., 2023), and marine biology studies (Nalmpanti et al., 2023). Unlike acoustic sensors, optical cameras capture high-resolution color and texture information, providing rich visual cues that are critical for detecting fine surface details, corrosion, or debris. However, their performance is constrained by turbidity, low light, and scattering, which reduce image quality at depth or in unclear waters. Despite these challenges, the combination of high resolution and rich appearance data makes optical imagery particularly suitable for automatic object detection and recognition.

Recent advances in deep learning have significantly improved the capabilities of underwater optical detection. ACM Computing Surveys and Multimedia Tools and Applications publish comprehensive surveys highlighting the evolution of convolutional neural networks (CNNs), region-based detectors such as Faster R-CNN and YOLO, and more recently transformer-based architectures for underwater object detection (Dakhil & Khayeat, 2022), (Wang et al., 2020), (Fayaz et al., 2022), (Jian et al., 2024), (Fayaz et al., 2024). These works emphasize the importance of domain adaptation, color correction, contrast enhancement, and physics-aware preprocessing to mitigate underwater image degradation before detection. The integration of image restoration and detection into unified pipelines has also been shown to significantly improve robustness in real deployments.
Individual studies further demonstrate the effectiveness of deep learning for underwater object detection.

(Mbani & Greinert, 2025) presents automated underwater image analysis for sediment pattern characterization and megafauna distribution, illustrating how CNN-based feature extraction can capture subtle seabed texture variations relevant to environmental assessment. (Zhao et al., 2022) proposes improvements to YOLO-based networks tailored for underwater environments, achieving faster and more accurate detection through architectural refinement and dataset-specific training strategies. An extensive performance comparison is presented based on the Brackish fish dataset (Pedersen et al., 2019). (Fayaz et al., 2024) introduces an integrated framework combining image restoration and object detection for autonomous underwater vehicles (AUVs), demonstrating that joint optimization of enhancement and detection improves performance under scattering and low-contrast conditions. (Chen & Er, 2024) proposes Dynamic YOLO, a tailored object detector for underwater imagery that enhances small object detection by combining a lightweight deformable-convolution backbone with a unified multi-scale attention-driven feature fusion framework and a decoupled detection head. (Chen et al., 2026) provides a comprehensive analysis of AI-driven underwater optical object detection, outlining current challenges such as limited annotated datasets, domain shift across water types, and reduced generalization in highly turbid conditions.

Currently, there are few published papers or publicly available datasets specifically targeting underwater UXO detection using optical cameras. Also in remote sensing, a recent survey stated that about the 20% of the methods are based on optical images (Samadzadegan et al., 2025). In contrast to the more mature body of work on acoustic sensing—particularly side-scan and synthetic aperture sonar—optical UXO research remains comparatively limited, largely due to the operational constraints imposed by turbidity and light attenuation.

In this context, (Dahn et al., 2024) introduced a multimodal dataset combining acoustic and optical data for underwater UXO perception. The dataset includes synchronized forward-looking sonar imagery, optical images, calibration parameters, and ground-truth pose information. This multimodal design enables cross-domain learning, sensor fusion strategies, and comparative evaluation between acoustic and optical detection pipelines. From a research perspective, it represents a bridge between traditionally sonar-dominated UXO detection workflows and emerging vision-based approaches.

Complementing this effort, (Tavaris et al., 2025) provides an annotated optical dataset focused on small underwater object detection, including UXO-relevant targets. A web crawling approach has been followed in order to further enrich the dataset with images of UXO objects downloaded from the Web (Foresti & Scagnetto, 2022). Unlike acoustic datasets that emphasize sonar intensity patterns and shadow geometry, the optical dataset captures high-resolution appearance features such as color, texture, surface corrosion, and shape irregularities. These characteristics are particularly relevant during the identification and classification stage following initial detection. The dataset therefore supports the development of AI-based optical classifiers capable of discriminating between true UXO objects and benign debris under realistic seabed conditions.

The characteristics emphasized by acoustic datasets, like shadow geometry just mentioned, have been exploited together with appearance information in a multi-modal fashion in (Abu & Diamant, 2023), where both optical images and SAS have been used jointly. The optical images are converted in SAS images with a Shape-From-Shading algorithm, and then integrated with the SAS images to enrich the input information. On these features, a multi-label classifier based on quadratic discriminant analysis (QDA) have been builts. The method has been compared on a small dataset provided by the authors (86 SAS images and 82 optical images) against other three different SOTA methods obtaining good performance w.r.t the tested ones.

Optical camera data made possible also to work on image-based 3D reconstruction using Structure-from-Motion (SfM) techniques. Some approaches to use SfM of the seafloor to classify military munitions underwater are provided in (Wilbur & Jaffee, 2021).

The production of large UXO dataset based on optical images is challenging also because sometime the annotation of the objects of interest is difficult on the RGB images of the seafloor. In (Bryan et al., 2022), the authors investigate the usage of optical images vs SAS images as ground truth data. The conclusion is that buried objects are more easily identifiable thanks to the SAS images, and that it is better to use optical images for high confident data (i.e. non-buried objects). Grouping different classes in “simpler” classes is also useful to improve the performance of the recognition.

One of the largest datasets created for image-based UXO classification, is the UXORD-10k dataset (Begkas et al., 2022), with over 10000 annotated images, including eight categories of objects. This data has the capacity to serve as a benchmark dataset for UXO image recognition. The baseline of this benchmark dataset has been achieved by testing CNN and VIT-based networks, obtaining an accuracy of 86.57% with the best solution even if it contains highly overlapping UXO categories.

Compared to FLS-based UXO datasets such as (Dahn et al., 2024) and (Ściegienka & Blachnik, 2024), optical datasets remain smaller in scale and more environment-dependent, but they provide complementary semantic richness that acoustic imagery lacks. While sonar excels in wide-area search and low-visibility conditions, optical cameras can offer superior fine-grained object characterization when environmental conditions permit.

Overall, although UXO-specific optical datasets are still scarce, the methodological advances developed for general underwater optical detection—transfer learning from large-scale terrestrial datasets, multi-scale detection architectures, enhancement-aware training, and domain adaptation—constitute a solid technical foundation. As multimodal datasets expand and sensor fusion techniques mature, optical AI-based detection is expected to increasingly complement acoustic systems within integrated UXO detection and identification frameworks.

Table 3: Summarization of optical camera approaches with used data.

Method	Data	Metrics
CNN-based detection (Mbani & Greinert, 2025)	Seafloor images from tropical North Atlantic. (Non-UXO)	No explicit metrics reported.
YOLO-UOD which is based on Yolov4 tiny. (Zhao et al., 2022)	Brackish underwater dataset. (Non-UXO)	mAP 87.88%
Integrated enhancement + detection (Fayaz et al., 2024)	Underwater Robot Picking contest dataset.	mAP 94.35% Recall 90.6% UCIQUE 3.09
Dynamic YOLO for underwater (Chen & Er, 2024)	DUO dataset	AP 68.6%
Public UXO dataset (Chen et al., 2026)	Collection UXO dataset	Not applicable.
Integrated optical and SAS images (Abu & Diamant, 2023)	Collected by the authors and made publicly available	Confusion matrix is provided, the overall accuracy is 0.88
Geometric features extracted from image-based reconstruction (Wilbur & Jaffee, 2021)	Case study collected by the authors – Mascoma Lake, New Hampshire (US)	Qualitative analysis. No aggregated metrics reported.
ResNet, EfficientNet and ViT-based network (Begkas et al., 2022)	UXORD-10k dataset (10000 images, 8 classes)	ResNet 78.92% of accuracy; EfficientNet 81.53% of accuracy; ViT-based 86.57%

Conclusions

The report compared a wide spectrum of approaches, ranging from traditional signal processing pipelines to different deep learning architectures. Regarding the hydroacoustic object detection methods, the analysis shows that no single method dominates across all conditions; instead, each approach presents specific advantages and limitations depending on the operational constraints, data availability, seabed characteristics, and computational budget. Overall, the comparison indicates that modern one-stage detectors from YOLO family currently provide the best balance between accuracy, speed, and operational feasibility, especially when enhanced through transfer learning or synthetic data generation. Traditional methods still offer competitive performance with minimal computational cost, making them practical for low-power embedded systems. In contrast, two-stage detectors remain the preferred choice for high-precision, detail-sensitive analysis, if sufficient computational resources are available.

Underwater optical imagery complements hydroacoustic detection by providing high-resolution visual and textural information that is particularly valuable for confirming potential detections and refining classification. Modern deep learning approaches, such as CNNs, Faster R-CNN, and YOLO, have substantially improved automated optical detection, particularly when combined with domain-specific preprocessing including color correction, contrast enhancement, and physics-aware image restoration. While UXO-specific optical datasets remain smaller, only one large dataset have been released since now, and more environment-dependent than sonar datasets, they capture fine-grained features such as surface texture, corrosion, and shape irregularities, which aid in discriminating true UXO objects from benign seabed clutter. Integrated pipelines that jointly optimize image enhancement and detection further increase robustness, enabling reliable performance under varying turbidity and lighting conditions.

The evidence suggests that an integrated workflow leveraging both modalities is most effective. Sonar is ideal for initial wide-area surveys and rapid detection of potential objects, even under low-visibility conditions, while optical cameras can then confirm detections, reduce false positives, and provide detailed characterization for classification. Multimodal datasets and cross-domain learning support this complementary approach, enhancing detection accuracy, generalization, and operational robustness. Together, these strategies form the foundation for scalable, AI-driven UXO detection systems that balance broad coverage, efficiency, and precise object verification.

References

Abu, A., & Diamant, R. (2023). Underwater object classification combining SAS and transferred optical-to-SAS imagery. *Pattern Recognition, 144*, 109868. https://doi.org/10.1016/j.patcog.2023.109868

Agrawal, A., Sikdar, A., Makam, R., Sundaram, S., Besai, S. K., & Gopi, M. (2025). Syn2Real domain generalization for underwater mine-like object detection using side-scan sonar. *IEEE Geoscience and Remote Sensing Letters, 22*, 1–5. https://doi.org/10.1109/LGRS.2025.3550037

Armstrong, R. A., Pizarro, O., & Roman, C. (2019). Underwater robotic technology for imaging mesophotic coral ecosystems. In Y. Loya, K. Puglise, & T. Bridge (Eds.), *Mesophotic coral ecosystems* (Vol. 12). Springer. https://doi.org/10.1007/978-3-319-92735-0_51

Aubard, M., Madureira, A., Teixeira, L., & Pinto, J. (2025). Sonar-based deep learning in underwater robotics: Overview, robustness, and challenges. *IEEE Journal of Oceanic Engineering, 50*(3), 1866–1884. https://doi.org/10.1109/JOE.2025.3531933

Begkas, G., Giannakeris, P., Ioannidis, K., Kalpakis, G., Tsikrika, T., Vrochidis, S., & Kompatsiaris, I. (2022). Automatic visual recognition of unexploded ordnances using supervised deep learning. In *Proceedings of the 2022 International Conference on Multimedia Retrieval* (pp. 286–294). https://doi.org/10.1145/3512527.3531383

Blueprint Subsea. (2026). *Oculus sonar*. https://blueprintsubsea.com/oculus

Bobkov, V., Shupikova, A., & Inzartsev, A. (2023). Recognition and tracking of an underwater pipeline from stereo images during AUV-based inspection. *Journal of Marine Science and Engineering, 11*, 2002. https://doi.org/10.3390/jmse11102002

Bryan, O., Hansen, R. E., Haines, T. S. F., Warakagoda, N., & Hunter, A. (2022). Challenges of labelling unknown seabed munition dumpsites from acoustic and optical surveys. *Remote Sensing, 14*(11), 2619. https://doi.org/10.3390/rs14112619

Caricchio, C., Mendonça, L. F., Lima, A. T. C., & Lentini, C. A. D. (2025). Operational assessment of side-scan sonar data applied to naval mine detection. *IEEE Geoscience and Remote Sensing Letters, 22*, 1–5. https://doi.org/10.1109/LGRS.2025.3596852

Chen, J., & Er, M. J. (2024). Dynamic YOLO for small underwater object detection. *Artificial Intelligence Review, 57*, 165. https://doi.org/10.1007/s10462-024-10788-1

Chen, L., Huang, Y., Dong, J., Xu, Q., Kwong, S., Lu, H., Lu, H., & Li, C. (2026). Underwater optical object detection in the era of artificial intelligence: Current, challenge, and future. *ACM Computing Surveys, 58*(3), Article 62. https://doi.org/10.1145/3759243

Chen, Z., Xie, G., Deng, X., Peng, J., & Qiu, H. (2024). DA-YOLOv7: A deep learning-driven high-performance underwater sonar image target recognition model. *Journal of Marine Science and Engineering, 12*, 1606. https://doi.org/10.3390/jmse12091606

Dahn, N., et al. (2024). An acoustic and optical dataset for underwater unexploded ordnance (UXO). In *OCEANS 2024* (pp. 1–6). https://doi.org/10.1109/OCEANS55160.2024.10754316

Dakhil, R., & Khayeat, A. (2022). Review on deep learning techniques for underwater object detection. *Computer Science & Information Technology*, 49–63. https://doi.org/10.5121/csit.2022.121505

Dobeck, G. J. (2001). Algorithm fusion for automated sea mine detection and classification. In *OCEANS 2001* (pp. 130–134). https://doi.org/10.1109/OCEANS.2001.968690

Du, L., Zhang, R., & Wang, X. (2020). Overview of two-stage object detection algorithms. Journal of Physics: Conference Series, 1544, 012033. https://doi.org/10.1088/1742-6596/1544/1/012033

Einsidler, D., Dhanak, M., & Beaujean, P.-P. (2018). A deep learning approach to target recognition in side-scan sonar imagery. In *OCEANS 2018* (pp. 1–4). https://doi.org/10.1109/OCEANS.2018.8604879

Fayaz, S., Parah, S. A., Qureshi, G. J., Lloret, J., Ser, J. D., & Muhammad, K. (2024). Intelligent underwater object detection and image restoration. *IEEE Transactions on Vehicular Technology, 73*(2), 1726–1735. https://doi.org/10.1109/TVT.2023.3318629

Fayaz, S., Parah, S. A., & Qureshi, G. J. (2022). Underwater object detection: Architectures and algorithms. *Multimedia Tools and Applications, 81*, 20871–20916. https://doi.org/10.1007/s11042-022-12502-1

Foresti, G. L., & Scagnetto, I. (2022). An integrated low-cost system for object detection in underwater environments. *Integrated Computer-Aided Engineering, 29*(2), 123–139. https://doi.org/10.3233/ICA-220675

Franchi, M., Ridolfi, A., & Allotta, B. (2021). Underwater navigation with 2D forward looking sonar. *Journal of Field Robotics, 38*, 355–385. https://doi.org/10.1002/rob.21991

Hurtós, N., Palomeras, N., Carrera, A., & Carreras, M. (2017). Autonomous detection and mapping of an underwater chain using sonar. *Ocean Engineering, 130*, 336–350. https://doi.org/10.1016/j.oceaneng.2016.11.072

Jian, M., Yang, N., Tao, C., et al. (2024). Underwater object detection and datasets: A survey. *Intelligent Marine Technology Systems, 2*, 9. https://doi.org/10.1007/s44295-024-00023-6

Jiang, L., Cai, T., Ma, Q., Xu, F., & Wang, S. (2020). Active object detection in sonar images. *IEEE Access, 8*, 102540–102553. https://doi.org/10.1109/ACCESS.2020.2999341

Kongsberg Discovery. (2026). *Flexview multibeam sonar*. https://www.kongsberg.com/discovery/seafloor-mapping/sonars/flexview/

Li, J., Cao, X. (2022). Target recognition in side-scan sonar images based on YOLOv3. In *Chinese Control Conference* (pp. 7186–7190). https://doi.org/10.23919/CCC55666.2022.9902742

Li, J., Kaess, M., Eustice, R. M., & Johnson-Roberson, M. (2018). Pose-graph SLAM using forward-looking sonar. *IEEE Robotics and Automation Letters, 3*(3), 2330–2337. https://doi.org/10.1109/LRA.2018.2809510

Long, H., Shen, L., Wang, Z., & Chen, J. (2023). Underwater sonar image detection via speckle reduction. *IEEE Transactions on Geoscience and Remote Sensing, 61*. https://doi.org/10.1109/TGRS.2023.3248605

Mbani, B., & Greinert, J. (2025). Automated underwater image analysis reveals sediment patterns. *Scientific Reports, 15*, 27481. https://doi.org/10.1038/s41598-025-12723-y

Mu, X., Yue, G., Zhou, N., & Chen, C. (2022). Occupancy grid-based AUV SLAM method. *Journal of Marine Science and Engineering, 10*, 1056. https://doi.org/10.3390/jmse10081056

Nalmpanti, M., Chrysafi, A., Meeuwig, J. J., et al. (2023). Monitoring marine fishes using underwater video. *Reviews in Fish Biology and Fisheries, 33*, 1291–1310. https://doi.org/10.1007/s11160-023-09799-y

Parsons, M. J. G., Fenny, E., & Lucke, K. (2017). Imaging marine fauna with sonar. *Acoustics Australia, 45*, 41–49. https://doi.org/10.1007/s40857-016-0076-1

Pedersen, M., Haurum, J., Gade, R., Moeslund, T., & Madsen, N. (2019). Detection of marine animals in underwater datasets.

Samadzadegan, F., Qaderi, F., Haghighi, T. L., et al. (2025). UAV remote sensing for unexploded ordnance detection. *Sensors Imaging, 26*, 137. https://doi.org/10.1007/s11220-025-00668-5

Ściegienka, P., & Blachnik, M. (2024). Acoustic dataset for unexploded ordnance classification. *Sensors, 24*(18), 5946. https://doi.org/10.3390/s24185946

Tavaris, D., Scandino, L., Foresti, G. L., & Scagnetto, I. (2025). Cost-effective underwater object detection system. *Integrated Computer-Aided Engineering, 32*(3), 258–271. https://doi.org/10.1177/10692509251336668

Terven, J., et al. (2023). Review of YOLO architectures. *Machine Learning and Knowledge Extraction, 5*(4), 1680–1716. https://doi.org/10.3390/make5040083

Tritech. (2026). *Gemini sonar products*. https://www.tritech.co.uk/products/gemini

Tulsook, S., et al. (2018). Pipeline extraction using sonar images. In *ECTI-CON* (pp. 584–587). https://doi.org/10.1109/ECTICon.2018.8619973

Ultralytics. (2026a). *One-stage object detectors*. https://www.ultralytics.com/glossary/one-stage-object-detectors

Ultralytics. (2026b). *Ultralytics GitHub repository*. https://github.com/ultralytics/ultralytics

Wang, N., Wang, Y., & Er, M. J. (2020). Review on deep learning for marine object recognition. *Control Engineering Practice, 118*, 104458. https://doi.org/10.1016/j.conengprac.2020.104458

Wang, Y., Liu, J., Yu, S., Wang, K., Han, Z., & Tang, Y. (2021). Underwater object detection based on YOLOv3. In *ICUS* (pp. 571–575). https://doi.org/10.1109/ICUS52573.2021.9641489

Wang, Z., Guo, J., Zhang, S., & Xu, N. (2025). Marine object detection in sonar images. *Frontiers in Marine Science, 12*. https://doi.org/10.3389/fmars.2025.1539210

Wen, X., & Zhang, F. (2023). Underwater target detection based on YOLOv7. In *ACAIT* (pp. 1536–1542). https://doi.org/10.1109/ACAIT60137.2023.10528458

Wilbur, J., & Jaffee, J. (2021). Optical detection of military munitions underwater. https://apps.dtic.mil/sti/html/trecms/AD1217113/

Williams, D. P. (2015). Fast target detection in sonar imagery. *IEEE Journal of Oceanic Engineering, 40*(1), 71–92. https://doi.org/10.1109/JOE.2013.2294532

Xie, K., Yang, J., & Qiu, K. (2022). Dataset with multibeam forward-looking sonar. *Scientific Data, 9*, 739. https://doi.org/10.1038/s41597-022-01854-w

Yang, H., Zhou, T., Jiang, H., Yu, X., & Xu, S. (2024). Lightweight underwater detection network. *IEEE Transactions on Instrumentation and Measurement, 73*. https://doi.org/10.1109/TIM.2024.3425490

Yang, Z., Zhao, J., Yu, Y., & Huang, C. (2024). Sample augmentation for sonar images. *IEEE Transactions on Geoscience and Remote Sensing, 62*. https://doi.org/10.1109/TGRS.2024.3371051

Yulin, T., Jin, S., Bian, G., & Zhang, Y. (2020). Shipwreck detection using YOLOv3. *IEEE Access, 8*, 173450–173460. https://doi.org/10.1109/ACCESS.2020.3024813

Zacchini, L., et al. (2020). CNN-based sonar target recognition. In *AUV Symposium* (pp. 1–6). https://doi.org/10.1109/AUV50043.2020.9267902

Zhang, Y., et al. (2022). Submarine pipeline tracking using sonar. *Applied Ocean Research, 122*, 103128. https://doi.org/10.1016/j.apor.2022.103128

Zhao, S., Zheng, J., Sun, S., & Zhang, L. (2022). Improved YOLO for underwater detection. *Symmetry, 14*, 1669. https://doi.org/10.3390/sym14081669

MMinE-SwEEPER

www.mminesweeper-munition.eu