IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 December 2023, accepted 8 January 2024, date of publication 12 January 2024, date of current version 26 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3353688

SURVEY

Underwater Target Detection Using Deep Learning: Methodologies, Challenges, Applications, and Future Evolution

ANWAR KHAN[1], MOSTAFA M. FOUDA[2], (Senior Member, IEEE), DINH-THUAN DO[3], (Senior Member, IEEE), ABDULAZIZ ALMALEH[4], ABDULLAH M. ALQAHTANI[5], AND ATIQ UR RAHMAN[6]

[1]Department of Electronics, University of Peshawar, Peshawar 25120, Pakistan
[2]Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID 83209, USA
[3]School of Engineering, University of Mount Union, Alliance, OH 44601, USA
[4]College of Computer Science, Information Systems Department, King Khalid University, Abha 61421, Saudi Arabia
[5]School of Computer Science and Information Technology, Jazan University, Jazan 45142, Saudi Arabia
[6]Faculty of Computer Information Science, Higher Colleges of Technology, Ras Al Khaimah Campus, Abu Dhabi, United Arab Emirates

Corresponding author: Anwar Khan (arkhan@uop.edu.pk)

ABSTRACT This paper provides a study of the latest target (object) detection algorithms for underwater wireless sensor networks (UWSNs). To ensure selection of the latest and state-of-the-art algorithms, only algorithms developed in the last seven years are taken into account that are not entirely addressed by the existing surveys. These algorithms are classified based on their architecture and methodologies of operation and their applications are described that are helpful in their selection in a diverse set of applications. The merits and demerits of the algorithms are also addressed that are helpful to improve their performance in future investigation. Moreover, a comparative analysis of the described algorithms is also given that further provides an insight to their selection in various applications and future enhancement. A depiction of the addressed algorithms in various applications based on publication count over the latest decade (2023-2013) is also given using the IEEE database that is helpful in knowing their future application trend. Finally, the challenges associated with the underwater target detection are highlighted and the future research paradigms are identified. The conducted study is helpful in providing a thorough analysis of the underwater target detection algorithms, their feasibility in various applications with future challenges and defined strategies for further investigation.

INDEX TERMS Underwater target detection, deep learning, underwater object detection, YOLO, convolutional neural networks, ConVNNs.

I. INTRODUCTION

Underwater wireless sensor networks (UWSNs) is one of the latest realms of research that aims to explore the underwater environment for a number of applications. These applications include Tsunami detection and prediction [1], military surveillance [2], underwater navigation [3], secure communications [4], oil detection [5], fault monitoring in underwater cables [6], water quality monitoring [7], detection of crashed ships and boats [8] and target/object detection [9], to mention a few. Underwater target/object detection processes signals (image, audio, video, acoustic vibrations, radio/optical radiations) and extracts the information content that provides an ultimate insight to the underwater environment. This information is useful in a number of applications such as object detection and tracking by underwater robots [10], [11], objects detections by radars for military and civilian purposes [12], surveillance systems [13], [14], [15], [16], ship tracking [17], [18], mine detection [19], [20], [21], water quality and environmental

The associate editor coordinating the review of this manuscript and approving it for publication was Xuebo Zhang.

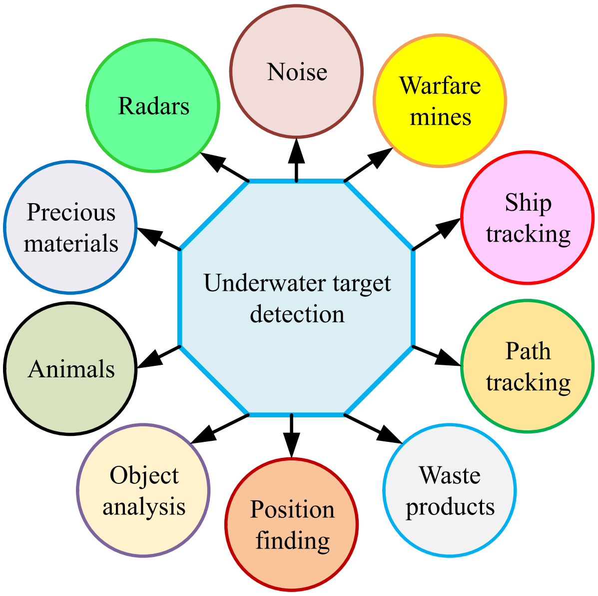

FIGURE 1. **Applications of underwater object detection in various fields.**

impacts [22], [23], [24], and toxins in water [25], [26], to mention a few. Figure 1 summarizes the various applications of underwater object detection ranging from noise detection in underwater channel to waste products, precious materials, mines, marine animals, radar signals, ships and their paths and position finding, to mention a few. There are a number of challenges inherently carried by the underwater environment [27]. For instance, the electric and magnetic fields of radio waves interact with water and other particles and transfer their energy to them. Due to this, radio energy is lost and, consequently, radio waves are not generally preferred for underwater communications instead acoustic waves are used. However, the speed of acoustic waves is almost five times slower than the speed of radio waves in water due to which underwater communications bear long delay. In addition, the acoustic spectrum is limited (almost to 100 kHz) that further narrows the available bandwidth for underwater communications. Consequently, underwater communications have low data rates. Besides, underwater nodes have a limited battery lifetime. This restricts the life span of UWSNs and demands for smart, efficient, reliable and optimized operation and working strategies.

The use of deep learning algorithms for underwater object detection has recently captured the attention of researchers due to a number of advantages over the traditional detection techniques in terms of accuracy of prediction, speed, generalization and automatic processing of tasks [28], [29]. Figure 2 shows the the basic concept of object detection by deep learning and traditional algorithms. Both types of algorithms perform signal operations on the input signals (such as de-noising, filtering, image enhancement, to mention a few). However, the operations of features extractions, features selection and classification of signals/objects are performed in an automated fashion by the deep learning algorithms.

There exists a number of surveys in literature related to underwater target detection and identification [30], [31], [32]. However, the addressed algorithms in these surveys are not state-of-the-art to address the requirements of the latest applications. Besides, they lack a thorough and in-depth insight and comparative analysis of the addressed algorithms. Keeping in view the importance and scope of the underwater target detection, challenges, potential applications and its direct and in-direct effects on the planet earth, this paper studies the latest underwater target detection algorithms designed in the last seven years so that only the most recent algorithms are taken into account. The algorithms are classified into various categories based on their architecture. These categories include algorithms based on you only look once (YOLO) architecture, convolutional neural networks (ConVNNs) and their various types for varied applications and hybrid algorithms that combine various techniques to construct a single target detection architecture. Such a classification makes algorithms selection convenient for the appropriate applications. The algorithms are also described in terms of their operational strategies, merits and demerits that helps in knowing their working mechanisms and improving their demerits in future enhancement. In addition, the operation of the architecture of each class of algorithm is described with suitable figures that helps in not only understanding of these algorithms but provides the areas where the performance of these algorithms can excel.

A comparative analysis of the classified algorithms is also provided that helps in selection of the specific class of algorithms for various applications as well as provides a path for their future improvement. For instance, a bar chart is provided that shows the calculated accuracy in target detection of the compared algorithms. This helps not only in the use of these algorithms in various applications depending upon the accuracy requirement but also provides clues for the further investigation towards enhanced accuracy. The described mathematical models of the algorithms further provide an insight to their work, operation and object detection strategies. Moreover, the use of the classified algorithms in diverse object/target detection applications in the latest decade (2023-2013) is graphically depicted that provides an idea of the latest trends of these classes of algorithms in terms of applications to real world problems. Finally, the challenges associated with the underwater target detection are revealed and future research directions are specified.

In summary, this paper provides a study of the underwater target detection algorithms developed in the last seven years by addressing their architectures, operational strategies, merits, demerits, comparative analysis, target detection accuracy and applications in various fields in the latest decade (2023-2013). Moreover, the challenges with underwater

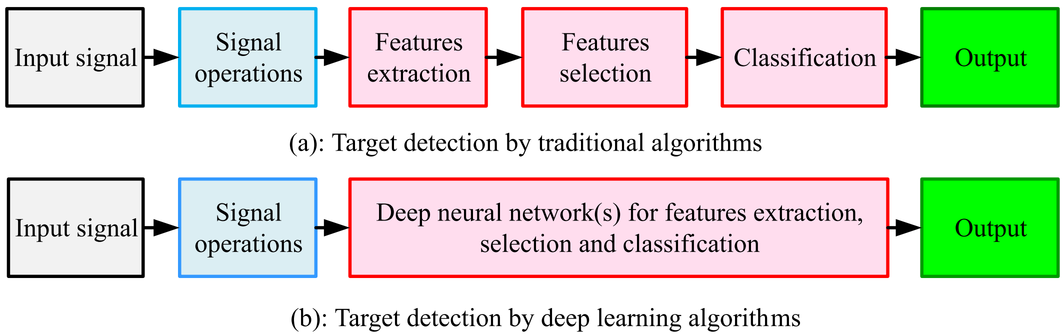

FIGURE 2. **Object detection by (a): traditional algorithms and (b): deep learning algorithms. The latter perform features extraction, selection and classification processes automatically by the deep networks rather than by varied techniques in the former.**

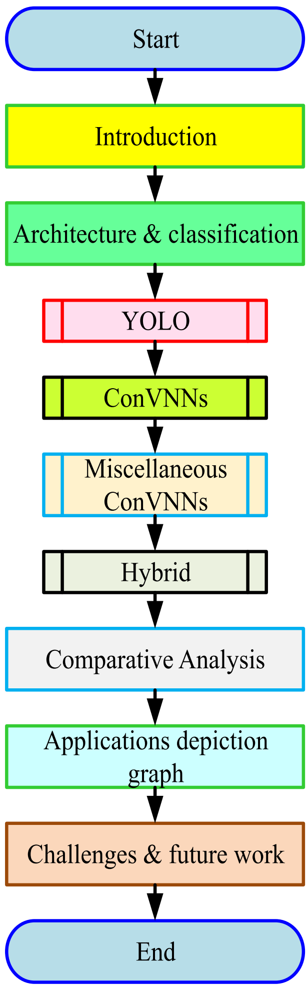

FIGURE 3. **Organization and flow of the manuscript.**

target detection are highlighted and the possible solution strategies are described for future investigation.

Figure 3 shows the organization and flow of the manuscript. The introduction discusses the importance of underwater object detection, its applications and the contributions of this paper. Sections II deals with the description of YOLO, its basic architecture and various versions and their comparison as well as object detection algorithms based on YOLO. Section III discusses algorithms using ConVNNs for underwater target detection, their architecture, operation and sub-classification into various applications. Section IV focuses on discussion of hybrid algorithms for underwater target detection. A comparative analysis of the classified algorithms is performed in Section V. Finally, Section VI discusses the challenges in underwater target detection while Section VII concludes the paper with directions for future investigation.

II. UNDERWATER TARGET DETECTION USING YOLO

A. THE BASIC ARCHITECTURE OF YOLO SERIES

YOLO considers real time object detection and was first designed by [33]. It divides an image into an $S \times S$ grid with bounding box regression applied on each cell of the grid with a confidence score C that measures the probability $P(Obj)$ of existence of an object in each box and is defined as:

$$C = P(Obj) \times IoU^{truth}_{predicted}, \tag{1}$$

where IoU is the intersection over union operation having values between 0 and 1 with the latter being the ideal value. Union represents the total area of the predicted bounding box and the ground truth while the intersection signifies the overlapping area of the predicted bounding box and the ground truth. The conditional class probability of each cell of the grid given that it has an object is denoted by $P(Class_i \mid Obj)$ and is mathematically defined as:

$$\begin{aligned} P(Class_i \mid Obj) \times P(Obj) \times IoU^{truth}_{predicted} \\ = P(Class_i) \times IoU^{truth}_{predicted}. \end{aligned} \tag{2}$$

To ensure accurate object detection, the bounding box and center of each prediction is corrected by a loss function given by:

$$Loss = \lambda_{Coord} \sum_{i=0}^{s^2} \sum_{j=0}^{A} \mathbb{1}_{ij}^{Obj}[(b_{x_i} - b_{\hat{x}_i})]^2 + [(b_{y_i} - b_{\hat{y}_i})]^2$$

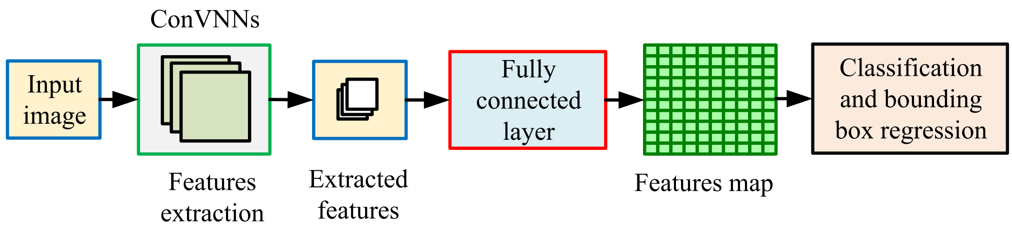

FIGURE 4. Basic architecture of the YOLO series.

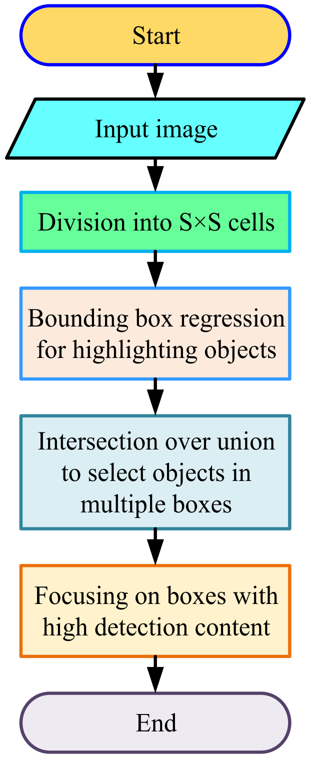

FIGURE 5. Flow chart of the basic YOLO architecture. An input image is divided into a grid of $S \times S$ cells with each cell processed by bounding box regression to detect object. It is followed by the intersection over union operation for selection of objects in multiple boxes and boxes with high detection threshold are identified.

$$
\begin{aligned}
&+\lambda_{Coord}\sum_{i=0}^{s^2}\sum_{j=0}^{A}\mathbb{1}_{ij}^{Obj}[(b_{w_i}-b_{\hat{w}_i})]^2+[(b_{h_i}-b_{\hat{h}_i})]^2\\
&+\sum_{i=0}^{s^2}\sum_{j=0}^{A}\mathbb{1}_{ij}^{Obj}(C_i-\hat{C}_i)^2\\
&+\lambda_{Noobj}\sum_{i=0}^{s^2}\sum_{j=0}^{A}\mathbb{1}_{ij}^{Noobj}(C_i-\hat{C}_i)^2\\
&+\sum_{i=0}^{s^2}\mathbb{1}_{i}^{Obj}\sum_{c\in Classes}(p_i(c)-\hat{p}_i(c))^2, \qquad (3)
\end{aligned}
$$

where A is the number of boxes assigned to each grid, b_x and b_y are the center co-ordinates of each prediction, b_w and b_h represent the dimensions of the bounding box, the parameters λ_{Coord} and λ_{Noobj} are used to give more weight to boxes with objects and less weight to boxes with no objects and $p(c)$ represents classification prediction. The parameter $\mathbb{1}_{ij}^{Obj}$ has a unity value in case the j^{th} bounding box in the i^{th} cell predicts an object, otherwise its value is zero. Similarly, if the predicted object is in the i^{th} cell then $\mathbb{1}_{i}^{Obj}$ is unity else it is zero. Figure 4 shows the basic architecture of the YOLO series. The input images are processed by the ConvNNs for features extraction. The fully connected layer reduces the dimensions of the extracted features and obtains a features map and classifies it. The bounding box regression assigns attributes to the classified features. Figure 5 shows flow chart of the basic YOLO architecture in which an input image is divided into an $S \times S$ grid with bounding box regression to identify the objects. The intersection over union operation is then applied to find objects in multiple boxes followed by identifying boxes with high detection content.

In the lines to follow, the algorithms using the YOLO series for objects detection are considered. Table 1 shows a comparison among the various versions of the YOLO architecture with enhancements and added features as the architecture evolved from the the initial to the latest version. It shows that the evolution of YOLO started with real-time object detection and with the development and progression, it continued to include more features. Each version puts emphasis on the overall better, more efficient, reliable and accurate detection of objects that the earlier version(s).

B. OBJECT DETECTION USING YOLOV8 AND V7

A comparison of the various YOLO algorithms is made in [34] for synthetic and real world data and concludes that YOLOv5 exhibits the best results on synthetic data while YOLOv8 outperforms all the other versions on real datasets. The authors in [35] conclude that YOLOv7 is the best when compared with faster region-based ConVNNs (R-ConVNNs), single-shot detector (SSD) and Centernet for land and aquatic small object detection. The authors in [36] compare YOLOv7 with YOLOv5 series and find it better in terms of object detection accuracy and performance in image challenging conditions. A loss function is proposed in [37] for YOLOv7 based on the concept of a bag of features to optimize the error and enhance the accuracy and speed of marine object detection and classification. The concept of supervised features learning is introduced in [38] based on efficient

TABLE 1. A comparison of the various YOLO versions.

YOLO Version	Year of Development	Architecture	Uniqueness than predecessor(s)
YOLOv1	2015	Consisted of a single ConVNN	Fast response than some existing models, less accurate that the existing two-stage detection algorithms
YOLOv2	2016	Improved YOLOv1	Introduced anchor boxes to previous version for enhanced detection accuracy and an up-sampler layer for improving the resolution of the features map
YOLOv3	2018	Improved accuracy and speed of YOLOv2	Involved Darknet-53 architecture, a modification of ResNet architecture specialized for object detection, improved accuracy and stability of detection due to the use of feature pyramid networks, specialized loss functions and varied range of aspect ratio and object size
YOLOv4	2020	Improved YOLOv3	Introduced a new backbone to YOLOv3 with improved training and model capacity and cross mini-batch normalization for improved stability
YOLOv5	2020	Improved YOLOv4	Used an efficient network architecture with enhanced and improved features
YOLOv6	2022	Improved YOLOv5	Introduced a novel ConVNN architecture known as spatial pyramid pooling network (SSP-Net) with varied size and aspect ratios for effective object detection
YOLOv7	2022	Improved YOLOv6	Introduced a novel ConVNN known as ResNeXt trainable with images at various scales and their combination for effective detection. To counteract the class imbalance problem in object detection, focal loss technique was used
YOLOv8	2023	The most improved YOLO	It has better prediction accuracy and capable of segmentation and classification of images along with other features

aggregation network to improve the scaling calculation on objects.

C. OBJECT DETECTION USING YOLOv5

The images are first processed by the deep wave net scheme in [40] that uses ConVNNs for enhancement and then feeds the output to YOLOv5 for object detection. The authors in [41] first enhance the images using Gridmask method followed by adding intersection over union to the non-maximum suppression method to improve detection accuracy when the detected objects overlap. The features pyramid network (FPN) in the main architecture is modified to detect small objects. Drones are used in [42] to utilize YOLOv5s to detect submerged objects in water with an effective accuracy and precision. The authors in [43] analyse the vocal behaviour of mammals by processing their signals using YOLOv5. The obtained information, such as central frequency of the signals, duration and bandwidth are effective in knowing the behaviour of these mammals. The network parameters of YOLOv5 are first reduced by using GhostBottleneck in [44] and then important features weight is increased by the addition of a convolutional block attention layer to the final layer of the backbone architecture and the intersection over union feature is modified to enhance the accuracy of the overlapped objects. The methodology in [45] uses the coordinate attention mechanism and a bidirectional feature pyramid network to improve the target detection accuracy of YOLOv5 for ship target detection. A comparison of YOLOv5 is made with YOLOv3 for seaweeds detection in [46] and it is found that the former is faster than the later but with a reduced detection accuracy of 3-5 %. The authors in [47] make three changes to the YOLOv5s to improve its performance. First, they use a multi-head self-attention technique having contextual information that replaces the convolutional block for better features extraction. Second, a hybrid convolutional module is added for reduction in parameters number. Third, a path aggregation network is used to collect features from the shallow and deep layers. Sonar images are first pre-processed in [48] to overcome the internal and external noise and then the improved YOLO5 is used for enhanced accuracy of object detection, especially in overlapped objects.

The concept in [49] embeds a camera with an autonomous underwater vehicle (AUV) that captures underwater images of the target in a swimming pool, which are then detected by the YOLOv5 and the information is extracted. The training of the deep learning module is performed by the images processed on the Google-Collab over the cloud and the output is then processed by the on-board computer of the AUV that consists of Raspberry-Pi4 having a coral USB accelerator. In [50], the realization of detection and tracking of the target is performed by the deep sort algorithm. Due to poor lightening conditions, the obtained images are also enhanced leading to an overall 96% detection accuracy. The concept in [51] optimizes the performance of the original YOLOv5 by optimizing its main architecture (CSPDarknet). The features are selected and extracted using the cross stage partial (CSPNet), a convolutional layer that utilizes the contextual block streaming (CBS) as its fundamental architecture for the recognition of useful information. The elementary layers are changed by ConVNNs followed by swin transformer.

TABLE 2. Target Detection using YOLOv8 [34], YOLOv7 [35], [36], [37], [38] and YOLOv5 [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60]. The symbol x represents the unspecified value.

Reference	Key Concept	Achievement	Limitations/Cost	Detection Rate (%)	Year
[34]	Compares various YOLO algorithms on synthetic and real datasets and finds YOLOv5 as the best model for object detection for synthetic data and YOLOv8 for real data	Provides a comparative analysis of various YOLO algorithms for real and synthetic datasets	Requires further validation for diverse datasets	x	2023
[35]	Compares YOLOv7 with object detectors working in one and two stages and finds it the best	Effective for small object detection on land and in marine environment	Struggles with detection of small objects below a certain size	x	2023
[36]	Compares YOLOv7 with YOLOV5 series and finds it better in object detection accuracy	Reliable and accurate detection in image challenging conditions such as occlusion, image blurring and color distortion	Involves optical communications that are sensitive to dispersion	x	2023
[37]	Models a loss function to optimize error and enhance accuracy and speed of detection	Speed and accuracy of detection	Error optimization and features bag require constant computation	x	2023
[38]	Uses an efficient aggregation network to improve the scaling calculation on objects	Time-efficient training and detection	Struggles in detection of overlapped objects	x	2023
[39]	First obtains the image of an underwater resource then locates position and distance of the resource from object in the image using a depth chart and a weighted average model	Useful in identification of underwater resources	Requires constant computation of the positions of the camera to find the relative distance of the resource and its position	x	2023
[40]	First processes the images using Wavenet algorithm for enhancement then applies YOLOv5	Effective for distorted and faded images due to enhancement	High computation during training	x	2023
[41]	Enhances images using Gridmask method and adds intersection over union to the non-maximum suppression for accurate detection of overlapped objects and modifies the features pyramid network to detect small objects	Effective for overlapped and small objects detection	Complexity of modifying the basic architecture of YOLOv5s	x	2023
[42]	Uses drones that capture images of submerged objects by processing them using YOLOv5s to detect drowned/drowning objects	Effective for life rescue team	Drones need constant battery power to provide effective monitoring of the water body	89	2023
[43]	Performs an analysis of the vocal behaviour of mammals	The obtained vocal information is effective in obtaining knowledge about the behaviour of the mammals	Requires heavy datasets of mammals' vocal information	93.87	2023
[44]	Modifies the backbone and features weight of YOLOv5 using convolutional block attention and intersection over union to detect overlapped objects	Lightweight with enhanced accuracy of detection	Information in attention mechanism varies significantly with varied sea conditions	Enhanced by 1.53 of YOLOv5 with significant reduction in calculation	2023
[45]	Combines coordinate attention mechanism with bidirectional FPN to improve the accuracy of YOLOv5	Improved accuracy of 3.3% higher than YOLOv5	Involves extra computational steps	99.1	2023
[46]	Compares YOLOv5 with YOLOv3 for seaweeds detection	Investigates that YOLOv5 is faster than YOLOv3 but with lower accuracy	Requires a diverse comparison	x	2023
[47]	Modifies the YOLOv5s by a self-attention contextual layer for features extraction, hybrid convolution layer for parameters reduction and path aggregation network for features extraction from shallow and deep layers	Improved performance	High computational complexity	x	2023
[48]	Pre-processes sonar images to overcome internal and external noise and then uses improved YOLOv5 for enhanced accuracy of objects detection	Enhanced accuracy of object detection, especially overlapped objects	Requires intensive training on internal and external noise factors	x	2023
[49]	A camera embedded in an AUV takes images of target for processing by YOLOv5. The training phase of the algorithm is performed on the Google-Collab through cloud and the output is then processed by the on-board computer of the AUV consisting of Raspberry-Pi4 with a coral USB accelerator	Good for detection of target location	Low processing capability due to the limited battery power of the AUV	80.5	2022
[50]	Uses the YOLOv5 for moving fish as a target detection and the deep sort algorithm to realize the detection and tracking of the target followed by image enhancement	Better performance than YOLOv3 in real-time tracking	Additional steps of image enhancement	96	2022
[51]	Uses the ConVNNs and the swin transformer to replace the backbone of the YOLOv5 object detector	Improved detection accuracy	Replacing the backbone of YOLOv5 by the swin transformer diminishes the lightweight architecture of the former	2.2% better than YOLOv5	2022
[52]	Modifies the YOLOv5 in three ways: enhancing bottlenecks count from one to three in C3 module, replacing bottleneck module by a coordinated attention mechanism and introducing the SE module to the backbone of the model for enhanced attention	Improved accuracy of prediction	The light computing architecture of YOLOv5 is transformed to heavy computing	3.7% better than the baseline model	2022
[53]	Modifies the YOLOv5 for time and accuracy efficient target detection and finds the position of the target by the coordinates of the extracted features and the involved imaging system	Accurate and fast target detection	Underwater coordinates of the features and imagining system change regularly	54.2 % better than YOLO series	2023
[54]	First applies image enhancement and then processes images by a lightweight ghost module based on YOLOv5. The features extraction is performed by the selective kernel convolution and the positive and negative sample imbalance is overcome by the focal loss optimization	Accurate and quick target tracking	Involves multiple steps	86.12	2022
[55]	Applies various image enhancement and detection techniques and finds YOLOv5 as the best method	Advocates that image enhancement techniques make objection detection and differentiation convenient	Real-time and automatic operation of the compared algorithms requires a further insight	x	2022
[56]	Modifies the YOLOv5 for small target detection by re-clustering the initial frames, adding attention, layers, attention mechanism and connections to extract features and overcome model overfitting	Improved accuracy of detection	Deviates from the simple computation of YOLOv5	96.1	2022
[57]	Uses transfer learning on YOLOv5 to detect two similar objects in underwater images	Ease of computation due to transfer learning	Requires training the network prior to transferring learning	x	2023
[58]	Performs underwater plastic waste detection by modifying backbone, features extraction and loss function of theYOLOv5n	Efficient underwater plastic waste detection	Involves multiple steps and deviates from the light structure	12.25% better than YOLOv5n	2022
[59]	Performs a real-time comparison of YOLOv5, YOLOv5-TR and YOLOX with the latter exhibiting the best detection efficiency	Real-time comparison for target detection	Does not provide an insight for the selection of the chosen schemes for comparison	91.3	2022
[60]	Improves the YOLOv5 model by pre-training, clustering, fine-tuning the pre-training model and adding features extraction layer for high-level features	Detection speed and accuracy	Deviation from the light weight structure of YOLOv5 for computation	56.95	2022

The authors in [52] argue that underwater target tracking requires precision and accuracy and for this purpose the architecture of the YOLOv5 is modified in three ways. Firstly, the bottleneck count is increased from one to three.

Secondly, the bottleneck module is replaced by a module with coordinated attention so as to enhance the level of attention in the image of the target object. Thirdly, a module is introduced to the backbone of the model for further enhancing the attention level of the image and ignoring the unimportant features of the image. The approach adopted in [53] first detects and then locates the position of underwater objects. It modifies the YOLOv5 model for detection time and accuracy efficiency. First it downsamples the target image and extracts the features followed by the fusion to obtain a map of features. Then the image features coordinates and the imaging system coordinates determine the position of the target. The method given in [54] first applies the image enhancement on the objects followed by the YOLOv5 with ghost modules (instead of backbone features extraction) to detect objects with reduced parameters and computation. Moreover, the selective kernel convolution is applied for features extraction that has better results than the traditional convolution. Finally, the positive and negative sample imbalance is overcome by the focal loss optimization. The authors in [55] first apply the various image enhancement algorithms on various detection techniques and find that the YOLOv5 is the best in terms of objection detection and differentiation from background in underwater images. The small target objects are detected in side-scan sonar images in [56] by the modified YOLOv5. The re-clustering of the anchor frames of the target is performed in the first stage using K-means. It is followed by a new layer for capturing shallow features and then an attention mechanism for extracting deep features. Small samples overfitting is overcome by new connections. The authors in [57] use the YOLOv5 model to detect two similar objects in an underwater sonar image. The knowledge obtained for training images is used to identify similar images and the objects in them. A method for underwater plastic waste is proposed in [58]. It modifies the YOLOv5n by reducing its backbone size and the problem of insufficient features is overcome by modifying the feature pyramid, followed by inserting a loss function in the bounding box regression loss of the model. A comparison of the YOLOv5, YOLOv5-TR and YOLOX is performed in [59] for real-time detection and localization of target (a harbor's wall) and proves that the YOLOX has the best detection rate of 91.3 %. The concept given in [60] modifies the YOLOv5 by pre-training, clustering, fine-tuning the pre-training and adding features extraction for high-level features for forward-looking sonar images.

D. TARGET DETECTION USING YOLOv4 AND YOLOv3

The authors in [61] modify YOLOv4. They add a deep separable convolutional layer to the backbone of the network with a feature that allows detection of small objects followed by K-clustering of the bounding box of the dataset with improved size of the box according to the clustering. Also, a spatial pyramid pooling module is added that increases the complexity but also enhances the accuracy. In the last phase, multi-scale training of the model is performed for effective results. The concept in [62] combines YOLOv4 with a fusion mechanism that uses attention mechanism for multiple features; that learns and obtains the features of a number of characteristics utilized for object detection. This approach provides a balance between speedy of detection and accuracy. The framework in [63] also uses YOLOv4 and obtains encouraging detection rate. The technique in [64] studies the YOLOv3 and the deep-sort-multi-target tracking algorithm for fish detection. The research considers the coordinates of a fish to track the path it follows.

The authors in [65] use the YOLOv3 model that makes use of marine pasture biological targets and uses open source images to train the model. It is followed by the testing phase along with optimizing the tuning parameters of the learning process. The output of the learning then results in knowing the objects in the images, their locations and the classification. The image detection module is based on the Pytorch framework and is trained by the open source SeaCLEF image database until the desired optimization level is achieved. The underwater images have usually low light conditions and contrast and, therefore, they are treated based on Fuzzy contrast and enhanced in [66] using a self-adaptive technique followed by the application of YOLOv3 for object detection. The target detection is performed in [67] using the YOLOv3 model in an underwater sonar image and its position is also identified followed by the recurrent neural networks for tracking the path of a dynamic target. The YOLOv3 model is modified in [68] for real time target detection by adding the feature clear and pooling layers to achieve enhanced and effective extraction of features. Moreover, the images are processed by the augmentation, enhancement and equalization processes for improved accuracy of recognition. The idea in [69] uses YOLOv3 for the recognition and detection of objects in side-scan sonar images. The image features are extracted using the various maximum bounding boxes of high credibility and Darknet53 is used as the backbone network for extraction. The algorithm in [70] reduces the detection scale of YOLOv3 by a single decrement and re-clusters the anchor boxes so as to make them appropriate for the considered datasets during the training process. This reduces computational complexity and still maintains a certain degree of accuracy in detection of garbage in water by robots. The authors in [71] up-sample the down-sampling rate and add splicing and features fusion techniques to YOLOv3 to enhance its performance for small target detection.

E. OBJECT DETECTION USING YOLO

The authors in [72] apply the YOLOX algorithm to recognize and detect underwater objects for forward looking SONAR. The algorithm first extracts the features from the images followed by obtaining enhanced features using the FPN and the detected images are then recognized. A mechanism given in [73] dynamically chooses feature layer channels, termed as DC block and is combined with YOLOX to make YOLOX-DC. A network establishment concept with defined local points in underwater environment is given in [74]

TABLE 3. Target detection using YOLOv4 [61], [62], [63], YOLOv3 [64], [65], [66], [67], [68], [69], [70], [71] and YOLO [72], [73], [74], [75], [76]. The symbol x represents the unspecified value.

Reference	Key Concept	Achievement	Limitations/Cost	Detection Rate (%)	Year
[61]	Modifies the YOLOv4 by adding a separable convolution layer with feature map, performing clustering of dataset, introducing spatial pyramid pooling layer and performing multi-scale training of the model	Adds to the accuracy of the model	Enhanced computational complexity	4.8% higher than YOLOv4	2021
[62]	Obtains various features through attention mechanism and combines it with YOLOv4	Provides a balance between speed detection and accuracy	Detection accuracy varies with datasets	92.65	2021
[63]	Uses YOLOv4 and obtains encouraging detection accuracy	Lightweight as does not involve additional operations	Struggles for accuracy in diverse datasets	x	2021
[64]	Uses the deep-sort-multi-target tracking algorithm with YOLOv3 for fish target detection with position tracking	Improved detection accuracy	Involves constant position and coordinates knowledge of the fish	96.16	2023
[65]	Develops an automated system for detection, recognition and classification of marine living organisms in sea bed using YOLOv3. Marine images are captured by cameras, enhanced and processed for object detection	Multiple tasks performance on images	Complexity of operations due to multiple processes involvement	x	2022
[66]	First applies a self-adaptive image enhancement technique to adjust the contrast of the images followed by YOLOv3 for object detection in the images	Fast and accurate target detection	Requires multi-steps processing	x	2021
[67]	Detects objects in an underwater sonar image using YOLOv3 and tracks the path of a dynamic target using recurrent neural networks	Accuracy of detection and tracking	Requires constantly knowing the position of the dynamic target	x	2022
[68]	Adds the feature clear and pooling layers to YOLOv3 and processes the images by augmentation, enhancement and equalization processes for accurate target detection	Effective for real time target detection in complex underwater environment	Deviates from the fast processing of YOLOv3 by adding more features	x	2020
[69]	Uses YOLOv3 with the maximum bounding boxes to extract features and detect target effectively	Low rate of missed detection	Additional features extraction adds to complexity of computation	x	2022
[70]	Modifies the YOLOv3 by re-clustering its anchor boxes and reducing its detection scale for garbage detection by underwater robots	Reduced computational complexity	Clustering is performed for specialized datasets	x	2020
[71]	Adds up-sampling, splicing and features fusion map to modify YOLOv3 for small target detection	Efficient detection of small targets	Enhanced computation	x	2020
[72]	Detects and recognizes the forward looking SONAR images	Effective and precise detection and recognition of objects due to enhanced features of YOLOX	The enhanced features slow down the computational speed	92.6	2022
[73]	The feature layer channels are chosen that further use the feature layers having varied values of receptive field to precisely select features in combination with the single stage YOLOX model for target detection	Fast target detection	Detection accuracy is a fraction of the counterpart scheme	0.9	2022
[74]	Establishes the concept of an underwater network with defined position and performing automatic detection of objects using YOLO model	Automatic detection of objects	Permanent establishment of underwater position is challenging due to the environment dynamism	90	2020
[75]	Modification in YOLO is made with histogram equalization and frames structure similarity is made with transfer learning	Improved detection accuracy	Requires additional steps than YOLO	x	2019
[76]	Uses three different datasets to detect fish	Shows that a diverse dataset is required in the training process to detect objects in data not used during training	Requires further investigation of datasets	x	2018

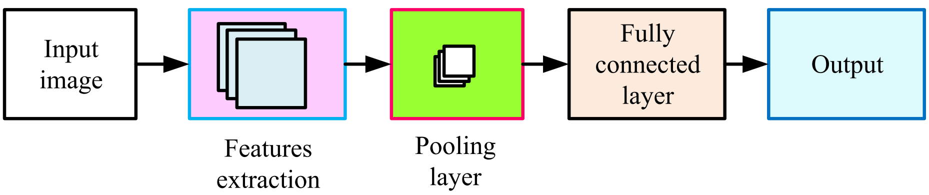

FIGURE 6. Basic architecture of convolutional neural networks. Features are extracted by the convolutional layer that are further reduced in dimensions by the pooling layer. The fully connected layer makes a bigger features map by combining the collected features.

that uses the YOLO version for automatic target detection and eases the manual measurement in future trials. The YOLO algorithm is modified in [75] and transfer learning is adopted to ease the complexity of training and target detection. The concept of histogram equalization is used to deal degradation of image quality. The similarity in structures of frames is utilized for enhancing the frame detection rate. The authors in [76] use three different datasets to train YOLO for detection of fish. The results showed that the model did not detect fish in datasets that were not used in the training process that advocated the use of diverse datasets during training.

III. TARGET DETECTION USING ConVNNs

This section describes the algorithms that make use of the ConVNNs or their variants for underwater target detection.

Figure 6 shows the basic architecture of a ConVNN for object detection. Features are extracted from an object of interest using the convolutional operation between an image (or any signal of interest) I of size $m \times n$ and a kernel or filter F of size $L \times L$ as:

$$O(i,j) = F * I = \sum_{k=1}^{L}\sum_{l=1}^{L} I(i+k-1, j+1-l)F(k,l), \tag{4}$$

where $O(i,j)$ is the output of the convolution matrix at the i^{th} row and j^{th} column considering a single channel convolution and the symbol $*$ represents the convolution operation. The pooling layer removes unnecessary information content from the information it receives from the features extraction layer and reduces its dimension that is further processed by the fully connected layer that combines all the features in a

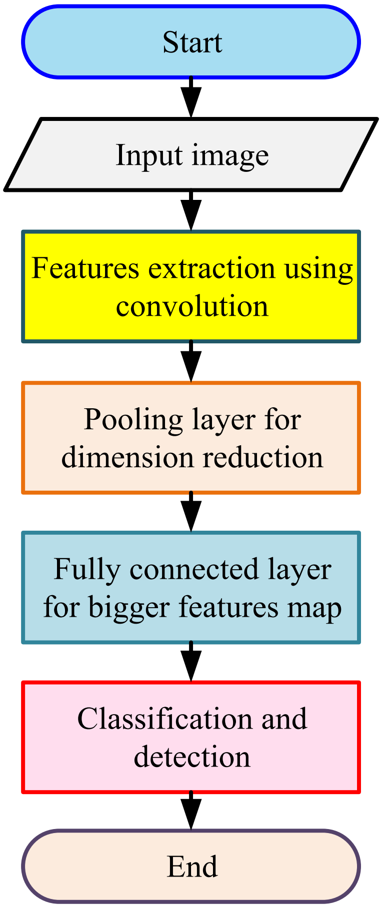

FIGURE 7. Flow chart of the the basic ConVNNs. The features are extracted from an input signal (image, for instance) through convolution operation that are further processed by the pooling layer for dimensions reduction and the fully connected layer to obtain a diverse features map and classify/detect objects.

single array to construct a bigger features map for information classification and object detection. The architecture of the ConVNNs for object detection involves features extraction from an input image (signal) using convolution operation that are further processed by the pooling layer to reduce dimensions and fully connected layer to obtain a features map and classify and detect objects. These operations are further elaborated in the flow chart of the ConVNNs depicted in Figure 7. The target detection using ConVNNs is further classified as given in the lines to follow.

A. UNDERWATER ANIMALS/MOVING OBJECTS DETECTION

The scheme designed in [77] uses deep learning with improved regression network to detect the target followed by prewitt feature enhancement to minimize the features loss and their uneven distribution. The binocular vision is used to determine the position of the target (fish swarm) and their spatial information are displayed on a radar map. The concept given in [78] studies the underwater camera imaging with light refraction and the various internal and external calibration parameters. Moreover, it uses the feature pyramid ConVNNs for detection of the target in the image. The various videos of the motion of the target are optimized for the trajectory calculation as well. The method in [79] identifies the jellyfish and its density. A camera takes the real-time picture of jellyfish and processed by the ConVNNs. The obtained image is enhanced and its edges and their closure are detected and holes are filled in the gray-scale and the binary image is obtained. To detect a lobster in the image in [80], the initial data enhancement is performed in the pre-processing stage by the generative adversarial network and increment method followed by the use of the separable ConVNNs that compress the fully connected layer and make the model light for computation. The voice calls of the underwater mammal species are processed for features extraction by the fractional Fourier transform in [81] as they behave like modulated pulses with a linear frequency. The obtained features are then input to the ConVNNs for voice recognition and, therefore, detect the corresponding mammals. The authors in [82] use limited training data for fish detection. It involves various convolutional layers and residual blocks to detect and segment the target. The weights of the features of interest are increased. The features are effectively presented by the residual blocks after the concatenation of the shallow and deep layers of the model. The authors in [83] design a model that send signals towards the moving fish and their reflections are collected and analyzed using the ConVNNs to detect the moving target. A trade-off existed between the detection bound and accuracy for training the network with real and synthesized data. A recurrent neural network model is also given for online processing with low accuracy. The combination of improved faster region-based ConVNNs and FPN is performed for target detection in [84]. The accuracy and speed of detection are improved by replacing the intersection over union by the distance intersection over union. The authors in [85] improve the detection accuracy and training speed of the faster R-ConVNNs in detection of jellyfish in underwater images. The obtained images are preprocessed to improve their brightness and contrast followed by the integration of the restnet50 into the network for effective features extraction. The training speed is improved by using the semi-precision floating point method.

B. TARGET DETECTION FOR SAFETY PROVISION

The authors in [86] make use of the time and time-frequency spectra of underwater images and process them by ConVNNs to recognize these objects. In addition, the neural network parameters such as the pooling rate, learning rate and batch size are studied for optimal performance. To detect divers and underwater intruders, the authors in [87] first obtain a background image of the surrounding using ConVNNs and then the underwater moving objects are detected using the difference of the current image and predicted image using plan position indication. The ConVNNs are used in [88] to process the images of the drowning objects and detect them for rescue robots. They also focus on reducing the cost and difficulties in the deployment of the existing rescue robots. A deep ConVNNs algorithm is proposed in [89] for the automatic detection and classification of underwater mines in sonar images captured by the synthetic aperture radar. The realistic images are synthetically generated to train the model.

C. TARGET DETECTION FOR SONAR IMAGERY

An approach based on ConVNNs is adopted in [90] that enhances the features of the target in a sonar image and weakens the background so that the false detection is overcome and missed detection is minimized. A ConVNN that has prior training knowledge of the features uses pixel intensity as the features extraction metric to recognize objects/anomalies in the seabed images with small to no false detection is proposed in [91]. The authors in [92] first identify the target region in a forward-looking sonar echoscope image during the pre-processing stage using graph-based manifold ranking and then processed by the deep ConVNNs for the extraction of features and recognition of the target. The method of transfer learning is used to cope with the requirement of the availability of sufficient data. The optimization of the network is performed by the gradient descend method.

D. TARGET DETECTION FOR MISCELLANEOUS APPLICATIONS

The authors in [93] first apply the method of convolutional downsampling to the features of underwater acoustic target to be recognized, which reduces the cost of computational processing. It is followed by learning about the local and global features using a varied set of conformer blocks. Finally, the splicing of the output of the various conformer blocks is carried out and the features of the speech are extracted by the mechanism combining pooling layer and the attention statistics. An algorithm is developed in [94] for the detection and modeling of underwater propeller noise in acoustic signals. The ConVNNs architecture is used to detect and classify the noise for various signal-to-noise ratio (SNR) levels. The authors in [95] argue that sonar images have limited availability of data and are not suitable for deep learning. To overcome this problem, the striation images are used that minimize the need for training data of the ConVNNs, the optimization of fuzzy or discontinuous fringes is performed and the shared latent sparse scheme is used to represent the interference fringes. These features are then correspondingly used to train the network. A method proposed in [96] combines the spatial and spectral features of the target obtained through 3D ConVNNs that are further fine-tuned using the depth information, as these features vary with the water depth. The depth information is also used to predict the accurate depth of the target. The network synthetically generates a copy of the actual hyperspectral data that automatically removes the noise in the real data and is then used to train the model.

The authors in [97] propose an algorithm that blind detects the acoustic signals in underwater communications. First it pre-processes the noise using the generative adversarial network that mitigates the noise in the signal and then the ConVNNs are used to extract the features from the signals and differentiate them from the noise. In addition, a data transfer model is used to overcome the issue of insufficient underwater data for training the respective model. The authors in [98] improve the detection accuracy of the single-shot multibox detector algorithm, which is based on ConVNNs. It obtains the position and detail information of the object using channel-spatial attention mechanism for high value features to improve detection.

The authors in [99] use the reflection signals of an active acoustic emitter to localize, detect and track moving underwater targets with convolutional denoising auto-encoder. The concept of faster R-ConVNNs is used in [100] for object detection that involves the use of the swin transformer as the backbone of the architecture, a path aggregation network for fusing the deep and shallow features maps, online hard mining and using an improved pooling layer to remove quantization error and improve detection.

IV. TARGET DETECTION USING HYBRID ALGORITHMS

The authors in [101] combine YOLO, ConVNNs and SSD for object detection with a comparative analysis to detect even small objects. It is argued in [102] that the existing underwater object detection algorithms involve frequent human-computer interaction, which is not feasible for the automatic operation of the detecting devices. Therefore, a long short-term memory (LSTM) deep learning model based algorithm is utilized to extract and classify the features of the target noise by an underwater glider. It first obtains the data input samples including various noise frequencies and signals, normalizes them and then inputs them to the LSTM for the extraction and classification of features. The authors in [103] explore the resonant behavior of the low frequency sound waves when they are incident on unexploded ordnance. So two deep learning approaches are modeled to detect the unexplored ordnance in the sonar imagery of synthetic aperture radar. These algorithms use sequence models to correlate the spatial features in the resonant sound waves.

A method of automatic detection of underwater objects is given in [104] that uses Fuzzy C-means and K-means global clustering of the images to get many regions of interest followed by local segmentation using the pulse coupled neural network to differentiate the boundary of the target. Extraction of multiple features from the target area is performed and are input to the nonlinear converter to increase the distance of the features and the Fisher discrimination is used to compute a classification threshold and detect the target. A shallow neural network algorithm is proposed in [105] that considers the temporal variations in the amplitude and frequency of the target and clutter signals from

TABLE 4. ConVNNs for detection of Animals/moving objects [77], [78], [79], [80], [81], [82], [83], [84], [85], safety [86], [87], [88], [89], sonar imagery [90], [91], [92] and miscellaneous applications [93], [94], [95], [96], [97], [98], [99], [100].

Reference	Key Concept	Achievement	Limitations/Cost	Detection Rate (%)	Year
[77]	Deep learning with improved regression and prewitt feature enhancement method is used to detect target and later identify its position with the binocular vision	Real-time target detection	Fish motion requires their position to be calculated constantly	83.2	2018
[78]	Detects the objects in an underwater image using feature pyramid ConVNNs and optimizes the trajectory of the moving target by combining images from various fixed cameras	Target detection with its 3D optimized trajectory calculation	Camera functionality is sensitive to underwater light conditions	x	2020
[79]	The detection and count of jellyfish is performed by processing the real-time obtained images through the ConVNNs and performing various operations	Informs about the invasion of marine animals on objects in water	Performance is sensitive to light conditions, especially in deep water	x	2020
[80]	Real-time lobster detection by robots is performed using the generative adversarial network with increment method as data enhancement and the depth-wise separable ConVNNs for prediction of objects with minimized overfitting of the model	Lightweight computation model	Network training is required for the robots prior to real-time detection	90.32	2022
[81]	The voice calls of the mammal species are processed by the fractional Fourier transform due to their behavior as modulated pulses of linear frequency and the obtained features are processed by the ConVNNs to recognize the voices and detect the corresponding specie	Effective for recognizing and detecting mammal species	Training the network with the voices of the mammals is cumbersome as the required dataset has limited existence	97.8	2021
[82]	Performs fish detection and segmentation with limited training data by combining various convolutional layers and residual blocks	Does not require the usually large datasets required for training the model	Addition of more layers adds complexity to the implementation	95.04	2022
[83]	Uses reflection from the moving fish target and processes them with ConVNNs by ignoring the clutter reflection for detection. Also proposes a recurrent neural network version for online processing with low detection accuracy	Real-time moving target detection	In deep water, detecting moving object becomes challenging	x	2020
[84]	Detects fish in sonar images using improved faster R-CoNVNNs and FPN with distance over union intersection method	Accuracy and speedy target detection	High computational complexity	x	2022
[85]	Detects jellyfish in underwater images by first adjusting the brightness and contrast of the images and then using ResNet50 with the backbone of the faster R-ConVNNs with improved training speed	Accurate detection of jellyfish	Complexity of operations due to a number of algorithms involvement	x	2023
[86]	Processes time and frequency-time spectra of underwater objects through the convolutional neural networks to recognize them	Better features detection than the time spectra of the target	High computational processing is involved	98.29	2020
[87]	A mechanism is designed for divers and underwater intruders detection that obtains a background image using deep learning and its difference with the current image using plan position indication	Effective for surveillance of harbors and low depth water bodies	Sensitive to false alarm due to low depth, simple architecture for deep water	x	2018
[88]	Uses ConVNNs on the images of the drowning objects for their detection by the rescue robots	Reduced cost and ease of deployment	Challenging functionality in deep water	x	2022
[89]	Discovers the underwater hidden warfare mines using their specific sound reflection property	Detection of underwater mines in naval warfare	Synthetic generation of images may affect the performance of the model for real world input images	x	2017
[90]	Uses ConVNNs to enhance the features of the target and weaken the background in an underwater image to accurately detect target	Improved accuracy of detection	Complexity of computation	x	2022
[91]	A ConVNN that has prior training knowledge of the features of the objects based on pixel intensity of seabed images uses small datasets for objects/anomalies detection	Small to no false detection, effective for seabed imaging	Requires prior training transfer of the convolutional network	x	2018
[92]	Echoscope images are first pre-processed to identify the target in a forward-looking sonar image followed by the gradient descend optimized deep ConVNNs with attention mechanism to reduce the requirement of large data availability for network training	Focuses target during data preprocessing and reduces the requirement of large data by attention mechanism, automatic recognition	Involves multiple steps of computation	97.3	2019
[93]	Downsamples the features of acoustic target, differentiates the features into the local and global using various conformal blocks whose output is spliced and the features are extracted by using pooling layer and attention statistics	Achieves the reduced error rate with precision of recognition	Involves multiple processing steps	98.65	2022
[94]	Detects and models noise due to an underwater propeller and classifies it to various levels depending upon the ambient conditions	Requires few training samples to train the model	Requires the prior knowledge of the probability distribution of the received signal	x	2020
[95]	Uses striation images, optimizes the fuzzy and discontinuous fringes and differentiates the interference fringes to classify surface objects	Does not require the need for the usual large amount of training data	Challenging performance in deep water	93.34	2019
[96]	Obtains the spectral and spatial features of the underwater target with depth information and synthetically generates the obtained data to overcome noise in training the model and target detection	Accurate and effective target detection due to noise removal	Requires additional steps to combine the spatial and spectral features	x	2023
[97]	Designs a model for blind detection of acoustic signals by denoising the acoustic signals with generative adversarial network and extracting features using the ConVNNs	Effective for high quality underwater communications	Noise frequencies in the communication frequencies range are hard to remove	x	2020
[98]	Uses channel-spatial attention mechanism in the deep layers of the network to obtain high value features	Reduced missed and false detection rate	More computational steps are involved than simple ConVNNs	79.76	2022
[99]	Detects underwater moving targets using reflection signals of an acoustic emitter and a convolutional denoising autoencoder	Energy and computational-efficient detection of moving targets	Localization of moving targets is challenging in underwater environment	x	2019
[100]	Uses swin transformer as the backbone of the architecture, path aggregation network for fusing shallow and deep features, online mining and improved pooling layer	Overcomes missed and false detection in complex underwater environment	Involves multiple steps of computation	80.54	2022

pre-processed spectrographs. The authors in [106] estimate the monocular depth to restore the image affected by the underwater channel properties. This helps later in the target detection based on the depth learning. First images are enhanced in [107] using the max-RGB and shades of grey techniques and then a correlation filter tracking method is combined with the R-ConVNNs to extract the regions of interest and detect objects.

TABLE 5. **Target Detection using hybrid deep learning techniques [102], [103], [104], [105], [106].**

Reference	Key Concept	Achievement	Limitations/Cost	Detection Rate (%)	Year
[101]	Addresses the problem of detecting multiple small objects in underwater images; especially located in the deep sea environment, by comparing YOLO, ConVNNs and SSD	Identification of all objects in an image with tagging	High computation involvement. Moreover, requires a thorough and in-depth analysis of specifying which algorithm is the best for which application(s) as the application in which an algorithm is applied plays a key role to determine the performance metrics of the algorithm	>78	2023
[102]	The LSTM model is used to detect and recognize the underwater noise by a glider. The noise signals and frequencies are obtained, normalized and input to the LSTM to classify the target noise	Automated operation bypassing the human-computer interaction	LSTM suffers from vanishing gradient problem in which the weight of the information decreases with the increasing length of the input sequence that results in information loss. Moreover, underwater noise prevails over a broad spectrum so the noise beyond the designed filter range cannot be detected	94	2018
[103]	Detects the unexploded underwater ordnance by the low frequency incident sound waves that resonate and their spatial correlations are determined by the sequence models	Provides safety and security to the underwater environment from unexploded ordnance	Requires sophisticated circuitry	x	2022
[104]	Small targets are automatically detected by first performing the clustering of the input signals and then segmenting the desired areas of interest in the signals. It is followed by extraction of features and classification based on the similarity in the properties	Low false alarm of detection	Involves multiple steps of computation	x	2022
[105]	The variations in the amplitude and frequency of the target and clutter signals from pre-processed spectrograms are processed to obtain the features of interest and process them for object detection	Time-efficient and fast detection	Not effective for diverse underwater environment due to shallow learning. Moreover, the underwater noise changes with frequency and noise beyond the designed filter limit is challenging to cope with	x	2022
[106]	Uses estimation of monocular depth for image restoration and depth learning for target detection	Ease of target detection due to image restoration	Challenging performance in complex water environment	x	2021
[107]	Preprocesses images using max-RGB and shades of grey techniques and then applies the R-ConVNNs with correlation filter tracking algorithm to extract the regions of interest and detect objects	Convenience of object detection and tracking due to the use of the R-ConVNNs as they specifically target the regions of interest in the signals (images) to extract the features of interest and process them for further decision	Involves multiple computational steps	x	2023

TABLE 6. **Comparative analysis of the classified categories of algorithms.**

Technique	Key Idea	Pros	Cons	Applications
YOLO	Converts an input signal (usually image) into a grid of $S \times S$ cells with bounding box regression applied on each cell for objects detection. Intersection over union is applied to find out objects in multiple boxes and emphasizing on boxes with high detection content	Fast and real-time detection and prediction with a processing speed of 45 frames per second (and even higher with compromised accuracy), simple architecture with minimal training data	Poor detection and generalization problem when objects are small, close or in groups because only two boxes (five in latest version) are predicted in a single grid	Self-driving cars to detect objects, obstacles others cars and predict and follow traffic path, video-surveillance and monitoring
ConVNNs	The convolution operation is used to obtain features from the input signal that are further passed through a varied set of pooling and fully connected layers for dimensions reduction and obtaining a diverse features map (to further perform classification and detection of objects), respectively	Features extraction from input signals and dimensions reduction lead to efficient computation and processing, effective and accurate object detection as they require large datasets for training and obtaining the information patterns from them, robust to noise due to the use of various filters, automated features extraction and once trained, the trained pattern can be efficiently used for other related tasks	Computationally intensive due to the use of various layers, require large datasets for training and, hence, the performance is poor with minimal training datasets, diversity in the training datasets leads to poor performance due to poor ability to generalize	Medical image analysis, object detection, image recognition, speech recognition, obstacles detection and spatio-temporal traffic information collection in intelligent transportation, document analysis
3D ConVNNs	Uses a 3D kernel or filter to perform convolution operation than the traditional 2D version	Able to extract more features due to the extra filter dimension than the 2D ConVNNs version that results in increased accuracy of detection and recognition, can extract spectral and spatial features from the images at the same time	Requires more storage and processing resources	Medical imaging, video surveillance, best for volumetric (3D) data
Faster R-ConVNNs	Obtains features from input signals using ConVNNs that are then processed by a separate network to identify regions of interest instead of using selective search algorithm that are then processed by a reshaping pooling layer that classifies the image and forecasts the values of the bounding boxes	High speed of object detection than R-ConVNNs and fast R-ConVNNs	Struggles with small object detection and requires large training data, high computational complexity and increasing processing time, especially while processing high resolution images	Medical imaging, image recognition, image enhancement, traffic data patterns acquisition in intelligent transportation
LSTM	Considers the object detection and recognition as a regression problem and uses weighted values of the current and previous information with gated operation to predict the input	Capable of retaining short and long-term information in the data	Requires large dataset for training, subject to the well known vanishing (or exploding) gradient problem when information weights decrease with increasing length of data sequence	For prediction and detection applications in medical imaging, intelligent transportation, video surveillance and acoustic signals processing

V. COMPARATIVE ANALYSIS OF THE TARGET DETECTION ALGORITHMS

Based on the description of the classified algorithms, Table 6 shows a comparative analysis of the classified algorithms for underwater target detection. It shows that the YOLO architecture provides a fast and real-time object detection but struggles with the detection of small objects. The ConVNNs and 3D ConVNNs are effective in extraction of features but they have enhanced computational complexity and require intensive training data. The LSTM is capable of retaining

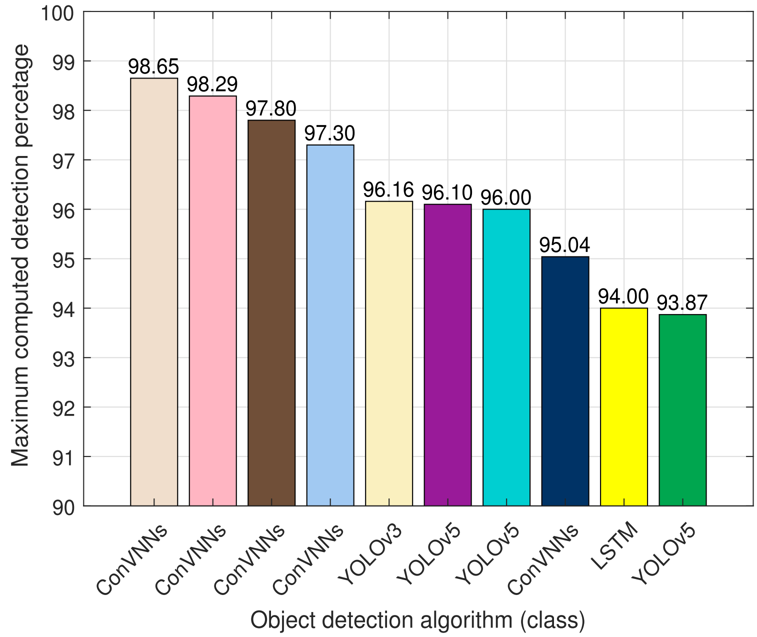

FIGURE 8. **The maximum computed detection percentage of the compared techniques.**

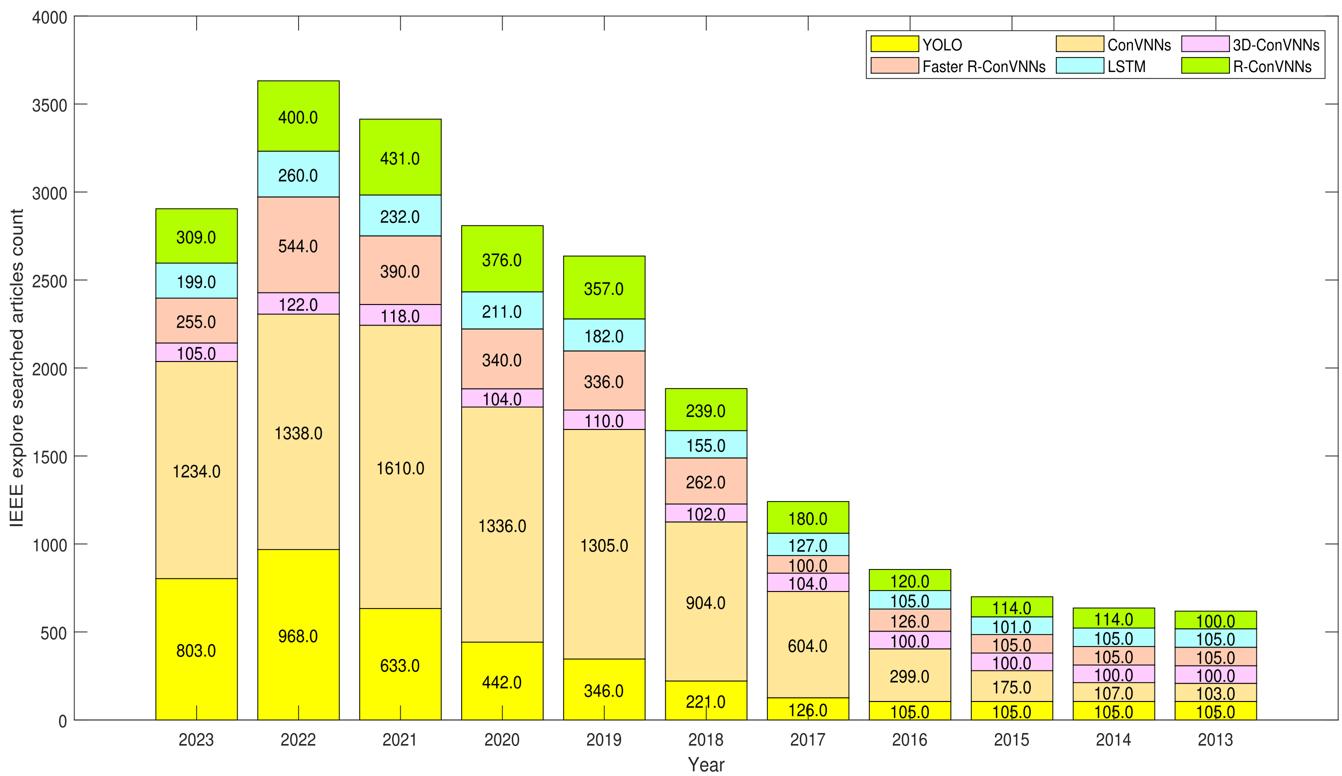

FIGURE 9. **The (normalized) publication count for object (target) detection as searched in the IEEE Explore database for a diverse set of applications for the latest decade (2023-2013).**

short and long-term information of the target detection but it suffers from the vanishing gradient problem, where the weight of the information gradually decreases with the length of information sequences. Figure 8 shows a comparison of the percentage target detection by the compared classes of algorithms. The plotted values indicate the recorded percentage detection as mentioned by the researchers. The algorithms of the researchers that have not measured and mentioned the detection percentage are not included. The plot shows that the maximum computed percentage detection of 98.65 is achieved by the ConVNNs due to the use of multiple features processing layers with enhanced complexity. The second and third highest percentage detection values are also achieved by the ConVNNs, which are 98.29 and 97.80, respectively.

Figure 9 shows the normalized number of articles (for better presentation) published in the latest decade (2023-2013) involving various applications in object detection using the IEEEE Explore database. These applications include, for instance, defects in bicycles, outdoor smoking, wood pith, X-rays, traffic data, security warning, music instruments, railways, microalgae and remote sensing, to mention a few. The ConVNNs and YOLO techniques exhibit a major boom in application to object detection tasks.

VI. CHALLENGES IN UNDERWATER TARGET DETECTION

Keeping in view the challenging underwater conditions, there are a number of challenges associated with underwater target detection, as described in the lines to follow [108], [109].

- The underwater environment has poor light conditions and there is complete darkness beyond a certain depth. This challenges the target detection and identification, especially when the target is in motion. Due to these conditions, the underwater target resembles with its background that challenges the detection probability.
- The size of the underwater target is generally smaller than its surrounding that makes the detection process a challenging task.
- Underwater target are subjected to noise from various sources; such as thermal, shipping, wave and ambient environment, due to which the images of the targets are blurry and distorted. The spectrum of these noise types varies with frequency. Therefore, specific frequencies filters design is required to cope with it. As a result, data pre-processing and enhancement become necessary for underwater target detection.
- When underwater targets are in motion, sophisticated and fast response circuitry design is required to capture the attributes of the target well before they change with motion.
- The presence of various occlusions; such as full or partial covering of the objects by marine life, debris and accumulated waste products, to mention a few, challenges the target detection.
- The differentiation and separation of overlapped objects is specifically challenging as the bounding box approach usually counts all the objects in a box as a single object. This becomes further challenging when the overlapped objects are small or tiny.

VII. CONCLUSION AND FUTURE WORK

A survey of the latest and state-of-the-art underwater target detection algorithms is addressed. These algorithms were classified into various categories depending upon their architecture and their operational mechanisms, merits and demerits were identified for further future enhancements. A comparative analysis is also performed for further providing an insight to the understanding of the classified algorithms. The applications of the described algorithms for the recent decade (2023-2013) in object detection is graphically depicted that provided their scope and importance. The classified algorithms and techniques are effective and useful in underwater object detection in a diverse set of applications such as underwater military and civil radars, precious materials, debris, mines and submarine detection, tracking the path of autonomous underwater vehicles and position calculation of mission robots. They are also beneficial to analyze underwater objects for military and civilian purposes, detect noise over underwater channel for communications, water quality monitoring and ensuring underwater exploration. The conducted study is useful to provide a thorough analysis of underwater target detection algorithms and their mutual comparison in terms of methodologies, structure and operation to highlight their effectiveness and robustness. The challenges in these algorithms are useful to provide future enhancement paths. The discussion of the merits and demerits of the algorithms provides an insight to differentiate them for utilization in specific underwater applications. It also provides clues to outline strategies in the development of more robust, sophisticated, efficient and effective algorithms than the existing algorithms.

The following strategies are effective in future investigation to cope with the challenges in underwater target detection [30].

- **Requirement of a Diverse and Balanced Dataset**. Deep learning models need to be trained to acquire the information patterns hidden in the input datasets so as to predict the objects in the testing phase. Future research investigation needs to have thorough, diverse, robust, balanced and comprehensive datasets owing to the diversified zones and regions of the sea environment so that object detection is performed at a diverse level.
- **Deep Transformer for Efficient Processing**. The use of deep learning techniques such as transformer [110] would reduce the computational delay due to its parallel processing capability unlike the use of the already prevailing algorithms that struggle with computational efficiency.
- **Transfer Learning for Ease of Training and Detection/Prediction**. The transfer learning techniques have the capability of training the deep networks on datasets and then using the information obtaining during training to detect/predict similar and related objects without training the deep network again. This avoids the need for computational rigor and, consequently, results in time-efficient processing.
- **Development of Hybrid Detection Techniques**. Combining the advantages and merits of various data processing and object detection techniques could results in a bulk performance enhancement. For instance,

the techniques for features extraction of convolutional neural networks could be used for multiple bounding box regressions within a single box to detect the tiny and overlapped objects that are challenging to detect with the traditional methods.

- **Multiple Signal Processing Techniques**. Statistical signal processing techniques such as entropy, Fourier transform and Wavelet transform; to mention a few, could be utilized to extract only the informative parts of the signals that could significantly reduce the computational cost.
- **Development of Sophisticated Cameras and Data Processing Circuitries**. With the involvement of big data and heavy training data requirement by deep learning techniques, future investigation needs to consider fast, efficient, modular, reliable and adaptive circuitries to detect the changes in the objects orientation, position and status and include them in the actual status of the objects before the changes happened. This will lead to more reliable, accurate and trustworthy detection.

REFERENCES

[1] L. L. C. Benisto, R. Sukumaran, and M. Saravanan, "Architecture, localization techniques, routing protocols and challenges for UWNS," in *Proc. Int. Conf. Data Sci. Netw. Secur. (ICDSNS)*, Tiptur, India, Jul. 2023, pp. 1–7.

[2] N. Nomikos, P. K. Gkonis, P. S. Bithas, and P. Trakadas, "A survey on UAV-aided maritime communications: Deployment considerations, applications, and future challenges," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 56–78, 2023.

[3] E. Coccolo, C. Delea, F. Steinmetz, R. Francescon, A. Signori, C. N. Au, F. Campagnaro, V. Schneider, F. Favaro, J. Oeffner, B.-C. Renner, and M. Zorzi, "System architecture and communication infrastructure for the RoboVaaS project," *IEEE J. Ocean. Eng.*, vol. 48, no. 3, pp. 1–24, Jul. 2023.

[4] Z. A. Zukarnain, O. A. Amodu, C. Wenting, and U. A. Bukar, "A survey of Sybil attack countermeasures in underwater sensor and acoustic networks," *IEEE Access*, vol. 11, pp. 64518–64543, 2023.

[5] M. Chaudhary, N. Goyal, A. Benslimane, L. K. Awasthi, A. Alwadain, and A. Singh, "Underwater wireless sensor networks: Enabling technologies for node deployment and data collection challenges," *IEEE Internet Things J.*, vol. 10, no. 4, pp. 3500–3524, Feb. 2023.

[6] K. S. Kumar and V. Anand, "Underwater acoustic sensor network: Architecture, challenges, future possibilities in the perspective of IoT," in *Proc. 2nd Int. Conf. Electr., Electron., Inf. Commun. Technol. (ICEEICT)*, Trichirappalli, India, Apr./Jun. 2023, pp. 1–7.

[7] C. Lin, G. Han, J. Jiang, C. Li, S. B. H. Shah, and Q. Liu, "Underwater pollution tracking based on software-defined multi-tier edge computing in 6G-based underwater wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 2, pp. 491–503, Feb. 2023.

[8] A. Khan, M. Imran, A. Alharbi, E. M. Mohamed, and M. M. Fouda, "Energy harvesting in underwater acoustic wireless sensor networks: Design, taxonomy, applications, challenges and future directions," *IEEE Access*, vol. 10, pp. 134606–134622, 2022.

[9] Y. Cheng, H. Wu, Z. Yang, and H. Wang, "Underwater target detection by measuring water-surface vibration with millimeter-wave radar," *IEEE Antennas Wireless Propag. Lett.*, vol. 22, no. 9, pp. 2260–2264, Sep. 2023.

[10] M.-F.-R. Lee and Y.-C. Chen, "Artificial intelligence based object detection and tracking for a small underwater robot," *Processes*, vol. 11, no. 2, p. 312, Jan. 2023.

[11] Y. Liu, Z. Liu, H. Yang, Z. Liu, and J. Liu, "Design and realization of a novel robotic manta ray for sea cucumber recognition, location, and approach," *Biomimetics*, vol. 8, no. 4, p. 345, Aug. 2023.

[12] Y. Zeng, S. Shen, and Z. Xu, "Water surface acoustic wave detection by a millimeter wave radar," *Remote Sens.*, vol. 15, no. 16, p. 4022, Aug. 2023.

[13] J. Yang, Y. Yao, and D. Yang, "Particle filter based on Harris hawks optimization algorithm for underwater visual tracking," *J. Mar. Sci. Eng.*, vol. 11, no. 7, p. 1456, Jul. 2023.

[14] M. Vadivel, S. P. Vimal, V. G. Sivakumar, V. V. Baskar, and M. Selvi, "Internet based defence surveillance robot to prevent intruder activities and auto combat system using SONAR technology," in *Proc. Int. Conf. Innov. Data Commun. Technol. Appl. (ICIDCA)*, Uttarakhand, India, Mar. 2023, pp. 1055–1059.

[15] N. Morozs, P. D. Mitchell, and Y. Zakharov, "Target detection using underwater acoustic networking," in *Proc. OCEANS*, Limerick, Ireland, Jun. 2023, pp. 1–5.

[16] T. Rong, C. Wang, Q. Luo, and Y. Li, "Underwater shipwreck target detection based on audio-visual heterogeneous data," in *Proc. OCEANS*, Limerick, Ireland, Jun. 2023, pp. 1–6.

[17] H. Park, S.-H. Ham, T. Kim, and D. An, "Object recognition and tracking in moving videos for maritime autonomous surface ships," *J. Mar. Sci. Eng.*, vol. 10, no. 7, p. 841, Jun. 2022.

[18] H. Li, X. Cui, and S. Chen, "PolSAR ship detection with optimal polarimetric rotation domain features and SVM," *Remote Sens.*, vol. 13, no. 19, p. 3932, Sep. 2021, doi: 10.3390/rs13193932.

[19] D. Munteanu, D. Moina, C. G. Zamfir, Ş. M. Petrea, D. S. Cristea, and N. Munteanu, "Sea mine detection framework using YOLO, SSD and EfficientDet deep learning models," *Sensors*, vol. 22, no. 23, p. 9536, Dec. 2022.

[20] S. Hożyń, "A review of underwater mine detection and classification in sonar imagery," *Electronics*, vol. 10, no. 23, p. 2943, Nov. 2021.

[21] V. Padmaja, V. Rajendran, and P. Vijayalakshmi, "Retraction note to: Study on metal mine detection from underwater sonar images using data mining and machine learning techniques," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. S1, p. 377, Apr. 2023.

[22] D. Zhao, B. Tan, H. Zhang, and R. Deng, "Monitoring marine oil spills in hyperspectral and multispectral remote sensing data by the spectral gene extraction (SGE) method," *Sustainability*, vol. 14, no. 20, p. 13696, Oct. 2022.

[23] C. Şerban, C. Maftei, and G. Dobrică, "Surface water change detection via water indices and predictive modeling using remote sensing imagery: A case study of Nuntasi-Tuzla Lake, Romania," *Water*, vol. 14, no. 4, p. 556, Feb. 2022.

[24] V. Assumma, M. Bottero, C. Caprioli, G. Datola, and G. Mondini, "Evaluation of ecosystem services in mining basins: An application in the Piedmont region (Italy)," *Sustainability*, vol. 14, no. 2, p. 872, Jan. 2022.

[25] H. I. Farhat, A. Gad, A. Saleh, and S. M. A. El Bakey, "Risks assessment of potentially toxic elements' contamination in the Egyptian red sea surficial sediments," *Land*, vol. 11, no. 9, p. 1560, Sep. 2022.

[26] R. Ou, L. Cai, J. Qiu, H. Huang, D. Ou, W. Li, F. Lin, X. He, L. Wang, and R. Wu, "Simulation experiment of environmental impact of deep-sea mining: Response of phytoplankton community to polymetallic nodules and sediment enrichment in surface water," *Toxics*, vol. 10, no. 10, p. 610, Oct. 2022.

[27] R. Astya and C. K. Jha, "UWSN survey: Challenges and application issues," in *Proc. 4th Int. Conf. Adv. Comput., Commun. Control Netw. (ICACN)*, Greater Noida, India, Dec. 2022, pp. 1855–1858.

[28] J. R. Prasad, A. Parikh, and H. K. Prasanth, "Exploration of deep learning based underwater image processing techniques," in *Proc. 10th Int. Conf. Comput. for Sustain. Global Develop. (INDIACom)*, New Delhi, India, Mar. 2023, pp. 1222–1225.

[29] T. Yin, X. Jiang, Q. Cen, and Z. Zhang, "An efficient real-time marine object detection algorithm based on convolutional attention," in *Proc. 8th Int. Conf. Intell. Comput. Signal Process. (ICSP)*, Xi'an, China, Apr. 2023, pp. 1496–1501.

[30] S. Fayaz, S. A. Parah, and G. J. Qureshi, "Underwater object detection: Architectures and algorithms—A comprehensive review," *Multimedia Tools Appl.*, vol. 81, no. 15, pp. 20871–20916, Mar. 2022.

[31] R. A. Dakhil and A. R. H. Khayeat, "Review on deep learning techniques for underwater object detection," in *Proc. 3rd Int. Conf. Data Sci. Mach. Learn.*, Copenhagen, Denmark, Sep. 2022, pp. 49–63.

[32] S. Mittal, S. Srivastava, and J. P. Jayanth, "A survey of deep learning techniques for underwater image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 6968–6982, Oct. 2023.

[33] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of YOLO algorithm developments," in *Proc. Elseier 8th Int. Conf. Inf. Technol. Quant. Manag.*, Chengdu, China, Feb. 2022, pp. 1066–1073.

[34] N. R. Nandyala and R. K. Sanodiya, "Underwater object detection using synthetic data," in *Proc. 11th Int. Symp. Electron. Syst. Devices Comput. (ESDC)*, Sri City, India, May 2023, pp. 1–6.

[35] J. Li, L. Shi, and S. Guo, "YOLOv7-based land and underwater target detection and recognition," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Harbin, China, Aug. 2023, pp. 1437–1442.

[36] S. Wang, W. Wu, X. Wang, Y. Han, and Y. Ma, "Underwater optical image object detection based on YOLOv7 algorithm," in *Proc. OCEANS*, Limerick, Ireland, Jun. 2023, pp. 1–5.

[37] A. Liu, X. Li, and K. Xu, "Machine learning-based image detection of deep-sea seamounts creatures," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Pasadena, CA, USA, Jul. 2023, pp. 5735–5737.

[38] Y. Wang, Z. Dai, J. Liu, and P. Wu, "Underwater target detection technology based on YOLOv7," in *Proc. IEEE Int. Conf. Sensors, Electron. Comput. Eng.*, Jinzhou, China, Aug. 2023, pp. 265–270.

[39] X. Li, Y. Kou, and Y. Liang, "Binocular matching method for detecting and locating marine resources using binocular cameras and deep learning," in *Proc. 4th Int. Conf. Comput. Vis., Image Deep Learn. (CVIDL)*, Zhuhai, China, May 2023, pp. 381–384.

[40] A. Balaji, S. Yogesh, C. K. Kalyaan, R. Narayanamoorthi, G. Dooly, and S. Dhanalakshmi, "Deep WaveNet-based YOLOv5 for underwater object detection," in *Proc. OCEANS*, Limerick, Ireland, Jun. 2023, pp. 1–5.

[41] P. Li, A. Zhao, Y. Fan, and Z. Pei, "Research on underwater robust object detection method based on improved YOLOv5s," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Harbin, China, Aug. 2023, pp. 1185–1189.

[42] D. A. R. Ellen, P. Kristalina, M. Z. S. Hadi, and A. Patriarso, "Effective searching of drowning victims in the river using deep learning method and underwater drone," in *Proc. Int. Electron. Symp. (IES)*, Denpasar, Indonesia, Aug. 2023, pp. 569–574.

[43] C. Escobar-Amado, M. Badiey, and L. Wan, "Computer vision for bioacoustics: Detection of bearded seal vocalizations in the Chukchi shelf using YOLOV5," *IEEE J. Ocean. Eng.*, early access, pp. 1–12, Sep. 2023.

[44] Y. Wei, Y. Fang, F. Cheng, M. Zhang, M. Cao, and H. Zhang, "A lightweight underwater target detection network for seafood," in *Proc. 42nd Chin. Control Conf. (CCC)*, Tianjin, China, Jul. 2023, pp. 8381–8387.

[45] A. Zhang and X. Zhu, "Research on ship target detection based on improved YOLOv5 algorithm," in *Proc. 5th Int. Conf. Commun., Inf. Syst. Comput. Eng. (CISCE)*, Guangzhou, China, Apr. 2023, pp. 459–463.

[46] E. Ranolo, K. Gorro, A. Ilano, H. Pineda, C. Sintos, and A. J. Gorro, "Underwater and coastal seaweeds detection for fluorescence seaweed photos and videos using YOLOV3 and YOLOV5," in *Proc. 2nd Int. Conf. Innov. Technol. (INOCON)*, Bangalore, India, Mar. 2023, pp. 1–5.

[47] Y. Zhang, T. Liu, and X. Yu, "Contextual and lightweight network for underwater object detection with self-attention mechanism," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Harbin, China, Aug. 2023, pp. 1644–1649.

[48] Z. Yin, S. Zhang, R. Sun, Y. Ding, and Y. Guo, "Sonar image target detection based on deep learning," in *Proc. Int. Conf. Distrib. Comput. Electr. Circuits Electron. (ICDCECE)*, Ballar, India, Apr. 2023, pp. 1–9.

[49] M. F. Ercan, N. I. Muhammad, and M. R. N. Bin Sirhan, "Underwater target detection using deep learning," in *Proc. IEEE Region Conf. (TENCON)*, Hong Kong, Nov. 2022, pp. 1–5.

[50] C. Xing, B. Sun, and W. Zhang, "Image-enhanced YOLOv5 and deep sort underwater multi-moving target tracking method," in *Proc. 5th Int. Symp. Auton. Syst.*, Hangzhou, China, Apr. 2022, pp. 1–6.

[51] R. Chen, S. Zhan, and Y. Chen, "Underwater target detection algorithm based on YOLO and Swin transformer for sonar images," in *Proc. OCEANS*, VA, USA, Oct. 2022, pp. 1–7.

[52] G. Wen, F. Liu, X. Luo, and S. Li, "Underwater target detection based on modified YOLOv5," in *Proc. 5th Int. Conf. Intell. Auton. Syst. (ICoIAS)*, Dalian, China, Sep. 2022, pp. 7–12.

[53] Y. Li, W. Liu, L. Li, W. Zhang, J. Xu, and H. Jiao, "Vision-based target detection and positioning approach for underwater robots," *IEEE Photon. J.*, vol. 15, no. 1, pp. 1–12, Feb. 2023.

[54] B. Le, H. Ying, and C. H. Kun, "A marine benthic detection algoritm based on improved lightweight network," in *Proc. 34th Chin. Control Decis. Conf. (CCDC)*, Hefei, China, Aug. 2022, pp. 5795–5800.

[55] Z. Zhang, Q. Tong, C. Yi, X. Fu, J. Ai, and Z. Wang, "The appropriate image enhancement method for underwater object detection," in *Proc. IEEE 22nd Int. Conf. Commun. Technol. (ICCT)*, Nanjing, China, Nov. 2022, pp. 1627–1632.

[56] S. Fu, F. Xu, J. Liu, Y. Pang, and J. Yang, "Underwater small object detection in side-scan sonar images based on improved YOLOv5," in *Proc. 3rd Int. Conf. Geol., Mapping Remote Sens. (ICGMRS)*, Zhoushan, China, Apr. 2022, pp. 446–453.

[57] I. Shogo and S. Masahiko, "Similar object identification in sonar images by machine learning," in *Proc. IEEE Underwater Technol. (UT)*, Tokyo, Japan, Mar. 2023, pp. 1–6.

[58] Z. Hu and C. Xu, "Detection of underwater plastic waste based on improved YOLOv5n," in *Proc. 4th Int. Conf. Frontiers Technol. Inf. Comput. (ICFTIC)*, Qingdao, China, Dec. 2022, pp. 404–408.

[59] M. Aubard, A. Madureira, L. Madureira, and J. Pinto, "Real-time automatic wall detection and localization based on side scan sonar images," in *Proc. IEEE/OES Auton. Underwater Vehicles Symp. (AUV)*, Singapore, Sep. 2022, pp. 1–6.

[60] H. Zhang, M. Tian, G. Shao, J. Cheng, and J. Liu, "Target detection of forward-looking sonar image based on improved YOLOv5," *IEEE Access*, vol. 10, pp. 18023–18034, 2022.

[61] W. Hao and N. Xiao, "Research on underwater object detection based on improved YOLOv4," in *Proc. 8th Int. Conf. Inf., Cybern., Comput. Social Syst. (ICCSS)*, Beijng, China, Dec. 2021, pp. 166–171.

[62] M. Zhang, S. Xu, W. Song, Q. He, and Q. Wei, "Lightweight underwater object detection based on YOLO v4 and multi-scale attentional feature fusion," *Remote Sens.*, vol. 13, no. 22, p. 4706, Nov. 2021.

[63] W. Wang and Y. Wang, "Underwater target detection system based on YOLOv4," in *Proc. 2nd Int. Conf. Artif. Intell. Inf. Syst.*, Chongqing, China, May 2021, pp. 1–5.

[64] L. Zhang, G. Zhai, B. Hu, Z. Qiao, and P. Zhang, "Fish target detection and speed estimation method based on computer vision," in *Proc. IEEE 6th Int. Conf. Electron. Inf. Commun. Technol. (ICEICT)*, Qingdao, China, Jul. 2023, pp. 1330–1336.

[65] P. Yang, Y. Yu, and P. Yao, "An automatic detection system for marine ranch biological targets based on deep learning," in *Proc. 3rd Int. Conf. Mach. Learn., Big Data Bus. Intell. (MLBDBI)*, Taiyuan, China, Dec. 2021, pp. 696–699.

[66] T. Guo, Y. Wei, H. Shao, and B. Ma, "Research on underwater target detection method based on improved MSRCP and YOLOv3," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Takamatsu, Japan, Aug. 2021, pp. 1158–1163.

[67] X. Cao, L. Ren, and C. Sun, "Dynamic target tracking control of autonomous underwater vehicle based on trajectory prediction," *IEEE Trans. Cybern.*, vol. 53, no. 3, pp. 1968–1981, Mar. 2023.

[68] Z. Wang, X. Ye, Y. Han, S. Guo, X. Yan, and S. Wang, "Improved real-time target detection algorithm for similar multiple targets in complex underwater environment based on YOLOv3," in *Proc. Global Oceans*, Biloxi, MS, USA, Singapore: U.S. Gulf Coast, Oct. 2020, pp. 1–6.

[69] J. W. Li and X. Cao, "Target recognition and detection in side-scan sonar images based on YOLOv3 model," in *Proc. 41st Chin. Control Conf.*, Hefei, China, Jul. 2022, pp. 7191–7196.

[70] X. Li, M. Tian, S. Kong, L. Wu, and J. Yu, "A modified YOLOv3 detection method for vision-based water surface garbage capture robot," *Int. J. Adv. Robot. Syst.*, vol. 17, no. 3, pp. 1–11, May 2020.

[71] F. Jie and W. Pingbo, "Application of improved YOLO V3 algorithm for target detection in echo image of sonar underwater reverb," *J. Phys., Conf. Ser.*, vol. 1748, pp. 1–9, Nov. 2020.

[72] B. Xie, S. He, and X. Cao, "Target detection for forward looking sonar image based on deep learning," in *Proc. 41st Chin. Control Conf. (CCC)*, Hefei, China, Jul. 2022, pp. 7191–7196.

[73] W. Ouyang, Y. Wei, and D. Wang, "YOLOX-DC: A small target detection network up to underwater scenes," in *Proc. OCEANS*, Hampton Roads, VA, USA, Oct. 2022, pp. 1–6.

[74] L. Beaudoin and L. Avanthey, "Underwater field equipment of a network of landmarks optimized for automatic detection by AI," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Waikoloa, HI, USA, Sep./Oct. 2020, pp. 1572–1575.

[75] L. Qingzhong, L. Yibing, and N. Jiong, "Real-time detection of underwater fish based on improved YOLO and transfer learning," *Pattern Recognit. Artif. Intell.*, vol. 32, no. 3, pp. 193–203, 2019.

[76] W. Xu and S. Matzner, "Underwater fish detection using deep learning for water power applications," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Las Vegas, NV, USA, Dec. 2018, pp. 313–318.

[77] L. Xu, Y. Wei, X. Wang, A. Wang, and L. Guan, "Binocular vision of fish swarm detection in real-time based on deep learning," in *Proc. OCEANS MTS/IEEE Charleston*, Charleston, SC, USA, Oct. 2018, pp. 1–6.

[78] T. Liu, N. Wang, L. Zhang, S. Ai, and H. Du, "A novel visual measurement method for three-dimensional trajectory of underwater moving objects based on deep learning," *IEEE Access*, vol. 8, pp. 186376–186392, 2020.
[79] Y. Liu, W. Meng, and H. Zong, "Jellyfish recognition and density calculation based on image processing and deep learning," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Hefei, China, Aug. 2020, pp. 922–927.
[80] S. Wang, J. Guo, S. Guo, Q. Fu, and J. Xu, "Study on real-time recognition of underwater live shrimp by the spherical amphibious robot based on deep learning," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Guilin, China, Aug. 2022, pp. 917–922.
[81] X. Zhang, Z. Liu, W. Wang, and J. Xu, "Automated detection of marine mammal species based on short-time fractional Fourier transform," in *Proc. OCEANS*, San Diego, CA, USA, Sep. 2021, pp. 1–5.
[82] H.-Y. Lin, S.-L. Tseng, and J.-Y. Li, "SUR-Net: A deep network for fish detection and segmentation with limited training data," *IEEE Sensors J.*, vol. 22, no. 18, pp. 18035–18044, Sep. 2022.
[83] A. Testolin, D. Kipnis, and R. Diamant, "Detecting submerged objects using active acoustics and deep neural networks: A test case for pelagic fish," *IEEE Trans. Mobile Comput.*, vol. 21, no. 8, pp. 2776–2788, Aug. 2022.
[84] D. Zhao, B. Yang, Y. Dou, and X. Guo, "Underwater fish detection in sonar image based on an improved faster RCNN," in *Proc. 9th Int. Forum Electr. Eng. Autom. (IFEEA)*, Zhuhai, China, Nov. 2022, pp. 358–363.
[85] B. Weihong, J. Yun, L. Jiaxin, S. Lingling, F. Guangwei, and J. Wa, "In-situ detection method of jellyfish based on improved faster R-CNN and FP16," *IEEE Access*, vol. 11, pp. 81803–81814, 2023.
[86] S. Li, X. Jin, S. Yao, and S. Yang, "Underwater small target recognition based on convolutional neural network," in *Proc. Global Oceans*, Biloxi, MS, USA, Singapore: U.S. Golf Coast, Oct. 2020, pp. 1–7.
[87] M. Ogawa, H. Tanaka, J. Muramatsu, M. Nakano, K. Yoshida, and T. Asakura, "Application of deep learning to underwater invasion warning system," in *Proc. MTS/IEEE Kobe Techno-Oceans (OTO)*, Kobe, Kobe, Japan, May 2018, pp. 1–5.
[88] Y. Wang, Z. Guo, and J. Xu, "Underwater search and rescue robot based on convolutional neural network," in *Proc. IEEE 4th Int. Conf. Civil Aviation Saf. Inf. Technol. (ICCASIT)*, Dali, China, Oct. 2022, pp. 786–790.
[89] K. Denos, M. Ravaut, A. Fagette, and H.-S. Lim, "Deep learning applied to underwater mine warfare," in *Proc. OCEANS*, Aberdeen, U.K., Jun. 2017, pp. 1–7.
[90] Y. Liu, R. Wang, K. Cao, J. Wang, Z. Shi, Y. Wang, and Y. Zhou, "Sonar target detection based on a dual channel attention convolutional network," in *Proc. 12th Int. Conf. Inf. Sci. Technol. (ICIST)*, Kaifeng, China, Oct. 2022, pp. 255–264.
[91] D. Einsidler, M. Dhanak, and P.-P. Beaujean, "A deep learning approach to target recognition in side-scan sonar imagery," in *Proc. OCEANS MTS/IEEE Charleston*, Charleston, SC, USA, Oct. 2018, pp. 1–4.
[92] L. Jin, H. Liang, and C. Yang, "Accurate underwater ATR in forward-looking sonar imagery using deep convolutional neural networks," *IEEE Access*, vol. 7, pp. 125522–125531, 2019.
[93] X. Wang, C. Zhang, and R. Wang, "Underwater acoustic target recognition technology based on MFA-conformer," in *Proc. 2nd Int. Conf. Electron. Inf. Eng. Comput. Technol. (EIECT)*, Yanan, China, Oct. 2022, pp. 214–217.
[94] M. Thomas, F. Lionel, and D.-P. Laurent, "Propeller noise detection with deep learning," in *Proc. ICASSP - IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 306–310.
[95] X. Zhou, K. Yang, and R. Duan, "Deep learning based on striation images for underwater and surface target classification," *IEEE Signal Process. Lett.*, vol. 26, no. 9, pp. 1378–1382, Sep. 2019.
[96] Q. Li, J. Li, T. Li, Z. Li, and P. Zhang, "Spectral–spatial depth-based framework for hyperspectral underwater target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, May 2023, Art. no. 4204615.
[97] Y. Li, B. Wang, G. Shao, S. Shao, and X. Pei, "Blind detection of underwater acoustic communication signals based on deep learning," *IEEE Access*, vol. 8, pp. 204114–204131, 2020.
[98] Q. Li, M. J. Er, L. Li, J. Chen, and J. Wu, "Underwater object detection based on improved SSD with convolutional block attention," in *Proc. 5th Int. Conf. Intell. Auton. Syst. (ICoIAS)*, Dalian, China, Sep. 2022, pp. 37–42.
[99] A. Testolin and R. Diamant, "Underwater acoustic detection and localization with a convolutional denoising autoencoder," in *Proc. IEEE 8th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, Le Gosier, Guadeloupe, Dec. 2019, pp. 281–285.
[100] J. Liu, S. Liu, S. Xu, and C. Zhou, "Two-stage underwater object detection network using Swin transformer," *IEEE Access*, vol. 10, pp. 117235–117247, 2022.
[101] S. Raavi, P. B. Chandu, and T. SudalaiMuthu, "Automated recognition of underwater objects using deep learning," in *Proc. 7th Int. Conf. Trends Electron. Informat. (ICOEI)*, Tirunelveli, India, Apr. 2023, pp. 1055–1059.
[102] Z. Shao-Kang, T. De-Yan, W. Chao, and Z. Xiao-Chuan, "Intelligent recognition of underwater acoustic target noise on underwater glider platform," in *Proc. Chin. Autom. Congr. (CAC)*, Xi'An, China, Nov. 2018, pp. 4189–4193.
[103] T. Hoang, K. S. Dalton, I. D. Gerg, T. E. Blanford, D. C. Brown, and V. Monga, "Domain enriched deep networks for munition detection in underwater 3D sonar imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Kuala Lumpur, Malaysia, Jul. 2022, pp. 815–818.
[104] T. Zhou, J. Si, L. Wang, C. Xu, and X. Yu, "Automatic detection of underwater small targets using forward-looking sonar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Art. no. 4207912, 2022.
[105] Y. Choo, K. Lee, W. Hong, S.-H. Byun, and H. Yang, "Active underwater target detection using a shallow neural network with spectrogram-based temporal variation features," *IEEE J. Ocean. Eng.*, early access, pp. 1–15, Jul. 2022.
[106] H. Chen, J. Lin, L. Zhuge, and X. Xia, "Underwater image restoration and target detection based on monocular depth estimation," in *Proc. China Autom. Congr. (CAC)*, Beijng, China, Oct. 2021, pp. 5597–5601.
[107] M. S. Sangari, K. Thangaraj, U. Vanitha, N. Srikanth, J. Sathyamoorthy, and K. Renu, "Deep learning-based object detection in underwater communications systems," in *Proc. 2nd Int. Conf. Elect., Electron., Inf. Commun. Technol.*, Trichirappalli, India, Apr. 2023, pp. 1–6.
[108] 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.
[109] Z. Cao, "Deep learning methods for objective detection," in *Proc. IEEE 2nd Int. Conf. Data Sci. Comput. Appl. (ICDSCA)*, Dalian, China, Oct. 2022, pp. 1353–1357.
[110] A. B. Amjoud and M. Amrouch, "Object detection using deep learning, CNNs and vision transformers: A review," *IEEE Access*, vol. 11, pp. 35479–35516, 2023.

ANWAR KHAN received the M.Sc. degree in electronics from the Department of Electronics, University of Peshawar, in 2006, the M.Phil. degree in electronics from Quaid-i-Azam University, Islamabad, Pakistan, in 2009, and the Ph.D. degree in electronics from the Department of Electronics, Quaid-i-Azam University, in 2018, in underwater wireless sensor networks. He joined the Department of Electronics, University of Peshawar, in 2009, as a Faculty Member. He has published articles with IEEE, Springer, Elsevier, Taylor & Francis, MDPI, and *KSII Transactions on Internet and Information Systems*. He is also a reviewer with these publishers and one of the members of the technical program committee of several conferences. He has supervised and co-supervised more than ten graduate and master's students. His current research interests include underwater wireless sensor networks, energy harvesting, deep learning, machine learning, object detection, game-theoretic optimization, and intelligent transportation.

MOSTAFA M. FOUDA (Senior Member, IEEE) received the B.S. degree (as the valedictorian) and the M.S. degree in electrical engineering from Benha University, Egypt, in 2002 and 2007, respectively, and the Ph.D. degree in information sciences from Tohoku University, Japan, in 2011. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Idaho State University, ID, USA. He is currently a Full Professor with Benha University. He was an Assistant Professor with Tohoku University and a Postdoctoral Research Associate with Tennessee Technological University, TN, USA. He has (co)authored more than 160 technical publications. His current research interests include cybersecurity, communication networks, signal processing, wireless mobile communications, smart healthcare, smart grids, AI, and the IoT. He has guest-edited a number of special issues covering various emerging topics in communications, networking, and health analytics. He is also serving on the Editorial Board of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE INTERNET OF THINGS JOURNAL, and IEEE ACCESS. He has received several research grants, including NSF Japan–U.S. Network Opportunity 3 (JUNO3).

DINH-THUAN DO (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from Vietnam National University Ho Chi Minh City (VNU-HCM), in 2007 and 2012, respectively. Prior to joining academia, he was a Senior Engineer in the telecommunications industry with VinaPhone Mobile Network (the biggest cellular network provider in Vietnam), from 2003 to 2009. Before joining the University of Mount Union, he was a Research Scientist with the University of Colorado Denver, in 2022, The University of Texas at Austin, in 2021, and an Assistant Professor with Asia University, Taiwan, from 2020 to 2021. He is currently serving as an Associate Editor for IEEE TRANSACTION ON VEHICULAR TECHNOLOGY and *Computer Communications* (Elsevier). He has also served as a Lead Guest Editor/Guest Editor in more than 20 special issues of journals, such as *Physical Communications* (Elsevier) and *Annals of Telecommunications* (Elsevier). His publications include over 120 SCIE/SCI-indexed journal articles, five edited books (IET, Springer), and over 50 international conference papers. He was a recipient of the 2015 Golden Globe Award by Vietnamese Ministry of Science and Technology (top 10 outstanding scientists nationwide). He also received the Medal of Creative Young Talents, in 2015. He was named in top 14 highly cited scientists at Asia University, Taiwan, in 2021 (Stanford's list of top 2% scientists in the world, in October 2020, October 2021, and October 2022).

ABDULAZIZ ALMALEH received the master's and Ph.D. degrees in information science from the prestigious University of Pittsburgh. He is currently a Renowned Academician and a Researcher in the field of information science, specializing in machine learning, smart cities, and smart infrastructure. With the master's and Ph.D. degrees, he has established himself as an Expert in his domain, with a strong understanding of emerging technologies and research methodologies. Throughout his academic journey, he has demonstrated a deep commitment to advancing the field of information science, focusing on the practical applications of machine learning and its implications for the development of smart cities and infrastructure. His research interests include a wide variety of subjects, including data analytics, artificial intelligence, the Internet of Things (IoT), and the integration of these technologies into urban planning and development processes. He has an impressive publication record that attests to his dedication and expertise in his chosen field. His numerous works have been published in well-respected journals and conference proceedings, contributing to the collective knowledge and understanding of machine learning applications in the creation of intelligent urban environments. His research findings have greatly influenced the way in which advanced technologies are applied to urban planning, offering practical solutions to the modern challenges faced by cities around the world.

ABDULLAH M. ALQAHTANI received the B.Sc. degree in communication and network engineering from King Khalid University, Abha, Saudi Arabia, in 2012, the M.Sc. degree in digital communications networks engineering from Southern Methodist University, Dallas, TX, USA, in 2016, and the Ph.D. degree from the School of Electronic and Electrical Engineering, University of Leeds, Leeds, U.K., in 2023. He is currently an Associate Professor in computer and network engineering with the School of Computer Science and Information Technology, Jazan University. His research interests include the optimization techniques of edge computing, the Internet of Things (IoT), artificial intelligence, and optical access networks.

ATIQ UR RAHMAN received the Bachelor of Science (B.S.) degree in computer science from the University of Peshawar, in 2004, the Master of Science (M.S.) degree in computer system engineering from the GIK Institute of Engineering Sciences and Technology, Pakistan, in 2008, and the Ph.D. degree in information technology from Universiti Teknologi PETRONAS, Malaysia, in 2013. He was a System Engineer with Pakistan's National Database and Registration Authority, from 2004 to 2006. Later, he was an Assistant Professor, the Head of the Computer Science Department, and the Head of the Quality and Academic Accreditation Unit, Northern Border University, Saudi Arabia, from 2013 to 2019. Currently, he is an Assistant Professor with the Higher Colleges of Technology, United Arab Emirates. His research interests include energy-efficient design, routing, deployment, coverage, and the quality of service in wireless sensor networks and underwater sensor networks. He is a Distinguished Senior Member of esteemed professional organizations, such as the Universal Association of Computer and Electronics Engineers (UACEE), the Institute of Research Engineers and Doctors (IRED), and the American Society for Engineering Education (ASEE). With a significant number of research publications in renowned journals and conferences, he actively contributes as a reviewer to several journals. He is a technical program committee member at international conferences. Additionally, he also participated in the "Fundamental and Advanced Assessment Workshop" organized by ABET in the USA.