IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 7 September 2023, accepted 27 September 2023, date of publication 29 September 2023,
date of current version 9 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3320949

RESEARCH ARTICLE

Fish Monitoring in Aquaculture Using Multibeam Echosounders and Machine Learning

JÓHANNUS KRISTMUNDSSON[1,2], ØYSTEIN PATURSSON[3], JOHN POTTER[4], AND QIN XIN[1]

[1]Faculty of Science and Technology, University of the Faroe Islands, Tórshavn 110, Faroe Islands
[2]Department of Fjord Dynamics, Fiskaaling, Hvalvík 430, Faroe Islands
[3]RAO, Kirkubøur 175, Faroe Islands
[4]Centre for Geophysical Forecasting, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway

Corresponding author: Jóhannus Kristmundsson (johannusk@setur.fo)

This work was supported in part by Research Council Faroe Islands, in part by Fiskaaling, in part by the University of the Faroe Islands, in part by Waive (Formerly Aquabio), and in part by Føroyagrunnurin. The work of John Potter was supported by the Centre for Geophysical Forecasting (CGF), Norwegian University of Science and Technology (NTNU), under Grant 309960 (Research Council of Norway).

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT Offshore salmon aquaculture is a growing industry that faces challenges such as sea lice infestations and varying environmental conditions, necessitating the development of new monitoring systems to improve fish welfare and sustainability. In this paper, we propose and test a machine learning based method for underwater detection and localisation using multibeam echosounders (MBES) in fish farming applications. We demonstrate a three-stage process involving data acquisition, pre-processing, and object detection. We then compare the performance of four different vision based deep learning object detection algorithms in different signal-to-noise scenarios by artificially adding noise to the pre-beamformed signals. This method successfully detects fish in MBES images, which has potential applications in optimising feeding schedules, behaviour analysis, and fish health monitoring. Furthermore, this method holds potential for the detection and tracking of other objects within fish farms, such as cages and mooring lines. This study paves the way for further development of MBES data being used as a non-invasive and automated monitoring method in aquaculture.

INDEX TERMS Salmon, observation, echosounder, multibeam, automatic object detection, target detection.

I. INTRODUCTION

Aquaculture is driving the increase in global aquatic food production, offering crucial contributions to food security and fuelling economic growth worldwide [1]. To keep up with global trends, offshore salmon farming in the northern Atlantic has moved from sheltered locations to include exposed farming sites [2]. However, larger waves and stronger currents at exposed sites set high requirements for the equipment needed for fish farming. This study addresses the predominant form of offshore salmon farming: the gravity-type fish farm, where salmon are housed in large, flexible net cages [3]. An example is shown in Figure 1. The cages can be up to 60m in diameter and 30m deep. The shape of the cage is often square or round [3]. These cages move around due to mooring flexibility and undergo significant deformation when subjected to strong forcing from currents or large waves. As wave height increases, fish dive deeper to avoid unsteady water at the surface and swim farther from the net, reducing the available volume inside the cage that the fish can occupy [4].

Various challenges arise during the offshore phase of the growth cycle which can adversely affect the health and welfare of farmed fish. Among these issues is sea lice infestations [5], which can directly affect fish health, expose them to infection, and indirectly affect fish welfare through delousing treatments [6]. Welfare concerns may also arise due to low oxygen conditions [7], causing fish to seek swimming

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

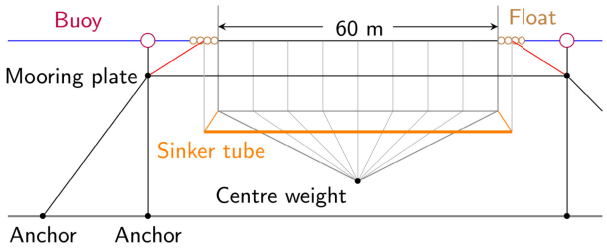

FIGURE 1. Cross section of the main components of an example offshore cage, including moorings to the mooring plates, frames, anchors, and fastening buoys.

depths with more favourable oxygen levels [6]. Moreover, other factors such as temperature [8], time of day [9], currents [4], [9], and waves [4] can also influence fish diving depth. Each of these factors can serve as an indicator of potential welfare issues for salmon. To enhance the welfare of fish and manage these challenges, fish farmers garner location-specific experiences through their work. This knowledge is frequently hard-earned, but extremely valuable as it supports better decision-making that can prevent incidents that lead to stress or harm to the farmed fish.

To address these issues, a new approach called Precision Fish Farming has been proposed [10]. This method suggests a shift from experience-based to knowledge-based fish farming, based on observable data. Observing what is happening below the water line is crucial for data-driven and knowledge-based approaches, as it provides the accurate and reliable data necessary for informed decision-making.

Monitoring the underwater environment in fish farming is a difficult task, as traditional monitoring methods such as underwater cameras and the capture of live fish for welfare scoring can be limited in range and/or sample size, labour-intensive and stressful for fish. There is a clear need for non-invasive monitoring tools that can overcome these challenges and provide accurate and reliable estimates of key parameters [11]. Among the various non-invasive monitoring tools available, acoustic observations are particularly promising. This is primarily because this technology has the range to capture information on the scale of fish farms [4], contrary to alternatives such as submerged camera-based systems where the maximum range varies between 0.5 - 25m [10]. The simplest form of acoustic observation is the active single-beam sonar which, by producing acoustic pulses and measuring the backscatter, can produce information about the range and target strength of objects in the field of view. In single-beam sonars, target detection has been performed by applying a threshold on received echo level intensity [12], and by extension, counting the number of targets provides a count of the fish [13]. However, one limitation of using single-beam sonars is that only targets within the beam can be observed, since there is no way to steer the beam. Therefore, a single beam system can only observe a part of the fish farm, and the information acquired is limited to target strength and range. This limitation also means that individual or multiple targets at the same distance within the beam can not be distinguished.

Multibeam echosounder (MBES) is a technology that electronically steers beams in different directions using array processing and which can provide high-resolution images of the underwater environment. A MBES can provide real-time location and target strength information over many beams, covering a wide field of view. Additionally, secondary information can be estimated from the backscatter data, such as biomass, spatial distribution, swimming behaviour and health indicators, in addition to the presence and position of other objects such as cage outline, mooring lines, predators, or intruders.

Multibeam echosounders have rarely been used in fish farming applications. To the authors' knowledge, only a few other studies have used a MBES to extract information from an offshore fish farm, e.g. [14]. Other studies have investigated the acoustic effects inside an offshore fish farm, including metroscopic wave physics and ultra-slow acoustic energy transport [15], [16].

The use of MBES in fish farming poses significant challenges due to the large amount of data generated and the labour intensive nature of manual analysis of backscatter data. To address these challenges and investigate the utility of MBES in fish farming, there is a need to develop methods for automatic data analysis. Object detection and classification are key tasks for extracting useful information from the data produced by a sonar looking into a fish cage. Object detection involves locating and identifying objects of interest in an image, while object classification involves assigning labels to detected objects based on their features or characteristics. Efforts have been made to implement object detection and classification based on multibeam imagery. For example in [17], a method that can detect and classify diving seabirds, single fish, and fish schools in a very intense backscatter environment near a power-generating tidal turbine was presented. This method uses a threshold and filtering to detect objects and provides tracking with a nearest-neighbour algorithm. Classification is achieved by extracting key metrics such as size, shape, backscatter intensity, and velocity, and then setting thresholds for these parameters. Another study employed a Gabor-based feature pyramid network to effectively detect mine-like objects in side-scan sonar images, demonstrating its potential for handling multi-scale object detection tasks [17]. Some studies have explored the potential of machine learning-based approaches for object detection in MBES data. For instance, a recent study [18] proposed a target detection and classification workflow that demonstrated improved efficiency and reliability in detecting and classifying objects in sonar images.

This study aims to investigate the potential of vision-based machine learning to detect targets in multibeam images. On this basis, we make the following contributions:

- We present a novel approach that leverages machine vision techniques, instead of conventional threshold-based detection to detect targets in MBES images.

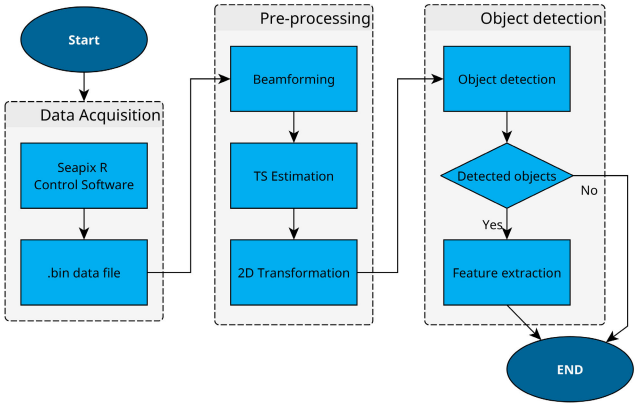

FIGURE 2. Proposed method.

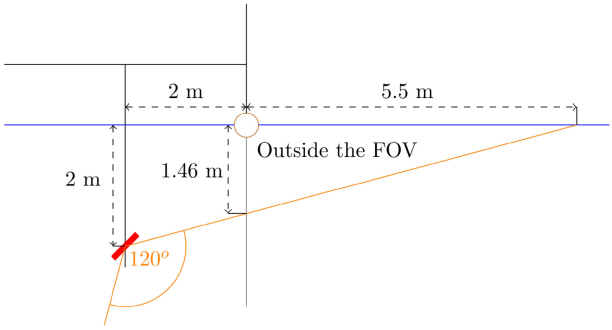

FIGURE 3. Cross section of the measurement setup. The MBES is drawn in red, and the field of view of the MBES is orange.

- We test the approach by identify targets imaged from a MBES.
- We demonstrate the precision, adaptability, and potential of our approach for the automatic monitoring of fish farms.
- We evaluate the robustness of our approach in noisy environments.

This paper is organised as follows: In Section II our approach is described, including our experimental setup, data acquisition and pre-processing. The results are presented in Section III, Section IV discusses the results and limitations of this approach, and Section V concludes this paper.

II. PROPOSED METHOD

The proposed method for object detection and classification of MBES data involves a three-stage process consisting of data acquisition, pre-processing, and object detection. Data was acquired from a MBES attached to the side of a fish farm, the specifications and operating parameters of which are detailed in Section II-C. During the data acquisition stage, system parameters such as the source level, range, and transmit beamforming angles were configured to acquire baseband data per transducer. The acquired data were saved and passed to the pre-processing stage. During this stage, the data were beamformed, target strengths estimated and a transformation applied from polar to Cartesian coordinates to transform 3D sonar data into 2D sonar images. These images were then sent to the object detection stage of the process. A summary of this process is visualised in Figure 2.

A training dataset must be created to train the object detection algorithms. This is accomplished by manually labelling some of the data that have been captured with different settings, at different ranges and orientations. Subsequently, various machine learning-based object detection algorithms were trained on the dataset to identify and locate targets in the scans.

A. MULTIBEAM ECHOSOUNDER

The MBES used in these trials was a Seapix-R [19]. The Seapix-R consists of a reversible steerable symmetric Mills cross containing 2×64 transceivers. It can steer acoustic beams at angles of $120° \times 120°$ and operates at a frequency range of 145-155 kHz.

The Seapix-R Mills cross consists of a vertical transmitting array which beamforms in a fan-like pattern in the vertical and a receiving array that beamforms received data in a horizontal fan. This process is made possible by Seapix-R's fully digital receiving front-end, which includes individual receiving chains for each transducer, thus enabling simultaneous beamforming across all angles during reception. The independence of the two arrays also allows data to be received at short distances without ringing caused by the transmitted pulse.

The beam aperture of the Seapix-R is $1.6^o \times 1.6^o$ aligned with the element axis. Seapix-R is connected to a Beam-Former Unit (BFU) that stores the data locally [20]. 4G cellular networking allows remote control and real-time data retrieval.

B. MEASUREMENT SETUP

During the experiment a MBES was placed outside the cage on a mounting bracket with a 45^o downward tilt, such that most of the cage was within the vertical beamforming capability of the system, with the exception of a wedge-shaped section near the net, 1.5 meters high, extending 5.5 meters into the cage. Since salmon generally avoid being near the water surface and net [4], we did not consider this limitation to be a significant issue. A mounting bracket was attached to the walkway of the farming cage and oriented such that the MBES was pointed perpendicular to the walkway. The MBES was mounted 2 m from the side net and submerged 2m below the water surface, as shown in Figure 3. The rotation of the MBES on the mounting bracket is such that one of the element arrays is horizontal, and the other lies in the vertical plane. With this setup, the cage was insonified from one side. Due to the extremely high biomass, the power-delay profile is expected to decrease severely as the distance from the MBES increases because of biomass scattering and absorption. Fish closer to the MBES are exptected to be clearly visible, while those at greater distances become increasingly obscured due to interference from multiple wave

TABLE 1. MBES Configuration parameters during data acquisition.

Parameter	Value	Unit
Transmission mode	CW	
Transmission frequency	150	kHz
Transducer drive voltage	600	[V]
Transmission duration	50	[μS]
Reception duration	25	[ms]
Idle duration	10	[ms]
Transmission swath start angle	10	deg
Transmission swath stop angle	-10	deg
Transmission swath step	1	deg

paths. Such an effect is a well-documented characteristic in environments with high biomass [21]. Consequently, only the portion of the cage nearest to the MBES is likely to be observable. The cage can be insonified by either transmitting with the horizontal or vertical array and receiving with the other. This means that for each swath, we can either obtain a bird's eye view or a cross-sectional view of the cage at the angles determined by the beamforming of the transmitted beam. Naturally, by combining the data from multiple swaths, the responses can be combined to obtain data from the other orientation, at the cost of the view not being captured simultaneously.

C. DATA ACQUISITION

The dataset used in this study was obtained during a single data collection event. The sampling was conducted at an offshore salmon farming site operated by Lingelaks in Bergadalen, Norway. The data capture started at 09:22 on April 15 2021, during daylight hours and outside the routine feeding schedule, ensuring minimal disturbances and deviations in behavioural patterns. During the sampling event, 15,300 swaths were collected over a 10-minute period. Each swath is the result of one beamformed transmission pulse sent by the vertical array and then sampled by the horizontal array. These data consist of samples from 64 different channels, each sampled at a rate of 36 kHz. The reception duration of a swath was 25ms, resulting in 865 samples from each element for each swath. Samples captured at distances over 10 m were not processed, due to the deterioration of the signal passing through the high biomass environment. Each sample consists of an in-phase and quadrature component, representing information spanning 4.2 cm in distance and within the beam aperture of $1.6^o \times 1.6^o$. The description of the configuration of the MBES is reported in Table 1. During the capture events, the vertical array was transmitting and the horizontal array was receiving, resulting in a bird's eye view snapshot.

D. BEAMFORMING AND TARGET STRENGTH ESTIMATION

To perform beamforming, a steering vector was applied to the transmitting and receiving signal. Steering vector v describes the phase delay applied to each transducer for a plane wave arriving from an angle of arrival (AOA) or departure the from an angle of departure (AOD). The steering vector can be derived from the following wave vector:

$$k(\theta, \phi) = [k_x, k_y, k_z] = \frac{2\pi}{\lambda}[\cos\phi \sin\theta, \sin\theta, \sin\phi, \cos\theta] \tag{1}$$

This describes the phase change rate in any given direction. The transducer array R is defined in terms of the transducer coordinates (x, y, z):

$$R = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \\ z_1 & z_2 & \dots & z_n \end{bmatrix} \tag{2}$$

where n denotes the number of transducers in the transducer array. After combining the wave vector (1) with the transducer coordinates (2), we obtain the phase for each transducer as follows:

$$\alpha_e(\theta, \phi) = kR = \begin{bmatrix} x_1k_x + y_1k_y + z_1k_z \\ x_2k_x + y_2k_y + z_2k_z \\ \vdots \\ x_nk_x + y_nk_y + z_nk_z \end{bmatrix} \tag{3}$$

This produces the steering vector of the array:

$$v(\theta, \phi) = \begin{bmatrix} e^{j\alpha_{e1}} \\ e^{j\alpha_{e2}} \\ \vdots \\ e^{j\alpha_{eN}} \end{bmatrix} \tag{4}$$

Beamforming is accomplished by applying the phase shift defined by the steering vector $v(\theta, \phi)$ (4) at the time of transmission and to the received signal after sampling. The received signal per transducer is denoted by $\mathbf{y}$, and captured by the horizontal transducer array, and the transmitted waveform is denoted by $\mathbf{x}$ and is sent with the vertical array. The steering vector for the receiving case is:

$$\mathbf{w}^T = v(\theta, 0)^H \tag{5}$$

and in the transmitting case:

$$\mathbf{f} = v(0, \phi) \tag{6}$$

The received echo level (EL) after beamforming the transmitted and received signals can be expressed as:

$$\text{EL} = \mathbf{w}^T\mathbf{H} \cdot \mathbf{f} \cdot \mathbf{x} + \mathbf{w}^T\mathbf{n} \tag{7}$$

where the pulse $\mathbf{x}$ to be transmitted is beamformed with the transmitting steering vector $\mathbf{f}$ and passes through the acoustic channel $\mathbf{H}$. The echo signal, which results from the interaction of the transmitted waveform with the surrounding environment, is shaped by the channel $\mathbf{H}$. This channel imposes the effects of the echo on the signal, characterising the multiple-input, multiple-output (MIMO) relationship between the transmitting and receiving transducers. The signal from the receiving transducer array then applies a steering vector $\mathbf{w}^T$ to the received signal, including potential noise $\mathbf{n}$ in the system.

The EL, which can be presented in terms of energy, can be determined using the sonar equation [22]. The beamformed data provide a voltage vs. delay for each angle θ, ϕ integrated over the sampling period. The received EL in logarithmic form can be expressed as:

$$\mathrm{EL} = \mathrm{SL} + 2\mathrm{AF}(\theta, \phi) + \mathrm{TS} - 2 \cdot 20 \log_{10}(r) - 2r\alpha \quad (8)$$

where SL is the source level, TS is the target strength, AF is the array factor, r is the distance to the target, and α is the absorption coefficient. The model is based on the assumption of spherical spreading from the source and target; thus, we employed the term $20 \cdot \log_{10}(r)$ [23]. This range correction aligns with the methodology employed in [14]. To obtain the target strength at each distance, we take the received signal Pa, translate it to V and compensate for the source level (SL) and two-way path loss by applying time-varied-gain (TVG) and adding a constant for the absorption coefficient.

$$\mathrm{TS} = \mathrm{EL} - \mathrm{SL} - 2\mathrm{AF}(\theta, \phi) + 40 \log_{10}(r) + 2r\alpha \quad (9)$$

This provides a value for target strength versus distance for each angle (θ, ϕ).

The methodology is demonstrated here in the context of salmon observation using a specific sonar system. Nevertheless, since our approach is based on fundamental underlying principles of the sonar equation and array processing it is broadly applicable to many other configurations and species.

E. ADAPTING 3D ACOUSTIC DATA FOR IMAGE RECOGNITION ALGORITHMS

To facilitate object detection and classification, multiple readily-accessible image recognition algorithms were employed for performance comparison. These algorithms have been primarily developed to work on images with multiple colour channels which are fundamentally different from the beamformed data target strength images resulting from our data. To improve the performance of image-based algorithms, we take the logarithm of the target strengths (TS). This transformation maps the TS into a space that better aligns with the implicit assumptions of optical computer vision algorithms, given their typical expectation of a linear relationship between sensor irradiance and the corresponding image intensity [24].

$$\mathrm{TS_{dB}} = 10 \log_{10}(\mathrm{TS}) \quad (10)$$

Once the target strengths (TS) are transformed into decibels (dB), the distribution of TS across different azimuth angles can be evaluated to get an insight into the frequency of backscatter strengths across the swath, as can be seen in Figure 4. Notably, the target strength distribution closely approximates a Gaussian distribution, as indicated by an R^2 value of 0.99981. This can be attributed to the combined reflections from different fish sizes, orientations and locations, coupled with multiple scattering events and system noise. Such aggregated independent measurements often produce a Gaussian-like distribution due to the Central Limit Theorem.

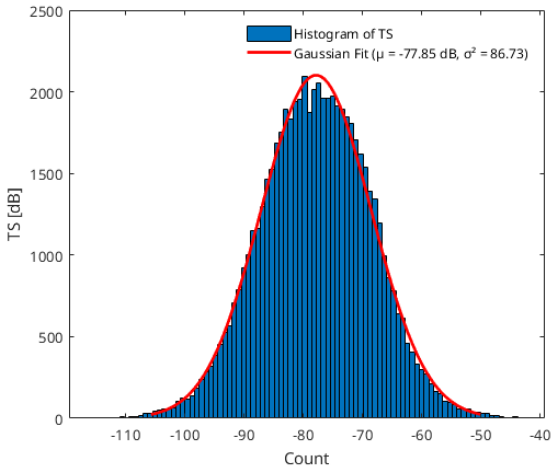

FIGURE 4. Histogram of target strengths (TS) from a single swath ($\theta = 0^o$) showing the distribution of backscatter strengths.

The raw data consists of multiple swaths resulting in a 3D representation of the area in front of the MBES. Due to the limitations of the image recognition algorithms, we need to flatten the data from 3D to 2D to analyse each swath separately. Thus, each image consists of delay vs. either θ or ϕ with TS as intensity. To maintain relative angles and distances across the image and thereby ensure coherent interpretation, the images were transformed from polar to Cartesian coordinates. This transformation was performed by constructing a polar grid with Equation11.

$$\begin{aligned} X &= r_{\mathrm{Norm}} \cdot \sin(\theta) \\ Y &= r_{\mathrm{Norm}} \cdot \cos(\theta) \end{aligned} \quad (11)$$

where r_{Norm} is the normalised radius, θ is the azimuth, and X and Y are the Cartesian coordinates.

This results in a 2D image, where the backscatter from targets near the MBES is near the origin (the vertex at the bottom) and distances between targets in angle are relative to the physical distance. An example of a resulting image is shown in Figure 5. Given the rapid speed of sound in water and the comparatively slow movement of fish, any fish movement within this short timeframe is minimal and doesn't significantly affect our analysis.

F. TRAINING DATA

To use the 2D images obtained from the pre-processing steps as training data for the object detection algorithm, each image was carefully analysed, and the positions of any clearly distinguishable targets were marked. The only significant source of backscatter in the measurements recorded during this trial was the fish inside the cage. This process was meticulous and labour-intensive, as great care had to be taken to ensure that the quality of the training data was good.

Of the 15,300 swaths measured, 300 (1.96%) were chosen at random times and angles to produce the training, validation

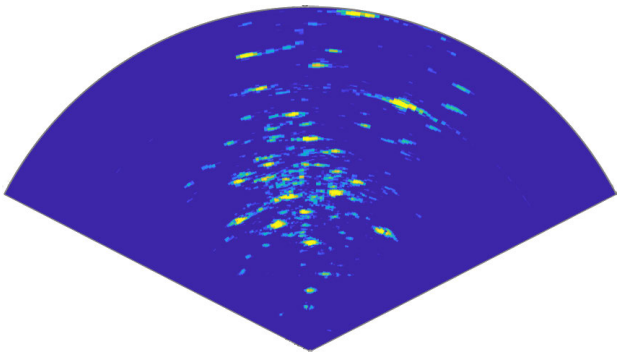

FIGURE 5. Target strength of a single swath ($\theta = 0^o$) from the MBES. Backscatter from fish at different angles and distances can be seen. The targets observed in these measurements had a TS of <-65dB.

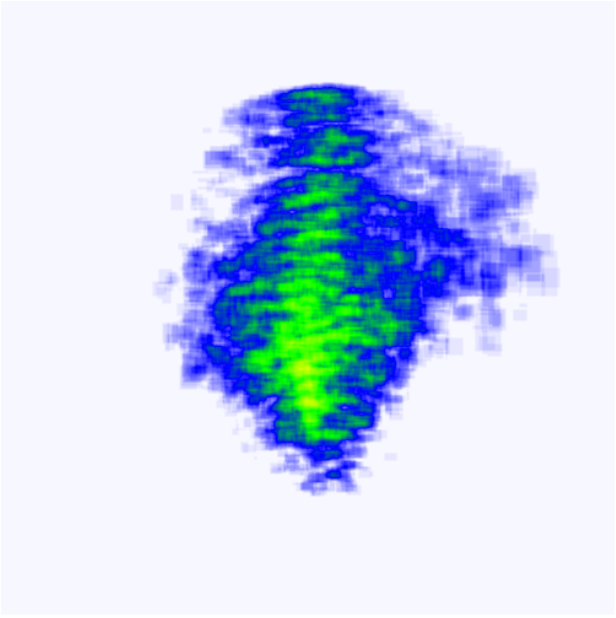

FIGURE 6. Spatial distribution of targets in training data.

and test data. The targets in these 300 swaths were carefully annotated, resulting in a total of 10,352 labelled objects. The spatial distribution of all 10,352 targets manually identified in the training data is illustrated in Figure 6.

Of the 300 labelled swaths, the data were then split into training (70%), validation(20%) and test data(10%). The training set was augmented by flipping the images horizontally to double the amount of training data to 600 swaths to increase the amount of training data. Some of the chosen object detection algorithms have limitations regarding the image resolution and maximum number of targets per image. Therefore to ensure compatibility, the data set was tiled in a 2 × 2 format, transforming one image into four smaller sub-images. This tiling approach not only reduces the resolution of each sub-image but also effectively distributes the objects present, so that the number of detections needed in each sub-image is decreased. The number of targets in each original image varied between 27 and 100 before tiling, and between 3-48 after tiling. The

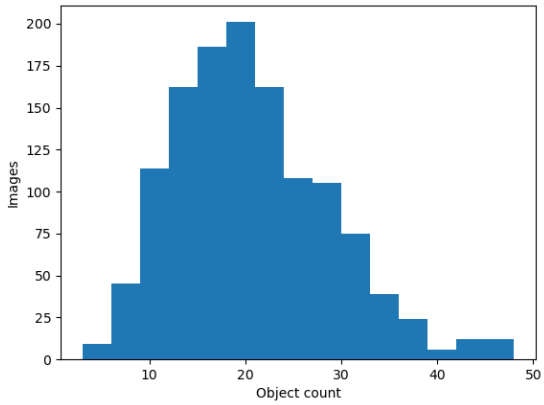

FIGURE 7. Distribution of number of targets in the training data before tiling.

frequency of the number of targets is illustrated in Figure 7. Following the tiling, we ended up with a total of 2,400 images. Then we manually removed 98 swaths where the targets were not distinctly discernible. Ultimately, 70% of the refined dataset, amounting to 1,582 swaths after augmentation, was used for training the algorithms.

G. TARGET DETECTION

Departing from conventional threshold-based detection methods, this study leveraged machine vision techniques to detect objects in the collected acoustic data. These techniques offer significant advantages in terms of adaptability and precision and provide an innovative approach for underwater object detection and analysis.

In this study, YOLOv5 [25] (v7.0, 2022), YOLOv6 [26] (v4.0, 2023), YOLOv8 [27](v8.0.0, 2023), and SSD [28] were employed as representative algorithms to demonstrate the effectiveness of the proposed framework for object detection using acoustic data. Each algorithm operates by setting bounding boxes to detect objects, classifying the object within the box and provides a confidence score, an example of which can be seen in Figure 8. These algorithms were selected to demonstrate the applicability and efficacy of machine vision techniques for this application. More sophisticated target detection algorithms can be considered to further enhance performance in future research.

The original YOLO algorithm, released in 2015 [29], paved the way for new single-stage object detectors, including YOLOv5, YOLOv6, and YOLOv8. The latter versions have introduced various enhancements to the original model. For instance, YOLOv8 applies a compound scaling method that simultaneously adjusts the network depth and width, employs a multi-scale feature fusion technique to combine features from different network layers, and incorporates a novel loss function [27]. A recent update of YOLOv6 introduced a new network design and training methodology that significantly improved its performance, achieving state-of-the-art accuracy [26]. In comparison to these advancements, YOLOv5 has remained a popular choice for similar detection

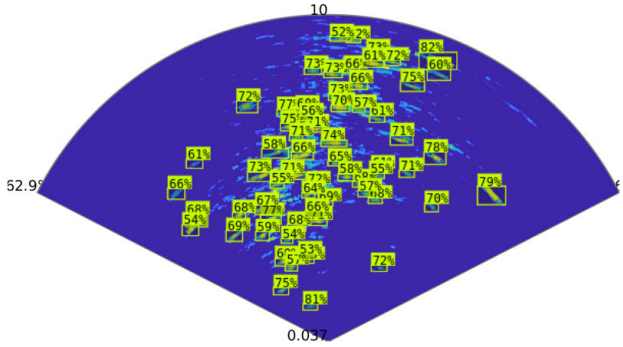

FIGURE 8. YOLOv8-L inference on a unseen swath.

tasks, owing to its robust performance [30], [31], including automatic single fish detection with data acquired from a single-beam sonar [32]. The YOLO models were trained with minimal tweaking of hyper-parameters and were quickly able to achieve satisfactory performance. Larger models of YOLO (X and L) were chosen because of the primary goal of testing the detection accuracy, with less importance placed on latency and throughput. Each model was trained for 10,000 epochs. However, the patience parameter was set to 600, which means that if there was no improvement after 600 consecutive epochs, the training process would terminate. The best-performing model from multiple training rounds was selected for further comparison.

On the other hand, SSD or "Single Shot MultiBox Detector" is a renowned object detection algorithm recognised for its balance between speed and accuracy. Similar to YOLO, SSD employs a deep neural network composed of convolutional layers of varying sizes to perform detections at multiple scales and aspect ratios. The SSD is easy to train and offers an effective trade-off between speed and accuracy. The SSD implementation used in this study was provided in [33]. The training and validation data were converted to the COCO format, and training was performed for 64 epochs. Hyper-parameters such as the learning rate, momentum, and weight decay were adjusted, and the model with the highest detection accuracy was selected for comparison.

III. RESULTS

This section presents the findings from our experimental assessment of the efficacy of the proposed method in object detection and classification within MBES swaths of fish farming cages. Additional tests were conducted to evaluate the detection accuracy and the impact of noise on the detection performance.

A. DETECTION ACCURACY

To evaluate the performance of the algorithms, we employed three metrics: Precision, Recall, and Mean Average Precision (mAP).

TABLE 2. Detection accuracy for the tested object detection methods.

Method	mAP_{50}	P	R
YoloV5 - x	0.713	0.748	0.702
YoloV6-L	0.773	0.824	0.72
YoloV8 - x	0.697	0.781	0.626
SSD300 - ResNet-50	0.283	0.136	0.151

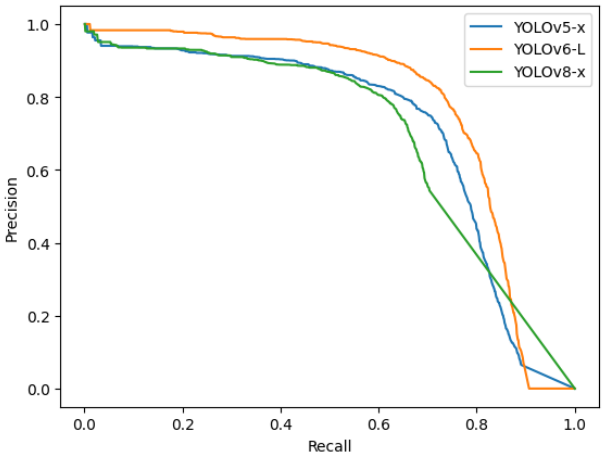

FIGURE 9. Precision recall curve of the different YOLO based models.

Precision is defined as the proportion of identified positives (i.e., detected objects) that are correct. A model with perfect precision would only produce correct detections, but it might miss many objects. Recall is defined as the proportion of true positives that are identified correctly. A model with perfect recall can detect every instance of an object in an image, but it may also produce many false positives. mAP_{50} (Mean Average Precision at 50% Intersection over Union) is a combined metric. A detected bounding box is considered correct if it overlaps $\geq$ 50% with the ground truth box. Essentially, mAP_{50} evaluates both the model's detection ability and the precision of its bounding box placement.

Table 2 presents a comparative analysis of the detection accuracy for the various object detection methods tested in our study. With the exception of SSD, the overall mAP_{50} achieved was generally above 70%, with small variations depending on the model used, demonstrating a good ability to detect objects in the MBES images.

The Precision-Recall curve (Figure 9) shows Precision against Recall. A model with perfect classification yields a curve that reaches the top-right corner of the plot, signifying both high Precision and Recall. Conversely, a model with poor classification performance produces a curve closer to the diagonal, indicating a random or arbitrary prediction. Among the methods evaluated in this study, the Yolo-based algorithms performed reasonably well. The models were able to correctly identify a high percentage of objects of interest while minimising the number of false detections.

While processing time is a crucial metric in many applications, our study primarily focussed on accuracy and feasibility. Previous literature, such as [26], [29], and [28], has

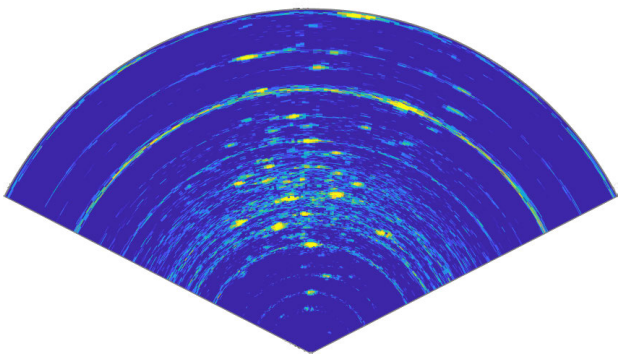

FIGURE 10. Beamformed image after adding noise. This illustrates the deterioration of the image as noise is added before beamforming. Notice the side-lobe effects on areas with high backscatter.

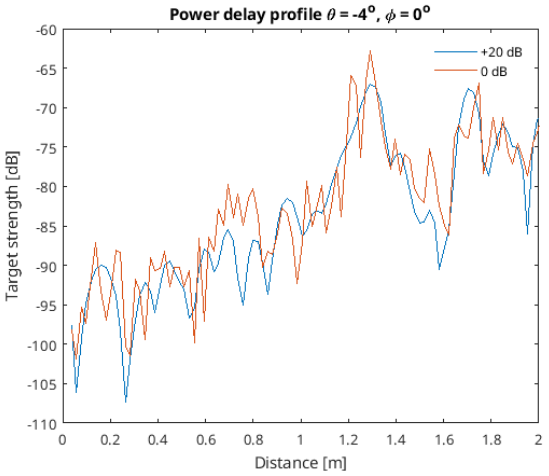

FIGURE 11. Power delay profile of the beamformed signal with two different noise levels applied. The figure illustrates that while features at high TS remain observable, lower intensity features become obscured at higher NL.

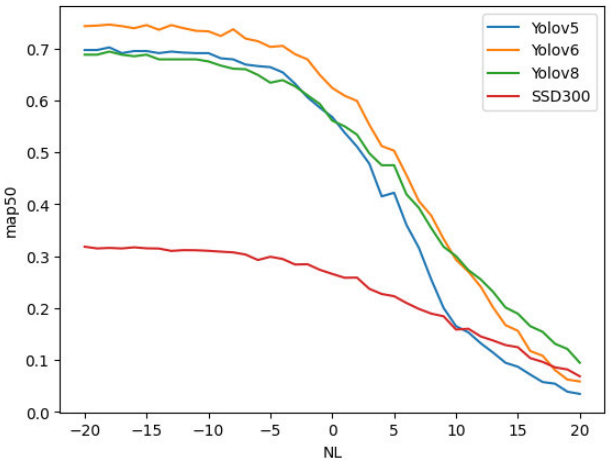

FIGURE 12. mAP_{50} at different artificially added noise levels.

extensively covered processing time comparisons for various models.

B. DETECTION ACCURACY IN THE PRESENCE OF NOISE

In real-world scenarios, image and object detection systems often encounter noisy environments which can affect their accuracy. For our experiment, set in real-world conditions, we wanted to explore the impact of noise on detection. To provide a consistent and controlled investigation of the robustness of object detection algorithms under varying signal-to-noise ratio (SNR) conditions, we introduced an artificial deterioration of the SNR using Normally-distributed random noise. This method is commonly employed in simulations to approximate the random noise typically found in real-world environments and is very similar to models that aim to model the spectrum of background noise underwater [34]. Further, as can been seen in Figure 4, this distribution also fits the TS distribution observed in our experiment.

To artificially deteriorate the signal-to-noise ratio before beamforming, we add Normally-distributed noise to the received signal. The expression for the noise-augmented received signal, derived from Equation 7, is:

$$\mathrm{y} = \mathrm{H} \cdot \mathrm{f} \cdot \mathrm{x} + \mathrm{n} + \mathcal{N}(0, \sigma) \tag{12}$$

We introduced a new parameter to set the noise level NL.

$$\sigma = 20 \log_{10}(\mathrm{NL}) \tag{13}$$

We used noise levels (NL) ranging from -20 dB to 20 dB. The noise was added as a vector n in equation 7: for each level of noise, Gaussian noise was randomly added to the input images, with the standard deviation of the noise determined by the NL parameter. The detection accuracy of the model was evaluated on noisy images using the mAP_{50}. Examples of the effects of added noise are shown in Figure 10 and Figure 11. The results of this evaluation are shown in Figure 12 which shows that the Detection Accuracy of the different models decreased as the NL increased, with a significant drop in performance at NL levels above 5 dB. At 0 dB, the model's mAP_{50} dropped by approximately 20% compared to its performance on noise-free images. At low NL, the model's Detection Accuracy reaches levels near to its noise-free performance at -20 dB NL. We also observe that the YOLO-based models perform significantly better than the SSD300 model at low NL, with a smaller difference at high NL.

Given that the Detection Accuracy remains unaffected until a substantial degree of noise is introduced, it is plausible to suggest that the acoustic emission power we used may be excessive for the employed range. Consequently, reducing the power level could not only contribute to a reduction in MBES power consumption but also potentially enhance fish welfare.

C. BOUNDING BOX ANALYSIS

In our results, we observed that the average size of the bounding boxes is approximately $0.11\mathrm{m}^2$, as can be seen in Figure 13. While bounding box dimensions generally correlate with fish size, with larger fish producing larger bounding boxes, it's important to understand this correlation

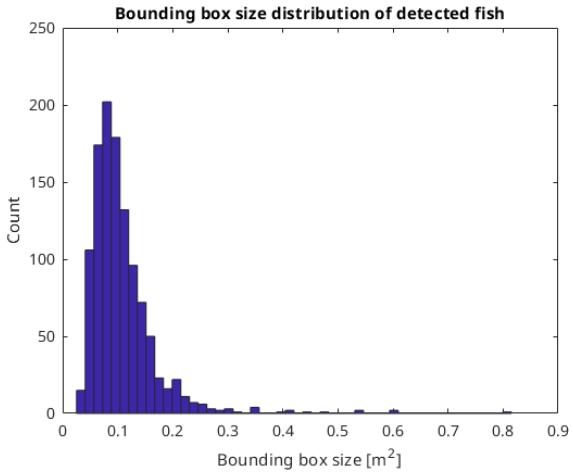

FIGURE 13. Bounding box size distribution of targets detected with YOLOv6.

is not direct. Equating bounding box dimensions to fish size directly would be misleading as the heading of the fish can significantly influence the size of the bounding box. For example, if the fish is swimming in the direction of the Cartesian axis of the image the bounding box will appear small, but if the fish is swimming diagonally the bounding box becomes larger. Furthermore, it may happen that not all parts of a fish are equally insonified in the sonar image. Situations may arise where only a portion of the fish is visible, leading to an underestimation of its size based on the bounding box dimensions. This issue is further complicated when considering that the beam angle of the MBES is only 1.6 deg, meaning that there may be many fish in each image that are only partially illuminated. While most bounding boxes corresponded to the anticipated size range, we did note the presence of some that were significantly larger. This indicates the potential occurrence of scenarios where multiple fish are near each other, resulting in overlap and a consequent larger bounding box. Overlapping object detection is a recognised challenge in this field [35]. However, given the relatively consistent shape of a fish, there are techniques available to fit known shapes to overlapping objects, aiding in their differentiation, such as [36]. Nevertheless, in our dataset, instances of overlapping fish were uncommon. Approximately 95% of the bounding boxes were smaller than 0.2m^2, suggesting they likely encapsulated a single fish.

IV. DISCUSSION

Our results demonstrate the potential of using machine-learning algorithms for object detection in images as an alternative to the traditional method of selecting a threshold. We demonstrate that a machine vision algorithm can exploit information from multiple delays and angles to provide accurate target detection. However, some disadvantages must be considered. For example, this method is currently limited to detecting objects within a specific range of sizes, and may not perform well in detecting objects that are significantly smaller or larger than the sizes present in the training data.

Exposing the algorithms to more training data, captured in more varied situations, may improve their performance, as it would allow them to better recognise and differentiate between different objects.

This work serves as an initial step in showcasing some of the possibilities that a MBES offers to fish farming, particularly by establishing the feasibility of employing machine vision for object detection in sonar images. While the current study provides insights into the feasibility of this method, we recognise the limitations in using vision-based algorithms, and the merit in developing a dedicated algorithm for this task. This is consequently a direction we are considering for future research. One promising avenue for future work is the exploration of temporal tracking of targets. This approach would involve analysing consecutive swaths to track the movement of individual fish over time, allowing us to account for fish movement during longer periods. Correlating the positions of fish over time provides a more comprehensive understanding of fish behaviour and movement patterns, addressing concerns related to temporal dynamics not captured in isolated swaths.

Another promising direction for future work is to explore the potential of 3D expansion to improve the detection performance. By incorporating additional azimuth information, we can potentially improve the ability of our method to detect objects from different viewpoints and under different lighting conditions.

Overall, although our proposed method demonstrates promising results for object detection in images, there is certainly still room for improvement. Future work should focus on exploring new approaches and techniques that can help address the current limitations and further improve the performance of our method.

V. CONCLUSION

In this study, we successfully trained different vision-based machine learning algorithms to detect targets based on images obtained from an MBES pointed at a fish farming cage.

Successful implementation of this method may have significant implications for fish farming. Reliable methods for fish detection can potentially improve the efficiency and reduce welfare risks of fish during fish farming operations. This can include optimising feeding schedules, analysing behaviour, and detecting fish health issues.

Although the present work focuses on detecting fish, this method of using vision-based object detection algorithms to detect targets in fish farms can be extended to detect and classify other objects present in and around fish farms, such as cages, mooring lines, and mooring plates allowing for live monitoring of the dynamics of fish farms.

In conclusion, our study demonstrates the feasibility of using machine learning algorithms to successfully detect targets from MBES images of fish farming cages. The incorporation of this technology has the potential to develop sustainability and welfare practices in aquaculture by introducing new observation methods. Moreover, it paves the

way for further research and advancements in precision aquaculture, contributing to the development of intelligent management systems that can revolutionise how we approach marine farming practices.

ACKNOWLEDGMENT

Waive (Formerly Aquabio) provided and set up the Seapix-R MBES.

REFERENCES

[1] *The State of World Fisheries and Aquaculture 2022, Towards Blue Transformation*, FAO, Rome, Italy, 2022.

[2] H. V. Bjelland, M. Føre, P. Lader, D. Kristiansen, I. M. Holmen, A. Fredheim, E. I. Grøtli, D. E. Fathi, F. Oppedal, I. B. Utne, and I. Schjølberg, "Exposed aquaculture in Norway," in *Proc. OCEANS*, Oct. 2015, pp. 1–10.

[3] P. McIntosh, L. T. Barrett, F. Warren-Myers, A. Coates, G. Macaulay, A. Szetey, N. Robinson, C. White, F. Samsing, F. Oppedal, O. Folkedal, P. Klebert, and T. Dempster, "Supersizing salmon farms in the coastal zone: A global analysis of changes in farm technology and location from 2005 to 2020," *Aquaculture*, vol. 553, May 2022, Art. no. 738046.

[4] Á. Johannesen, O. Patursson, J. Kristmundsson, S. P. Dam, M. Mulelid, and P. Klebert, "Waves and currents decrease the available space in a salmon cage," *PLoS ONE*, vol. 17, no. 2, 2022, Art. no. e0263850.

[5] S. C. Johnson, S. Bravo, K. Nagasawa, Z. Kabata, J. Hwang, J. Ho, and C. T. Shih, "A review of the impact of parasitic copepods on marine aquaculture," *Zoological Stud.*, vol. 43, no. 2, pp. 229–243, 2004.

[6] D. Johansson, K. Ruohonen, A. Kiessling, F. Oppedal, J.-E. Stiansen, M. Kelly, and J.-E. Juell, "Effect of environmental factors on swimming depth preferences of Atlantic salmon *(Salmo salar* L.) and temporal and spatial variations in oxygen levels in sea cages at a Fjord site," *Aquaculture*, vol. 254, nos. 1–4, pp. 594–605, Apr. 2006.

[7] M. Remen, M. Sievers, T. Torgersen, and F. Oppedal, "The oxygen threshold for maximal feed intake of Atlantic salmon post-smolts is highly temperature-dependent," *Aquaculture*, vol. 464, pp. 582–592, Nov. 2016.

[8] R. D. Hedger, A. H. Rikardsen, J. F. Strøm, D. A. Righton, E. B. Thorstad, and T. F. Næsje, "Diving behaviour of Atlantic salmon at sea: Effects of light regimes and temperature stratification," *Mar. Ecol. Prog. Ser.*, vol. 574, pp. 127–140, Jul. 2017.

[9] Á. Johannesen, Ø. Patursson, J. Kristmundsson, S. P. Dam, and P. Klebert, "How caged salmon respond to waves depends on time of day and currents," *PeerJ*, vol. 8, p. e9313, Jun. 2020.

[10] M. Føre, K. Frank, T. Norton, E. Svendsen, J. A. Alfredsen, T. Dempster, H. Eguiraun, W. Watson, A. Stahl, L. M. Sunde, C. Schellewald, K. R. Skøien, M. O. Alver, and D. Berckmans, "Precision fish farming: A new framework to improve production in aquaculture," *Biosystems Eng.*, vol. 173, pp. 176–193, Sep. 2018.

[11] T. A. Beddow and L. G. Ross, "Predicting biomass of Atlantic salmon from morphometric lateral measurements," *J. Fish Biol.*, vol. 49, no. 3, pp. 469–482, Sep. 1996.

[12] J. L. Stewart and E. C. Westerfield, "A theory of active sonar detection," *Proc. IRE*, vol. 47, no. 5, pp. 872–881, May 1959.

[13] K. G. Foote, "Linearity of fisheries acoustics, with addition theorems," *J. Acoust. Soc. Amer.*, vol. 73, no. 6, pp. 1932–1940, Jun. 1983.

[14] B. Tallon, P. Roux, G. Matte, J. Guillard, and S. E. Skipetrov, "Acoustic density estimation of dense fish shoals," *J. Acoust. Soc. Amer.*, vol. 148, no. 3, pp. 234–239, Sep. 2020.

[15] B. Tallon, P. Roux, G. Matte, and S. Skipetrov, "Mesocopic wave physics in a dense fish school," *J. Acoust. Soc. Amer.*, vol. 146, p. 3076, Oct. 2019.

[16] B. Tallon, P. Roux, G. Matte, J. Guillard, J. H. Page, and S. E. Skipetrov, "Ultra slow acoustic energy transport in dense fish aggregates," *Sci. Rep.*, vol. 11, no. 1, p. 17541, Sep. 2021.

[17] B. J. Williamson, S. Fraser, P. Blondel, P. S. Bell, J. J. Waggitt, and B. E. Scott, "Multisensor acoustic tracking of fish and seabird behavior around tidal turbine structures in Scotland," *IEEE J. Ocean. Eng.*, vol. 42, no. 4, pp. 948–965, Oct. 2017.

[18] A. Minelli, A. N. Tassetti, B. Hutton, G. N. Pezzuti Cozzolino, T. Jarvis, and G. Fabi, "Semi-automated data processing and semi-supervised machine learning for the detection and classification of water-column fish schools and gas seeps with a multibeam echosounder," *Sensors*, vol. 21, no. 9, p. 2999, Apr. 2021.

[19] G. Matte, D. Charlot, O. Lerda, T.-K. N'Guyen, Giovanini, M. Rioblanc, and F. Mosca, "SeapiX: An innovative multibeam multiswath echosounder for water column and seabed analysis," in *Proc. Hydro17 Conf.*, 2017, pp. 21–23.

[20] F. Mosca, G. Matte, O. Lerda, F. Naud, D. Charlot, M. Rioblanc, and C. Corbières, "Scientific potential of a new 3D multibeam echosounder in fisheries and ecosystem research," *Fisheries Res.*, vol. 178, pp. 130–141, Jun. 2016.

[21] I. Røttingen, "On the relation between echo intensity and fish density," *FiskDir. Skr. Havunders*, vol. 16, pp. 301–314, Jan. 1976.

[22] R. J. Urick, *Principles of Underwater Sound*, 3rd ed. Westport, CT, USA: Peninsula, 1983.

[23] J. M. Hovem, *Marine Acoustics: The Physics of Sound in Underwater Environments*. Los Altos, CA, USA: Peninsula, 2012.

[24] K. Ikeuchi, *Computer Vision: A Reference Guide*. Berlin, Germany: Springer, 2021.

[25] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, Y. Kwon, K. Michael, J. Fang, Z. Yifu, C. Wong, D. Montes, and Z. Wang, "ultralytics/yolov5: v7.0—YOLOv5 SOTA realtime instance segmentation," Ultralytics, Los Angeles, CA, USA, Tech. Rep., Nov. 2022, doi: 10.5281/zenodo.7347926.

[26] C. Li, L. Li, Y. Geng, H. Jiang, M. Cheng, B. Zhang, Z. Ke, X. Xu, and X. Chu, "YOLOv6 v3.0: A full-scale reloading," 2023, *arXiv:2301.05586*.

[27] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by ultralytics," Ultralytics, Los Angeles, CA, USA, Version 8.0.0, Jan. 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[28] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Computer Vision ECCV*. Berlin, Germany: Springer, 2016, pp. 21–37.

[29] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[30] M. Lamane, M. Tabaa, and A. Klilou, "Classification of targets detected by mmWave radar using YOLOv5," *Proc. Comput. Sci.*, vol. 203, pp. 426–431, Jan. 2022.

[31] 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.

[32] J. Tong, W. Wang, M. Xue, Z. Zhu, J. Han, and S. Tian, "Automatic single fish detection with a commercial echosounder using YOLO v5 and its application for echosounder calibration," *Frontiers Mar. Sci.*, vol. 10, Jun. 2023, Art. no. 1162064.

[33] NVIDIA. (2022). *Deeplearningexamples/SSD300 v1.1 for PyTorch*. [Online]. Available: https://github.com/NVIDIA/DeepLearningExamples/tree/master/PyTorch/Detection/SSD

[34] M. Bouvet and S. C. Schwartz, "Underwater noises: Statistical modeling, detection, and normalization," *J. Acoust. Soc. Amer.*, vol. 83, no. 3, pp. 1023–1033, Mar. 1988.

[35] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *Multimedia Tools Appl.*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023.

[36] S. Zafari, T. Eerola, J. Sampo, H. Kälviäinen, and H. Haario, "Segmentation of overlapping elliptical objects in silhouette images," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5942–5952, Dec. 2015.

JÓHANNUS KRISTMUNDSSON received the B.Sc. degree in electronic engineering and IT with a specialization in communication systems and the M.Sc. degree in engineering wireless communication from Aalborg University, Denmark, in 2016 and 2018, respectively, and the Ph.D. degree in computer science and ocean engineering from the University of the Faroe Islands, in collaboration with Waive, Norway. He was with Fiskaaling as a Researcher, where he was involved in processing and analyzing data used for evaluations of new fish farming sites, developing data processing software for signal processing from Acoustic Doppler Current Profilers (ADCPs), and CTD data analysis. His research interests include developing signal processing methods for multibeam sonars to be used in fish farming, beam alignment methods for terminals in millimeter-wave wireless networks, satellite communication, and channel analysis.

ØYSTEIN PATURSSON received the B.Sc. degree in petroleum engineering from the University of the Faroe Islands, in 2002, and the Ph.D. degree in ocean engineering from the University of New Hampshire, U.K., in 2008. He was a Senior Researcher, the Director, and the Head of Research with Fiskaaling, Faroe Islands. He is currently the Director and an Owner of the Research and Development Company ÍVF RAO, which focuses on engineering aspects applied to sea-based aquaculture and oceanography, such as the design of flexible structures in the ocean (e.g., fish farming cages), the effect of fish farming on the physical ocean environment, development of field measurement protocols. His research interests include the assessment of exposed fish farming locations and systems, measurements of exposed salmon farming in high currents and waves, and machine learning for underwater detection.

JOHN POTTER received the joint B.Sc. degree (Hons.) in mathematics and physics from the University of Bristol, U.K., in 1979, and the Ph.D. degree in glaciology and oceanography from the University of Cambridge, U.K., in 1985, on research conducted with the British Antarctic Survey. He is currently a Professor with the Norwegian University of Science and Technology (NTNU), Norway, working across departments and with external organizations to develop interdisciplinary innovative projects in the domain of ocean, space, and sustainable futures. He was formerly an Associate Professor, the Founder, and the Head of the Acoustic Research Laboratory, NUS Tropical Marine Science Institute, Singapore, where he moved after several years as a Researcher with SIO and UCSD, USA. He was a Principal Strategic Development Officer, a Principle Scientist, and the Project Leader of NATO STO CMRE, Italy, where he led a high-performing team of skilled scientists and engineers in designing and testing maritime communications technology, creating baseline critical technologies for an Internet of Underwater Things (IoUT). His main research interests include acoustic environmental monitoring, communication, and physical oceanography. It is no longer true that he neither owns nor operates a television.

QIN XIN received the Ph.D. degree from the Department of Computer Science, University of Liverpool, U.K., in December 2004. He is currently a Professor of computer science with the Faculty of Science and Technology, University of the Faroe Islands (UoFI), Faroe Islands. Prior to joining UoFI, he had held various research positions in world-leading universities and research laboratories, including a Senior Research Fellowship with Université Catholique de Louvain, Belgium, a Research Scientist/Postdoctoral Research Fellow with the Simula Research Laboratory, Norway, and a Postdoctoral Research Fellow with the University of Bergen, Norway. Moreover, he also investigates combinatorial optimization problems with applications in bioinformatics, data mining, and space research. He is serving on the Management Committee Board of Denmark for several EU ICT projects. He has produced more than 150 peer-reviewed scientific papers. His works have been published in leading international conferences and journals, such as ICALP, ACM PODC, SWAT, IEEE MASS, ISAAC, SIROCCO, IEEE ICC, *Algorithmica*, *Theoretical Computer Science*, *Distributed Computing*, IEEE TRANSACTIONS ON COMPUTERS, *Journal of Parallel and Distributed Computing*, IEEE TRANSACTIONS ON DIELECTRICS AND ELECTRICAL INSULATION, IEEE TRANSACTIONS ON SUSTAINABLE COMPUTING, and *Advances in Space Research*. His main research interests include the design and analysis of sequential, parallel, and distributed algorithms for various communication, optimization problems in wireless communication networks, and cryptography and digital currencies, including quantum money. He has been very actively involved in the services for the community in terms of acting (or acted) in various positions, such as the Session Chair, a member of the Technical Program Committee, a Symposium Organizer, and the Local Organization Co-Chair, for numerous international leading conferences in the fields of distributed computing, wireless communications, and ubiquitous intelligence and computing, including IEEE MASS, IEEE LCN, ACM SAC, IEEE ICC, IEEE Globecom, IEEE WCNC, IEEE VTC, IFIP NPC, and IEEE Sarnoff. He is the Organizing Committee Chair for the 17th and 18th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2020 and SWAT 2022, Tórshavn, Faroe Islands). He also serves on the editorial board for more than ten international journals.