Computer Vision for Bioacoustics: Detection of Bearded Seal Vocalizations in the Chukchi Shelf Using YOLOV5

Christian Escobar-Amado, Mohsen Badiey, and Lin Wan

***Abstract*—Year-round recordings of bearded seal calls were collected in the northeastern edge of the Chukchi Continental Slope (Alaska, within the Arctic Circle) in 2016–2017, 2018–2019, and 2019–2020. While the underwater vocalizations of bearded seals are often analyzed manually or using automatic detections manually validated, in this article, a detection and classification system (DCS) based on the You Only Look Once Version 5 (YOLOV5) algorithm is proposed. With YOLOV5, the network learns how to detect and classify these marine mammals' calls using the principle of computer vision for object detection in images where bounding boxes enclose the objects of interest. During training, validation, and testing, YOLOV5 achieved an accuracy of 96.54%, 93.36%, and 93.87%, respectively. The DCS was applied to the three-year-long dataset, and an analysis of the vocal behavior of the bearded seals showed that there exists a geographical dependence where this species prefers shallower water depths in the Chukchi Continental Slope. Another advantage of using YOLOV5 over other typical DCS is that the predicted bounding boxes have embedded statistical information about the vocalization, such as the duration, bandwidth, and center frequency of the signals. This additional information equips biologists with statistical data that facilitate the analysis of animal vocal behavior.**

***Index Terms*—Arctic, bearded seals, computer vision, deep learning, marine mammals, You Only Look Once Version 5 (YOLOV5).**

I. Introduction

THE Arctic Ocean is rapidly changing due to the climate change that is making a big impact on the sea ice conditions [1], [2], [3] with several ecological and economical implications. For example, new commercial opportunities are emerging, such as the potential for opening new trans-Arctic shipping routes [3], [4], which can affect the marine life. On the other hand, the sea ice decline is affecting the marine mammal distributions [5]. With all of these factors, a renewed interest in Arctic acoustics inspired the Canada Basin Acoustic Propagation Experiment (CANAPE) [6]. The CANAPE was focused on investigating the spatial and temporal variability and coherence of the acoustic environment and propagation in the Arctic Ocean over one year from October 2016 until October 2017. The experiment was separated into two areas: 1) deep water in the Canada Basin and 2) shallow water (SW) in the Chukchi shelf. In the subsequent years, the Ocean Acoustics and Engineering Laboratory (OAELAB) at the University of Delaware (UD) deployed acoustic and environmental sensors in the same SW area in September 2018–September 2019 and November 2019–September 2020. Positions of the acoustic arrays in the SW site are shown in Fig. 1.

During the data analysis of the Shallow Water Canada Basin Acoustic Propagation Experiment (SW-CANAPE), it was found that several marine mammal's vocalizations were recorded. Bearded seals, in particular, are one of the main contributors to the marine soundscape in the Arctic [8], and their calls were consistently found in the recordings during the months when the ice was present. In the past few years, extensive studies of these marine mammal's vocalizations have been possible thanks to the large acoustic datasets collected using passive acoustic monitoring [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19]. These data allow researchers to analyze the spatial and temporal behavior of marine mammal's vocalizations, which are indicatives of their presence, population, and distribution.

Commonly, for bearded seals, the calls are manually identified by trained analysts by visualizing the spectrograms of the audio recordings. However, for large datasets, this laborious task is not feasible, and automatic detection systems become necessary [19]. Recently, detection and classification systems (DCSs)—commonly based on spectrograms—have been implemented using deep learning for identifying several species of marine mammals, mainly for whale calls [20], [21], [22], [23], [24], [25], [26].

In a previous work, we have proposed an automatic DCS of bearded seal vocalizations in the Arctic Ocean [19]. This DCS was a two-step process where regions of interest (ROIs) were first detected by a spectrogram correlation method, and then, convolutional neural networks (CNNs) were in charge of classifying the ROIs among several different classes. This method was applied for detecting and classifying two well-known types of vocalizations from the bearded seal's repertoire using two representative masks. Even though the algorithm proved to have

Manuscript received 16 September 2022; revised 5 June 2023; accepted 13 August 2023. Date of publication 25 September 2023; date of current version 8 February 2024. This work was supported by the Office of Naval Research Ocean Acoustics Program (ONR OA322) under Grant N00014-15-1-2110, Grant N00014-18-1-2140, and Grant N00014-21-1-2760. *(Corresponding author: Christian Escobar-Amado.)*

Associate Editor: J. Potter.

The authors are with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716 USA (e-mail: escobarc@udel.edu; badiey@udel.edu; wan@udel.edu).

Digital Object Identifier 10.1109/JOE.2023.3307175

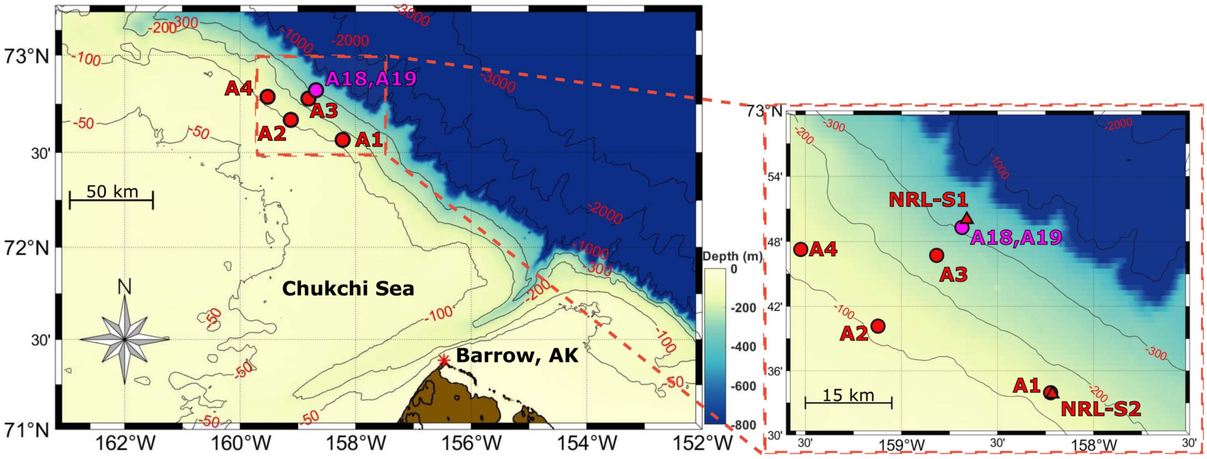


Fig. 1. (a) Wide area map showing the locations of AMAR1 (A1), AMAR2 (A2), AMAR3 (A3), and AMAR4 (A4) deployed in 2016–2017, and AMAR18 (A18) and AMAR19 (A19) deployed in 2018–2019 and 2019–2020, respectively, in the Chukchi continental slope. (b) Close-up of the locations of the receivers relative to the sound sources. Maps are overlaid on IBCAO v3.0 bathymetric data [7].

a precision above 89.2% in the generalization stage, its main limitation is that the signals of interest are stereotypical, and it is necessary to manually define the masks based on prior knowledge. Now, we propose a methodology that uses only deep learning techniques without the need for knowing specifics of the signal, meaning that no ROIs or masks are required.

In recent years, deep learning algorithms, particularly CNNs, have been utilized for vocalization event detection in large audio datasets [19], [27], [28], [29]. However, in this study, we have taken a different approach by using a state-of-the-art object detection system algorithm called “You Only Look Once (YOLO)” [30], [31], [32], [33] in its fifth version, YOLOV5 [34]. While YOLO, a computer vision method, was originally designed for object detection in images using bounding boxes, we adapt it for our specific task of detecting and classifying vocalizations by inputting spectrogram representations of audio files instead of images.

One significant advantage of the YOLO approach, in comparison to previous deep learning algorithms used for vocalization event detection on bioacoustic datasets, is its ability to predict bounding boxes of variable sizes around the sounds of interest. This means that our system can not only detect and classify vocalizations but also provide valuable insights into the frequency, bandwidth, and duration of the signals of interest. This additional information equips biologists with statistical data that facilitate the analysis of vocal behavior in animals.

The data used for this work were recorded in the northeastern edge of the Chukchi Continental Slope in 2016–2017 during the SW-CANAPE. Details about the position of the recorders and whereabouts of the experiment can be found in [6], [35], and [36]. The data collected by the OAELAB at the UD in the two subsequent deployments in 2018–2020 are also used for testing the algorithm.

II. Methods

The DCS proposed for detecting and classifying bearded seal vocalizations is inspired by the computer vision algorithms for object detection on images. Audio recordings are treated as images by converting them into their frequency versus time representation—also known as spectrograms—because of their 2-D nature. In this case, the YOLOV5 [34] algorithm is applied to the spectrograms to enclose bearded seal calls and classify them among several classes. The general workflow of the implemented methodology can be divided into three stages. First, for the input data, a dataset is generated by extracting several spectrograms, where the bearded seal vocalizations are labeled by manually placing a bounding box around the sounds of interest. For all the labeled bounding boxes (see Fig. 4), a K-mean clustering technique is used for selecting several representative boxes named “anchor boxes,” which will act as priors to simplify the problem by making it easier for the networks to learn [31]. Then, a 76%/12%/12% split is used for training, validating, and testing the YOLOV5 algorithm. Finally, in the inference stage, nonlabeled spectrograms are passed to the trained YOLOV5 model to generate bounding boxes enclosing the bearded seal vocalizations of interest.

A. Input Data

The data used in this article were collected in the North of Alaska on the northeastern edge of the Chukchi Shelf. During the SW-CANAPE in 2016–2017, the Defense Research and Development Canada agency deployed four Autonomous Multichannel Acoustic Recorder (AMAR) arrays at positions 72.566 °N 158.223 °W, 72.669 °N 159.122 °W, 72.779 °N 158.817 °W, and 72.788 °N 159.524 °W labeled as AMAR1, AMAR2, AMAR3, and AMAR4, respectively. One year later, in 2018–2019, an AMAR array (AMAR18) was deployed by the UD at 72.817 °N 158.703 °W close to the AMAR3 position (∼5.6 km away). In continuation to this recordings, in 2019–2020, the UD deployed the AMAR19 at 72.822 °N 158.685 °W.

On the other hand, during the SW-CANAPE, the Naval Research Laboratory (NRL) deployed two sound sources: NRL-S1 and NRL-S2. The anthropogenic noise from these sources possesses a challenge for the DCS since it is present in most of the

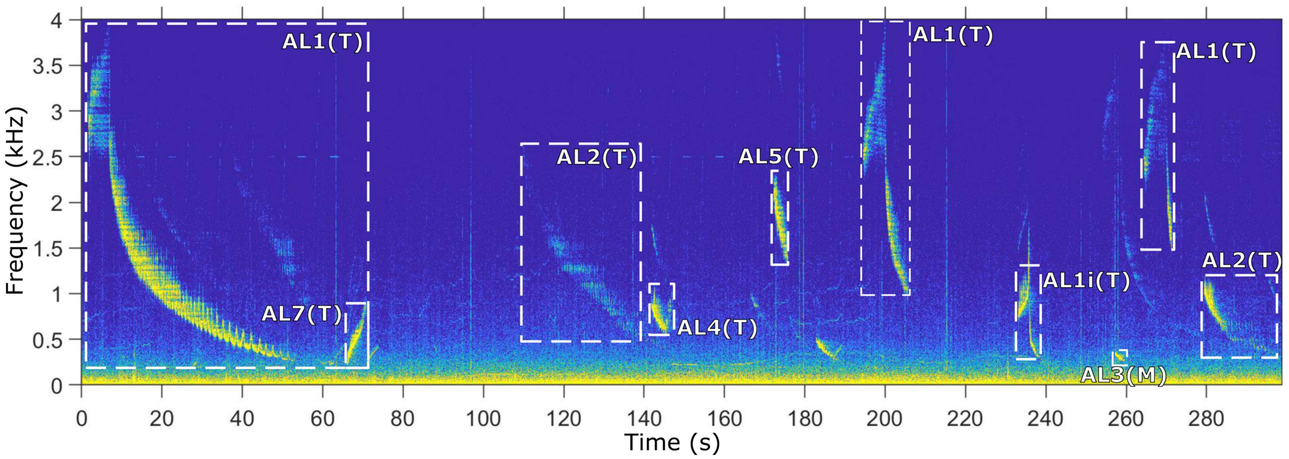


Fig. 2. Representative examples of bearded seal vocalization repertoire. Data measured on AMAR4 on April 25, 2017, from 20:08 until 20:13.

TABLE I
BEARDED SEAL VOCALIZATION CLASSES USED FOR TRAINING THE YOLOV5 ALGORITHM

Class	Vocalization	Description
1	AL1(T)	Vocalization with ascend and descend components in the higher frequency, AL1(T)
2	AL1i(T)-H	AL1i(T) with center frequency higher than 940 Hz
3	AL1i(T)-L	AL1i(T) with center frequency lower than 940 Hz
4	AL2(T) & AL5i(T)	Long downsweeps. Both have similar behavior but different duration.
5	AL7(A)	Ascend plume
6	Head of AL1(T)	Head of AL1(T) vocalization
7	Head of AL1i(T)-H	Head of AL1i(T) with center frequency higher than 1200 Hz
8	Head of AL1i(T)-L	Head of AL1i(T) with center frequency lower than 1200 Hz

2016–2017 recordings in the 1.5–4 kHz frequency band, where the bearded seals generate several of their calls. Details about the signals transmitted by NRL-S1 and NRL-S2 are explained in more detail in [19]. Locations of the recorders and sound sources overlaid on IBCAO v3.0 bathymetric data are shown in Fig. 1.

In total, ~ 8824 h were recorded on the six receivers. AMAR1, AMAR2, AMAR3, and AMAR4 were on 3.58 h per day for ~ 356 days. AMAR18 was on 4.8 daily hours for ~ 371 days, and AMAR19 was on 6.47 daily hours for ~ 300 days. These data are represented in the form of 86.4 s-long spectrograms (in dB re 1 μPa^2/Hz) generated in the 0–4 kHz frequency band with an 80 ms time spacing and a 3.9 Hz frequency step. The whole dataset is normalized to have values in between 0 and 255 to be quantized to 8-bit integers. Each data sample, i.e., each spectrogram, is reshaped to 640 × 640, which is a typical size of images for object detection. To train the YOLOV5 algorithm, several spectrograms recorded at AMAR4 and AMAR1 are labeled by manually placing bounding boxes around bearded seal vocalizations. AMAR4 and AMAR1 have been chosen because they are the farthest and closest receivers to the anthropogenic sound sources, respectively, providing a rich variety of background noise (BN) signals.

1) Data Labeling: The vocal repertoire of bearded seals from Alaska (AL) can be categorized into trills (T), moans (M), and ascents (A), as explained in [9], [10], and [11]. An example spectrogram of these calls is shown in Fig. 2, where most of the vocalizations were found in a 5 min frame on April 25, 2017 on AMAR4. For this work, the data have been labeled among the eight classes shown in Table I. Short vocalizations corresponding to AL3(M), AL4(T), and AL6(T) were omitted for this study case due to the time-consuming task of labeling all of them. From the bounding boxes (which represent each different call), some statistics of the labeled signals can be obtained; for example, the center frequency is computed as the center of the bounding box height, while the bandwidth and duration correspond to the height and width of the Bbox, respectively. The class with the longer duration, bandwidth, and center frequency corresponds to AL2(T), which can be as long as 80 s, approximately.

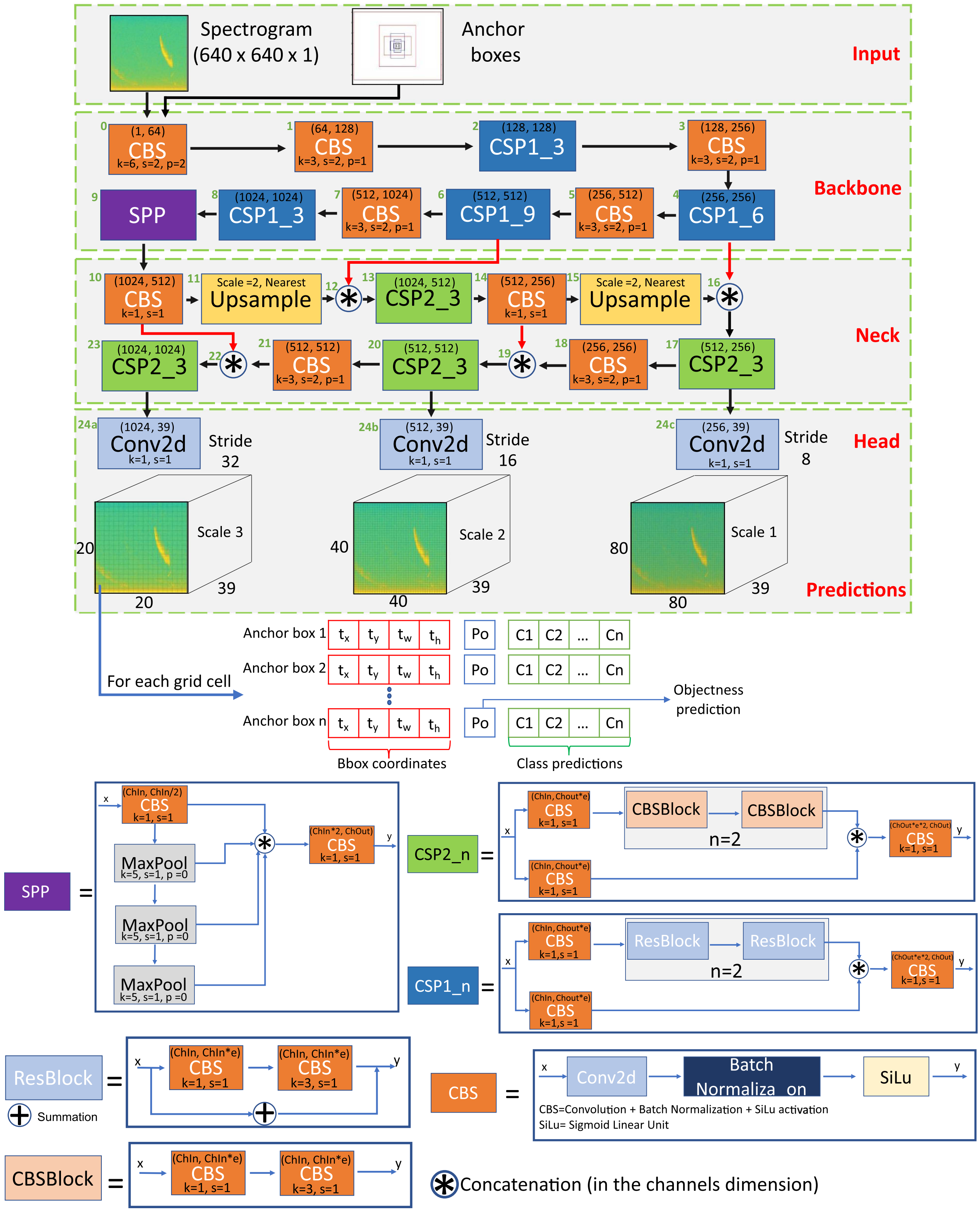

Spectrogram
(640 x 640 x 1)
Anchor
boxes
Input
0 (1, 64) CBS k=6, s=2, p=2
1 (64, 128) CBS k=3, s=2, p=1
2 (128, 128) CSP1_3
3 (128, 256) CBS k=3, s=2, p=1
Backbone
9 SPP
8 (1024, 1024) CSP1_3
7 (512, 1024) CBS k=3, s=2, p=1
6 (512, 512) CSP1_9
5 (256, 512) CBS k=3, s=2, p=1
4 (256, 256) CSP1_6
10 (1024, 512) CBS k=1, s=1
11 Scale =2, Nearest Upsample
12
13 (1024, 512) CSP2_3
14 (512, 256) CBS k=1, s=1
15 Scale =2, Nearest Upsample
16
Neck
23 (1024, 1024) CSP2_3
22
21 (512, 512) CBS k=3, s=2, p=1
20 (512, 512) CSP2_3
19
18 (256, 256) CBS k=3, s=2, p=1
17 (512, 256) CSP2_3
24a (1024, 39) Conv2d k=1, s=1
Stride 32
24b (512, 39) Conv2d k=1, s=1
Stride 16
24c (256, 39) Conv2d k=1, s=1
Stride 8
Head
Scale 3
20
20
39
Scale 2
40
40
39
Scale 1
80
80
39
Predictions
For each grid cell
Anchor box 1
Anchor box 2
Anchor box n
tx ty tw th
Po
C1 C2 ... Cn
Objectness prediction
Bbox coordinates
Class predictions
SPP =
x
(ChIn, ChIn/2) CBS k=1, s=1
MaxPool k=5, s=1, p =0
MaxPool k=5, s=1, p =0
MaxPool k=5, s=1, p =0
(ChIn*2, ChOut) CBS k=1, s=1
y
CSP2_n =
x
(ChIn, Chout*e) CBS k=1, s=1
CBSBlock
CBSBlock
n=2
(ChIn, Chout*e) CBS k=1, s=1
(ChOut*e*2, ChOut) CBS k=1, s=1
y
CSP1_n =
x
(ChIn, Chout*e) CBS k=1,s =1
ResBlock
ResBlock
n=2
(ChIn, Chout*e) CBS k=1,s =1
(ChOut*e*2, ChOut) CBS k=1,s =1
y
ResBlock =
x
(ChIn, ChIn*e) CBS k=1, s=1
(ChIn, ChIn*e) CBS k=3, s=1
y
Summation
CBS =
x
Conv2d
Batch Normaliza on
SiLu
y
CBS=Convolution + Batch Normalization + SiLu activation
SiLu= Sigmoid Linear Unit
CBSBlock =
x
(ChIn, ChIn*e) CBS k=1, s=1
(ChIn, ChIn*e) CBS k=3, s=1
y
Concatenation (in the channels dimension)

Fig. 3. YOLOV5 architecture. Numbers in parenthesis on top of each block represent the (input and output) number of channels. The kernel size (k), stride (s), and padding (p) are shown in the lower part of the CBS and Conv2d blocks. The green numbers in the upper left corner represent the layer number of their corresponding block. Asterisk blocks represent a concatenation in the channels dimension.

Since the AL1i(T) vocalizations—in the manually labeled spectrograms—are clustered at distinct center frequencies separated at 940 Hz, this class has been divided into AL1i(T)-H and AL1i(T)-L. Similarly, the head of AL1i1(T) can be clustered into two classes separated at 1200 Hz.

2) Anchor Boxes: In object detection, one approach to simplify the target localization is to predict bounding boxes based on priors. This way, the networks do not have to predict a specific location or dimension of the bounding box; instead, the networks predict offsets with respect to the priors, which are known as anchor boxes. However, these anchor boxes need to be selected such that they represent the entire dataset. In the most recent YOLO algorithms [32], three different scales of outputs are predicted at the end of the model. Each scale is in charge of predicting bounding boxes for small-, medium-, and large-size objects, and in this case, they have grid sizes of 80 × 80, 40 × 40, and 20 × 20, respectively. Therefore, a set of three anchor boxes is proposed for each of the three scales for a total of nine priors. For the small scale, the anchor boxes have pixel sizes of 10 × 13, 16 × 30, and 33 × 23. For the medium scale, the anchor boxes have sizes of 30 × 61, 62 × 45, and 59 × 119, and for the large scale, the anchor boxes have sizes of 116 × 90, 156 × 198, and 373 × 326.

B. YOLOV5

The YOLO algorithm is a computer vision technique proposed for one-stage object detection, which performs a one-pass regression of target localization (bounding box position) and classification. In YOLOV3 [32], three different scales are predicted to provide bounding boxes of different sizes. In this case, for each scale, we predict three bounding boxes (corresponding to the three anchor boxes per scale) per grid cell, as shown in the prediction stage in Fig. 3. For each cell, the network will output the vector: $[(t_x, t_y, t_w, t_h), (P_o), (C_1, C_2, \ldots, C_8)]$ for each of the three bounding boxes. Therefore, at each scale ($S = 80$, $S = 40$, and $S = 20$), the network predicts a tensor of size $S \times S \times [3 * (4 + 1 + 8)]$ for the four bounding box offsets (t_x, t_y, t_w, t_h), one objectness prediction (P_o), and eight class predictions $(C_1, C_2, \ldots, C_8)$. The output of the YOLO algorithm corresponds to the offset and scaling of the bounding boxes, where the set of coordinates (t_x, t_y, t_w, t_h) can be converted to (x, y, w, h) as $x = \sigma(t_x) * 2 - 0.5$, $y = \sigma(t_y) * 2 - 0.5$, $w = P_w * (\sigma(t_w) * 2)^2$, and $h = P_h * (\sigma(t_h) * 2)^2$, where $\sigma(.)$ is a sigmoid function and P_h and W_h correspond to the height and the weight, respectively, of the corresponding anchor box.

1) YOLOV5 Architecture: The YOLOV5 architecture is divided in three main sections: the backbone, the neck, and the head of the network, as shown in Fig. 3. The backbone is in charge of feature extraction, and it is composed of cross-stage partial network (CSPNet) [37] blocks to form the CSPDarkNet53 [33]. The backbone and the neck are connected through a spatial pyramid pooling (SPP) [38] block to improve the receptive field of the network. For getting the semantic representation of extracted features, the neck is a path aggregation network [39] with CSP2 blocks [40], [41]. Finally, the head has the same structure as YOLOV3 [32] to predict bounding boxes at three different scales.

In the architecture shown in Fig. 3, the CBS (Convolution + Batch Normalization + SiLu activation function [42]) block is the smallest and most used unit throughout the network. The structure of the CBS is shown in Fig. 3, where x and y represent the input and output of the unit, respectively.

The residual CSP1_n blocks are used in the backbone stage as the main component of the CSPDarknet53 for improving the feature extraction. Inspired on the CSPNet, the CSP1_n divides the input in two branches that are concatenated at the end, as shown in Fig. 3. The first part is composed of a CBS block followed by n Resblocks (Residual network blocks) to extract deeper features. The second part is a CBS block of the input. The concatenated branches are then passed to another CBS module.

To connect the backbone and the neck of the network, an SPP block is implemented, as shown in Fig. 3. In the SPP block, first, a CBS unit is applied to the input followed by three stacked maximum pooling operations. The output of these four components is then concatenated and passed to a CBS block. With this procedure, the receptive field of the network is significantly enhanced.

The CSP2_n block follows the same principle of CSP1_n, but CBS blocks are replacing the Resblocks to help enhancing and speeding up the flow of feature information [40] in the neck stage.

In the prediction stage, the bounding boxes with objectness prediction higher than a given threshold are filtered by the nonmaximum suppression postprocessing algorithm to provide the final predictions. These object detections contain the location of the selected bounding boxes along with a confidence value computed as

$$C_{\text{Pred}} = Pr(Object) * Pr(Class) \tag{1}$$

where $Pr(Object)$ is the probability that an object exists in the bounding box, and $Pr(Class)$ is the probability of the class with the maximum score.

2) Cost Function: For bounding box regression, the intersection over union (IoU) is used for measuring the similarity between the ground truth bounding box (B^{gt}) and the predicted bounding box (B^P), and it is computed as

$$\text{IoU} = \frac{B^{\text{gt}} \cap B^P}{B^{\text{gt}} \cup B^P}. \tag{2}$$

The issue with this method is that the IoU only considers the overlap of the bounding boxes. To also penalize the aspect ratio and center point distance between the bounding boxes, the complete intersection over union (CIoU) is used instead. This measure was defined in [43] as

$$\text{CIoU} = \text{IoU} - \frac{\rho^2(C^{gt}, C^P)}{c^2} - \alpha\nu \tag{3}$$

where C^{gt} and C^P correspond to the center points of B^{gt} and B^P, respectively, as shown in Fig. 4, $\rho(.)$ is the Euclidean distance, c is the diagonal of B^s, α is the positive tradeoff parameter defined in (4), and ν measures the consistency of aspect ratio

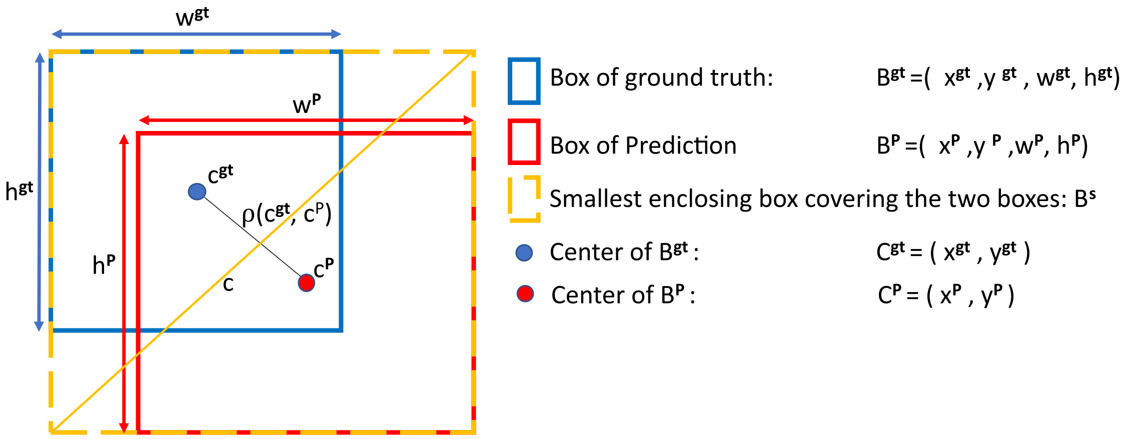

w^gt
w^P
h^gt
h^P
c^gt
ρ(c^gt, c^P)
c
c^P
Box of ground truth: B^gt =(x^gt ,y^gt , w^gt, h^gt)
Box of Prediction B^P =(x^P ,y^P ,w^P, h^P)
Smallest enclosing box covering the two boxes: B^s
Center of B^gt : C^gt = (x^gt , y^gt)
Center of B^P : C^P = (x^P , y^P)

Fig. 4. Dimensions of ground truth and predicted bounding boxes.

defined as (5)

$$\alpha = \frac{\nu}{(1-\mathrm{IoU})+\nu} \tag{4}$$

$$\nu = \frac{4}{\pi^3}\left(\arctan\left(\frac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}}\right) - \arctan\left(\frac{w^P}{h^P}\right)\right)^2. \tag{5}$$

The loss function of the YOLO algorithm is composed of three error components to penalize the main aspects of the predictions

$$L = L_{\mathrm{CIoU}} + L_{\mathrm{obj}} + L_{\mathrm{class}}. \tag{6}$$

Here, the CIoU loss (or box loss) is defined as

$$L_{\mathrm{CIoU}} = \sum_{k=0}^{O_s}\sum_{i=0}^{S_k^2}\sum_{j=0}^{B} I_{i,j,k}^{\mathrm{obj}}\left[1-\mathrm{CIoU}_{i,j,k}\right] \tag{7}$$

where O_s denotes the number of output scales ($O_s = 3$), S_k^2 denotes the number of cells in the grid of the kth scale ($S_1 = 80$, $S_2 = 40$, and $S_3 = 20$), B denotes the number of bounding boxes in each cell ($B = 3$), and $\mathrm{CIoU}_{i,j,k}$ represents the CIoU of the jth bounding box in the ith grid cell of the kth output scale. $I_{i,j,k}^{\mathrm{obj}}$ is 1 when there is an object in the Bbox and 0, otherwise.

The object loss is computed using the binary cross-entropy loss with a sigmoid layer as

$$L_{\mathrm{obj}} = -\sum_{k=0}^{O_s}\sum_{i=0}^{S_k^2}\sum_{j=0}^{B} C_{i,j,k}\log\left(\sigma\left(\hat{C}_{i,j,k}\right)\right) + \left(1-C_{i,j,k}\right)\log\left(1-\sigma\left(\hat{C}_{i,j,k}\right)\right). \tag{8}$$

Here, $\hat{C}_{i,j,k}$ is the predicted confidence score, and $C_{i,j,k}$ is the true confidence of the prediction determined by the CIoU as

$$C_{i,j,k} = I_{i,j,k}^{\mathrm{obj}} * \mathrm{CIoU}_{i,j,k}. \tag{9}$$

Finally, the classification error is also computed with the binary cross-entropy loss as

$$L_{\mathrm{class}} = -\sum_{k=0}^{O_s}\sum_{i=0}^{S_k^2}\sum_{j=0}^{B} I_{i,j,k}^{\mathrm{obj}} \sum_{c\epsilon \mathrm{classes}} P_{i,j,k}(c)\log\left(\sigma\left(\hat{P}_{i,j,k}(c)\right)\right) + \left(1-P_{i,j,k}(c)\right)\log\left(1-\sigma\left(\hat{P}_{i,j,k}(c)\right)\right) \tag{10}$$

where $P_{i,j,k}(c)$ is the true probability of class c and $\hat{P}_{i,j,k}(c)$ is the predicted probability score.

III. Results

The metrics used for evaluating the performance of the networks during the training/validation stage are accuracy, precision, recall, and mean average precision (mAP). Accuracy is calculated by counting the number of times the CNN predicted the correct class. Precision is the ratio between correctly predicted observations for a given class and the total of predicted observations for that class, and it is defined as

$$\mathrm{Precision} = \frac{\#\mathrm{True\ positives}}{\#\mathrm{True\ positives} + \#\mathrm{False\ positives}}. \tag{11}$$

Recall is the ratio between correctly predicted observations and the total number of observations of a given class, and it is defined as

$$\mathrm{Recall} = \frac{\#\mathrm{True\ positives}}{\#\mathrm{True\ positives} + \#\mathrm{False\ negatives}}. \tag{12}$$

The mAP metric is computed based on the IoU between the predicted and true bounding boxes [see (2)]. The mAP is the area under the curve of the precision versus recall curve. mAP@0.5 means that a prediction is considered correct if the IoU is greater than 0.5. mAP@0.5: 0.95 is the mean mAP computed for IoU thresholds in between 0.5 and 0.95, i.e., mean of mAP@0.5, mAP@0.55,..., mAP@0.95.

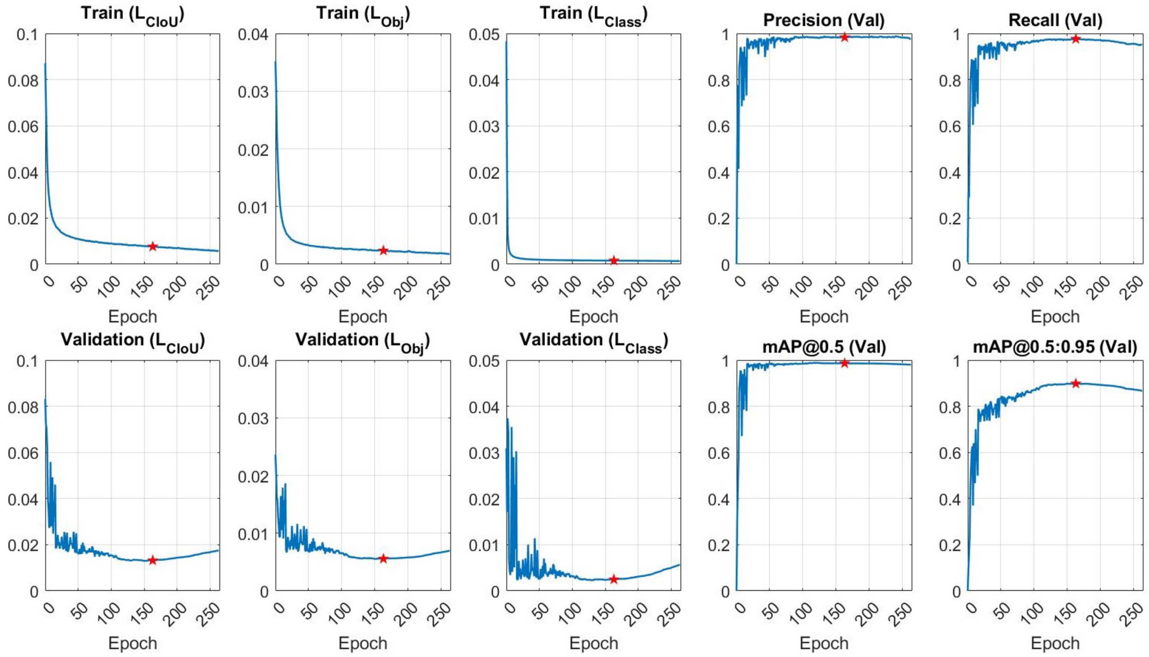
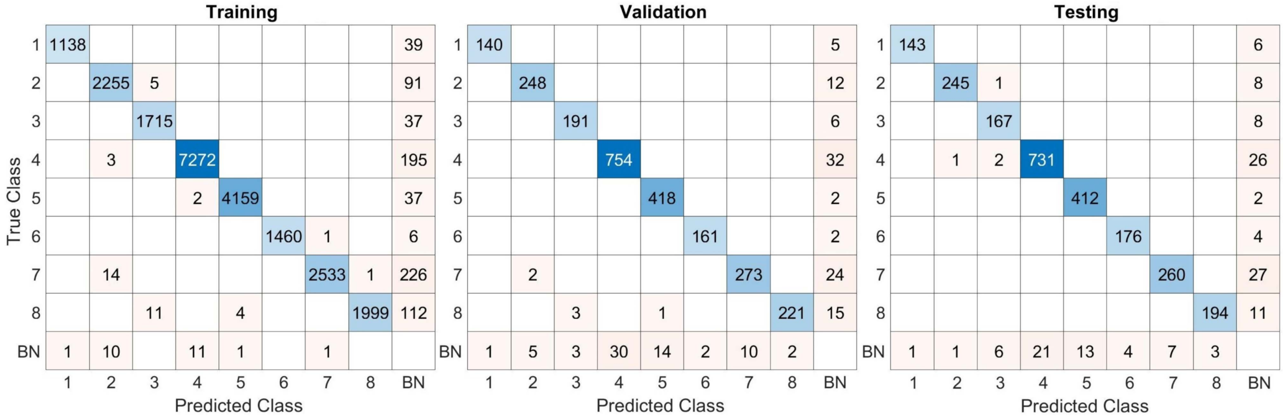


Fig. 5. Training and validation errors and metrics during training per epoch. Red star marker represents the best epoch, i.e., point at which the network achieved the lowest validation error.


True Class
Predicted Class

Training

	1	2	3	4	5	6	7	8	BN
1	1138								39
2		2255	5						91
3			1715						37
4		3		7272					195
5				2	4159				37
6						1460	1		6
7		14					2533	1	226
8			11		4			1999	112
BN	1	10		11	1		1		

Validation

	1	2	3	4	5	6	7	8	BN
1	140								5
2		248							12
3			191						6
4				754					32
5					418				2
6						161			2
7		2					273		24
8			3		1			221	15
BN	1	5	3	30	14	2	10	2	

Testing

	1	2	3	4	5	6	7	8	BN
1	143								6
2		245	1						8
3			167						8
4		1	2	731					26
5					412				2
6						176			4
7							260		27
8								194	11
BN	1	1	6	21	13	4	7	3	

Fig. 6. Training, validation, and testing confusion matrices for eight classes and BN.

A. *Training, Validation, and Testing results*

The YOLOV5 algorithm described in the previous section was trained on 6400 labeled spectrograms with 500 epochs. An early stopping technique with a patience of 100 epochs was used. The error metrics during training and validation are shown in Fig. 5. The CIoU, object, and class losses are shown individually for training and validation. The best network performance was obtained at epoch number 163, point at which the network started to overfit, as shown in the validation plots where the error started to increase (see red star marker in Fig. 5). The best metrics at this point were precision = 0.9832, recall = 0.9762, mAP@0.5 = 0.9831, and mAP@0.5: 0.95 = 0.8982.

The confusion matrices with the predicted bounding boxes versus the true bounding boxes for the training and validation stages are shown in Fig. 6. During training, the network achieved a precision and recall above 98.8% and 91.3%, respectively, for all the classes. When validating, all the classes achieved a precision and recall above 96.2% and 91.3%, respectively. These validation results show that the YOLOV5 algorithm has learned representative patterns to simultaneously detect and classify the eight classes of bearded seal vocalizations using the spectrogram representation of these acoustic signals.

Furthermore, the trained network was applied to several spectrograms that were not used during training to test how well the algorithm generalizes on new data samples. The confusion matrix with these predictions is shown in Fig. 6, where an accuracy of 93.87% is achieved. For all the classes, the precision and recall were above 94.9% and 90.6%, respectively. This

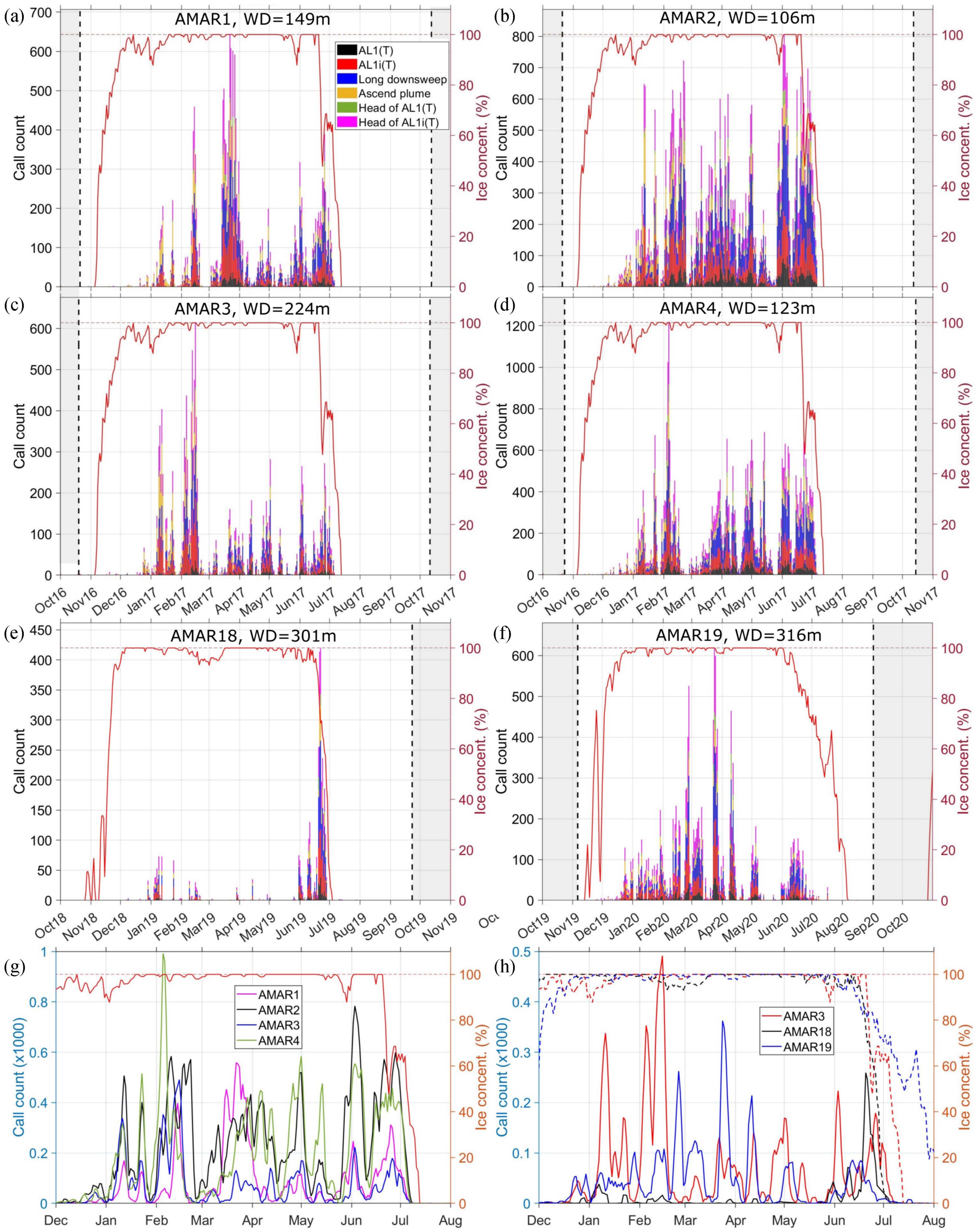


Fig. 7. Number of bearded seal vocalizations recorded by (a)–(d) DRDC AMARs deployed in the SW-CANAPE in 2016–2017, (e) AMAR18 in 2018–2019, and (f) AMAR19 in 2019–2020. Red lines indicate the sea ice concentration. Shadow areas correspond to the days where no recordings are available. (g) Envelop of the bearded seal vocalizations using a median moving window of three days for the four DRDC AMARs deployed in the SW-CANAPE in 2016–2017. Ice concentration is depicted by the red line. (h) Envelop of the bearded seal vocalizations using a median moving window of three days for AMAR3, AMAR18, and AMAR19. Ice concentration is depicted by the dashed line with the same color as the detections.

indicates that the YOLOV5 algorithm is generalizing well on data not seen by the network. The main advantage with respect to other DCSs is that, with YOLOV5, flexible bounding boxes are automatically placed around each detection, which allows researchers to analyze the statistics of the predicted signals in large datasets.

B. Bearded Seal's Vocal Activity and Geographical/Temporal Variations

Now that we have shown that the network is performing well, we apply the algorithm to the full dataset recorded on the six arrays. The YOLOV5 algorithm was trained on 86.4 s-long

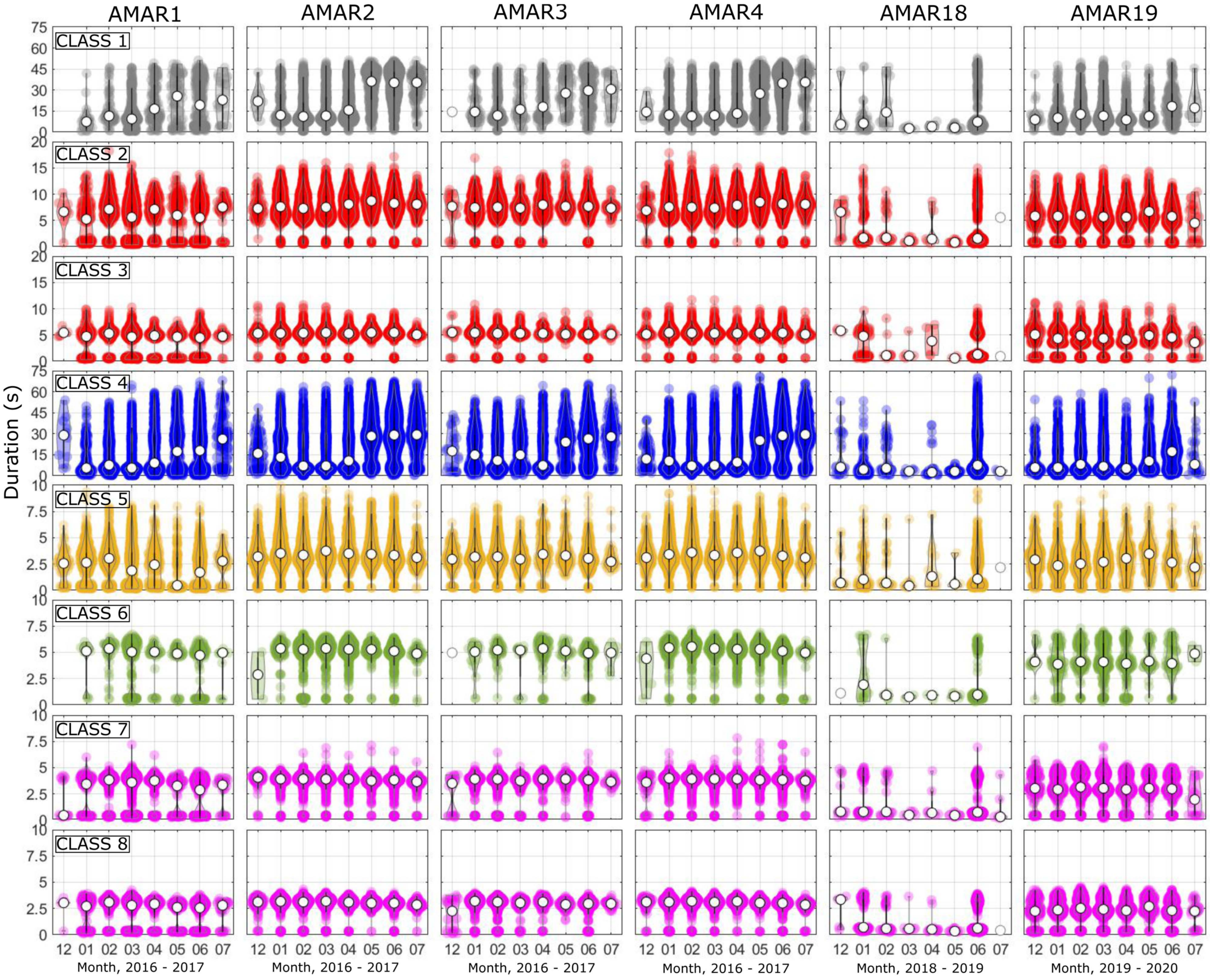


Fig. 8. Violin plots with inserted box plots for the time duration (width of the Bbox) in seconds for the eight classes of vocalizations detected by the YOLOV5 algorithm in the six recorders.

spectrograms (or 1080 samples in the time domain); however, the recordings are much longer than that. To solve this issue, a moving window of 8 s (or 100 samples) steps is applied. With this method, there is an overlap of 92.6% in the input spectrograms, which allows the network to detect the bearded seal calls from different parts of the spectrogram. This will yield several predictions for the same signal when the long spectrograms are passed to the network. At the end, using nonmax suppression, only the bounding boxes with the highest confidence are considered.

With the YOLOV5 algorithm, eight classes of vocalizations from bearded seals were detected and classified at six recorders, where the center of the arrays was located approximately 175 km to the northwest of Barrow, AK at the northeastern edge of the Chukchi Shelf. To assess the spatial dependence of the bearded seals in the Chukchi continental slope, the detections for the four DRDC AMARs deployed in the SW-CANAPE in 2016–2017 are used. For evaluating the temporal fluctuations of their vocal behavior, we use the arrays that were located at similar positions during the three years, i.e., AMAR3 and the recorders from the two subsequent deployments, AMAR18, and AMAR19.

To analyze the spatial and temporal dependence of the bearded seal's vocal behavior across the Chukchi continental slope, a summary of the detections is shown in Fig. 7(a)–(f) for the eight types of vocalizations of interest shown in Table I. The classes that were separated in the frequency components for training are now combined again, i.e., classes 1 and 2 and classes 6 and 7 are merged. Each bin in the histograms represents the daily count of the sounds of interest. Given that the vocal presence of bearded seals is directly correlated with the formation of pack ice, the sea ice concentration is shown on top of the detections. Ice data were collected from the European Organization for the Exploitation of Meteorological Satellites Ocean and Sea Ice Satellite Application Facility and were reported daily with a 10 × 10 km spacing resolution [44]. Daily mean sea ice concentration values were extracted for an area that covers all the arrays.

At all the stations, the bearded seals were highly vocally active in late June coinciding with detections reported by Frouin-Mouy et al. [9] and Hannay et al. [16] for the Chukchi Sea in 2007–2010. This timing corresponds to the breeding season [13], [45] when the males vocalize louder and longer trills [9], [13], [16], which explains the increase in the count of long downsweep calls represented by the blue bars in Fig. 7. The envelop of the detections for the DRDC AMARs shown in Fig. 7(g) shows the spatial dependence of the bearded seals in the Chukchi continental slope, where the recorders located at deeper water depths (WDs) exhibit lower vocalization densities. One possible reason for this is that bearded seals prefer depths below 100 m on the continental slope [11], [46]. The larger vocalization count occurs at AMAR2 (WD = 106 m) followed by AMAR4 (WD = 123 m), AMAR1 (WD = 149 m), and AMAR3 (WD = 224 m), which have a distance to Barrow, AK of 169.3, 187.7, 144.5, and 174.5 km, respectively.

To analyze the temporal variation of bearded seal vocalizations, the detections in 2016–2017 (AMAR3, WD = 224 m), 2018–2019 (AMAR18, WD = 301 m), and 2019–2020 (AMAR19, WD = 316 m) are presented in Fig. 7(c), (e), and (f), respectively. These arrays were located at deeper WDs and exhibit lower vocal activity. The envelop of the detections for the three-year-long recordings is shown in Fig. 7(h). In general, very few vocalizations are detected in 2018–2019 by AMAR18. However, at the end of June, the number of calls detected in AMAR18 was higher than in the other years. In addition, the periods of time where there is a larger count of vocalizations are consistent for AMAR3 (2016–2017) and AMAR19 (2019–2020), and they also match the small peaks of AMAR18 (2018–2019). Furthermore, it can be observed that the vocal behavior of the bearded seals is strongly related with the ice concentration, which is represented by the dashed lines of the same color as the detections. For the three years of recordings, as soon as the ice concentration starts decreasing, the number of calls increases and goes to zero right before the ice is completely melted. When the ice is not present, no vocal activity is recorded in the receivers.

In a previous study conducted by Jones et al. [11] on acoustic data recorded in 2006–2009 at the Chukchi shelf break, 120 km northwest of Barrow, AK, it was hypothesized that the recorder deployed at a WD of 240 m may have been located on the edge of bearded seal habitat due to the low-intensity levels of the vocalizations and the few detected calls. The combined observations at different years and locations, presented in this article, support the hypothesis that the edge of the bearded seal habitat is possibly located somewhere between 100 and 400 m isobath in the Chukchi continental slope.

Another advantage of using the YOLOV5 algorithm is that the predicted bounding boxes have embedded statistical information about the vocalizations. The width of the box represents the duration of the signals, the height corresponds to the bandwidth, and the center of the bounding box indicates the center frequency. The availability of this additional information equips biologists with statistical data that greatly facilitate the analysis of vocal behavior in animals. As an example, violin plots with inserted box plots for the duration of all the signals are shown in Fig. 8. Each panel contains the statistical distribution of every class for each of the six receivers.

When comparing the duration of the vocalizations throughout the year, AL1(T) calls (class 1) and long downsweeps (class 4) have an increase in duration in May, June, and July when the ice starts breaking. This has been observed in previous works, where Frouin-Mouy et al. [9] hypothesized that AL1(T) and AL2(T) calls are used to advertise their breeding condition and can be used as an indicator of the mating period based on their fluctuations in duration. Van Parijs et al. [47] also suggested that the trill duration may be a useful indicator of male “quality.” This type of behavior is observed in the data where only the signals containing long downsweeps present strong fluctuations not only in duration but also in bandwidth. The rest of the classes have a similar statistical distributions throughout the year. Bearded seal vocalizations recorded from 2018–2019 in AMAR18 exhibit a short duration behavior, which is an indicator that mostly roaming males were present in the area during that period [47].

IV. Conclusion

In this article, we showed the potential of computer vision for detecting and classifying marine mammal’s vocalizations recorded in large databases. We showed that by using the principles of object detection in computer vision, we can find acoustic signals of interest by treating the spectrogram representation of the sound as an image. For this purpose, we used the object detection algorithm YOLOV5 [34], where we detected and classified eight different types of bearded seal vocalizations without the need for using handpicked features, such as representative masks, frequencies, or contours. With this method, as long as we have enough labeled data of well-known stereotypical vocalizations, the YOLOV5 algorithm will be able to learn representative features about the signals of interest. It is important to mention that in the deep learning stage, the training data must have enough information content for the networks to learn how to distinguish the signals of interest from other possible signals that might be present in the testing scenarios.

Another advantage of using YOLOV5 over other typical DCS is that we not only detect and classify the signals of interest but also extract statistical information of the sound. This way, researchers can quickly analyze the vocal behavior of marine mammals in the ocean without the tedious task of visually assessing the spectrograms of large acoustic datasets.

Furthermore, an analysis of the spatial and temporal dependence of the bearded seal vocalizations showed that this species seems to prefer shallower WDs in the Chukchi Continental Slope. In addition, it was found that one common factor across all the receivers was the increase of vocalizations in late June during the breeding season when the ice is breaking and the trills are louder and longer, possibly, for fitness display purposes [13], [48].

References

[1] J. Stroeve, M. M. Holland, W. Meier, T. Scambos, and M. Serreze, “Arctic sea ice decline: Faster than forecast,” *Geophys. Res. Lett.*, vol. 34, no. 9, pp. 1–5, 2007.

[2] R. Kwok, “Arctic sea ice thickness, volume, and multiyear ice coverage: Losses and coupled variability (1958–2018),” *Environ. Res. Lett.*, vol. 13, no. 10, 2018, Art. no. 105005.

[3] D. Llaveria, J. F. Munoz-Martin, C. Herbert, M. Pablos, H. Park, and A. Camps, “Sea ice concentration and sea ice extent mapping with l-band microwave radiometry and GNSS-R data from the FFSCAT mission using neural networks,” *Remote Sens.*, vol. 13, no. 6, 2021, Art. no. 1139.

[4] L. C. Smith and S. R. Stephenson, “New trans-arctic shipping routes navigable by midcentury,” *Proc. Nat. Acad. Sci. United States Amer.*, vol. 110, no. 13, pp. 6–10, 2013.

[5] R. M. Mattmüller, K. Thomisch, I. Van Opzeeland, K. L. Laidre, and M. Simon, “Passive acoustic monitoring reveals year-round marine mammal community composition off Tasiilaq, Southeast Greenland,” *J. Acoust. Soc. Amer.*, vol. 151, no. 2, pp. 1380–1392, 2022.

[6] M. D. Collins, A. Turgut, R. Menis, and J. A. Schindall, “Acoustic recordings and modeling under seasonally varying sea ice,” *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 8323.

[7] M. Jakobsson et al., “The International Bathymetric Chart of the Arctic Ocean (IBCAO) version 3.0,” *Geophys. Res. Lett.*, vol. 39, no. 12, pp. 1–6, 2012.

[8] A. F. Heimrich, W. D. Halliday, H. Frouin-Mouy, M. K. Pine, F. Juanes, and S. J. Insley, “Vocalizations of bearded seals (Erignathus barbatus) and their influence on the soundscape of the western Canadian Arctic,” *Mar. Mammal Sci.*, vol. 37, no. 1, pp. 173–192, 2020.

[9] H. Frouin-Mouy, X. Mouy, B. Martin, and D. Hannay, “Underwater acoustic behavior of bearded seals (Erignathus barbatus) in the northeastern Chukchi Sea, 2007–2010,” *Mar. Mammal Sci.*, vol. 32, no. 1, pp. 141–160, 2016.

[10] D. Risch et al., “Vocalizations of male bearded seals, Erignathus barbatus: Classification and geographical variation,” *Animal Behav.*, vol. 73, no. 5, pp. 747–762, 2007.

[11] J. M. Jones et al., “Ringed, bearded, and ribbon seal vocalizations north of Barrow, Alaska: Seasonal presence and relationship with sea ice,” *Arctic*, vol. 67, no. 2, pp. 203–222, 2014.

[12] I. Parisi et al., “Underwater vocal complexity of arctic seal Erignathus barbatus in Kongsfjorden (Svalbard),” *J. Acoust. Soc. Amer.*, vol. 142, no. 5, pp. 3104–3115, 2017.

[13] H. J. Cleator, I. Stirling, and T. G. Smith, “ Underwater vocalizations of the bearded seal (Erignathus barbatus),” *Can. J. Zool.*, vol. 67, no. 8, pp. 1900–1910, 1989.

[14] K. Q. MacIntyre, K. M. Stafford, C. L. Berchok, and P. L. Boveng, “Year-round acoustic detection of bearded seals (Erignathus barbatus) in the Beaufort Sea relative to changing environmental conditions, 2008–2010,” *Polar Biol.*, vol. 36, no. 8, pp. 1161–1173, 2013.

[15] T. K. Boye, M. J. Simon, K. L. Laidre, F. Rigét, and K. M. Stafford, “Seasonal detections of bearded seal (Erignathus barbatus) vocalizations in Baffin Bay and Davis Strait in relation to sea ice concentration,” *Polar Biol.*, vol. 43, no. 10, pp. 1493–1502, 2020.

[16] D. E. Hannay et al., “Marine mammal acoustic detections in the northeastern Chukchi Sea, September 2007-July 2011,” *Continental Shelf Res.*, vol. 67, pp. 127–146, 2013.

[17] W. D. Halliday, S. J. Insley, T. de Jong, and X. Mouy, “Seasonal patterns in acoustic detections of marine mammals near Sachs Harbour, Northwest Territories,” *Arctic Sci.*, vol. 4, no. 3, pp. 259–278, 2017.

[18] W. D. Halliday, M. K. Pine, S. J. Insley, R. N. Soares, P. Kortsalo, and X. Mouy, “Acoustic detections of arctic marine mammals near Ulukhaktok, northwest territories, Canada,” *Can. J. Zool.*, vol. 97, no. 1, pp. 72–80, 2019.

[19] C. D. Escobar-Amado, M. Badiey, and S. Pecknold, “Automatic detection and classification of bearded seal vocalizations in the northeastern Chukchi Sea using convolutional neural networks,” *J. Acoust. Soc. Amer.*, vol. 151, no. 1, pp. 299–309, 2022.

[20] S. Liu, M. Liu, M. Wang, T. Ma, and X. Qing, “Classification of cetacean whistles based on convolutional neural network,” in *Proc. 10th Int. Conf. Wireless Commun. Signal Process.*, 2018, pp. 1–5.

[21] W. Luo, W. Yang, and Y. Zhang, “Convolutional neural network for detecting odontocete echolocation clicks,” *J. Acoust. Soc. Amer.*, vol. 145, no. 1, pp. EL7–EL12, 2019.

[22] M. Zhong, M. Castellote, R. Dodhia, J. L. Ferres, M. Keogh, and A. Brewer, “Beluga whale acoustic signal classification using deep learning neural network models,” *J. Acoust. Soc. Amer.*, vol. 147, no. 3, pp. 1834–1841, 2020.

[23] Y. Shiu et al., “Deep neural networks for automated detection of marine mammal species,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, 2020.

[24] M. Thomas, B. Martin, K. Kowarski, B. Gaudet, and S. Matwin, “Marine mammal species classification using convolutional neural networks and a novel acoustic representation,” in *Machine Learning and Knowledge Discovery in Databases*, U. Brefeld, E. Fromont, A. Hotho, A. Knobbe, M. Maathuis, and C. Robardet, Eds. Cham, Switzerland: Springer, 2020, pp. 290–305.

[25] O. S. Kirsebom, F. Frazao, Y. Simard, N. Roy, S. Matwin, and S. Giard, “Performance of a deep neural network at detecting North Atlantic right whale upcalls,” *J. Acoust. Soc. Amer.*, vol. 147, no. 4, pp. 2636–2646, 2020.

[26] P. C. Bermant, M. M. Bronstein, R. J. Wood, S. Gero, and D. F. Gruber, “Deep machine learning techniques for the detection and classification of sperm whale bioacoustics,” *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 12588.

[27] L. Nanni, G. Maguolo, and M. Paci, “Data augmentation approaches for improving animal audio classification,” *Ecol. Informat.*, vol. 57, 2020, Art. no. 101084.

[28] E. Dufourq, C. Batist, R. Foquet, and I. Durbach, “Passive acoustic monitoring of animal populations with transfer learning,” *Ecol. Informat.*, vol. 70, 2022, Art. no. 101688.

[29] M. Zhong et al., “Acoustic detection of regionally rare bird species through deep convolutional neural networks,” *Ecol. Informat.*, vol. 64, 2021, Art. no. 101333.

[30] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.

[31] J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger,” in *Proc. IEEE 30th Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6517–6525.

[32] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” 2018, *arXiv:1804.02767.*

[33] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection,” 2020, *arXiv:2004.10934.*

[34] *YOLOV5*, Ultralytics, Los Angeles, CA, USA, 2022. [Online]. Available: https://github.com/ultralytics/yolov5

[35] M. S. Ballard et al., “Temporal and spatial dependence of a year-long record of sound propagation from the Canada Basin to the Chukchi Shelf,” *J. Acoust. Soc. Amer.*, vol. 148, no. 3, pp. 1663–1680, 2020.

[36] M. Badiey, L. Wan, S. Pecknold, and A. Turgut, “Azimuthal and temporal sound fluctuations on the Chukchi continental shelf during the Canada Basin Acoustic Propagation Experiment 2017,” *J. Acoust. Soc. Amer.*, vol. 146, no. 6, pp. EL530–EL536, 2019.

[37] C. Y. Wang, H. Y. Mark Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, “CSPNet: A new backbone that can enhance learning capability of CNN Chien-Yao,” in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 1571–1580.

[38] P. Msonda, S. A. Uymaz, and S. S. Karaağaç, “Spatial pyramid pooling in deep convolutional networks for automatic tuberculosis diagnosis,” *Traitement du Signal*, vol. 37, no. 6, pp. 1075–1084, 2020.

[39] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “PANet: Path aggregation network for instance segmentation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8759–8768.

[40] J. Yu and W. Zhang, “Face mask wearing detection algorithm based on improved YOLO-v4,” *Sensors*, vol. 21, no. 9, 2021, Art. no. 3263.

[41] Q. Song et al., “Object detection method for grasping robot based on improved YOLOV5,” *Micromachines*, vol. 12, no. 11, 2021, Art. no. 1273.

[42] S. Elfwing, E. Uchibe, and K. Doya, “Sigmoid-weighted linear units for neural network function approximation in reinforcement learning,” *Neural Netw.*, vol. 107, no. 2015, pp. 3–11, 2018.

[43] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, “Distance-IoU loss: Faster and better learning for bounding box regression,” in *Proc. 34th AAAI Conf. Artif. Intell.*, 2020, no. 2, pp. 12993–13000.

[44] *Global Sea Ice Concentration (SSMIS)*, EUMESAT OSI SAF. Accessed: Aug. 29, 2023. [Online]. Available: https://osi-saf.eumetsat.int/products/osi-401-d

[45] I. A. McLaren, “Some aspects of growth and reproduction of the bearded seal, Erignathus barbatus (Erxleben),” *J. Fisheries Res. Board Canada*, vol. 15, no. 2, pp. 219–227, 1958.

[46] M. C. Kingsley, I. Stirling, and W. Calvert, “The distribution and abundance of seals in the Canadian High Arctic, 1980–1982,” *Can. J. Fisheries Aquatic Sci.*, vol. 42, pp. 1189–1210, 1983.

[47] S. M. Van Parijs, C. Lydersen, and K. M. Kovacs, “Vocalizations and movements suggest alternative mating tactics in male bearded seals,” *Animal Behav.*, vol. 65, no. 2, pp. 273–283, 2003.

[48] S. M. V. Parijs, “Aquatic mating strategies of male bearded seals. I,” 2000. [Online]. Available: https://www.researchgate.net/profile/Sofie-Van-Parijs/publication/242424514_AQUATIC_MATING_STRATEGIES_OF_MALE_BEARDED_SEALSi/links/00b49535ae1e356e89000000/AQUATIC-MATING-STRATEGIES-OF-MALE-BEARDED-SEALSi.pdf

Christian Escobar-Amado received the Bachelor of Science degree in electronics engineering from the Francisco de Paula Santander University, Cúcuta, Colombia, in 2016, and the M.Sc. degree in electrical and computer engineering in 2022 from the University of Delaware, Newark, DE, USA, where he is currently working toward the Ph.D. degree in electrical and computer engineering.

From 2017 to 2018, he was a Junior Researcher with the Administrative Department of Science, Technology, and Innovation, also known as Colciencias, Bogotá, Colombia. He was an intern with the Ocean Acoustic Engineering Laboratory, University of Delaware. His research interests include shallow water acoustics, Bayesian optimization methods, and physics-based deep learning techniques applied to ocean acoustics.

Mr. Escobar-Amado is a Member of the Acoustical Society of America.

Mohsen Badiey received the Ph.D. degree in applied marine physics and ocean engineering from the Rosenstiel School of Marine and Atmospheric Science, University of Miami, Coral Gables, FL, USA, in 1988.

From 1988 to 1990, he was a Postdoctoral Fellow with the Port and Harbour Research Institute, Ministry of Transport, Yokosuka, Japan. After his postdoctoral research, he became a Faculty Member with the University of Delaware, Newark, DE, USA, where he is currently a Professor of Electrical and Computer Engineering and a joint Professor in Physical Ocean Science and Engineering. From 1992 to 1995, he was a Program Director and a Scientific Officer with the Office of Naval Research, Arlington, VA, USA, where he was the team leader to formulate long-term naval research in the field of Acoustical Oceanography. His research interests include physics of sound and vibration, shallow water acoustics and oceanography, underwater acoustic communications, acoustic signal processing and machine learning, seabed acoustics, and geophysics.

Dr. Badiey is a Fellow of the Acoustical Society of America.

Lin Wan received the Ph.D. degree in mechanical engineering from the Georgia Institute of Technology (Georgia Tech), Atlanta, GA, USA, in 2010.

He was a Postdoctoral Fellow with the School of Mechanical Engineering, Georgia Tech. After his postdoctoral research with Georgia Tech, he joined the Ocean Acoustics Laboratory, University of Delaware, Newark, DE, USA, where he is currently a Faculty Member with the Department of Electrical and Computer Engineering. He conducts experimental and theoretical research in ocean acoustics, acoustical oceanography, and acoustic signal processing. He is the Principal Investigator of various Office of Naval Research (ONR) grants to study the geoacoustic properties in marine sediments using broadband acoustic signals. His research interests include geoacoustic inversion, internal wave effects on 3-D sound propagation, and Arctic acoustics.

Dr. Wan was a recipient of the U.S. Navy ONR Graduate Traineeship Award supported by the special research award from the Ocean Acoustics program. He is a Member of the Acoustical Society of America.