Deep Learning-Based Identification of Arctic Ocean Boundaries and Near-Surface Phenomena in Underwater Echograms

Femina Senjaliya*, Melissa Cote*, Amanda Dash†, Alexandra Branzan Albu*, Andrea Niemi‡
Stéphane Gauthier‡, Julek Chawarski†, Steve Pearce†, Kaan Ersahin†, Keath Borg†
* Electrical and Computer Engineering, University of Victoria, Victoria, Canada
† ASL Environmental Sciences, Victoria, Canada ‡ Fisheries and Oceans Canada
Email: {feminasenjaliya, mcote, aalbu}@uvic.ca, {adash, jchawarski, spearce, kersahin, kborg}@aslenv.com,
{Andrea.Niemi, Stephane.Gauthier}@dfo-mpo.gc.ca

Abstract

Monitoring marine environments is a crucial part of understanding the impact of oceans on global climate and their importance for biodiversity and ecological systems, particularly in the Arctic region. Underwater active acoustic surveys with moored multi-frequency echosounders allow for the continuous collection of valuable data reflecting the complex dynamics of these environments. This paper addresses the automatic identification of sea surface boundaries and near-surface phenomena in echograms using deep learning methods to support researchers such as biologists in their work, who typically rely on time-consuming manual analyses. We propose a two-step process that first characterizes echograms according to the surface conditions using an image classification paradigm and then identifies the sea surface boundary and near-surface bubbles and their extent in the water column using a semantic segmentation paradigm. Segmentation is carried out using surface type-specific models, which perform better than a single global segmentation model. We also propose learning strategies, such as a custom boundary loss function, that further improve performance. Experiments with various image classification and semantic segmentation architectures allow us to select the most efficient models for Arctic echogram analysis that, when used in conjunction within our proposed pipeline and our learning strategies, offer excellent results.

1. Introduction

Monitoring marine environments is of prime importance to understand the impact of oceans on global climate and their importance for ecological systems. Underwater active acoustic surveys play a key role in examining the multifaceted dynamics of marine ecosystems, providing insights into the complex interactions among ecological elements. These surveys typically rely on data collected by echosounders, which emit a series of acoustic pulses (pings) at different frequencies and listen for echoes from potential targets to generate visualizations of the water column in a minimally invasive manner. Echosounders can be deployed in two setups: moored to the sea floor (looking upwards), or attached to ships (looking downwards). Ship-based active acoustics have provided critical information about forage species in the Western Canadian Arctic [13], yet they lack a complete annual perspective required to understand population and food web dynamics. The recent development of moored equipment to monitor key Arctic species relative to the surface ocean structure over an entire annual cycle is thus critical for understanding ecosystem responses to Arctic change; we favor such a setup in this paper.

Underwater acoustic imaging is based on the principle that different materials and boundaries reflect sound waves differently according to their acoustic properties [22]. Data are visualized as echograms, *i.e*. sets of single-frequency 2D images capturing the reflected echoes for series of pings over time, with the x-axis representing temporal units and the y-axis depicting distance units, *i.e*. the depth or range from the instrument in the water column. Each pixel intensity corresponds to the amplitude of the reflected echo at a given time and distance over a sampling volume (volume backscattering strength S_v). Biologists rely predominantly on time-consuming manual or semi-automatic echogram analyses, utilizing commercial software such as Echoview [9]. These analyses primarily focus on the statistical characteristics of organism aggregations [43]. There is a critical need for efficient and accurate automatic methods targeting echograms extending beyond species abundance tracking to include other crucial features such as sea surface boundaries and phenomena, which are key to understanding marine environments and to accurately assessing datasets as a whole.

This research is dedicated to advancing methods for automatically detecting near-surface ocean boundaries and

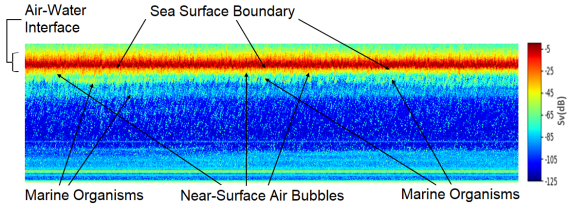

Figure 1. Sample one-hour echogram (125 kHz) covering the period from 9:00 pm to 10:00 pm on August 13, 2017, under windy conditions with open water (no ice), illustrating some of the challenges involved in distinguishing between near-surface bubbles and marine organisms. From the CBASSA dataset (see Sec. 3.1).

phenomena, focusing on entrained air bubbles and surface boundaries, using multi-frequency hourly echograms and deep learning (DL) approaches. The goals are to *identify the sea surface type* (*i.e*. surface condition, such as open water, solid ice, etc.) and *locate the sea surface boundary and any near-surface bubble layer* (downwelling of air driven by wave action), in particular its extent in the water column. Identifying and quantifying structure and dynamics at the atmosphere-ice-ocean interface is needed to understand the transfer of energy within changing Arctic ecosystems. The precise classification of sea surface types affects habitat availability and species interactions including the critical transfer of sympagic (ice-associated) carbon to the pelagic food web. Furthermore, bubble layers can present challenges to discerning biology from physics, as bubbles clouds can resemble aggregations of fish and plankton. Adding further to the challenge is that the (lower) boundary of the entrained air penetration within the water column can be indistinct and discontinuous [21]. Fig. 1 illustrates some of the challenges in distinguishing between near-surface bubbles (and their extent) and biology under windy open water conditions, as they both appear near the sea surface with similar patterns and strength.

Our contributions are as follows: 1) We propose a DL-based two-step process that allows for the automatic identification of the sea surface type via an image classification paradigm, followed by the automatic identification of the sea surface boundary and near-surface bubbles via a semantic segmentation paradigm using sea surface type-specific models. These specific models provide a better boundary and bubble detection compared to a single, global segmentation model. 2) We propose three learning-related strategies that further improve the segmentation performance: a) training first on the region of interest (sea surface) followed by training on full echograms; b) focusing on segment boundaries via a custom boundary loss function; c) addressing the class imbalance problem via a weighting scheme. 3) We provide extensive experiments on a variety of image classification and semantic segmentation architectures, leading to the selection of the most efficient models for Arctic echogram analysis.

These contributions enhance the methodological approaches available to experts in oceanography and marine biology. Additionally, the two-step pipeline has the potential to be applied beyond marine research, *e.g*. in medical imaging, where an initial classification could identify potential anomalies, and a subsequent anomaly-specific segmentation could delineate precise areas for diagnosis.

The remainder of the paper is divided as follows: Sec. 2 reviews relevant related works on echogram analysis, Sec. 3 describes our dataset and presents our proposed two-step methodology, Sec. 4 discusses experimental results, and Sec. 5 provides concluding remarks.

2. Related works

There is a long tradition of underwater echogram analysis, from early analog methods to sophisticated machine learning (ML) techniques and a growing use of DL techniques. Most research has focused on the acoustic classification of pelagic species and biomass estimates. Yassir *et al*. [48] review acoustic fish species identification using ML and DL. Conventional multi-frequency approaches [7, 16, 17, 40], particularly for zooplankton, look at the differential or relative frequency response and forgo any learning.

ML-based methods require hand-crafted features that typically relate to energetic, behavioral, and/or morphometric characteristics [14, 34]. Various ML classifiers have been utilized to identify fish species, such as: support vector machines [36], decision trees and random forests [10, 11, 23, 32, 38], minimum distance classifiers [3, 18], shallow artificial neural networks [2, 36, 45], and general Gaussian mixture models [47]. Of particular relevance here are the works of Minelli *et al*. [27], which used gradient boost classifiers to distinguish fish schools from other targets including gas bubbles, and of Sandy *et al*. [39], which used self-organizing maps to distinguish between open water and sea ice and to characterize statistical properties of surface wave height envelopes and ice draft.

While ML methods leverage expert knowledge via hand-crafted features, DL methods automatically learn relevant features from the data. This is one of many advantages of DL models over ML ones for automating fish species echo classification, which also tend to outperform ML ones even with few annotated data [48]. Existing DL methods for underwater echogram analysis can be classified according to the image analysis paradigm that they use: image classification [5, 35], object detection [25], instance segmentation [24], and semantic segmentation [1, 6, 21, 26, 30, 31, 41, 42, 46]. Of particular interest here are the works from Slonimer *et al*. [41, 42] that covered the detection of air bubbles with U-Net networks and from Lowe *et al*. [21] that detect the extent of entrained air bubbles in the water column in tidal energy streams using a U-Net-based architecture.

Existing DL works use a single image analysis paradigm; our approach leverages a combination of paradigms for im-

proved results. Few works in the literature have tackled the identification of sea surface types and near-surface bubbles; the works closest to our approach are that of Sandy *et al.* [39] (Arctic ocean boundary identification), of Slonimer *et al.* [41, 42] (air bubble identification), and of Lowe *et al.* [21] (extent of entrained air identification). Our paper differs from [39] in its use of DL as opposed to ML and in the more fine-grained level of sea surface type classification. It differs from [41, 42] in the targeted region (Arctic vs. coastal British Columbia) and the characteristics of involved marine organisms (unlike coastal British Columbia, the Arctic ecosystem possesses unique features that make it more complex to distinguish near-surface features from biology, as surface meltwater and the underside of ice provide distinct habitat for plankton and fish). It differs from [21] in its targeted region (Arctic vs. tidal energy demonstration site in the Bay of Fundy), its absence of data pre-processing ([21] utilized a data cleaning process in Echoview [9]), and its two-step process. Our proposed learning strategy that first focuses on the water-air interface followed by full echograms was inspired by [21], which used different zoom levels for training, starting with a full image followed by a zoomed-in image of the water column only.

3. Method

Fig. 2 shows the flowchart of the proposed method. In the first step, input echograms are fed to a DL image classification model which classifies them into one of six sea surface boundary types: open water (OW), windy open water (WOW), ice with keels (IK), ice without keels (INK), slushy conditions (SC), and mixed conditions (MC). In the second step, dedicated DL semantic segmentation models for each type yield segmented masks for the sea surface boundary and near-surface bubbles. Such a process allows for the first step to set the stage for a more nuanced and targeted analysis in the second step. Details on the dataset (including the surface types and the annotation process) and on the proposed method's two steps are given next.

3.1. CBASSA dataset

Our data come from the Cape Bathurst Arctic Sea Surface Acoustics (CBASSA) dataset. It consists of 15 months of one-hour multi-frequency echograms, collected from August 2017 to October 2018 using an upward-looking Acoustic Zooplankton Fish Profiler (AZFP) [19] echosounder moored to the sea floor, located near Cape Bathurst in the Northwest Territories of Canada (data from Fisheries and Oceans Canada). This area is a dynamic, productive region, and the CBASSA dataset provides a unique annual perspective of acoustic backscatter coupled to sea surface type, supporting the ecosystem-based approach to Arctic monitoring. The AZFP pinged the water column (about 51 m) at four frequencies (38, 125, 200, and 455 kHz). It was calibrated

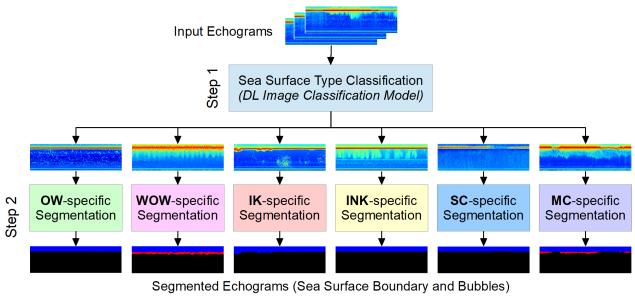

Figure 2. Flowchart of the proposed method. Input echograms are first classified into one of six sea surface types and then segmented by a surface type-specific segmentation model (DL semantic segmentation model), which outputs sea surface boundary and near-surface bubble segmentation masks. OW: open water, WOW: windy open water, IK: ice with keels, INK: ice without keels, SC: slushy conditions, MC: mixed conditions.

by the manufacturer prior to deployment and deployed at a 15 degree angle from the vertical axis.

The collected data are visualized as 201×712-pixel 1-hour echograms, where each pixel represents approximately 25.3 cm by 5 seconds. The echograms display the volume backscattering strength (S_v), calculated from the raw acoustic data and deployment metadata as [19]:

$$S_v = EL_{max} - \frac{2.5}{a} + \frac{N}{26214a} - SL + 20\log R + 2\alpha R - 10\log(\frac{c\tau\Psi}{2}). \quad (1)$$

Here, EL_{max} is the maximum echo level (in dB re 1μPa) that the 16-bit A/D converter can handle before reaching saturation, N the count value from the raw data, a the detector response's gradient (V/dB), α the seawater absorption coefficient (dB/m), R the distance from the instrument (m), SL the source level in dB re 1μPa at 1 m, c the speed of sound in the water (m/s), τ the duration of the transmitted pulse (s), and Ψ the two-way solid angle of the acoustic beam. In CBASSA, S_v values, typically ranging from around -125 to 0 dB, are converted to red-green-blue (RGB) integers using the "jet" colormap. Jet is appealing as it shows large changes in chroma and luminance, highlighting the smallest image features with high contrast, helping distinguish between co-occurring bubbles and biology.

The absence of actual ground truth data poses some challenges for training DL models. Our annotation process relies on contextual clues derived from information such as the deployment location, time of day, time of year, echo strength (absolute and relative between frequencies and visible targets in the echograms), target morphology, etc., and sometimes external relevant measurements such as wind speed and sampling. The sea surface type classification task requires one label per image, whereas the echogram segmentation task requires one label per pixel.

For the sea surface type classification task, we manually labeled 3,529 echograms, categorized into one of the fol-

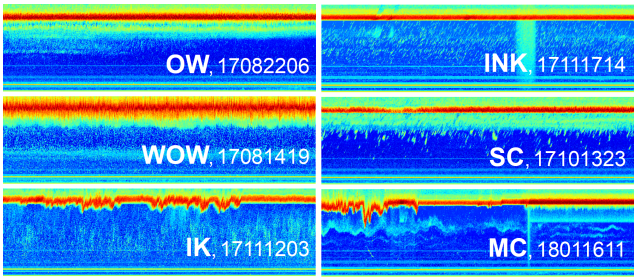

Figure 3. Sample one-hour echograms (125 kHz) illustrative of each sea surface type (top third region). 8-digit timestamp: YYMMDDHH, with HH in the 24-hour clock format. OW: open water, WOW: windy open water, IK: ice with keels, INK: ice without keels, SC: slushy conditions, MC: mixed conditions.

lowing six sea surface types that encompass the varied conditions in the Arctic environment: OW (646), WOW (882), IK (770), INK (506), SC (514), and MC (211). The disparity in class representation mirrors the natural occurrence rates of these conditions over several months. We focused on the 125 kHz frequency as it tends to offer the clearest representation to discern surface conditions. Fig. 3 shows typical echograms for each type. OW shows a strong air-water interface echo (red) and minimal below-surface scattering, indicating calm, ice-free water, typical in summer or when ice disperses. WOW displays irregular (jagged) air-water interface echoes and increased backscatter from wind-induced waves and entrained bubbles, creating a scattered appearance below the surface; biology can be mingling with the bubbles, creating a dynamic interplay appearing as cyan and yellowish hues. IK presents a locally smooth and strong surface echo with some protrusions into the water column of varied depth (ice keels), common in cold months. In contrast, INK presents a strong surface echo that looks flat, typical of winter continuous ice cover. SC represents the formation or melting of ice during shoulder seasons or temperature fluctuation events, creating a semi-solid (slushy) surface layer (diffuse, indistinct air-water interface echoes with a "smeared" or "fuzzy" appearance). Finally, MC combines the aforementioned features within the 1-hour window. To support the annotation process, we relied on the time of year and on satellite imagery from NASA Worldview [28] to observe surface conditions. We also referred to hourly wind data from Cape Parry, the nearest weather station, to corroborate wind conditions.

For the echogram segmentation task, our approach leveraged fused echograms that combine data from the 125, 200, and 455 kHz frequencies, excluding 38 kHz due to its lower sensitivity to small features. This is a pixel-level image fusion technique similar to that of [46], summing the S_v values from all frequencies pixel-wise before transposing the results to the jet colormap. Fused data enhance the perception of the image, either for human observers or for automated analysis systems, by incorporating the strengths of individual frequencies [20], although they are not typically

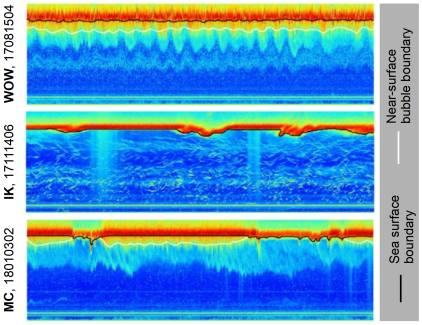

Figure 4. Sample sea surface and near-surface bubble boundary annotations superimposed on fused multi-frequency echograms. Timestamp: YYMMDDHH. WOW: windy open water, IK: ice with keels, MC: mixed conditions.

used by biologists. We developed a semi-automatic approach using traditional computer vision techniques to annotate the surface boundary and bubbles at the pixel level, requiring two user-set parameters each. The red band signifies the immediate reflection of acoustic energy where the air meets the ocean's surface. Its lower edge typically represents the sea surface boundary. As a general rule, bubbles occur below the ocean surface due to wind events, which form ripples or breaking waves. They can be characterized by their continuity and the way they taper off with depth, following the sea's undulating surface, with shades of amber, yellow, and lime yellow. Co-occurring biology often share similar shades of yellow, lime yellow (and cyan), but tend to be less predictable and may aggregate in patches. Bubbles appear in WOW and sometimes in MC; their conspicuous absence in OW, IK, INK, and SC can be attributed to the lack of surface agitation. The surface boundary annotation process makes use of horizontal Gaussian blurring and region growing to obtain a mask of the air-water interface, which is then extended to the top of the echogram in a customary way that represents the region outside of the water column to be excluded for any biological analysis; the boundary is the lower border (lowest pixel in each column) of the mask. The bubble annotation process makes use of the annotated sea surface boundary, Gaussian filtering, colormap conversions, k-means clustering, and connected component labeling, to obtain a bubble mask; the extent of entrained air bubbles is the lower border (lowest pixel in each column) of the mask. Fig. 4 shows examples of annotated sea surface and near-surface bubble boundaries. While the semi-automatic annotation process is faster than manual annotation, it remains time-consuming; 1,691 echograms (out of 3,529) were annotated for the segmentation task: OW (300), WOW (299 – 264 with bubbles), IK (298), INK (299), SC (318), MC (177 – 50 with bubbles).

3.2. Sea surface type classification (step 1)

Sea surface type classification is a critical first step, as it segregates the echograms into homogeneous groups that ex-

hibit (more) similar acoustic properties. Such grouping allows for a more targeted and effective subsequent analysis, also providing relevant information for biologists and acousticians studying the Arctic in itself.

We experimented with several DL image classification frameworks known for their efficacy in various applications to find the most suitable for underwater Arctic echogram analysis: ResNet-101 [12], Darknet-53 [33], DenseNet-201 [15], and Inception-v3 [44]. ResNet-101 employs residual learning, Darknet-53 is optimized for speed, DenseNet-201 packs densely connected layers, and Inception-v3 follows a modular approach, with inception modules that adaptively capture information at multiple scales within a single layer. This positions Inception-v3 as a versatile model for echograms, capable of accommodating the diverse sizes and scales inherent in marine acoustic imaging. Experiments (see Sec. 4.1.1) have shown Inception-v3 to be the most effective overall in our case; our first step is therefore based on Inception-v3. We also employed transfer learning (with the classification models pre-trained on the ImageNet dataset [8]), standard cross entropy loss, and a cost-sensitive class weighting approach to address class imbalance, in which penalties are assigned to each class via a cost matrix to increase the weight of the minority group.

3.3. Echogram segmentation (step 2)

For detecting bubbles and delineating the sea surface boundary, we favor a semantic segmentation approach that assigns a label to each pixel (either "surface", "bubble", or "background"). The background class can be quite varied as it covers everything else, *i.e.* anything that is not related to the air-water interface or near-surface physical phenomena, which may include biological signals. From the predicted masks, we can infer the sea surface boundary and the extent of entrained air bubbles as the lower border of the masks.

We experimented with several DL semantic segmentation frameworks renowned for their efficacy in various applications to find out the most suitable for underwater Arctic echogram analysis: U-Net [37], Attention U-Net [29], DeepLabV3 [4], UNet++ [49]. We mainly focused on U-Net-like architectures due to their proven success in several semantic segmentation-based echogram studies (see Sec. 2). U-Net follows an encoder-decoder architecture in which upsampled feature maps in the decoder are concatenated with corresponding feature maps from the encoder via skip connections. Attention U-Net adds attention gate mechanisms to focus on specific regions of interest. DeepLabV3 leverages atrous spatial pyramid pooling. UNet++ builds upon U-Net by adding nested and dense skip pathways. The nested skip connections facilitate the integration of features from different levels of the network hierarchy, enhancing the model's ability to capture fine details and global context simultaneously, whereas the dense skip connections promote feature reuse and propagation throughout the network. These enhancements improve the performance in scenarios with complex image structures and varying object scales, making UNet++ the best performing architecture in our case (see Sec. 4.1.2). Our second step is therefore based on UNet++. We also employed three learning strategies to improve results, the effect of which are shown in an ablation study (see Sec. 4.1.4): 1) zoomed-in first ("Zoom"), 2) custom boundary loss ("BL"), and 3) class weighting ("CW"). In the "Zoom" strategy, zoomed-in 128×128 tiles, covering the water-air interface and near surface regions are fed to the model for the first part of the training, before continuing with the full echogram. The rationale is for the model to first learn the subtleties of the near-boundary region and then learn the global context. The "BL" strategy introduces a custom boundary loss to augment the model's ability to capture fine-grained details at class boundaries. We add a custom boundary loss term ($\mathcal{L}_B$) to the loss function (which also uses the standard cross entropy). This new term, based on $\mathcal{L}1$ loss and the Sobel operator, allows the model to quantify the alignment between predicted and true boundaries:

$$\mathcal{L}_B = \frac{1}{N}\sum_{i=1}^{N} |\text{edge\_pred}_i - \text{edge\_GT}_i| \tag{2}$$

where N is the total number of pixels in the image, and edge_pred and edge_GT are the magnitude of the gradient for each pixel calculated as the convolution between the Sobel kernels and the one-hot encoded predicted and ground truth labels, respectively. The "CW" strategy is a bit more complex than that used in step 1, as it incorporates both class pixel counts within an echogram and echogram class presence counts within the dataset, to address the inherent imbalance between bubble, surface, and background pixels. The total weight for class i (w_i^{tot}) is calculated as follows:

$$w_i^{tot} = \max(\hat{w}_{p_i}, \hat{w}_{e_i}) \tag{3}$$

$$\hat{w}_{p_i} = \frac{w_{p_i}}{\sum w_{p_i}}, \quad w_{p_i} = \frac{N}{p_i} \tag{4}$$

$$\hat{w}_{e_i} = \frac{w_{e_i}}{\sum w_{e_i}}, \quad w_{e_i} = \begin{cases} \frac{M}{e_i} & \text{if } e_i > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $\hat{w}_{p_i}$ is the normalized pixel count weight (w_{p_i}) of class i, p_i is the number of pixels of class i in the echogram, $\hat{w}_{e_i}$ is the normalized echogram class presence count weight (w_{e_i}) of class i, M is the total number of echograms, and e_i is the number of echograms in which class i is present.

4. Experimental results

The experimental results cover the sea surface type classification step with all compared architectures, the echogram segmentation step with all compared architectures, the full proposed end-to-end pipeline, and an ablation study. The

sea surface type classification task was implemented in MATLAB and trained for 20 epochs using Stochastic Gradient Descent with Momentum (SGDM) optimization, a mini-batch size of 10, an initial learning rate of 0.0001, and data augmentations of random reflections along the x-axis and translations along both axes within a [-30, 30] pixel range. The echogram segmentation task was implemented in Python (PyTorch) and trained for 200 epochs using the Adam optimizer, a learning rate of 0.001, a batch size of 2, with random horizontal flips. When in use, the "Zoom" strategy covered the first 50 epochs. 20% of the CBASSA echograms were set aside for testing.

4.1. Quantitative evaluation

The classification task is assessed using the standard accuracy, recall, and precision metrics. The segmentation task is assessed using a combination of standard metrics that compare the predicted segmentation masks with the ground truth masks (intersection over union (IoU), recall, and precision) and of metrics that are more informative of the tracing of the sea surface boundary and the extent of the bubbles over time: the overall mean vertical distance (OMVD), the relative error (RE), false positives time-wise (FP-T), and false negatives time-wise (FN-T). OMVD computes the vertical distance between the lower border of the predicted and ground truth masks (with one pixel corresponding to 25.3 cm), while RE normalizes this distance with respect to the predicted vertical extent of the mask. FP-T and FN-T track pings (*i.e.* columns) where a model incorrectly predicts a segment or misses a segment, respectively, emphasizing the model's temporal consistency. For instance, a FP-T value of 13 for bubbles would indicate that for 13 pings out of 712 (712 being the total number of pings in one hour, or the width of the echogram), the model predicted bubbles when there were none in the ground truth. Here, all metrics are computed echogram-wise then averaged over the entire test set for each task. Arrows in the following tables indicate whether higher (↑) or lower (↓) metric values are desirable.

4.1.1 Sea surface type classification

Table 1 shows the performance of various image classification architectures (ResNet-101 [12], Darknet-53 [33], DenseNet-201 [15], and Inception-v3 [44]) for the sea surface type classification problem on the test set, for each of the six classes and overall. ResNet-101 excels in the OW class with the highest precision and accuracy, and leads in accuracy and recall for SC. Darknet-53 does not secure the top position in any case. DenseNet-201 achieves the highest recall for OW and the highest precision for WOW and SC. Inception-v3 outperforms others in IK, INK, and MC across all three metrics, indicating its robustness in complex classifications. Additionally, Inception-v3 achieves the overall highest scores in recall, precision, and accuracy, with a notable second-highest performance in accuracy and recall for OW and accuracy and recall for WOW, confirming its comprehensive proficiency for the task and making it a clear choice for step 1 of our end-to-end pipeline.

	Metric	OW	WOW	IK	INK	SC	MC	Overall
RN101	Acc ↑	**0.943**	**0.966**	0.942	0.955	**0.946**	0.965	0.858
	Rec ↑	0.791	**0.943**	0.870	0.832	**0.854**	0.738	0.838
	Prec ↑	**0.887**	0.922	0.865	0.848	0.793	0.689	0.834
DN53	Acc ↑	0.916	0.950	0.952	0.963	0.930	0.963	0.835
	Rec ↑	0.814	0.875	0.929	0.832	0.378	0.643	0.805
	Prec ↑	0.750	0.922	0.861	0.848	0.776	0.692	0.814
DN201	Acc ↑	0.917	0.955	0.952	0.966	0.944	0.970	0.852
	Rec ↑	**0.899**	0.875	0.955	0.812	0.698	0.714	0.826
	Prec ↑	0.716	**0.939**	0.845	0.943	**0.902**	0.769	0.852
IncV3	Acc ↑	0.930	0.956	**0.967**	**0.977**	0.943	**0.976**	**0.877**
	Rec ↑	0.837	0.926	**0.974**	**0.842**	0.777	**0.837**	**0.853**
	Prec ↑	0.794	0.911	**0.888**	**1.000**	0.825	**0.821**	**0.873**

Notes: OW: open water, WOW: windy open water, IK: ice with keels, INK: ice without keels, SC: slushy conditions, MC: mixed conditions, Acc: accuracy, Rec: recall, Prec: precision, RN101: ResNet-101 [12], DN53: Darknet-53 [33], DN201: DenseNet-201 [15], IncV3: Inception-v3 [44].

Table 1. Performance evaluation of various image classification architectures for the sea surface type classification problem on the test set. Best results in bold font, selected architecture underlined.

Class	Metric	U-Net [37]	A-UNet [29]	DLV3 [4]	UNet++ [49]
Surf	OMVD ↓	10.603	1.304	0.920	**0.571**
	RE ↓	0.382	0.032	0.022	**0.014**
	FP-T ↓	**0.000**	**0.000**	**0.000**	**0.000**
	FN-T ↓	**0.000**	**0.000**	**0.000**	**0.000**
	IoU ↑	0.751	0.969	0.978	**0.986**
	Recall ↑	0.751	0.980	0.986	**0.995**
	Prec ↑	**1.000**	0.989	0.992	0.991
Bub	OMVD ↓	6.712	4.432	2.039	**1.112**
	RE ↓	0.300	0.416	0.363	**0.294**
	FP-T ↓	**0.366**	8.666	132.030	25.710
	FN-T ↓	210.150	177.430	**0.560**	25.300
	IoU ↑	0.243	0.417	0.424	**0.605**
	Recall ↑	**0.960**	0.817	0.470	0.663
	Prec ↑	0.243	0.445	0.576	**0.773**

Notes: Surf: sea surface, Bub: bubble, OMVD: overall mean vertical distance, RE: relative error, FP-T: false positives time-wise, FN-T: false negatives time-wise, IoU: intersection over union, Prec: precision, A-UNet: Attention U-Net, DLV3: DeepLabV3.

Table 2. Performance evaluation of various architectures for the echogram segmentation problem on the WOW test set. Best results in bold font, selected architecture underlined.

4.1.2 Echogram segmentation

Table 2 shows the per-class performance of various semantic segmentation architectures (U-Net [37], Attention U-Net [29], DeepLabV3 [4], and UNet++ [49]) for the echogram segmentation problem on the WOW test set. To select the best architecture for the proposed method, the experiments targeted WOW conditions as these include all three pixel classes (surface, bubble, background). UNet++ distinguishes itself with the lowest OMVD and RE, indicating precise boundary capture and shape consistency. It also has the highest IoU and precision for bubbles and highest IoU

Model	Pixel Class	OMVD ↓ (pixels)	RE ↓	FN-T ↓ (columns)	FP-T ↓ (columns)	IoU ↑	Recall ↑	Precision ↑
OW-specific	Surface	0.423	0.010	0.000	0.000	0.990	0.996	0.995
WOW-specific	Surface	0.592	0.014	0.000	0.000	0.986	0.990	0.996
	Bubble	1.093	0.197	7.030	54.080	0.663	0.908	0.700
IK-specific	Surface	0.945	0.021	0.000	0.000	0.979	0.992	0.987
INK-specific	Surface	0.423	0.010	0.000	0.000	0.990	0.997	0.993
SC-specific	Surface	0.615	0.015	0.000	0.000	0.985	0.997	0.988
MC-specific	Surface	0.596	0.014	0.000	0.000	0.985	0.991	0.994
	Bubble	1.329	0.306	24.930	41.760	0.533	0.777	0.650
Overall (Step 1 GT)	Surface	0.413	0.010	0.000	0.000	0.990	0.995	0.995
	Bubble	1.170	0.232	13.000	49.970	0.620	0.864	0.683
Overall (End-to-end, proposed)	Surface	**0.650**	**0.015**	**0.000**	**0.000**	**0.985**	**0.993**	**0.992**
	Bubble	**1.231**	**0.223**	**29.730**	41.680	**0.598**	**0.813**	**0.663**
Single	Surface	2.380	0.054	**0.000**	0.336	0.972	0.983	0.989
	Bubble	6.319	1.946	222.358	**8.896**	0.243	0.270	0.610

Notes: OW: open water, WOW: windy open water, IK: ice with keels, INK: ice without keels, SC: slushy conditions, MC: mixed conditions, GT: ground truth annotations, OMVD: overall mean vertical distance, RE: relative error, FN-T: false negatives time-wise, FP-T: false positives time-wise, IoU: intersection over union.

Table 3. Performance evaluation of the six sea surface type-specific models per model, overall (utilizing sea surface boundary classification ground truth) and overall (full end-to-end pipeline, *i.e.* proposed) vs. a single global model for the echogram segmentation problem on the test set. Best results for each pixel class, between overall (proposed) and single, shown in bold font.

Exp	Zoom	BL	CW	OMVD ↓ (pixels)	RE ↓	FN-T ↓ (columns)	FP-T ↓ (columns)	IoU ↑	Recall ↑	Precision ↑
1				1.112	0.294	25.710	**25.300**	0.605	0.663	0.773
2	✓			1.076	0.257	21.230	25.880	0.659	0.765	**0.804**
3		✓		**1.036**	0.245	23.860	28.150	0.628	0.716	0.797
4			✓	4.102	0.541	**2.166**	169.58	0.561	0.856	0.591
5	✓	✓		1.102	0.243	23.310	29.980	0.647	0.737	0.765
6	✓		✓	1.060	0.250	28.35	28.65	0.636	0.731	0.771
7		✓	✓	1.149	0.198	15.066	47.280	0.622	0.809	0.676
8 (proposed)	✓	✓	✓	1.093	**0.197**	7.030	54.080	**0.663**	**0.908**	0.700

Notes: Exp: experiment, Zoom: learning features on zoomed-in echogram tiles first, BL: custom boundary loss, CW: class weighting, OMVD: overall mean vertical distance, RE: relative error, FN-T: false negatives time-wise, FP-T: false positives time-wise, IoU: intersection over union.

Table 4. Ablation study of the proposed segmentation model on the WOW test set for the bubble pixel class. Best results in bold font. Experiment 1: baseline UNet++, experiment 8: proposed approach.

and recall for surfaces. Interestingly, U-Net performs very differently for surfaces and bubbles, with the highest precision and lowest recall for surfaces, and the highest recall and lowest precision for bubbles. Attention U-Net does not yield any of the best metrics, but generally improves upon U-Net's performance for surfaces. DeepLabV3 performs best only in terms of FN-T for bubbles. The bubble class appears harder to segment with overall lower metrics values, also illustrated by all architectures being able to prevent false detections time wise (FP-T and FN-T) for surfaces. UNet++'s performance across key metrics positions it as the optimal model for step 2 of our end-to-end pipeline.

4.1.3 Single vs. multiple segmentation models

Table 3 compares the performance of the six sea surface type-specific segmentation models (per model, overall utilizing the sea surface boundary classification ground truth (called "Step 1 GT"), and overall utilizing the full pipeline, (called "End-to-end, proposed")), to that of a single global model trained on all data (all sea surface types at once), for the echogram segmentation problem on the test set. All are based on Inception-v3 and UNet++ and include the "Zoom", "BL", and the "CW" improvements. The "Step 1 GT" case allows us to remove any error propagation from step 1, whereas the "End-to-end, proposed" case showcases the actual end results. All models related to classes without bubbles have high recall, precision, and IoU. The performance is nuanced for models associated with bubbles (WOW and MC), as the bubble class has proven more difficult to segment (see Sec. 4.1.2). MC model results are less remarkable, attributable to the complexity of learning a class combining multiple conditions. Comparing our approach to "Step 1 GT", there are minor variations (2-3 points) for bubbles in IoU and precision, with a larger change in recall due to some step 1 misclassifications. For surface pixels, the metrics remain consistent (negligible difference of 0.05 points). There is thus an error propagation effect, with errors in sea surface type classification affecting the end results, but of limited scope. *Utilizing a single model applied globally without the classification step shows a significant drop in performance, with OMVD increasing fivefold, decreases in recall and IoU (significant for bubbles), and a marked increase in FN-T for bubbles, compared to the proposed end-to-end approach.* This illustrates the advantages of our approach's specialized models over a generalized single model.

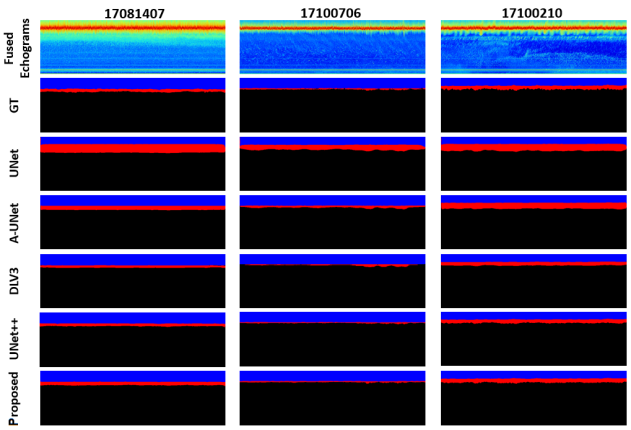

Figure 5. Sample segmentation results (rows 3 to 7) for fused echograms (row 1) across models, with ground truth (GT) masks in row 2. Color code of pixel masks: background (black), blue (surface), red (bubbles). Timestamp: YYMMDDHH.

4.1.4 Ablation study

Table 4 presents a full ablation study of the proposed echogram segmentation model on the WOW test set, for the more difficult bubble class. Experiments #1 to 8 cover all combinations of adding (or not) the three proposed learning strategies while training the retained UNet++ architecture: "Zoom", "BL", and "CW". #1 corresponds to the baseline UNet++, whereas #8 corresponds to the proposed method. The baseline UNet++ (#1) minimizes FP-T, the "Zoom" feature (#2) maximizes precision, the "BL" feature (#3) yields the most accurate boundary tracing (OMVD), while the "CW" (#4) minimizes FN-T. The combination of all three learning strategies enhances the performance, making our approach (#8) the overall best: best RE, IoU, and recall. *This study validates our end-to-end approach's superiority over the baseline UNet++ for segmenting intricate echogram features, particularly for the nuanced task of bubble detection, which is central to the thorough analysis of Arctic underwater imagery.*

4.2. Qualitative evaluation

Fig. 5 shows typical segmentation results of fused multi-frequency echograms for all compared architectures and for the proposed segmentation model (UNet++ with custom learning strategies). Our model exhibits the closest alignment with the ground truth, particularly in the accurate delineation of boundaries, essential for the purpose of finding the sea surface line and bubble extent. Fig. 6, showing additional sample results of our method for both the classification step and the end results (segmentation), demonstrates our method's proficiency in classifying and segmenting various sea ice conditions. Complex scenarios like WOW and MC (Fig. 6(b,f,g)) are particularly well processed. Fig. 6(h) is a noteworthy exception, where a misclassification in step 1 led the (wrong) segmentation model to falsely detect a small layer of bubbles, highlighting a limitation due to error propagation. However, the infrequency of such errors underscores the reliability and overall accuracy of the pipeline.

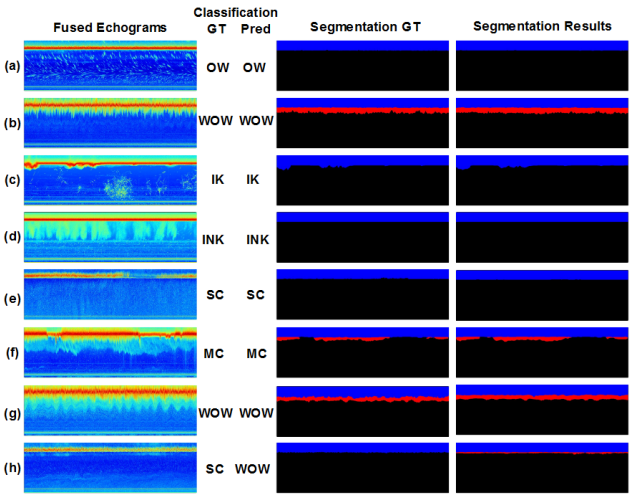

Figure 6. Additional sample results of the proposed method for both classification (GT: ground truth, pred: predicted) and segmentation. Color code of pixel masks: background (black), blue (surface), red (bubbles). Timestamps (YYMMDDHH): (a) 17082416, (b) 17100614, (c) 17112019, (d) 18011204, (e) 17100904, (f) 18010302, (g) 17081505, (h) 17081607. OW: open water, WOW: windy open water, IK: ice with keels, INK: ice without keels, SC: slushy conditions, MC: mixed conditions.

5. Conclusion

This paper tackles the automatic identification of ocean boundaries and near-surface phenomena under challenging Arctic conditions in multi-frequency echograms using DL models, to advance the monitoring of coupled ice-ocean ecosystems. It characterizes the problem as a two-step process, first identifying the sea surface type present in a given echogram via an image classification paradigm, and then locating the sea surface boundary and identifying any near-surface bubble cloud and its vertical extent in the water column via a semantic segmentation paradigm. Several widely used DL models for both paradigms are compared on the CBASSA datset, with Inception-v3 and UNet++ standing out. The paper also proposes three learning-related strategies for the segmentation step that, used in conjunction UNet++, further improves the performance. Future work will put an emphasis on differentiating near-surface biology signals from bubbles, as these get typically discarded from biomass estimates when they co-occur with bubbles, providing more reliable biological analyses, and on testing how well our models generalize to other Arctic locations.

Acknowledgments

This work was enabled by NSERC Canada and ASL Environmental Sciences (Alliance-Mitacs Grants program).

References

[1] O. Brautaset, A.U. Waldeland, E. Johnsen, K. Malde, L. Eikvil, A.-B. Salberg, et al. Acoustic classification in multifrequency echosounder data using deep convolutional neural networks. *ICES Journal of Marine Science*, 2020. 2

[2] A.G. Cabreira, M. Tripode, and A. Madirolas. Artificial neural networks for fish-species identification. *ICES Journal of Marine Science*, 66(6):1119–29, 2009. 2

[3] A. Charef, S. Ohshimo, I. Aoki, and N. Al Absi. Classification of fish schools based on evaluation of acoustic descriptor characteristics. *Fisheries Science*, 76(1):1–11, 2010. 2

[4] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017. 5, 6

[5] C. Choi, M. Kampffmeyer, N.O. Handegard, A.-B. Salberg, O. Brautaset, L. Eikvil, and R. Jenssen. Semi-supervised target classification in multi-frequency echosounder data. *ICES Journal of Marine Science*, 78(7):2615–2627, 2021. 2

[6] C. Choi, M. Kampffmeyer, N.O. Handegard, A.-B. Salberg, and R. Jenssen. Deep semisupervised semantic segmentation in multifrequency echosounder data. *IEEE Journal of Oceanic Engineering*, 2023. 2

[7] A. De Robertis, D.R. McKelvey, and P.H. Ressler. Development and application of an empirical multifrequency method for backscatter classification. *Canadian Journal of Fisheries and Aquatic Sciences*, 67(9):1459–1474, 2010. 2

[8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255. IEEE, 2009. 5

[9] Echoview Software Pty Ltd. Hydroacoustic Data Processing - Echoview — Echoview. https://echoview.com/. Accessed: 2024-02-14. 1, 3

[10] N.G. Fallon, S. Fielding, and P.G. Fernandes. Classification of Southern Ocean krill and icefish echoes using random forests. *ICES Journal of Marine Science*, 73(8):1998–2008, 2016. 2

[11] S. Gauthier, J. Oeffner, and R.L. O'Driscoll. Species composition and acoustic signatures of mesopelagic organisms in a subtropical convergence zone, the New Zealand Chatham Rise. *Marine Ecology Progress Series*, 503:23–40, 2014. 2

[12] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 5, 6

[13] J. Herbig, J. Fisher, C. Bouchard, A. Niemi, M. LeBlanc, A. Majewski, S. Gauthier, and M. Geoffroy. Climate and juvenile recruitment as drivers of Arctic cod (Boreogadus saida) dynamics in two Canadian Arctic seas. *Elementa: Science of the Anthropocene*, 11(1):00033, 2023. 1

[14] J.K. Horne. Acoustic approaches to remote species identification: A review. *Fisheries Oceanography*, 9(4):356–71, 2000. 2

[15] G. Huang, Z. Liu, L. Van Der Maaten, and K.Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017. 5, 6

[16] R.J. Korneliussen and E. Ona. An operational system for processing and visualizing multi-frequency acoustic data. *ICES Journal of Marine Science*, 59(2):293–313, 2002. 2

[17] R.J. Korneliussen, Y. Heggelund, G.J. Macaulay, D. Patel, E. Johnsen, and I.K. Eliassen. Acoustic identification of marine species using a feature library. *Methods in Oceanography*, 17:187–205, 2016. 2

[18] P. LeFeuvre, G.A. Rose, R. Gosine, R. Hale, W. Pearson, and R. Khan. Acoustic species identification in the Northwest Atlantic using digital image processing. *Fisheries Research*, 47(2-3):137–47, 2000. 2

[19] D. Lemon, P. Johnston, J. Buermans, E. Loos, G. Borstad, and L. Brown. Multiple-frequency moored sonar for continuous observations of zooplankton and fish. In *IEEE Oceans*, pages 1–6. IEEE, 2012. 3

[20] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin. Pixel-level image fusion: A survey of the state of the art. *Information Fusion*, 33:100–112, 2017. 4

[21] S.C. Lowe, L.P. McGarry, J. Douglas, J. Newport, S. Oore, C. Whidden, and D.J. Hasselman. Echofilter: A deep learning segmention model improves the automation, standardization, and timeliness for post-processing echosounder data in tidal energy streams. *Frontiers in Marine Science*, 9:867857, 2022. 2, 3

[22] X. Lurton. *An introduction to underwater acoustics: principles and applications*. Springer Science & Business Media, 2002. 1

[23] L. Mannocci, F. Baidai, Y.and Forget, M.T. Tolotti, L. Dagorn, and M. Capello. Machine learning to detect bycatch risk: Novel application to echosounder buoys data in tuna purse seine fisheries. *Biological Conservation*, 255:109004, 2021. 2

[24] T.P. Marques, M. Cote, A. Rezvanifar, A.B. Albu, K. Ersahin, T. Mudge, and S. Gauthier. Instance segmentation-based identification of pelagic species in acoustic backscatter data. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4378–4387, 2021. 2

[25] T.P. Marques, A. Rezvanifar, M. Cote, A.B. Albu, K. Ersahin, T. Mudge, and S. Gauthier. Detecting marine species in echograms via traditional, hybrid, and deep learning frameworks. In *25th International Conference on Pattern Recognition (ICPR)*, pages 5928–5935. IEEE, 2021. 2

[26] T.P. Marques, M. Cote, A. Rezvanifar, A. Slonimer, A.B. Albu, K. Ersahin, and S. Gauthier. U-MSAA-Net: a multi-scale additive attention-based network for pixel-level identification of finfish and krill in echograms. *IEEE Journal of Oceanic Engineering*, 2023. 2

[27] 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*, 21(9):2999, 2021. 2

[28] NASA. EOSDIS Worldview. https://worldview.earthdata.nasa.gov/. Accessed: 2024-02-20. 4

[29] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B.

Kainz, et al. Attention U-Net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018. 5, 6

[30] A. Ordonez, I. Utseth, O. Brautaset, R. Korneliussen, and N.O. Handegard. Evaluation of echosounder data preparation strategies for modern machine learning models. *Fisheries Research*, 254:106411, 2022. 2

[31] A. Pala, A.and Oleynik, I. Utseth, and N.O. Handegard. Addressing class imbalance in deep learning for acoustic target classification. *ICES Journal of Marine Science*, 80(10): 2530–2544, 2023. 2

[32] R. Proud, R. Mangeni-Sande, R.J. Kayanda, M.J. Cox, C. Nyamweya, C. Ongore, V. Natugonza, I. Everson, M. Elison, L. Hobbs, et al. Automated classification of schools of the silver cyprinid Rastrineobola argentea in Lake Victoria acoustic survey data using random forests. *ICES Journal of Marine Science*, 77(4):1379–1390, 2020. 2

[33] J. Redmon and A. Farhadi. YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 5, 6

[34] D. Reid, C. Scalabrin, P. Petitgas, J. Masse, R. Aukland, P. Carrera, et al. Standard protocols for the analysis of school based data from echo sounder surveys. *Fisheries Research*, 47(2-3):125–36, 2000. 2

[35] A. Rezvanifar, T.P. Marques, M. Cote, A.B. Albu, A. Slonimer, T. Tolhurst, K. Ersahin, T. Mudge, and S. Gauthier. A deep learning-based framework for the detection of schools of herring in echograms. In *NeurIPS Workshop Tackling Climate Change with Machine Learning*, 2019. 2

[36] H. Robotham, P. Bosch, J.C. Gutiérrez-Estrada, J. Castillo, and I. Pulido-Calvo. Acoustic identification of small pelagic fish species in Chile using support vector machines and neural networks. *Fisheries Research*, 102(1-2):115–22, 2010. 2

[37] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In *18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer, 2015. 5, 6

[38] S. Rousseau, S. Gauthier, C. Neville, S. Johnson, and M. Trudel. Acoustic classification of juvenile Pacific salmon (Oncorhynchus spp) and Pacific herring (Clupea pallasii) schools using random forests. *Frontiers in Marine Science*, 9:857645, 2022. 2

[39] S.J. Sandy, S.L. Danielson, and A.R. Mahoney. Automating the acoustic detection and characterization of sea ice and surface waves. *Journal of Marine Science and Engineering*, 10 (11):1577, 2022. 2, 3

[40] M. Sato, J.K. Horne, S.L. Parker-Stetter, and J.E. Keister. Acoustic classification of coexisting taxa in a coastal ecosystem. *Fisheries Research*, 172:130–136, 2015. 2

[41] A.L. Slonimer, M. Cote, T.P. Marques, A. Rezvanifar, S.E. Dosso, A.B. Albu, K. Ersahin, T. Mudge, and S. Gauthier. Instance segmentation of herring and salmon schools in acoustic echograms using a hybrid U-Net. In *19th Conference on Robots and Vision (CRV)*, pages 8–15. IEEE, 2022. 2, 3

[42] A.L. Slonimer, S.E. Dosso, A.B. Albu, M. Cote, T.P. Marques, A. Rezvanifar, K. Ersahin, T. Mudge, and S. Gauthier. Classification of herring, salmon, and bubbles in multifrequency echograms using U-Net neural networks. *IEEE Journal of Oceanic Engineering*, 2023. 2, 3

[43] T.K. Stanton. 30 years of advances in active bioacoustics: a personal perspective. *Methods in Oceanography*, 1:49–77, 2012. 1

[44] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016. 5, 6

[45] S.A. Villar, A. Madirolas, A.G. Cabreira, A. Rozenfeld, and G.G. Acosta. Ecopampa: A new tool for automatic fish schools detection and assessment from echo data. *Heliyon*, 7 (1):e05906, 2021. 2

[46] R. Vohra, F. Senjaliya, M. Cote, A. Dash, A.B. Albu, J. Chawarski, S. Pearce, and K. Ersahin. Detecting underwater discrete scatterers in echograms with deep learning-based semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 375–384, 2023. 2, 4

[47] M. Woillez, P.H. Ressler, C.D. Wilson, and J.K. Horne. Multifrequency species classification of acoustic-trawl survey data using semi-supervised learning with class discovery. *The Journal of the Acoustical Society of America*, 131 (2):EL184–EL190, 2012. 2

[48] A. Yassir, S.J. Andaloussi, O. Ouchetto, K. Mamza, and M. Serghini. Acoustic fish species identification using deep learning and machine learning algorithms: A systematic review. *Fisheries Research*, 266:106790, 2023. 2

[49] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang. UNet++: A nested U-Net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018*, pages 3–11. Springer, 2018. 5, 6