JASA EXPRESS LETTERS ARTICLE asa.scitation.org/journal/jel

Detecting ringed seal vocalizations using deep learning

Karlee E. Zammit,[1,a] William D. Halliday,[1,2,b] Amalis Riera,[3,c] Annika Heimrich,[3,d] Mariana Barbosa,[1,e] Nikoletta Diogou,[1,f] Olokhaktomiut Hunters and Trappers Committee[4,g] Stephen J. Insley,[2,3,h] Fabio Frazao,[5,i] Sebastien Fabbro,[6,7,j] and Stan E. Dosso[1,k]
[1]School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia, V8P5C2, Canada
[2]Wildlife Conservation Society Canada, Whitehorse, Yukon, Y1A0E9, Canada
[3]Department of Biology, University of Victoria, Victoria, British Columbia, V8P5C2, Canada
[4]Ulukhaktok, Northwest Territories, X0E0S0, Canada
[5]Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia, B3H4R2, Canada
[6]National Research Council of Canada, Herzberg Astronomy & Astrophysics Research Centre, Victoria, British Columbia, V9E2E7, Canada
[7]Department of Physics, University of Victoria, Victoria, British Columbia, V8P5C2, Canada

Abstract: Ringed seals (*Pusa hispida*) in the Arctic face habitat loss from climate change, making monitoring critical for conservation. Passive acoustic monitoring provides essential data; however, large data volumes require automated analysis. We developed a ResNet-based detector for ringed seal barks in passive acoustic recordings. The detector achieved F1 scores above 0.90 on development data and held-out datasets from different locations and recording periods. In continuous recordings, it maintained high recall (>0.90), with lower precision (∼0.5) reflecting a trade-off between detecting most vocalizations and false positives from acoustically similar non-target sounds. The open-source detector supports large-scale bioacoustic monitoring. *© 2026 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).*

[Editor: Annebelle CM Kok] https://doi.org/10.1121/10.0046176

Received: 28 May 2026 **Accepted:** 8 August 2026 **Published Online:** 28 August 2026

1. Introduction

The Canadian Arctic is experiencing accelerated climate change, characterized by thinning sea ice and shorter ice-cover duration (Comiso, 2012; Markus *et al.*, 2009). These changes threaten ice-dependent species, including ringed seals (*Pusa hispida*), through habitat loss and reduced foraging opportunities (COSEWIC, 2019; Stirling and Derocher, 2012). Furthermore, the Arctic soundscape is transitioning from a historically low-anthropogenic-noise environment to one increasingly shaped by human activity (Halliday *et al.*, 2020a). Vessel traffic, resource exploration, and military activity are projected to increase as sea ice diminishes, potentially leading to acoustic masking, behavioral disturbance, and physiological stress for endemic marine mammals (Clark *et al.*, 2009; Dawson *et al.*, 2018; Erbe and Farmer, 2000). Effective conservation therefore requires a comprehensive understanding of specific species' critical habitats and the mitigation of these increasing threats.

The underwater soundscape of the Western Canadian Arctic is acoustically complex and varies seasonally in response to geophony, anthropophony, and biophony (Halliday *et al.*, 2020a). High winter ice concentrations are generally associated with lower ambient sound levels, whereas open-water conditions are characterized by increased sound from wind, waves, and vessel activity (Dingwall *et al.*, 2024; Halliday *et al.*, 2019). Biological contributions also vary seasonally and include vocalizations from beluga whales (*Delphinapterus leucas*), bowhead whales (*Balaena mysticetus*), bearded seals (*Erignathus barbatus*), and ringed seals. Previous studies have reported variable spatial and seasonal patterns in ringed seal acoustic activity across the region (Halliday *et al.*, 2022; Halliday *et al.*, 2017; Halliday *et al.*, 2019; Halliday *et al.*, 2020b).

[a]Corresponding author: karlee.zammit@gmail.com
[b]Email: whalliday@wcs.org
[c]Email: amalis.riera@gmail.com
[d]Email: afheimrich@gmail.com
[e]Email: marianabarbosafa@gmail.com
[f]Email: niki.diogou@gmail.com
[g]Email: ohtc2015@outlook.com
[h]Email: sinsley@wcs.org
[i]Email: fsfrazao@dal.ca
[j]Email: sebastien.fabbro@nrc.ca
[k]Email: sdosso@uvic.ca

The ringed seal is a culturally and ecologically significant species, currently assessed as a species of special concern in Canada (COSEWIC, 2019). Ringed seals produce five vocalization types: barks, yelps, growls, woofs, and chirps, with barks and yelps being the most prevalent (Barbosa *et al.*, 2024; Stirling, 1973). These vocalizations are typically short and low amplitude (Barbosa *et al.*, 2024; Halliday *et al.*, 2019). Mean call durations are under 1 s, with barks averaging 0.2 s, yelps 0.3 s, and growls 0.8 s (Barbosa *et al.*, 2024). Barks are characterized by low frequencies (approximately 50–400 Hz) and little frequency modulation, whereas yelps occur at higher frequencies (approximately 500–2500 Hz) and exhibit strong frequency modulation (Stirling, 1973). Ringed seal vocalizations have low source levels, which make them highly susceptible to masking by ambient noise, and the specific functions of these calls remain largely unknown (Barbosa *et al.*, 2024; Stirling *et al.*, 1983). Despite these challenges, passive acoustic monitoring (PAM) is an important tool for studying the spatial and temporal distribution of ringed seals because it circumvents many of the logistical difficulties associated with visual surveys in remote, harsh environments (Hannay *et al.*, 2013). Ringed seal vocalizations have therefore been used to investigate spatial and temporal patterns of the species' occurrence throughout the Arctic (Calvert and Stirling, 1985; Halliday *et al.*, 2022; Halliday *et al.*, 2017; Halliday *et al.*, 2019; Halliday *et al.*, 2020b; Hannay *et al.*, 2013; Jones *et al.*, 2014; Stirling *et al.*, 1983). However, these studies relied on manual detection of vocalizations, requiring hundreds of hours of analyst time and limiting the volume of acoustic data that could be examined.

As large volumes of PAM data are collected annually in the Arctic, comprehensive manual analysis has become increasingly infeasible. Automated detectors can improve the efficiency and consistency of acoustic analyses, and detectors using convolutional neural networks (CNNs) have been successfully developed for several other marine mammal species (Bergler *et al.*, 2019; He *et al.*, 2016; Padovese *et al.*, 2023). However, automated detection of ringed seal vocalizations remains challenging because these calls are brief in duration (<1 s), generally have low signal-to-noise ratios (SNRs), and are frequently obscured by or confused with louder sound sources such as ice cracking and bearded seal vocalizations (Barbosa *et al.*, 2024; Halliday *et al.*, 2020b). In this study, we develop and evaluate a supervised deep-learning detector based on a custom 15-layer ResNet architecture to identify ringed seal barks in spectrograms of passive acoustic recordings from the Western Canadian Arctic. The detector performs binary classification, distinguishing segments containing ringed seal barks from those without them. We focus on barks because they are consistently present across datasets, often produced in sequences with yelps, and have both a consistent spectral signal and a relatively higher amplitude compared to yelps (Barbosa *et al.*, 2024; Halliday *et al.*, 2019). The detector is evaluated using recordings collected across multiple locations, deployment configurations, and acoustic conditions to assess its applicability to large-scale Arctic PAM datasets.

2. Data and methods

2.1 Acoustic data

PAM data were collected by Wildlife Conservation Society Canada and Fisheries and Oceans Canada (DFO) at 18 locations within the range of ringed seals in the Western Canadian Arctic between 2017 and 2023. These sites, spanning approximately 864 km, were grouped into four regions: Cape Bathurst, Cambridge Bay, Pearce Point, and Ulukhaktok (Fig. 1). Recordings collected between 2017 and 2019 were obtained from long-term deployments lasting 348–396 days using Ocean Instruments SoundTrap ST500 (Ocean Instruments, Fall City, WA) or Wildlife Acoustics Song Meter SM3M recorders (Wildlife Acoustics, Maynard, MA) with varying duty cycles and deployment depths. These deployments captured ringed seal vocalizations opportunistically across a range of source distances and ambient noise conditions, resulting in signals of variable quality and SNR. In contrast, the 2022–2023 deployments in Ulukhaktok consisted of short-term, continuous recordings lasting 1–5 days at a sampling rate of 96 kHz. SoundTrap ST300 or ST600 recorders were deployed through sea-ice cracks near known ringed seal haul-out sites, generally resulting in higher SNR. Together, the recordings comprised 95 954 WAV files, totaling 5832 h and approximately 3 TB of data.

2.2 Annotations

Manual review of the complete acoustic dataset was not feasible; therefore, a subset of the available recordings was annotated to provide labeled data for detector development and evaluation. Seven experienced marine bioacousticians annotated all identifiable ringed seal vocalization types using Raven Pro (Version 1.6; K. Lisa Yang Center for Conservation Bioacoustics, 2024). To promote consistency among annotators, all annotations were subsequently verified by a single senior analyst. This study focused on barks, which are highly consistent across datasets and typically exhibit high amplitudes. Annotations of other vocalization types were therefore excluded, although sequences containing both barks and yelps were retained because the calls could not always be separated reliably. An additional analyst refined the retained annotations to ensure accurate bark start and end times and to limit gaps within annotated sequences to a maximum of 1 s. Ambiguous signals for which the annotator could not confidently determine whether the sound was a ringed seal vocalization were excluded because the use of clear, high-SNR training examples has been shown to improve model performance (Cai *et al.*, 2022; Mouy *et al.*, 2024). The retained annotations included individual barks and bark-yelps, both of which were assigned to the positive binary class. Segments that did not contain these vocalizations were assigned to the negative class. The final dataset consisted of both discrete vocalizations and call sequences of variable length.

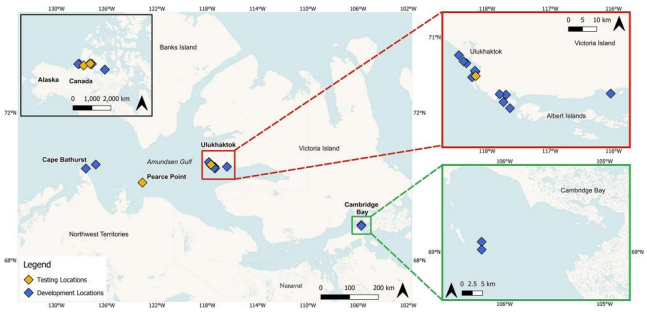

Fig. 1. Recorder locations within the four regions: Cape Bathurst, Pearce Point, Ulukhaktok, and Cambridge Bay. Overlapping sites in Ulukhaktok and Cambridge Bay are shown in further detail in the zoom-in panels on the right (projection WGS84). See Table 1 for a description of testing and development locations.

2.3 Data processing

Audio segments were generated by referencing the annotation tables to define time periods containing ringed seal vocalizations. From these periods, 1 s segments were extracted; annotations exceeding 1 s were split into multiple 1 s segments, while shorter calls were encompassed within a single 1 s window. These "positive" segments (i.e., segments containing barks, labeled 1) were not forcibly centered to increase data variability, following established methods for marine mammal detectors (Kirsebom *et al.*, 2020). All waveforms were resampled to 3000 Hz and normalized to zero mean and unit standard deviation. The waveforms were then converted to magnitude spectrograms using a short-time Fourier transform with a 0.032 s window size, 0.001 s window step, and Hamming window function. The frequency range was restricted to 100–1200 Hz to avoid low-frequency noise and capture the primary spectral content of the barks. Tonal and slowly varying background noise, which frequently obscured ringed seal barks, was reduced by subtracting the median value from each row of the spectrogram. Although this preprocessing may have reduced the representation of tonal background sounds among the negative training samples, it helped improve target-signal clarity given the limited amount of available training data. The resulting input spectrogram had dimensions of 750 time bins $\times$ 31 frequency bins $\times$ 1 intensity channel (Fig. 2).

An equal number of "negative" segments (labeled 0) were automatically generated from non-annotated periods using the same normalization and spectrogram configurations. Incorrectly classified samples within negative segments were possible due to the potential of missed vocalizations. Manual inspection indicated that most negative samples contained background noise or non-target sounds, and any containing ringed seal vocalizations were replaced. Manual inspection of these confirmed negative samples identified four categories: non-biological signals (14%), bearded seal vocalizations (5%), unidentified signals (6%), and background noise (75%). In total, the dataset comprised 10 280 samples, consisting of 5140 positive and 5140 negative segments.

The acoustic data were partitioned into five datasets (A–E) to facilitate model development and spatial-temporal testing (Table 1). Datasets A–C consisted of manually verified 1 s spectrograms in HDF5 format, referred to as "segmented" datasets, while datasets D–E comprised continuous WAV files for real-world deployment testing, referred to as "continuous" datasets.

Dataset A included the majority of annotated data, containing 9090 positive and negative samples from 2017–2022 across Cape Bathurst, Cambridge Bay, and Ulukhaktok. These samples were split into training, validation, and testing subsets using a 70/20/10 temporal split (Kirsebom *et al.*, 2020). To account for differences in recording context between targeted short-term and opportunistic long-term deployments, their relative proportions (32% and 68%, respectively) were preserved across all three subsets. Short-term recordings, typically captured near known haul-out locations, provided higher SNR vocalizations compared to the opportunistic long-term recordings.

Datasets B–E were used exclusively for testing the generalizability of the detector on data unseen during training and validation. Dataset B (Pearce Point, 2018–2019) was intentionally withheld from development to evaluate performance at a new location. Dataset C comprised 1022 labeled spectrograms collected in Ulukhaktok in 2023 and served as a temporal test set, as it originated from the same region as the training data but from a later recording period not included in model development. Finally, Datasets D and E consisted of continuous long-form audio recordings from Pearce Point and Ulukhaktok, respectively, to assess detector performance in a continuous monitoring context. Rather than being provided as pre-segmented 1 s clips, the recordings were supplied as continuous audio and automatically divided into 1 s segments during processing.

2.4 Neural network architecture

The detector was developed using custom scripts written in Python 3.9.17 and the Ketos package version 2.7 (MERIDIAN, 2026). Binary classification was implemented using a custom ResNet model (He *et al.*, 2016), building on the successful application of ResNet-based models to the development of acoustic detectors for other marine mammal species (Bergler *et al.*, 2019; Jeantet and Dufourq, 2023; Kirsebom *et al.*, 2020). The network comprised 15 convolutional layers with two residual connections, batch normalization (momentum = 0.99), and ReLU activation. An initial 3×3

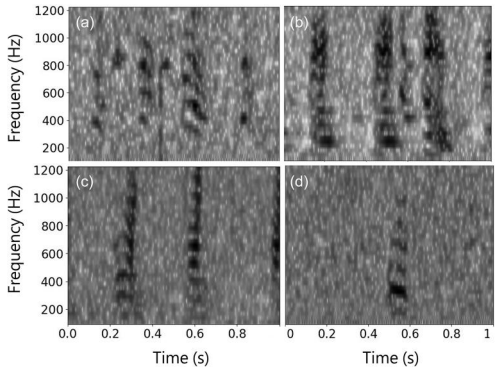

Fig. 2. Examples of normalized positive spectrograms with tonal noise reduced. Panels (a) and (b) display spectrograms containing bark-yelp sequences, and panels (c) and (d) display spectrograms containing only barks. Spectrograms are of dimensions 750 time bins × 31 frequency bins × 1 intensity channel. Short-time Fourier transform parameters: 3000 Hz downsample rate, 0.032 s window rize, 0.001 s window step, Hamming window function, frequency range 100-1200 Hz.

kernel with 16 filters was used, followed by an average pooling layer, a fully connected layer, and a softmax output. Training was performed in TensorFlow using an Adam optimizer (learning rate = 0.001, decay = 0.1) and an F-score-based loss function (Frazao *et al.*, 2024).

Training was conducted on a desktop computer equipped with a 12th Gen Intel® Core™ i9-12900 processor (2.40 GHz; Intel, Santa Clara, CA), 16 GB of RAM, and an NVIDIA GeForce RTX 3080 GPU (NVIDIA, Santa Clara, CA). Models were trained for 40 epochs using a batch size of 16, which was sufficient for the accuracy, precision, recall, and loss metrics to plateau. To account for the stochasticity in the training process, an ensemble of ten structurally identical models was trained, with variation introduced solely via unique global random seeds in TensorFlow to govern layer weight initialization and batch shuffling. The detector used all ten models to process each input, leveraging their consensus to reduce the variance of individual random realizations, and the final classification score was calculated as the median across the outputs. For deployment on continuous audio data, files were automatically segmented into 1 s windows with 50% overlap during processing. Overlapping or adjacent detections were merged using median scores (Fig. 3), and positive detections were retained based on a user-defined confidence threshold. The classification threshold defines the boundary between the positive and negative classes. Inputs with classification scores greater than or equal to the selected threshold are classified as positive, whereas inputs with lower scores are classified as negative. Adjusting this threshold changes the balance between precision and recall.

2.5 Model performance evaluation and visualization

The classification performance of the detector is summarized in terms of precision, recall, and F1 score, defined as

$$\text{precision} = \frac{TP}{TP + FP}, \tag{1}$$

$$\text{recall} = \frac{TP}{TP + FN}, \tag{2}$$

$$\text{F1 score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}, \tag{3}$$

Table 1. Summary of datasets used for model development (training and validation; "Dev") and testing ("Test"). Continuous datasets were supplied as unsegmented audio recordings rather than pre-extracted segments and were automatically segmented during processing.

Type	Dataset	Site[a]	WAV files	Hours	Segments	Dev/Test
Segmented	A	Cape Bathurst	78	6.50	380	Dev
Segmented	A	Cambridge Bay	636	32.41	3848	Dev
Segmented	A	Ulukhaktok 2017-19	133	11.08	1910	Dev
Segmented	A	Ulukhaktok 2022	1080	90.0	2952	Dev
Segmented	B	Pearce Point	17	1.42	168	Test
Segmented	C	Ulukhaktok 2023	125	39.6	1022	Test
Continuous	D	Pearce Point	20	1.67	N/A	Test
Continuous	E	Ulukhaktok 2023	40	3.33	N/A	Test

[a]All deployment years per site apply if no year is indicated.

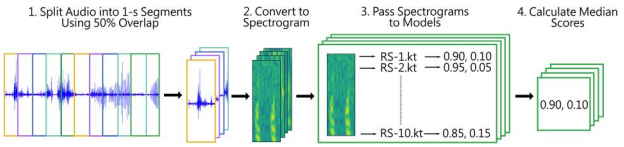

Fig. 3. Processing pipeline for continuous audio data (".kt" is the extension used for trained model files created using Ketos).

where *TP*, *FP*, and *FN* are the number of true positives, false positives, and false negatives, respectively. True positives and true negatives represent correctly classified inputs, while false positives and false negatives denote misclassifications.

Uniform Manifold Approximation and Projection (UMAP) (McInnes *et al.*, 2020) was used to visualize the feature space learned by the detector. Feature representations were extracted for each pre-labeled segment and projected into two dimensions using UMAP. Each point in the resulting embedding represents one input segment and was labeled according to its known class and predicted classification outcome. UMAP was used as a qualitative visualization of the learned feature space rather than as a quantitative validation of clusters or class separability.

3. Results

3.1 Classification performance

Figure 4 illustrates precision and recall curves alongside UMAP dimensionality reduction of features extracted for Datasets A–C. For the validation subset of Dataset A, the detector achieved balanced performance at a threshold of 0.45, with a precision of 0.924, recall of 0.926, and F1 score of 0.925. High performance was maintained across testing datasets, with Datasets B and C (Pearce Point and Ulukhaktok 2023) achieving F1 scores of 0.970 and 0.932, respectively, at a 0.55 threshold. The features learned by the ResNet architecture were visualized by extracting the output of the final average pooling layer. These high-dimensional feature arrays were projected into two dimensions using the UMAP algorithm with nine nearest neighbors and a minimum distance of 0.7. As illustrated in Fig. 4, the results show distinct spatial clustering of positive and negative classes across all datasets. However, localized overlap within these clusters likely reflects similarities among low-SNR ringed seal vocalizations, acoustically similar vocalizations from other species, and non-biological ambient noise, which may contribute to the precision-recall trade-off observed during deployment.

3.2 Detection performance on continuous audio data

The detector was deployed on the continuous audio data in Datasets D (Pearce Point) and E (Ulukhaktok 2023). The output detections were visualized in Raven Pro, and manually classified as true or false positives, with their score being the median of the median scores for each 1 s window within the detection. The continuous files were also reviewed to identify potentially missed ringed seal barks. This allowed false negatives and recall to be estimated for these test recordings. Table 2 presents the total number of true and false classifications, the number of false negatives, and the precision and recall at each confidence threshold for Dataset D. At a confidence threshold of 0.9, the network achieves a precision of 0.436 and recall of 0.944 for Dataset D. Table 3 presents the same metrics for Dataset E. At a confidence threshold of 0.9, the network achieves a precision of 0.532 and recall of 0.979.

4. Discussion

The deep learning-based detector successfully identified ringed seal bark and bark-yelp sequences across PAM deployments and diverse environments in the Western Canadian Arctic. On manually verified segmented spectrograms (Datasets A–C), the model achieved high performance (precision, recall, F1 score >0.90), demonstrating that the ResNet architecture effectively learned the distinct features necessary for class separation. Notably, the model generalized well to seen and unseen locations within these curated datasets, maintaining high metrics even at Pearce Point despite no data from this region being included in model development.

Evaluations on continuous audio revealed a drop in precision (0.436 at Pearce Point, 0.532 at Ulukhaktok), although recall remained high (>0.94). This discrepancy highlights the limitations of training on "cleaned" 1 s segments, which exclude low-SNR signals or overlapping sounds to avoid mislabeling. In continuous audio, the detector encountered novel sound sources, including bowhead whale vocalizations (20–5000 Hz; most social calls approximately 1 s, but some songs up to 146 s; Cummings and Holliday, 1987), that were absent from the training set. This indicates that while the model extrapolates effectively within its training domain, performance degrades when faced with increased environmental variability (Meyer and Pebesma, 2021). In this case, high recall was prioritized to ensure that rare vocalizations in low-density areas were not missed (important for ecological interpretations), accepting the requirement for a higher manual verification workload for false positives. False positives were primarily driven by non-target sounds with similar frequency signatures, such as bearded seal vocalizations, scratching noises from ringed seals, and water-movement sounds. Additionally, the detector processes segments individually and therefore it lacks the temporal context, such as surrounding aural information, that human analysts use to differentiate species. Incorporating such context into future architectures could significantly reduce these errors. For broad monitoring applications where missed detections can be costly, a

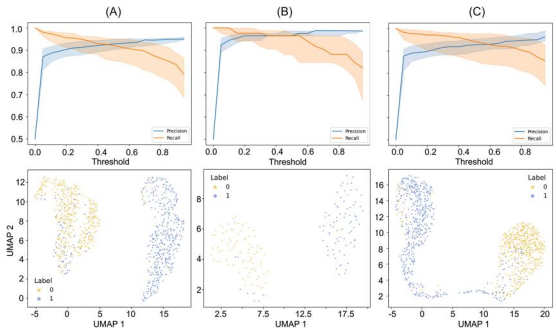

Fig. 4. Results for datasets A (Development), B (Pearce Point), and C (Ulukhaktok 2023). Top panels: Precision (blue) and recall (orange) curves against confidence threshold. Shaded regions represent the 10th and 90th percentiles, and the solid lines represent the median of the 10 model scores. Bottom panels: UMAP dimensionality reduction of features extracted for Datasets A–C. Coloring is set by the annotated label for the input data point, where 0 (orange) represents negative spectrograms not containing ringed seal vocalizations, and 1 (blue) represents positive spectrograms containing ringed seal vocalizations.

relatively low detection threshold (e.g., 0.5) may be preferable to maximize the number of candidate detections, followed by manual verification. Higher thresholds can reduce the number of detections requiring manual review but may increase the risk of missed detections, which can be detrimental in some applications.

Future work with more available training data could compare models trained with and without tonal-noise reduction to determine whether retaining these background features improves generalization. Future developments could also address spatial and temporal imbalances in the training data, as the over-representation of specific sites (e.g., Ulukhaktok) and months (May) may bias predictions by limiting generalization to underrepresented or out-of-distribution conditions (Lauha *et al.*, 2022; van Merriënboer *et al.*, 2024). As more acoustic data are collected, retraining with a more balanced and noise-diverse dataset could improve the detector's precision and generalizability.

The large volume of acoustic data that is often collected by acoustic monitoring programs renders manual analysis infeasible. This challenge is further compounded by analyst fatigue and differing annotation procedures. In contrast, the automated detector developed in this study processed 3.3 h of data (2.14 GB) in approximately 6 min using a GPU workflow. Extrapolated to a full year of data collection, the detector could analyze 7 TB in roughly 14 days. This represents a significant increase in processing speed, even when accounting for the time required to verify false positives, and provides a practical means of keeping pace with annual data intake.

Developing deep learning-based tools, such as automated detectors, typically requires an experimental and iterative approach (Sarker, 2021). While the current detector achieved strong results with limited data, performance is expected to improve through a continuous feedback loop during ongoing use, with model predictions employed to identify misclassified patterns (e.g., specific noise sources) and incorporating those into future training sets. Initial domain-adaptation tests using fine-tuning at Pearce Point produced only a marginal reduction in the number of false positives, suggesting that larger volumes of site-specific training data may be required to improve generalizability.

A primary constraint of the current ResNet architecture is its inability to utilize temporal context. Unlike human analysts who rely on surrounding acoustic information to distinguish ambiguous signals, this model classifies segments in isolation. Future development should explore hybrid architectures such as CNN-LSTMs (CNN-long short-term memory), which leverage multiple time scales to identify sequential patterns in marine mammal vocalizations (Madhusudhana *et al.*, 2021). Furthermore, performance could be improved by adopting insights from similar PAM applications. For instance, testing transfer learning from recent open-source fish grunt detectors (Mouy *et al.*, 2024) or experimenting with class-imbalanced datasets may improve the model precision. Future work could also extend the detector to distinguish additional ringed seal vocalization types and other Arctic marine mammal species.

Table 2. Number of false positives, true positives, false negatives, and corresponding performance metrics for Dataset D (continuous Pearce Point data).

Threshold	FP (#)	FP (%)	TP (#)	TP (%)	Total	FN	Precision	Recall	F1 Score
0.5	152	63.33	88	36.67	240	4	0.367	0.957	0.530
0.6	140	62.78	83	37.22	223	4	0.372	0.954	0.535
0.7	127	62.25	77	37.75	204	4	0.377	0.951	0.540
0.8	99	59.64	67	40.36	166	3	0.404	0.957	0.568
0.9	66	56.41	51	43.59	117	3	0.436	0.944	0.596

Table 3. Number of false positives, true positives, false negatives, and corresponding performance metrics for Dataset E (continuous Ulukhaktok 2023 data).

Threshold	FP (#)	FP (%)	TP (#)	TP (%)	Total	FN	Precision	Recall	F1 Score
0.5	344	54.43	288	45.57	632	7	0.456	0.976	0.621
0.6	306	52.94	272	47.06	578	7	0.471	0.975	0.635
0.7	263	50.38	259	49.62	522	7	0.496	0.974	0.657
0.8	215	47.25	240	52.75	455	6	0.527	0.976	0.685
0.9	162	46.82	184	53.18	346	4	0.532	0.979	0.689

5. Conclusions

This study successfully developed a deep learning–based detector capable of identifying ringed seal vocalizations (barks and bark-yelp sequences) across diverse environments in the Western Canadian Arctic. By processing 100 GB of data in approximately 5 h, the tool substantially reduces the volume of recordings requiring manual review, thereby supporting more efficient large-scale acoustic monitoring. Although performance on continuous recordings reflected a trade-off between high recall (>0.90) and lower precision (0.43–0.53), largely due to false positives caused by acoustically similar biological vocalizations and non-biological ambient sounds, the detector remains a practical tool for prioritizing recordings for human verification. This capacity to efficiently process large acoustic datasets can support analyses of ringed seal spatial and temporal occurrence, which are important for understanding Arctic ecosystems and assessing environmental change. By identifying recordings likely to contain ringed seal vocalizations, the detector can help researchers examine patterns in acoustic occurrence across locations and time periods and prioritize data for further analyses of environmental and anthropogenic influences. The detector therefore provides a foundation for more efficient long-term acoustic monitoring of ringed seals in the Arctic.

Supplementary Material

See supplementary material for an analysis of classification uncertainty across the ensemble models; examples of tonal-noise reduction; annotation proportions; tabular results for precision, recall, and F1 scores for Datasets A–C; and recorder metadata for all study locations.

Acknowledgments

This data collection for this research took place in the Inuvialuit Settlement Region and the Cambridge Bay region. The authors are grateful to the Inuvialuit and Inuit for their assistance and for allowing this research to be conducted on their traditional lands. *Koana, quyanainni*, thank you. We also acknowledge and respect the Lək̓ʷəŋən Peoples on whose territory the University of Victoria stands (where model development was completed), and the Lək̓ʷəŋən and W̱SÁNEĆ Peoples whose historical relationships with the land continue to this day. We thank data collectors and local community partners Tony Alanak, Shane Alikamik, Roy Inuktalik, Adam Kudlak, and Humfrey Melling. We also thank Andrea Niemi, Svein Vagle, and Bill Williams for access to data and manuscript review. Our gratitude is also extended to the captain and crew of the CCGS Sir Wilfrid Laurier. Finally, we acknowledge the project funders for their support: Mitacs (Grant No. IT30379), Fisheries and Oceans Canada (Grant No. 2022-NF-OPR-011), DFO through the Canada Nature Fund for Aquatic Species at Risk, Transport Canada, the Inuvialuit Fisheries Joint Management Committee, and the Weston Family Foundation.

Author Declarations

Conflict of Interest

The authors have no conflicts to disclose.

Data Availability

The data that support the findings of this study are openly available in an open-source command-line interface version of the detector at https://github.com/karlzam/ringed-seal-detector. A user guide and tutorial are also available at https://ringed-seal-detector-docs.readthedocs.io/en/latest/. The annotated ringed seal data are available from the corresponding author upon reasonable request.

References

Barbosa, M., Halliday, W. D., Insley, S. J., and Dosso, S. E. (**2024**). "Characterisation of the ringed seal (*Pusa Hispida*) acoustic repertoire during spring in the Western Canadian Arctic," Bioacoustics **33**(5), 433–447.

Bergler, C., Schröter, H., Cheng, R. X., Barth, V., Weber, M., Nöth, E., Hofer, H., and Maier, A. (**2019**). "ORCA-SPOT: An automatic killer whale sound detection toolkit using deep learning," Sci. Rep. **9**, 10997.

Cai, W., Zhu, J., Zhang, M., and Yang, Y. (**2022**). "A parallel classification model for marine mammal sounds based on multi-dimensional feature extraction and data augmentation," Sensors **22**(19), 7443.

Calvert, W., and Stirling, I. (**1985**). "Winter distribution of ringed seals (*Phoca hispida*) in the Barrow Strait area, Northwest Territories, determined by underwater vocalizations," Can. J. Fisheries Aquatic Sci. **42**(7), 1238–1243.
Clark, C. W., Ellison, W. T., Southall, B. L., Hatch, L., Van Parijs, S. M., Frankel, A., and Ponirakis, D. (**2009**). "Acoustic masking in marine ecosystems: Intuitions, analysis, and implication," Mar. Ecol. Prog. Ser. **395**, 201–222.
Comiso, J. C. (**2012**). "Large decadal decline of the Arctic multiyear ice cover," J. Clim. **25**, 1176–1193.
COSEWIC (**2019**). "COSEWIC assessment and status report on the Ringed Seal Pusa hispida in Canada" (Committee on the Status of Endangered Wildlife in Canada, Ottawa, Canada).
Cummings, W. C., and Holliday, D. V. (**1987**). "Sounds and source levels from bowhead whales off pt. Barrow, Alaska," J. Acoust. Soc. Am. **82**(3), 814–821.
Dawson, J., Pizzolato, L., Howell, S. E. L., Copland, L., and Johnston, M. E. (**2018**). "Temporal and spatial patterns of ship traffic in the Canadian Arctic from 1990 to 2015," Arctic **71**(1), 15–26.
Dingwall, J. T., Halliday, W. D., Diogou, N., Niemi, A., Steiner, N., and Insley, S. J. (**2024**). "The Arctic marine soundscape of the Amundsen Gulf, Western Canadian Arctic," Mar. Pollut. Bull. **204**, 116510.
Erbe, C., and Farmer, D. M. (**2000**). "Zones of impact around icebreakers affecting beluga whales in the Beaufort Sea," J. Acoust. Soc. Am. **108**(3), 1332–1340.
Frazao, F., Kirsebom, O., Padovese, B., and Matwin, S. (**2024**). "Ketos: Fscore loss," https://docs.meridian.cs.dal.ca/ketos/generated/ketos.neural_networks.dev_utils.losses.FScoreLoss.html (Last viewed 2026-08-17).
Halliday, W. D., Brittain, S. A., Niemi, A., Majewski, A. R., Mouy, X., and Insley, S. J. (**2022**). "The underwater soundscape of Minto Inlet, Northwest Territories, Canada," Arctic **75**(4), 462–479.
Halliday, W. D., Insley, S. J., de Jong, T., and Mouy, X. (**2017**). "Seasonal patterns in acoustic detections of marine mammals near Sachs Harbour, Northwest Territories," Arct. Sci. **4**(3), 259–278.
Halliday, W. D., Pine, M., Insley, S., Soares, R., Kortsalo, P., and Mouy, X. (**2019**). "Acoustic detections of Arctic marine mammals near Ulukhaktok, Northwest Territories, Canada," Can. J. Zool. **97**(1), 72–80.
Halliday, W. D., Pine, M. K., and Insley, S. J. (**2020a**). "Underwater noise and Arctic marine mammals: Review and policy recommendations," Environ. Rev. **28**(4), 438–448.
Halliday, W. D., Pine, M. K., Mouy, X., Kortsalo, P., Hilliard, R. C., and Insley, S. J. (**2020b**). "The coastal Arctic marine soundscape near Ulukhaktok, Northwest Territories, Canada," Polar Biol. **43**(6), 623–636.
Hannay, D. E., Delarue, J., Mouy, X., Martin, B. S., Leary, D., Oswald, J. N., and Vallarta, J. (**2013**). "Marine mammal acoustic detections in the northeastern Chukchi Sea, September 2007–July 2011," Cont. Shelf Res. **67**, 127–146.
He, K., Zhang, X., Ren, S., and Sun, J. (**2016**). "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778.
Jeantet, L., and Dufourq, E. (**2023**). "Improving deep learning acoustic classifiers with contextual information for wildlife monitoring," Ecol. Inf. **77**, 102256.
Jones, J. M., Thayre, B. J., Roth, E. H., Mahoney, M., Sia, I., Merculief, K., Jackson, C., Zeller, C., Clare, M., Bacon, A., Weaver, S., Gentes, Z., Small, R. J., Stirling, I., Wiggins, S. M., and Hildebrand, J. A. (**2014**). "Ringed, bearded, and ribbon seal vocalizations north of Barrow, Alaska: Seasonal presence and relationship with sea ice," Arctic **67**(2), 203–222.
Kirsebom, O. S., Frazao, F., Simard, Y., Roy, N., Matwin, S., and Giard, S. (**2020**). "Performance of a deep neural network at detecting North Atlantic right whale upcalls," J. Acoust. Soc. Am. **147**(4), 2636–2646.
Lauha, P., Somervuo, P., Lehikoinen, P., Geres, L., Richter, T., Seibold, S., and Ovaskainen, O. (**2022**). "Domain-specific neural networks improve automated bird sound recognition already with small amount of local data," Methods Ecol. Evol. **13**(12), 2799–2810.
Madhusudhana, S., Shiu, Y., Klinck, H., Fleishman, E., Liu, X., Nosal, E., Helble, T., Cholewiak, D., Gillespie, D., Širović, A., and Roch, M. A. (**2021**). "Improve automatic detection of animal call sequences with temporal context," J. R Soc. Interface **18**(180), 20210297.
Markus, T., Stroeve, J. C., and Miller, J. (**2009**). "Recent changes in Arctic sea ice melt onset, freezeup, and melt season length," J. Geophys. Res. Oceans **114**, C12024, https://doi.org/10.1029/2009JC005436.
McInnes, L., Healy, J., and Melville, J. (**2020**). "UMAP: Uniform manifold approximation and projection for dimension reduction," arXiv:1802.03426.
MERIDIAN (**2026**). "Ketos: Meridian Python package for deep-learning based acoustic detectors and classifiers," https://git-dev.cs.dal.ca/meridian/ketos (Last viewed May 19, 2026).
Meyer, H., and Pebesma, E. (**2021**). "Predicting into unknown space? Estimating the area of applicability of spatial prediction models," Methods Ecol. Evol. **12**(9), 1620–1633.
Mouy, X., Archer, S. K., Dosso, S., Dudas, S., English, P., Foord, C., Halliday, W., Juanes, F., Lancaster, D., Parjis, S. V., and Haggarty, D. (**2024**). "Automatic detection of unidentified fish sounds: A comparison of traditional machine learning with deep learning," Front. Remote Sens. **5**, 1439995.
Padovese, B., Kirsebom, O. S., Frazao, F., Evers, C. H. M., Beslin, W. A. M., Theriault, J., and Matwin, S. (**2023**). "Adapting deep learning models to new acoustic environments - A case study on the North Atlantic right whale upcall," Ecol. Inf. **77**, 102169.
Sarker, I. H. (**2021**). "Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions," SN Comput. Sci. **2**(6), 420.
Stirling, I. (**1973**). "Vocalization in the ringed seal (*Phoca hispida*)," J. Fish. Res. Board Can. **30**(10), 1592–1594.
Stirling, I., Calvert, W., and Cleator, H. (**1983**). "Underwater vocalizations as a tool for studying the distribution and relative abundance of wintering pinnipeds in the High Arctic," Arctic **36**(3), 262–274.
Stirling, I., and Derocher, A. E. (**2012**). "Effects of climate warming on polar bears: A review of the evidence," Global Change Biol. **18**(9), 2694–2706.
van Merriënboer, B., Hamer, J., Dumoulin, V., Triantafillou, E., and Denton, T. (**2024**). "Birds, bats and beyond: Evaluating generalization in bioacoustics models," Front. Bird Sci. **3**, 1369756.