7194

Classification of Herring, Salmon, and Bubbles in Multifrequency Echograms with a U-Net Neural Network

Alex L. Slonimer[1, 3], Stan E. Dosso[1], Alexandra Branzan Albu[2], Melissa Cote[2], Tunai Porto Marques[2], Alireza Rezvanifar[2], Steve Pearce[3], and Stephane Gauthier[4]

[1]School of Earth and Ocean Science, University of Victoria, [2]Electrical and Computer Engineering, University of Victoria, [3]ASL Environmental Sciences, Victoria BC, [4]Fisheries and Oceans Canada, Victoria BC,

Emails: aslonimer@aslenv.com, sdosso@uvic.ca, aalbu@uvic.ca, mcote@uvic.ca, tunaip@uvic.ca, arezvani@uvic.ca, spearce@aslenv.com, stephane.gauthier@dfo-mpo.gc.ca

ASL Environmental Sciences

1. Introduction

The manual classification of fish within acoustic echograms is possible based on morphological characteristics resulting from the behaviour of fish schools (scattering strength, density, size, etc) and scattering relationships between acoustic frequencies [1]. However, manual classification is time consuming, prone to analyst error, and complicated by the presence of noise.

The primary goal of this study is to automate the classification of salmon (sockeye and chum, *Oncorhynchus nerka* and *O. keta*), Pacific herring (*Clupea pallasii*), air bubbles, and the sea surface in echogram images. Accurately monitoring these fish species is important for maintaining their role in ecosystems and the economy [2]. Bubbles and the sea surface are included to improve confidence and aid in manual verification of the results. The automated classification was accomplished using a U-Net[3] convolutional neural network (CNN).

2. AZFPs

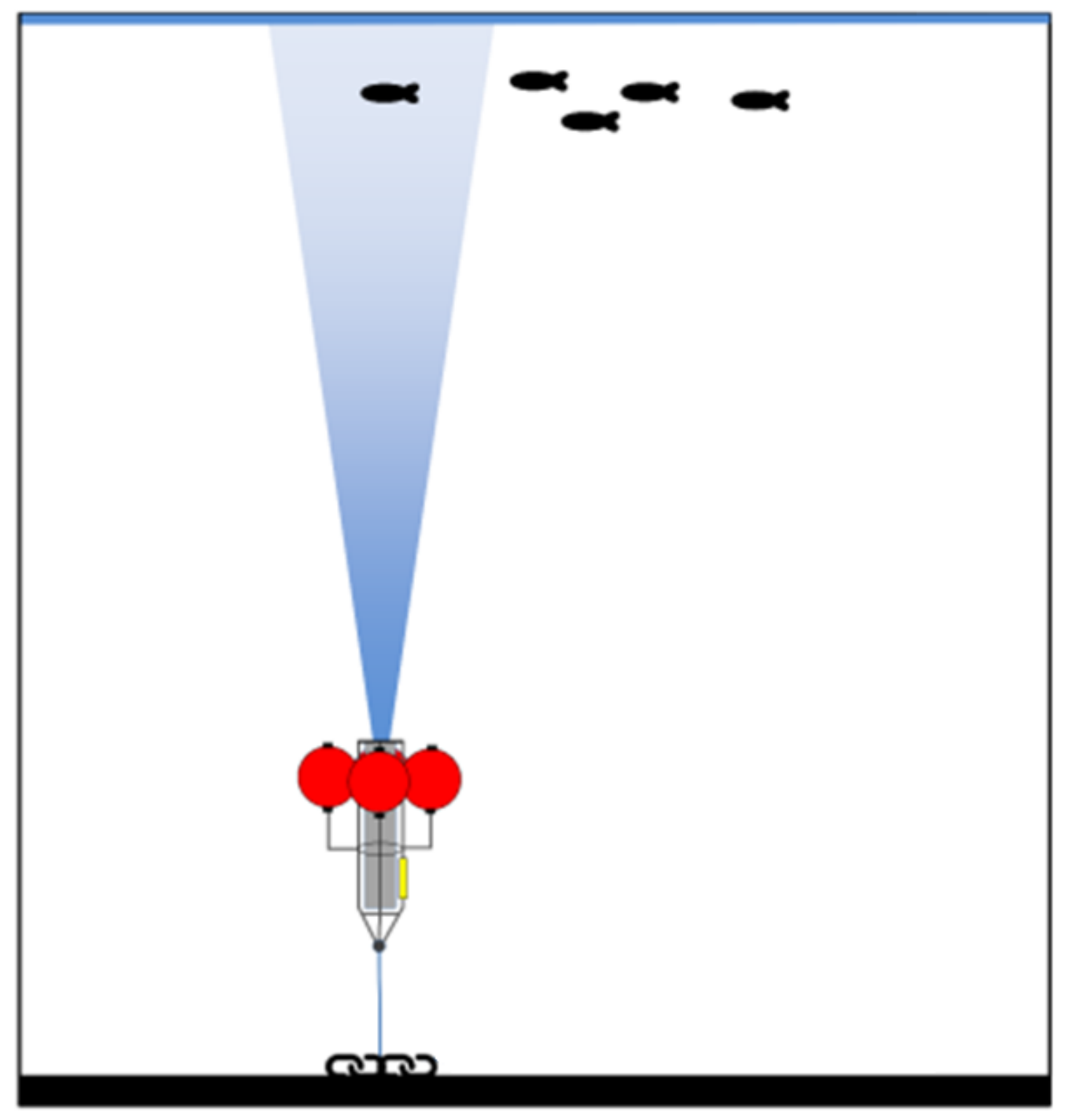


Figure 1. The data used in this study were collected by upward facing multi-frequency Acoustic Zooplankton Fish Profiler (AZFP) echosounders. Echosounders use a narrow acoustic beam to vertically profile the water column and measure volume backscatter, S_v, from biological and physical phenomena.

3. Location

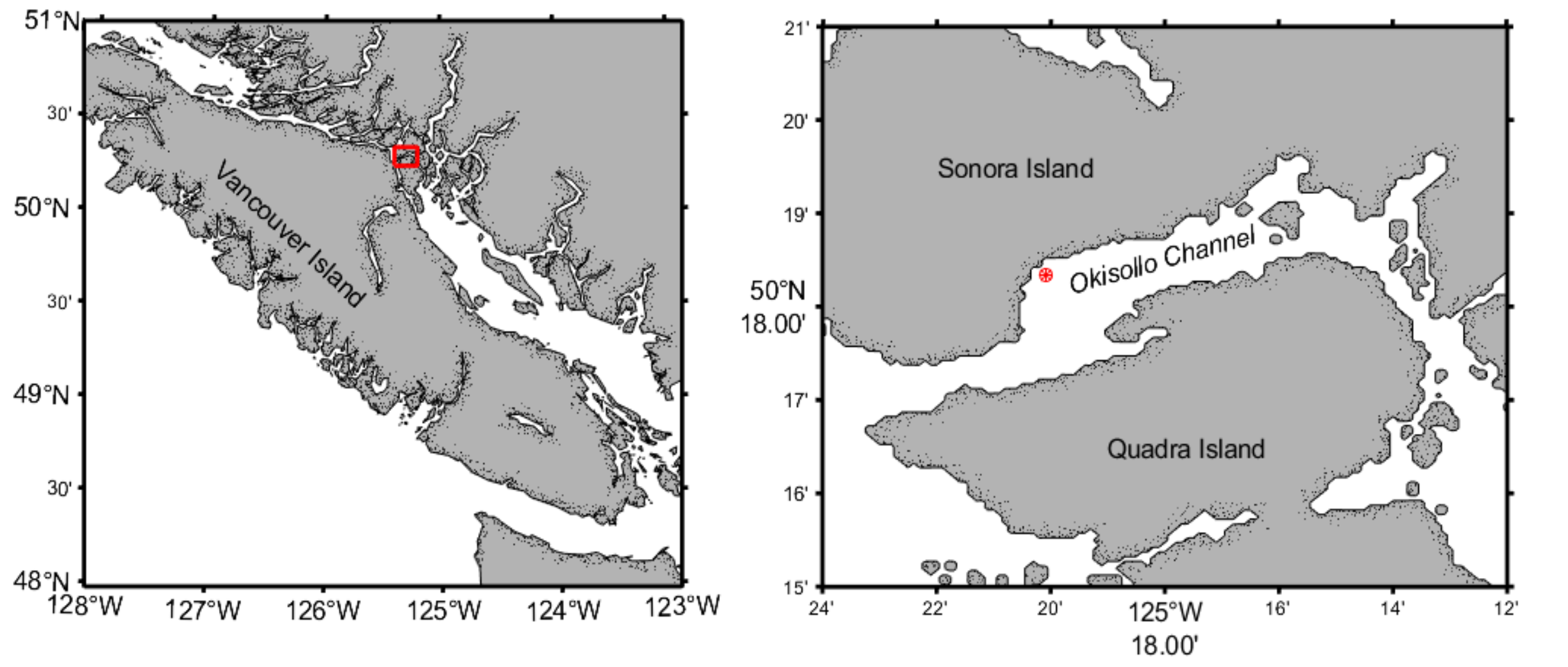


Figure 2. The AZFPs were moored on the sea-floor, during a multi-year study by Fisheries and Oceans Canada[4] to study juvenile salmon migration in Okisolla Channel on the central coast of British Columbia, Canada.

4. Model Inputs

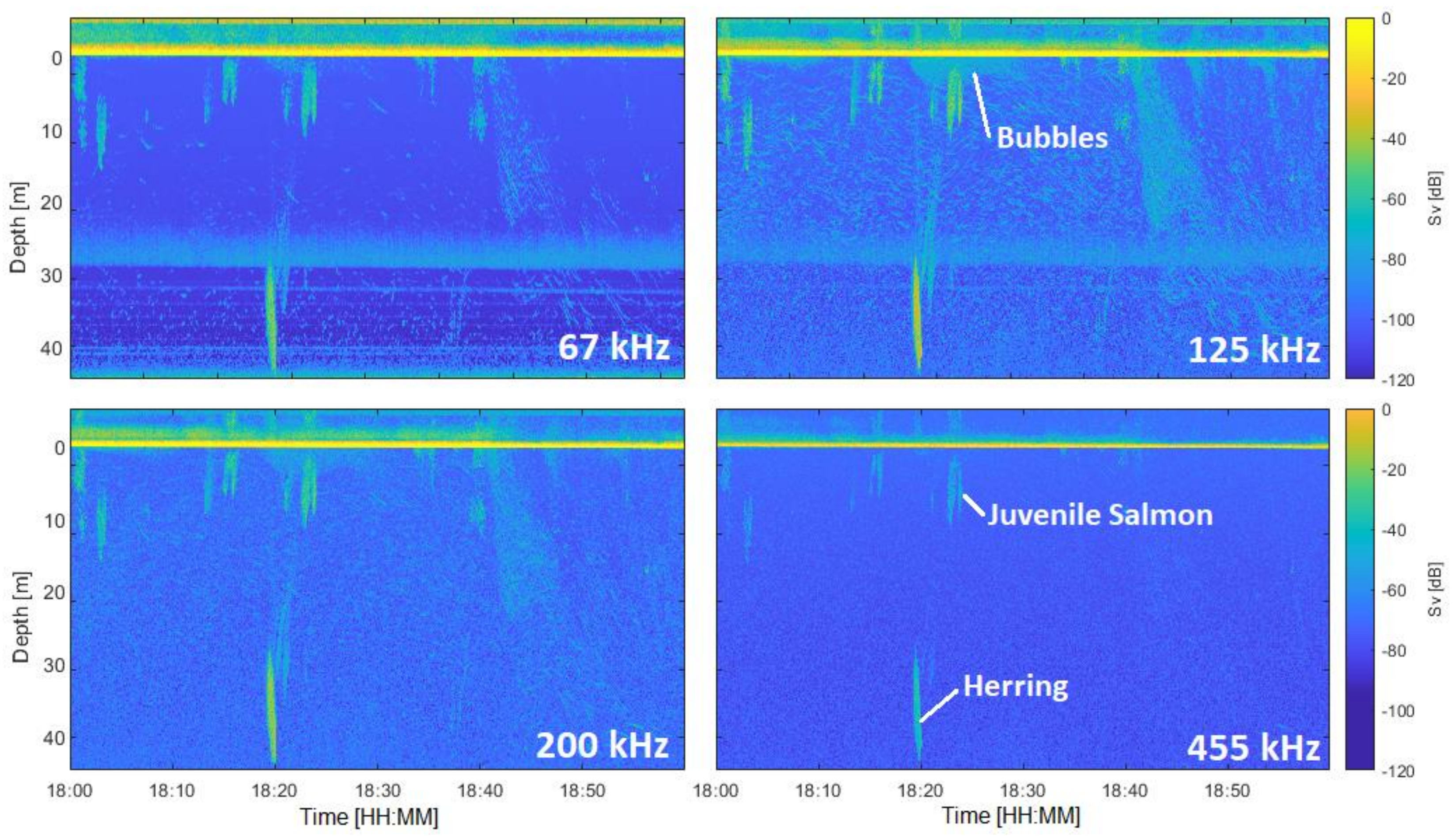


Figure 3. The AZFPs were equipped with 67, 125, 200, and 455 kHz transducers. Profiles were collected in bursts, initiated every 3.0 s, when each transducer would ping in sequence. These profiles are concatenated into a sequence of continuous time series and viewed as an image known as an echogram. The profiles were recorded at a range resolution of 9.31 cm. The variability in morphology and material properties amongst species results in multi-frequency scattering signatures that can be used to differentiate them [5].

Each acoustic frequency is sensitive to certain sizes of particles and different material properties, as visible in the four panels. Air bubbles scatter most intensely at 125 kHz [6], whereas salmon and herring have a more consistent strength across frequencies. Juvenile salmon group in loose aggregations, generally in the upper 25 m of the water column, whereas herring school in dense narrow groups generally below 10 m depth. These characteristics are used to manually annotate the data which is used to train the CNNs to recognize each class.

In addition to the four echogram frequency channels, two simulated context channels for water depth and solar elevation angle (a proxy for sunlight) are provided as input to the U-Net. These provide spatial and temporal context.

5. U-Net CNNs

CNNs treat echograms like images, and are able to "learn" morphological features like shape, texture, and relationships between input channels [7]. A U-Net style CNN was used because it is capable of producing pixel-level classification with a small training set. The four (or six) channels of data are fed into the model, and the output is 5 channels corresponding to the herring, salmon, bubbles, sea-surface, and background class. Convolutional kernels are activated by different features in the images, and subsequent layers of convolutions learn to recognize more complex features in the data. The training data set consisted of 71 one-hour echograms, and the test set consisted of 31 echograms.

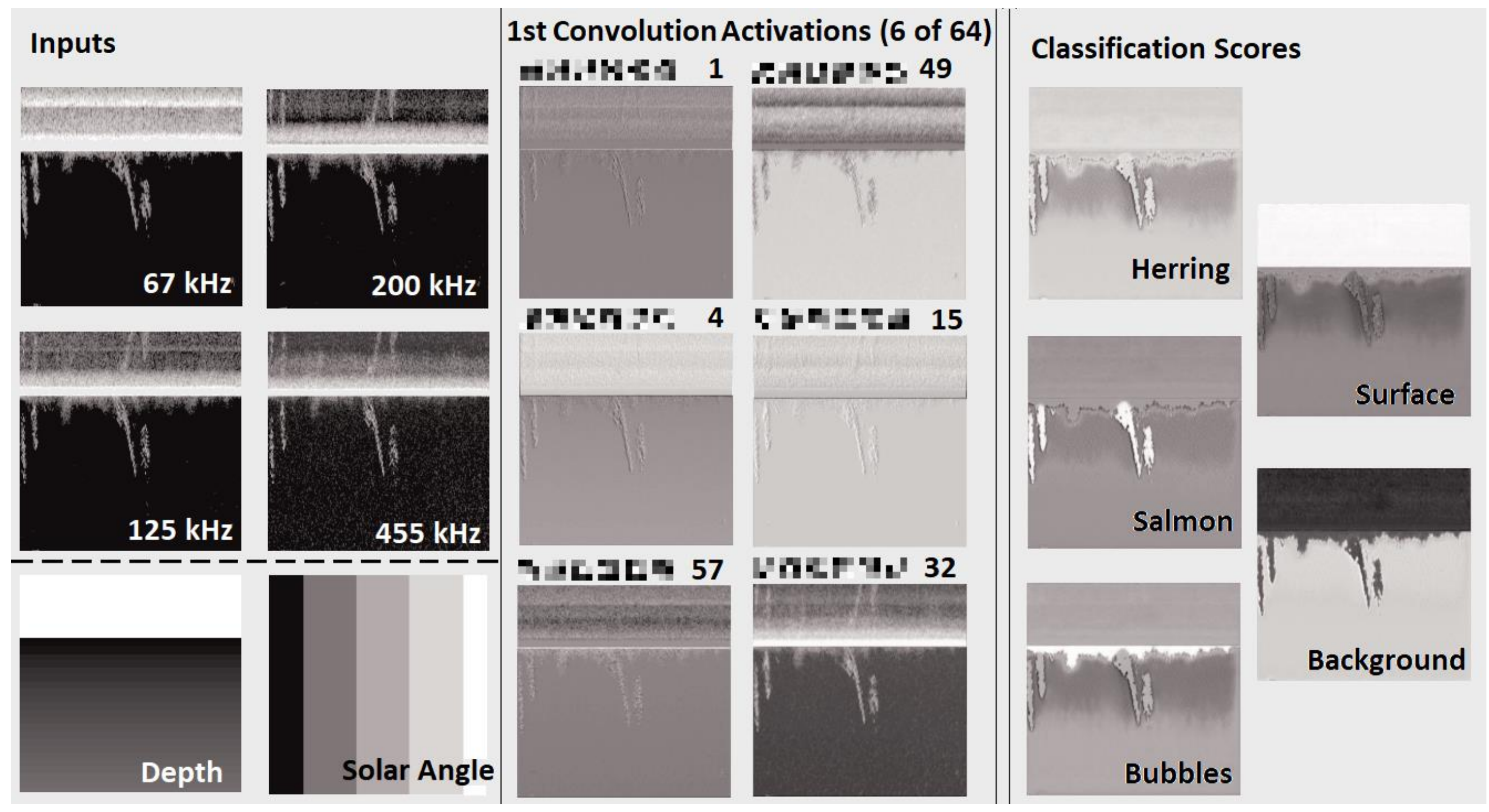


Figure 4. Left to right: (i) Inputs to the U-Net, tiles of size 256 x 256 pixels, (ii) Example activation from six filters at the first convolution layer of the U-Net, (iii) Classification scores output by the U-Net. For (ii) and (iii), bright pixels indicate strong activations, or high scores for a given class.

6. Classification

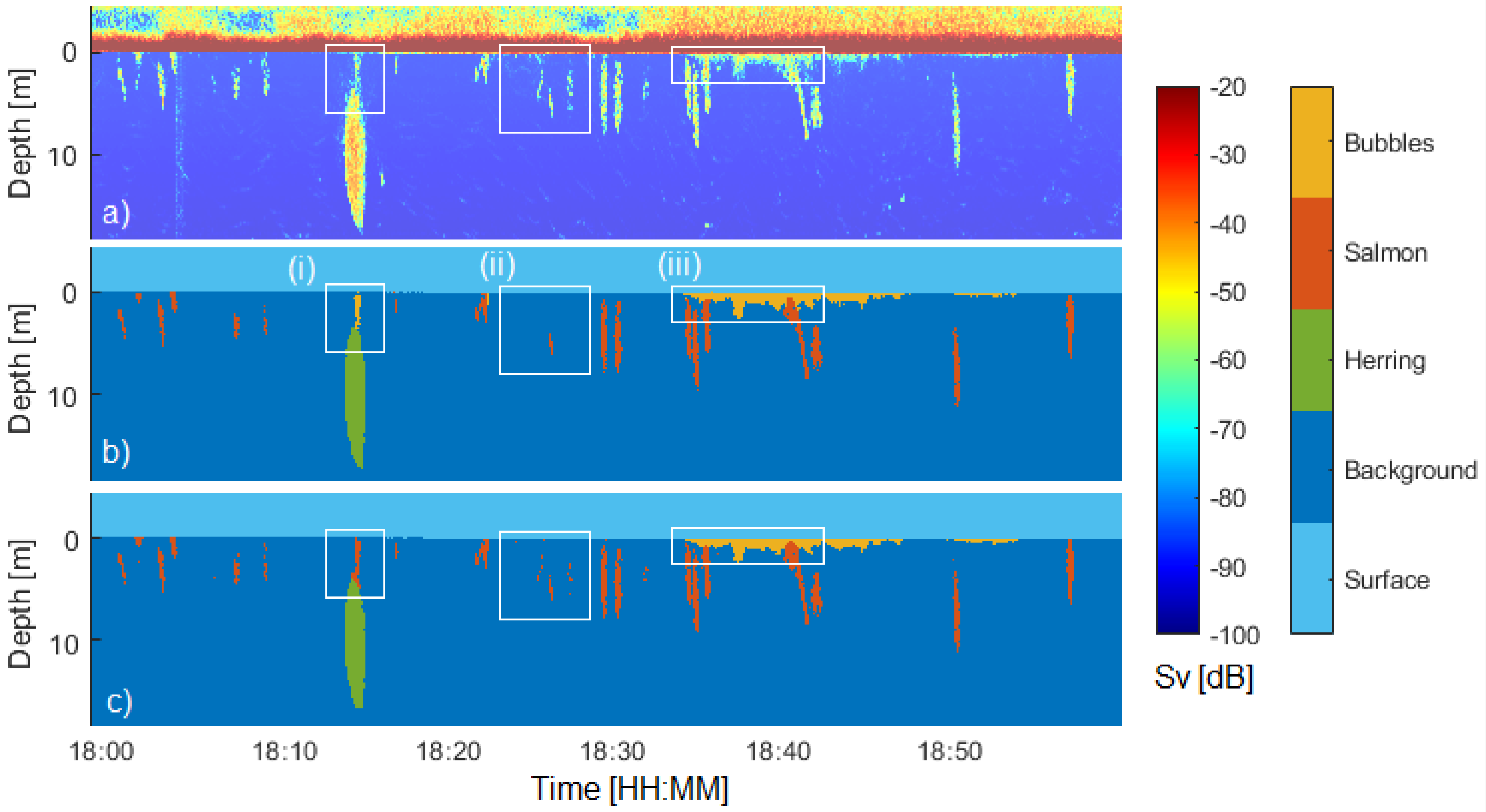
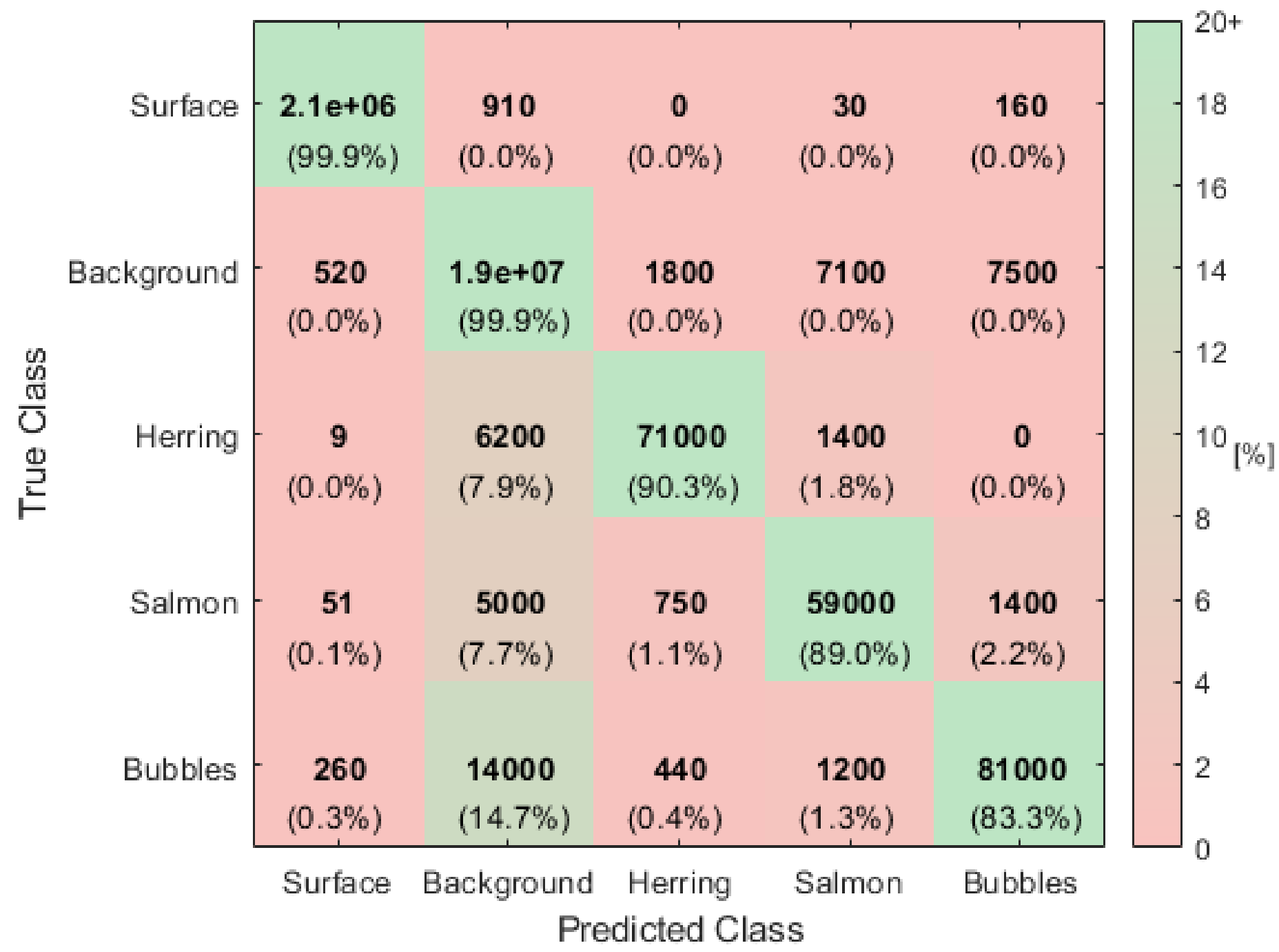


Figure 5. (a) corresponds to the 67 kHz S_v data, **(b)** is the manual annotations of the data **(c)** is the classifications output by the U-Net. This example illustrates three areas where the U-Net has performed remarkable well.
(i) The scattering group above this herring school is annotated as bubbles but the U-Net predicts this as salmon. This is a scenario where the U-Net has out-performed the annotation, such that the manual interpretation was incorrect.
(ii) Salmon are predicted by the U-Net but annotated manually as background. The manual annotation ignored groups less than 6 pixels, so the U-Net out-performs the manual annotation.
(iii) Aggregations of salmon are in contact with bubbles at the surface. The U-Net can accurately predict the boundary between the two classes.

7. Results

The classification is quantified with the F_1-score, a harmonic mean between precision and recall. The six-channel U-Net classified 31 echograms with **F_1-scores of 0.865 for bubbles, 0.873 for salmon and 0.930 for herring**. This outperforms the F_1-score of the four-channel U-Net by 0.01, 0.03, and 0.02, for bubbles, salmon, and herring, respectively.

To get a clearer understanding of the performance, we look to the confusion matrix. This shows how well the model has predicted each class (true positives on the diagonal) and which classes have issues with mis-classification (off-diagonal). Salmon are very rarely misclassified as herring, and vice-versa. Salmon are also rarely classified as bubbles, which is very challenging to do manually. The biggest issue is the number of false negatives of salmon and herring that have been misclassified as background.



True Class \ Predicted Class	Surface	Background	Herring	Salmon	Bubbles
Surface	**2.1e+06** (99.9%)	**910** (0.0%)	**0** (0.0%)	**30** (0.0%)	**160** (0.0%)
Background	**520** (0.0%)	**1.9e+07** (99.9%)	**1800** (0.0%)	**7100** (0.0%)	**7500** (0.0%)
Herring	**9** (0.0%)	**6200** (7.9%)	**71000** (90.3%)	**1400** (1.8%)	**0** (0.0%)
Salmon	**51** (0.1%)	**5000** (7.7%)	**750** (1.1%)	**59000** (89.0%)	**1400** (2.2%)
Bubbles	**260** (0.3%)	**14000** (14.7%)	**440** (0.4%)	**1200** (1.3%)	**81000** (83.3%)

Figure 6. Confusion matrix for 31 one hour echograms for the six-channel U-Net. Bold values represent the total number of pixels, and the percent corresponds to the amount of true pixels per class.

8. Conclusions

- U-Nets enable high-resolution pixel-level classification of echogram data with few (<100) training examples.
- Context channels are shown to improve the classification results. Animal behaviour is related to environment, so providing proxies for environmental context improves the classification results.
- Classification of bubbles and small aggregation of salmon by a U-Net is very precise, and shown to out-perform manual annotations. This is a challenging problem in echogram analysis and often results in discarding valuable data. CNN's are able to efficiently learn the morphological differences and frequency relationships to rapidly classify fish, even in the presence of bubbles.

References

[1] ICES. Report on Echo Trace Classification. ICES Cooperative Research Report, 2000.
[2] Government of Canada. Wild Salmon Policy 2018 to 2022 Implementation Plan | Pacific Region | Fisheries and Oceans Canada, Oct. 2018b.
[3] Ronneberger, O., P. Fischer, and T. Brox. U-Net: Convolutional Networks For Biomedical Image Segmentation. arXiv:1505.04597 [cs], May 2015.
[4] Rousseau. S., S. Gauthier, S. Johnson, C. Neville, and M. Trudel. Juvenile salmon acoustic monitoring in the Discovery Islands, British Columbia. Can. Tech. Rep. Fish. Aquatic Sci. 3277. Canadian Technical Report of Fisheries and Aquatic Sciences, 2018.
[5] Lavery, A.C, P.H. Wiebe, T.K. Stanton, G.L. Lawson, M.C. Benfield, and N. Copley. Determining dominant scatterers of sound in mixed zooplankton populations. The Journal of the Acoustical Society of America, 122(6):3304–3326, Dec. 2007.
[6] Trevorrow, M.V., S. Vagle, and D.M. Farmer. Acoustical measurements of microbubbles within ship wakes. The Journal of the Acoustical Society of America, 95(4):1922–1930, Apr. 1994.
[7] Kriegeskorte, N. Deep neural networks: A new framework for modeling biological vision and brain information processing. 1(1):417–446, Nov. 2015.

Acknowledgments

This work was supported by Mitacs through the Mitacs Accelerate program in partnership with the University of Victoria and the industry partner, ASL Environmental Sciences.
This research was enabled in part by support provided by Westgrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca).