
A Deep Learning-based Framework for the Detection of Schools of Herring in Echograms

Alireza Rezvanifar ^{*1} Tunai Porto Marques ^{*1} Melissa Cote ¹ Alexandra Branzan Albu ¹ Alex Slonimer ²
Thomas Tolhurst ² Kaan Ersahin ² Todd Mudge ² Stéphane Gauthier ³

Abstract

Tracking the abundance of underwater species is crucial for understanding the effects of climate change on marine ecosystems. Biologists typically monitor underwater sites with echosounders and visualize data as 2D images (echograms); they interpret these data manually or semi-automatically, which is time-consuming and prone to inconsistencies. This paper proposes a deep learning framework for the automatic detection of schools of herring from echograms. Experiments demonstrated that our approach outperforms a traditional machine learning algorithm that uses hand-crafted features. Our framework could easily be expanded to detect more species of interest to sustainable fisheries.

1. Introduction

The study of acoustic backscatter is commonly utilized by biologists as a means to monitor underwater sites in a thorough and non-invasive manner. Different species possess different acoustic properties, thus producing different responses. Acoustic survey data are typically interpreted via time-consuming manual or semi-automatic approaches prone to inconsistencies, using expensive softwares such as Echoview (<https://www.echoview.com/>). In this paper, we apply deep learning methods for automatically detecting specific biological targets in acoustic survey data. Our goals are to improve data processing and interpretation for species abundance tracking and environmental monitoring, and to provide a tool that reliably supports analyses of the effects of climate change. We present a case study on schools of herring.

^{*}Equal contribution ¹University of Victoria, Victoria, Canada
²ASL Environmental Sciences, Victoria, Canada ³Fisheries and Oceans Canada, Victoria, Canada. Correspondence to: Alexandra Branzan Albu <aalbu@uvic.ca>.

Tackling Climate Change with Machine Learning Workshop at the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

Acoustic data are obtained via multifrequency echosounders, such as the Acoustic Zooplankton Fish Profiler (AZFP) [1], and visualized in 2D, with depth along the y-axis and time along the x-axis, producing an echogram. A color code represents the echo amplitude, usually in the form of volumetric backscatter strength. Visual structures indicate geological and biological elements. Acoustic-scattering models provide a link between echograms and the size/abundance distribution of the animals [2]. Echograms are traditionally analyzed using statistical characteristics of the aggregations of organisms [3]. Feature-based classification methods commonly favor a classical machine learning [4, 5, 6, 7, 8, 9] paradigm and use hand-crafted features. Deep learning, which has been shown to be very effective at various tasks in computer vision such as object detection and recognition, has yet to permeate echogram analysis. Only a few methods deal with identifying fish from echograms utilizing deep learning and are quite limited in the nature of their experiments [10] or data [11]. A few additional works have applied deep learning to the detection of mobile marine life in sonar data [12, 13, 14], which typically have a higher resolution than echograms.

Our contributions are two-fold. Firstly, we propose a novel application of deep learning techniques for fish detection from echograms. Secondly, our framework automates acoustic survey analyses; an automatic approach will reduce processing times, required man-power, and inconsistencies in the results. Our approach also has the potential to be scaled to handle additional underwater species. Biologists collect large quantities of echograms so the amount of data is not a limitation for the training of a deep learning-based framework; however, the annotation process is a tedious task. We show that, even using a smaller annotated dataset, deep learning-based solutions can outperform a classical machine learning method. Our promising results indicate that with the annotation of more data, deep learning can be efficiently applied in underwater acoustic analysis.

2. Technical Approach

Thanks to the emergence of convolutional neural networks (CNNs) and the existence of large annotated datasets like

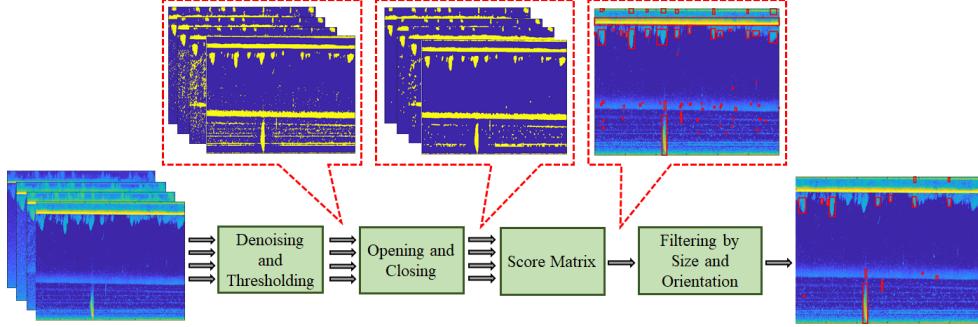


Figure 1. ROI extractor pipeline and sample processed echograms at each stage.

MS COCO [15], PASCAL VOC [16], and ImageNet [17], object detection is now well addressed for natural scenes. CNNs have been utilizing different architectures to find regions of interest (ROIs) in images and to classify them. While earlier networks like R-CNN [18] and Fast R-CNN [19] were using selective search [20] for finding ROIs, more recent ones like Faster R-CNN [21] use dedicated networks (i.e., Region Proposal Networks (RPNs)) to that end. Unfortunately, this approach is not feasible in applications outside the natural scene scope, including ours, where there is insufficient box-annotated data for training RPNs. We thus designed an ROI extractor based on high level information on the intensity and morphology of schools of herring. We introduce the proposed ROI extractor in Section 2.2 and then describe the deep learning-based classification phase in Section 2.3. Our data are described in Section 2.1.

2.1. Dataset

Our dataset is composed of 100 multifrequency echograms generated from data collected with an AZFP device deployed on the surface of the water looking downward, in the Discovery Passage off Vancouver Island, British Columbia, Canada in 2015. The AZFP transducer was calibrated in four frequency channels (67, 125, 200, and 455 KHz), generating a total of four echograms. We use all four echograms as input, as they may convey complementary information. Each echogram resolution is 1200 pixels in the x-axis (representing one hour of measurements) by 571 pixels in the y-axis (corresponding to a depth range of 50 meters). The dataset is divided into 70 echograms for training (which itself is divided in 80% training, 20% validation) and 30 echograms for testing.

2.2. ROI Extractor

Two high level features common to all schools of herring are presumed: 1) a strong intensity core visible in most echograms, 2) a vertical elongated shape. Figure 1 shows the ROI extractor pipeline and sample intermediate outputs after each step. We first apply a median filter along the time direction of the echograms to smooth out the spiky variations and remove some of the undesirable signals such

as small fish. The denoised echograms are then binarized using adaptive thresholding [22], and processed via morphological opening and closing to remove small isthmuses and protrusions and fill in small holes. We create a binary score matrix, indicating the strength of the potential ROIs, by summing the four processed binarized echograms and retaining the pixels that have a value greater than 2, thus retrieving objects present in at least three frequency channels. Connected components are then filtered for size and orientation: we discard very small regions (less than 50 pixels) as well as regions with an orientation smaller than 60 degrees with respect to the horizontal axis, as those cannot be schools of herring. The remaining connected components constitute the final ROIs. The ROI bounding boxes are passed to the classification stage (Section 2.3) for further analysis.

We do not seek a high precision for ROIs at this stage, as they will be further refined in the next step. However, we do seek a high recall; if a school of herring is not encompassed by an ROI, it cannot be detected in subsequent steps.

2.3. Classification

The classification stage determines whether each ROI represents a school of herring or background. We compare three deep learning-based architectures, which automatically extract features (without any contextual information) from ROIs and classify them: ResNet [23], DenseNet [24] and InceptionV3 [25].

Each echogram might contain multiple instances of schools of herring (*positive* samples) and background (*negative* samples). To obtain meaningful and realistic negative samples (avoiding the use of random crops), we used the ROI extractor outputs for the 70 training echograms as follows. The ROI bounding boxes are compared with the positive samples from the ground truth and the intersection-over-union (IoU), i.e. ratio between areas of overlap and union, is calculated. ROI bounding boxes with less than 0.4 of IoU score are used as negative samples and the others as positive samples. We follow the 1:2 ratio between positive and negative samples used in the popular work of Girshick et al. [19].

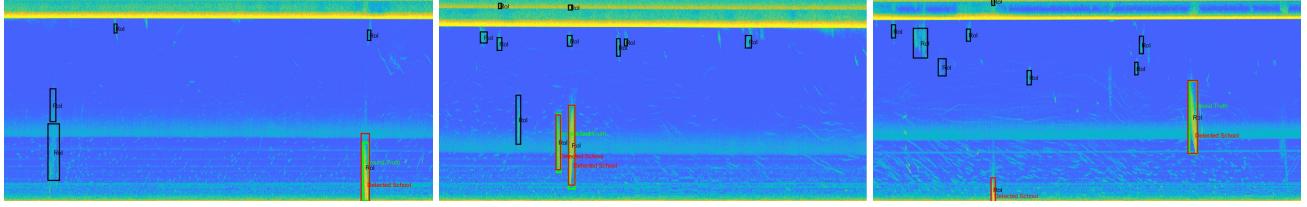


Figure 2. Results of the detection framework using DenseNet. Black: regions classified as background; green: ground truth bounding boxes; red: regions classified as schools of herring. All detections are correct in the first two samples (left and center), while one ROI bounding box was incorrectly classified in the right-most echogram.

3. Results and Discussion

The two main components of the framework (ROI extractor and classifier) are evaluated separately using precision, recall, and F1-score metrics.

3.1. ROI Extractor Evaluation

Since we are dealing with bounding boxes, we make use of an IoU threshold between the ROIs and the ground truth to determine true positives, false positives, and false negatives. We consider a true positive when an ROI bounding box has an IoU with the ground truth *higher* than the threshold. Results for the 100 echograms are shown in Table 1.

Table 1. Precision, recall and F1-score of the ROI extractor for different IoU thresholds.

IoU Threshold	Precision	Recall	F1-Score
0.0	0.173	0.931	0.292
0.2	0.171	0.917	0.288
0.4	0.155	0.834	0.262

The most relevant metric here is the recall, as ROIs are further classified in the next stage and the classifier cannot recover a school missed by the ROI extractor. Table 1 shows that the performance increases as the IoU threshold is lowered. The small difference in recall (less than 10 p.p.) between IoU thresholds 0.0 and 0.4 shows that most of the ROI bounding boxes have a significant overlap with the ground truth. Ideally, the incorrect bounding boxes will be identified as background in the next step, classification.

3.2. Framework Evaluation

This section assesses the complete framework as the output of the ROI extractor are used as input to the classifiers. Table 2 shows the classification results for each tested network architecture over the test set (30 echograms) for an IoU threshold of 0.4, while highlighting the best results in bold. We compare our results with a baseline method based on traditional hand-crafted features engineered for fisheries research (ratio between minor to major axis, mean intensity, eccentricity, and circularity of the region) and classical machine learning (support vector machine (SVM) classifier [26] with a linear kernel).

The DenseNet architecture achieves the best balance be-

tween metrics, and outperforms the baseline method. Figure 2 shows sample results of the framework using DenseNet. In most of the echograms (as exemplified by the left and center images of Figure 2), the ROI extractor was able to detect bounding boxes around the positive samples, followed by a correct classification by the DenseNet. However, the classifier also incorrectly classified some samples, as illustrated in the right-most image of Figure 2. Experiments where the IoU threshold was set to 0.0 in the classification phase saw a recall by the framework that reached the maximum value of 0.93 (bounded by the ROI extractor), attesting to the efficiency of the trained classifier. However, the usage of such a low IoU threshold is not recommended, given that any ROI bounding box that at least touches a positive sample could be considered as a school of herring candidate. Higher IoU thresholds guarantee that the detections are more reliable.

Table 2. Precision, recall and F1-score of the full pipeline using different CNN-based architectures for IoU thresholds of 0.4.

Architecture	Precision	Recall	F1-Score
ResNet50 [23]	0.77	0.85	0.81
DenseNet201 [24]	0.78	0.85	0.82
InceptionNet [25]	0.81	0.81	0.81
Baseline (SVM) [26]	0.51	0.78	0.62

4. Conclusion

We propose a novel framework that assists the study of species' abundance in underwater sites using acoustic data. Our first case study is focused on the detection of schools of herring in echograms using two main components: a novel ROI extractor and a deep learning-based image classifier. The framework achieved a good performance for IoU thresholds (i.e., 0.4) which guaranteed meaningful detections. This framework will assist in the interpretation of large amounts of raw acoustic data. Although this proof-of-concept focuses on schools of herring, it is scalable to other aquatic species that can be monitored using echosounders (e.g., zooplankton, salmon). The future substitution of our ROI extractor by generic ROI calculation systems (e.g., RPNs [21]) will simplify the expansion of detectable classes to the gathering of new visual samples. The ability to measure the abundance of such subjects is paramount for the study of the effects of water temperature shifts caused by climate change-related phenomena.

References

[1] David Lemon, Paul Johnston, Jan Buermans, Eduardo Loos, Gary Borstad, and Leslie Brown. Multiple-frequency moored sonar for continuous observations of zooplankton and fish. In *2012 Oceans*, pages 1–6. IEEE, 2012.

[2] Mark V Trevor. The use of moored inverted echo sounders for monitoring meso-zooplankton and fish near the ocean surface. *Canadian Journal of Fisheries and Aquatic Sciences*, 62(5):1004–1018, 2005.

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

[4] P LeFeuvre, GA 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):137–147, 2000.

[5] Ariel G Cabreira, Martín Tripode, and Adrián Madirolas. Artificial neural networks for fish-species identification. *ICES Journal of Marine Science*, 66(6):1119–1129, 2009.

[6] Aymen Charef, Seiji Ohshima, Ichiro Aoki, and Natheer Al Absi. Classification of fish schools based on evaluation of acoustic descriptor characteristics. *Fisheries Science*, 76(1):1–11, 2010.

[7] Hugo Robotham, Paul Bosch, Juan Carlos Gutiérrez-Estrada, Jorge Castillo, and Inmaculada Pulido-Calvo. Acoustic identification of small pelagic fish species in Chile using support vector machines and neural networks. *Fisheries Research*, 102(1-2):115–122, 2010.

[8] Stéphane Gauthier, Johannes Oeffner, and Richard L ODriscoll. 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.

[9] Niall G Fallon, Sophie Fielding, and Paul G Fernandes. Classification of Southern Ocean krill and icefish echoes using random forests. *ICES Journal of Marine Science*, 73(8):1998–2008, 2016.

[10] Yudai Hirama, Soichiro Yokoyama, Tomohisa Yamashita, Hidenori Kawamura, Keiji Suzuki, and Masaaki Wada. Discriminating fish species by an Echo sounder in a set-net using a CNN. In *21st Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES)*, pages 112–115. IEEE, 2017.

[11] Yue Shang and Jianlong Li. Study on echo features and classification methods of fish species. In *10th International Conference on Wireless Communications and Signal Processing (WCSP)*, pages 1–6. IEEE, 2018.

[12] L. Liu, H. Lu, Z. Cao, and Y. Xiao. Counting fish in sonar images. In *25th IEEE International Conference on Image Processing (ICIP)*, pages 3189–3193. IEEE, 2018.

[13] Geoff French, Michal Mackiewicz, Mark Fisher, Mike Challiss, Peter Knight, Brian Robinson, and Angus Bloomfield. Jellymonitor: automated detection of jellyfish in sonar images using neural networks. In *14th IEEE International Conference on Signal Processing (ICSP)*, pages 406–412. IEEE, 2018.

[14] Dmitry Glukhov, Rykhard Bohush, Juho Mäkiö, and Tatjana Hlukhava. A joint application of fuzzy logic approximation and a deep learning neural network to build fish concentration maps based on sonar data. In *2nd International Workshop on Computer Modeling and Intelligent Systems (CMIS)*, pages 133–142, 2019.

[15] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European Conference on Computer Vision*, pages 740–755. Springer, 2014.

[16] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The Pascal Visual Object Classes (VOC) challenge. *International Journal of Computer Vision*, 88(2):303–338, 2010.

[17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[18] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 580–587, 2014.

[19] Ross Girshick. Fast r-cnn. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1440–1448, 2015.

[20] Jasper RR Uijlings, Koen EA Van De Sande, Theo Gevers, and Arnold WM Smeulders. Selective search for object recognition. *International Journal of Computer Vision*, 104(2):154–171, 2013.

- [21] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, pages 91–99, 2015.
- [22] Derek Bradley and Gerhard Roth. Adaptive thresholding using the integral image. *Journal of Graphics Tools*, 12(2):13–21, 2007.
- [23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [24] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017.
- [25] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.
- [26] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.