Computer Vision-based Systems for Environmental Monitoring Applications

PhD Candidate: **Tunai Porto Marques**

Academic Supervisor: Dr. Alexandra Branzan Albu

University of Victoria

2021 Winter Conference on Applications of Computer Vision - Doctoral Consortium

Research statement

This doctoral thesis explores the potential for the use of Computer Vision techniques in diverse environmental monitoring (EM) efforts. As a result, it aims to provide automatic methods that aid specialists and the general public alike in efficiently and timely interpreting large amounts of EM data. This streamlined and consistent interpretation of years-worth of visual and acoustic monitoring samples helps informing better conservation initiatives, local and federal legislation, as well as providing strong research and education tools.

In order to study such potential while considering different EM layouts and data types, this thesis has been divided into three self-contained projects connected by the same overarching theme, "the employment of computer vision aiming to aid environmental monitoring". A first project relies on physically-based models of underwater environments to enhance low-lighting marine images without the use of any learning process. The other two efforts explore hybrid approaches that involve hand-crafted visual features and deep learning (DL)-based frameworks, ultimately offering detection mechanisms for visual targets in different types of EM data (i.e., visual and acoustic). Each project and their associated publications are summarizes in this poster.

Detection of Marine Species in Echograms

This project aims to provide a framework that accelerates and standardizes the analysis of echosounder-based data. These acoustic data are typically visualized in the form of time/depth images (*Echograms*), and interpreted by Biologists and Acouticians in time-consuming and error-prone tasks. Echograms are able to indicate the presence and inform about the behavior of large groups of specimen (*schools* in the case of herring fish).

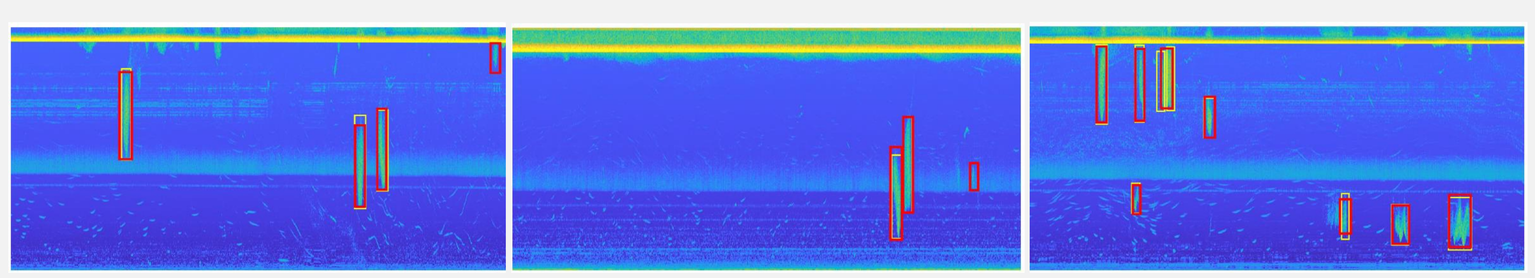
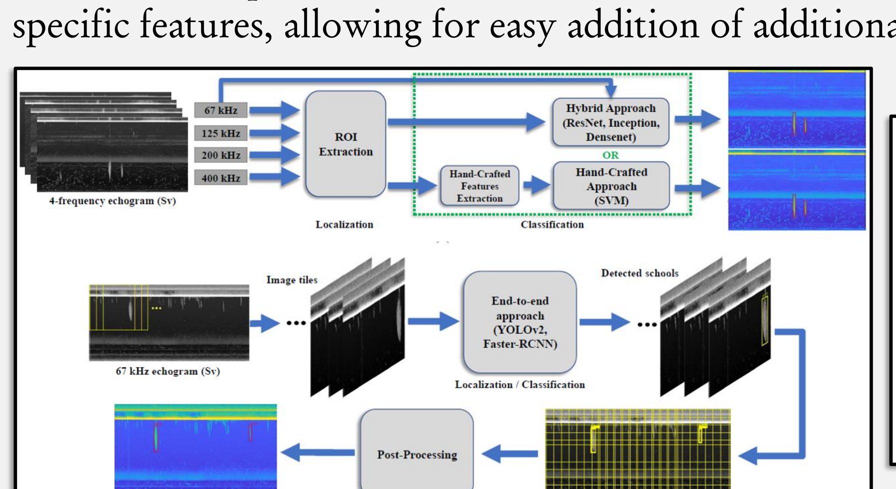


Figure 1. Detection of schools of herring (red bounding boxes) in echograms using the tiling approach in the proposed system. Ground truth is represented by yellow bounding boxes. Retrieved from [6].

In [5] we detail a novel, non-data-driven Region of Interest (ROI) extractor that determines regions where there is a high likelihood of the presence of schools of herring. These regions are classified as *herring* or *background* following two approaches, namely: 1) four hand-crafted features are calculated from a given ROI and used to drive a custom-trained Support Vector Machine (SVM); 2) in a *hybrid approach* each ROI is used as the input of a custom-trained, DL-based image classifier. Our best-performing detector in this publication was obtained using the hybrid approach with a DenseNet-201 (Huang *et al.*, 2017) as feature extractor (F1-score of 0.82 in a custom echogram dataset).

In [6] we expand the dataset of annotated echograms introduced in [5] to 358 samples, and compare three detection approaches aiming to explore the machine-learning-based solution space in the context of schools of herring detection. Using the custom ROI extractor [5], experiments with the aforementioned SVM and hybrid approaches were combined with a custom-trained, end-to-end object detector-based detection strategy. This publication also introduced the *tiling* strategy, which mitigates the scale discrepancy issues observed when training/testing the end-to-end detection approach. Our end-to-end approach yielded the best detection performance on the dataset introduced (see Figure 2) without the use of class-specific features, allowing for easy addition of additional biological targets.



Method	IoU	P	R	F1
1) Hand-crafted (SVM [29])	0.3	45.41	65.63	53.67
1) Hand-crafted (SVM [29])	0.5	41.62	60.16	49.20
2) Hybrid (ResNet-50 [22])	0.3	72.37	85.94	78.57
2) Hybrid (ResNet-50 [22])	0.5	62.50	74.20	67.80
2) Hybrid (DenseNet-201 [23])	0.3	64.37	87.50	74.17
2) Hybrid (DenseNet-201 [23])	0.5	55.75	75.78	64.24
2) Hybrid (InceptionV3 [24])	0.3	**78.99**	85.16	**81.95**
2) Hybrid (InceptionV3 [24])	0.5	**68.12**	73.44	70.68
3) End-to-end (YOLOv2 [26])	0.3	73.08	**89.06**	80.28
3) End-to-end (YOLOv2 [26])	0.5	67.31	**82.03**	**73.94**
3) End-to-end (Faster R-CNN [25])	0.3	66.90	74.22	70.37
3) End-to-end (Faster R-CNN [25])	0.5	45.07	50	47.41

P = precision, R = recall, F1 = F1-score

Figure 2. **Left:** Hybrid DL-based and Hand-Crafted Machine Learning (ML)-based detection approaches (top), End-to-End DL-based Detection approach (bottom). **Right:** performance comparison between the three approaches considering different feature extractor backbones and Intersection-over-Union (IoU) thresholds. Retrieved from [6].

Detection of Marine Vessels in Visual Time Series

In this project we seek to automate the identification of marine vessels travelling in critical habitats of Southern Resident Killer Whales (SRKW) living off the coast of Vancouver Island, BC, as marine traffic has been associated with significant environmental impact in the region. These visual targets might take vastly different shapes, colors, sizes and are expected to be sighted with varying levels of visibility (e.g., changing weather). In particular, we focus on the more challenging aspect of identifying pixel-level vessels (i.e., those with mean area of approximately 80 pixels or less) using small time series of only three images that replicate remote EM layouts.

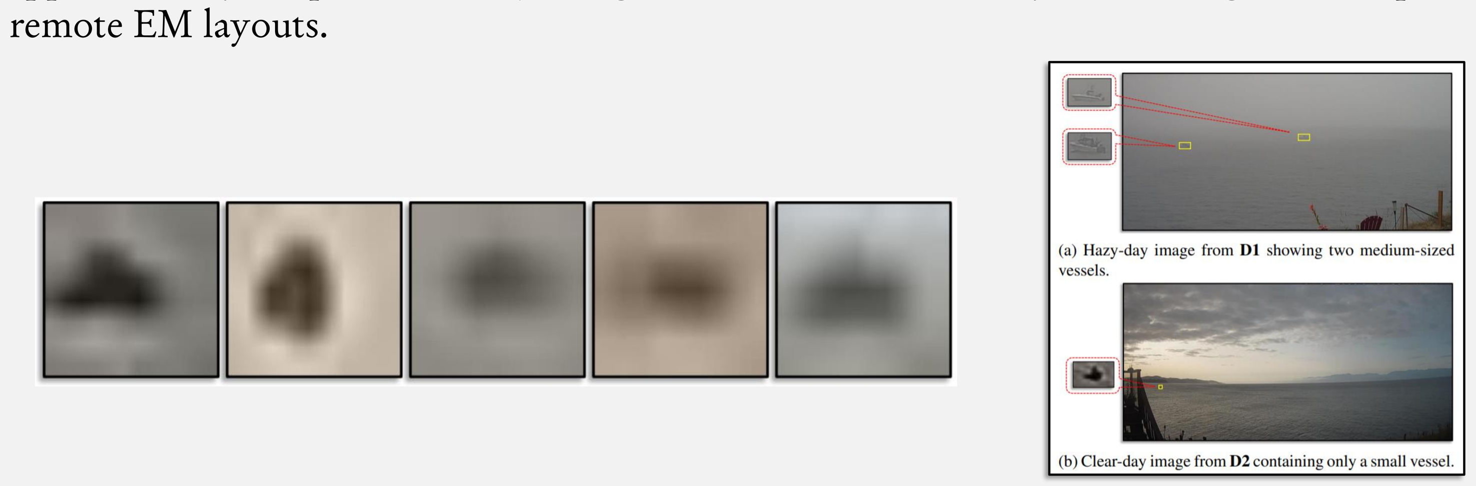


Figure 3. **Left:** Examples of small marine vessels (mean area of 79 pixels) illustrating the challenging task of small-vessel recognition. **Right:** marine vessels datasets offered in [4] and used to evaluate the proposed system. Retrieved from [4].

When considering a single image, identifying small vessels represents a challenging task even for human operators (see Figure 3 left). We then propose to consider *motion* from small time series as a proxy for such visual targets. The Detector of Small Marine Vessels (DSMV) system created and introduced in [4] employs a novel bi-directional Gaussian Mixture Model (GMM), four environmental assumptions based on the EM data curated, as well as a DL-based image classifier (several backbones are evaluated). The output of this small-vessel detector are combined with those of DL-based end-to-end object detectors---our studies show that they are efficient in identifying medium- and large-sized vessels (see Figure 4). The proposed system is therefore capable of identifying boats of any size using small time series.

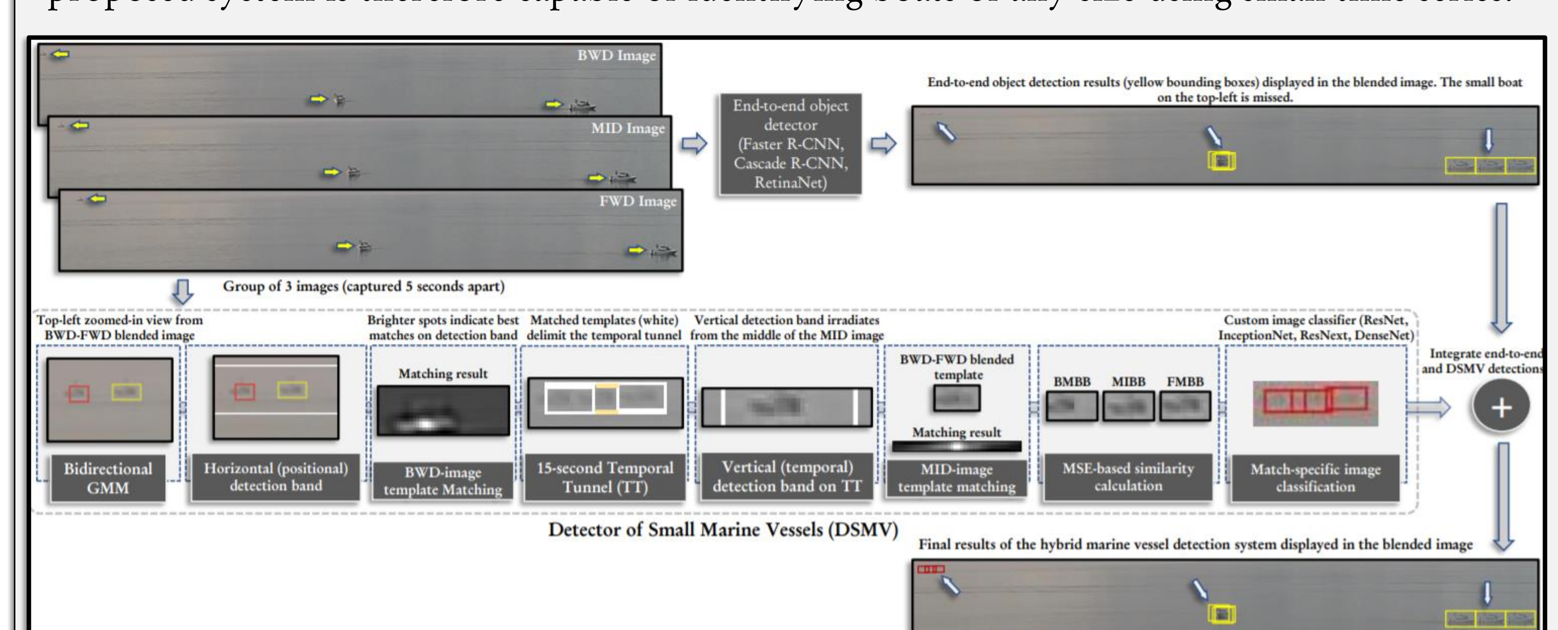


Figure 4. Hybrid marine vessel detector proposed. The detection results of the end-to-end object detector and the DSMV are combined for enhanced detection capabilities (i.e., vessels of all sizes). Retrieved from [4].

The DSMV starts with a novel bi-directional GMM that considers motion from the three input images in the *forward* and *backward* motion. Morphological operations ensue and the resulting connected components determine bounding boxes (BB) from the first and third frames. These motion-triggered BBs are matched inside a narrow horizontal detection band representing the initial and final positions of a single vessel. This same boat is identified in the second frame by determining a vertical detection band ("temporal tunnel") where it is expected to be placed. A similarity-based filtering follows, and a final custom-trained DL-based image classifier determines, among the filtered BBs, instances of *vessels* and *background*.

Two datasets are introduced in [4]: D1 (1,056 boats of diverse sizes) and D2 (165 exclusively small marine vessels). We perform detection in D1 and D2 using the proposed detection system and numerous state-of-the-art (SOTA) end-to-end object detectors (Ren *et al.* 2015, Cai and Vasconcelos 2018, Lin *et al.* 2017). The use of the proposed hybrid detection system outperforms the SOTA methods in up to 112% for D2 [4].

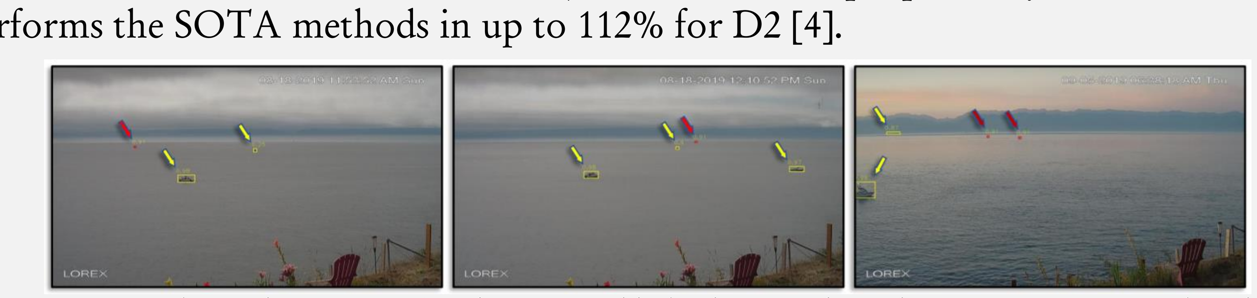


Figure 5. Detection results on dataset D1 using the proposed hybrid approach. Red BBs represent DSMV-based detection, while yellow ones indicate those from DL-based object detectors (combined for the final output). Retrieved from [4].

Low-Light Underwater Image Enhancement using Local Contrast and Multi-Scale Fusion

The single-image framework proposed in this project enhances the human-perceived quality (i.e., visibility, contrast, color fidelity) of underwater monitoring images; these data, when captured in cabled observatories with subpar quality, might have their scientific value severely reduced (see Figure 6).

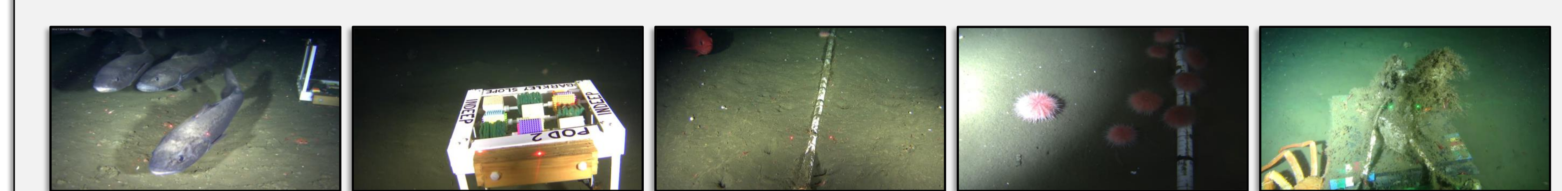


Figure 6. Low-light underwater images from the OceanDark dataset. The goal of this project is to enhance their visibility and overall human-perceived quality (i.e., visibility, contrast, color fidelity). Retrieved from [2].

Initial iterations of the framework were based on the similarity between the inverted version of low-light images and hazy ones (Dong *et al.* 2011). In [1] we introduce a dataset of synthetic hazy images based on haze/haze-less pairs of inputs (Silberman *et al.* 2012), as well as a thorough study on distinct methods for the refinement of *transmission maps*. In [2] we offered a dataset of 183 low-light underwater images, OceanDark. This publication also debuts our observation that the local standard deviation of different-sized patches can be used as a proxy to local contrast information and guide the choice of patch size in Dark Channel-based (He *et al.* 2011) image enhancement systems (see Figure 7). Experiments in [2] using OceanDark and four SOTA underwater-specific methods showed a significant enhancement in global contrast (Matkovic *et al.* 2005), number of visual features (Bay *et al.* 2008), visibility (Hautiere *et al.* 2011) and reduction in overall darkness (Choi *et al.* 2015).

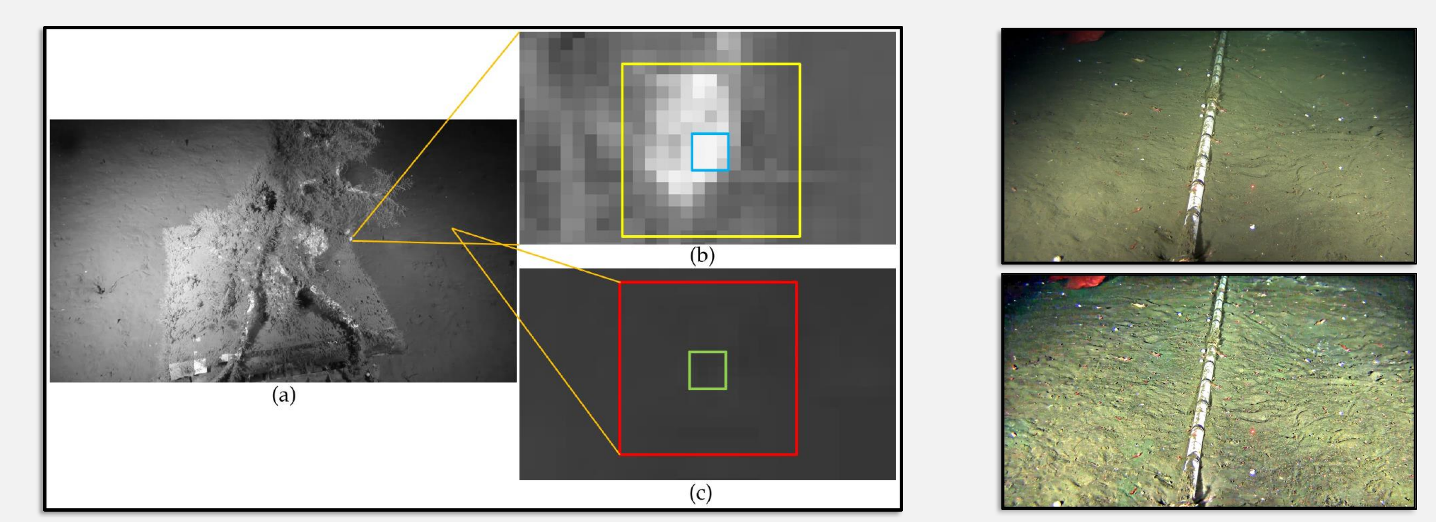


Figure 7. **Left**: dynamic patch sizes—regions of higher contrast (b) present different transmission profiles, justifying the use of bigger patches in the Dark Channel and Transmission Map calculations. **Right**: OceanDark sample enhanced using the proposed L^2UWE [3]. Retrieved from [2,3].

The current version of this enhancement system, L^2UWE [3], used the OceanDark dataset and the aforementioned contrast-guided approach to derive two atmospheric lighting models of an underwater scene, then employed multi-scale fusion (Ancuti and Ancuti, 2013) to combine the two enhanced outputs obtained with them. This new approach mitigates the over-illumination issue observed in some regions of the output from [2] stemming from the use of a simplified atmospheric lighting model. We create an evaluation suite composed of seven metrics to compare L^2UWE with several SOTA methods (both low-light- and underwater-specific). Our experiments show that L^2UWE can enhance the illumination of darker regions while preserving the finer details of the input (see Figure 8 for a summary of the results on the OceanDark dataset).

Method	UIQM [37]	PCQI [45]	GCF [34]	e-score [23]	r-score [23]	FADE [17]	SURF [9]
Original	0.88 ± 0.13	1	3.28 ± 0.62	N/A	1	1.91 ± 0.79	340 ± 293
Marques [39]	0.99 ± 0.12	0.85 ± 0.03	3.41 ± 0.71	0.28 ± 0.32	2.75 ± 0.76	0.46 ± 0.18	705 ± 470
Berman [10]	1.00 ± 0.18	0.78 ± 0.15	3.84 ± 1.07	0.25 ± 0.50	2.91 ± 1.96	1.15 ± 0.40	425 ± 317
Fu [20]	0.93 ± 0.15	0.93 ± 0.09	3.28 ± 0.57	0.09 ± 0.39	1.72 ± 0.35	1.75 ± 0.25	865 ± 478
Cho [16]	1.24 ± 0.15	0.87 ± 0.04	4.11 ± 0.84	0.89 ± 0.54	1.72 ± 0.09	1.81 ± 0.52	751 ± 428
Drews [19]	1.38 ± 0.14	0.85 ± 0.05	4.70 ± 0.89	1.06 ± 0.83	1.29 ± 0.31	2.08 ± 0.90	589 ± 324
Zhang [49]	1.28 ± 0.08	1.03 ± 0.07	$\mathbf{6.34 \pm 0.74}$	1.48 ± 0.88	3.70 ± 1.00	0.72 ± 0.39	1719 ± 620
Guo [22]	0.93 ± 0.13	0.86 ± 0.03	3.50 ± 0.73	0.16 ± 0.14	2.21 ± 0.64	0.60 ± 0.24	607 ± 452
L^2UWE	$\mathbf{1.38 \pm 0.11}$	$\mathbf{1.17 \pm 0.07}$	4.89 ± 0.66	$\mathbf{1.82 \pm 0.83}$	$\mathbf{4.61 \pm 0.58}$	$\mathbf{0.42 \pm 0.20}$	$\mathbf{1856 \pm 655}$

Figure 8. Mean and standard deviation of seven metrics computed on all samples of OceanDark. Best result are bolded. See [3] for detailed references on the metrics and methods utilized.

Publications

[1] **T. Porto Marques**, A. B. Albu, and M. Hoeberechts, "Enhancement of low-lighting under-water images using dark channel prior and fast guided lters," in *International Conference on Pattern Recognition Workshops* (ICPRW). Springer, 2018, pp. 55–65.

[2] **T. Porto Marques**, A. Branzan Albu, and M. Hoeberechts, "A contrast-guided approach for the enhancement of low-lighting underwater images," *MDPI Journal of Imaging*, vol. 5, no. 10, p. 79, 2019.

[3] **T. Porto Marques** and A. Branzan Albu, "L2UWE: A framework for the efficient enhancement of low-light underwater images using local contrast and multi-scale fusion," in Proceedings of the *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (CVPRW), 2020.

[4] **T. Porto Marques**, A. B. Albu, *et al.*, "Size-invariant detection of marine vessels from visual time series," in *2021 IEEE/CVF Winter Conference on Applications of Computer Vision* (WACV).

[5] A. Rezvanifar *, **T. Porto Marques** *, *et al.*, "A deep learning-based framework for the detection of schools of herring in echograms," in *2019 NeurIPS Workshop - Tackling Climate Change with Machine Learning* (arXiv preprint arXiv:1910.08215), 2019. *: equal contribution.

[6] **T. Porto Marques**, A. Rezvanifar, M. Cote, A. B. Albu, K. Ersahin, T. Mudge, and S. Gau-thier, "Detecting marine species in echograms via traditional, hybrid, and deep learning frameworks," in 2020 International Conference on Pattern Recognition (ICPR).