

KNOWLEDGE TRANSFER FOR DEEP-LEARNING GAS-BUBBLE DETECTION IN UNDERWATER ACOUSTIC WATER COLUMN DATA: EXPLORING DATA IN THE MOZAMBIQUE CHANNEL

T Perret Ifremer, Geo-Ocean, Plouzané, France
S Dupré Ifremer, Geo-Ocean, Plouzané, France
A Gaillot Ifremer, Geo-Ocean, Plouzané, France
Y Lacroix Kongsberg Discovery, Norway; IMAS, University of Tasmania
G Le Chenadec ENSTA Bretagne, Brest, France

1 INTRODUCTION

Concerns related to fluid emissions from the seabed are relevant to both the biosphere and the geosphere. They include marine geohazards such as earthquakes, sedimentary instabilities, and the release of large amounts of methane. These fluids escape from the seafloor and could potentially migrate to the ocean-atmosphere interface through the water column. In addition, precipitation from hydrothermal sources (e.g., metal sulphides)¹ can offer mineral resources. For the above reasons, it is crucial to detect and locate fluid emissions. Methane is produced through thermogenic and biogenic processes² and methane seeps can be found worldwide in different geological environments. The gas is either dissolved or present as isolated bubbles or "megaplumes".

Multibeam Echo Sounders (MBESs) are underwater mapping devices used to capture acoustic backscatter data from targets within the water column. MBESs, active sonars equipped with transmitting and receiving antennas, are primarily utilized for bathymetric surveys³. These sonars are typically mounted on the hull of ships and can create a substantial number of beams (e.g., 288 and 880 beams for the EM302 and EM122 Kongsberg and the Reson Seabat 7150 MBES, respectively). With their wide swath (usually between 120 and 170 degrees), MBESs can effectively cover extensive areas of the seabed and large volumes within the water column. This process creates Water Column Images (WCIs) depicting the acoustic backscatter from the water column during each ping cycle (Fig. 1). The utilization of water column data for fluid detection has been gaining traction since 2009⁴. However, the substantial amount of recorded data makes manual interpretation a time-consuming task. In addition, automatic methods are challenged in the region under the top specular sidelobe in WCIs, namely the Minimum Slant Range (MSR)⁵ (Figs. 1 and 2). Finally, automated methods must be able to adapt to WCIs from various sounders with different acoustic configurations, resulting in distinct fluid features and WCI acoustic characteristics.

To address this challenge, convolutional neural networks, like YOLOv5⁶, are increasingly used to detect objects in images due to their ability to learn features and classifiers from large datasets⁷. Perret et al. (2023)⁸ conducted a study exploring the use of YOLOv5 to detect fluids in images of water column, intending to develop a versatile and transferable approach to this issue. It has also shown interesting detection performances when learning the model with WCIs from two different MBESs. Based on this latter conclusion, the present study aims to investigate how to train a YOLOv5 model based on: i) a small number of WCIs without fluid, acquired at the start of a marine expedition operated with a MBES Kongsberg EM122, and ii) a set of labelled WCIs with and without fluids acquired with other MBESs (i.e. Kongsberg EM302 and Reson Seabat 7150).

The first set of WCIs is composed of non-fluid related echoes from the first acquired lines of the PAMELA-MOZ01 marine expedition (first day of this one-month-long marine expedition). The latter WCIs are those used in Perret et al. (2023). The aim was to train the neural network on WCIs with the acoustic characteristics of the EM122. This network was thus used for inference over the entire PAMELA-MOZ01 EM122 data, including challenging data acquired during coring operations and unsynchronized acoustic surveys (i.e. SubBottom Profiler, SBP). Our results show that the YOLOv5 trained with this method can effectively extract fluid features from other sounders without making too many false detections, facilitating its use during marine expeditions.

2 MATERIAL AND PROCEDURES

2.1 Water Column MBES Data

A MBES uses wide-beam emission for coverage and a narrow beam for high resolution along the track. The signals are processed to form beams on the port and starboard angles and are then digitised and reprojected in polar geometry. In this study, the water column images are obtained by cutting the water column data below seafloor detection. These images are successively captured as the vessel progresses.

Marine expedition	GAZCOGNE1	GHASS2	PAMELA-MOZ1
Survey information	Aquitaine Basin (Bay of Biscay) 07-08 2013	Offshore Romania (Black Sea) 08-09 2021	Mozambique Channel 09-10 2014
MBES	Kongsberg EM302	Reson Seabat 7150	Kongsberg EM122
Frequency (kHz)	[28.25-29.5]	[22.5–24.5]	[12-12.125]
WC sampling frequency (Hz)	1623	6022	[78-202]
Number of beams	288	880	288
Total pulse length [ms]	[1.1-7.5]	[2-4.6]	[6.8-11]
Beam aperture (°)	1 (Tx) x 2 (Rx)	0.5 (Tx) x 0.5 (Rx)	1 (Tx) x 1 (Rx)
Mean depth (+std) (m)	532 ±354	1022 ±452	2992 ±860
Acoustic and environmental conditions	multiple transmission sectors, presence of biomass	presence of dolphins, strong backscattering seabed	multiple transmission sectors, interference with subbottom profiler and coring operations

Table 1: Key information on the GAZCOGNE1, GHASS2 and PAMELA-MOZ1 MBES dataset.

This study utilises data from three campaigns (Table 1). The first two campaigns (GAZCOGNE1 and GHASS2) involved the observation and labelling of fluids (methane bubbles). The GAZCOGNE1 project⁹ facilitated exhaustive acoustic mapping of gas seeps in the Aquitaine Basin (Bay of Biscay). During this expedition, a 30 kHz Kongsberg EM302 MBES was operated. The GHASS2 marine expedition¹⁰ investigated the dynamics of gas emissions in the Black Sea and the relationship between gas hydrates, sedimentary deformations, and submarine instabilities. A 24 kHz Reson Seabat 7150 was employed accordingly. These two sets of data were used exclusively to detect potential fluids on the third dataset from the PAMELA-MOZ1¹¹ campaign exploring seamounts, carbonate platforms, and fluid systems in the Mozambique Channel. The Kongsberg EM122 (12 kHz) was used during this marine expedition for seafloor mapping and fluid detection purposes. During PAMELA-MOZ1, the EM122 MBES insonified a wide seafloor area of nearly 5 000 km² corresponding to 477 847 acquired water column pings.

The first two datasets were manually labelled by operators (Figs 1 and 2). Then, additional fluid labels were included in the training with the help of YOLOv5 (pseudo labelling) and verified by an expert. Experts also examined the PAMELA-MOZ1 dataset, and the fluid outlets were located. However, we chose not to provide any information to the network from the EM122 fluid signature in EM122 WCIs. This information was only used for final performance evaluation, and not for training. Instead, a limited part of the PAMELA-MOZ1 dataset, without any fluid records (corresponding to the first 20 acquired MBES lines, out of 733), was used to train the network, taking into account the acoustic and environmental context of the expedition.

These three datasets show significant variability from an acoustic perspective. There are differences in the parameters used for acoustic acquisition, including variations in the MBES used, their operating frequency, and the aperture for the formed beams (Table 1). Additionally, it is important to note that the acquisition parameters of an MBES can vary even during the same mission. During the PAMELA-MOZ1 marine expedition, the EM122 MBES acquisition parameters were adjusted based on water depth. Three distinct acoustic acquisition modes were used, resulting in different acoustical characteristics, like the number of transmission sectors, angular apertures, and pulse length. The acoustic acquisition modes are classified as 'deep' (57% of the dataset; Fig. 3), 'medium' (31%; Fig. 4), and 'shallow' (12%) modes. The EM122 MBES compensates for pitch and roll by creating steered emission sectors to optimise the geographical coverage of the survey area (Fig. 3).

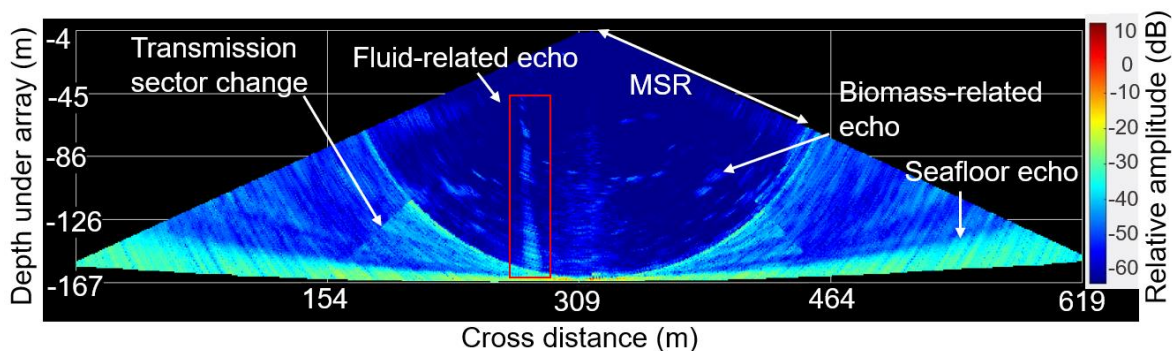


Figure 1: Gas-bubble-related echo in a GAZCOGNE1 (Kongsberg EM302) WCI.

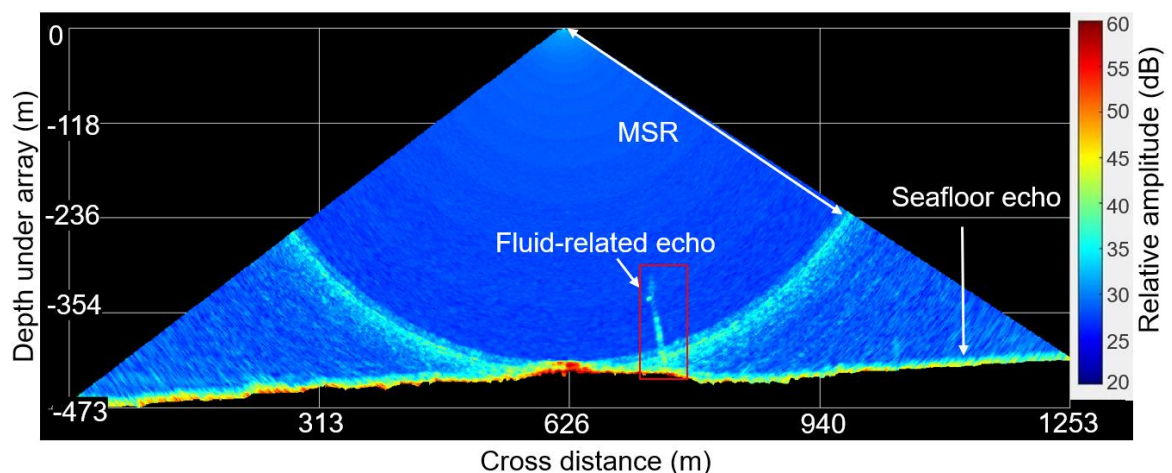


Figure 2: Gas-bubble-related echo in a GHASS2 (Reson Seabat 7150) WCI.

Automatically analysing the PAMELA-MOZ1 dataset is particularly challenging due to the high level of acoustic interference from the subbottom profiler (Ixlblue Echoes, 1.8-5.3 kHz). This interference contaminates the quasi-totality of the survey as the SBP was in acquisition mode for ~97% of EM122 acquisition time (Fig. 3). There was no synchronisation to optimise the seismic acquisition. EM122

acquisition was maintained during coring operations resulting in ~17% of the water column data recorded at “fixed” stations.

To summarize, the WCIs obtained during PAMELA-MOZ1 differ significantly from those of the two marine expeditions previously used to learn fluid features (Table 1 and Figs 1- 3). These differences are due to variations in acoustic and environmental characteristics. Furthermore, SBP-related interference and operations at fixed stations, in particular coring, coupled with the lack of knowledge on the appearance of the fluids on EM122 WCIs during training, make the automatic processing of these WCIs extremely complex for a conventional algorithm.

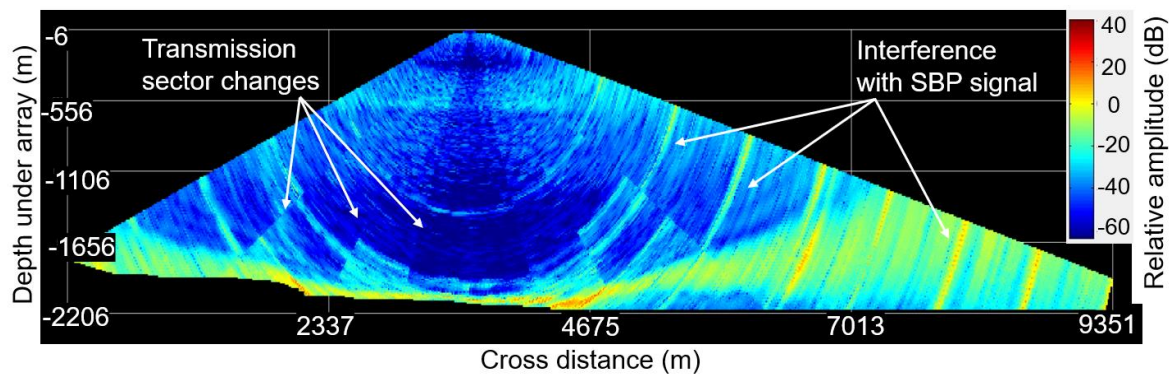


Figure 3: PAMELA-MOZ1 WCI from “deep” mode acquisition with transmission sector changes and interference from the subbottom profiler signal.

2.2 Deep-learning detection with YOLOv5

Deep-learning algorithms are now more commonly used in object detection in images, provided that labelled training datasets are available¹². Object detection is a computer vision task that involves identifying and localising objects of interest in images. In this study, we used the YOLO version 5 (YOLOv5) algorithm (developed by Ultralytics and released in June 2020) as it offers both accuracy and computational efficiency⁶. The YOLO algorithm was the first to propose the one-stage approach using anchor boxes. YOLO's architecture consists of a feature extractor and a detection and localisation head. The feature extractor hierarchically decomposes the spatial, intensity and texture information in images into features. By combining these multiscale features, the detection and localisation head answers three questions: Are there objects in the image? If so, what class do they belong to? Where are these objects located in the image? Subsequent improvements led to the fifth version. In this study, we employed the small (S) version of the YOLOv5 algorithm, which has 7.2 million parameters.

2.3 WCI training dataset

To train an efficient YOLOv5 network, it is essential to control the information used by the network during training. These WCIs can be divided into three types: WCIs with fluid, WCIs without fluid and WCIs with environmental and acoustic artefacts. As demonstrated in Perret et al. (2023), the incorporation of WCIs without fluid and those exhibiting acoustic and environmental artefacts (in a distinct class) particularly reduced the incidence of false positives, in comparison to a network trained exclusively with WCIs exhibiting fluid characteristics. In the construction of our training set for PAMELA-MOZ1, we employed analogous methodologies.

First, to reduce the number of false detections of acoustic interference and artefacts, hard negative mining¹³ was employed. This task involves making an inference using a model trained with GHASS2 and GAZCOGNE1 data. The inference was performed in this study on all WCIs from the first twenty

EM122 acquired lines out of 733. The network produces errors when dealing with artefacts or interferences that share characteristics with fluids. To address this issue, we selected some of these detections (Table 2) that were reclassified as 'acoustic artefacts/environmental phenomena' to build a new two-class training set (fluid and artefact classes). This helped the network distinguish artefacts from fluids. Both PAMELA-MOZ1 and GHASS2 artefacts were included in the dataset. GHASS2 artefacts were considered complementary, such as 'cetacean' echoes or bottom detection failure, which we believe would be interesting for our network to not detect as fluid if this type of echo appears in the PAMELA-MOZ1 data.

To limit false positive detections, another approach is to present the network with WCIs that do not contain fluid and artefacts. The network can learn the acoustic characteristics of EM122 WCIs in our case by including PAMELA-MOZ1 and GAZCOGNE1 data in the 'deep' and 'medium' configurations. This aids the network in generalising to these acoustic configurations, even though they are not from the same MBES. Additionally, features from EM302 acoustic configurations, such as transmission sectors, may be used to generalise to acoustic configurations on EM122.

Set	GAZCOGNE1	GHASS2	PAMELA-MOZ1	Total WCIs
~55% Training set WCIs with fluid	3 500	1 750	0	5 250
~20% Training set WCIs without fluid	633	0	1 267	1 900
~25% Training set WCIs with artefacts	0	1 185	1 185	2 370
Inference (test set)	0	0	477 847	477 847

Table 2: Dataset composition used in this study for the training (percentage refers to the rate of this WCI category in the training set) and test set.

The training set comprises 9 520 WCIs and is presented in Table 2. The percentages of WCIs with and without fluid-related echoes are based on a previous study conducted by Perret et al. (2023). To limit the number of WCIs with fluid-related echoes from GAZCOGNE1, we decided to include only 3 500, as exceeding this number did not improve performance on this particular marine expedition⁸. We included more GAZCOGNE1/EM302 fluid data as this MBES is technically closer to the EM122 than the Reson Seabat 7150, which means that fluid-related echoes may be more similar in terms of features in EM122 WCIs than in Reson Seabat 7150 WCIs.

The training was conducted over 100 epochs and inference was performed on all PAMELA-MOZ1 WCIs.

3 RESULTS

3.1 Fluid detection performance

The training of the network took 6 hours and 10 minutes and the inference on all PAMELA MOZ-1 pings (477 847 WCIs) took 5 hours and 27 minutes with 1 304 detections using a 0.6 confidence threshold. This represents a detection rate of 0.27% of the marine expedition, greatly reducing the number of WCIs that operators observed. The network detected typical elongated fluid-related echoes in WCIs (Fig. 4) and even some fluid-related echoes with a relatively small height above the seafloor (e.g. 100 m in Fig. 5), demonstrating its high sensitivity to this type of echo, regardless of interference or differences in frequency, aperture, and number of sectors in the WCIs.

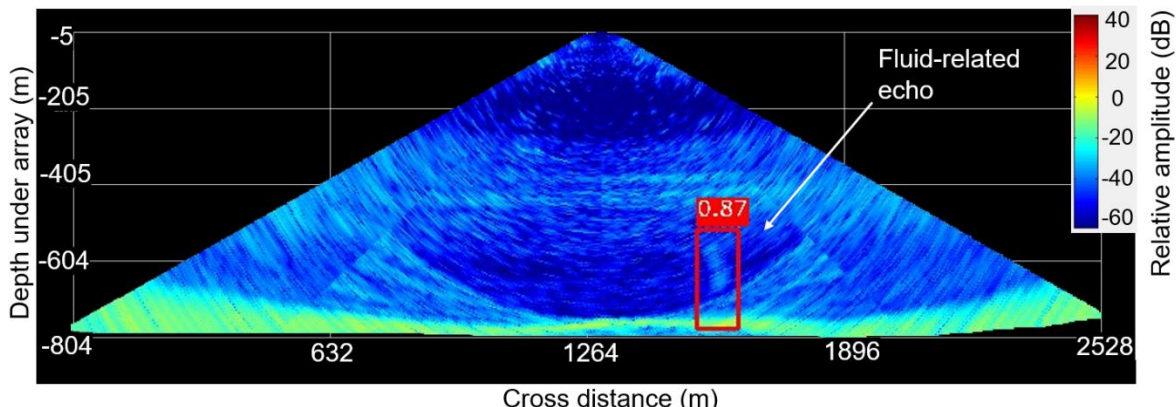


Figure 4: Good fluid detection made by the network (“medium” mode acquisition) on a typical elongated fluid-related echo which reaches 210 m above the seafloor.

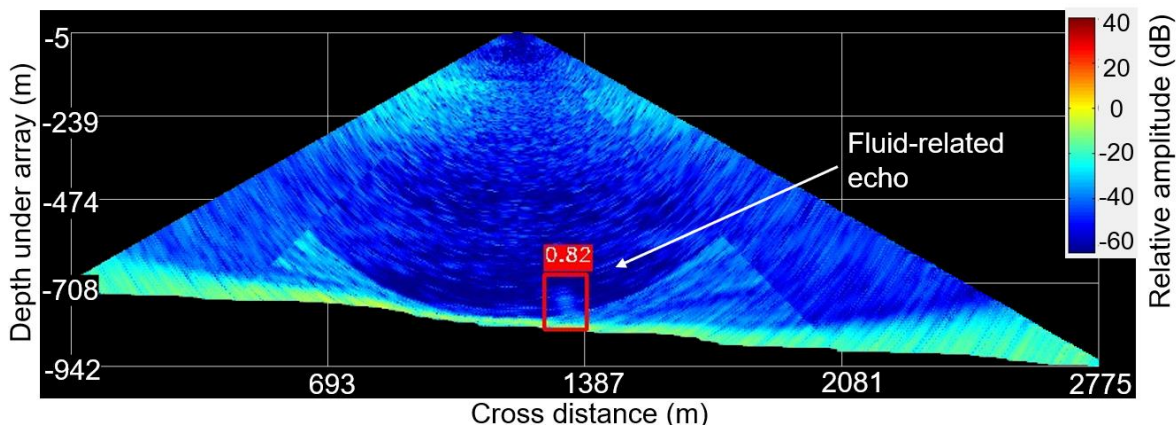


Figure 5: Good fluid detection made by the network (“medium” mode acquisition) on a small-height fluid-related echo which reaches only 100 m above the seafloor.

3.2 Errors made by the network on acoustic data

The majority of fluid outlet locations (88 %) identified by the human expert were found by the network (with a threshold of 0.6), except for four of them. However, the network identified three of these locations with a lower confidence threshold (0.27 to 0.34), with only one undetected echo, that is under MSR (Fig. 6). If the confidence threshold were lowered to 0.27 to include these three lower confidence detections, the PAMELA-MOZ1 dataset would contain 5 470 detections, accounting for 1.14% of the total dataset. Out of all the pings detected by the network with a 0.6 confidence threshold, 498 were true positives and 806 were false positives. The network is capable of detecting fluid under MSR (with 63 true positives at a threshold of 0.6). It is worth noting that false detections are mainly caused by 1) EM122 nadir emission sector changes (43% of false detections) (Fig. 7) and 2) coring operation (Fig. 8) (11% of false detections).

The composition of the training sets can easily explain these false positives. It is worth noting that medium and shallow configurations are not well represented in the training set, with only one line in medium mode and none in shallow mode. Similarly, WCIs acquired during coring operations are absent from the training set. The number of false positives for EM122 transmission sector changes (Fig. 7) is negligible, with 347 false detections including 21 in deep mode, 133 in medium mode and 193 in shallow mode. This is insignificant given the large number of water column images in these configurations: 146 521 WCIs for medium mode and 58 735 WCIs for shallow mode (while 272 591 for deep mode).

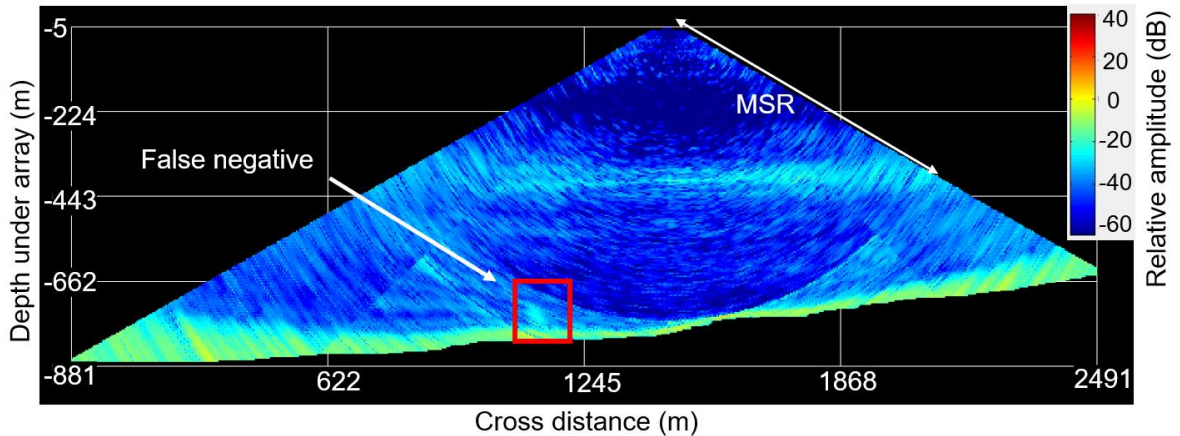


Figure 6: Single fluid-related echo not detected by YOLOv5. The echo was below the minimum slant range where noise from antenna sidelobes becomes visible (“medium” mode acquisition).

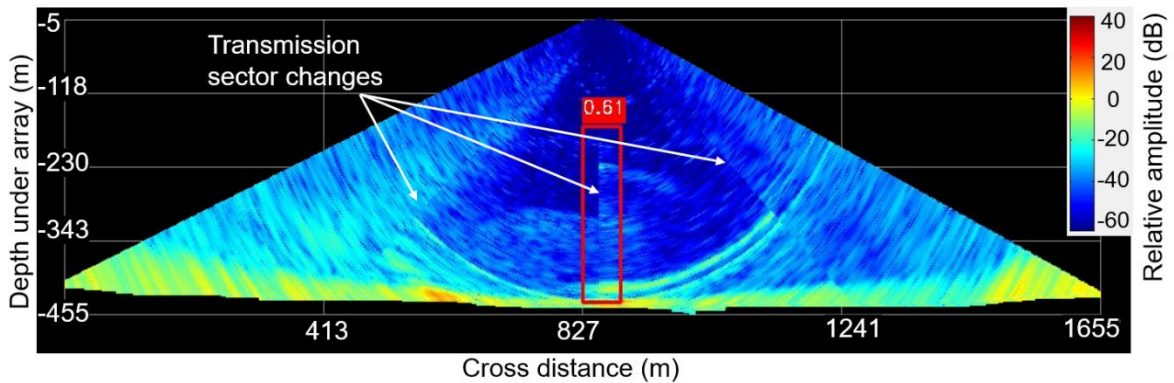


Figure 7: False positive made by the network on transmission sector changes at the nadir (“medium” mode acquisition configuration).

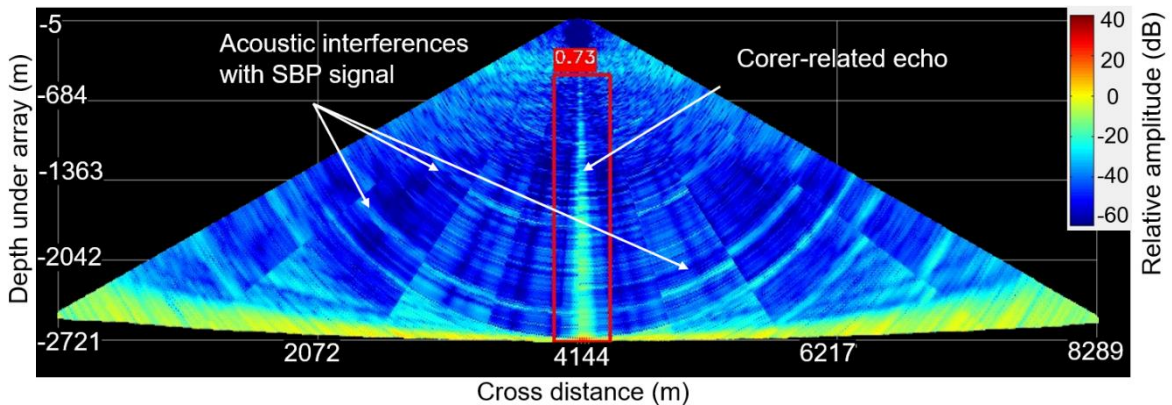


Figure 8: False positive made by the network on corer-related echo (“deep” mode acquisition configuration).

4 CONCLUSION

This study demonstrates that a YOLOv5 model, trained with data adapted to the acoustic context, can successfully transfer knowledge of 'fluid-related' echoes from one MBES to another while minimising false detections. The network was able to detect the majority of fluid outlets present in the

dataset, even under the Minimum Slant Range. Errors made by the network were mostly due to configurations not seen during training. Including these water column images in the training network will enhance its robustness for future marine expeditions with the same equipment deployment. An accurate representation of the range of acoustic signatures would consequently induce minimal false positives. Additionally, the network training and inference time of 11 hours and 37 minutes makes it suitable for use on-board. For instance, the network can be trained on data acquired at the beginning of the marine expedition and then utilised for the rest of the mission while further data is being acquired.

ACKNOWLEDGEMENTS

We wish to express our gratitude to the officers and crews of the research vessels *Le Suroît*, *Pourquoi pas ?*, and *L'Atalante*, as well as the technical staff from Genavir and IFREMER. The GAZCOGNE1 and PAMELA-MOZ1 marine expeditions were part of the PAMELA project and were co-funded by TotalEnergies and IFREMER for the exploration of continental margins. The GHASS2 marine expedition was co-funded by the Agence Nationale de la Recherche for the BLACK sea METHANE (BLAME) project and IFREMER. This study presents the findings of a PhD project that was funded by IFREMER and the Brittany region through an ARED grant. The author expresses sincere gratitude to Alison Chalm for English language revision.

REFERENCES

1. Y Fouquet, E Pelleter, C Konn, G Chazot, S Dupré, AS Alix, S Chéron, JP Donval, V Guyader, J Etoubleau JL Charlou, S Labanieh and C Scalabrin, "Volcanic and hydrothermal processes in submarine calderas: The Kulo Lasi example (SW Pacific)", *Ore Geology Reviews* **99**, 314-343. (2018)
2. G Judd and M. Hovland, *Seabed Fluid Flow: The Impact of Geology, Biology and the Marine Environment*, Cambridge University Press (2007)
3. X Lurton, G Lamarche, C Brown, V Lucieer, G Rice, A Schimel, and T Weber, "Backscatter Measurements by Seafloor-mapping Sonars. Guidelines and Recommendations", *GeoHab*, (2015).
4. K Colbo, T Ross, C Brown and T Weber, "A review of oceanographic applications of water column data from multibeam echosounders", *Estuarine, Coastal and Shelf Science* **145**(0) 41-56. (2014)
5. P Urban, K Köser, and J Greinert, "Processing of Multibeam Water Column Image Data for Automated Bubble/Seep Detection and Repeated Mapping: Processing of Multibeam Water Column Image Data", *Limnology and Oceanography: Methods* **15**(1) 1-21. (2017).
6. G Jocher, "ultralytics/yolov5: v6.2 - YOLOv5 Classification Models, Apple M1, Reproducibility, ClearML and Deci.ai integrations", *Zenodo* (2022)
7. Y LeCun, Y Bengio. and G Hinton, Deep learning. *Nature* **521**, 436–444 (2015).
8. T Perret, S Dupré, G Le Chenadec, A Gaillot and Y Lacroix, "Automatic detection of fluid emission echoes in water column acoustic data using Deep Learning approaches", *Proceedings of the 2023 International Symposium on Marine Geological and Biological Habitat Mapping (GeoHab '23)*, Saint-Gilles-Les-Bains, La Réunion Island, France, 66. (2023)
9. B Loubrieu, 2013. GAZCOGNE1 cruise, *Le Suroît R/V*. (2013).
10. V Riboulot, S Dupré, S Ker and N Sultan. GHASS2 cruise, *Pourquoi pas ? R/V*. (2021).
11. K Olu, PAMELA-MOZ01 cruise, *L'Atalante R/V*. (2014)
12. Z Zou, K Chen, Z Shi, Y Guo and J Ye, "Object Detection in 20 Years: A Survey", *Proceedings of the IEEE* **111**(3) 257-76. (2023)
13. F. Schroff, D. Kalenichenko and J. Philbin, FaceNet: A unified embedding for face recognition and clustering, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 815-823. (2015)