Supplementary Material

Deep-learning-based detection of underwater fluids in multiple multibeam echosounder data

Tyméa Perret*, Gilles Le Chenadec, Arnaud Gaillot, Yoann Ladroit, Stéphanie Dupré

*** Correspondence:** Corresponding Author: tymea.perret@ifremer.fr

Supplementary Texts

Text S1 YOLOv5 version - data augmentation applied

Version 1.12.0 (at least) of PyTorch and version 6.2 of YOLOv5 (at least) are required to guarantee reproducibility of previous tests. The training of YOLOv5 involves data augmentation techniques such as rotations (±45°), translations, scaling, mosaics, left-right image flipping, and HSV (Hue, Saturation, and Value) augmentation.

Text S2 Execution speed on different processing units

These networks were trained with an Nvidia Volta V100 GPU 32 GB. We used both a GPU and a more accessible and deployable work laptop for test inferences (11th Gen Intel(R) Core (TM) i7-11850H processor). YOLOv5 can process 41.7 Images Per Second (IPS) with one GPU and 3.8 IPS with the CPU. These IPSs are given by the total time it takes our algorithm to process the image (from loading WCIs to saving the results) and are averaged over the processing of 1000 images.

Supplementary Tables S1-S2

Table S1: Hyperparameter values used for Haar-LBP and YOLOv5 training. The threshold is used to merge the bounding boxes from the Haar and Local Binary Pattern cascades in Haar LBP. The average position of the detection boxes was calculated (Zhao et al. 2020) to retain one detection box only. The hyperparameters for Haar Cascade (on the left) and LBP cascade (on the right) are provided. For the Haar-Local Binary Pattern, hyperparameters were optimized through a grid search in our datasets. Other Adaboost hyperparameters can be found in Freund et al (1995). In the case of YOLOv5, LR refers to the learning rate, with 0 and f representing the initial and final learning rates, respectively. Further details on these hyperparameters can be found in Jocher (2021).

Dataset **Hyperparameters**	**GAZCOGNE1**	**GHASS2**	**Combined (GAZCOGNE1 and GHASS2)**
for Haar-LBP			
Threshold	150	300	300
Feature size	32 32	32 64	80 80
Number of stages	50 50	30 40	50 20
MiniHitRate	0.95 0.90	0.95 0.90	0.999 0.999
Acceptance Ratio Break	0.00001 0.0001	0.00001 0.00001	0.00001 0.00001
Max False Alarm Rate	0.5 0.5	0.5 0.5	0.5 0.5
for YOLOv5			
Number of epochs	Max 15 (overfitting if more)	Max 50 (early stopping)	Max 50 (early stopping)
Number of batches	16	16	16
LR0	0.00683	0.00773	0.00712
LRf	0.01094	0.01632	0.01095
Momentum	0.94614	0.92401	0.92773
Weight decay	0.00041	0.00051	0.00045
Warmup epochs	2.27960	2.44700	2.06290
Warmup momentum	0.68217	0.75565	0.77183
Warmup bias LR	0.11924	0.13766	0.10261

Table S2: Details of partitioning performed on the datasets of GAZCOGNE1 (Kongsberg EM302) and GHASS2 (Reson 7150).

Dataset Part	Number of water column images	Number of fluids manually pointed by operators	Remarks
GAZCOGNE1 Part 1	174 134	577	Shallow and medium acquisition modes. Fluid-related echoes in the northern transit, at the upper slope and at the continental shelf edge. Generally, few fish-related echoes with however abundance towards the end of transit.
GAZCOGNE1 Part 2	77 337	581	Shallow acquisition mode. Fluid-related echoes at the edge of the continental shelf.
GAZCOGNE1 Part 3	175 012	499	Shallow and medium acquisition modes. Fluid-related echoes at the edge of the continental shelf.
GAZCOGNE1 Part 4	69 769	556	Shallow acquisition mode. Fluid-related echoes at the edge of the continental shelf.
GAZCOGNE1 Part 5	349 370	544	Shallow acquisition mode. Fluid-related echoes and lots of biomass-related echoes.
GHASS2 Part 1	142 568	469	Five different pulse lengths used. Presence of dolphins. Survey mainly at the upper slope and occasionally at the shelf edge.
GHASS2 Part 2	83 089	472	Five different pulse lengths used. Presence of dolphins. Survey at the upper slope.
GHASS2 Part 3	161 416	455	Eight different pulse lengths used. Presence of dolphins. Survey mainly at the slope domain.
GHASS2 Part 4	24 486	463	Nine different pulse lengths used. Survey mainly at the slope and basin domains. Strong amplitude echoes on a hydrate ridge, and a moderate echo amplitude at a mud volcano.
GHASS2 Part 5	220 056	456	Eight different pulse lengths used. Survey mainly at the slope and basin domains. Strong amplitude echoes on a hydrate ridge.

Supplementary Figures

Figure S1

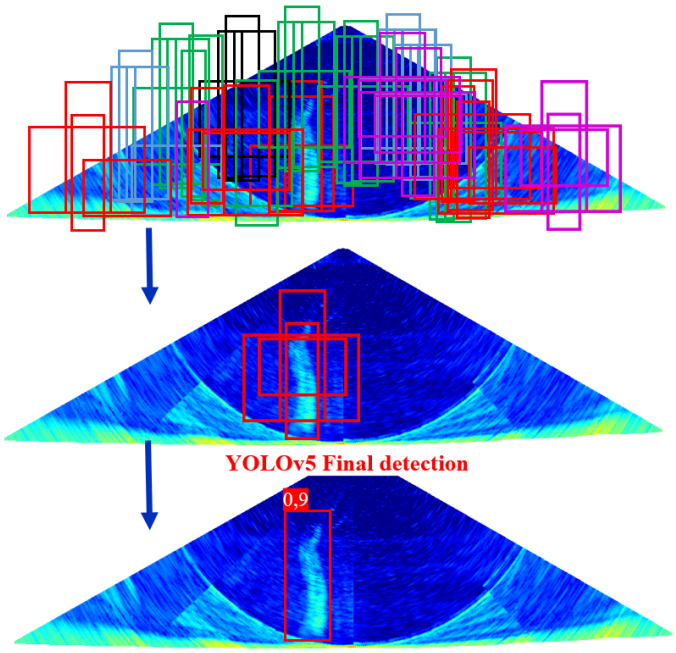
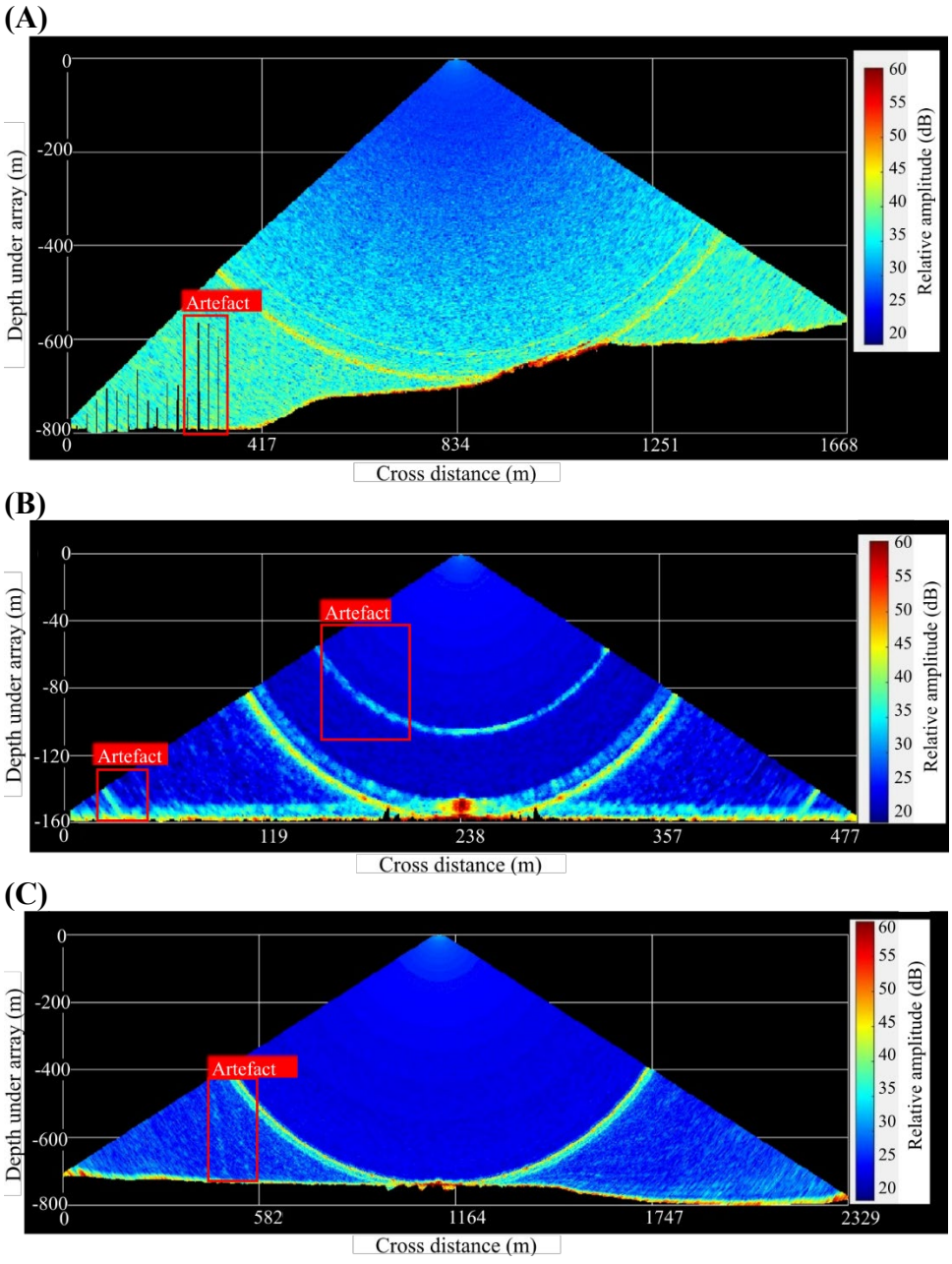


Figure S1 Example of box formations (inference) on a water column image with a fluid-related echo through three successive steps. **(A)** Anchor box propositions (one color for each x) at different scales and aspect ratios over the image grid cells, **(B)** YOLOv5 prediction of a confidence score with a probability that it contains an object and class probabilities for each anchor box. **(C)** Bounding box selection with the highest confidence score (0.9 in this case) for each detected object using Non-Maximum Suppression algorithm.

Figure S2


(A)
Depth under array (m)
0
-200
-400
-600
-800
Artefact
0
417
834
1251
1668
Cross distance (m)
Relative amplitude (dB)
60
55
50
45
40
35
30
25
20
(B)
Depth under array (m)
0
-40
-80
-120
-160
Artefact
Artefact
0
119
238
357
477
Cross distance (m)
Relative amplitude (dB)
(C)
Depth under array (m)
0
-200
-400
-600
-800
Artefact
0
582
1164
1747
2329
Cross distance (m)
Relative amplitude (dB)

(D)

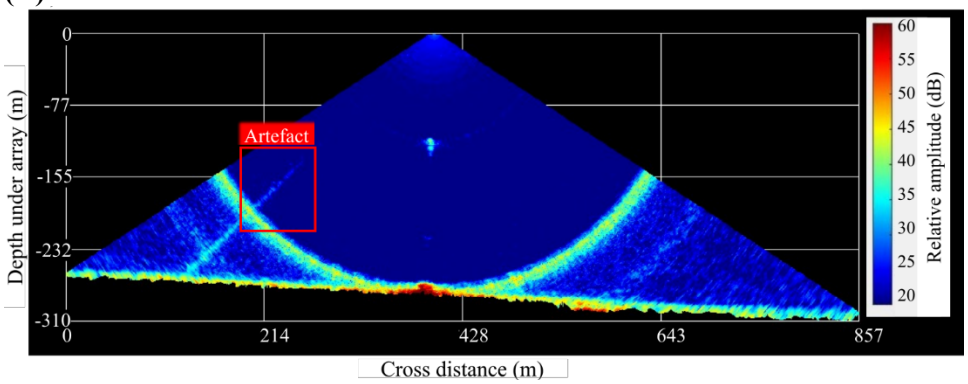


Figure S2 Examples of labels generated through inference performed by YOLOv5 in Reson Seabat 7150 multibeam echograms (GHASS2) initially detected by the model as fluids (on Experiments #1-3) and subsequently labelled as artefacts (Experiment #4). **(A)** Data loss during Sonarscope preprocessing, **(B)** Multiple seafloor echoes recorded by the sidelobes of the Reson Seabat 7150 receiving antenna, **(C)** Echoes resembling those caused by the presence of fluids but due to the sidelobes of the receiving antenna and linked to the seafloor echo. These sidelobes are higher than in the initially planned directivity. The defective sensors were verified at Ifremer by measuring the impedance of the transducers **(D)** Increased intensity over a full beam (sounder artefact, temporary malfunctioning of a transducer).

Figure S3

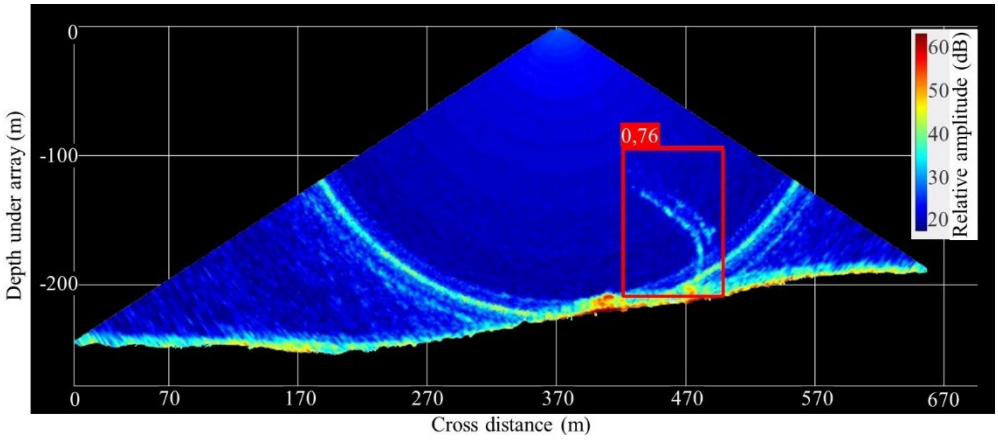


Figure S3 Example of a detection generated by YOLOv5 on Reson Seabat 7150 echograms, showing a fluid shape affected by bottom currents yet successfully detected. The model was trained using the basic GHASS2 training set, consisting exclusively of manually annotated fluid feet (Table 1), serving as the foundation for strategies #1–4.

Supplementary Videos S1-S5

To illustrate the performance of our model developed for fluid detection, we provide five videos of pings from two marine expeditions, GAZCOGNE1 and GHASS2. Results from the "basic training set" and the "enhanced training set" are displayed. The basic training set is composed solely of manually picked fluid feet (Table 1) and constitutes the starting point for strategies #1-4. The "enhanced training sets" result from combining the best YOLOv5 performance for each strategy (#2-5) (Table 6) with regard to the number of added WCIs with fluid, the percentage of added WCIs without fluid, and the percentage of added WCIs with acoustic and environmental artefacts. These "enhanced training sets" correspond to strategy #6 (first two columns in Table 7). Videos are available in MOV format at one frame per second for S1-3 and 0.6 frame per second for S4 and S5.

Video S1. GAZCOGNE1 data (basic training set only)

- Video S1 with 100 pings from GAZCOGNE1. This video displays polar echograms with many of the fluid-related echoes mainly located within the MSR (Minimum Slant Range). The basic training set model is already sensitive to fluids and confident in its detection score.

Video S2. GAZCOGNE1 data

- Video S2 with 60 pings from GAZCOGNE1. This video displays polar echograms with no fluid-related echoes. Beyond the MSR, the data become noisy. There are relatively few false positives detected with the basic training set and none with the enhanced training set.

Video S3. GHASS2 data (basic training set only)

- Video S3 with 100 pings from GHASS2. This video contains polar echograms with fluid-related echoes and some dolphin echolocation signals (Fig. 2E). The basic training set performs well on fluid detection without being impacted by echolocations.

Videos S4-S5. GHASS2 data

- Video S4 with 75 pings from GHASS2. This video displays polar echograms with fluid-related echoes. Both models (with basic and enhanced training sets) correctly detect fluid-related echoes that are located under the secondary lobes. The second model gives better detections with greater confidence on average. There are no errors on the acoustic artefacts in the nadir beams compared to the model trained with the initial configuration.
- Video S5 with 80 pings from GHASS2. This video displays polar echograms with no fluid-related echoes but with many typical echolocation signals, strong acoustic artefacts in the nadir beams and erroneous sea-bottom detections. The second model shows a superior ability to correctly not detect acoustic artefacts compared to the model trained with the initial configuration training set.