International Journal of Data Warehousing and Mining
Volume 18 • Issue 2

Boat Detection in Marina Using Time-Delay Analysis and Deep Learning

Romane Scherrer, ISEA, Universite de la Nouvelle-Caledonie, New Caledonia*
Erwan Aulnette, L2K Innovation, New Caledonia
Thomas Quiniou, ISEA, Universite de la Nouvelle-Caledonie, New Caledonia
Joël Kasarherou, L2K Innovation, New Caledonia
Pierre Kolb, L2K Innovation, New Caledonia
Nazha Selmaoui-Folcher, ISEA, Universite de la Nouvelle-Caledonie, New Caledonia
https://orcid.org/0000-0003-1667-3819

ABSTRACT

An autonomous acoustic system based on two bottom-moored hydrophones, a two-input audio board, and a small single-board computer was installed at the entrance of a marina to detect entering/exiting boats. Windowed time lagged cross-correlations are calculated by the system to find the consecutive time delays between the hydrophone signals and to compute a signal which is a function of the boats' angular trajectories. Since its installation, the single-board computer performs online prediction with a signal processing-based algorithm which achieved an accuracy of 80%. To improve system performance, a convolutional neural network (CNN) is trained with the acquired data to perform real-time detection. Two classification tasks were considered (binary and multiclass) to both detect a boat and its direction of navigation. Finally, a trained CNN was implemented in a single-board computer to ensure that prediction can be performed in real time.

KEYWORDS

Autonomous Real-Time System, Deep Learning, Passive Boat Detection, Time Series Classification, Underwater Noise

INTRODUCTION

Since the invention of sonar, a growing interest was given to underwater sounds that radiated from boats. Although the first applications were related to military purposes, several studies aim at using the underwater sounds for non-military uses, such as maritime traffic management (Fillinger, 2009; Zwemer, 2018), underwater surveillance (Fillinger et al., 2010), assessment of the impact of noise pollution on marine life (Codarin, 2009; Holles, 2013), among others.

Since the acoustic signal produced by a ship has many sources (propeller, machinery, hydrodynamic, vibrations, etc.) that produced tones at different frequencies, most developed methods to detect, classify or track boats are applied in the frequency domain or time-frequency domain. Several approaches were developed to detect the harmonic frequencies in the signals and to extract the acoustic

DOI: 10.4018/IJDWM.298006 *Corresponding Author

signature of the ships. In many cases, these methods are based on the spectrum (Guo et al., 2020), DEMON spectrum (Chung et al., 2011) and Cepstrum (Das, 2013; Santos-Domínguez, 2016) and the automatic detection is usually performed by detecting the peaks with a threshold (Reis et al., 2019).

In recent years, deep neural networks have seen a lot of successful applications in many different domains. One successful deep learning architecture used in computer vision is a convolutional neural network (CNN). This architecture is known to automatically learn complex feature representations using its convolutional layers and has led to impressive results in many problems such as in image classification (Krizhevsky et al., 2012), speech recognition (Palaz et al., 2015) or time series classification (Cui, 2016; Guennec, 2016; Zhao, 2017). More recently, several studies (Li, 2019; Yamaguchi, 2019) aimed at detecting boats have used machine learning methods, which have the advantage of automatically extracting characteristic attributes to classify the data. These methods are often more robust to noise and do not require hand designed features. However, all the proposed methods performed the detection on a 2D-signal in the frequency domain of the recorded sound. In this article, the researchers aimed to show than the detection itself can be performed with a simpler 1D signal that does not required the computation of the spectrum or the use of a 2 dimensional CNN. With a 1D signal, the detection can be performed with a time series classifier.

The research work is part of a project which aims to improve the safety management of a marina by automatically detect boats departure or arrival using a low-cost acoustic system based on two hydrophones and a single-board computer. The system is still in development and since its installation in July 2019, it had saved thousands of signals that need to be processed to improve the true detection rate. In this paper, the researchers investigated the possibility of replacing the initial algorithm with a trained time series classifier to perform online real-time classification with an embedded system.

The paper is organized as follows. The first section introduces the methodology to track a boat and gives an overview of the data acquisition process performed by the acoustic system. The next section describes the data preparation. The last two sections present the different classifiers evaluated to perform the detection on the autonomous acoustic system.

METHODOLOGY

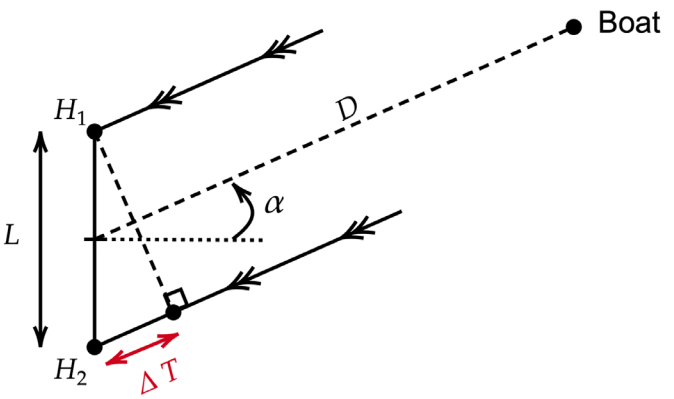
Passive acoustic systems used to track an underwater sound are frequently based on several hydrophones connected to a central unit (Fillinger, 2010; Guo, 2020). In order to detect a boat without computing the spectrum, the time delay between the signals received by the hydrophones can be measured to estimate a boat's position. Unlike the previous cited methods that use the spectrum to detect the presence of boats in a noisy environment, the authors of this article aimed to exclusively use the time delay between the hydrophones, which is faster to compute than the spectrum by the single-board computer on their acoustic system. In this section, the methodology to compute the signal and their acoustic system that have been developed to improve the security of a marina are presented.

Time Delay

Let us consider two hydrophones and H_2H_1 that are positioned at the entrance of a marina (Figure 1). They are separated by a distance L and we denote $h_i\left(t\right)$ the signal received by the hydrophone H_i. The position of the boat makes an angle $\alpha\left(t\right)$ with the normal to the segment between the hydrophones. Considering that $D \gg L$ (far field conditions) and denoting c the velocity of sound in sea water (approx. 1500 m/s), it can be shown that the sound wave emitted by the boat reaches the hydrophones with a time delay:

$$\Delta T\left(t\right) = \frac{L}{c}\sin\left(\alpha\left(t\right)\right)$$

Figure 1. The signals recorded by the hydrophones are delayed versions of the same signal



One can notice that the time delay ΔT depends on the boat angular position and changes with the movement of the source. Thus, in order to determine the direction of a boat (arrival or departure), one can exclusively examine how the time delay between $h_1(t)$ and $h_2(t)$ is changing over time.

The well-known cross-correlation is a measure of similarity between two signals and can be used to find the time delay (displacement) of one signal relative to the other. For two continuous functions $h_1(t)$ and $h_2(t)$, the authors recall that the cross-correlation is defined as:

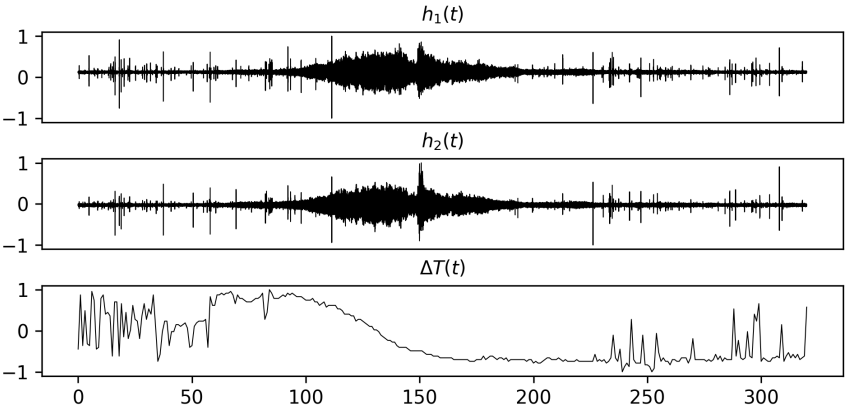
$$Xcorr(h_1, h_2)(\tau) = \int_{-\infty}^{+\infty} h_1^*(t) h_2(t+\tau) dt$$

where $*$ denotes the complex conjugate and τ is the displacement or lag. If the signals $h_1(t)$ and $h_2(t)$ are delayed by ΔT, then the cross-correlation has a maximum at $\tau = \Delta T = \frac{L}{c} \sin(\alpha)$. Thus, to track the boat's angular trajectory $\alpha(t)$, the signals $h_1(t)$ and $h_2(t)$ are cross correlated every N points. Then, the argument of the maximum in the cross-correlation is detected and appended in an array. In this paper, this array will be called the "**trajectory signal**".

Note that, in order to save computing time, the cross-correlation $Xcorr(\tau)$ does not need to be calculated for every lag τ. Indeed, (ignoring multi-path propagation) the noise radiated by a boat reaches the hydrophones with a maximum time delay equal to $|\Delta T| \leq L/c = \tau max$. As a result, the cross-correlation can be calculated for every $\tau \in [\![-\tau max, \tau max]\!]$ and the argument of the maximum is only picked in this window.

Figure 2 shows an example of the described process for a recorded boat. The trajectory signal computed by cross-correlation has a sinusoidal shape which is consistent with the expression of the time delay. Moreover, the direction of the boat can be deducted from the sign of the signal before and after the zero-crossing. In this example, the sign of the signal changes over time from a positive value to a negative one. Thus, for our hydrophone configuration, this example corresponds to a boat on departure. It is worth noting that the shape of the trajectory signal depends on many factors such as the boat speed and its trajectory. Furthermore, the ambient noise and potential inferences also contribute to the cross-correlation and make the resulting computed final signal very noisy (see figure 5 for more examples).

Figure 2. h1,h2 and the corresponding trajectory signal for a departing boat



Autonomous System and Data Acquisition

A low-cost and autonomous system (Figure 3) is under development to improve the management of a marina. The main purpose is to determine in real time boats departure or arrival and to quantify the ship traffic inside a marina. Since July 2019, the system is operational and had saved thousands of data that are still be processed to improve the system accuracy and robustness. The passive acoustic system uses two omnidirectional hydrophones (H1A, Aquarian hydrophones) that have been bottom-moored at the entrance of Port Brunelet marina (Figure 4) in Noumea, New Caledonia. The hydrophones are separated by a distance $L = 0.8$ m and are placed 2 m below sea level. They are connected to an ultra-low latency audio cape (Bela: https://bela.io/products/) and a small single-board computer (BeagleBone Black: https://beagleboard.org/black) that perform the recording at a sampling frequency of 44.1 kHz. Every 186 ms, the cross-correlation between the hydrophones signals is performed and the corresponding maximum lag is stored in a file which is saved on a SD card after 10 minutes. The acoustic system is fully autonomous and relies on a solar panel that charges a battery. Thus, in order to save the battery and because there is no departure or arrival after nightfall, the system shuts down at 6pm and reboots at 7am. During daytime, the system saves 6 trajectory files per hour.

During the early stages of the project development, a first algorithm was implemented to detect boats from the background noise and to determine their trajectories. The algorithm proceeds as follows. First, for every change of sign in the trajectory signal, a window of temporal span equal to N centered

Figure 3. Autonomous system

Figure 4. Port Brunelet - red dots show hydrophones position. From http://carto.noumea.nc

on the zero-crossing is applied. Then, in this window, the number of points having the same sign before and after the zero-crossing was calculated. Finally, if this number is greater than a manually set threshold then the algorithm predicts that a boat is departing or arriving at the marina. However, since the trajectory signals are very noisy and depend on the boat speed, a static threshold could not lead to a high true detection rate. For this reason, the accuracy of this algorithm was evaluated at 80%. To improve the system robustness, the data acquired by the system from July 2019 to November 2019 was processed to train a time series classifier to perform a rapid and better prediction.

Data Preparation

For each saved file, the trajectory signal is first split up into small time series of length 322 points (corresponding to 1 minute of observation) and overlapping by 50 points. The length of the time series was set to 322 for several reasons. Firstly, depending on its speed, the departure or arrival of a boat takes 20 to 30 seconds, which correspond to approximately 100 – 160 points in the trajectory signal. Thus, a time series of length 322 points is sufficient to detect the boat's trajectory. Secondly, the main disadvantage of using exclusively the cross-correlation to track boats is the difficulty to separate the contributions of several boats (for e.g. two boats arriving at the marina at the same time). Even though the trajectories of several ships can be separated if their contributions to the cross-correlation do not overlap (Fillinger et al., 2011), this separation needs to be done before storing the trajectory signal. However, in the first stage of development, the autonomous system only tracked in real time one boat and did not store the raw hydrophones signals. Hence, in the presence of several boats, the stored trajectory signal tracks the louder boat and can "jump" between the boats when one becomes louder than the other. As a result, taking a time series of length 322 helped decrease the likelihood of having several departures/arrivals during the same period of observation. However, all the time series that did show several boat's trajectories were removed from the dataset (approx. 40).

Approximately 10,700 time series were manually labeled depending on their natures (background noise, arriving boat or departing boat). The background noise here refers to white noise, biological noise and all other noises that radiated from boats outside or inside the marina but that do not cross the acoustic barrier. In order to increase the number of time series that show a departing/arriving boat (under-represented classes in the dataset), the authors performed a trivial data augmentation, which consists of taking the opposite sign of the time series. Indeed, the angular trajectory of a departing boat is the opposite of that of an arriving boat. After performing this data augmentation, the trained CNN gives better classification performances for all classes with a significant improvement concerning the background class. Indeed, as previously explained, the time series in this class can highly vary in shape depending on the background noise. For example, for a noisy moored boat inside the marina,

the trajectory signal keeps an overall negative sign while for a boat that navigates outside the marina, the trajectory signal keeps a positive sign. For both scenarios, the boats clearly contribute on the cross-correlation but the corresponding trajectory signal does not show the typical sinusoidal shape that characterizes a moving boat that crosses the acoustic barrier. As a result, taking the opposite sign of the all the time series during training improved the overall robustness and performances of the model on the validation set.

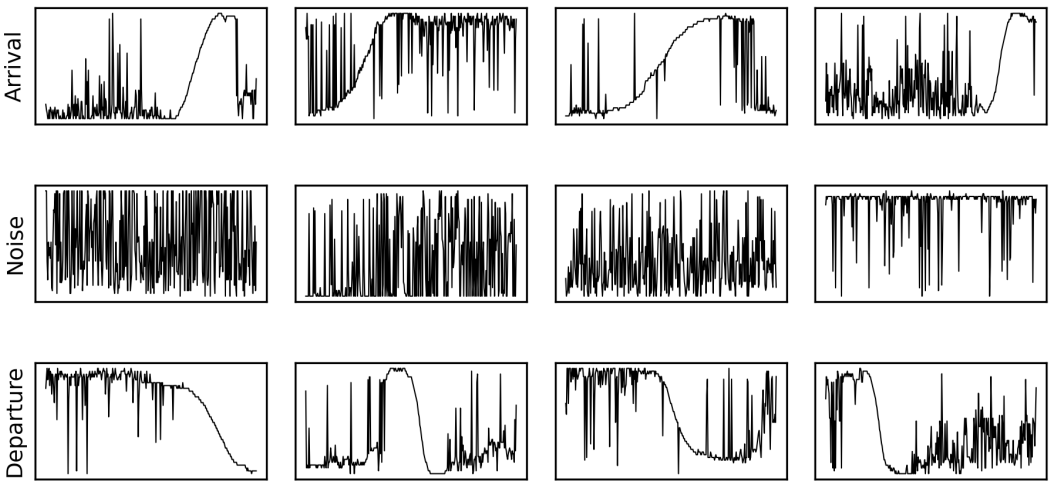
After performing the data augmentation, the dataset consists of 19,844 background noise time series and 1,534 boat radiated noises (767 departures and 767 arrivals). Finally, the time series were normalized using a MinMax scaler (between 0,1). No further processing/transformation were performed to remove the noise or to improve the quality of the signals. Figure 5 presents some examples of time series per class.

Time Series Classifiers

The main objective of the project is to detect a boat from the background noise with a short 1D signal so that the detection algorithm is light (few weights that need to be stored in the memory of the single-board computer) and has a low computational complexity so that the prediction can be performed in real time.

To find an algorithm that meets these two constraints and has a high detection rate, 5 classifiers de time series were considered: (1) the Bag-Of-SFA-Symbols in Vector Space classifier (BOSS VS) proposed by (Schäfer, 2015) that combines a BOSS model with a vector space model, (2) the time series forest classifier (TSF) proposed by (Deng el al., 2013) that employs a combination of entropy gain and a distance measure, (3) the KNeighborsClassifier (KNN) build on sklearn (Predregosa et al., 2011) that employs a DTW distance, (4) the Time-CNN proposed by (Zhao et al., 2017) and (5) a modified t-LeNet (Guennec et al., 2016). The Time-CNN model is composed of two 1D-convolutional layers with a kernel size of 7 and a sigmoid activation function. The layers use 6 and 12 filters for the first and second layer, respectively. Each convolutional layer is followed by an average-pooling layer with a pool size of 3. The classification is performed with a softmax layer. The t-LeNet model, which is a time-series specific version of LeNet model (LeCun et al., 1989), consists of two 1D-convolutional layers followed by a fully connected layer (FC) and a final Softmax classifier. For both convolutions, the ReLU activation function is used with a filter of temporal span equal to 5. The first convolutional

Figure 5. Four time series examples per class



layer uses 5 filters and is followed by a max pooling of size 2. The second convolutional layer uses 20 filters and is followed by a max pooling of size 4 and a dropout layer with a rate equal to 0.5. The last layer is followed by a flattening layer and a fully connected layer. Finally, the output layer has a number of neurons equal to the number of classes and uses a Softmax activation. Note that the Time-CNN and t-LeNet models have the same number of convolutional layers but the activation functions, number of filters and kernel sizes are different. Moreover, t-LeNet has an extra FC layer before the softmax layer and is trained to minimize the cross entropy loss during training while Time-CNN is trained to minimize the mean square error.

The problem of detecting a boat in the marina can be done using two different classification tasks. On the one hand, this problem can be seen as a multiclass classification with three classes: departing boat, arriving boat and background noise. This classification task allows the detection of the boat and gives its direction so that the traffic inside the marina can also be quantified. On the other hand, all the time series that show a boat event (regardless of the boat's trajectory) can be grouped in the same class and one can consider this problem as a binary classification task (background noise vs boat). Even though this classification task cannot be used to quantify the traffic inside the marina, its simplicity compared to the multiclass classification gives an overview of the performance of the different classification models. For this reason, the best classifier, among the 5 tested in this article, was selected based on its performances on the binary classification task only. The best classifier was then retrain on the multiclass classification task for a deeper evaluation of its performances.

To choose the best classifier that will run on the acoustic system, all the models have been evaluated based on a typical train-validation-test approach. The dataset was randomly split into three subsets with a ratio of 60:20:20 for training, validation and testing, respectively.

The two CNN-based models were trained using the Adam optimizer with an initial learning rate set at 0.001 (0.0001 was also considered) and a decay of 0.9. During training, the loss (cross entropy for t-LeNet and mean square error for Time-CNN) is computed and heightened to force the model to "pay more attention" to samples from the under-represented classes. The number of epochs was initially set to 500 but to prevent the model from over-fitting, the training process is stopped when the loss on the validation set does not improve for 10 consecutive epochs. For each scenario, 40 t-LeNets have been trained with different hyperparameters (learning rate, batch size, number of neurons in the FC layer for t-LeNet and kernel initializers) and their performances on the validation set are compared to select the best model. Finally, tables 1, 2 and 3 report the test set performance of the best model for each classification task.

RESULTS

Binary Classification

For this scenario, all the time series that detect a boat from the background noise (regardless of its direction) were grouped in the same class called "boat". This classification task was used to select the best classifier among the 5 tested. A set of 4 hyperparameters with different values was considered during the training of the t-LeNet models: number of neurons in the fully connected layer (500, 400, 300, 200 and 100), batch size (32 and 64), initial learning rate (0.001 and 0.0001) and kernel initializer (random normal and glorot uniform). Since the Time-CNN model does not have a fully connected layer to tune, only 3 hyperperameters were tested with the same batch size, initial learning rate and kernel initializer sets used to tune the t-LeNet models. The BOSS SV, TSF and KNN models were trained and evaluated once on the training set and on the testing set, respectively.

All the classifiers were evaluated on five metrics:

1. **Recall** or **True Positive rate** which is defines as the fraction of time series from the "boat event" class which are correctly predicted by the model.

2. **True Negative rate (TNR)** which is defined as the fraction of time series from the "background noise" class which are correctly predicted by the model.
3. **Precision** that quantifies the number of positive class prediction that actually belong to the "boat event" class.
4. **F1-score** which combines precision and recall:

$$F1 - score = 2 \times \ precision \times \ recall \ / \left(precision + recall \right)$$

5. **Balanced accuracy** which is defined as the average of the True Positive Rate and True Negative Rate.

The best t-LeNet and Time-CNN models were selected based on their F1-score on the validation set. Figure 6 summarizes the performances of the 5 classifiers on the same testing set. The BOSS SV, Time-CNN and KNN classifiers performed poorly on the classification task and obtained a precision score of less than 80%. The two best classifiers are TSF and t-LeNet that obtained a 0.895 and 0.952 precision score, respectively. Since t-LeNet obtained a better score on each of the 5 metrics used to compared the classifiers, it was selected to perform the boat detection with the acoustic system and was retrain on the multiclass classification for a deeper evaluation of its performances.

Figure 7 presents the binary classification results of 40 t-LeNets that have been trained with different hyperparameters. As a result, the chosen model was built with 200 neurons in the FC layer and a glorot uniform kernel initializer (also called Xavier normal initializer). An initial learning rate of 0.001 and a batch size of 32 was set for the training.

Table 1 shows the confusion matrix obtained with the best t-LeNet on the test set. For 304 time series in the boat class, 280 are correctly predicted by the model while 24 are mistaken with background noise. Concerning the background noise class, only 14 out of 3972 time series are mistaken with a departure or an arrival. This model gives a balanced accuracy of 0.96, a F1-score of 0.94, a precision of 0.95, a recall of 0.92 and a TNR of 0.99.

Figure 6. Classifiers performances on the testing set

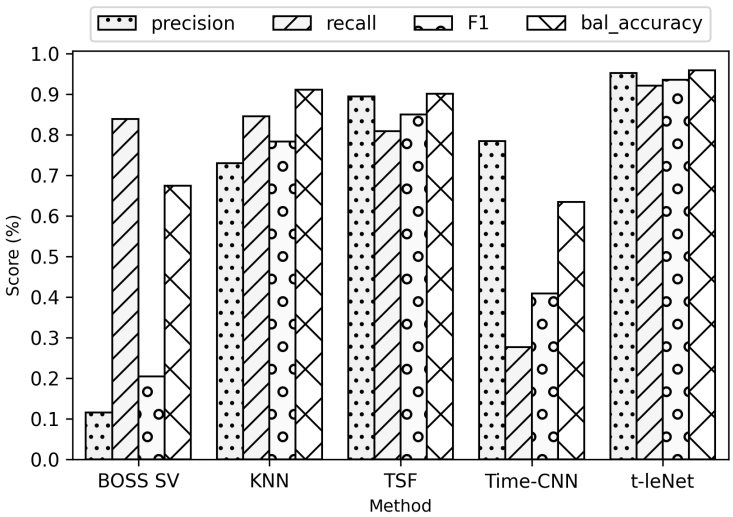


Figure 7. Binary classification scores for different hyperparameters obtained on the validation set [Number of neurons, batch size, learning rate, kernel initializers ('RN': random normal, 'GU': glorot uniform)]

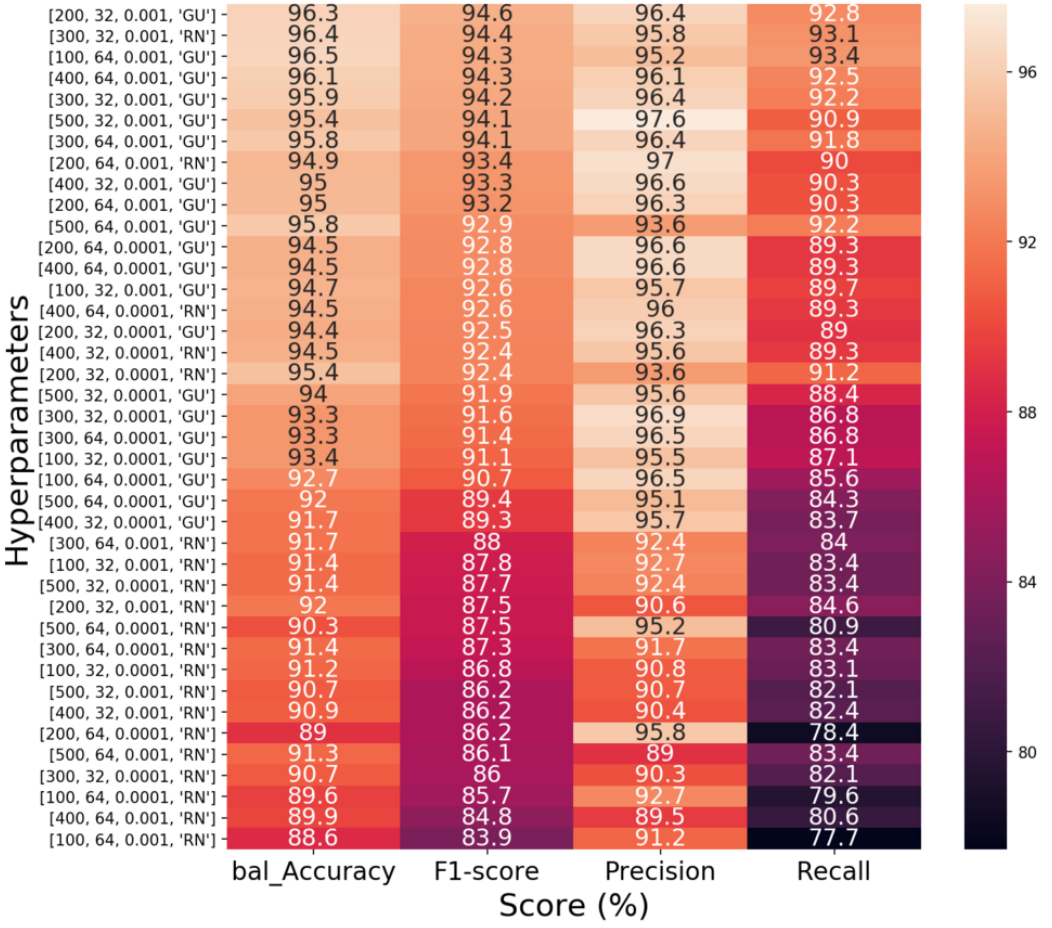


Table 1. Binary classification confusion matrix on the test set

True Class	Predicted Noise	Predicted Boat
Noise	3956	14
Boat	24	280

Multiclass Classification

For this scenario, the models have are trained to detect boats from the background noise and to determine their directions. Compared to the previous binary task, this scenario can be interesting to evaluate the traffic in the marina. Thus, three classes have been considered: "departure", "arrival" and "background noise". Since t-LeNet performed better than the other classifiers on the binary classification task, it was retrained for this task and tuned with different hyperparameters. The best one was chosen based on its F1-score on the validation set (see Figure 8).

Table 2 shows the multiclass confusion matrix obtained with the best t-LeNet on the test set. The ability of the model to identify the presence or absence of boats is indicated by the fact that only

Figure 8. Multiclass classification scores for different hyperparameters obtained on the validation set [Number of neurons, batch size, learning rate, kernel initializers ('RN': random normal, 'GU': glorot uniform)]

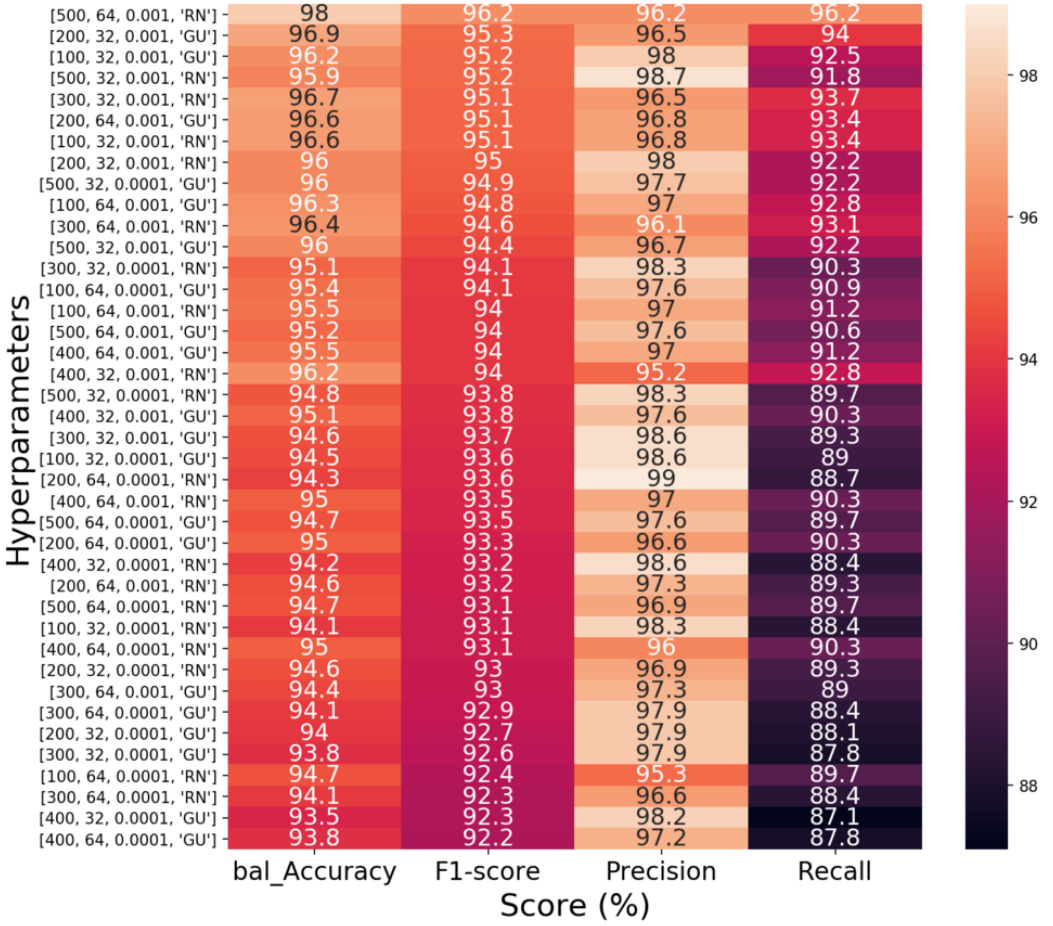


Table 2. Multiclass classification confusion matrix on the test set

TrueClass	Predicted Departure	Predicted Noise	Predicted Arrival
Departure	147	10	3
Noise	5	3960	7
Arrival	3	9	132

19 out of 304 time series that show an arrival or a departure are mistaken with background noise. Moreover, only 12 out of 3,972 time series from the background noise class are mistaken with a boat. As a result, this model is better at detecting a boat from the background noise than the previous model that performs the binary classification. Furthermore, this model is capable of predicting the direction of boats since there are only 6 confusions between the departure class and the arrival class.

To further evaluate the classifier performance on the test set, a one-vs-rest transformation is performed on the multiclass confusion matrix. Thus, for every class i a binary confusion matrix is

calculated such that the class *i* is considered as the positive class and the classes $j \neq i$ as the negative class. Finally, for each class, the balanced accuracy, precision, recall, F1-score and TNR are presented on Table 3. One can notice that the t-LeNet performances on the departure and arrival classes are substantially similar and give the following mean scores: a balanced accuracy of 0.96, a recall of 0.91, a precision of 0.94 and a F1-score of 0.93.

Model Performances on Multiple Boats Trajectories

During the data preparation, the time series with multiple boat trajectories were removed from the training set due to the low number of instances. Among the labeled signals (Figure 9), 12 are representing two boats that are arriving at the marina, 14 are representing two boats that are departing from the marina and 14 are representing two boats that are crossing each other.

Since the trajectory signal is computed with a windowed-cross correlation between the hydrophone signals, the system can only tracked in real time the louder boat. Hence, in the presence of several boats the signal "jumps" between the boats when one becomes louder than the other. If the boats are sufficiently spaced from each other, their individual trajectories are clearly visible and separable, but otherwise the trajectory of one boat may obscure the other (for e.g. the 4th time series in the class "2 arrivals").

Although the models were not trained for this scenario due to the small number of labeled signals, their ability to detect the boats was evaluated on this dataset that was augmented by flipping and reversing the time series and by adapting the ground-truth labels accordingly. The results are given in Table 4 and Table 5 for the binary model and multiclass model, respectively.

Table 3. Multiclass classification scores on the test set

	Bal.Acc	Recall	Precision	F1-Score	TNR
Departure	0.958	0.919	0.948	0.933	0. 998
Noise	0.968	0.997	0.995	0.996	0.938
Arrival	0.958	0.917	0.930	0.923	0.998

Figure 9. Time series with multiple boats

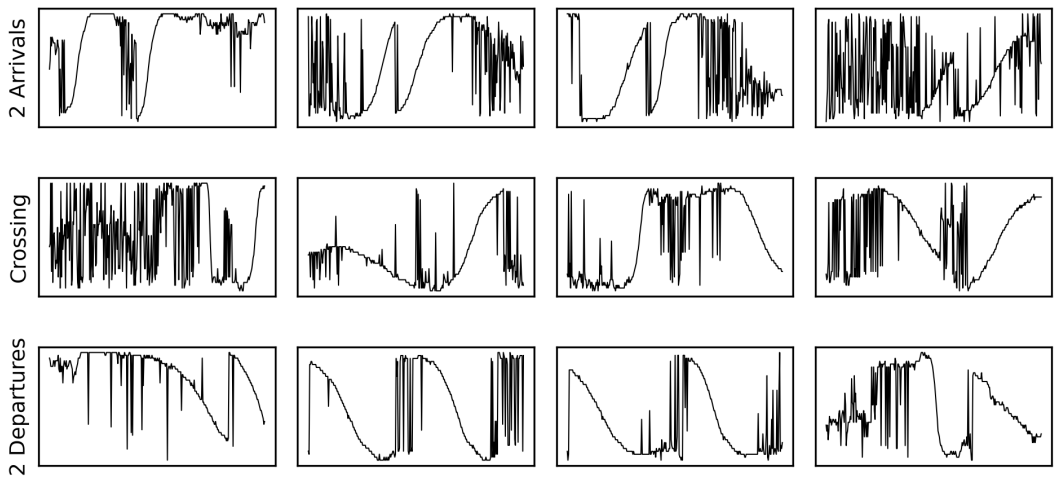


Table 4. Binary model - confusion matrix on multiple boats data

True Class	Predicted Noise	Predicted Boat
2 Arrivals	4	48
Crossing	5	51
2 Departures	5	47

Table 5. Multiclass model - confusion matrix on multiple boats data

True Class	Predicted Departure	Predicted Noise	Predicted Arrival
2 Arrivals	15	3	34
Crossing	22	6	28
2 Departures	31	5	16

The model trained to perform a binary classification (noise vs boat) is able to detect the boats with only 14 time series with multiple boats that are mistaken with a background noise. The model that performs the multiclass classification is also able to detect boats. When several boats sail in the same direction, the model tends to determine the right direction of navigation. When two boats are crossing each other in opposite directions, the model predicts either a departing or arriving boat, which was expected. If the time series are grouped together in the same class regardless of the direction of navigation, the performances of the multiclass model are similar to those of the binary model with 14 time series among the 160, which are confused with noise. These results show that the two models are able to detect the presence of boats for this scenario but that they cannot determine the direction of navigation or the number of boats that are passing simultaneously since they were not trained for this task. It can be problematic to accurately quantify the traffic in the marina because the count of departures/arrivals during the day may be wrong.

In order to train a model that is able to detect multiple boat, a large amount of data is needed. To increase the number of signals that show several boats simultaneously, 3,000 random boat trajectories with noise have been simulated with the equation $\Delta T\left(t\right) = \frac{L}{c}\sin\left(\alpha\left(t\right)\right)$ (see Figure 10). All the cases have been considered for the simulations, namely, two boats departing, two boats arriving or two boats crossing each other. A t-LeNet model was trained to predict 3 classes (“noise”, “1 boat” and “2 boats”) with only simulated time series for the “2 boats” class and only real time series acquired by the acoustic system for the other two classes in the training set. By testing the performances (Table 6) on real time series, the results showed a lot of confusion between the classes “1 boat” and “2 boats”, which is probably due to the fact that the simulated data of several boat trajectories are not sufficiently close to the real data (different noise, non-rectilinear trajectories, etc.). However, it can be noted that no real test time series belonging to the “2 boats” class is confused with noise, which shows that the simulated data used during the training step helped to reduce the false negative rate of the classifier.

The results obtained with the simulated data tends to indicated that the detection of multiple boats might be possible if enough and accurate data are obtained by the system. However, during the data preparation, 21 days of recorded signals have been processed and manually labeled. Among the data, only 40 time series show multiple boats, which proves that this scenario does not happen very often in a small marina. For an accurate detection inside a larger marina where the traffic might be higher, more data will be needed to train a model. To acquire this data, the model that performs the multiclass

Figure 10. Comparison of real data and simulated data

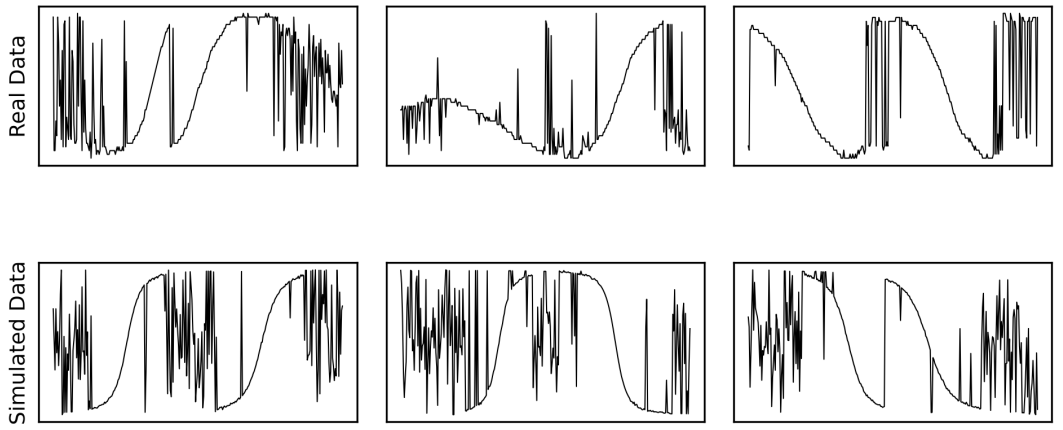


Table 6. Multiple boat trajectories model – confusion matrix on test data.

True Class	Predicted Noise	Predicted 1 boat	Predicted 2 boats
Noise	3955	14	3
1 boat	16	253	35
2 boats	0	96	64

classification will be implemented into the autonomous acoustic system and its prediction will trigger the recording of the raw signals perceived by the hydrophones. These signals will subsequently make it possible to classify boats with their acoustic signatures by detecting characteristic peaks in the spectrum and will increase the size of the presented dataset. By adding several noises radiated from different boats, it will also be possible to simulate a very large variety of H1 and H2 signals perceived by the hydrophones that present several boat trajectories. By using the same algorithm used in the system that computes the trajectory signal from H1 and H2, it will be possible to simulate times series with multiple boats that are more accurate than the previous simulated signals that have been used to train the model in this section. The future acquired data could be used to train a model to make a multiclass classification that takes into account the detection and the direction of navigation of one or more boats.

Trained CNN Implementation on a Single-Board Computer

The t-LeNet model which performs the multiclass classification was selected to replace the signal processing-based algorithm that was used during the first stage of development of the system. To ensure that a trained CNN can run on a small single-board computer, the CNN was loaded on a raspberry PI 3 model B+ running with the PI OS operating system (previously called Raspbian). The board equips a Broadcom BCM2837B0 SoC which has a quad-core ARM Cortex-A53 cluster, running at 1.4GHz. To measure the execution time, only one CPU core of the Raspberry Pi was used by the TensorFlow library when performing the CNN inference operations. The inference time to perform the classification of one time series was measured 10,000 times in a control loop. As a result, the mean inference time to predict whether a boat is departing or is arriving at the marina is

9.75ms. This result shows that the prediction can be done in real-time with an embedded system. Further works will be needed to implement the CNN in our developed software (memory allocation, task parallelism, etc.) but the CNN prediction will be used as a trigger to record the raw hydrophones signals when a boat is detected.

CONCLUSION

In this paper, the authors have presented their autonomous system, which is based on two hydrophones connected to a two-input audio board and a small single-board computer. The article describes how the system uses the hydrophone signals to build a passive acoustic barrier that detect the boat traffic inside a marina. The signals from the hydrophones are cross-correlated by the system to find the time delay and to compute a signal that depends on the boat trajectory. By examining how this signal is changing over time, the system can detect a boat and its direction. The presented method can be considered as a dimensionality reduction technique where the two 1D signals recorded by the hydrophones at a frequency of 44.1 kHz are reduced to a single 1D signal of 322 points for one minute of listening. This reduction greatly simplifies the problem because the signal can be calculated in real time on a single-board computer. Moreover, the detection itself can be carried out more rapidly on a 1D signal than on a 2D spectrum, which is necessary in the case of a detection applied in the time-frequency domain as it can be seen in the literature.

Initially, a threshold-based detection algorithm was implemented into the acoustic system to detect boats in real time. To improve the robustness of the system, the researchers have processed 21 days of saved signals and manually labeled them depending on their natures (background noise, arriving boat and departing boat) to train several different classifier. As a result, the built dataset consists of 21,378 time series with 19,844 background noise and 1,534 boat radiated noise. Three different time series classification tasks are experimented and the results presented show that the signal calculated by cross-correlation, which is a function of the trajectory of the boat, is sufficient to detect the presence of a boat in the marina and its direction with an accuracy of 0.96 with a t-LeNet model. The scenario where several boats are crossing or following each other can be problematic for the precise quantification of traffic in the marina because the trajectory signal tracks only one boat at a time. This problem is addressed and an experiment conducted with simulated time series with multiple boat trajectories showed that the false negative detection rate could be drastically reduced with enough and accurate data. Prospects for improving the simulations are presented.

The next step in the development of the acoustic system is the implementation of the trained CNN on the already developed software. The result of the classification performed by the CNN will be used as a trigger to automatically save the hydrophone signals when a boat is detected. The goal is to acquire enough data to perform ship signature classification and to detect intruder inside a marina.

ACKNOWLEDGMENT

The authors would like to thank Sodemo which supported this work and allowed the installation of the acoustic system at the entrance of Port Brunelet.

FUNDING AGENCY

Publisher has waived the Open Access publishing fee.

REFERENCES

Chung, K. W., Sutin, A., Sedunov, A., & Bruno, M. (2011). DEMON acoustic ship signature measurements in an urban harbor. *Advances in Acoustics and Vibration*, *2011*(i), 1–13. doi:10.1155/2011/952798

Codarin, A., Wysocki, L. E., Ladich, F., & Picciulin, M. (2009). Effects of ambient and boat noise on hearing and communication in three fish species living in a marine protected area (Miramare, Italy). *Marine Pollution Bulletin*, *58*(12), 1880–1887. doi:10.1016/j.marpolbul.2009.07.011 PMID:19666180

Cui, Z., Chen, W., & Chen, Y. (2016). *Multi-Scale Convolutional Neural Networks for Time Series Classification.* CoRR, abs/1603.0.

Das, A., Kumar, A., & Bahl, R. (2013). Marine vessel classification based on passive sonar data: The cepstrum-based approach. *IET Radar, Sonar & Navigation*, *7*(1), 87–93. doi:10.1049/iet-rsn.2011.0142

Deng, H., Runger, G. C., Tuv, E., & Martyanov, V. (2013). A time series forest for classification and feature extraction. *Inf. Sci.*, *239*, 142–153. doi:10.1016/j.ins.2013.02.030

Fillinger, L., De Theije, P., Zampolli, M., Sutin, A., Salloum, H., Sedunov, N., & Sedunov, A. (2010). Towards a passive acoustic underwater system for protecting harbours against intruders. *2010 International Waterside Security Conference, WSS 2010*. doi:10.1109/WSSC.2010.5730243

Fillinger, L., Sutin, A., & Sedunov, A. (2009). Cross-correlation of ship noise for water traffic monitoring. *The Journal of the Acoustical Society of America*, *126*(4), 2251. doi:10.1121/1.3249259

Fillinger, L., Sutin, A., & Sedunov, A. (2011). Acoustic ship signature measurements by cross-correlation method. *The Journal of the Acoustical Society of America*, *129*(2), 774–778. doi:10.1121/1.3365315 PMID:21361436

Guo, W., Piao, S., Guo, J., Lei, Y., & Iqbal, K. (2020). Passive detection of ship-radiated acoustic signal using coherent integration of cross-power spectrum with doppler and time delay compensations. Sensors, 20(6).

Holles, S. H., Simpson, S. D., Radford, A. N., Berten, L., & Lecchini, D. (2013). Boat noise disrupts orientation behaviour n a coral reef fish. *Marine Ecology Progress Series*, *485*, 295–300. doi:10.3354/meps10346

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery*, *33*(4), 917–963. doi:10.1007/s10618-019-00619-1

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, *25*, 1097–1105.

Le Guennec, A., Malinowski, S., & Tavenard, R. (2016). Data Augmentation for Time Series Classification using Convolutional Neural Networks. *ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data*.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to digit recognition. *Neural Computation*, *1*(4), 541–551. doi:10.1162/neco.1989.1.4.541

Li, G., Song, Z., & Fu, Q. (2019). Small Boat Detection via Time-Frequency Analysis and DenseNet. *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, 410-414.

Palaz, D., Magimai, M., & Collobert, R. (2015). Analysis of CNN-based speech recognition system using raw speech as input. *Interspeech*, 3–7.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Louppe, G., Prettenhofer, P., Weiss, R., Weiss, R. J., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Reis, C. D., Padovese, L. R., & Oliveira, M. C. (2019). Automatic detection of vessel signatures in audio recordings with spectral amplitude variation signature. *Methods in Ecology and Evolution*, *10*(9), 1501–1516. doi:10.1111/2041-210X.13245

Santos-Domínguez, D., Torres-Guijarro, S., Cardenal-López, A., & Pena-Gimenez, A. (2016). ShipsEar: An underwater vessel noise database. *Applied Acoustics*, *113*, 64–69. doi:10.1016/j.apacoust.2016.06.008

Schäfer, P. (2015). Scalable time series classification. *Data Mining and Knowledge Discovery*, *30*(5), 1273–1298. doi:10.1007/s10618-015-0441-y

Yamaguchi, H., & Muto, K. (2019). *Detection of Boat Noise by a Convolutional Neural Network for a Boat Information System*. Academic Press.

Zhao, B., Lu, H., Chen, S., Liu, J., & Wu, D. (2017). Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, *28*(1), 162–169. doi:10.21629/JSEE.2017.01.18

Zwemer, M. H., Wijnhoven, R. G. J., & De With, P. H. N. (2018). Ship detection in harbour surveillance based on large-Scale data and CNNs. *VISIGRAPP 2018 - Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 5*, 153–160.

Romane Scherrer received the Diplôme d'ingénieur from Les Arts et Métiers, France and the master's degree in mechanical engineering from the University of Auckland, New Zealand. She is pursuing the Ph.D. degree with ISEA, University of New-Caledonia. Her current research interests include embedded system, computer vision and holography.

Erwan Aulnette, after a Master 2 in Computer Engineering (University of South Brittany in France and Aalborg Universitet in Denmark), acquired a rich and varied experience in the field of connected objects. Erwan first acquired 9 years of development and operating experience in the field of agtech (the world's first connected cow). He worked in the design of tools for large-scale production and then specialized in server architecture and development. He then specialized in the design of IT architecture for the management of parks of connected objects, mastering the entire chain of measurement acquisition until its visualization. Invested in entrepreneurship, he is also co-manager of the startup L2K Innovation since 2020 which provides innovative solutions to its clients in various fields (Smart Marina, IoT, etc.).

Thomas Quiniou, after graduating from the University of Rennes 1, received his PhD in electronics and telecommunications from the National Institute of Applied Sciences (INSA) of Rennes in 2001. He then worked during three years in the telecommunication industry for Alcatel and Thomson. In 2005, he joined the University of New-Caledonia, ISEA, as a research engineer in electronics where he has been involved in many research projects in environmental science.

Joël Kasarherou has a PhD in Fluid Mechanics at Paris VI, France, and a postdoc at the University of New South Wales (Sydney, Australia). He holds 3 patents and he's a co-founder of a start-up company, L2K innovation, based in New-Caledonia, focused on marine activities. He is a member of the Board of Directors of several Caledonian companies.

Pierre Kolb is the chairman of A2EP, the New Caledonia leading engineering and consulting firm since 2002. He has gained previously a 12-year experience in the non-ferrous metallurgical industry as managing director of several Eramet Manganese business units in Europe (1998-2001). He also served as plant manager in the Doniambo SLN nickel refinery in New Caledonia (1990-1998). He has managed geophysical exploration research projects in the oil and gas industry (Elf Aquitaine) for 5 years (1984-1989). Having integrated innovation as a strategic lever for development and export, Pierre is involved In L2K innovation, a new Caledonian start-up company designing IoT systems for marine and coastal monitoring. Pierre is actively involved in the main socio-professional organizations in New Caledonia. He is presently an elected member of the Chamber of Commerce and Industry (Board member in charge of innovation and education). Pierre is a board & founding member of the Maritime Cluster of New Caledonia.

Nazha Selmaoui-Folcher is Associate Professor (HDR) at the University of New Caledonia since 1998. She received her Ph.D. degree in 1992 from the "Institute National des Sciences Appliques" at Lyon (France) and her Habilitation degree in 2012 from the University of Lyon I. Her current research interests are image analysis, machine learning, artificial intelligence, graph mining, data mining on temporal series of satellite images, spatio-temporal data and its application to health and environmental sciences. She is involved in the program committees of many data mining and machine learnong conferences and Journals. She was the leader of a multidisciplinary laboratory on matter and environment 2012-2016 (PPME). She was the coordinator for the ANR FOSTER projects dedicated to Knowledge Discovery on spatio-temporal Databases and Application to Soil Erosion (2011-2014). She participates to ANR project SpiRAL (Soils, rainfall and leptospirosis Understanding leptospirosis environmental contamination). She is member of LABEX-CORAIL (laboratory of excellence, Funded by French Agency) and associate member of IPAL (International Research Lab in Singapore).