CHALMERS
UNIVERSITY OF TECHNOLOGY

Deep learning for deep waters: An expert-in-the-loop machine learning framework for marine sciences

Downloaded from: https://research.chalmers.se, 2024-04-26 05:53 UTC

Citation for the original published paper (version of record):
Ryazanov, I., Nylund, A., Basu, D. et al (2021). Deep learning for deep waters: An expert-in-the-loop machine learning framework for marine
sciences. Journal of Marine Science and Engineering, 9(2): 1-18.
http://dx.doi.org/10.3390/jmse9020169

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology.
It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004.
research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Journal of
Marine Science and Engineering

MDPI

Article

Deep Learning for Deep Waters: An Expert-in-the-Loop Machine Learning Framework for Marine Sciences

Igor Ryazanov [1,†], Amanda T. Nylund [2,†], Debabrota Basu [1,‡], Ida-Maja Hassellöv [2,*] and Alexander Schliep [1,*]

[1] Department of Computer Science and Engineering, University of Gothenburg | Chalmers University of Technology, SE-412 96 Gothenburg, Sweden; gusryaig@student.gu.se (I.R.); basud@chalmers.se (D.B.)
[2] Department of Mechanics and Maritime Science, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden; amanda.nylund@chalmers.se
* Correspondence: ida-maja@chalmers.se (I.-M.H.); alexander@schlieplab.org (A.S.)
† These authors contributed equally to this work.
‡ Current address: School, Inria, Université de Lille, CNRS, UMR 9189-CRIStAL, F-59000 Lille, France.

Abstract: Driven by the unprecedented availability of data, machine learning has become a pervasive and transformative technology across industry and science. Its importance to marine science has been codified as one goal of the UN Ocean Decade. While increasing amounts of, for example, acoustic marine data are collected for research and monitoring purposes, and machine learning methods can achieve automatic processing and analysis of acoustic data, they require large training datasets annotated or labelled by experts. Consequently, addressing the relative scarcity of labelled data is, besides increasing data analysis and processing capacities, one of the main thrust areas. One approach to address label scarcity is the expert-in-the-loop approach which allows analysis of limited and unbalanced data efficiently. Its advantages are demonstrated with our novel deep learning-based expert-in-the-loop framework for automatic detection of turbulent wake signatures in echo sounder data. Using machine learning algorithms, such as the one presented in this study, greatly increases the capacity to analyse large amounts of acoustic data. It would be a first step in realising the full potential of the increasing amount of acoustic data in marine sciences.

Keywords: machine learning; marine sciences; deep learning; expert-in-the-loop; turbulent ship wake; environmental impact of shipping

check for updates

Citation: Ryazanov, I.; Nylund, A.T.; Basu, D.; Hassellöv, I.-M.; Schliep, A. Deep Learning for Deep Waters: An Expert-in-the-Loop Machine Learning Framework for Marine Sciences. *J. Mar. Sci. Eng.* **2021**, *9*, 169. https://doi.org/10.3390/jmse9020169

Academic Editor: Kristen Splinter
Received: 4 December 2020
Accepted: 2 February 2021
Published: 7 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The United Nations (UN) has declared the next ten years, 2021–2030, as the UN Decade of Ocean Science for Sustainable Development. The primary reason is that healthy and productive oceans, which relies on and requires sustainable management, are essential for humanity. One of the core themes in the UN Ocean Decade is data: in particular the transformative possibilities resulting from the generation of big data within marine science and monitoring [1–4]. This is also reflected in the coordinated work within the European Union (EU) Copernicus and EMODnet platforms [5,6].

This growing volume of data, often referred to as "big data", invokes new issues and calls for new approaches to manage and analyse data [2]. Machine learning methods are often used to analyse "big data" and to detect statistically significant patterns in them [1,2]. However, state-of-the-art machine learning methods, such as deep learning, typically require a large number of labelled examples to train on. Therefore, one of the principal bottlenecks in using machine learning to facilitate analysis of marine data, is the current lack of annotated data [1,2,7].

One example of "big data" in marine science, is acoustic data obtained from different types of sonars/echo sounders: i.e., Multi-Beam Echo Sounders, Single Beam Echo

Sounders, Side Scan Sonars, and Acoustic Doppler Current Profilers (ADCP) [8–11]. Acoustic data has eclectic applications in marine science [10], and the Marine Strategy Framework Directive (MSFD) stresses the importance of using acoustic methods to improve environmental monitoring in the future [12]. For example, acoustic methods can be used: (a) to map geological properties of the sea bed and benthic habitats (e.g., [11,13,14], and refs therein), (b) to monitor fish schools and individual fishes [8,9,15], zooplankton [16,17], and marine mammals [18,19], and (c) to study physical properties of the water column [20–22]. Here, we present a novel deep learning-based framework for automatic detection of wake signatures in echo sounder data (echograms). The advantages of our deep learning framework are demonstrated by a case study using acoustic data of turbulent ship wakes and their introduction of energy by vertical mixing, as outlined in the following sections.

Our goal was to automatically detect turbulent ship wakes in echograms, with the aim of assessing the spatiotemporal extent of the wakes, as well as the energy input from ships through vertical mixing. Of the eleven MSFD descriptors used to assess the environmental status of the marine environment (MSFD, 2008/56/EC) [23], the eleventh descriptor concerns the introduction of energy, including underwater noise, into the marine environment. Environmental impact from shipping is often mentioned in relation to descriptor 11, as ships cause both noise [12] and erosion [24,25]. However, there is another energy input from shipping which is rarely mentioned, namely the energy input through ship-induced vertical mixing.

Behind all ships, a turbulent wake is induced by the ship propeller and hull [26–29]. The wake is characterised by an increase in turbulence and a dense bubble cloud. The bubbles in the wake region can be detected using an echo sounder [30,31]. In regions with intense ship traffic the turbulence and bubbles from repeated ship passages have the potential to affect air-sea gas-exchange [30–32], sea floor integrity and turbidity [33], and the dispersion and distribution of pollutants and contaminants from ship discharges [34,35]. Hypothetically, water column stratification and nutrient supply could also be affected, if the ship-induced vertical mixing is deep and frequent enough. There is currently a lack of knowledge regarding the parameters determining the temporal and spatial development of the turbulent wake, which is needed to estimate the environmental impact from ship-induced vertical mixing. Hence, there is a need for suitable and efficient methods to measure the temporal and spatial scales of the turbulent wake, as well as the intensity of the vertical mixing. Echo sounder data from bottom mounted ADCPs or multibeams, provides an excellent data source for ship wake characterisation, as the bubbles in the turbulent wake create an elevated echo amplitude in the turbulent wake region in the echogram [26,29–31,36,37]. However, manually identifying and annotating ship wakes in echograms requires expert knowledge and considerable time. The efficiency of the analytical process can be greatly improved by using machine learning to automatically identify and quantify ship wakes. This provides the core motivation for this study.

Our contributions: We propose adopting the state-of-the-art deep neural networks, which are widely used for image classification also for applications in marine science [38], to detect wakes using the echo sounder data from the bottom mounted ADCP. The principal bottleneck is that these deep neural networks, namely Convolutional Neural Networks (CNN), require significant amount of labelled data to train and thus, to achieve the desired detection accuracy. To the best of the authors knowledge, there currently exists only one publicly available annotated dataset of ship wakes in acoustic data, namely the one included in this study. Hence, in order to enable training of a deep neural network in absence of a sufficiently large labelled dataset, data augmentation is necessary.

Data augmentation was performed by generating synthetic wake samples based on the existing ones. A deep learning model was then trained on the populated dataset. Several important steps, including the results of synthetic data generation, were supervised by a marine science expert, using the expert-in-the-loop approach [39]. In particular, for uses of machine learning in health-care, the efficacy of expert-in-the-loop or human-in-the-loop has been demonstrated for a range of tasks, such as interpreting and assessing the uncertainty

of machine learning predictions, as an oracle in active learning settings, or, similar to here, to assess goodness of fit of auxiliary models and confirm labelling suggestions from the machine learning method [40,41].

We experimentally show that the proposed approach of using deep neural networks combined with the expert-in-the-loop framework achieves significant accuracy for detecting ship wakes from the echos sounder data with limited affirmative instances.

Though we use the proposed framework for a specific case study and data type, it should be noted that this framework provides a blueprint to address related problems, i.e., as long as the collected data is similar in nature, e.g., acoustic data, we expect that both the proposed deep learning model for identifying different signals in the data and the expert-in-the-loop framework to address label scarcity, will prove beneficial.

2. Background: Bridging Machine Learning and Marine Science

Its successes over the past decade in tasks such as object and face detection, natural language processing, or text mining, have established deep learning as a suitable paradigm of machine learning for signal, image, and video analysis [42]. Therefore, we chose to use deep learning to analyse acoustic data for the automatic detection of ship wakes in echo sounder data. This setting provided a suitable scenario to evaluate the potential benefits of using deep learning to assist the analysis of marine acoustic data. The following section provides a short background on the use of machine learning techniques for acoustic signal processing, with a particular focus on the applications in marine science.

2.1. Deep Learning for Signal and Image Analysis

Machine learning techniques are efficient in processing image and signal data and detecting patterns in them [43]. These techniques adapt numerical weights and other parameters, in general according to features of a set of known training data. Such methods can often be smoothly adapted to a domain-specific task, provided there is a sufficient amount of data. When training data is scarce, the machine learning models tend to lack generalisation and instead *overfit* to the known examples used for training. As scarce data can impede machine learning techniques, several approaches, including data generation and augmentation, are proposed to improve the performance. For example, data generation has previously been used in a study by Allken et al. [7], where synthetic samples were leveraged to expand the training dataset for a deep learning algorithm identifying fish species from video data.

2.2. Machine Learning for Acoustic Detection in Marine Science

In recent decades, machine learning has successfully been applied to analyse acoustic data in order to automatically classify the geological properties of the seafloor [44–46], constituting an exemplary application of machine learning in marine sciences. In contrast, machine learning methodologies are still at the developing stage for identifying biota and physical properties of the water column in acoustic data. However, the potential benefit of, and necessity for incorporating machine learning in the analysis of the increasing amount of acoustic data, have been pointed out [2,14,47]. Currently, there exists methodologies for automatic target tracking and identification of e.g., fish schools, marine mammals, and birds, in multibeam and echo sounder data [8,48–51]. Still, the analysis of echo sounder data often relies on manual scrutiny by experts at some step of the process [52], which is time consuming and introduces subjectivity to the analysis [49,51].

Twenty years ago, during the previous boom of neural network popularity, a few studies used Artificial Neural Networks (ANN) to automatically identify fish schools in echo sounder data [53–56]. Other machine learning methods, such as random forests have also been applied [52]. However, all of these models were small and applied to the pre-defined features extracted from the already detected fish schools, hence do not strictly classify as "deep learning". Nevertheless, there is one later example of the application of "deep learning" in acoustic classification. Brautaset et al. [57] recently applied deep

learning techniques to a problem comparable to the identification of ship wakes—acoustic classification of fish schools. Similarly, they used echo sounder data, although collected from surface vessels and at greater depths, compared to the ship wake dataset. The authors applied a CNN model to their problem and achieve good results for isolated regions in the echogram.

3. Materials and Methods

Here the proposed framework is presented, which is essentially expert-in-the-loop with a deep neural network for classification, and the methodology to deploy this framework for wake detection from acoustic data is specified.

The development of the proposed deep learning model for wake detection can be structured as follows:

1. In situ collection of acoustic data (Section 3.4);
2. Initial data labelling by the domain expert (Section 3.5);
3. Data augmentation and its approval by the expert (Section 3.6);
4. Training baseline deep learning model (Section 3.7);
5. Evaluating baseline model's performance (Section 3.8);
6. Implementing algorithms to make the model robust against data imbalance and noise with an additional input from the expert (Sections 3.7 and 3.8);
7. Evaluating the final model (Section 4).

A schematic view of the proposed pipeline is given in Figure 1. Following subsections explain the expert-in-the-loop framework, which is used in model development, and also elaborate all the steps.

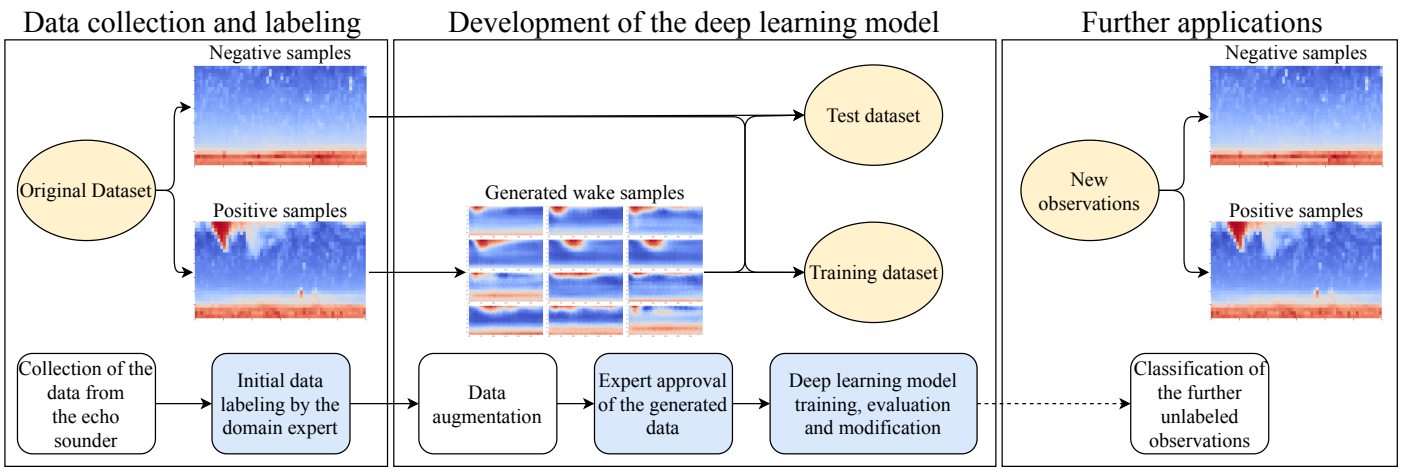

Figure 1. The framework pictured as a pipeline, showing all major steps from data collection to the final model evaluation. Potential further applications are also included in the diagram. Steps shaded in blue indicate the involvement of the domain expert.

3.1. Expert-in-the-Loop

Expert-in-the-loop is a framework for Artificial Intelligence (AI) training [39,41], where an expert oversees the training and gives additional feedback based on the results after a training iteration. It can be beneficial in terms of both achieving better results and speeding up the training. However, in the traditional machine learning approach, the participation of a human expert is usually limited to preparing the training task (e.g., labelling). This is because large datasets are typically very expensive to create and require the work of many annotators.

In the case of this paper, the dataset was relatively small (1 month of observation and 165 identified wakes), and the data was very domain-specific. The expert who performed the initial labelling was also involved in the development, so it was possible to introduce

intermediate expert evaluation in some form. The stages with the expert involvement are highlighted in Figure 1.

3.2. Machine Learning: Classification

Classification is a machine learning problem where the goal is to assign a new observation to one of two, or more, predefined categories. The prediction is based on the training dataset of observations for which the categories are known. Classification problems can be solved by a large variety of algorithms ranging from decision trees to neural networks, deep or shallow.

In the case of wake detection, the proposed model must solve a binary classification task, where the two classes will be labelled as 'wake' and 'background'. More formally, if $X = \{x_1, x_2, \ldots, x_n\}$ is a set of objects and $Y = \{y_1, y_2, \ldots, y_n\}$ is a set of classes, and f is a classifier, which maps X into $Y : f(x_i) = y_j$. For wake recognition, X is the set of frames and $Y = \{\text{'}wake\text{'}, \text{'}background\text{'}\}$.

The formulation of the principal task in terms of classification allows the use of a range of machine learning classification algorithms. The nature of the available data, however, suggests using models based on Convolutional Neural Networks (CNNs) [58]. CNNs are known to be effective in image recognition and finding patterns. By treating the acoustic data in the form of time frames, the task can be mapped directly to pattern recognition, where the target pattern is the trace of bubbles in the turbulent wake behind the ship.

3.2.1. Convolutional Neural Network

Neural networks are a class of machine learning models that consist of multiple interconnected nodes (or neurons). Nodes perform simple processing of incoming information and pass it further down the network. Nodes in neural networks are commonly organised in layers, which can be connected in various ways. Neural Networks are trained using backpropagation, which calculates the gradient of the loss function with respect to the weights of the model for each input-output and updates the weights accordingly.

The most common type of neural network used in image analysis is a CNN. The main feature of a CNN is convolutional layers. Each neuron in such a layer processes only the information from the receptive field corresponding to this neuron. It can be perceived as a sliding window moving over a 2-dimensional image and checking if certain patterns of the size of the window are present. Typical CNN architecture involves stacking several convolutional layers with other types of layers, such as pooling (aggregating information from neighbouring neurons) and fully connected layers. One example of a well-known and successful deep neural network for image recognition is VGG (VGGNet by Visual Geometry Group) [59]. CNNs were also successfully used for analysis of photographic images in marine science, e.g., monitoring waves [38].

3.2.2. Residual Neural Network

A more recent CNN architecture, which is often used as a benchmark, is the Residual Neural Network (ResNet) [60]. ResNet solves the vanishing gradient problem, which affects the training of very large neural networks with gradient-based methods. In large networks, the gradient can become vanishingly small, and the weights can stop changing at all during training. To address this, ResNet introduces skip-connections. In simpler architectures, layers are always connected sequentially, while in ResNet some layers can be skipped during the first stages of training. For instance, if there are three layers in the network: A, B, C, and they are connected in order A to B to C, then the direct connection from A allows skipping B during initial training (Figure 2). This architecture not only addresses the vanishing gradient problem, but also greatly improves training time. As variations of ResNet obtain close to state-of-the-art results on several image analysis task, we choose it for a different type of data, superficially resembling photographic image data, namely the data we collected from an ADCP.

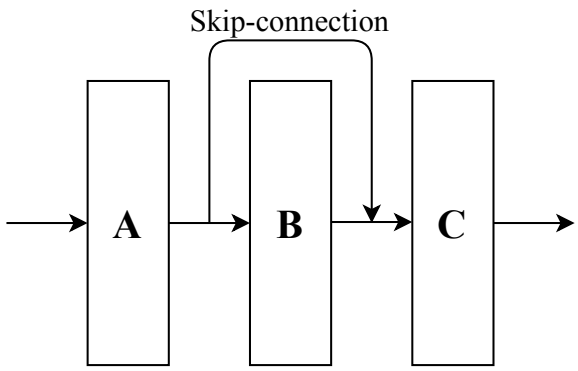

Figure 2. An example of a skipped layer in a Residual Neural Network. There is an additional connection between layers *A* and *C* that allows skipping layer *B* when propagating.

3.3. Approaches for Scarce Data

One of the main challenges encountered in this research was the lack of labelled data and, more specifically, the lack of labelled positive wake examples. Performing supervised machine learning algorithms typically require a considerable number of labelled samples. This is especially important for a neural network model, as it includes many parameters to adjust. There are several possible approaches to address the lack of data.

3.3.1. Data Augmentation

Data augmentation is a common strategy to use when the training dataset is too small and more samples are needed [61]. To augment a dataset, original samples are changed in a minor way, while keeping the designated output value. Even simple augmentation is known to improve the performance of machine learning models. For example, image datasets can be augmented with rotated, cropped, or mirrored samples (Note, we did not utillize these standard augmentation steps; see below and Section 3.6). Data augmentation is also a step that can be overseen by a domain expert. If generated or modified samples have to be approved as plausible by an expert, it reduces the chance of introducing errors during the augmentation.

However, this type of augmentation might not be applicable to stationary profiler data, primarily because wakes have a fixed position relative at the top of the frame. There are, nevertheless, more complex approaches, which can also serve as a means of regularisation and increasing robustness.

3.3.2. Probabilistic Models

New samples can also be generated using statistical methods. They preserve the patterns in the original samples while adding diversity to the dataset.

A Gaussian Mixture Model (GMM) [62] is a simple probabilistic model that fits the data to a convex combination of several uni- or multi-variate Gaussian distributions. It can be an efficient sample generating or clustering tool when the data structure is not very complex. The optimal number of components can be determined using relative model quality estimators, such as Akaike Information Criterion (AIC) [63]. AIC penalises a large number of parameters to fit and rewards the goodness of fit using the likelihood function, thus, finding a balanced number of distributions.

GMMs are not very efficient when applied to high-dimensional data such as images, so in order to apply them to echo sounder data, some dimensionality reduction is needed. Compression should also be reversible, so the generated samples can be restored.

Principal Component Analysis (PCA) [64] is a method that transforms the data into a new coordinate system. The greatest sample variance in the data lies on the first component, the second-largest lies on the second component, and so on. The ratio of explained variance can regulate the number of dimensions after compression.

While PCA is efficient, it does not specifically try to preserve the existing patterns in the data. This can be addressed by applying a more meaningful transformation first. CNNs perform well in preserving image patterns, and they can also be used for efficient compression. An autoencoder is a special type of unsupervised neural network that consists of two components: an encoder, which compresses the input and decoder that reverts the process (Figure 3). The training process of an autoencoder consists of attempting to reconstruct a set of samples. A trained CNN-based autoencoder can create a compressed representation of image-like data while preserving patterns such as the shape of the bubble trace of a ship's wake.

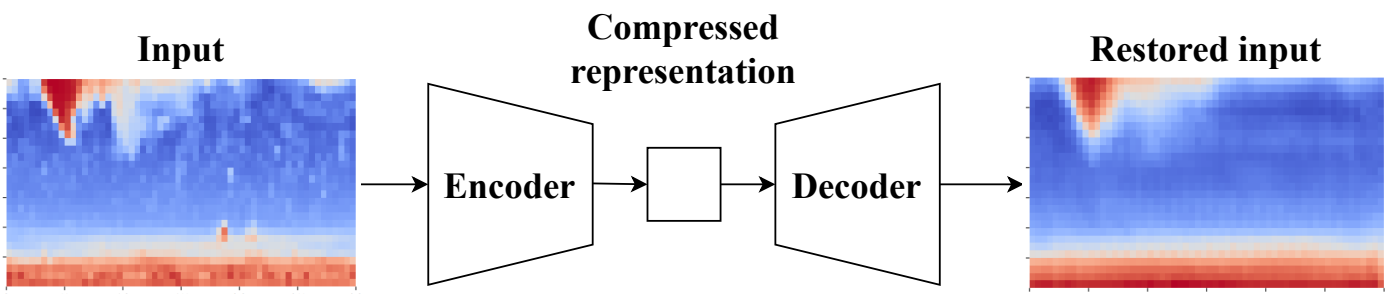

Figure 3. The architecture of an autoencoder.

3.4. In-Situ Data Collection

The data collection was conducted in the large ship lane outside Gothenburg harbour, which is the largest harbour in Scandinavia [65]. The 4 week measurement period included 165 clearly visible ship wakes and varying weather conditions.

3.4.1. ADCP Measurements

The ship wake dataset was collected using a bottom-mounted Nortek Signature 500 kHz broadband Acoustic Doppler Current Profiler (ADCP). The instrument was deployed under the ship lane during 4 weeks (28 August to 25 September 2018), at approximately 30 m depth (57.61178 N, 11.66102 E). The ADCP had four slanted beams (25° angle) and one vertical beam, all with a measured cell size of 1 m and ping frequency of 1 Hz. The measured echo amplitude was used to identify the ship wake region, as the bubble cloud in the ship wake reflects sound more efficiently than water, and is clearly visible as an elevation in the signal strength [26,29–31,36,37].

3.4.2. AIS Data

A dataset of the ships passing the study area during the measurement period, was used to identify periods without ship passages, to use as negative controls when training the algorithm. The dataset was purchased form he Swedish Maritime Administration, and originates from the Baltic Marine Environment Protection Commission (HELCOM) Automatic Information System (AIS) database. The data was processed according to the procedure described in the annex of the HELCOM Assessment on maritime activities in the Baltic Sea 2018 [66]. Additional files from the same HELCOM database was provided by the Swedish Institute for the Marine Environment (SIME).

3.5. Data Labelling and Preparation

The raw dataset from each ADCP beam was used in the analysis, resulting in 5 time series, with observations for 79,023 timestamps in total. For each timestamp, 28 data points were given, corresponding to depth levels from 3.5 m to 30.5 m with 1-meter interval. Due to side lobe interference, the 2.5 m closest to the surface did not have reliable data, hence the measurements start at 3.5 m depth. One of the slanted beams (beam 2) was malfunctioning, resulting in corrupted data, and was thus excluded from the analysis. The signal data was

normalised, so that all the values fell between 0 and 1. The nighttime data contained very high levels of noise, and several nighttime wakes were marked as ambiguous by the expert. Thus, for training and testing purposes, only daytime negative samples were used.

3.5.1. Data Labelling

High resolution figures of the echo amplitude of the vertical beam were used to manually identify the ship wakes in the ADCP dataset. Wake signatures in the ADCP dataset were then compared with the AIS data of passing ships, to confirm that the wake signatures could be connected to vessel passages. Next, the echo amplitude in each confirmed wake region was compared to the daily/nightly mean, and all measurements in the wake frame ∼15% higher than the mean was annotated as part of the wake.

3.5.2. Data Representation and Visualisation

Acoustic data can be visualised in the form of an echogram, as shown in Figure 4. The echogram displays the intensity of the reflected signal; wakes and other objects are visually recognisable in this representation. A total of 165 wakes were identified by an expert and marked on echograms. AIS ship tracking data was checked during initial labelling, to confirm that all the detected wakes could be related to a passing ship. This allowed the assumption that all data outside the marked time frames could be used as negative samples.

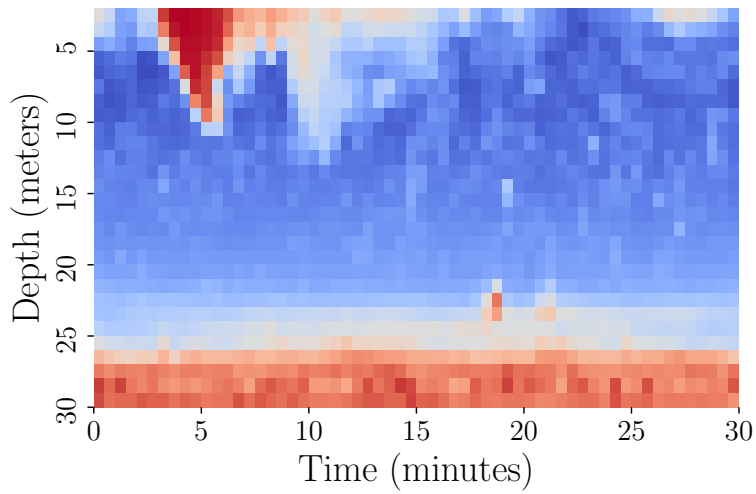

Figure 4. A 30-min frame of observations starting around 15:50 on the 9th of September 2018. It displays a clearly visible wake in the top-left corner.

To perform binary classification, the echogram was split into fixed-size frames. The length of the time frame was 60 data points or 30 min, and the size of one such sample was $4 \times 28 \times 60$ data points. The first dimension corresponds to the number of functional beams. One example of such a frame is displayed in Figure 4. The wakes in the dataset ranged from large to barely noticeable, as can be seen in Figure 5.

All of the examples in this paper use the data from the central vertical beam (beam 5) because it consistently showed stronger signals. The difference was confirmed by comparing Wasserstein distances between the data from different beams and an "empty" frame with a flat distribution of signals. Wasserstein distance computes the cost of transforming one distribution into another, where the cost is the size of the part that has to be changed. The Wasserstein distance from the flat distribution was consistently larger for the central beam than for the other three.

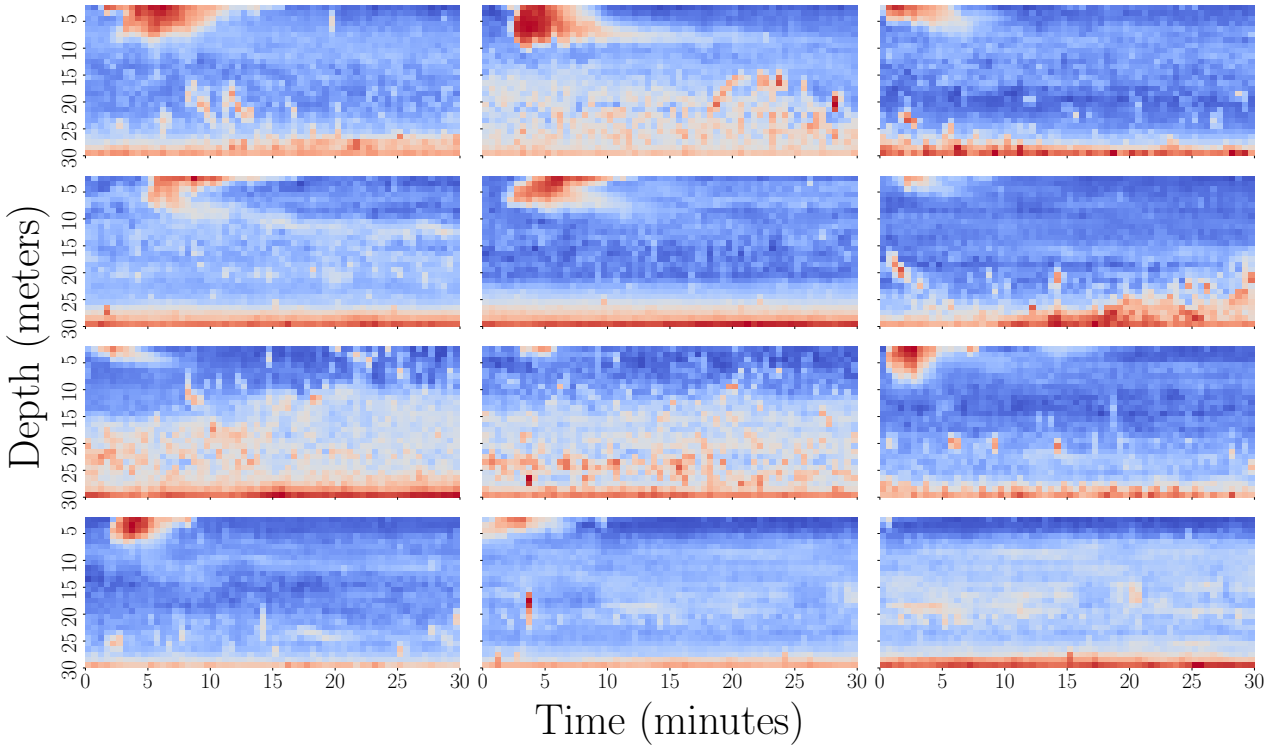

Figure 5. Examples of wakes labeled by the expert.

3.5.3. Set-Aside Dataset

Before proceeding with any computational experiments, a subset covering four days (22 September to 25 September 2018) was set aside for separate evaluation. The period included 23 wakes of varying sizes and notably worse weather conditions, according to the expert. This subset was selected as a more difficult and completely independent additional evaluation task for the model. The remaining dataset containing 142 labelled wakes was used in the development of the predictive model.

3.6. Data Augmentation

To successfully perform deep learning training, a sufficient number of positively and negatively labelled data samples is needed. The ship wake dataset was heavily unbalanced and noisy, so data augmentation was necessary to generate more wake samples and allow the use of neural networks.

Data augmentation could only be based on the labelled 142 wake frames. This number is clearly not large enough to apply deep learning models, such as generative adversarial networks (GANs) [67], so a more robust approach of fitting a GMM was chosen. Note, we decided against the often used simple geometric transformations such as stretching, flipping, rotating, cutting for augmentation, as the positioning of wakes is fixed and can be meaningful.

3.6.1. Data Compression

Since one frame containing a wake has relatively high dimensionality, a reduction had to be applied before fitting a GMM. To make the dimensionality reduction more meaningful and preserve the patterns of the wakes, a CNN-based approach was taken.

The first step of data compression was, therefore, training an autoencoder on all 142 positively labelled frames with wakes. A simple autoencoder model with three convolutional layers in both encoder and decoder was trained on these examples for 70 epochs. The autoencoder then was able to compress and restore frames while keeping the major patterns and cancelling most of the noise. In Figure 6 the frame from Figure 4 is shown after

passing the autoencoder. Most of the noise was removed, but the wake is preserved and clearly visible. After the training, the encoder was used to compress the same 142 samples.

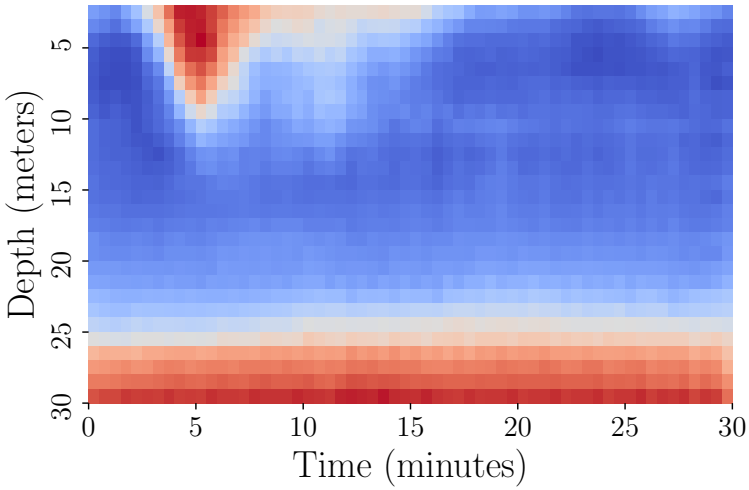

Figure 6. The frame displayed in Figure 4 after being passed through the trained autoencoder.

After the frames were passed through the network, the dimensionality remained relatively high, and as the second compression step, PCA was applied. The model was set to keep 99.9% of the variance, and it reduced the sizes to only 48 dimensions each.

3.6.2. Sample Generation

A GMM model was subsequently fitted to the set of 142 compressed examples. The number of Gaussian components was set to 21 after computing Akaike Information Criterion for fitted GMMs with 1 to 60 components.

The GMM model allows the generation of new samples according to the learned distributions. For future experiments, a total of 1000 48-dimensional samples were generated. The samples were passed through the inverse PCA model and then through the decoder, restoring them to the $4 \times 28 \times 60$ size.

Some of the generated wakes are shown in Figure 7. Synthetic data has noticeably less noise and is much smoother than the original samples (Figure 5), but all the defining features have been successfully preserved. Following the pipeline (Figure 1), the generated set was approved by the expert as containing plausible shapes that could only be identified as wakes in the original dataset. The expert agreed that while being smoother than the wakes in the original observations, most of the generated data qualifies as a set of positive examples. As an additional step, to confirm the quality of augmentation, pairwise Wasserstein distances were computed between the original wakes and a set of generated wakes. The average value of 0.0526 was smaller than the average pairwise distance between the original wake frames and a random subset of frames (0.0544).

3.7. Deep Learning Models

Two deep learning models were implemented and tested. The base model used ResNet architecture. Since the structure of data was not very complex, and the resolution was small, a relatively small ResNet18 model was implemented. ResNet18 consists of 18 layers in total: 1 convolutional layer followed by 8 residual blocks with 2 convolutional layers each, and a fully connected output layer with sigmoid activation. The model treats the data from the four beams as four channels of the same image. A single input, therefore, has the shape of $4 \times 28 \times 60$. The output consists of two values, which can be interpreted as probabilities of the input belonging to *wake* and *background* classes.

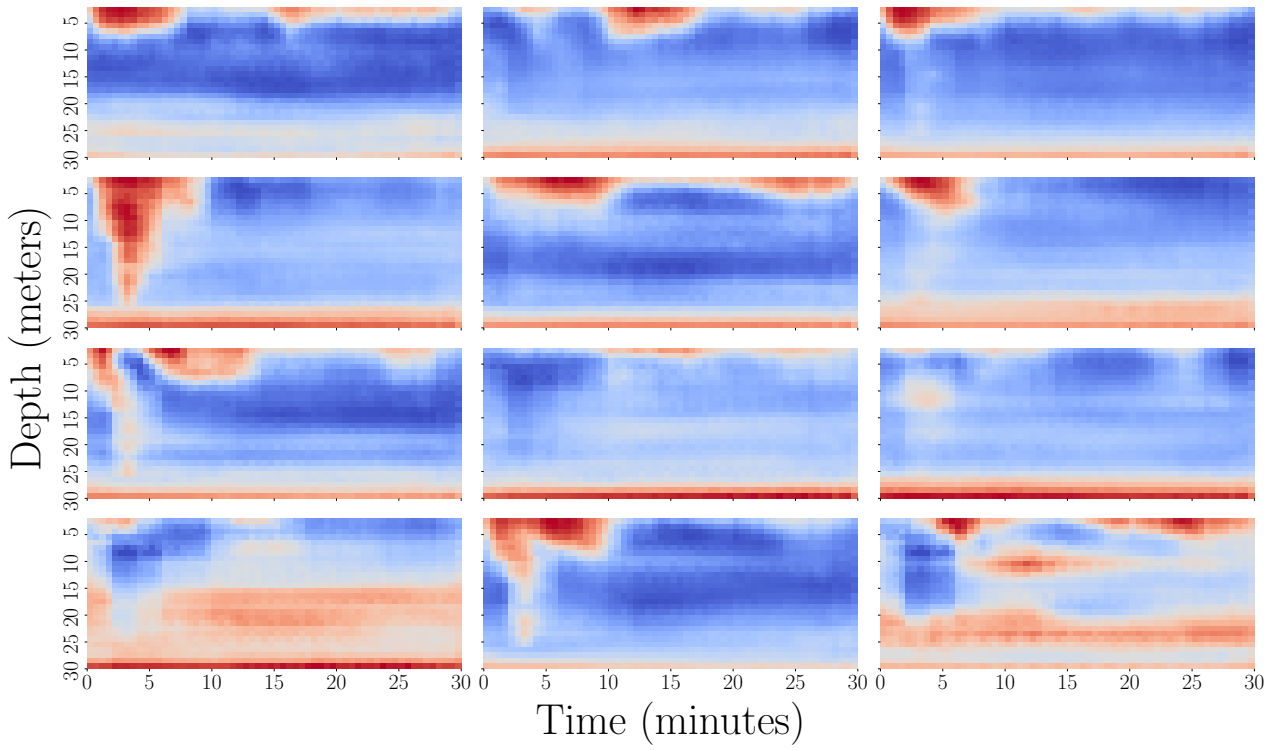

Figure 7. Wake frames generated using GMM and autoencoder.

For the second model, exactly the same ResNet18 architecture was re-implemented using the example reweighting technique. It is a strategy for addressing unbalanced data by identifying more or less valuable samples within the training dataset and assigning corresponding weights to them. Typically these weights are initialised offline, once during the training process, but there is an alternative approach proposed by Ren et al. [68]. It involves using a small subset of perfectly labelled clean data as a hyper-validation set. At each step of training, the gradient of the hyper-validation loss with respect to the weights of samples in the current mini-batch is computed and the weights are updated accordingly. The reweighting model can be implemented for most deep learning architectures and showed good performance both for unbalanced datasets and datasets with noisy labels. Perhaps the most important feature that makes this method applicable to the wake detection problem is that the size of the hyper-validation set can be as small as 10 samples while keeping a high level of performance. It also aligns well with the expert-in-the-loop framework (Figure 1). Since only a small number of clean examples are needed, they can be hand-picked or approved by the expert.

For the experiment, the hyper-validation dataset was set to the size of 10:5 positive and 5 negative examples. The negative samples were hand-picked from the dataset by the expert as clean with confidence. The purpose of this model was to test its performance compared to the baseline when the training data is unbalanced.

3.8. Evaluation Metrics

The evaluation of the models was based on the standard metrics for binary classification. Accuracy (the ratio of correct guesses) indicates the performance on balanced datasets, and the False Negative Rate (FNR) the ratio of wakes missed by the algorithm. Since the purpose of the model was to assist in the detection of wakes in the datasets, the FNR was the most relevant metric. The Area Under Receiver Operating Characteristic Curve (AUC ROC) is a metric that measures how well the model distinguishes between classes, as the threshold to make a decision varies.

To achieve stability in results, each experiment was performed 10 times. For each metric, the average and the standard deviation were calculated. For all experiments, the test

dataset included 150 samples, 29 of which were randomly selected original wake frames, 46 were randomly selected generated wake frames, and 75 were randomly sub-sampled frames without wakes. Test-time augmentation [69] has recently received attention for its capability to obtain more reliable or informative evaluation metrics [70], and even better results when used to obtain classification as an ensemble average over augmented data [71]. We chose to include augmented data to check for generalization, a real concern given the very low number of positive samples.

An additional experiment was performed using the imbalanced set-aside dataset covering four days with 23 known wakes. It was passed to the model sequentially in the form of overlapping frames. The purpose of this test was to show the performance with a completely independent set of more noisy observations.

4. Results

The performance of the two main models was evaluated: the baseline ResNet18 model and the example reweighting model. In all experiments, the main metrics were the accuracy score and FNR, since the test dataset was always balanced. Additionally, for the baseline experiment, the AUC ROC score was computed. Each experiment was performed 10 times to achieve stability in the results, and average values were used. As an additional evaluation, the performance of the main model is shown for the set-aside subset of four days.

4.1. Baseline ResNet Model

To perform the baseline experiment, the ResNet18 model was trained on a balanced dataset which included: 500 negative samples, 113 positive samples from the original data, and 387 generated positive samples. The test dataset was also balanced and included 75 negative samples, 29 positive samples from the original data, and 46 generated positive samples. All samples were selected randomly from their respective sets. The experiment showed the mean average accuracy of $93.40 \pm 1.80\%$, AUC ROC value of 0.97 ± 0.01, and false negative rate of $9.87 \pm 3.28\%$ over 10 repetitions. The results can be interpreted in the following way: 93.4% of predictions were correct on average, but most of the wrong predictions came from false negatives. The model had a high AUC ROC score, meaning it performed well in ranking samples by the likelihood of containing wakes.

4.2. Example Reweighting Model

The second evaluated model was the example reweighting model. Its main purpose was to allow training on unbalanced data while keeping test performance from quickly degrading. The reweighting model had exactly the same architecture as the baseline model but took an additional step in training to adjust the example weights during training.

The experiment included training the models on five training datasets with different class balance: 50%, 20%, 10%, 5%, and 2.5% of positive samples. The total size of the training dataset was the same as in the previous experiment, namely 1000 samples. The test dataset was kept balanced with 75 positive and 75 negative samples. The results are shown in Figure 8. With the decrease of positive samples in the training data, the quality of predictions of both models predictably dropped. The example reweighting models showed noticeably better robustness to class imbalance.

4.3. Set-Aside Dataset Experiment

As a secondary evaluation of the baseline model, an experiment was performed on the set-aside dataset. The whole dataset was passed through the ResNet18 model in the form of overlapping time frames of the $4 \times 28 \times 60$ size starting at each timestamp. Unlike the previous experiments, the whole set-aside dataset was used, including extremely noisy nighttime data and data from bad weather conditions during the measurement period. The 23rd of September was specifically marked by the expert as the day with the worst weather over the month.

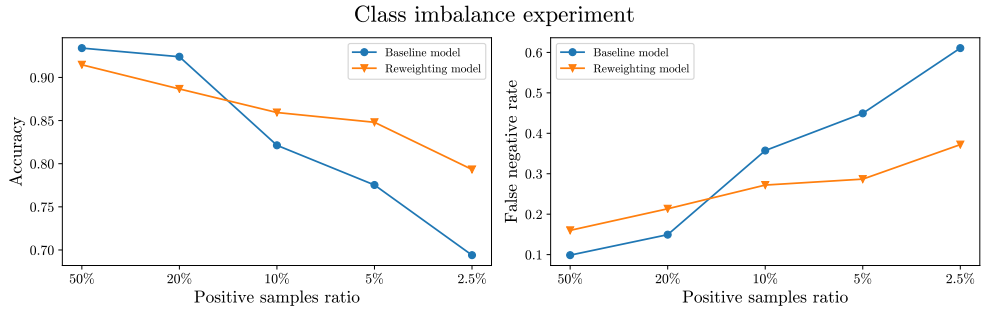

Figure 8. The class imbalance experiment with the ResNet18 model. Accuracy and false negative rate scores on the test dataset are visualised for the baseline and the reweighting model. The training sets had five different ratios of positive samples from 50% to 2.5%. The test set contained 75 positive (29 original + 46 generated) and 75 negative samples. The results are the mean average over 10 experiments.

There were 23 known wakes in this subset. For each wake, only the starting timestamp was known. It was assumed that the wake was visible in a 30-min frame if the frame starts at most 25 min before the wake start or at most 5 min later. The total number of frames in this experiment was 9780, out of which only 1338 were expected to be classified as wakes due to overlaps.

For this setup, the mean accuracy score over 10 experiments was as low as 60.6%, while the mean false negative rate was at 38.06%, which is comparable to test performance with a significant level of noise. Overall, the results show a significant drop in performance compared to experiments on cleaner and smaller test sets.

5. Discussion

The looming ocean crisis, for example the threat of near extinction of commercially and fundamentally ecosystem relevant species, has been in part caused by unsustainable use enabled by regulatory omissions. One contributing factor is the wide gap between the resolution and extent of information provided by current approaches to monitoring the ocean, including the effects of commercial activities and what would be needed to derive actionable, policy-supporting insights. In response, data has been designated a core theme of the UN Ocean Decade and a priority for other marine science platforms, such as EU Copernicus and EMODnet, acknowledging the opportunity for transformation made possible by machine learning. This growing demand for data-driven technologies to enhance marine sciences, require development of machine learning techniques that can simultaneously leverage the available data with its limitations, and the experts' knowledge from the scientific community. For acoustic data, the analysis and processing has traditionally been performed manually [2]. As a consequence, there is a huge potential for increased analytical capacity within the marine science field, if machine learning can be applied to acoustic data. Acoustic data is currently used in a wide array of applications within the field, benefiting from the increasing use of high-quality echo sounders [2]. The need to further increase the use of acoustic data within marine monitoring was pointed out in the recent European Marine Board (EMB) Future Science Brief №6: *Big Data in Marine Science* [1]. However, the increasing volumes of acoustic data have not been mirrored by a similar increase in analytical capacity, making data analysis a bottleneck [1,2]. Furthermore, there is still a scarcity of annotated data to train machine learning algorithms.

Here, we demonstrated the advantage of machine learning in marine science by detecting ship wakes from acoustic marine data in the form of echograms, using an expert-in-the-loop framework. Since the annotated data was limited in size and quality, we adopted machine learning algorithms to augment the data, by generating new data samples similar to the data available. The expert was then included to validate the quality of the generated data and send it forward in the framework to build a deep learning model for ship wake detection. Hence, the proposed collaboration of machine learning algorithms

and a domain expert can circumvent the limited quality and size of acoustic marine data. This framework also allows development of accurate and scalable deep learning models that can be used for detecting patterns of interest in marine sciences. When applied to the ship wake detection problem, it showed an average accuracy of ∼93.40% given a very limited and unbalanced dataset. The framework also serves as a demonstration of general applicability of the deep learning approach in marine sciences.

The experimental results demonstrate the potential of machine learning expert-in-the-loop approaches in analysing acoustic marine data, but also show that there is room for improvement. Since this is the first study of its kind, the performance cannot be compared to the result of previous studies. Increasing the algorithm's robustness to noisy and unbalanced data would be a direct path to improvement without requiring significantly more data. This could be achieved by modifying both the predictive algorithm itself or the data augmentation process. Furthermore, the proposed machine learning framework can be adopted to identify further objects of interest, which are potential sources of miss-classification in the current data set: schools of fish, vegetation, marine mammals, water mixing zones, currents, or stratification. With suitable data available, knowledge transfer [72] provides a machine learning approach to proceed.

The method presented here can be considered a stepping-stone solution for emerging fields plagued by the lack of labelled training data, which can be used and developed as the field is established. This will contribute to the effort of further integrate machine learning and marine science, as urged by Guidi et al. [1] and Malde et al. [2].

This demand for more, and more insightful data, can—similar to other disciplines where monitoring biodiversity is paramount [73]—be most effectively satisfied by the use of artificial intelligence or machine learning methods. Approaches differ depending on the exact phenomenon which needs to be recognised or quantified. To put things into perspective, it is important to recall that commercial machine learning methods e.g., for recognising objects in images, are often trained with billions of labelled examples. However, as this study demonstrates, the initial, labelled data a human expert provides, together with expert supervision of the often incremental model-improvement process, can rapidly lead to useful tools. This demonstration establishes that this particular type of application can be addressed with machine learning, which should be further explored and improved in future studies.

6. Conclusions

This study presents a novel deep learning-based expert-in-the-loop framework for automatic detection of turbulent ship wake signatures in echo sounder data. The proposed framework enables collaboration of experts and data-driven algorithms to create novel machine learning methods, which can be adapted to work on multiple types of signal data, may it be acoustic or visual. The suggested machine learning algorithms can fill the gap between the data available today and its use to rapidly address marine science problems, such as the effect commercial shipping has on marine ecosystem. In addition, it is a step towards further integrating machine learning and marine science, to propel a sustainable management of our oceans.

Author Contributions: A.S. and I.-M.H. conceived and supervised the study. All authors contributed to the design of the study. A.T.N. and I.-M.H. acquired the acoustic data, and A.T.N. performed the data annotation. I.R. and D.B. developed the method, the testing procedure and performed data analysis; the method was implemented by I.R. under supervision by D.B. The initial draft was written by I.R. and A.T.N. All authors discussed, read, edited and approved the article. All authors have read and agreed to the published version of the manuscript.

Funding: This research received seed project funding from the Area of Advance Transport, Chalmers University of Technology.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: I.-M.H. and A.S. acknowledge a seed grant for "Deep Learning for Deep Waters" from the Transport Area of Advance at Chalmers University of Technology. Acknowledgment to the Swedish Institute for the Marine Environment (SIME), for supplying the AIS dataset. D.B. acknowledges the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation for funding a part of his tenure during this work.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADCP	Acoustic Doppler Current Profiler
AI	Artificial Intelligence
AIC	Akaike Information Criterion
AIS	Automatic Information System
ANN	Artificial Neural Networks
AUC ROC	Area Under Receiver Operating Characteristic Curve
CNN	Convolutional Neural Networks
EMB	European Marine Board
EMODnet	The European Marine Observation and Data Network
EU	European Union
FNR	False Negative Rate
GAN	Generative Adversarial Network
GMM	Gaussian Mixture Model
HELCOM	Baltic Marine Environment Protection Commission
IOC	Intergovernmental Oceanographic Commission
ML	Machine Learning
MSFD	Marine Strategy Framework Directive
PCA	Principal Component Analysis
ResNet	Residual Neural Network
SIME	Swedish Institute for the Marine Environment
UN	United Nations
VGG	Visual Geometry Group

References

1. Guidi, L.; Fernandez Guerra, A.; Canchaya, C.; Curry, E.; Foglini, F.; Irisson, J.O.; Malde, K.; Marshall, C.T.; Obst, M.; Ribeiro, R.P.; et al. *Big Data in Marine Science*; Future Science Brief 6 of the European Marine Board: Ostend, Belgium, 2020.
2. Malde, K.; Handegard, N.O.; Eikvil, L.; Salberg, A.B. Machine intelligence and the data-driven future of marine science. *ICES J. Mar. Sci.* **2020**, *77*, 1274–1285. [CrossRef]
3. Intergovernmental Oceanographic Commission (IOC). *The Science we Need for the Ocean We Want: The United Nations Decade of Ocean Science for Sustainable Development (2021–2030)*; (English) IOC Brochure 2018-7 (IOC/BRO/2018/7 Rev); IOC: Paris, France, 2019; 24p.
4. GOOS. Global Ocean Observing System-GOOS. 2020. Available online: https://www.goosocean.org/ (accessed on 22 June 2020).
5. The European Marine Observation and Data Network (EMODnet). Central Portal | Your Gateway to Marine Data in Europe. 2020. Available online: https://www.emodnet.eu/ (accessed on 22 June 2020).
6. Copernicus-The European Earth Observation Programme. Copernicus-Marine Environment Monitoring Service. 2020. Available online: https://marine.copernicus.eu/ (accessed on 22 June 2020).
7. Allken, V.; Handegard, N.O.; Rosen, S.; Schreyeck, T.; Mahiout, T.; Malde, K. Fish species identification using a convolutional neural network trained on synthetic data. *ICES J. Mar. Sci.* **2018**, *76*, 342–349. [CrossRef]
8. Horne, J.K. Acoustic approaches to remote species identification: A review. *Fish. Oceanogr.* **2000**, *9*, 356–371. [CrossRef]
9. Godø, O.R.; Handegard, N.O.; Browman, H.I.; Macaulay, G.J.; Kaartvedt, S.; Giske, J.; Ona, E.; Huse, G.; Johnsen, E. Marine ecosystem acoustics (MEA): Quantifying processes in the sea at the spatio-temporal scales on which they occur. *ICES J. Mar. Sci.* **2014**, *71*, 2357–2369. [CrossRef]
10. Benoit-Bird, K.J.; Lawson, G.L. Ecological Insights from Pelagic Habitats Acquired Using Active Acoustic Techniques. *Annu. Rev. Mar. Sci.* **2016**, *8*, 463–490. [CrossRef]

11. Bennion, M.; Fisher, J.; Yesson, C.; Brodie, J. Remote Sensing of Kelp (Laminariales, Ochrophyta): Monitoring Tools and Implications for Wild Harvesting. *Rev. Fish. Sci. Aquac.* **2019**, *27*, 127–141. [CrossRef]
12. Zampoukas, N.; Palialexis, A.; Duffek, A.; Graveland, J.; Giorgi, G.; Hagebro, C.; Hanke, G.; Korpinen, S.; Tasker, M.; Tornero, V.; et al. *Technical Guidance on Monitoring for the Marine Strategy Framework Directive*; Publications Office of the European Union: Luxembourg, 2014; ISSN 1831-9424. ISBN 978-92-79-35426-7.
13. Brown, C.J.; Smith, S.J.; Lawton, P.; Anderson, J.T. Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuar. Coast. Shelf Sci.* **2011**, *92*, 502–520. [CrossRef]
14. Gumusay, M.U.; Bakirman, T.; Tuney Kizilkaya, I.; Aykut, N.O. A review of seagrass detection, mapping and monitoring applications using acoustic systems. *Eur. J. Remote Sens.* **2019**, *52*, 1–29. [CrossRef]
15. Misund, O.A. Underwater acoustics in marine fisheries and fisheries research. *Rev. Fish Biol. Fish.* **1997**, *7*, 1–34. [CrossRef]
16. Ross, T.; Gaboury, I.; Lueck, R. Simultaneous acoustic observations of turbulence and zooplankton in the ocean. *Deep. Sea Res. Part I Oceanogr. Res. Pap.* **2007**, *54*, 143–153. [CrossRef]
17. Lombard, F.; Boss, E.; Waite, A.M.; Vogt, M.; Uitz, J.; Stemmann, L.; Sosik, H.M.; Schulz, J.; Romagnan, J.B.; Picheral, M.; et al. Globally Consistent Quantitative Observations of Planktonic Ecosystems. *Front. Mar. Sci.* **2019**, *6*, 196. [CrossRef]
18. Williamson, B.J.; Blondel, P.; Armstrong, E.; Bell, P.S.; Hall, C.; Waggitt, J.J.; Scott, B.E. A self-contained subsea platform for acoustic monitoring of the environment around Marine Renewable Energy Devices-Field deployments at wave and tidal energy sites in Orkney, Scotland. *IEEE J. Ocean. Eng.* **2015**, *41*, 67–81.
19. Francisco, F.; Sundberg, J. Detection of Visual Signatures of Marine Mammals and Fish within Marine Renewable Energy Farms using Multibeam Imaging Sonar. *J. Mar. Sci. Eng.* **2019**, *7*, 22. [CrossRef]
20. Stranne, C.; Mayer, L.; Jakobsson, M.; Weidner, E.; Jerram, K.; Weber, T.C.; Anderson, L.G.; Nilsson, J.; Björk, G.; Gårdfeldt, K. Acoustic mapping of mixed layer depth. *Ocean Sci.* **2018**, *14*, 503–514. [CrossRef]
21. Stranne, C.; Mayer, L.; Weber, T.C.; Ruddick, B.R.; Jakobsson, M.; Jerram, K.; Weidner, E.; Nilsson, J.; Gårdfeldt, K. Acoustic mapping of thermohaline staircases in the arctic ocean. *Sci. Rep.* **2017**, *7*, 1–9. [CrossRef]
22. Lavery, A.C.; Geyer, W.R.; Scully, M.E. Broadband acoustic quantification of stratified turbulence. *J. Acoust. Soc. Am.* **2013**, *134*, 40–54. [CrossRef] [PubMed]
23. European Parliament. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). *Off. J. Eur. Union L* **2008**, *164*, 19–40.
24. Kelpšaite, L.; Parnell, K.; Soomere, T. Energy pollution: The relative influence of wind-wave and vessel-wake energy in Tallinn Bay, the Baltic Sea. *J. Coast. Res.* **2009**, *56*, 812–816.
25. Zaggia, L.; Lorenzetti, G.; Manfé, G.; Scarpa, G.M.; Molinaroli, E.; Parnell, K.E.; Rapaglia, J.P.; Gionta, M.; Soomere, T. Fast shoreline erosion induced by ship wakes in a coastal lagoon: Field evidence and remote sensing analysis. *PLoS ONE* **2017**, *12*, e0187210. [CrossRef] [PubMed]
26. NDRC. *The Physics of Sound in the Sea*; United States Office of Scientific Research and Development, National Defense Research Committee, Division 6: Washington, DC, USA, 1946. Available online: https://www.loc.gov/item/2015490953/ (accessed on 27 April 2020).
27. Soloviev, A.; Gilman, M.; Young, K.; Brusch, S.; Lehner, S. Sonar measurements in ship wakes simultaneous with TerraSAR-X overpasses. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 841–851. [CrossRef]
28. Voropayev, S.; Nath, C.; Fernando, H. Thermal surface signatures of ship propeller wakes in stratified waters. *Phys. Fluids* **2012**, *24*, 116603. [CrossRef]
29. Francisco, F.; Carpman, N.; Dolguntseva, I.; Sundberg, J. Use of Multibeam and Dual-Beam Sonar Systems to Observe Cavitating Flow Produced by Ferryboats: In a Marine Renewable Energy Perspective. *J. Mar. Sci. Eng.* **2017**, *5*, 30. [CrossRef]
30. Trevorrow, M.V.; Vagle, S.; Farmer, D.M. Acoustical measurements of microbubbles within ship wakes. *J. Acoust. Soc. Am.* **1994**, *95*, 1922–1930. [CrossRef]
31. Weber, T.C.; Lyons, A.P.; Bradley, D.L. An estimate of the gas transfer rate from oceanic bubbles derived from multibeam sonar observations of a ship wake. *J. Geophys. Res. Ocean.* **2005**, *110*. [CrossRef]
32. Emerson, S.; Bushinsky, S. The role of bubbles during air-sea gas exchange. *J. Geophys. Res. Ocean.* **2016**, *121*, 4360–4376. [CrossRef]
33. Uliczka, K.; Kondziella, B. Ship-Induced Sediment Transport in Coastal Waterways (SeST). In Proceedings of the 4th MASHCON-International Conference on Ship Manoeuvring in Shallow and Confined Water with Special Focus on Ship Bottom Interaction, Hamburg, Germany, 23–25 May 2016; pp. 2–8.
34. Katz, C.; Chadwick, D.; Rohr, J.; Hyman, M.; Ondercin, D. Field measurements and modeling of dilution in the wake of a US navy frigate. *Mar. Pollut. Bull.* **2003**, *46*, 991–1005. [CrossRef]
35. Loehr, L.C.; Beegle-Krause, C.J.; George, K.; McGee, C.D.; Mearns, A.J.; Atkinson, M.J. The significance of dilution in evaluating possible impacts of wastewater discharges from large cruise ships. *Mar. Pollut. Bull.* **2006**, *52*, 681–688. [CrossRef]
36. Marmorino, G.; Trump, C. Preliminary side-scan ADCP measurements across a ship's wake. *J. Atmos. Ocean. Technol.* **1996**, *13*, 507–513. [CrossRef]
37. Ermakov, S.A.; Kapustin, I.A. Experimental study of turbulent-wake expansion from a surface ship. *Izv. Atmos. Ocean. Phys.* **2010**, *46*, 524–529. [CrossRef]

38. Kang, L. Wave Monitoring Based on Improved Convolution Neural Network. *J. Coast. Res.* **2019**, *94*, 186–190. [CrossRef]
39. Holzinger, A. Interactive machine learning for health informatics: When do we need the human-in-the-loop? *Brain Inform.* **2016**, *3*, 119–131. [CrossRef] [PubMed]
40. Holzinger, A.; Plass, M.; Holzinger, K.; Crisan, G.C.; Pintea, C.M.; Palade, V. A glass-box interactive machine learning approach for solving NP-hard problems with the human-in-the-loop. *Creat. Math. Inform.* **2019**, *28*, 121–134.
41. Guo, X.; Yu, Q.; Li, R.; Alm, C.O.; Calvelli, C.; Shi, P.; Haake, A. An expert-in-the-loop paradigm for learning medical image grouping. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*; Springer: Cham, Switzerland, 2016; pp. 477–488.
42. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]
43. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
44. Alexandrou, D.; Pantzartzis, D. A methodology for acoustic seafloor classification. *IEEE J. Ocean. Eng.* **1993**, *18*, 81–86. [CrossRef]
45. Marsh, I.; Brown, C. Neural network classification of multibeam backscatter and bathymetry data from Stanton Bank (Area IV). *Appl. Acoust.* **2009**, *70*, 1269–1276. [CrossRef]
46. Feldens, P.; Darr, A.; Feldens, A.; Tauber, F. Detection of boulders in side scan sonar mosaics by a neural network. *Geosciences* **2019**, *9*, 159. [CrossRef]
47. van Overmeeren, R.; Craeymeersch, J.; van Dalfsen, J.; Fey, F.; van Heteren, S.; Meesters, E. Acoustic habitat and shellfish mapping and monitoring in shallow coastal water—Sidescan sonar experiences in The Netherlands. *Estuar. Coast. Shelf Sci.* **2009**, *85*, 437–448. [CrossRef]
48. Lawson, G.L.; Barange, M.; Fréon, P. Species identification of pelagic fish schools on the South African continental shelf using acoustic descriptors and ancillary information. *ICES J. Mar. Sci.* **2001**, *58*, 275–287. [CrossRef]
49. Korneliussen, R.J.; Heggelund, Y.; Macaulay, G.J.; Patel, D.; Johnsen, E.; Eliassen, I.K. Acoustic identification of marine species using a feature library. *Methods Oceanogr.* **2016**, *17*, 187–205. [CrossRef]
50. Williamson, B.J.; Fraser, S.; Blondel, P.; Bell, P.S.; Waggitt, J.J.; Scott, B.E. Multisensor Acoustic Tracking of Fish and Seabird Behavior Around Tidal Turbine Structures in Scotland. *IEEE J. Ocean. Eng.* **2017**, *42*, 948–965. [CrossRef]
51. Fraser, S.; Nikora, V.; Williamson, B.J.; Scott, B.E. Automatic active acoustic target detection in turbulent aquatic environments. *Limnol. Oceanogr. Methods* **2017**, *15*, 184–199. [CrossRef]
52. Fernandes, P.G. Classification trees for species identification of fish-school echotraces. *ICES J. Mar. Sci.* **2009**, *66*, 1073–1080. [CrossRef]
53. Haralabous, J.; Georgakarakos, S. Artificial neural networks as a tool for species identification of fish schools. *ICES J. Mar. Sci.* **1996**, *53*, 173–180. [CrossRef]
54. Simmonds, J.E.; Armstrong, F.; Copland, P.J. Species identification using wideband backscatter with neural network and discriminant analysis. *ICES J. Mar. Sci.* **1996**, *53*, 189–195. [CrossRef]
55. Ramani, N.; Patrick, P.H. Fish detection and identification using neural networks-some laboratory results. *IEEE J. Ocean. Eng.* **1992**, *17*, 364–368. [CrossRef]
56. Cabreira, A.G.; Tripode, M.; Madirolas, A. Artificial neural networks for fish-species identification. *ICES J. Mar. Sci.* **2009**, *66*, 1119–1129. [CrossRef]
57. Brautaset, O.; Waldeland, A.U.; Johnsen, E.; Malde, K.; Eikvil, L.; Salberg, A.B.; Handegard, N.O. Acoustic classification in multifrequency echosounder data using deep convolutional neural networks. *ICES J. Mar. Sci.* **2020**, *77*, 1391–1400. [CrossRef]
58. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2012; pp. 1097–1105.
59. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
60. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
61. Van Dyk, D.A.; Meng, X.L. The art of data augmentation. *J. Comput. Graph. Stat.* **2001**, *10*, 1–50. [CrossRef]
62. Reynolds, D.A. Gaussian Mixture Models. *Encycl. Biom.* **2009**, *741*, 659–663.
63. Hurvich, C.M.; Simonoff, J.S.; Tsai, C.L. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **1998**, *60*, 271–293. [CrossRef]
64. Cao, L.; Chua, K.S.; Chong, W.; Lee, H.; Gu, Q. A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomputing* **2003**, *55*, 321–336. [CrossRef]
65. Port of Gothenburg. 2020. Available online: https://www.portofgothenburg.com/FileDownload/?contentReferenceID=12900 (accessed on 18 May 2020).
66. HELCOM. HELCOM Assessment on Maritime Activities in the Baltic Sea 2018. Baltic Sea Environment Proceedings No.152. 2018. Available online: https://www.helcom.fi/wp-content/uploads/2019/08/BSEP152-1.pdf (accessed on 14 January 2021).
67. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2014; pp. 2672–2680.
68. Ren, M.; Zeng, W.; Yang, B.; Urtasun, R. Learning to reweight examples for robust deep learning. *arXiv* **2018**, arXiv:1803.09050.
69. Shanmugam, D.; Blalock, D.; Balakrishnan, G.; Guttag, J. When and Why Test-Time Augmentation Works. *arXiv* **2020**, arXiv:2011.11156.
70. Wang, G.; Li, W.; Aertsen, M.; Deprest, J.; Ourselin, S.; Vercauteren, T. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing* **2019**, *338*, 34–45. [CrossRef]

71. Zheng, Q.; Yang, M.; Tian, X.; Jiang, N.; Wang, D. A Full Stage Data Augmentation Method in Deep Convolutional Neural Network for Natural Image Classification. *Discret. Dyn. Nat. Soc.* **2020**, *2020*, 4706576. [CrossRef]
72. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]
73. Lasseck, M. Bird song classification in field recordings: Winning solution for NIPS4B 2013 competition. In Proceedings of the Int. Symp. Neural Information Scaled for Bioacoustics, sabiod.org/nips4b, Joint to NIPS, Reno, NV, USA, 10 December 2013; pp. 176–181.