

CLASSIFICATION OF SHIPS FROM UNDERWATER SOUND USING A DEEP LEARNING NETWORK

Jari Kataja Patria, Tampere, Finland
 Petri Pirttikangas Patria, Tampere, Finland

1 INTRODUCTION

Ships and other watercraft produce sound that can be detected by underwater sensor systems such as sonar. Traditionally, the classification of ship noise has been a task of well-trained sonar operators, and it has been done based on human hearing. Therefore, the development of reliable identification and classification systems plays an important role. The focus of technical development has been on feature extraction methods and training the classifier¹. Feature extraction works as an essential part of a deep-learning network to identify the origin of detected underwater noise.

Features obtained from MFCC coefficients (Mel Frequency Cepstral Coefficients) or the logarithmic Mel spectrogram have been widely used in the classification of environmental sound². They, together with features extracted from LOFAR (Low Frequency Analysis and Recording) spectrograms, are also well suited for underwater acoustic identification^{3,4,5}. In the Mel spectrogram, the frequencies of the audio signal have been converted to the Mel scale, which better represents the frequencies perceived by humans. In this paper, Mel spectrograms were chosen as the input for classification.

The sound transmission path from the ship to the observation point is complex, which poses its own challenge for ship classification. Getting as comprehensive set of samples as possible for training the classifier is therefore crucial. Another important issue is the robustness of the system, including the ability to correctly classify targets and minimize false alarms.

2 UNDERWATER ACOUSTIC EMISSION OF SHIPS

Ship noise can be generally divided into machinery and hydrodynamic sound, as shown in Figure 1⁶. Machinery noise is generated by various rotating machines, such as main and auxiliary engines. It is essentially narrowband and has low-frequency content, typically below a few kilohertz. Hydrodynamic noise is caused by interaction of the ship hull and propellers with the surrounding water, and it is broadband by nature.

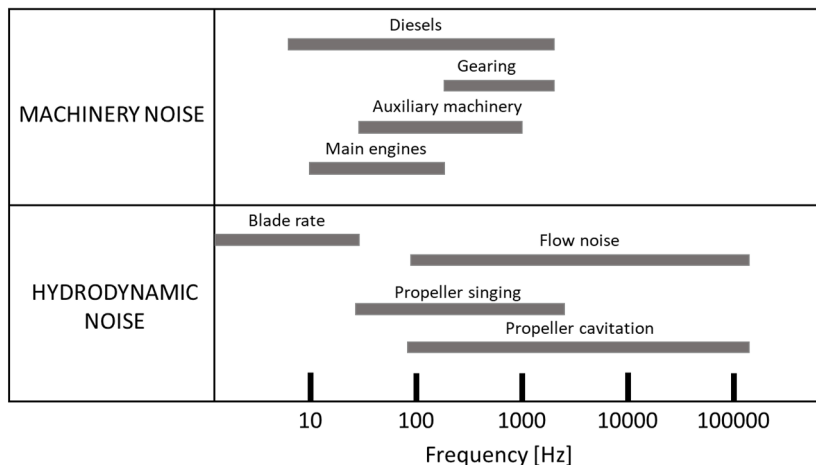


Figure 1. Ship noise components.

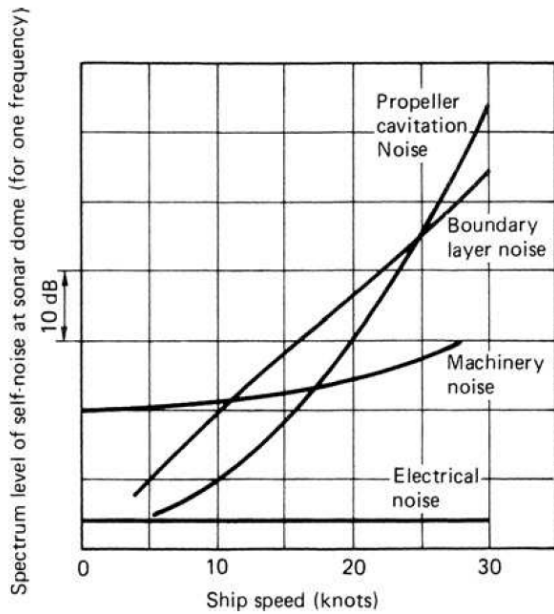


Figure 2. Variation of some noise components with respect to ship speed.

Ship noise sources generate sound at different intensities as a function of the speed (Figure 2)⁷. Source levels and nature depend on the type of both the vessel and its propulsion system. At slow speed, machinery noise dominates the underwater noise emission, while at high speed, flow noise due to boundary layer and propeller cavitation become dominant. In addition to this, the ship is not omnidirectional as a sound source, but the observed sound depends on the orientation of the ship relative to the observation point. Observing ship noise generally requires multiple measurement points and a moving vessel⁸.

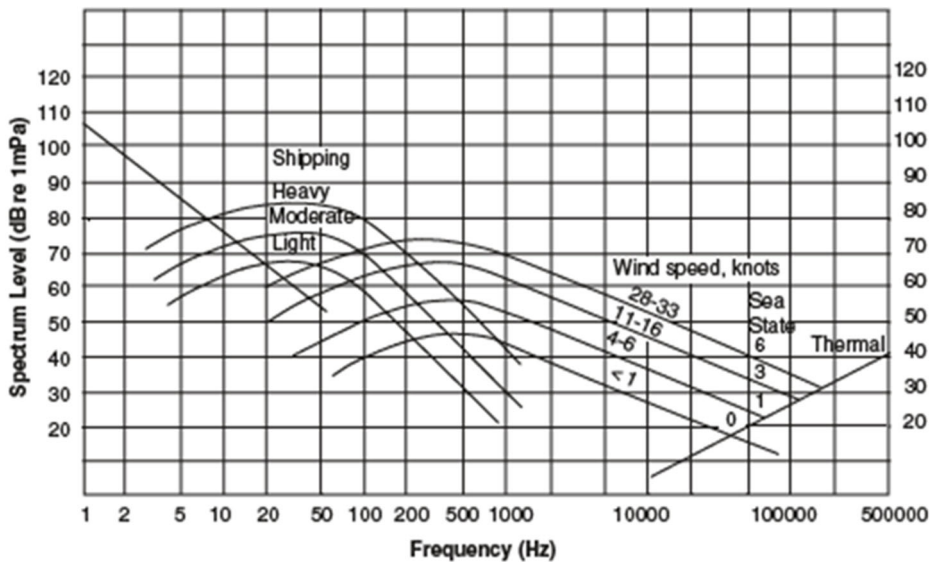


Figure 3. Background noise types.

The observed sound always contains background noise (Figure 3). At the infrasound range, noise is caused by seismic activity, temperature changes, ocean currents and tides. Low-frequency noise from distant shipping is dominant at a frequency range below a few hundred hertz. Noise caused by

wind consists of wind turbulence, water surface movement, wave interaction, spurts, and bubbling. It is often presented as curves, at which different wind speeds indicate different sea states, as developed by Knudsen. In addition to the noise components presented above, the observed sound is affected by acoustics of the underwater environment, involving distance attenuation as well as surface and bottom reflections.⁹

3 CLASSIFICATION OF SHIP NOISE

3.1 Ship noise data and pre-processing

For classification, data was first achieved using a ship noise simulator, which can generate realistic sound for various vessel types from a motorboat to a cargo ship. The simulated noise consists of the components shown in Figure 1, weighted characteristically for the vessel type. Simulation data was also filtered due to distance attenuation and background noise was added (Figure 3). By varying distance and background noise components, data variation was increased, although the effect of the ship's orientation was not considered. Data was generated for five vessel types, representing a ro-ro, a tanker, a cargo ship, a yacht, and a motorboat.

The second data set acquired for the classifier contained measurement data from various ships near the Spanish city of Vigo. The recordings were made with a set of hydrophones of many types of vessels, sixteen in total. The ships belonged to eight vessel types from a pilot boat to ocean cruisers. In addition, there were measurements of background noise, which was used as one class.¹⁰

Data preprocessing began with the creation and naming of consistent audio samples. Consistency means that every sample has the same data length, number of channels and sampling frequency. The data was also split into training and validation sets, in 80–20 ratio, the former of which is used to train the classifier and the latter to verify its performance. After that, the samples were augmented by shifting them randomly in time either forwards or backwards. Mel spectrograms were then computed from the samples, and they were augmented by masking random time or frequency ranges. With augmentation, data variability was increased. Data augmentation can also combat overfitting and improve the generalization ability of deep neural network¹¹.

3.2 Deep learning classifier

The classification of spectrogram images was carried out with the CNN (Convolutional Neural Network) structure shown in Figure 4. Such structure has been used e.g. for urban noise classification^{12,13}. Each CNN layer applies its filters to increase the image depth, i.e., the number of channels. The image width and height are reduced as the kernels and strides are applied. Finally, after passing through the four CNN layers, feature maps are obtained. They are then pooled and flattened in the adaptive layer and input to the linear layer, which outputs one prediction score per class. Class having the highest score is selected as the classification result for the image.

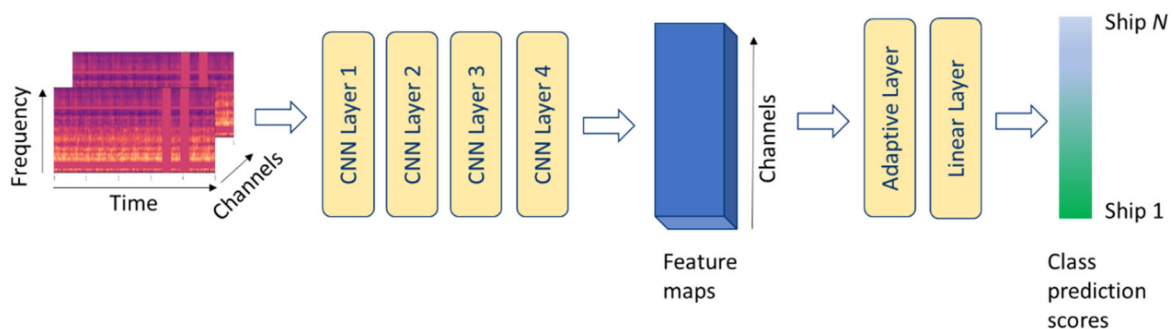


Figure 4. Structure of the classifier.

In both training and validation phases, data is processed in batches. A batch of spectrogram images is given as the input to the classifier as a 4-dimensional table with a size of $16 \times 2 \times 64 \times 344$ (*batch size* \times *number of channels* \times *Mel frequency bands* \times *time steps*). The feature maps being the output of the 4th CNN layer has a size of $16 \times 64 \times 4 \times 22$, indicating an increase in the number of channels and decrease in image size. The feature maps are processed in the Adaptive Layer to a size of 16×64 and fed to the Linear layer, which outputs one prediction score per class, i.e., the size is $16 \times N$.

The classifier was implemented with PyTorch, which is an open-source Python package suitable for tensor computation with GPU acceleration and deep neural networks¹⁴. For classifier training, functions for the optimizer, loss, and scheduler were defined to dynamically vary the learning rate as training progresses, which usually allows training to converge in fewer epochs. The classifier was trained for several epochs, processing a batch of data in each iteration. Track of a simple accuracy metric was kept, measuring the percentage of correct predictions. After training, an inference loop was run with the gradient updates disabled for validating the classifier.

4 CLASSIFICATION RESULTS

The classifier was first trained with the training data labeled with known classes. Training was carried out separately for the simulation and measurement data. In the case of measurement data, the classifier was trained to identify either vessel types or individual vessels. The classification results shown below were obtained with validation data.

Classification results obtained with the simulation data are presented as a confusion matrix in Figure 5. Confusion matrix is a table comparing the actual classes with those predicted by the classifier. The diagonal values of the confusion matrix indicate the percentage of correctly classified cases for each class. The motorboat has been classified correctly in all cases. In the other classes, misclassification has occurred in about 4-6 % of the cases. For example, the yacht has been identified as the motorboat and the cargo ship as the ro-ro. The overall classification accuracy is approximately 96%.

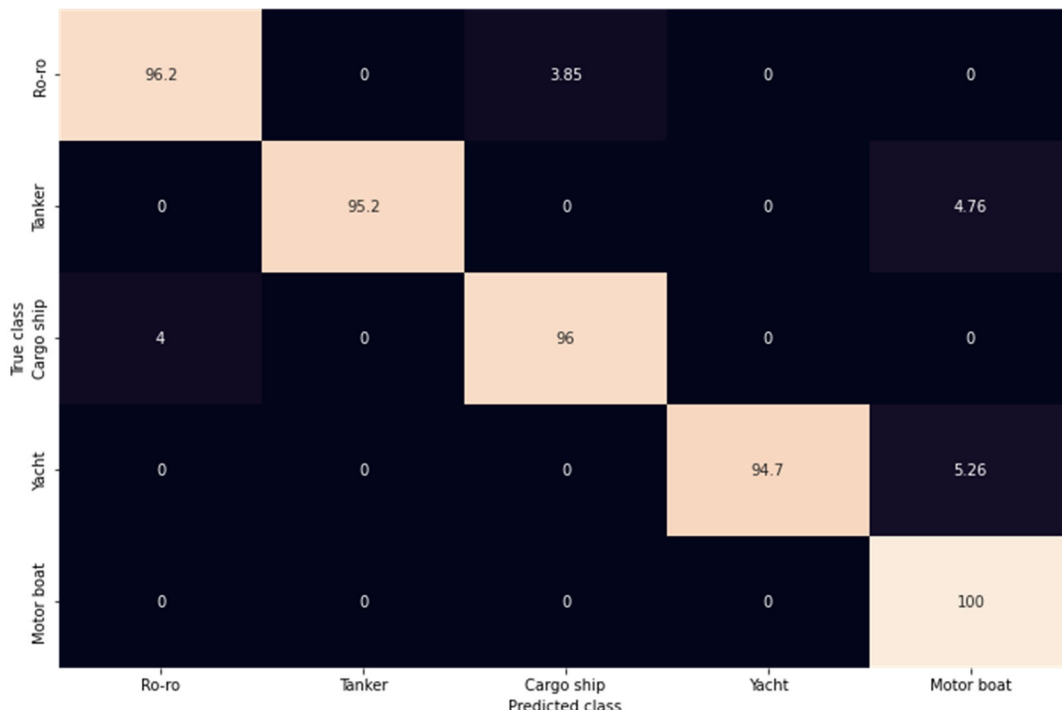


Figure 5. Classification results for simulated vessels.

With the measurement data, the results of classification of ship types are given in Figure 6. In the confusion matrix, values are amounts in samples. Since different numbers of samples were available for different vessel types, relative amounts would distort the results. In this classification case, there are misclassifications mostly with the ro-ro and the passenger ship. The dredger has been classified almost perfectly and only once misclassified. Ambient noise has been once identified as the ocean liner and the passenger ship once as the ambient noise. Reasons for that are discussed further in the end of this chapter. The overall classification accuracy is 93%.

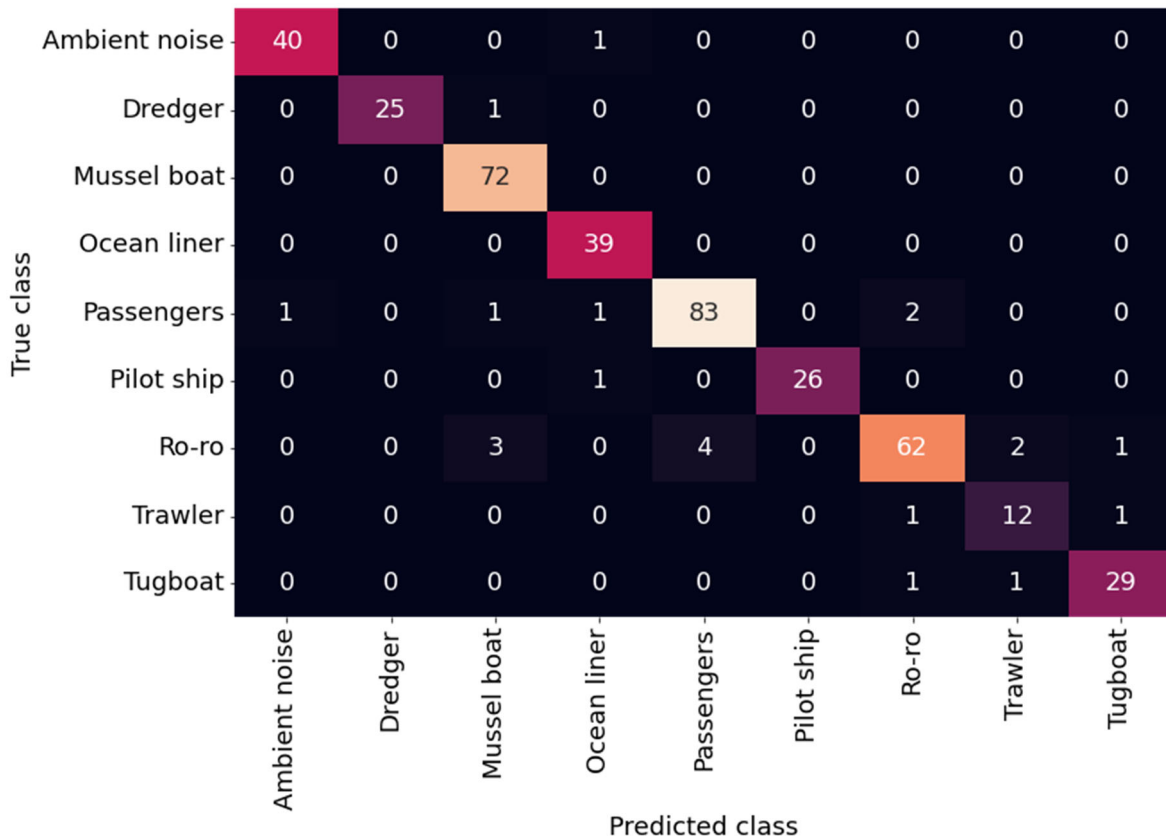


Figure 6. Classification results for vessel types.

Figure 7 shows the results for classification of single ships. Again, the values represent amounts. The most difficult to identify is a passenger ship named Mar de Onza, which has been falsely classified 5 times out of 32. Two times it has been interpreted as the other passenger ship, the Mar de Cangas. Mar de Cangas has also been mixed twice with Mar de Onza. A ro-ro called Viking Chance has been misclassified 2 times out of 12, as the other ro-ro Autopride. Several ships, such as both ocean liners, have been classified correctly in every case. The overall classification accuracy is 93%.

Measurement data consisted of passing vessels, meaning that audio samples further away from the vessel had a low signal-to-noise ratio (SNR). Those samples contained higher level of ambient noise, which was also used as one class. Ambient noise had been measured separately, and samples were used in the classifier training. A few samples of the vessel classes were classified as ambient noise, and vice versa, due to the low SNR.

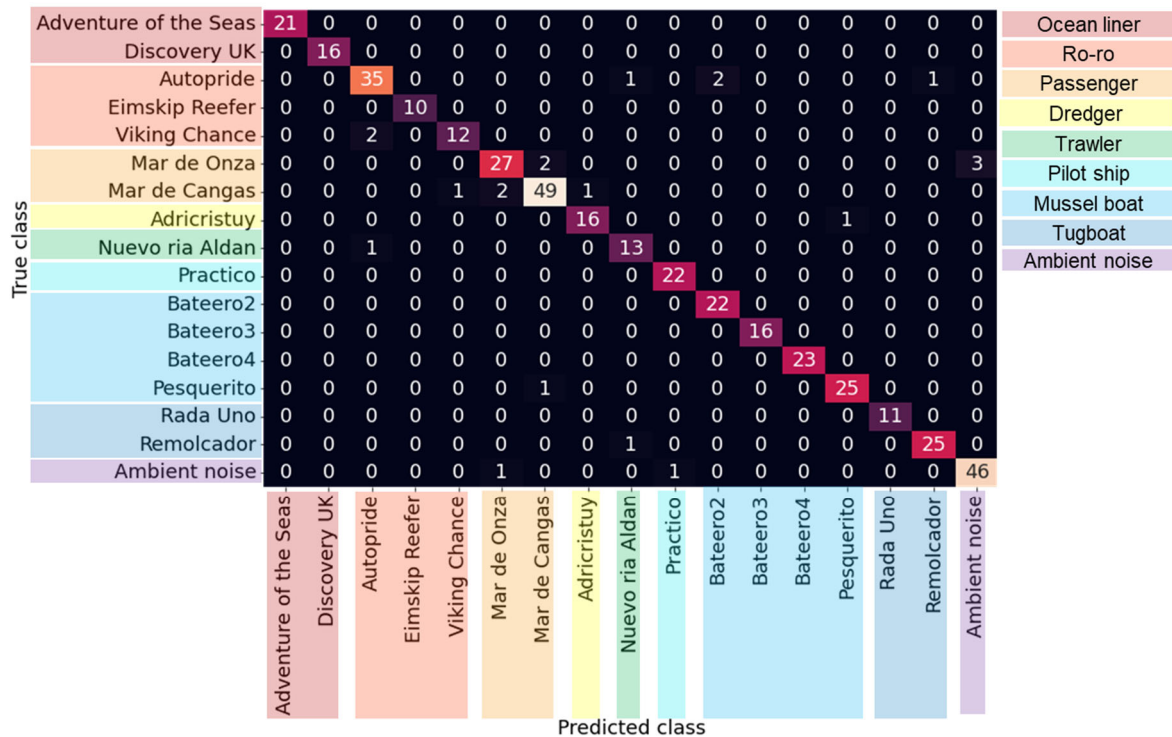


Figure 7. Classification results for individual vessels.

5 CONCLUSIONS

Underwater noise produced by ships can be reliably classified using a deep-learning CNN network and Mel spectrogram images. Although a relatively low amount of measurement data was available, accuracy obtained with validation data is above 90%. It should be noted that training of the classifier is a balancing act between classification accuracy and generalizability. By training the classifier to identify the training data too precisely leads easily to overfitting, in which case the classifier's performance in identifying other data suffers. On the other hand, data augmentation helps to tackle overfitting and improves the generalization ability.

The distance and orientation of the ship, together with underwater acoustics, also set challenges for classification. In the case of simulated ship noise, the orientation of the ship could not be varied, so the only variation came from the distance and background noise. The measurement data had been recorded of passing ships, whose orientation obviously varied in time. This increases the possibilities of the classifier to identify ships in a realistic environment. Verification of the classifier's performance would require yet separate testing data from the same ships present in the training and validation data.

6 ACKNOWLEDGEMENTS

We thank David Santos Domínguez from University of Vigo (Universidade de Vigo) for providing the ShipsEar database to us for academic purposes.

7 REFERENCES

1. F. Hong et al., Underwater Acoustic Target Recognition with a Residual Network and the Optimized Feature Extraction Method, *Applied Sciences* (2021)
2. J. Li et al., A comparison of deep learning methods for environmental sound detection, *Proc. ICASSP 2017*, 126-130, New Orleans, USA (2017)
3. L. Domingos et al., A Survey of Underwater Acoustic Data Classification Methods Using Deep Learning for Shoreline Surveillance, *Sensors* (2022)
4. L. Zhang et al., Feature extraction of underwater target signal using Mel frequency cepstrum coefficients based on acoustic vector sensor, *J.Sensors* (2016)
5. P. Wang and Y. Peng, Research on underwater acoustic target recognition based on LOFAR spectrum and deep learning method, *Proc. CACRE 2020*, Dalian, China (2020)
6. C. Audoly, AQUO Achieve QUIeter Oceans by shipping noise footprint reduction FP7 - Collaborative Project no. 314227, Task report (2015)
7. J. Carlton, *Marine Propellers and Propulsion*, 2nd ed Elsevier Ltd, Oxford, UK (2007)
8. T. Gaggero et al, Directivity patterns of ship underwater noise emissions. *Proc. 1st Int. Conf. and Exhibition on Underwater Acoustics*, 1295-1301, Greece (2013)
9. R. Urick, *Principles of Underwater Sound*, 3rd ed Peninsula Publishing (1983)
10. D. Santos-Domínguez et al., ShipsEar: An underwater vessel noise database, *Applied Acoustics* 113 64–69 (2016)
11. S. Wei, S. Zou, F. Liao and W. Lang, A Comparison on Data Augmentation Methods Based on Deep Learning for Audio Classification, *J. Phys.: Conf. Ser.* 1453 (2020)
12. Audio Deep Learning Made Simple: Sound Classification, Step-by-Step, web page, referred 19th Apr 2024, <https://towardsdatascience.com/audio-deep-learning-madesimple-sound-classification-step-by-step-cebc936bbe5>
13. M. Massoudi, S. Verma and R. Jain, Urban Sound Classification using CNN, *Proc. 6th International Conference on Inventive Computation Technologies (ICICT)* (2021)
14. PyTorch, web page, referred 19.4.2024. <https://pytorch.org>