ACOUSTICS 2025
JOONDALUP
SOUNDS OF THE SUNSET COAST

An Application of Machine Learning to the Classification of Vessels Using an Underwater Acoustic Sensor

Ming Jin [(1)] and Chaoying Bao [(1)]

(1) Defence Science and Technology Group, Department of Defence, Perth, Australia

ABSTRACT

This paper presents a study on vessel classification based on their underwater acoustic signatures. A hydrophone array was deployed on the seabed at Chowder Bay, Sydney and recorded underwater acoustic signals from various vessels. Of these vessels, four different ferries were targeted, and hundreds of their noise segments were selected and labelled for data classification. The time-domain signals were first converted into spectrograms, which were then classified using an image-based machine learning method: convolutional neural network (CNN). In the classification process, 77% of the data samples were used for training, while the remaining 23% were reserved for validation. The results show that pre-processing of spectrograms plays a critical role in the classification accuracy. The paper also examines the influence of pre-processing parameters on classification accuracy and computational efficiency. With appropriate pre-processing and model parameters, the proposed method achieved successful classification of ferry acoustic signals with an accuracy of up to 99%.

1 Introduction

Machine learning (ML) has been extensively applied for data classification across various acoustic domains. Some previous research has demonstrated its potential for underwater acoustic applications. Du et al. and Padfield reported in their work that classification of snapping shrimp and dolphins in shallow water using recorded underwater sound was feasible. We explore a similar approach for the recognition of vessels, which would prove practical and significant within the vicinity of a port. Vessel noises are primarily determined by their propulsion system. Theoretically, by extracting the unique signatures from the underwater sound of different types of vessels, it should be feasible to identify them using appropriate ML models. However, this application was rarely reported due to the scarcity of labelled acoustic data for vessels.

This paper aims to explore the use of ML for vessel classification. An image-based ML model, Convolutional Neural Network (CNN), was employed to identify four ferry types based on their underwater acoustic signals, using recordings taken from a single hydrophone. As a preliminary investigation, this research does not claim to be a comprehensive state-of-the-art study. Instead, its primary objective is to verify the feasibility of ML for vessel classification and investigate the impact of some key factors on this practical application.

2 Approach

2.1 Underwater Acoustic Data

Defence Science and Technology Group (DSTG) and a local industry partner, Midspar Systems, collaboratively executed a project aimed at monitoring marine traffic in Sydney. A hydrophone array was installed on the seabed at Chowder Bay, accompanied by a camera mounted on a nearby jetty. The camera was programmed to automatically track the loudest vessel based on the real-time feedback from the seabed array in order to provide the “ground truth” to the acoustic recording. The monitoring system was in operation for over 300 days, thereby accumulating an extensive audio and video database of local vessels. For the purpose of this study, a short period of audio data recorded by a single hydrophone within the array and corresponding video were solicited for the ML application.

Three ferry types, which were observed with high frequency and regularity in the dataset, were selected as the observations for classification. Comparing with other civilian vessels that were detected, these ferries were larger and noisier, were easy to track and therefore offered superior acoustic signals with high signal-to-noise ratio (SNR). By monitoring the video and audio simultaneously, the acoustic data segments dominated by a single ferry were successfully isolated and labelled from the dataset.

Figure 1 shows the ferries and their acoustic signals after pre-processing. For clarity, the three types of ferries were named by their colours as “blue ferry”, “red ferry” and “green ferry”. The green ferry has propulsion systems at both ends, resulting in direction-dependent acoustic signatures. Therefore, the acoustic data of the green ferry is classified into two groups as: “eastbound” and “westbound”. From Figure 1, it is clear that most of received acoustic energy of the ferries is concentrated in the low-frequency range below 300 Hz, and their harmonic distributions show obvious differences.

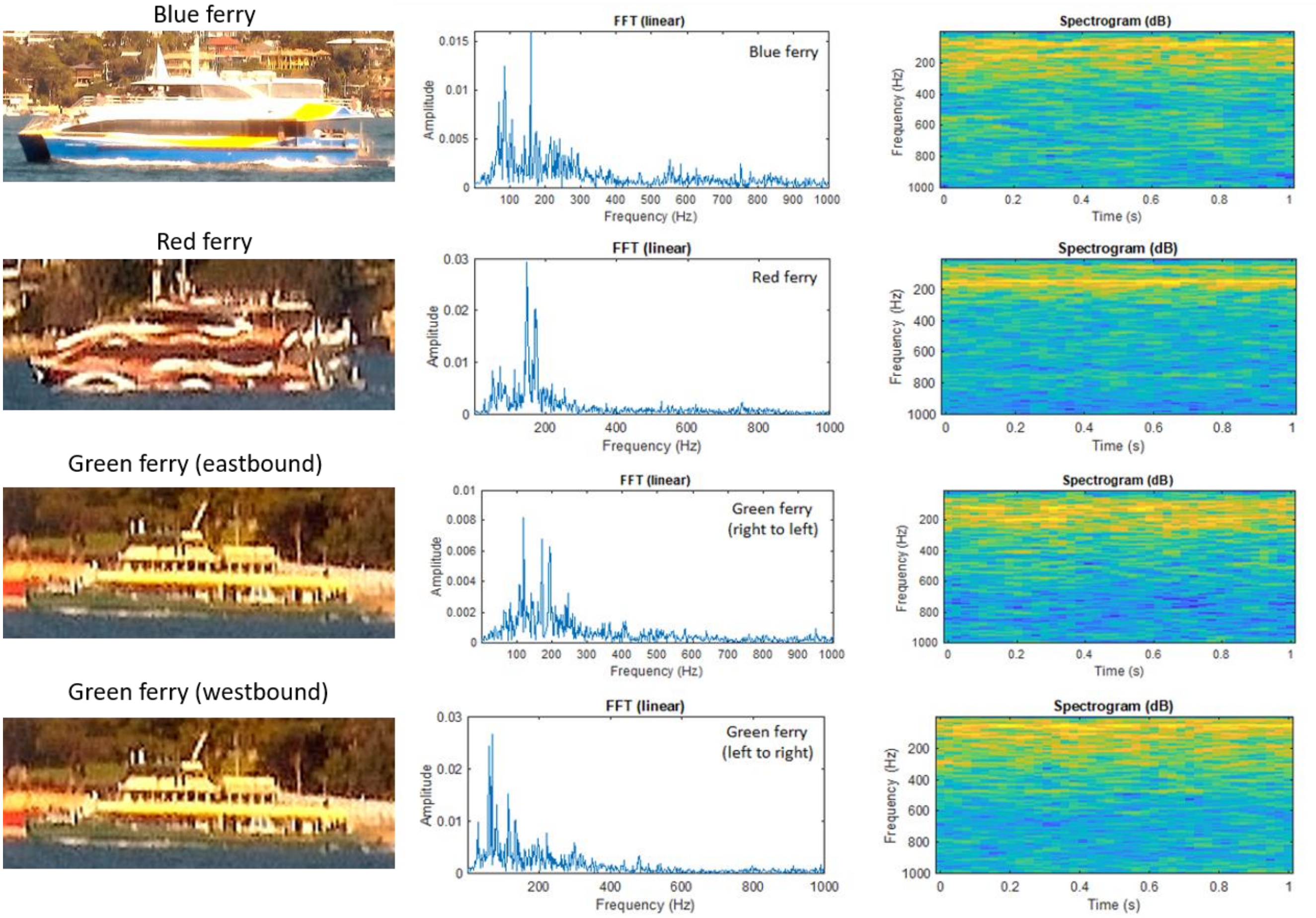

Figure 1: Four classes of ferries (left), their typical spectra (middle) and spectrograms (right)

2.2 Machine Learning Model

To describe the acoustics features of the ferries in both time-domain and frequency domain, the recorded underwater signals were converted into spectrograms, which could be classify by image-based ML models. CNN method, which is a widely used and effective deep learning network for image analysis, was utilised for classifying the ferries’ data here. Figure 2 displays the structure of the 10-layer CNN model coded by readily available ML tools in MATLAB. It consists of:

- one input layer for image data input;
- two 2D convolution layers for extracting features from input images;
- two Rectified Linear Unit (ReLU) layers for simplifying the feature matrices;
- one max pool layer for reducing the size of feature matrices after the first convolution layer which can improve the robustness of the network;

- one dropout layer for avoiding overfitting issue;
- one fully connected layer for combining the feature matrices based on the classification groups;
- one soft max layer for improving the accuracy of output; and
- one class output layer for computing necessary output coefficients.

The raw acoustic data of ferries were divided into hundreds of one-second-long samples. Then the time-domain samples were converted into spectrograms as the input images of the CNN model. Table 1 gives the number of samples in each class. In the classification process, 77% of the samples were used for training, while the remaining 23% were reserved for validation.

Table 1: Number of acoustic samples used for data classification

Blue ferry	Red ferry	Green ferry (eastbound)	Green ferry (westbound)	Total
462	297	484	319	1562

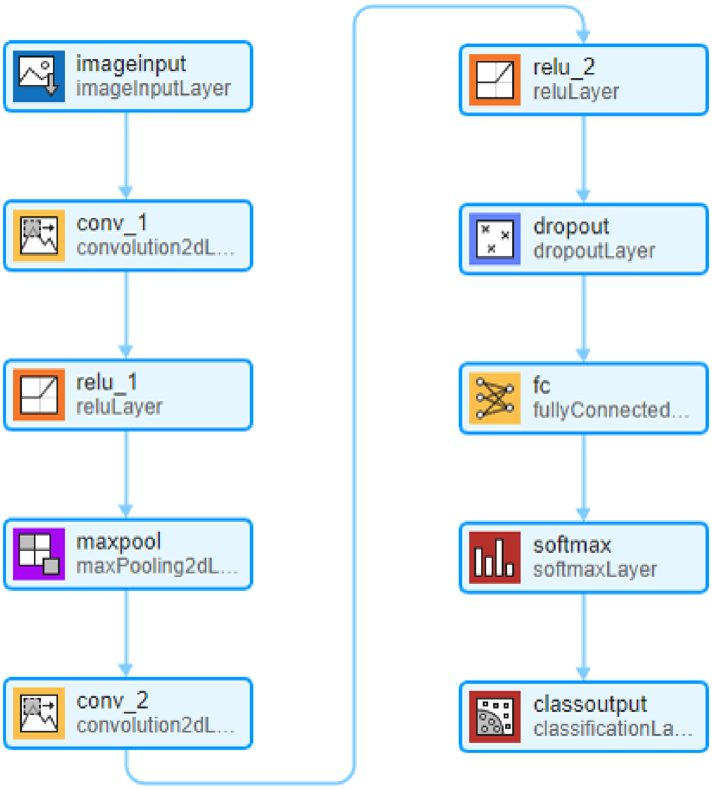

Figure 2: The 10-layer CNN model used for data classification

3 Results and Discussion

As a preliminary study, the ferry data selected for this project intentionally has high SNR and low interference from environment and other vessels. Therefore, the classification accuracy with appropriate settings is anticipated to be artificially high. Figure 3 shows the results of an optimised 10-layer CNN model, and the key parameters are listed in Table 2. It is clear that the accuracy is up to 99%.

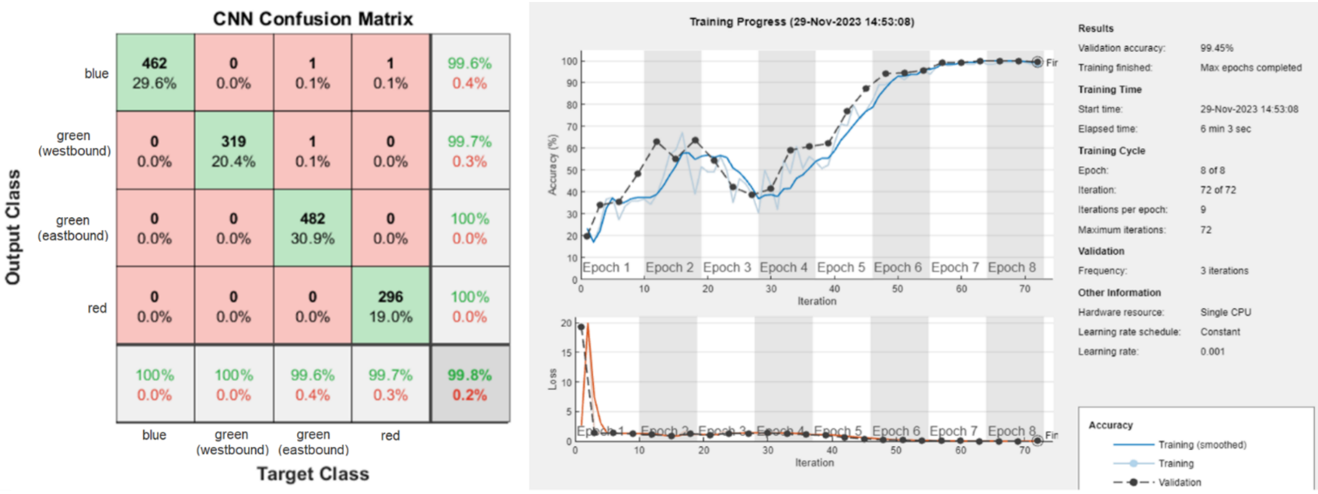

Figure 3: The confusion matrix (left) and training curves (right) of an optimised CNN model

Table 2: The parameters of the CNN model

Layer	Parameter	Value	Comments
Image Input layer	Number of images	1562	1200 for training, 362 for validation
	Image format	matrix	Unsigned 16-bit integer
	Image size	256×32×1	Frequency range 1~512 Hz, frequency resolution 2Hz, time resolution 1/32 s
First convolutional layer	Filter size	7×5	No padding. Stride is 1
	Number of filters	32	
Max pooling layer	Pool size	2×2	No padding
	Stride	[2 2]	
Second convolu-tional layer	Filter size	7×5	No padding. Stride is 1
	Number of filters	32	
Dropout layer	Dropout rate	0.5	
Fully connected layer	Class	4	
Global parameters	Initial learning rate	0.001	Fixed rate
	Max learning epochs	8	
	Validation fre-quency	1/3	

The classification accuracy of CNN models is primarily influenced by two factors: the quality of input images and the configuration of the model. It was observed that the quality of the input spectrograms is the predominant factor in the ferry classification task, while the model settings mainly impact the computational efficiency. When the input spectrograms meet the required quality, the CNN model demonstrates a high level of robustness, yielding satisfactory accuracy across a broad range of parameters. The model optimisation, on the other hand, has limited impact on enhancing classification accuracy when the quality of input images is poor. The essential requirement for the input images is that they must contain enough valid information of observations for data classification.

The study revealed that the quality of the ferries' spectrograms is mainly contingent on three parameters: amplitude display mode, frequency range and frequency resolution. Firstly, it is necessary to display the amplitude of spectrograms using logarithmic scale (dB mode). Figure 4 compares the spectrograms of the same sample with logarithmic and linear scales, elucidating that the logarithmic plot offers more details of the sample in the frequency domain, particularly for the weaker harmonics. The classification accuracy of the CNN model using spectrograms with linear scale was less than 30%, suggesting that the entire pattern consisting of harmonics was more significant for data classification than the several strong frequency components.

Furthermore, the input spectrograms need to be broadband, encompassing most of the valid signal in the frequency domain. Additionally, the frequency resolution should be sufficiently fine to display the details of harmonics. A sensitivity analysis over various values of frequency range and resolution was performed. Figure 5 illustrates the impact of the frequency range and resolution on the classification accuracy and efficiency. In terms of accuracy, even though most of the acoustic energy of the ferries is concentrated in the low-frequency range below 300 Hz (see Figure 1), the upper frequency limit of the input spectrograms needs to be expanded to cover more weak harmonics. The abundance of harmonics also necessitates a relatively finer frequency resolution. The left diagram of Figure 5 demonstrates that, to achieve ideal classification accuracy (>99%), the thresholds of upper frequency limit and frequency resolution were 500 Hz and 2 Hz, respectively. Further enhancements to these two parameters do not optimise the accuracy much but increase the computational time (see right diagram of Figure 5).

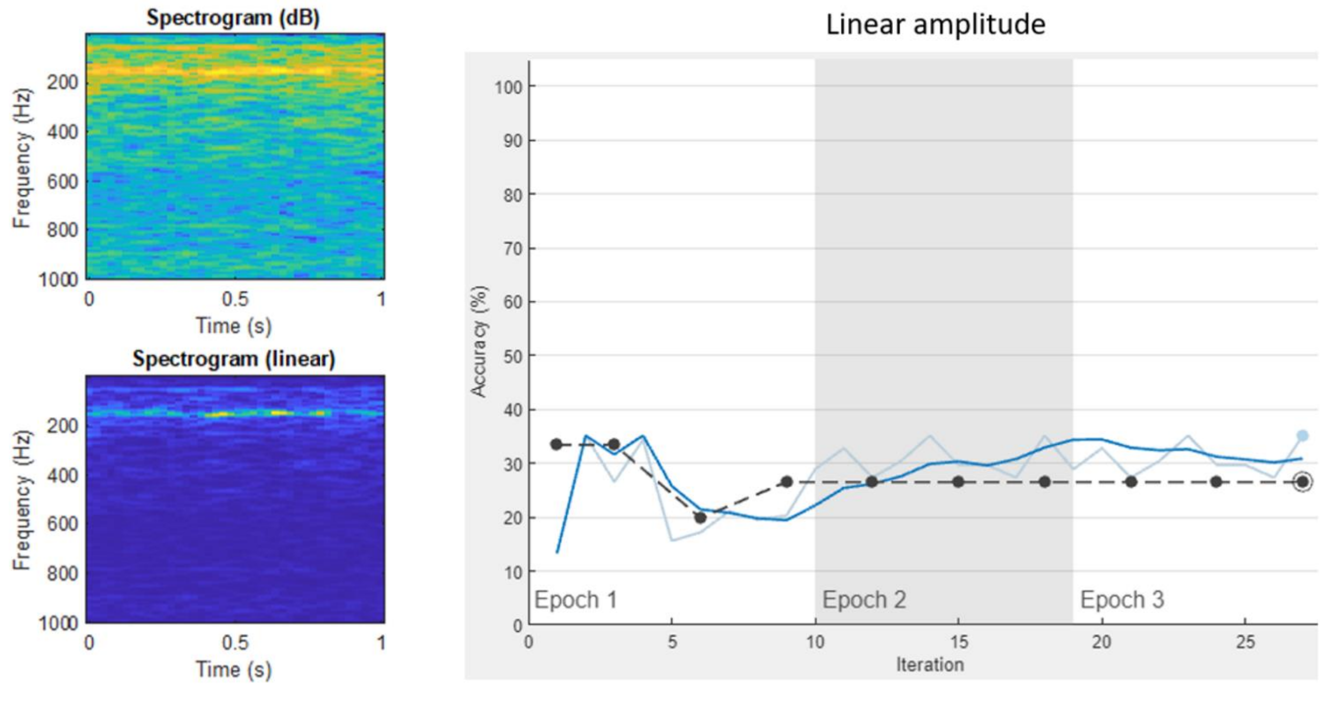

Figure 4: The spectrograms of the same sample with two different display modes (left) and the training result with the linear mode input (right)

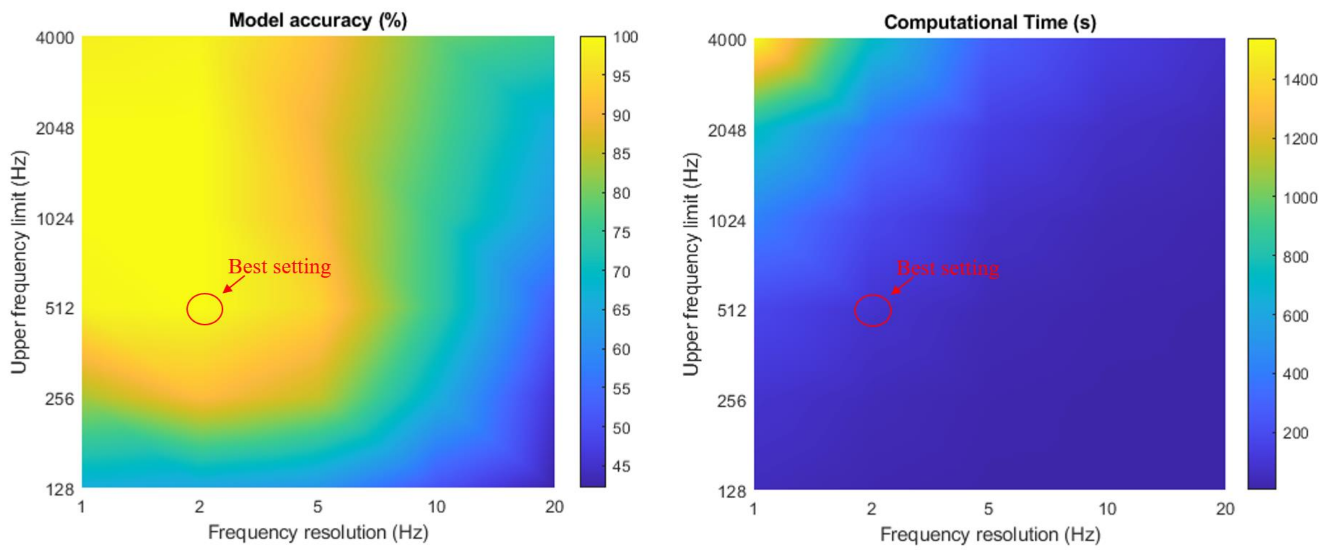

Figure 5: The impact of frequency range and resolution of the input spectrograms on the accuracy (left) and efficiency (right) of the CNN model

4 Conclusion and Future Work

This paper presented an initial exploration of ML applications for the classification of vessels based on underwater acoustic signals. An image-based 10-layer CNN model was utilised to successfully classify four distinct ferry types. The findings of this study provided valuable practical evidence supporting the viability of this method and highlighted several key factors contributing to its success. Notably, the pre-processing of input images was found to be critical for accurate classification. The input images used for this ferry classification task were the spectrograms converted from the recorded underwater signals. To ensure sufficient information for classification, the input spectrograms should meet the following criteria:

- Display the amplitude of sound with logarithmic scale to enhance the relatively weak harmonics.
- Cover an appropriate frequency range that includes most of harmonics.
- Utilise an appropriate frequency resolution based on harmonic density to provide clear details of harmonics.

Although this study reported an impressive classification accuracy rate of up to 99%, it was a preliminary investigation and did not account for potential interferences, such as those caused by multiple vessels and environmental noise. In real-world applications, the signal overlapping of multiple vessels can lead to signal pollution. The unavoidable environmental noise can significantly impact the classification accuracy by masking weak harmonics of valid signals. Therefore, further research is required to fully validate the use of ML for vessel classification. A comprehensive evaluation must consider various environmental factors and other potential interference. To address this challenge, signal processing techniques, such as beamforming, may be involved for pre-processing of the recorded data to focus on single vessel and enhance its SNR.

ACKNOWLEDGEMENTS

Many thanks to Jielai Zhang, Stuart Anstee and Matt Bigelow for the collection of the data used here. Also thanks Nicholas Padfield for his previous research which provided lots of useful experience for this project.

REFERENCES

Xuhao Du, Andrew Youssef, Yue Wang, Nicholas Padfield, David Matthews, 2018. “Detection of Snapping Shrimp Using Machine Learning”, *Australian Acoustical Society Conference Proceedings.*

Nicholas Padfield, 2019. “Exploring the classification of acoustic transients with machine learning”, *Australian Acoustical Society Conference Proceedings.*