Enhancing the Capacity of Detecting and Classifying Cavitation Noise Generated from Propeller Using the Convolution Neural Network

Hoang Nhat Bach[1](✉), Duc Van Nguyen[2], and Ha Le Vu[1]

[1] Institute of Electronics, Military Institute of Science and Technology, Hanoi, Vietnam
[2] Communication Engineering Departments, School of Electronics and Telecommunications, Hanoi University of Science and Technology, Hanoi, Vietnam

Abstract. One of the biggest concerns of underwater research is improving the ability to detect and classify sound sources. The Machine Learning and Deep Learning models often require a very large amount of data, while the data sources of the passive sonar system are limited; therefore, it is very important to pre-process data to improve data quality. This paper proposes a solution to improve the detection and classification of cavitation noise generated from propeller by improving the Detection of Envelope Modulation on Noise (DEMON) algorithm before using a modified Convolution Neural Network. The testing result shows that the accuracy of the modified model reaches around 90%, which is better than the results of existing methods, and it is prospectively developed and applied in practicalities.

Keywords: Passive sonar · DEMON · Convolution neuron network

1 Introduction

The role of marine control and defense for Vietnam, a country with 3,200-km coastline, is significantly important. The classification of underwater signals obtained from passive sonar systems is one of the challenges, due to the complex changes in time and spectral features in signals even from the same source. According to Nielsen [1] the typical noise sources for a ship include: noise from engines, machines and equipment on the ship while in motion (distant shipping), noise of hydrodynamic flows on the ship hull, propeller noise. Each type of signals has its own characteristics and can be detected by experienced surveyors by hearing or seeing the signal spectrum. During the movement, the main noise source of each ship is the cavitation of the propeller blades. The characteristics of this noise depend on the rotation frequency of the propeller blades, i.e. depend on frequency components that are varied by the speed of the blades. The cavitation noise increases proportionately with the speed of the blades and decreases as the depth increases. The repetition of such process produces vessel-specific features. Based on that

N.-S. Vo et al. (Eds.): INISCOM 2021, LNICST 379, pp. 268–275, 2021.
https://doi.org/10.1007/978-3-030-77424-0_22

linear relationship, the analysis of main frequency components will allow the calculation of the remaining parameters. The most popular and useful detection method is to use DEMON algorithm. DEMON algorithm was first proposed by Nielsen in 1991; since then, there have been many variations proposed to solve different specific problems, for example, tracking of multiple sources in a decoupled way [2], or 3/2D spectral analysis to extract propeller features from acoustic vector sensor data [3]. The basic DEMON algorithm has been tested in practice [4] and has also been used to detect the breathing pattern of divers from recorded data [5], etc. Based on the aforementioned research, we use a modified DEMON to analyse the propeller characteristic frequency components, and demonstrate the result under spectrogram (also called DEMONgram), which is a graphical representation of frequency in terms of time and magnitude. Spectrograms that possess characteristics of each object are fed into Convolution Neural Network (CNN) for analysis and processing.In recent years, Deep Learning (DL) has formed new breakthroughs; the DL model has the ability to process hidden features of the target signals through a multi-layer network. From the proposals of Fukushima (1980) [6] and LeCun (1989), the CNN completed in 2012 [7] was the first multi-layer structure using relative relationships in space to reduce the dimensions of parameters and improve training performance. LeNet [8] (1998) was the first network to apply 2-dimensional convolution. AlexNet [9] (2012) has broken the previous stereotype that learned features will not be as effecient as manually created features (through the SUFT, HOG, SHIFT algorithms). VGG-16 [10] (2014) formed a trend to improve the accuracy of DL networks by increasing the depth of the model. Variations of GoogleNet [11] (2014), by combining multiple filters of different sizes into the same block, produced the block architecture for the later CNN. ResNet-50 [12] (2015) used identified "short-cut" connection to map inputs from the previous layers to the following layers. It is a very deep network architecture, but has a smaller number of parameters, based on techniques from GoogleNet. DenseNet [13] (2016) is the next generation of ResNet which inherits the block architecture and develops the "short-cut" connection for a dense network.

The next parts of the paper will be organized as follows: part 2 will introduce the DEMON pre-processing method and its improvements, part 3 will analyze the CNN structure and the improved network model, and part 4 will be the conclusion.

2 Pre-processing by Modified DEMON

In the basic DEMON algorithm as defined by Nielsen, $x(t)$ is the acoustic signal that contains noise of the propeller and the environment, presented by:

$$x(t) = s(t) + n(t) \tag{1}$$

$$s(t) = m(f, t)\, w(t) \tag{2}$$

In which, $s(t)$ is a broadband signal formed by the modulation of a carrier waveform $\mathrm{w}(t)$ by a modulating waveform $m(f, t)$, while $n(t)$ is environmental noise. The modulating waveform $m(f, t)$ is periodic with frequency f, thus $m^2(f, t)$ is also periodic which can be expressed under a cosine formula as follows:

$$m^2(f,t) = \sum_{l=0}^{L} A_l \cos(lcft + l\theta) \tag{3}$$

Where $c = 2\pi/f_s$, f_s is the sampling frequency, A_l is the expansion coefficient of $m^2(t)$, θ is phase, and L is the number of coefficients. Because the square makes the left side of Eq. (3) always positive, the coefficient A_l must be selected to make the right side also positive (Fig. 1).

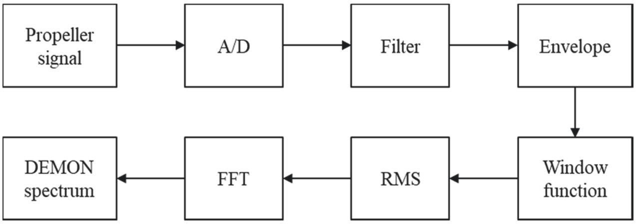

Fig. 1. DEMON algorithm

We propose the following solution: the signal spectrum is calculated by Short Time Fourier Transform (STFT). From that, we calculate a 2-dimensional spectral matrix, among which, one dimension is frequency, the other is the number of samples. The frequency amplitude of each segment is averaged to obtain a unique representative value. This technique divides the acoustic signal into consecutive overlapping segments. The result of this process is a set of filtered spectrogram images, which will be put into DL network for training. When the signal is unstable, the detection and classification accuracy will be reduced significantly. DL models can solve this problem more easily, because they extract hidden features using layers. On the other hand, as there are various types of noise, suitable selection of features plays an important role in guaranteeing the performance of the model. Thus, the result of modified DEMON reduces noise while retains sufficient features to increase detection accuracy. Our proposal can clearly separate characteristic frequencies and harmonics, as well as can decrease false alarm. Figure 3a, b are corresponding spectrograms of Fig. 2a, 2b.

In both methods, computation requires the definition of a target frequency window; unsuitable selection of input parameters can make the detection task unfeasible. Each sample is smoothed by window fuction, and the corresponding standard deviations are calculated. Signal is detected whenever the corresponding signal exceeds the corresponding detection threshold. Our simulation uses the dataset from the project: “An underwater vessel noise database” by Research center for Telecommunication Technologies – Universida de Vigo [14], as well as the dataset recorded by ourselves – Institute of Electronics, Military Institute of Science and Technology – in Lan Ha Bay, Hai Phong, Vietnam. Datasets include various types of underwater ship sounds. The sounds are recorded in shallow waters and in real conditions, which contain both natural and anthropogenic environment noise (Fig. 4).

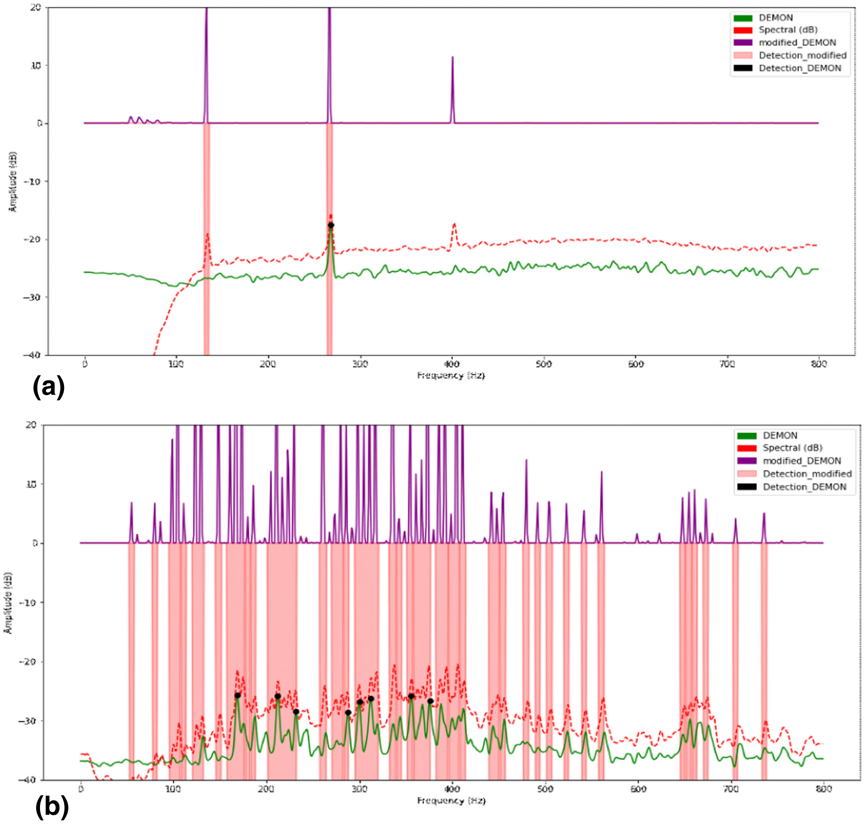

Fig. 2. (a) Comparision result between DEMON and Modified DEMON at Record-1 (b) Comparision result between DEMON and Modified DEMON at Record-2

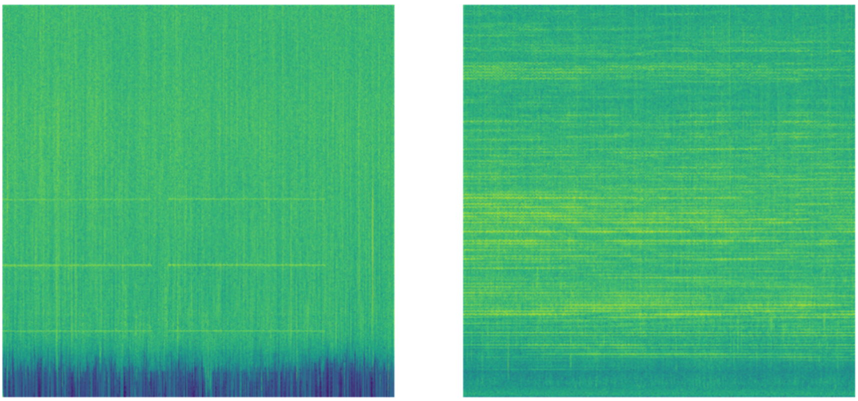

Fig. 3. a, b Spectogram after pre-processing with Modified DEMON

Fig. 4. Data recorded in Lan Ha Bay

Detection accuracy is calculated from the numbers and percentages of correct and incorrect ship detections. Table 1 shows two confusion matrices displaying the detection accuracy, and Table 2 summarizes the accuracy rates.

Table 1. Detection accuracy on a database of 3300 1-min audio samples

DEMON	Reality	No ship
	Ship	
Ship	1463	198
No ship	337	1302
Total samples	1800	1500
Modified DEMON	Reality	
	Ship	No ship
Ship	1768	45
No ship	32	1455
Total samples	1800	1500

Table 2. Accuracy rates and also the false-alarm rates

	DEMON (%)	Modified DEMON(%)
Detection accuracy	81.28%	98.22%
False alarm	13.2%	3%

3 CNN Comparision

We separate the samples into 70% for training set, 20% for validation set, and 10% for test set. We also use the spectrograms size of 3 × 224 × 224 to include in the CNN model for training. From analyzing results between the models, the accuracy of LeNet, AlexNet, VGG is only around 65–75% (Fig. 5).

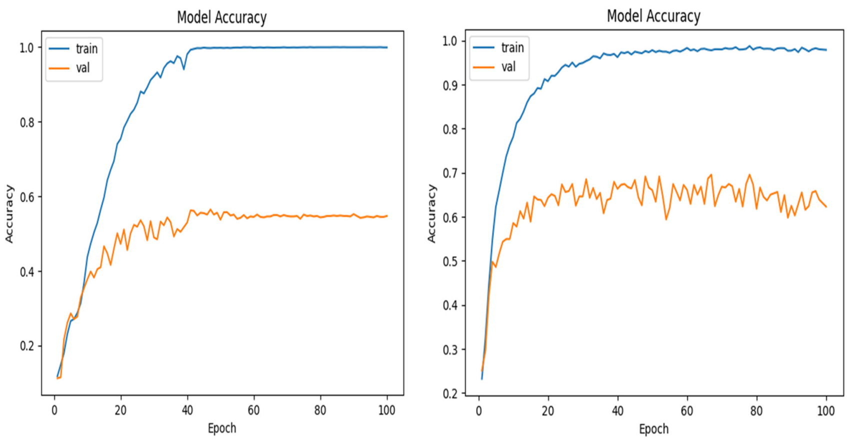

Fig. 5. a, b Training result our dataset with LeNet and VGG model

Proposed network model structure diagram (Fig. 6):

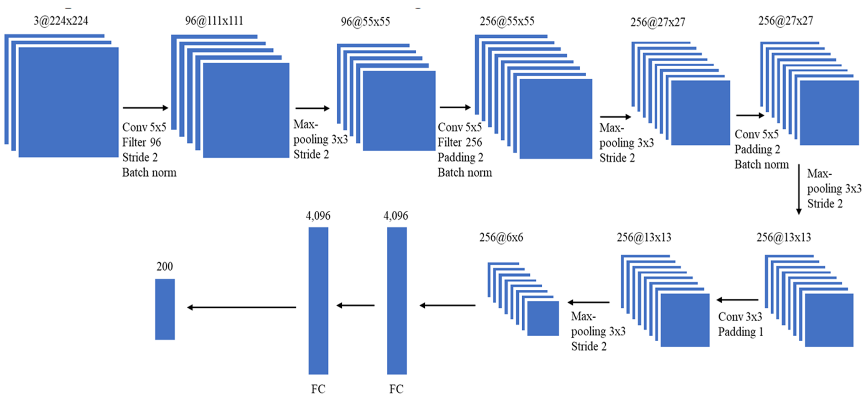

Fig. 6. Proposed convolution neural network architecture

The tuning is a challenge with a deep learning complicated structure. Because underwater datasets are insufficient, the deep model network is hard to be trained. Therefore, we propose a neural network using batch normalization with 1 input layer, 4 convolution layers, 4 maxpooling layers, and 2 fully connected layers. The batch normalization layers which are placed just after defining the sequential model and after the convolution layer will reduces the internal covariate shift of the model. The internal covariate shift is a change in the input distribution of an internal. The inputs received from the previous layer are always changed. Adding batch normalization layers ensure that the mean and standard deviation of the inputs will always remain the same, and minimize the fluctuation of the distribution. Batch norms don't compute the entire data, and the model's data distribution will make some noise. This can help overcome overfitting and help learn better. The first convolution layer has 1 convolution [5 × 5], the stride is 2, and 96 kernels. Using a smaller size of convolution matrix [5 × 5] will retain more information on the spectrogram. Filter size of the pooling layers is [3 × 3]; stride is 2. Extending the size of the convolution layers, reducing the dimensions of the feature map and making the filter size and stride smaller increase the accuracy of our model (Fig. 7).

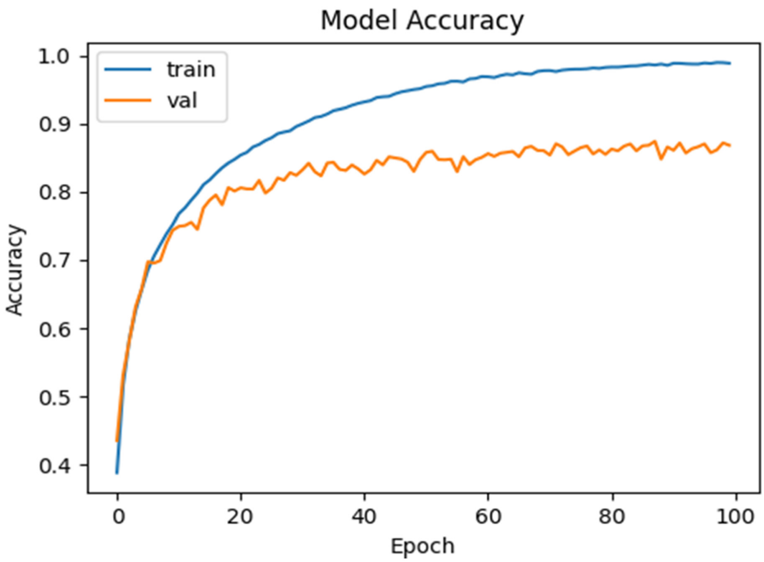

Fig. 7. Training result accuracy with proposed model

With our model, the accuracy of validation increases around 90%. After testing with the test set, the result shows that the accuracy has been greatly improved. It proves that the classification network can be further improved by the modification of the neural networks.

4 Conclusion

This paper describes a method for acoustic recognition of ships. It is a spectrogram domain analysis for passive sonar based on DEMON with a modified Convolution Neural Network which attains an accuracy percentage of 94.25%. The proposed model which is provided for cavitation noise from propeller, has a better performance in recognizing and preventing false detection. Based on the classification results, we conclude that: (1) deep learning models provide good results for detecting and classifying underwater and surface targets, and these models still process well in low SNR environments; (2) while DEMON algorithm focuses on fundamental frequency, our modified model additionally recognizes variations in the amplitude of fundamental frequencies; (3) the transformation of data from signal sequence to spectrogram enables the system to process a large amount of complicated data on a real-time basis; (4) datasets are still limited due to some security reasons. Therefore, preprocessing datasets and finding ways to increase the number of samples are the two main problems that shall be improved in the future.

Acknowledgment. This study was funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under the project number 102.04–2018.12.

References

1. Nielsen, R.O.: Sonar signal processing. Artech House Inc. (1991)
2. Fillinger, L., Sutin, A., Sedunov, A.: Acoustic ship signature measurements by cross-correlation method. J. Acoustical Soc. Am. **129**(2), 774–778 (2011)
3. Li, S.: DEMON feature extraction of acoustic vector signal based on 3/2-D spectrum. In: IEEE Conference on Industrial Electronics and Applications, pp. 2239–2243 (2007)
4. Tao, R., Feng, Y., Wang, Y.: Theoretical and experimental study of a signal feature extraction algorithm for measuring propeller acceleration in a port surveillance system. IET Radar Sonar Navig **5**(2), 172–181 (2011)
5. Lennartsson, R., Dalberg, E., Persson, L., Petrovic, S.: Passive acoustic detection and classification of divers in harbor environments. In: OCEANS 2009, Biloxi, pp. 1–7 (2009)
6. Fukushima, K.: Neocognitron: a hierarchical neural network capable of visual pattern recognition. Neural Netw. **1**, 119–130 (1988)
7. Alom, M.Z., et al.: A state-of-the-art survey on deep learning theory and architectures. Electronics **8**(3), 292 (2019)
8. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**, 2278–2324 (1998)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems NIPS, vol. 25 (2012)
10. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: arXiv:1409.1556 (2014)
11. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12, pp. 1–9 (2015)
12. 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, pp. 770–778 (2016)
13. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 4700–4708 (2017)
14. Research center for Telecommunication Technologies - Universida de Vigo homepage. http://atlanttic.uvigo.es/underwaternoise/