INTERNATIONAL JOURNAL OF
COASTAL, OFFSHORE & ENVIRONMENTAL ENGINEERING IJCOE Vol.8/No. 1/winter 2023 (10-15)

Available online at: www. Ijcoe.org

Underwater Ship-radiated Acoustic Noise Recognition Based on Mel-Spectrogram and Convolutional Neural Network

Mohammad Reza Khalilabadi

Faculty of Naval Aviation, Malek Ashtar University of Technology, Iran; khalilabadi@mut.ac.ir

ARTICLE INFO

Article History:
Received: 16 Jan. 2023
Accepted: 14 Feb. 2023

Keywords:
Underwater Acoustic
Deep Learning
Recognition
Noise
CNN

ABSTRACT

The abstract should include the One of the most exciting topics for researchers over the past few years is detecting underwater acoustic noises. Meanwhile, the complicated nature of the ocean makes this task very challenging. Also, making signals formatted data compatible with machine learning approaches needs much knowledge in signal processing for feature detection. This paper proposed a method to overcome these challenges, which extracts features with Convolutional Neural Network (CNN) and Mel-spectrogram (converting signal data to images). This method needless knowledge in signal processing and more knowledge in machine learning; because using CNNs find the hidden pattern and knowledge of the data automatically. The proposed approach detected the presence of the ships and categorized them into different kinds of them with 99% accuracy that is a noticeable improvement considering state of the art. The performed CNN models consist of 2 CNN layers for feature extraction and a Dense layer for classification the underwater ship noises.

1. Introduction

Underwater acoustic noise is one of the most attractive and widely used topics in acoustic oceanography. Several researches have been done by the author in the field of oceanography (e.g. [1]–[12]), and especially acoustic oceanography([13]-[14]). Recently, this path of development has progressed towards the use of artificial intelligence in predicting the environmental characteristics of the ocean[15].

Detecting underwater noises played an essential role in many oceanographic applications like navigation systems and defense systems.

The source of the underwater noises can be categorized as human maids like ships, drilling rigs, and aircraft; the other category is natural noises like raining noise, marine mammal's noise, and wind. There are many different radiated noise signals in the oceans, making the detection task quite tricky.

The most challenging part of this challenge is extracting efficient features that maid ship radiated noise feature extraction an active area in many years. In all of the research, three different feature extraction strategies alongside a classification model were used for this task.

The hand-designed features that need prior knowledge of data set and expertness in signal processing engineering are the first to extract features. For example, using spectral and cepstral, which used by Das et al. [16] and Santos-Domínguez et al.[17], Jain et al., used line spectrum density for feature extraction and support vector machine for classification. Wei et al. [18] used features based on wavelet packets. Yang et al. [19] dissimilarity-based evaluation-based features for classification. Zhang et al. [20] used Mel-Spectrum coefficient features for ship radiated recognition. Wei et al., used $1_{1/2}$D spectrum-based spectrum features for ship radiated noise classification. First-order alongside second-order differential MFFC were used by Zhang et al.[20] for underwater signal recognitions. However, these approaches suffer from low detection rates in noisy, shallow seas and unavailability to find time-series and complicated hidden patterns of data and need domain skill in signal engineering.

The second strategy for extracting the features from underwater noise signals is using automatically generated features like Shen et al., that used one dimensional CNN for extracting features directly from spectrums[21]. Cao et al. used a stack auto-encoder that extracted hidden and complicated information and patterns from data[22]. Yang et al. used Deep Belief Network

(DBN) for feature extraction[19]. Finally, yuan et al. used a multimodal Deep Learning approach for ship radiate noise recognition[23].
The 3rd strategy is the combination of the two first strategies, which means using the hand-designed features alongside the automatically generated features or feeding the hand-designed features to a machine learning feature extractor model. For example, Ke et al., used Restricted Boltzmann Machine (RBM) to generate automatic features and used these features alongside wavelet, spectral, etc.[24]. Finally, they feed these two kinds of features to two-dimensional feature fusion.
In this paper, we proposed a model for underwater ship radiated noise recognition based on 3rd strategy, which means we are not using the original spectrum as input of CNN like Shen et al., but using a pre-processing approach for transforming the spectrums into images named Mel-Spectrogram and then using Deep Neural Network with two, two dimensional CNN layers and one dense layer to find the hidden pattern of data and then detect the label[21]. Our proposed approach performed better or at least competitive to under introduced approaches. Also, using Mel-Spectrogram makes the model converge very fast (30 epochs).

2. Materials and Methods

In this section, an explanation of the proposed methodology for ship radiated noise has been provided.
In this research, we used Mel-Spectrogram as a preprocessing stage to make the spectrums compatible as the input of the 2-dimensional CNN. A Mel-spectrogram is a spectrogram where the frequencies are converted to the Mel scale.
In this method, the signal is first converted from time domain to frequency domain by using of FFT (Fast Fourier Transform). Then, the frequency converts to the mel scale.

After that, we applied power-to-DB scaling to convert a power spectrogram (amplitude squared) to decibel (dB) units. This algorithm is calculated by equation 1.

$$db_{units} = 10\log(\text{spectrogram}) \qquad (1)$$

This algorithm makes the spectrum more visible. the effect of power-to-DB scaling before and after applying this algorithm can be followed in the research of Lavanya et al. [25]. After that, a deep CNN model will be applied to the model for prediction. CNN is a class of deep neural networks, most commonly applied to analyze visual images[26]. Typical of a simple neural network, a CNN consists of three input, output, and hidden layers. Any neural network layers between input and output are named hidden because their inputs and outputs are transformed by the activation function, feature mapping, and convolution.
In a CNN, the hidden layers include layers that performed a transformation named convolutions. That is a layer that performs a feature mapping using the dot product of the convolution kernel with the layer's input.
As the kernel slides along the input matrix for the layer, a feature map will be generated using convolutional operation, which is devoted to the input of the next layer. The convolutional layers are the layers with the input shape of the number of inputs, input height, input width, and input channels. After input data passed through a convolutional layer, it is transformed to an abstracted feature map that is also named an activation map with a shape with axis similar to input but different dimensions.
A convolutional layer contains different hyper parameters which are:

1. Convolutional kernels: extract features from input data that is moved across the image and multiplied with the image.
2. The number of input and output channels.
3. Padding: defined the padding strategy to apply in the input like Valid Padding and Same Padding.
4. Stride: filters that modify the amount of movement over the image

Convolutional layers convolve the input and pass its result to the next layer. This is similar to the response of a neuron in the visual cortex to a specific stimulus[27].
The advantage of convolution is reducing the number of free parameters and allowing the network to be deeper[28].
The designed model includes four convolutional, two max-pooling, one Flatten, one Dropout, and one dense layer. Figure 1, shows the structure of the proposed model.

2.1. Experimental Dataset

All data of ship-radiated noise used in this paper come from a database called Ships-Ear[17].

During 2012 and 2013, sounds of many different ships were recorded on the Spanish Atlantic coast and included in the Ships-Ear database(available at http://atlanttic.uvigo.es/underwaternoise/).

The recordings were made with autonomous acoustic digitally SR-1 recorders manufactured by Mar Sensing LDA (Faro, Portugal). According to Santos-Domínguez et al. [17], 11 vessel types are merged into four practical classes (based on vessel size) and one background noise class, as shown in Table 1. Some ship pictures and ship-radiated noise of each class are demonstrated in figure 2.

Table 1. Ships Grouping

Class	Type
Class A	fishing boats, trawlers, mussel boats, tugboats, dredgers
Class B	motorboats, pilot boats, sailboats
Class C	passenger ferries
Class D	ocean liners and ro-ro vessels
Class E	background noise recordings

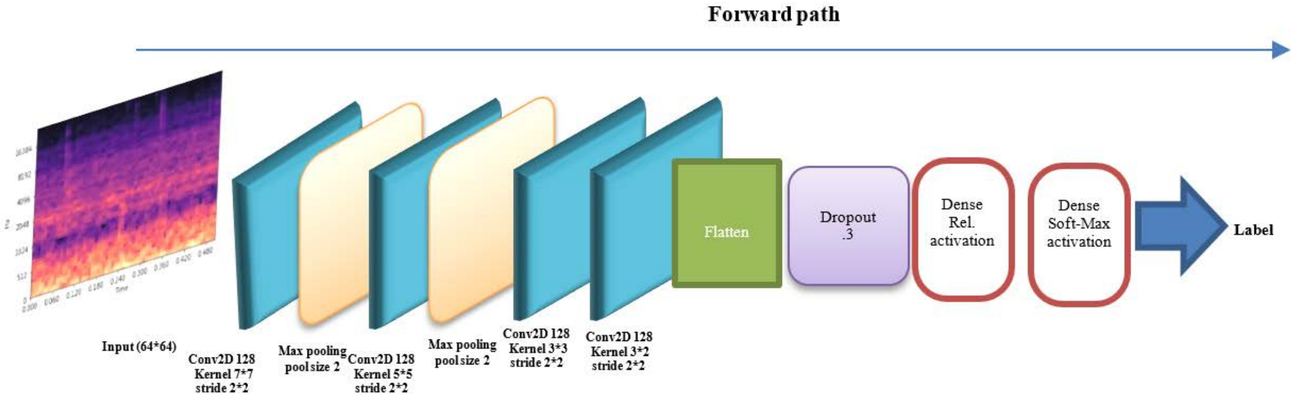
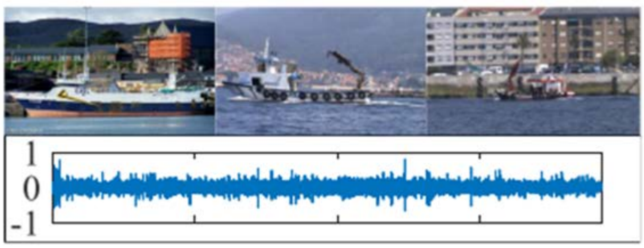
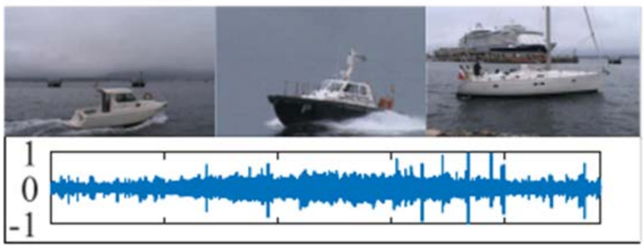


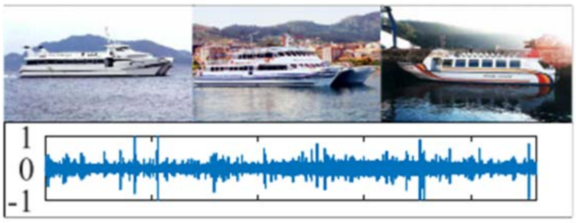
Figure 1. Proposed CNN Structure



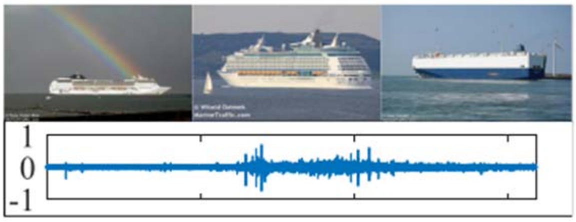
(a)



(b)



(c)



(d)

Figure 2. Ship pictures and ship-radiated noise of each class. (a) class A. (b) class B. (c) class C. (d) class D

2.2. Experimental Setup

All the codes are performed using google collaborator with python language and GPU enable Tensor Flow.
The hyper parameter of the framing and spectrogram generation have been provided in table 2.
All of these hyper-parameters have been calculated using cross-validation strategy. Also, we randomly select .8 of data as training set and .2 data as the testing set.

Table 2. Hyper-Parameters of the Pre-Processing Stages

Stage	Hyper parameters	
Loading data	Sampling rate	52,734
Framing	frame length	26,367
	hop length	5,273 (.8 overlapping)
Mel spectrogram	Sampling rate	52734
	n_mels	64
	hop length	415
	n_fft	1280

2.3 Evaluation Metrics

Our approach's performance was evaluated using the following metrics:

1. Accuracy rate (AC): Percentage of correctly classified records overall records Formula.

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2)$$

2. Precision rate (P): Percentage of correctly classified anomaly records over a total of classified anomalies[29].

$$P = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

3. Recall (R): Percentage of correctly classified anomalies divided by the number of attack entries[29].

$$R = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

4. F-Measure (F1-Score): Harmonic mean of precision and recall, which is considered.

$$F = \frac{2 \times P \times R}{P + R} \times 100\% \quad (5)$$

4. Results and Discussion

In this section, the evaluation results considering the mentioned evaluation metrics have been provided.
As a result, we achieved a noticeable accuracy in ship-radiated noise recognition (99.20%), which is quite an outstanding achievement.
Table 3 shows the classification report of the proposed approach on different class labels, which shows 99% precision, 99.25% recall, and 99.12% f1-score.

Also, Table 4 shows the confusion matrix of the proposed approach that shows how many instances of each class have been predicted correctly and, in the case of incorrect classification, what class it was detected.

Table 3. Classification report of proposed approaches in different classes

Class label	Precision%	recall%	f1-score%	Accuracy%
A	98.91	98.72	98.81	-
B	97.31	99.09	98.19	-
C	99.67	98.98	99.33	-
D	99.69	99.73	99.71	-
E	99.42	99.73	99.57	-
macro Average	99.00	99.25	99.12	99.20

Also, for showing that our model training convergence process and the effect of the process on training and testing data figure 3 shows the accuracy improvement during testing. Figure 4 shows the loss changing during training.

Table 4. Confusion matrix of the proposed model

Predicted label / True Label	A	B	C	D	E
A	3460	34	10	1	0
B	5	3033	14	0	9
C	22	48	8469	11	6
D	11	0	2	4885	0
E	0	2	2	3	2564

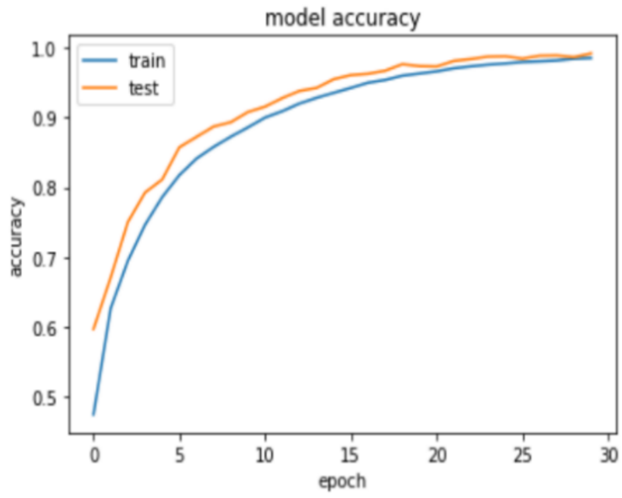


Figure 3. Changing of accuracy in training set and testing set during training

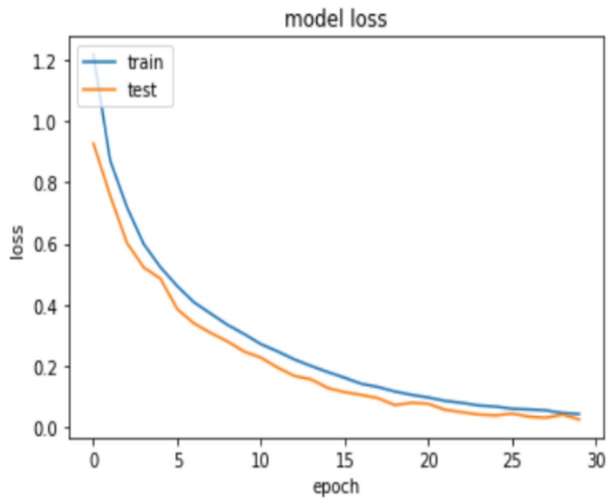


Figure 4. Changing of loss in the training set and testing set during training

3. CONCLUSION

In this research, we proposed a ship radiated noise recognition model based on Mel-Spectrogram and Convolutional Neural network. As an essential stage of studies in the same field is to provide features for the machine learning model, we used the Mel-Spectrogram for this process and fed the 64*64 generated images to the CNN. The result shows our proposed approach achieved a significant result considering accuracy, precision, recall, and f1-score.

References

[1] M. R. Khalilabadi, "Underwater Terrain and Gravity aided inertial navigation based on Kalman filter," *Int. J. Coast. Offshore Eng.*, vol. 5, no. 3, pp. 15–21, 2022.

[2] M. R. Khalilabadi, "The effect of meteorological events on sea surface height variations along the

northwestern Persian Gulf.," *Curr. Sci. 00113891*, vol. 110, no. 11, 2016.

[3] M. R. Khalilabadi, " 3D modeling of Circulation in the Oman Sea Using the MITgcm Model," *Hydrophysics*, vol. 2, no. 1, pp. 61–68, 2016.

[4] M. R. Khalilabadi, "Tide–surge interaction in the Persian Gulf, Strait of Hormuz and the Gulf of Oman," *Weather*, vol. 71, no. 10, pp. 256–261, 2016.

[5] M. R. Khalilabadi and M. Akbari Nasab, "Study of static stability and double diffusion in the Oman Sea," *Iran. J. Mar. Sci. Technol.*, vol. 18, no. 71, pp. 11–19, 2014.

[6] M. R. Khalilabadi and B. Behrooz, "Simulation of volumetric flow at the Arvandrood estuary using the MIKE21 model," *Hydrophysics*, vol. 5, no. 2, pp. 1–10, 2020.

[7] M. R. Khalilabadi and B. S. H. HASSANTABAR, "Investigation of magnetic field fluctuations due to sea waves in the Strait of Hormuz," 2016.

[8] M. R. Khalilabadi and D. Mansouri, "Effect of super cyclone 'GONU' on sea level variation along Iranian coastlines," 2013.

[9] M. R. Khalilabadi and H. Shahmirzaee, "Marine Magnetic Data Processing and Extracting Magnetic Anomaly," *Hydrophysics*, vol. 3, no. 1, pp. 1–10, 2017.

[10] M. R. Khalilabadi, M. Sadrinassab, V. Chegini, and M. Akbarinassab, "Internal Wave Generation in the Gulf of Oman (Outflow of Persian Gulf)," 2015.

[11] M. Akbarinasab, M. R. Khalilabadi, and M. M. Moghadam, "Impact of relative vorticity on sea level fluctuations on the South coast of Caspian Sea.," *Casp. J. Appl. Sci. Res.*, vol. 5, no. 4, 2016.

[12] A. Ghorbani and M. R. Khalilabadi, "Positioning Using Classification and Regression: Case study of Oman Sea," *Int. J. Coast. Offshore Eng.*, vol. 4, no. 3, pp. 35–41, 2020.

[13] S. H. Hosseini, M. Akbarinasab, and M. R. Khalilabadi, "Numerical simulation of the effect internal tide on the propagation sound in the Oman Sea," 2018.

[14] M. R. Khalilabadi, "2D Modeling of Wave Propagation in Shallow Water by the Method of Characteristics," *Arch. Acoust.*, vol. 47, no. 3, pp. 407–412, 2022.

[15] M. R. Khalilabadi, "An autonomous location prediction model for maritime transport applications: a case study of Persian Gulf," *Ships Offshore Struct.*, pp. 1–8, 2022.

[16] A. Das, A. Kumar, and R. Bahl, "Marine vessel classification based on passive sonar data: the cepstrum-based approach," *IET Radar Sonar Navig.*, vol. 7, no. 1, pp. 87–93, 2013.

[17] D. Santos-Domínguez, S. Torres-Guijarro, A. Cardenal-López, and A. Pena-Gimenez, "ShipsEar: An underwater vessel noise database," *Appl. Acoust.*, vol. 113, pp. 64–69, 2016.

[18] X. Wei, L. I. Gang-Hu, and Z. Q. Wang, "Underwater target recognition based on wavelet packet and principal component analysis," *Comput Simul*, vol. 28, pp. 8–290, 2011.

[19] H. Yang, S. Shen, X. Yao, M. Sheng, and C. Wang, "Competitive deep-belief networks for underwater acoustic target recognition," *Sensors*, vol. 18, no. 4, p. 952, 2018.

[20] L. Zhang, D. Wu, X. Han, and Z. Zhu, "Feature extraction of underwater target signal using mel frequency cepstrum coefficients based on acoustic vector sensor," *J. Sens.*, vol. 2016, 2016.

[21] S. Shen, H. Yang, and J. Li, "Improved auditory inspired convolutional neural networks for ship type classification," in *OCEANS 2019-Marseille*, 2019, pp. 1–4.

[22] X. Cao, X. Zhang, Y. Yu, and L. Niu, "Deep learning-based recognition of underwater target," in *2016 IEEE International Conference on Digital Signal Processing (DSP)*, 2016, pp. 89–93.

[23] F. Yuan, X. Ke, and E. Cheng, "Joint representation and recognition for ship-radiated noise based on multimodal deep learning," *J. Mar. Sci. Eng.*, vol. 7, no. 11, p. 380, 2019.

[24] X. Ke, F. Yuan, and E. Cheng, "Integrated optimization of underwater acoustic ship-radiated noise recognition based on two-dimensional feature fusion," *Appl. Acoust.*, vol. 159, p. 107057, 2020.

[25] K. Lavanya, C. L. Devi, M. D. Sree, and P. N. Shareef, "Predicting the Emotions Based on Emoji's and Speech Using Machine Learning Techniques," 2021.

[26] M. V. Valueva, N. N. Nagornov, P. A. Lyakhov, G. V. Valuev, and N. I. Chervyakov, "Application of the residue number system to reduce hardware costs of the convolutional neural network implementation," *Math. Comput. Simul.* vol. 177, pp. 232–243, 2020.

[27] I. M. Dheir, A. S. A. Mettleq, A. A. Elsharif, and S. S. Abu-Naser, "Classifying nuts types using convolutional neural network," *Int. J. Acad. Inf. Syst. Res. IJAISR*, vol. 3, no. 12, 2020.

[28] H. H. Aghdam and E. J. Heravi, "Guide to convolutional neural networks," *N. Y. NY Springer*, vol. 10, no. 978–973, p. 51, 2017.

[29] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *ArXiv Prepr. ArXiv201016061*, 2020.