www.dergipark.gov.tr
ISSN:2148-3736

El-Cezerî Journal of Science and Engineering
Vol: 10, No: 1, 2023 (57-65)
DOI :10.31202/ecjse.1149363

ECJSE

Research Paper

Ship Type Recognition using Deep Learning with FFT Spectrums of Audio Signals

Mustafa Eren YILDIRIM

Electrical and Electronics Engineering Department, Bahcesehir University, Istanbul, Turkey
mustafaeren.yildirim@eng.bau.edu.tr
Department of Electronics Engineering, Kyungsung University, Busan, Republic of Korea
meyeren3@ks.ac.kr

Received: 27.07.2022 **Accepted:** 20.01.2023

Abstract: Ship type recognition has gained serious interest in applications required in the maritime sector. A large amount of the studies in literature focused on the use of images taken by shore cameras, radar images, and audio features. In the case of image-based recognition, a very large number and variety of ship images must be collected. In the case of audio-based recognition, systems may suffer from the background noise. In this study, we present a method, which uses the frequency domain characteristics with an image-based deep learning network. The method computes the fast Fourier transform of sound records of ships and generates the frequency vs magnitude graphs as images. Next, the images are given into the ResNet50 network for classification. A public dataset with nine different ship types is used to test the performance of the proposed method. According to the results, we obtained a 99% accuracy rate.

Keywords: Signal processing, ship type recognition, ResNet50, MFCC

1. Introduction

Recognition of the existence and type of a ship is a very critical task in maritime surveillance, security, traffic control, and prevention of illegal fishing in restricted territories and periods. Detecting an unauthorized ship carrying hazardous loads can contribute to the protection of the environment, economy, and public health. Moreover, the classification of ship types in advance of their entrance to a sensitive or crucial territory helps the authorities to take necessary precautions to prevent possible accidents, damages or to manage the traffic. Automatic identification systems (AIS) and vessel traffic service (VTS) are used for gathering information about the vessels [1]. Both systems require personnel on the ship and administration sides, in which the ship officer reports to the administration office about the ship such as the weight, load type, direction, speed, and destination. In locations where the marine traffic load is heavy, these systems are feasible and efficient. Thus, more intelligent and computation light systems are at the focus of industry and academy.

Depending on the feature type they use, ship type recognition (STR) systems is divided into two major groups. These are audio-based systems and image-based systems. While some of the image-based systems operate on conventional visual features such as colour, contour, and shape, other systems use deep learning (DL) for feature extraction. In [2], the authors introduced an STR system on coastal radar images. In another study [3], researchers proposed a dual-stage feature selection method to conduct recognition among three types (cargo, fisher, and tanker). Authors in [4] used colour information in high-resolution satellite images for ship detection and recognition on the Spanish coast. In a similar study [5], authors used the geometric features extracted from synthetic

How to cite this article
Yıldırım, M. E., "*Ship Type Recognition using Deep Learning with FFT Spectrums of Audio Signals*" El-Cezerî Journal of Science and Engineering, 2023, 10(1); 57-65.
ORCID ID: 0000-0002-0662-2770

images of the Google 3D Warehouse dataset for testing on real images of ships. In [6], a model named BDA-KELM consisted of kernel extreme machine (KELM) and dragonfly algorithm in binary space (BDA) to select the features and optimal parameters automatically. They made classification among the types of bulk carrier, container ship, and oil tanker.

Recently, DL based methods are used also in maritime applications. Authors used it to recognize icebergs and ships by using synthetic aperture radar (SAR) images [7]. There are various studies in literature related to shipping detection and ship type recognition [8,1,9]. A study [8] presented a convolutional neural network (CNN) model for ship detection and classification on images. They achieved 95% accuracy for ship type classification. In [1], the authors introduced a coarse-to-fine cascaded CNN model for ship type recognition. A similar study [9] presented a CNN model for the recognition of commercial ships from satellite images. They obtained 90% recognition accuracy. Authors in [10] presented a rotation-based model for the recognition of three ship categories. They obtained 87.1% and 74.2% for detection and classification respectively. Authors in [11] used the you only look once (YOLO) network in their method for ship behaviour analysis. In [12], authors merged Zernike moment and CNN for recognition of three different types. Authors implemented multiple CNN architectures for different vessel types such as cruise, boats, war ships, cargo and achieved recognition accuracy over 90% in overall [13]. For DL based methods to perform well, a large dataset with many variations such as rotation, scale, translation, colour deformations of ship images must be used. Therefore, these methods may suffer in case of an insufficient dataset.

The other major group of systems for ship type recognition are audio-based. Authors in [14] introduced a model for ship type recognition by using raw acoustic data over five ship types and obtained a 79.2% recognition rate. Mel frequency cepstrum coefficients (MFCC) are commonly used in ship detection and classification [15]. Authors used the local binary pattern (LBP) for classification and investigated the accuracy with classifiers such as support vector machines (SVM), decision tree (DT), k-nearest neighbourhood (kNN), and linear discriminant analysis (LDA) and obtained 97.5% accuracy at most [16].

In this paper, we propose the use of frequency behaviour along with a DL model for ship type recognition. Fast Fourier transform (FFT) is applied to sound recordings of ships. The frequency vs magnitude plots are given into the ResNet50 model for recognition. The proposed method is tested on a public dataset and it gives a high recognition rate. This paper is organized as follows: In section 2, the proposed method is presented. In section 3, used dataset, experiments and discussions are given. The conclusion is presented in section 4.

2. Proposed Method

The proposed method has several steps. Figure 1 shows the block diagram of the proposed method. Initially, FFT is applied to the sound files in the original dataset to obtain the frequency magnitude plots. A new image dataset is generated from these plot images. The recognition is performed on the generated image dataset by using the ResNet50 network.

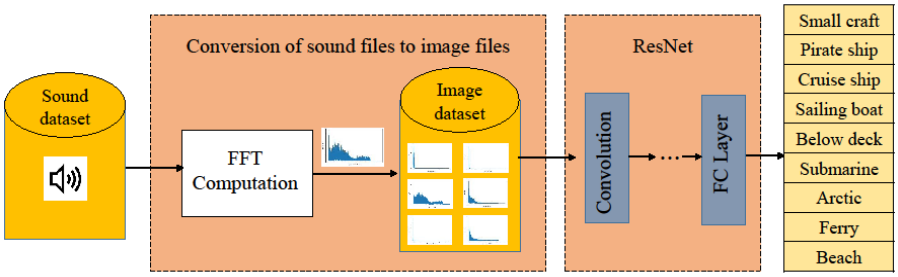

Figure 1. Block diagram of the proposed method

2.1. Fast Fourier Transform (FFT)

The first stage of the proposed method is to apply FFT to each ship's sound recording file. FFT is a popular algorithm in the signal processing field. Rather than the information, which we can gather from time-domain analysis, FFT supplies frequency or spectral based information about the audio signals. FFT implies that any continuous signal can be expressed in terms of the sum of delicately chosen sinusoidal waves with appropriate frequency, amplitude, and phase [17]. The Fourier transform is given in (1).

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-jwt} dt \tag{1}$$

Above expression can be written as a finite sum as below;

$$X(\omega) = \sum_{I=0}^{N-1} x_i(i\Delta t) e^{-j\omega(i\Delta t)} \Delta t \tag{2}$$

where N is the number of data points and $x_i(i\Delta t)$ is the data at i sampled at time $i\Delta t$. We can change the continuous variable ω, to a discrete number of samples by using (3).

$$\omega_k = (2\pi k)/(N\Delta t) \tag{3}$$

We can rewrite (2) which is the amplitude spectrum of the signal as;

$$\frac{X(\omega_k)}{\Delta} = \sum_{i=0}^{N-1} x_i(i\Delta t) e^{-j2\pi ik/N} = C(k) \tag{4}$$

where k is in $[0, N-1]$. $C(k)$ can be written in terms of sinusoidal functions as shown in (5).

$$C(k) = \sum_{i=0}^{N-1} x_i(i\Delta t) \left[\cos\frac{2\pi ik}{N} - j\sin\frac{2\pi ik}{N}\right] \tag{5}$$

In (5), every value of $C(k)$ is a complex number. The total array with N complex numbers of this series is called the discrete Fourier transform (DFT). The computational load of DFT is very high. In case of a signal with a length of N, the complexity of DFT is $O(N^2)$. The main objective of FFT is that it decreases the computational load of DFT. FFT operates in a divide and conquer manner, where it divides the whole signal into smaller sequences, computes their DFT, and merge them. The complexity of FFT on the same signal would be $O(N \log N)$. In case of a large amount of data, the application of FFT makes a huge difference in processing time.

In this study, we used the scipy and librosa libraries of Python 3.6 for audio signal processing and image generation. The signals are sampled at 22,050 samples per second. We did not perform any preprocessing on the original data to make the tests more realistic.

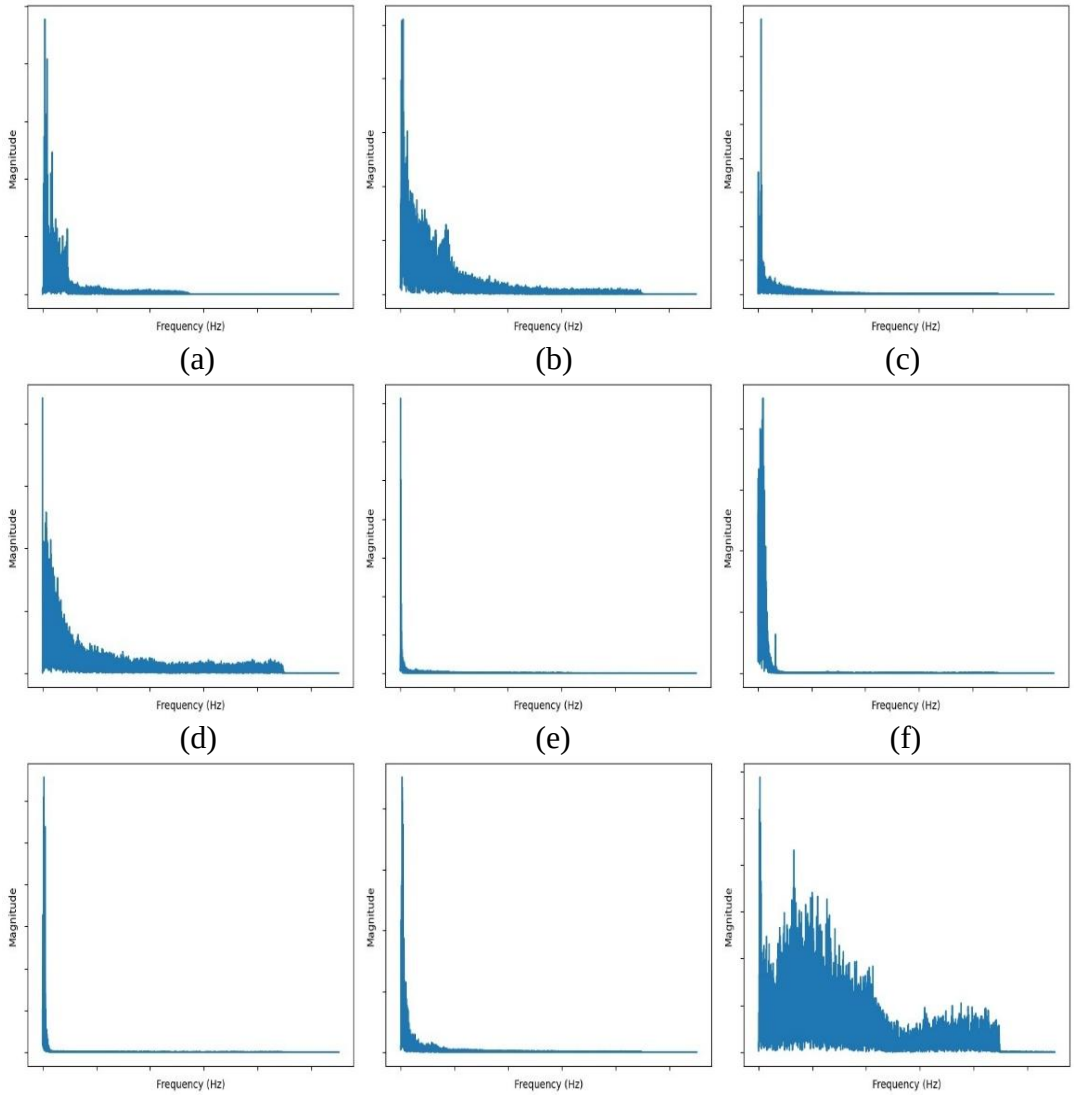

Figure 2. Frequency-magnitude graph of a sample from classes: (a) small craft, (b) pirate ship, (c) cruise, (d) sailing boat, (e) below-deck boat, (f) submarine, (g) arctic, (h) ferry and (i) beach

Figure 2 shows the resulting frequency vs magnitude plots of each ship type in the dataset. The frequency characteristics of different ship types differ from each other. However, some ship types such as below-deck boat and arctic boat are very similar to each other whose signal powers are concentrated to the very low-band. The pirate ship and sailing boat signal powers are distributed to a low and middle band of the spectrum. Ferry and cruise also show similar frequency behavior since they are similar types. The last class, which is the sound recording of the background beach, has the widest spectrum. The reason is that it has multiple sources of sound rather than a single type of vessel.

2.2. ResNet50 Model

DL models such as Alexnet [18], GoogleNet [19], and VGG [20] are widely used in studies. Beginning with AlexNet, the depth of the DL architectures have been increasing. Nevertheless, a deeper network does not necessarily lead to better training. Moreover, the deeper networks are more difficult to train since the gradient is backpropagated to previous layers and it may end up with an extremely small value. This phenomenon is called the vanishing gradient problem and affects the

performance of a deep network in a negative way. Residual networks (ResNet) brought a solution to these drawbacks by using skip connections, which are acting as gradient superhighways [21]. These superhighways lead the gradient to flow unhindered. ResNet50 [21] has 50 layers in its deep network. In addition to having a deep network for accurate training, it also has a much smaller number of training parameters compared to other deep networks [22]. Figure 3 shows the used network in this study. Although our images have the size of 640×480, they are resized to 224×224 before the first convolution stage.

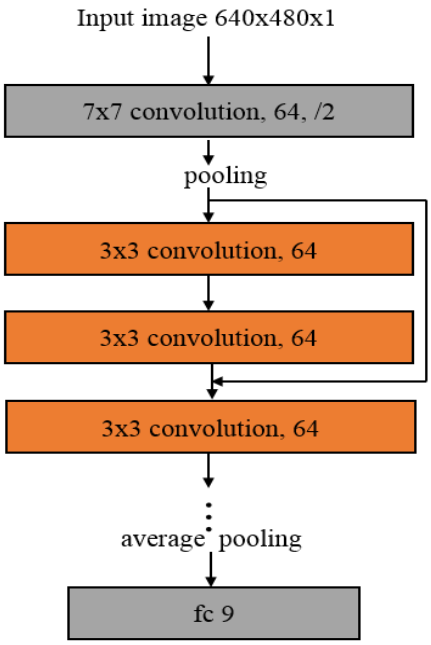

Figure 3. ResNet50 architecture

3. Experiments and Results

3.1. Dataset

Our experiments are performed on a publicly available dataset [16]. It contains sound recordings of nine different classes including eight ships and one background with 1,025 files in wav file format. Each recording has 2 seconds duration without repeating itself. The details of the dataset are shown in Table 1. It is a balanced dataset except for the small craft class, in which there is a fewer number of samples than in the other classes.

Table 1. Types and number of ship sound samples in the dataset.

Ship Type	Number of Samples
Small craft	77
Pirate ship	117
Cruise ship	119
Sailing ship	117
Below-deck sailboat	119
Submarine	116
Arctic	120
Ferry	122
Beach	118
Total	1025

3.2. Implementation details and training

We applied the proposed model to recognize the ship types in the above-mentioned dataset. Implementation was done using Tensorflow [21], Keras [21], Python 3.0 in Ubuntu OS 16.04 computer with Intel Xeon E5-2650 v3 CPU, NVIDIA Quadro M5000 dual-GPU, and 64GB RAM. We used stochastic gradient descent (SGD) optimizer with a mini-batch size of 12, a learning rate of 0.01, and 30 epochs. 20% of the dataset was used for validation and 20% was used for testing. Figure 4 shows the accuracy change during the training of the model. The model had 23,601,000 trainable parameters.

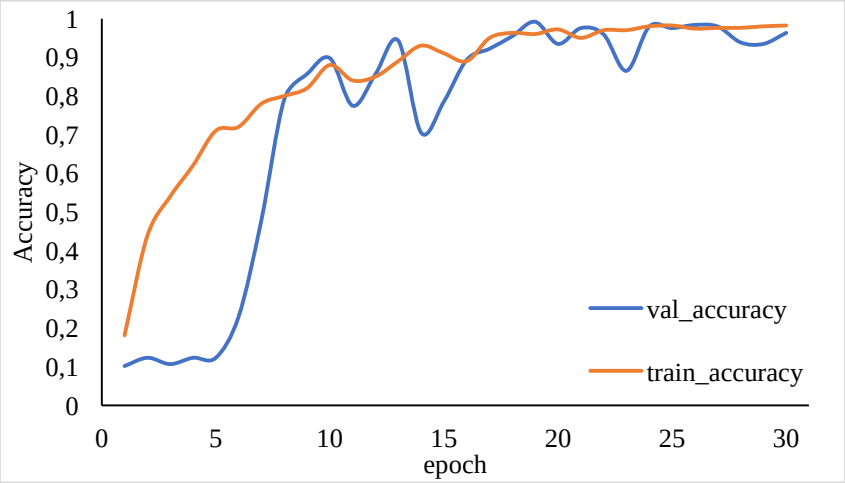

Figure 4. The model's accuracy through the epochs during the training

Although the validation accuracy reached over 94% after 13 epochs, it became stable above 90% after epoch number 17. At the end of the training, the accuracy for validation and training was 96.31% and 98.24%, respectively. Figure 5 shows the loss change during the training stage.

It can be seen that the losses in both validation and training decrease dramatically. At the end of the training, the loss for validation and training was 0.08 and 0.068 respectively. Even though it might be possible to decrease the loss more, it would cause the model to overfit. It would decrease the accuracy of the model.

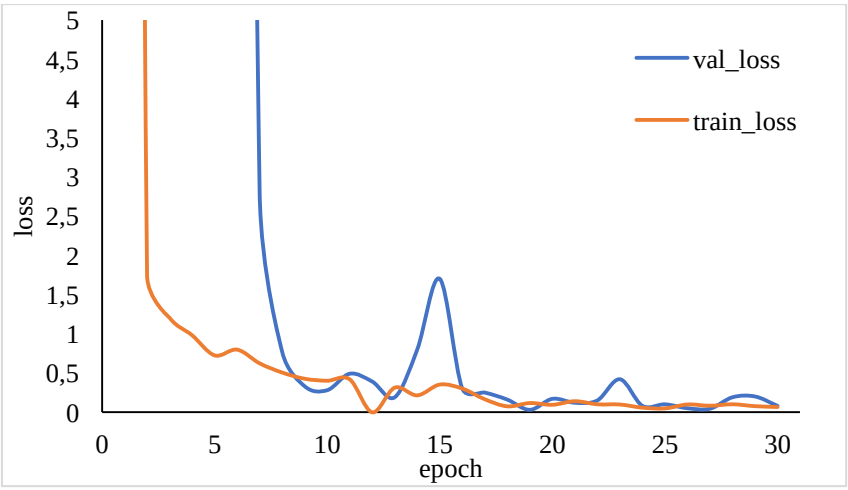

Figure 5. The model's loss through the epochs during the training

3.3. Performance of the proposed method

The performance of the proposed method is shown in Figure 5 based on the confusion matrix. Briefly, a confusion matrix evaluates the quality of the predictions of a classifier on a given dataset. The diagonal units mean the number of points for which the estimated labels represent true positives, while the off-diagonal elements are mislabeled by the classifier. The values in Figure 6 are normalized between zero and one. Thus, the closer values to one in the diagonal values of the confusion matrix are, the better the accuracy is. The proposed method misclassified sailing boat as a small craft and ferry as the arctic. The remaining ships were accurately classified.

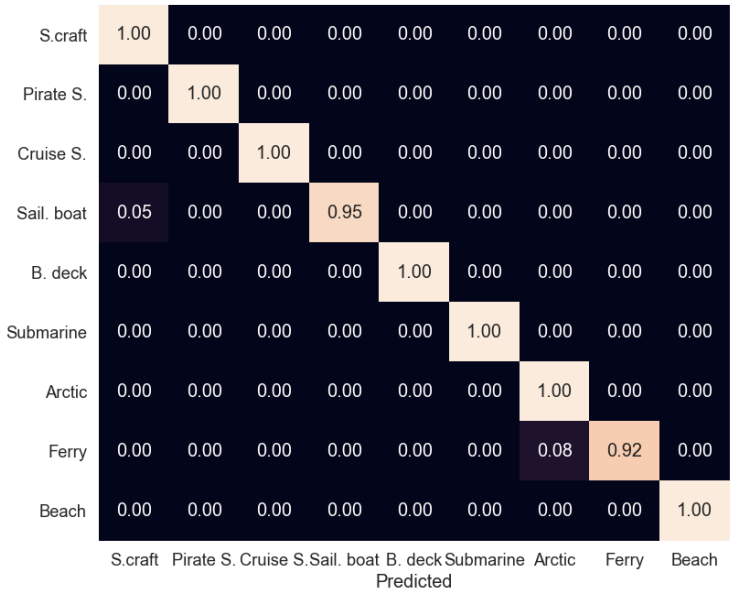

Figure 6. Confusion matrix of all classes in the dataset

The proposed method was compared with the results, which the authors of the dataset obtained, and with the statistical MFCC parameters. It is difficult to add more studies for the benchmark since the source codes or used datasets are not available. The accuracy results are given in Table 2. The proposed method outperforms the other studies by a recognition rate of 99%. The closest score is achieved by random forest base recognition using the mode of the MFCC coefficients during the sequence. Mode operation refers to the use of the most frequent value for each coefficient along the sequence, thus it eliminates the infrequent values that do not reflect the characteristic behavior.

Table 2. Recognition accuracy of the proposed method and other studies on the used dataset.

Method	**Accuracy (%)**
Decision tree [16]	95.1
kNN [16]	95.8
Linear discriminant[16]	94.9
SVM[16]	97.5
Mode - MFCC-RF	98.7
Standard deviation - MFCC - RF	88.6
Mean - MFCC-RF	98.2
Proposed method	99.0

2. Conclusions

Automatic ship type recognition is a required and useful task in marine surveillance and port management systems. In general, this task is performed either by using visual features extracted from ship image datasets or by using audio features.

In this study, we propose a new method for ship type recognition by using audio features in an image-based deep learning model. We use the idea that different types of ships have different distributions in the frequency domain. To observe the accuracy of this idea, we applied FFT on sound recordings of ships. The resulting frequency vs magnitude graphs of FFT are saved as images. This procedure is applied to each record in the dataset, which contains eight different ship types and one background. The saved images are given as input to the ResNet50 network for classification. According to the experimental results, the best accuracy achieved by the proposed method is 99%. However, the proposed method must be tested on datasets with more recordings and types of ships. Unfortunately, any dataset of real-world ship sound recording is inaccessible.

For future work, a well-trained deeper yet not complex DL network would be helpful in real-time recognition. If this method is applied to a dataset with thousands or more sound recordings of more classes to train a model, instantaneous recognition can be realized by using that pre-trained model at any time.

Authors' contributions

Author contributed to the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

References

[1]. Xinqiang, C., Yongsheng, Y., Shengzheng, W., Huafeng, W., Jinjun, T., Jiansen, Z., Zhihuan, W., "Ship Type Recognition via a Coarse-to-Fine Cascaded Convolution Neural Network", The Journal of Navigation, 2020, 73(4): 813-832.

[2]. Chuang, L. Z. H., Yujen, C., Tang, S. T., "A simple ship echo identification procedure with SeaSonde HF radar", Geoscience and Remote Sensing Letters IEEE, 2015, 12: 2491-2495.

[3]. Makedonas, A., Theoharatos, C., Tsagaris, V., Anastasopoulos, V., Costicoglou, S., "Vessel classification in Cosmo-Skymed SAR data using hierarchical feature selection", The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2015, 40: 975.

[4]. Antelo, J., Ambrosio, G., González-Jiménez, J., Galindo, C., "Ship Detection and Recognition in High-Resolution Satellite Images", In International Geoscience & Remote Sensing Symposium, Cape Town, South Africa, 514-517, (2009).

[5]. Kaçar, U., Kumlu, D., Kırcı, M., "A Novel Approach for Automatic Ship Type Classification", In 23rd Signal Processing and Communications Applications Conference, Malatya, Turkey, 2153-2156, (2015).

[6]. Wu, J., Zhu, Y., Wang, Z., Song, Z., Liu, X., Wang, W., Zhang, Z., Yu, Y., Xu, Z., Zhang, T., Zhou, J., "A novel ship classification approach for high resolution SAR images based on the BDA-KELM classification model", International Journal of Remote Sensing, 2017, 38: 6457-6476.

[7]. Bentes, C., Frost, A., Velotto, D., Tings, B., "Ship-iceberg Discrimination with Convolutional Neural Networks in High Resolution SAR Images", In 11th European Conference on Synthetic Aperture Radar, Hamburg, Germany, 1-4, (2016).
[8]. Dong, C., Liu, J., Xu, F., Liu, C., "Ship Detection from Optical remote-sensing Images Using Multi-Scale Analysis and Fourier HOG Descriptor", Remote Sensing, 2019, 11: 1529-1547.
[9]. Rainey, K., Reeder, J., Corelli, A., "Convolution neural networks for ship type recognition", SPIE Defense + Security, Maryland, United Sates, 9844, 984409, (2016).
[10]. Feng, Y.C., Wenhui, D., Sun, X., Yan, M.L., Gao, X., "Towards Automated Ship Detection and Category Recognition from High-Resolution Aerial Images", Remote Sensing, 2019, 11: 1901-1923.
[11]. Chen, X., Qi, L., Yang, Y.S., Luo, Q., Postolache, O., Tang, J.J., Wu, H., "Video-Based Detection Infrastructure Enhancement for Automated Ship Recognition and Behavior Analysis", Journal of Advanced Transportation, 2020, 7194342.
[12]. Cao, X., Gao, S., Chen, L. et al., "Ship recognition method combined with image segmentation and deep learning feature extraction in video surveillance", Multimedia Tools and Applications, 2020, 79: 9177-9192.
[13]. Lorencin, I., Anđelić, N., Mrzljak, V., Car Z., "Marine Objects Recognition Using Convolutional Neural Networks", NAŠE MORE, 2019, 66(3): 112-119.
[14]. Shen, S., Yang, H., Li, J., Xu, G., Sheng, M., "Auditory Inspired Convolutional Neural Networks for Ship Type Classification with Raw Hydrophone Data", Entropy (Basel), 2018, 20(12): 990-1003.
[15]. Zhang, L., Wu, D., Han, X., Zhu, Z., "Feature extraction of underwater target signal using Mel frequency cepstrum coefficients based on acoustic vector sensor", Journal of Sensors, 2016, 7864213.
[16]. Tuncer, T., Aydemir, E., "An Automated Local Binary Pattern Ship Identification Method by Using Sound", Acta Infologica, 2020, 4(1): 57-63.
[17]. Hladnik, A., Muck, T., Stanic, M., Cernic, M., "Fast Fourier Transform in Papermaking and Printing: Two Application Examples", Acta Polytechnica Hungarica, 2012, 9(5): 155-166.
[18]. Krizhevsky, A., Sutskever, I., Hinton, G. E., "ImageNet Classification with Deep Convolutional Neural Networks", In Conference of Neural Informarinmation Processing Systems, NIPS, Nevada, United States, 25, (2012).
[19]. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., "Going deeper with convolutions", In Conference on Computer Vision and Pattern Recognition, Boston, United States, 1-9, (2015).
[20]. Simonyan, K., Zisserman, A., "Very deep convolutional networks for large-scale image recognition", In International Conference on Learning Representations, San Diego, United States, 1-14, (2015).
[21]. Kaiming, H., Xiangyu, Z., Shaoqing, R., Jian, S., "Deep Residual Learning for Image Recognition", In Computer Vision and Pattern Recognition, Las Vegas, United States, 770-778, (2016).
[22]. Chu, Y., Yue, X., Yu, L., Sergei, M., Wang, Z., "Automatic Image Captioning Based on ResNet50 and LSTM with Soft Attention", Wireless Communications and Mobile Computing 2020, 8909458.
[23]. Abadi, M., et al., "Tensorflow: Large-scale machine learning on heterogeneous distributed systems", arxiv.org, 2016, 1603.04467.
[24]. Keras, (2015). https://github.com/keras-team/keras.