UNDERWATER ACOUSTIC SIGNAL ANALYSIS: PREPROCESSING AND CLASSIFICATION BY DEEP LEARNING

H. Wu,* *Q. Song*,* *G. Jin*,*

Abstract: The identification and classification is important parts of the research in the field like underwater acoustic signal processing. Recently, deep learning technology has been utilized to achieve good performance in the underwater acoustic signal case. On the other side, there are still some problems should be solved. The first one is that it cannot achieve high accuracy by the dataset that is transformed into audio spectrum. The second one is that the accuracy of classification on the dataset is still low, so that, it cannot satisfy the real demand. To solve those problems, we firstly evaluated four popular spectrums (Audio Spectrum, Image Histogram, Demon and LOFAR) for data preprocessing and selected the best one that is suitable for the neural networks (LeNet, ALEXNET, VGG16). Then, among these methods, we modified a neural network(LeNet) to fit the dataset that is transformed by the spectrum to improve the classification accuracy. The experimental result shows that the accuracy of our method can achieve 97.22 %, which is higher than existing methods and it met the expected target of practical application.

Key words: *underwater acoustic signal, deep learning, classification framework, spectrum*

Received: November 19, 2018
Revised and accepted: April 30, 2020

DOI: 10.14311/NNW.2020.30.007

1. Introduction

The ocean covers 71 % of the entire Earth's surface and is an essential part of the Earth's life supporting system. It contains huge underwater natural treasure like rich mineral and biological resources. Underwater target detection requires the use of appropriate underwater imaging techniques. Currently, underwater imaging technology mainly includes optical imaging and sonar imaging [1]. Sonar image has the advantages of representing long distance of action as it has strong penetrating ability, and is especially suitable for mixed waters. Therefore, it has been widely used in underwater address geomorphology survey, underwater-lost object searching, mine detection, dam foundation detection and other fields [2].

*Hao Wu, Qingzeng Song, Guanghao Jin – Corresponding author; School of Computer Science and Technology, Tiangong University, Tianjin, China, E-mail: moyingyao@foxmail.com, qingzengsong@tiangong.edu.cn, jingh\protect_research@163.com

Image processing for the characteristics of sonar images can clearly and realistically present the underwater scene information in front of the sonar operator, improve the accuracy of manual interpretation, and reduce the probability of missed and misjudged targets [3]. Under the same transmitting power, the sonar equipment has different echo intensities for different types of targets. The echo intensity carries the characteristic information of the target, which can be used for imaging display, target classification and recognition. Typically, sonar target recognition is based on characteristics such as radiated noise, timbre, and other factors belong to the ship, which depends on the experience or knowledge of the sonar operator. They identify the vessel by analyzing the radiation signal. The accuracy of those methods is greatly influenced by conditional factors such as experience and the training of voice workers who require a lot of time of practice and related financial support. Experts tried to classify underwater acoustics using machine learning methods such as k-NN [4]. In addition, Gaussian mixture models (GMMs) [5] and multilayer perceptron (MLP) [6] are also used for classification, but the accuracy of classification is still insufficient to satisfy the needs of actual applications. This is mainly caused by two reasons: Firstly, in the case of underwater acoustic signals, due to the complex environment, the quality of the obtained signals is very poor as it includes various noise, which causes the recognition rate of direct classification is very low; Secondly, traditional machine learning methods usually extract features manually, which greatly reduce the quality of the training set.

Hinton et al. proposed a deep learning framework [7]. Since then, it has proven to be effective in the classification of many applications, and attracts a large number of relevant researchers into related applications and actual system developments. In recent years, deep learning has made breakthroughs in the fields of speech analysis and image recognition. It captures hidden deep features of the target signal through a multi-level network without the effort to artificially design structural features. Compared with machine learning methods, deep learning can extract deeper features of the image through a multi-level network and achieve higher precision in like classification on large datasets. The convolutional neural network proposed by [8] is the first multi-layer structure that uses spatial relative relationships to reduce the parameters and improve training performance. In 2012, deep convolutional neural networks were applied to ImageNet and achieved amazing results [9]. Classification and positioning are the most important and challenging issues in the field like underwater acoustic signals. Researchers have done a lot of work in related area like the classification for marine animals and some unpredictable noise generated by objects. As the signal is unstable, it greatly reduces the recognition accuracy. Deep learning can be used to solve this problem as it can extract features based on big dataset. On the other side, as there are various noises, a good selection of the features plays a crucial role in its performance.

In this paper, we applied deep learning technology to the field of underwater acoustic signals. First of all, among the possible spectrums for the preprocessing of underwater acoustic signals, we evaluated the spectrums to find which is the most suitable one for neural network learning in the underwater signal case. Based on the selected spectrum, we modified an existing network to improve the accuracy of classification, so that, our framework meets the real demand of the identification and classification for aquatic signals which is to classify different types of ships and torpedoes. In the experiment, we compared the neural networks and selected the best one for classification. Each network uses the same training data and verification data. The results prove that our modified network can achieve the highest precision, which proves that the modification can further improve the classification accuracy.

The remainder of this paper is organized as follows: Section 2 provides a brief overview of the work in the field of underwater acoustic signal processing. Section 3 introduces our materials and methods, detailing the four spectrums that are used for data preprocessing and the classification networks. In Section 4, it describes our experimental process and discusses the results. Finally, it summarizes our work in Section 5 and introduces the direction of our future work.

2. Related works

Classification and the following positioning are the most important and challenging issues in the research field like underwater acoustic signals. Many researchers have done a lot of contributions in related area like the classification of marine animals and some noise generated by the objects. As the signal is unstable and unpredictable, it may reduce the accuracy of the classification models. Deep learning can be used to solve this kind problem as it can extract features based on big dataset. On the other side, as there are various noises, the good selection of the features plays an important role in its performance.

Existing methods of preprocessing for underwater acoustic signals are mainly to extract time domain or frequency characteristics, spectrum estimation, etc. In [10], Zhu, P. et al. identified sonar images by convolutional neural networks (CNN) by extracting features in underwater vehicle identification cases and using support vector machines (SMV). In [11], M. Valdenegro-Toro uses neural networks to perform object recognition on forward-looking sonar images and has achieved good results. In [12], Gang Hu et al. proposed a new method for feature extraction and recognition of underwater noise data based on deep learning framework, which applied CNN and ELM to the recognition and classification of underwater targets. In this paper, traditional machine learning methods are compared and the accuracy is greatly improved. In [13], Jager J. et al. uses the activation of convolutional neural networks to simulate the appearance of objects; thereby it can improve the accuracy of fish identification and tracking. In [14], Yue, H. et al. used convolutional neural networks (CNN) and deep belief networks (DBN) to classify underwater acoustic signals. The experimental results show that the deep learning method is suitable for the classification of underwater targets and can improve the recognition accuracy.

Paper [15] pointed out many important things for audio signal processing like the similarities and differences between domains, problems, methods, key references and potential for cross-fertilization between areas. In this work, key issues and future questions regarding deep learning applied to audio signal processing are identified. On the other side, there is still big gab between the state-of-art performance and real demands because of the limited samples or plenty types of noise in underwater case.

In this paper, the original audio signals are firstly converted into four kinds of spectrum samples to evaluate the efficiency of retaining key features. In order to find the best spectrum that can be used with deep learning network, the datasets that are derived from the four spectrums are input into some neural networks for training and testing. After comparison and verification, the best spectrum is selected and the corresponding samples are input into some neural models. Then the best model is selected for the classification task to achieve high accuracy.

3. Materials and methods

Our original materials are audio files which format is *.wav. In real applications, the radar operator makes the judgment based on not only the sound from the headphones, but also the map of the audio that is displayed on the machines. Thus the audio files of the target can be transformed into spectrums, so that, the original audio recognition task is converted to an image recognition task. Then the spectrum dataset is input into the deep neural networks, and after training and testing, it can achieve the purpose of assisting the radar operator.

There are many types of spectrums. In order to find the best spectrum that is suitable for deep neural network training, we selected four spectrums and converts audio data into spectrum images. We selected the most popular spectrums that are Audio Spectrum [16], Image Histogram [17], Demon [18] and LOFAR [19]. The datasets from related spectrum are input into LeNet [8] for training and testing. The experimental results show that the accuracy of LOFAR is much higher than the other three kinds of spectrum. Therefore, it proves that deep neural networks are easier to extract signal features in LOFAR format data. In order to improve the classification accuracy, we utilized LOFAR spectrum data through LeNet, ALEXNET [20], VGG16 [21]. The experiments show that their accuracy is similar, while it is not enough to the actual application. Based on the LeNet, we tried to improve the accuracy with some modification. The experimental results show that our modified network greatly improves the accuracy and can achieve the practical application goals.

The whole process is shown in Fig. 1, where here (a) is the original audio signal sample, and after the data pre-processing, the four types spectrum datasets (b) are obtained. After the four types of spectrum datasets were trained by LeNet, the accuracy of LOFAR was found to be the highest one by the comparison of accuracy. Thus our framework selected LOFAR spectrum and the related dataset is input into three classification networks as (d) shows. To further improve the accuracy, we modified LeNet. In step (e), the networks are evaluated on the test dataset.

3.1 Spectrums

In the field of underwater acoustic signals, the original dataset is to be classified by transforming it to audio spectrum. Because of the noise, the audio signals are difficult to be identified and classified by neural networks in under water case. Thus, we tried to convert them into different spectrums. There are many spectrums that can be used to converted underwater acoustic signals. We selected the most popular four types of spectrums: Audio Spectrum, Image Histogram, Demon and

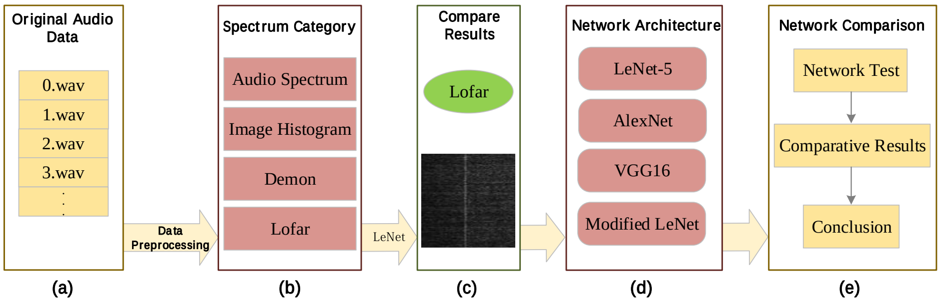

Fig. 1 *The entire process of our framework.*

LOFAR. Fig. 2 shows the samples of audio spectrum data. The horizontal axis is the frequency and the vertical axis is the intensity. The graphic data of the signal at various frequencies is recorded in a corrugated manner. There are 14,383 samples of the audio spectrum dataset, of which 10,080 are for training and 4,303 are for verification.

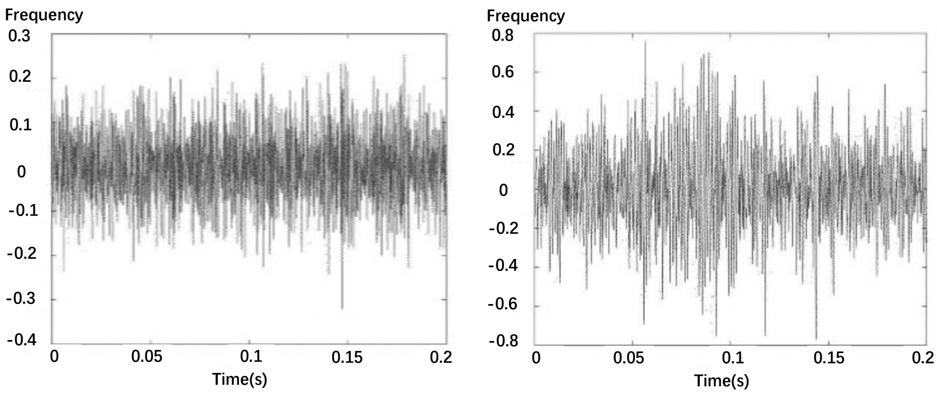

Fig. 2 *The samples in the audio spectrum dataset.*

Image Histogram is a histogram used to represent the distribution of brightness in an image, which plots the number of pixels for each luminance value of the image as Fig. 3 shows. In the Image Histogram, the left side of the abscissa is a pure black, darker area, and the right side is a brighter, whiter area. Therefore, the data in the histogram of an image of a darker picture is mostly concentrated on the left side and the middle part. In the field of computer vision, image histograms are often used to binarize in images. There are 14,383 samples of the Image Histogram dataset, of which 10,080 are for training and 4,303 for verification. We use a balanced histogram, which is a method of adjusting the contrast to the image-processing field. This method is often used to increase the global contrast of many images, especially when the contrast of useful data for an image is fairly close. In that way, the brightness can be better distributed over the histogram. This can be used to enhance local contrast without affecting the overall contrast, and histogram equalization achieves that by effectively extending the commonly used brightness. The normalized histogram is obtained by dividing the count of each attribute on the histogram by the sum of the counts of all the attributes. The reason for "normalization" is because the sum of all the attributes in the normalized histogram is 1, that is, the count corresponding to each attribute is a number (percentage) between 0 and 1.

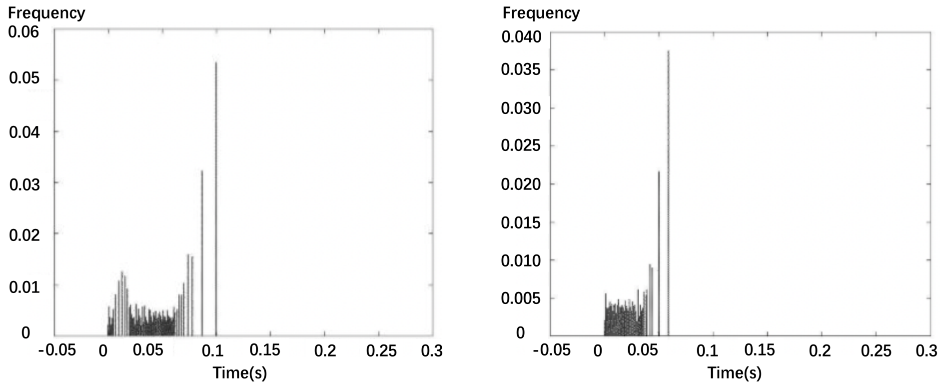

Fig. 3 *Part of the sample in the Image Histogram dataset.*

Demon spectrum extraction has stable features and clear physical meanings. It is one of the important analytical methods of ship target recognition. It can obtain line spectrum information through high-frequency broadband noise mediation analysis. The Demon analysis processing is performed on the underwater acoustic signal, and the underwater acoustic signal received by the passive sonar is selected by a high frequency band of square demodulation, and then subjected to low-pass filtering to obtain a corresponding Demon spectrum. Our Demon spectrum dataset has 11,796 samples, of which 8,292 are for training and 3,504 are for verification. Fig. 4 shows samples of the Demon spectrum dataset.

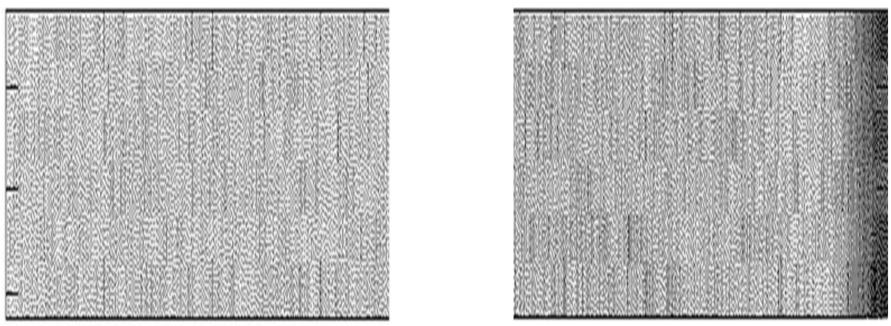

Fig. 4 *The samples in the Demon dataset.*

The LOFAR spectrum is a continuous time domain sample of the signal. The spectrum and frequency characteristics of the radiated noise is analyzed by using LOFAR [19]. Based on the local stability of the signal, it performs a short-time

Fourier transform and the obtained time-varying information. It is projected on the time and frequency planes, so it forms a three-dimensional mapping. It is usually used for target recognition in underwater signal case. In the field of underwater acoustic signal processing, target recognition and tracking tasks are achieved by using various methods of extracting and identifying line profile features. Fig. 5 shows a sample of the LOFAR spectral dataset.

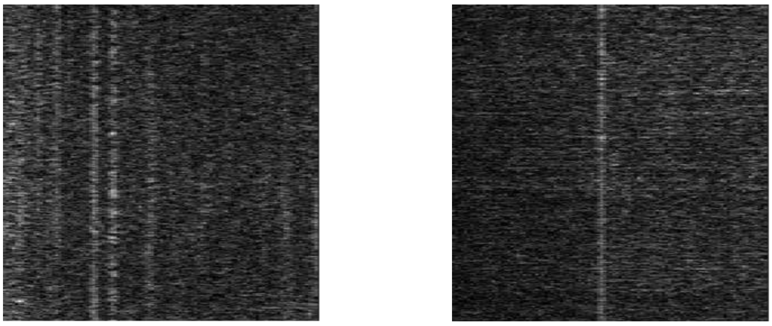

Fig. 5 *Part of the sample in the LOFAR dataset.*

3.2 Which spectrum is the best?

The classification problem is an important research issue in the field of underwater acoustic signals, and researchers have done a lot of work in this field. The underwater acoustic signal is unstable because of the reasons like noise, and such a phenomenon greatly reduces the recognition accuracy. One of the ways to solve this problem is feature extraction, and the features presented to the classifier play a crucial role in its performance. In fact, the selection of dataset may be more important than the modification the classifier when improving the accuracy. After the data is preprocessed, it gets the datasets of Audio Spectrum, Image Histogram, Demon and LOFAR. In order to determine which spectrum is the most suitable one for neural networks, we input these into the classification network LeNet for training. Fig. 6 shows the network architecture of LeNet [8].

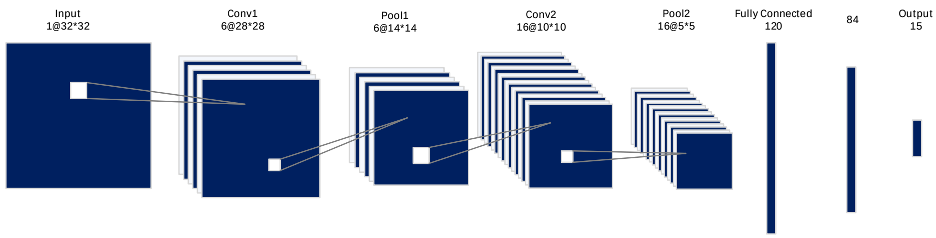

Fig. 6 *LeNet network architecture.*

LeNet has 7 layers, including 2 layers of convolution, 2 layers of pooling, and 3 layers of fully connected layers. The convolution operation is a weighted summation of pixels based on a convolution kernel. The purpose of the convolutional layer is to extract the features of the input image. The first convolutional layer can only extract some relatively low-level features, such as edges, lines, etc. The second convolutional layer extracts more complex features based on the low-level features extracted by the first layer. The convolutional layer in the LeNet network architecture uses 5×5 convolution kernels, and the convolution kernel has a moving step size of 1. The role of the pooling layer is equivalent to converting a picture with higher resolution to a picture with lower resolution, which can reduce the number of parameters in the neural network case. The most common operation of those methods is the largest pooling and average pooling. The pooling layer in the LeNet network uses a 2×2 input field with a moving step size of 2, that is, four pixels of the previous layer are input to one pixel of the next layer, and the input fields does not overlap. The pooling method adopted is the average pooling. The fully connected layer spatially transforms features, especially transforms high-dimensional features into low-dimensional features, and preserves useful information. The full connection of the last layer acts as a classifier. The datasets of the four types of spectrum are input into LeNet for training. By comparing the accuracy, we found that LOFAR spectrum is the most suitable one for training of neural networks.

3.3 Classification network comparison

Based on the evaluation on spectrums, our framework determined to use LOFAR in the field of underwater acoustic signals. Then our framework also inputs LOFAR spectrum data into some mainstream classification networks. The experimental results show that the difference of classification accuracy between LeNet, VGG16 and ALEXNET are not obvious, and they have not reached the standard that can be applied. Thus we designed a neural network derived from LeNet. It consists of an input layer, three convolution layers, three maximum pooling layers and three fully connected layers. Its network structure is shown in Fig. 7.

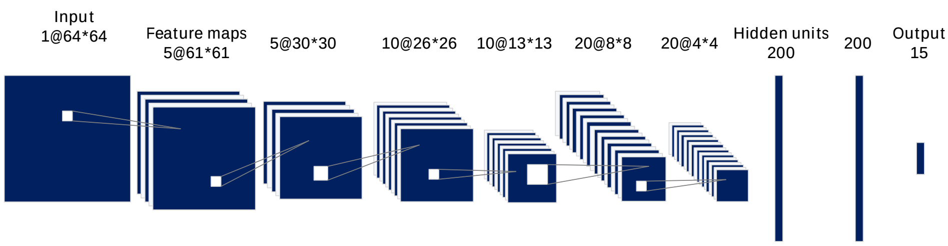

Fig. 7 *Network architecture.*

In order to facilitate the training of the neural networks when there are insufficient samples, we adjust the size of the LOFAR spectrum from 512×512 to 64×64, which is more suitable for network training. We input the 64×64 LOFAR spectrum into our improved classification models and the other ones (LeNet, VGG16, ALEXNET). In more details about our network, the first convolutional layer has five convolution kernels, each of which has a size of 4×4 and a convolution kernel with a sliding step size of 1, without padding. The size of the filter in the pooling layer is 2×2 and the step size is 2. The goal is to reduce the dimensions of the feature map and extend the perception domain. The effect of dimensionality reduction is to reduce the original image to a quarter and leave an average output. There are three combinations of such a convolution layer and a pooling layer. In order to make the network more suitable for the identification and classification of LOFAR spectrum in aquatic signals, it adjusted the size and number of convolution kernels in each convolutional layers. The specific parameters are shown in Tab. I.

Name	Channel	Kernel Size	Stride	Padding	Type
Conv1	5	4	1	Valid	–
Pool1	5	2	2	Valid	Max
Conv2	10	4	1	Valid	–
Pool2	10	2	2	Valid	Max
Conv3	10	4	1	Valid	–
Pool3	10	2	2	Valid	Max
FC1	200	–	–	–	–
FC2	200	–	–	–	–
FC3	15	–	–	–	–

Tab. I *Network parameter configuration.*

4. Experiment

4.1 Spectrum comparison

During the comparison experiment of spectrums, the training set and the test one of each dataset are using the same equipment. The experimental results are shown in Tab. II. We select the LeNet as the classification model to compare these spectrums.

Spectrum Category	Accuracy [%]
Audio Spectrum	54.93
Histogram	55.12
Demon	13.07
LOFAR	**83.64**

Tab. II *Spectrum comparison.*

The experimental result shows that the Demon spectrum has the lowest accuracy, which shows it is not suitable for the training of neural networks. The accuracy of Audio Spectrum and Image Histogram is similar, but it still cannot meet real requirements because of the low accuracy. LOFAR has the highest accuracy, greater than 80 %, and is suitable to deep learning framework training for further improvement. When the training samples are insufficient, the spectrums should help the models capture key features. LOFAR applies a short-time Fourier transform that can reduce the complexity of the signals while amplifying the most important features. This kind preprocessing can be a good choice to improve the accuracy of models when there are insufficient training samples.

4.2 Modified network

Based on the experiment, our framework determined that in the field of underwater acoustic signals, LOFAR spectrum is a suitable spectrum for the deep learning classification framework. On the other side, its classification accuracy is not enough for practical applications. Our framework inputs the LOFAR spectrum dataset to VGG16, ALEXNET. ALEXNET consists of 12 layers and consists of five convolutional layers, three pooling layers and four fully connected layers. The network uses 11×11 large convolution kernel with a total parameter of 60M. There are many versions of VGG, and this article uses VGG16 for comparison. VGG16 is a large network with a total of 138 million parameters. It consists of 13 convolutional layers, 5 pooled layers and 2 fully connected layers. The experimental results are shown in Tab. III.

Network Architecture	Accuracy [%]
LeNet	83.64
ALEXNET	89.30
VGG16	82.47
Our Net	**97.22**

Tab. III *Spectrum comparison.*

The experimental results show that LeNet, ALEXNET and VGG16 have similar accuracy, and they have not reached the standard of practical application. Considering that these networks are based on the network architecture derived from CNN, we also designed a suitable one based on LeNet. The classification structure is shown in Fig. 6. The experimental results show that the accuracy has been greatly improved, which reaches 97.22 % and meets the requirements of practical applications. It proves that the classification network can be further improved by the modification of the neural networks.

4.3 Discussion

Audio spectrum or Demon one can express details of the signals. The more information does not fit the models when the training samples are insufficient. On the contrary, Histogram reduces so much information which can improve the efficiency of training while the accuracy is limited as this loses some key features. LOFAR one can reduce some noises while retaining enough key features to ensure high accuracy.

The tuning of models is a challenge especially when the structure of these is complicated. When the training samples are insufficient, the model that has big structure is hard to be trained and achieve high accuracy because of the problems like gradient vanishing one. Thus we select a medium model to tune the parameters when the training samples are insufficient.

5. Conclusions and future work

By comparing the four spectrums of Audio Spectrum, Image Histogram, Demon and LOFAR, we found that in the field of underwater acoustic signals, LOFAR spectrum is the most suitable spectrum for neural network training among evaluated spectrums. Based on LOFAR spectrum, we designed a neural network architecture for the recognition and classification of underwater acoustic signals, which greatly improved the classification accuracy. In practical situations, datasets are not readily available due to some limitations and security reasons. Therefore, when adding classification labels to actual applications for in-depth analysis and finding details, there may be no large datasets for training. In future work, we will try to do some research on how to increase the number of samples by using methods such as Generative Adversarial Nets.

Acknowledgement

This work has been supported by the National Natural Science Foundation of China (Grant No. 61802279, 6180021345, 61702366) and Natural Science Foundation of Tianjin City (Grant No. 19JCYBJC15800, 19JCTPJC49200, 18JCYBJC44000).

References

[1] KOCAK D.M., DALGLEISH F.R., CAIMI F.M., et al. A Focus on Recent Developments and Trends in Underwater Imaging. *Marine Technology Society Journal*. 2008, 42(1), pp. 52–67, doi: 10.4031/002533208786861209.

[2] NGUYEN H.G., FABLET R., EHRHOLD A., et al. Keypoint-Based Analysis of Sonar Images: Application to Seabed Recognition. *IEEE Transactions on Geoscience & Remote Sensing*. 2012, 50(4), pp. 1171–1184, doi: 10.1109/tgrs.2011.2165848.

[3] DOBECK G.J., HYLAND J.C., SMEDLEY L. Automated detection and classification of sea mines in sonar imagery. *The International Society for Optical Engineering*. 1997, 3079, pp. 90–110, doi: 10.1117/12.280846.

[4] WALTERS T.C. Auditory-based processing of communication sounds. *University of Cambridge*. 2011, doi: 10.17863/CAM.16336.

[5] ATREY P.K., MADDAGE N.C., KANKANHALLI M.S. Audio Based Event Detection for Multimedia Surveillance. *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2006, doi: 10.1109/ICASSP.2006.1661400.

[6] SIDIROPOULOS P., MEZARIS V., KOMPATSIARIS I. On the Use of Audio Events for Improving Video Scene Segmentation. *International Workshop on Image Analysis for Multimedia Interactive Services*. 2010, pp. 1–4, doi: 10.1007/978-1-4614-3831-1_1.

[7] KRIZHEVSKY A., SUTSKEVER I., HINTON G. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in neural information processing systems*, 2012, 25(2), doi: 10.1145/3065386.

[8] KAYED M., ANTER A., MOHAMED H. Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture. *ITCE2020*, Aswan, Egypt, 2020, pp. 238–243, doi: 10.1109/ITCE48509.2020.9047776.

[9] KRIZHEVSKY A., SUTSKEVER I., HINTON G.E. ImageNet classification with deep convolutional neural networks. *International Conference on Neural Information Processing Systems*. 2012, pp. 1097–1105, doi: 10.1145/3065386.

[10] ZHU P., ISAACS J., FU B. Deep learning feature extraction for target recognition and classification in underwater sonar images. *Conference on Decision and Control*. 2018, pp. 2724–2731, doi: 10.1109/CDC.2017.8264055.

[11] VALDENEGRO-TORO M. Objectness Scoring and Detection Proposals in Forward-Looking Sonar Images with Convolutional Neural Networks. *IAPR Workshop on Artificial Neural Networks in Pattern Recognition*. 2016, pp. 209–219, doi: 10.1007/978-3-319-46182-3_18.

[12] HU G., WANG K., PENG Y. Deep Learning Methods for Underwater Target Feature Extraction and Recognition. *Computational Intelligence and Neuroscience*. 2018(3), pp. 1–10, doi: 10.1155/2018/1214301.

[13] JAGER J., WOLFF V., FRICKE-NEUDERTH K., et al. Visual fish tracking: Combining a two-stage graph approach with CNN-features. *Oceans. IEEE*. 2017, pp. 1–6, doi: 10.1109/OCEANSE.2017.8084691.

[14] YUE H., ZHANG L., WANG D. The Classification of Underwater Acoustic Targets Based on Deep Learning Methods. *International Conference on Control, Automation and Artificial Intelligence*. 2017, doi: 10.2991/caai-17.2017.118.

[15] PURWINS H., LI B., VIRTANEN T., SCHLUTER J., CHANG S.Y., SAINATH T.N. Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*. 2019, pp. 1-1, doi: 10.1109/JSTSP.2019.2908700.

[16] CHANDRA D., TESSLER M.J., USHER J. Audio spectrum and sound pressure levels vary between pulse oximeters. *Canadian Journal of Anaesthesia*, 2006, 53(1), pp. 26–32, doi: 10.1007/BF03021524.

[17] NAKIB A., OULHADJ H., SIARRY P. Image histogram thresholding based on multiobjective optimization. *Signal Processing*, 2007, 87(11), pp. 2516–2534, doi: 10.1016/j.sigpro.2007.04.001.

[18] ZHI-XIANG Y. Identifying two targets by a single vector hydrophone based on the DEMON spectrum. *applied acoustics*, 2005, doi: 10.1080/02533839.2005.9671036.

[19] VAN HAARLEM M.P, WISE M.W, GUNST A.W, et al. LOFAR: The low-frequency array. *Astronomy & Astrophysics*. 2013, 556, A2, doi: 10.1051/0004-6361/201220873.

[20] XIAOBING H., YANFEI Z., LIQIN C., et al. Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification. *Remote Sensing*, 2017, 9(8), 848, doi: 10.3390/rs9080848.

[21] HAN S., MAO H., DALLY W.J. Analysis of Content-Aware Image Compression with VGG16. *IWOBI*, San Carlos, Costa Rica, 2018, doi: 10.1109/IWOBI.2018.8464188.