Time-Frequency Feature-Based Underwater Target Detection with Deep Neural Network in Shallow Sea

Yunliang Zheng\(^1\), Qiyong Gong\(^2\,\ast\) and Shufang Zhang\(^3\)

\(^1\)China Rescue and Salvage of Ministry of Transport of the People's Republic of China, Beijing, China
\(^2\)Qingdao Institute of Marine Geology, Qingdao, China
\(^3\)National University of Defense Technology, Changsha, China

\ast Corresponding author’s e-mail: qygong@qnlm.ac

Abstract. In this paper, we propose a time-frequency feature-based signal detection method with deep neural network and apply it for target detection in shallow sea. The short time Fourier transform is employed to capture the time-frequency feature of the target signal. We input the time-frequency diagram of the signal into the deep neural network model, train the network model through the labeled data samples, and use the test set to estimate the signal detection probability. In addition, we analyze the influence of different learning rates on the detection performance. The experimental results show that the performance of the proposed method is better than that of the energy detector, and the detection performance of different learning rates is different.

1. Introduction
In the military task of underwater target detection and recognition, passive sonar system has better concealment and flexibility [1]. How to use the ship radiated noise obtained by passive sonar to detect and recognize underwater targets has become an urgent problem. However, due to the application of various Ship Stealth technologies and the complexity of marine environment, underwater target detection based on ship radiated noise is facing great challenges [2]. At present, we still rely on trained sonar for underwater target recognition. However, due to the instability of human physiological and psychological factors, it is difficult to achieve stable all-weather underwater target detection.

Automatic threshold estimation technology is a kind of detection technology based on airspace information. The detection technology based on airspace information is closely combined with the front-end of the system, and its technical theory is mature. It has been widely used in practical engineering [3]. The basic idea of automatic threshold estimation technology is to automatically select and calculate the original data within the limited window length according to the design criteria. The data representing noise in the data is lower than the calculated noise threshold, while the data representing the signal is higher than the calculated noise threshold, so as to achieve the purpose of suppressing noise and improving signal detection ability [4]. In the research, the signal noise or clutter model is generally simplified as a probability statistical model. For example, in the case of passive sonar, both ocean noise and ship radiated noise can be regarded as stationary random signals conforming to Gaussian distribution. Signal detection of passive sonar can be simplified as the optimal detection problem of Gaussian signal in Gaussian noise background [5]. In order to determine the threshold value of the detector, a series of criteria are derived. The most common criteria are the
maximum posterior probability criterion based on the prior probability. The Bayesian criterion that depends on the prior probability and the minimum detection loss, the minimum error probability criterion for determining the risk value of the loss function and minimizing the number of wrong decisions as far as possible, and minimizing the maximum risk when the prior probability is unknown [6]. The minimax criterion and the Newman Pearson criterion, which strictly restrict one kind of error and minimize the probability of other errors, are considered.

In order to further promote the development of signal detection technology, researchers have carried out a lot of research on signal characteristics, and put forward many detection methods based on signal characteristics. Instantaneous feature method [7] uses instantaneous frequency and instantaneous amplitude of signal as detection target, and its principle is simple, but it will be adversely affected by signal-to-noise ratio or signal-to-noise ratio (SNR/SCR). The zero-crossing method [8] has good performance at high SNR/SCR level, but requires high data sampling rate. Spectral line feature method [9] does not need prior knowledge and has strong robustness, but its range of action is limited by the spectrum characteristics of signals.

In recent years, deep learning has been widely concerned. Deep learning has achieved excellent performance in target classification and detection [10], and has been applied to speech recognition, fingerprint recognition, vehicle illegal photography and other fields. At present, researchers are trying to apply deep learning to the field of signal detection, and have made a series of achievements in signal recognition and classification. In reference [11], neural network is used to detect and classify jamming signals of synthetic aperture radar; in reference [12], deep learning is used to detect micro Doppler signals of human detection and Recognition Radar; and in reference [13], deep learning is used to detect echo signals of ultrasonic targets. In the field of communication, deep learning is applied to classification of modulation modes, and good results are achieved. Reference [14] designed a framework based on convolutional neural network, and used it to classify the short-time Fourier transform time-frequency spectrum of signals with different modulation modes, and the experimental results show that it has good recognition accuracy; reference [15] proposed a signal feature fusion method based on convolutional neural network, and achieved excellent performance in simulation experiments. There are also some domestic postgraduate projects using deep learning to carry out research in the field of signal processing, such as using neural network and feature factor coincidence in [16] to detect ultra wideband signal, and in [17]. Using BP network and convolution neural network to identify and analyze the signal through the characteristics of time domain, frequency domain and waveform wavelet domain.

In the field of underwater acoustic, a number of research results have emerged based on deep learning signal processing algorithm. [18] the achievements of deep learning in underwater acoustic signal processing, especially in underwater acoustic signal detection, and lists the authors, publication time and main contributions of the paper. It can be found that although the sonar signal detection research based on deep learning has achieved a series of ideal results, the related research has not been systematized. Behind the high accuracy results are the pre-processed experimental data or the experimental environment close to the ideal conditions. Strong interference and lack of samples still restrict the application of deep learning theory in sonar signal detection Application in the field of measurement. Domestic research is mainly based on master's project, and the research is not in-depth.

In this paper, the time-frequency characteristics of the signal are used to detect the target. The time-frequency characteristics of the signal are obtained by short-time Fourier transform. Google net is used as the classifier and the network parameters are trained to distinguish the background noise from the target signal. The influence of different learning rates on the detection performance is analyzed, and the detection performance differences under different SNR are compared. The results show that the proposed method can achieve high detection probability when SNR is -10dB.

2. The STFT of signal
Short time Fourier transform (STFT) is to add a short window function to the signal moving along the time axis. The slice each time of the signal is taken out from the short window, and the Fourier
transform is performed respectively to obtain the spectrum each time (called "local spectrum"). Let 
\(g(t)\) be a short window function, the STFT of signal \(x(t)\) is defined as

\[
STFT_x(t,f) = \int_{-\infty}^{\infty} x(u) g^*(u-t) e^{-j2\pi fu} du
\]

Set \(STFT_x(t,f)\) samples are taken at the equally spaced time-frequency grid points \((mT,nF)\), where \(T\) and \(F\) are the sampling periods of time variable and frequency variable respectively, and \(m\) and \(n\) are integers. In order to simplify the calculation, the symbol \(STFT(m,n) = STFT(mT,nF)\) is introduced. Thus, for discrete signals \(x(t)\), the discretization form of (1) is

\[
STFT(m,n) = \sum_{k=-\infty}^{\infty} x(k) g^*(kT-mT) e^{-j2\pi(nF)k}
\]

It can be seen that the result of STFT transformation is a two-dimensional complex matrix, the abscissa is time, the ordinate is frequency, and its amplitude matrix is

\[
M = |STFT(m,n)|
\]

For short-time Fourier transform, the choice of window function \(g(t)\) is the key, and the determination of window width should consider both time domain resolution and frequency domain resolution. At present, there is no fixed mode for the selection of window function. Through the comparative test of Hamming window, rectangular window and Blackman window with different widths, it is found that the main lobe width of rectangular window spectrum is the narrowest, the main lobe width of Blackman window function spectrum is slightly wider than Hamming window, and the attenuation amplitude of the first side lobe of Blackman window relative to the main lobe is much larger than that of rectangular window and Hamming window. Moreover, the side lobe attenuation rate is fast. The smaller the amplitude of the side lobe relative to the main lobe, the smaller the spectrum leakage error and the smaller the ripple of the transformed signal. In order to determine a suitable window function, it is necessary to compare the analysis results of power quality disturbance signals by using ST-transform with different window functions. The comparative study shows that for the analysis signal with 50 Hz power frequency and 1.6 kHz sampling frequency (i.e. 32 sampling points per power frequency cycle), selecting the Blackman window with a width of 65 points can obtain sufficient time-frequency resolution and flat harmonic spectrum distribution sequence barrier.

By dividing the steady-state signal amplitude matrix by the coefficients generated by windowing, the normalized spectrum amplitude matrix (referred to as the spectrum matrix) is obtained, and the spectrum amplitude of each frequency component is equal to the actual value. The column vector corresponding to the fundamental frequency in the matrix is called the fundamental frequency spectrum sequence.

### 3. Target Detection with Deep Neural Network

The task of target detection is viewed as a classification problem and is completed by the Googlenet. Googlenet is a convolutional neural network model proposed by the research team of Google in 2014, which won the first place in the ILSVRC (Imagenet large scale visual recognition challenge) competition that year. What's unique about googlenet is that there's an improved sparse network design called inception. Inception has been steadily improved across multiple versions (V1, V2, V3, and V4). Here, we introduce the version of inception V1 used by Googlenet.

In theory, the most direct way to enhance deep learning algorithm is to increase the number of network layers or neurons, but in practical application, there will be two problems: one is too many parameters, which will lead to the increase of computational complexity, and when the data set is limited, there will be over fitting problem; the other is that with the increase of depth, there will be "gradient disappear" problem. When the back propagation algorithm updates the parameters, the
shallow network level cannot update the weights. In order to reduce the number of parameters, sparse join can be used. However, most hardware optimization focuses on dense connections, so the computational complexity of sparse connections is not reduced. Inspired by the structure of human brain neurons, Google team developed concept V1, which is shown in Figure 1.

![Concept V1 structure in Googlenet.](image)

Figure 1. Concept V1 structure in Googlenet.

In concept V1, convolution (blue block) of 1×1, 3×3, 5×5 size and 3×3 Maximum pooling layer (purple block) are stacked together, and the output image size is the same after they are calculated. Convolution layer is used to extract image features of different scales, and maximum pooling layer is used to alleviate over fitting problem. After each convolution layer, a nonlinear calculation is performed to increase the nonlinearity of the network. A 1×1 convolution layer (orange block) is applied to concept V1 to reduce the thickness of the feature map being stacked and the number of parameters. After the calculation mentioned above, all the outputs are finally joined together to enter the next exceptio structure. There are 22 layers of Googlenet, including 9 interception V1 structures, which are modularized in the network. There are two branches in the third and sixth structure, i.e. auxiliary classifier. In the training process, the auxiliary classifier uses the output of the middle layer as the classification, and adds the decision result with a small weight (generally 0.3). This not only helps to alleviate the problem of gradient vanishing, but also realizes feature fusion. The target detection can be illustrated in Figure 2.

![The target detection scheme.](image)

Figure 2. The target detection scheme.

4. Experimental Results
In the framework of deep learning, this section compares the above three signal detection methods: time-frequency characteristic signal detection method based on deep neural network with the learning rate 0.001 and 0.0001, respectively; and energy detector.
Figure 3 shows the time-frequency characteristic color image samples of the simulation data set used in this study. Figure 3 (a) is the sample containing noise only, and figure 3 (b) to figure 2 (d) are the samples with signal, and the SNR is -11db, -15dB and -19dB, respectively. It can be seen from figure 3 that the non Gaussian simulation noise based on K distribution is more evenly distributed in the time domain and frequency domain. There is a trace of the target signal in the lower middle position of all the images in Fig. 3 (b) to (d), but it is blurred by the noise interference and is on the verge of disappearing with the continuous reduction of SNR.

| SNR(dB) | GoogleNet, learning rate $= 10^{-3}$ | GoogleNet, learning rate $= 10^{-4}$ | Energy detector |
|--------|-------------------------------------|-------------------------------------|-----------------|
| -14    | 0.007894737                        | 0.016                               | 0               |
| -12    | 0.011403509                        | 0.016                               | 0               |
| -10    | 0.020175439                        | 0.014666667                         | 0               |
| -8     | 0.045614035                        | 0.022666667                         | 0               |
| -6     | 0.101754386                        | 0.052                               | 0               |
| -4     | 0.249122807                        | 0.109333333                         | 0               |
| -2     | 0.446491228                        | 0.233333333                         | 0               |
| 0      | 0.628947368                        | 0.433333333                         | 0               |
| 2      | 0.796491228                        | 0.676                               | 0.0016          |
| 4      | 0.9289473                          | 0.866666667                         | 0.0931          |
| 6      | 0.992105263                        | 0.946666667                         | 0.2172          |
| 8      | 1                                  | 0.964                               | 0.4809          |
| 10     | 1                                  | 0.982666667                         | 0.728           |

Table 1 shows the detection performance results of different algorithms. It can be seen from the results that compared with the classic energy detector, the performance of Google net has been greatly improved. At the same time, the detection performance of Google net under different learning rates is different, and the detection performance under the learning rate of 0.001 is significantly better than that under the learning rate of 0.0001.

In the experiment, the Googlenet model is used to process the time-frequency characteristic color image to detect the target signal. For Googlenet, the transfer learning method is used to adjust the full connection layer to adapt to the detection problem in this paper. The data used for transfer learning and retraining are only noise samples and signal samples with signal-to-noise ratio of -15 dB, respectively. Each training uses 1000 noise only samples and 1000 signal samples with the same signal-to-noise ratio.

5. Conclusion
In this paper, a deep neural network signal detection method based on time-frequency characteristic graph is proposed. The method uses the time-frequency characteristics of signal and background noise to distinguish between target and noise. The model parameters are obtained by training the network model with labeled data samples, and the signal detection probability is estimated by test set. We
analyzed the influence of different learning rates on the detection performance. The experimental results show that the performance of the proposed method is better than that of the energy detector.

Reference
[1] Shunichi, A. (2004). Information geometry of statistical inference. *Systems Control & Information*, 48, 428-436.
[2] Nandi, A. K. and Azzouz, E. E. (1997). Algorithms for automatic modulation recognition of communication signals. *IEEE Transactions on Communications*, 46(4), 431-436.
[3] Lin, T. Y., Goyal, P., Girshick, R., He, K. and Dollár, Piotr. (2017). Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99), 2999-3007.
[4] Pan, H., Wang, B. and Jiang, H. (2015). Deep learning for object saliency detection and image segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 27(6), 1135-1149.
[5] Zeng, Y., Zhang, M., Han, F., Gong, Y. and Zhang, J. (2019). Spectrum analysis and convolutional neural network for automatic modulation recognition. *IEEE Wireless Communications Letters*, 8(3), 929-932.
[6] Dmitrieva, M., Brown, K., Heald, G. and Lane, D. (2019). Material recognition based on the time delay of secondary reflections using wideband sonar pulses. *IET Radar, Sonar and Navigation*, 13(11), 2034-2040.
[7] Peng, C., Zhang, X., Song, Z. and Meng, Z. (2018). Optimal tone detection for optical fibre vector hydrophone. *IET Radar, Sonar and Navigation*, 12(11), 1233-1240.
[8] Dong H., Wang H., Shen X. and He K. (2019). Parameter matched stochastic resonance with damping for passive sonar detection. *Journal of Sound and Vibration*, 458(11): 1206-1218.
[9] Peng C., Zhang X., Song Z. and Meng Z. (2018). Optimal tone detection for optical fibre vector hydrophone. *IET Radar Sonar and Navigation*, 12(11): 1233-1240.
[10] Yang Y., Xiao S. P., Wang X. S., Zhang W. M. and Li Y. Z. (2019). Generalised polarimetric whitening filter for polarimetric mimo radar detection. *IET Radar Sonar and Navigation*, 13(1): 1-7.
[11] Bouvet M. and Schwartz S. C. (2002). Comparison of adaptive and robust receivers for signal detection in ambient underwater noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(5): 621-626.
[12] Wang F. T., Chang S. H. and Lee C. Y. (2006). Signal detection in underwater sound using the empirical mode decomposition. *IEICE Transactions on Fundamentals of Electronics Communications and Computer*, 89(9): 2415-2421.
[13] Hemminger T. L. and Pao Y. H. (2002). Detection and classification of underwater acoustic transients using neural networks. *IEEE Transactions on Neural Networks*, 5(5): 712-718.
[14] Zhang Y., Li J., Zakharov Y., Li X. and Li J. (2019). Deep learning based underwater acoustic OFDM communications. *Applied Acoustics*, 154(11): 53-58.
[15] Jing S., Hall J., Zheng Y. R. and Xiao C. (2020). Signal detection for underwater IOT devices with long and sparse channels. *IEEE Internet of Things Journal*, 7(8): 6664-6675.
[16] Zhao G., Wang J., Chen W. and Song J. (2019). A novel signal detection algorithm for underwater mimo-OFDM systems based on generalized MMSE. *Journal of Sensors*, 2019(4): 1-10.
[17] Fan C., Yuan X. and Zhang Y. J. (2019). CNN-based signal detection for banded linear systems. *IEEE Transactions on Wireless Communications*, 18(9): 4394-4407.
[18] Cao X., Zhang X., Togneri R. and Yu Y. (2019). Underwater target classification at greater depths using deep neural network with joint multiple-domain feature. *IET Radar, Sonar and Navigation*, 13(3): 484-491.