Deep Learning based Tonal Detection for Passive Sonar Signals

D.-J. Jung (djjung@add.re.kr), J. Park, S.H. Lee and T. Reu
Agency for Defence Development (ADD), South Korea.

Abstract — Passive sonar is used to determine the presence of targets and classify the class of them. Generally, acoustic signals acquired from the passive sonar has been analysed manually by acoustic analysts. However, this approach depends on the personal abilities of acoustic analysts, which makes the analysis process hard and time consuming. To overcome such difficulties, an automatic method is needed for a fast and an accurate analysis. In this paper, we propose an automated tonal detection method which is based on deep learning. The deep neural network is adopted to effectively differentiating between tonal lines and environmental noises. This approach benefits from suppressing noise by picking tonal lines which contain frequency information. The test results showed that the proposed method can be expected to be utilized to develop automatic sonar signal analysis.

1 Introduction

Due to the severe attenuation of radio frequency and optical signals under the water, sonar (Sound Navigation and Ranging) is widely used to navigate, communicate with or detect objects on or under the surface of the water. Sonar systems record the sound waves using hydrophone and process them for detection, location, and classification of targets [1]. Two types (active and passive) of sonar systems are used; however, most of military systems use passive sonar systems for target detection/classification. Figure 1 depicts the brief flow of target classification process using passive sonar system. For the spectrogram analysis, LOFAR (Low Frequency Analysis and Recording) and DEMON (DEModulation Of Noise) are main tools, which utilize sonar signals, to detect the target. DEMON is for narrowband analysis that estimates the propeller characteristics such as number of shafts, shaft rotation frequency, and blade rate [2]. On the other hand, LOFAR provides broadband characteristics including noise vibration of the target machinery [3]. In any case, detecting tonal lines in time-frequency representation (spectrogram including LOFARgram and DEMONgram) is the base step for the target detection.

Generally, the detection of targets using spectrogram has been conducted manually by acoustic analysts. However, the manual approach is time-consuming and depends on the personal abilities of acoustic analysts. To overcome such difficulties, an automated method that can support analysts by processing sonar signals rapid and accurate is needed. However, automated signal analysis of spectrogram is known as a challenging problem due to severe environmental noises and frequency components related to a speed of ships.

In this paper, we propose a tonal detection method for automatic spectrogram analysis based on deep learning which has gained increasing attention for pattern recognition recently. Since tonal lines in spectrogram can be considered as a certain kind of patterns, the deep neural network is adopted to effectively differentiating between tonal lines and environmental noises. Especially, we focused on convolutional neural networks (CNN). The test results showed that the proposed method can be expected to be utilized to develop automatic sonar signal analysis.

The remainder of this paper is organized as follows: Section 2 describes the proposed method; the test results of the proposed method are presented in Section 3; and finally the conclusions are drawn in Section 4.

2 Method

Figure 2 presents overall procedure of our training of convolutional network for the tonal line detection. We used simulator and related parameters to create train/validation data set for CNN. From the sonar simulated wave signals, LOFARgrams were generated and related ground truth images were generated using simulation scenario parameters. Furthermore, the generated dataset was augmented and a certain portion of it was used to train the CNN. Finally, the trained CNN was validated using remaining dataset.

2.1 Simulator and Parameters

To generate the dataset for train/validation, we made a simulator which can output sonar wave signals from a scenario. The controllable parameters for a scenario include environmental noises, target/own-ship parameters (non-speed related components(NSRC) / speed related components(SRC)/trajectory/speed/etc.), and sensor configurations.

For the simplicity of train dataset creation, we used one single target which does not have any moving

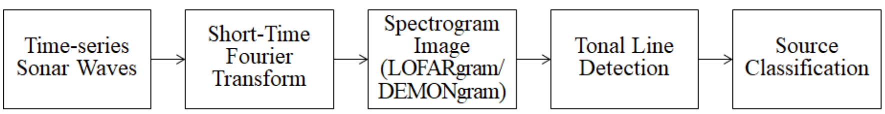

Fig. 1. Target classification process using passive sonar system.

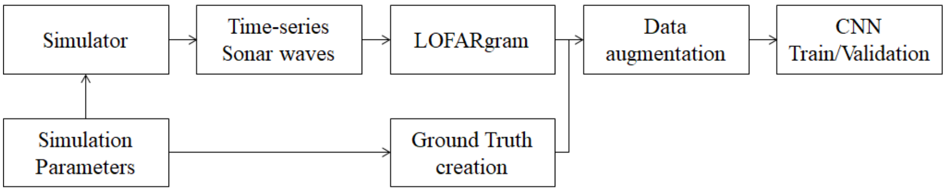

Fig. 2. Proposed procedure for tonal line detection.

trajectory. Only NSRC (including stability control parameter) was used to generate radiated target noises. It benefits from setting frequency consistent along the simulation time. We can make ground truth data easily without the consideration of SRC and Doppler Effect. We randomly set 10 frequencies between 30~200Hz per a ten-minute scenario and 312 scenarios were prepared for the train/validation data creation.

2.2 LOFARgram/Ground Truth and Data augmentation

The simulator generates ten one-minute wave files per a scenario. Using those ten wave files, we created one LOFARgram. For the LOFARgram creation, S3PM was adopted and window size and gap size were 17 and 3, respectively. The resolution of LOFARgram was 0.5Hz and integration time was two seconds. Furthermore, we made ground truth data corresponding each of LOFARgram. Each ground truth data has exactly same size of LOFARgram and consists of ten frequency straight-lines since we had chosen 10 random frequencies in the scenario creation step. Those lines were drawn using frequency data of the corresponding scenario.

We needed an enormous train dataset to avoid overfitting of CNN. The common minimum number of training data is order of 1,000. Since we only had 312 scenarios and they did not include SRC noises, we had to augment the dataset.

For the data augmentation, we utilized Gaussian probability density function written below:

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad (1)$$

where μ and σ represent mean and standard deviation, respectively. We could calculate the frequency-shift values at each time position (row index) in a LOFARgram utilizing the Gaussian probability density function. The frequency-shift function *f(x)* is defined as:

$$f(x) = d * m * g(x) + \mathrm{ms}\,, \quad (2)$$

where $d \in \{-1, 1\}$, m=[20, 40], and ms=[-10, 10], respectively. Every row of LOFARgram was shifted by the output value of frequency-shift function and the holes were filled by sampling environmental noise area. All parameters were selected randomly except $\mu(0)$ and $\sigma(1)$ to give some effects of target's movement. By using data augmentation, we made total 6,240 LOFARgarm-Ground Truth pairs.

2.3 CNN Train/Validation

To differentiate between tonal lines and environmental noises, CNN is adopted. To be more precise, we adopted U-Net which is fully convolutional network for semantic segmentation [4]. U-Net is widely used for image-to-image translation. In the case of tonal line detection, it can be considered as a problem of LOFARgram-to-frequency map translation. Therefore, U-Net is one of the most powerful candidates for the classifier of this problem.

We separated dataset into two classes, train and validation datasets. 2,640 data were randomly drawn for the train dataset, and the others were used for the validation dataset.

3 Test Results

The CNN train/validation was processed in a workstation which has 4 GPUs (NVIDIA Titan XP – 12GB). For better classification results, low frequency regions (0~20Hz) were cropped out. To minimize the effect of data imbalance in training, LOFARgrams and ground truth data were cropped to include minimum area of environmental noises before training. Table 1 enumerates the test results related to detection accuracy and speed. One remarkable point is that some of tonal lines which are almost invisible to naked human eyes are recognized.

Table 1. Test results.

	Train data	Validation data
Precision	0.9959	0.9618
Recall	0.9045	0.9206
Prediction time (sec)	0.3217 (10-minute LOFARgram)	

4 Conclusions

The proposed method achieved very good performance on simulated dataset. The automated sonar tonal detection can help speed and accuracy of analysis for undersea defense since current classification method is based on manual analysis. Our future work includes trying other CNNs in good performance and validation task on real sonar data.

References

[1] R.J. Urick, McGraw-Hill (1983)
[2] J.C.D. Martino, J.P. Haton, A. Laporte, Proceeding of Speech, and Signal Processing, 27-30 (1993)
[3] R.O. Nielsen, Artech Hoise Inc. (1991)
[4] O. Ronneberger, P. Fischer, T. Brox. MICCAI, 9351 (2015)