Effect of Multichannel Signal Sequence on Source Localization Using Convolutional Neural Network

Yinquan Zhang\textsuperscript{1, a}, Shuang Zhang\textsuperscript{2, *}, Kaiming Wu\textsuperscript{3, b}, Siyu Gao\textsuperscript{1, c}, Dong Li\textsuperscript{1, d}, Jie Liu\textsuperscript{2, e}, Guofu Li\textsuperscript{2, f}

\textsuperscript{1}National Marine Data and Information Service, Tianjin, China
\textsuperscript{2}National Ocean Technology Center, Tianjin, China
\textsuperscript{3}Science and Technology on Underwater Acoustic Antagonizing Laboratory, Zhanjiang, China

*Corresponding author e-mail: maymayed2007@126.com, \textsuperscript{a}zhyq\_ndis@126.com, \textsuperscript{b}kaimer2007@sohu.com, \textsuperscript{c}gsy\_ndis@126.com, \textsuperscript{d}lidong\_ndis@163.com, \textsuperscript{e}Liu1968Jie@sina.com, \textsuperscript{f}zhenglimuyun@sina.com

Abstract. In recent years, machine learning has become a promising data-driven method of source localization in underwater acoustics. Several algorithms have been developed by taking advantage of neural networks. This paper investigates the effect of multichannel signal sequence on the performance of source localization using a convolutional neural network (CNN). In this paper, source localization is solved as a classification problem. The performances of different sequences are demonstrated to be quite different. For a specific CNN, it is revealed that the multichannel sequence affects source localization through influencing the complexity of range classification. The complexity can be reasonably reflected by the conspicuousness of signal differences between adjacent range categories. The two-dimensional (2D) Fourier spectrum of the signal differences provides an intuitive approach to describe the conspicuousness. The multichannel sequence that could induce greater spectral amplitudes has better localization performance in noisy environments.

1. Introduction
Source localization in ocean waveguides has always been a research hotspot in the field of underwater acoustics. Numerous approaches to estimate source location have been proposed in the past several decades, including the well-known matched field processing (MFP) method [1-4]. MFP is achieved by calculating replicas using propagation model in a priori environment, and then matching the replicas with experimental results to locate sound source. Therefore, the performance of MFP is affected by the accuracy of the priori environment information. Inaccurate environment information could induce considerable performance degradation. However, accurate environmental information is often unavailable in practice.

In view of the above problem, several data-based methods [5-15] have been proposed to reduce the dependence on environmental information. Among these methods, machine learning is promising and has become a methodological reference in a lot of research fields such as computer vision and speech recognition. In recent years, several machine learning approaches has been developed for source
localization [8-12, 14, 15] by taking advantage of neural networks (NN). These approaches can be generally divided into two categories: sequence independent and sequence dependent. For the first category, the sequence of input neutrons has no effect on the performance of NN, such as feed-forward neural network [9, 10], generalized regression neural network [8]. For the sequence dependent category, as the name suggests, the input sequence contains some implicit factors affecting the performance of NN, such as CNN [15], deep neural network [11, 12].

In a multisensor acoustic observation system, each sensor can be regarded as one channel and the multisensor system obtains a multichannel input. In order to effectively utilize the multichannel input for source localization based on sequence dependent NN, one needs to answer the following question: How does the multichannel input sequence affect the performance of source localization for the sequence dependent NN? To our best knowledge, the question is first posed and investigated in this article.

In this paper, source localization is taken as a classification problem and solved by a specific CNN that has classification function. A horizontal line array (HLA) is adopted to simulate the multisensor acoustic observation system. It is demonstrated that the localization performances of different multichannel input sequences are quite different. In order to reveal how the multichannel sequence affects source localization, input differences between adjacent range categories (IDARC) and corresponding spectrums are investigated by taking advantage of 2D Fourier transform. It is demonstrated that:

1. The 2D spectral amplitudes of IDARC have obvious differences for different input sequences. The sequence that could induce greater spectral amplitudes has better localization performance in noisy environments.

2. The multichannel sequence affects source localization through influencing the complexity of range classification.

3. The complexity of range classification can be reasonably reflected by the conspicuousness of IDARC characteristics. The 2D Fourier spectrum of IDARC provides an intuitive approach to describe the conspicuousness.

2. Methodology
The approach to applying CNN in acoustic source localization is introduced in this section. Multichannel signal model and source range classification are first discussed. Then, the architecture of CNN and source localization algorithm are briefly introduced.

2.1. Multichannel signal model
Consider a single source is emitting a broadband signal in an ocean waveguide and the signal is received by an acoustic observation system with \(N\) sensors. The sound pressure measured by the \(n\)th sensor can be expressed in frequency domain as,

\[
p_n(f) = s(f)g(f,n) + \xi_n(f) , n = 1,2,3,...,N,
\]

where subscript \(n\) denotes sensor number, \(f\) is the signal frequency, \(s\) is the source spectrum, \(g\) is the Green’s function, \(p_n\) and \(\xi_n\) are the sound pressure and complex noise received by the sensor, respectively.

To reduce the effect of source spectrum, the pressure is normalized according to

\[
\bar{p}_n(f) = \frac{p_n(f)}{\sqrt{\sum_{i=1}^{N}|p_i(f)|^2}}
\]

The normalized pressure with multiple frequency components can be formulated by \(\bar{p}_n = [\bar{p}_n(f_1), \bar{p}_n(f_2), ..., \bar{p}_n(f_M)]^T\), \(\bar{p}_n \in \mathbb{C}^{M\times1}\), where superscript \(T\) represents matrix transpose, \(f_1, f_2, ..., f_M\) denote discrete frequencies and \(M\) is the number of frequencies. In this way, the amplitude of \(\bar{p}_n\) is
expressed as $a_0 = [|p_n(f_1)|, |p_n(f_2)|, ..., |p_n(f_M)|]^T$, $a_0 \in \mathbb{R}^{M \times 1}$. Let’s take $a_0$ as one channel. Then, the pressure field received by the multisensor system forms a multichannel signal $a = [a_1, a_2, ..., a_N]$, $a \in \mathbb{R}^{M \times N}$, where each column corresponds to a channel and these channels are arranged in a certain order. In this paper, $a$ is adopted as the input of CNN for source localization. This article aims to study the effect of multichannel sequence on localization performance.

2.2. Source range classification

In previous researches, source localization with NN has been solved as a regression (Ferguson et al. 2016, 2017; Huang et al. 2018; Lefort et al. 2017; Wang and Peng 2018) or a classification (Niu and Gerstoft 2016; Niu et al. 2017a, b) problem. The former treats source location as a continuous variable and the later discretizes source locations into a set of range bins. In this paper, we take source localization as a classification task in convenient for the analyses in the following section.

For the classification problem (Niu et al. 2017b), a set of source locations are discretized into $T$ range bins, $r_1, r_2, ..., r_T$, of equal width $\Delta r$. Each input matrix $a$, is labeled by a vector $y \in \mathbb{R}^{T \times 1}$ defined by,

$$y_t = \begin{cases} 1 & \text{if } |r - r_t| \leq \Delta r / 2 \\ 0 & \text{otherwise} \end{cases}, \ t = 1, 2, ..., T, \hspace{1cm} (3)$$

Where $r$ denotes the source location corresponding to $a$. The label vector is the expected output of the CNN.

2.3. Architecture of CNN

The CNN is composed of an input layer and an output layer, as well as convolution pooling blocks (CPB) and fully connected layers. The architecture of CNN is shown in Figure 1. Here, two CPB are used. Each CPB consists of a convolutional layer and a pooling layer.

At a convolutional layer, the inputs (previous layer’s outputs) are convolved with learnable kernels (filters) and put through the activation function to form the output feature maps. Every kernel performs convolution operation in a certain input region called the receptive field (solid rectangular in Figure 1. And the receptive field moves in stride across the entire input region with the kernel parameters unchanged. The operation of the convolutional layer can be described as,

$$h_k^c = \theta(w_k * v^c + b_k), \ k = 1, 2, ..., K, \hspace{1cm} (4)$$

where subscript $k$ denotes the $k^{th}$ kernel, $w_k$ and $b_k$ are the weights and bias of the $k^{th}$ kernel respectively, $K$ is the number of kernels, $v^c$ represents the inputs of the convolutional layer, $h_k^c$ is the $k^{th}$ output feature map, symbol $*$ denotes convolution operation, $\theta$ is the activation function. In this
paper, each convolutional layer has 40 kernels of size $3 \times 3$ and the stride is 1. Rectified linear units (ReLU) is adopted as the activation function and formulated by,

$$\theta(x) = \max(0, x). \quad (5)$$

Pooling layer is placed after the convolutional layer. The pooling layer takes a sliding window (dash rectangular in Figure 1 that moves in stride across the input, transforming the input into representative values. The transformation is performed by taking the maximum value from the values observable in the sliding window. In this article, the size of the sliding window is $2 \times 2$ and the stride is 2.

Fully connected layers are typically used in the last stages of the CNN to connect the CPB to the output layer. The operation of the fully connected layer can be described as,

$$h^f = \theta(w^f v^f + b^f). \quad (6)$$

where $w^f$ and $b^f$ are the weights and biases of the fully connected layer respectively, $v^f$ represents the inputs of the layer, $h^f$ denotes the outputs of the layer, $\theta$ is the activation function formulated by Eq.5. In this paper, the fully connected layers contain a single hidden layer with 500 neutrons.

In the output layer, softmax function is chosen as the activation function. The softmax function is suitable for multiclass classification problems and expressed as

$$y_t = \frac{\exp(v_t)}{\sum_{j=1}^{T} \exp(v_j)}, \quad t = 1, 2, \ldots, T, \quad (7)$$

Where $v_t$ and $y_t$ denote the input and output values of the $t^{th}$ neutron in the output layer, respectively. The output layer has $T$ neutrons in accordance with the number of range bins in Eq.3. Note that $y_t$ satisfies $\sum_t y_t = 1$ and $0 \leq y_t \leq 1$. The softmax function constrains the output $y_t$ to be the probability that the source is at range $r_t$.

As will be shown in section 3, the size of multichannel signal is $41 \times 41$ and source locations are classified into 140 categories. Accordingly, the input layer dimension of the CNN is $41 \times 41$ and the output layer dimension is $140 \times 1$.

2.4. Source localization algorithm

Similar to Niu's paper (Niu et al. 2017b), the localization problem solved by CNN is implemented as follows:

1. Data preprocessing. The recorded signals are Fourier transformed and the input matrix $A$ of CNN is constructed as described in subsection 2.1.

2. Divide the preprocessed data into training and testing data sets. Prepare data labels according to Eq.3.

3. Train the CNN model by using the training inputs and corresponding labels.

4. Using the model parameters trained in step 3, evaluate the performance of the CNN for the testing data.

3. Simulations and discussion

In this section, numerical simulations are used to study the effect of multichannel sequence on the performance of source localization with CNN.
3.1. Simulation setup

Figure 2 Schematic diagram of the simulated waveguide environment

Figure 3 Sound speed profile of water layer

The schematic diagram of the simulation environment is shown in Figure 2. Acoustic field in the range-independent ocean waveguide is simulated by KRAKEN. The depth of water layer is 128 m and the density is 1 g/cm³. The sound speed profile of water layer is shown in Figure 3, which is measured from an experiment conducted at the continental shelf of the South China Sea. Below the water layer is a fluid half space bottom where the density is 1.7 g/cm³, the attenuation is 0.5 dB/λ, and the sound speed is 1800 m/s.

A horizontal line array (HLA) is used to simulate the multisensor acoustic observation system. As shown in Figure 2, the HLA consists of 41 hydrophones spanning 400 m with uniform inter-sensor spacing 10 m. The depth of HLA is constant at 100 m. A source situated at 40 m depth is assumed to emit broadband signals. The signal bandwidth is 100-300 Hz with frequency resolution 5 Hz. So there are 41 frequency components for each sensor. According to the input model described in subsection 2.1, the received signals of HLA form the input of CNN, $a \in \mathbb{R}^{41 \times 41}$, where each column corresponds to a sensor and is regarded as one channel. Two arrangements of input columns are simulated to study the effect of multichannel sequence on localization performance, as shown in Table 1. For sequence A, the input columns are arranged according to sensor number (see Figure 2). For sequence B, the columns are arranged randomly.
Table 1: Arrangement of input columns for sequence A and sequence B

| Input column | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 |
|--------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| A Sensor number | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 |
| B Sensor number | 24 | 23 | 2  | 5  | 9  | 21 | 26 | 8  | 12 | 37 | 14 | 20 | 32 | 17 |

| Input column | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 |
|--------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| A Sensor number | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 29 | 11 | 33 |
| B Sensor number | 38 | 4  | 10 | 25 | 29 | 41 | 11 | 33 | 16 | 35 | 22 | 27 | 34 | 30 |

| Input column | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 |
|--------------|----|----|----|----|----|----|----|----|----|----|----|----|----|
| A Sensor number | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 |
| B Sensor number | 7  | 13 | 39 | 31 | 6  | 19 | 31 | 28 | 34 | 30 | 15 |

As illustrated in Figure 2, the range between the source and the HLA center varies from 1 km to 8 km with equal interval 10 m. Thus, there are 700 ranges. These ranges are classified into 140 range bins according to Eq.3 with equal width 10 m. At each range, the measured pressure field of HLA is generated by adding a complex white noise to the source signal. In this paper, the source level is assumed constant and the signal-to-noise ratio (SNR) is defined at the most distant range bin (Niu et al. 2017b),

$$SNR = 10\log_{10}\frac{\sum_{m=1}^{M} \sum_{n=1}^{N}|p_{mn}|^2}{\sigma^2},$$

where $p_{mn}$ is the $m^{th}$ frequency component of sound field received by the $n^{th}$ sensor at the longest source-HLA distance, $\sigma^2$ is the noise variance, $M$ and $N$ are the height and width of the input matrix, $M = 41, N = 41$. For each SNR, 4 training samples of noisy measurements are generated at each range and 100 testing samples are constructed in each range bin. Thus, there are a total of 2800 training samples and 14000 testing samples.

With the training samples, both weights and biases of the CNN are initialized randomly and updated by minimizing the cost function between the network output and the expected values, given by

$$E = \sum_{q=1}^{Q_{\text{train}}} \| (y_q - y'_q) \|^2$$

Where $y_q$ is the expected values, $y'_q$ is the network output, $Q_{\text{train}}$ is the number of training samples, “$\| \|$” denotes the 2-norm of vector.

Using the optimized CNN parameters, the performance of source localization is quantified by evaluating the mean accuracy ratio (MAR) over the testing samples, which is defined as

$$MAR = \frac{1}{Q_{\text{test}}} \sum_{q=1}^{Q_{\text{test}}} \delta(y_q - y'_q) \times 100\%,$$

Where $Q_{\text{test}}$ is the number of testing samples and $\delta$ is the Kronecker function expressed as

$$\delta(y - y') = \begin{cases} 1 & y = y' \\ 0 & y \neq y' \end{cases}$$

3.2 Simulation results and discussion

The performance of source localization based on CNN is investigated under 3 SNRs (0 dB, -5 dB, -10 dB).

The aforementioned two input sequences are individually trained. The convergence of the CNN algorithm is demonstrated by plotting the MAR versus optimization steps on training and testing samples.
(see Figure 4). It is shown that the CNN algorithm converges after about 600 steps at all SNRs for sequence A and B. The localization performance of the algorithm after 600 optimization steps is illustrated by evaluating the MAR over the testing samples (see Figure 5). For high SNR (0 dB), both sequences converge to high MAR at every range category, indicating good performance on range prediction. As SNR decreases, the MAR reduces and the CNN generates poor predictions, especially for the far distance source. Obviously, sequence A performs better than sequence B in low SNRs (-5 dB, -10 dB).

![Figure 4](image1.png)

**Figure.4** MAR versus optimization steps on training (solid) and testing data (dash) for sequence A (a, c, e) and B (b, d, f) with SNRs of 0dB (a, b), -5dB (c, d) and -10dB (e, f).

![Figure 5](image2.png)

**Figure.5** MAR of range prediction by CNN on testing data for sequence A (plus sign) and B (circle) with SNRs of 0dB, -5dB and -10dB.

It should be noted that the training and testing samples of sequence A and B are identical, just the multichannel (column) arrangements are different. How does the multichannel sequence affect source localization? In the authors' opinion, the CNN can be regarded as a classifier with learning function; the performance of the CNN is determined by two factors: CNN structure and sample data. The CNN structure determines the classification ability. More neurons and hidden layers generally mean stronger classification ability. The sample data determines the complexity of classification problem. Now that
the CNN structures for the two sequences are identical, the multichannel sequence could affect source localization only through influencing the complexity of range classification. Intuitively, IDARC is reasonable to reflect the distinction between neighboring range categories. If the IDARC is conspicuous, the range classification can be easily solved.

In view of the above analyses, it is assumed that the IDARC characteristics of sequence A are more conspicuous than that of sequence B. To certify this assumption, the IDARC and its spectrum are investigated. Let’s use matrix \( a(t) \) to represent the input signal of the \( t \)th range bin. Then, the IDARC can be expressed as,

\[
\Delta a(t) = a(t) - a(t + 1). \tag{12}
\]

The spectrum of IDARC is obtained by taking advantage of 2D Fourier transformation, described by

\[
\mathcal{F}\{\Delta a(t)\} = \iint \Delta a_{n,f}(t)e^{i(k_x n + k_y f)}dn df \tag{13}
\]

Where \( n \) denotes the column number of input matrix and \( f \) is frequency.

In a noise free environment, the IDARC of sequences A and B at ranges 1 km, 4 km, 7 km are shown in Figure 6. The spectrums of the IDARC are calculated according to Eq.13 and the spectral amplitudes are shown in Figure 7. Since the only difference between sequences A and B is the arrangement of input columns, the IDARC amplitudes for the two sequences are identical (see Figure 6). In contrast, the spectral amplitudes of IDARC for sequence A are obviously greater (see Figure 7), which implies that the IDARC characteristics of sequence A are more conspicuous than that of sequence B. Besides, greater spectral amplitudes imply that the IDARC characteristics are not easy to be covered by white noise. Therefore, range classification for sequence A is easier than sequence B. Sequence A has better localization performance in noisy environments.

![Figure 6](image-url)

Figure 6. Input difference of adjacent ranges (IDAR) for sequence A (a, c, e) and B (b, d, f) at ranges of 1km (a, b), 4km (c, d) and 7km (e, f).
Figure 7 Spectral amplitude of IDAR for sequence A (a, c, e) and B (b, d, f) at ranges of 1km (a, b), 4km (c, d) and 7km (e, f).

Through the analyses of this section, it has been revealed that the multichannel input sequence affects source localization through influencing the conspicuousness of IDARC characteristics. The spectral amplitudes of IDARC provides an intuitive way to describe the conspicuousness. In order to achieve high accuracy of range classification, the multichannel signal should be arranged to make the spectral amplitudes of IDARC as great as possible. It should be noted that the 2D Fourier spectrum is just one approach to describe the conspicuousness of IDARC characteristics. Other potential approaches should be further studied.

4. Conclusion
This paper investigates the effect of multichannel signal sequence on the performance of source localization using a specific CNN. Source localization is taken as a classification problem and the CNN is regarded as a classifier with learning function. Through the study of this article, the performances of range classification for different multichannel sequences are found to be quite different. It is revealed that the multichannel sequence affects source localization through influencing the complexity of range classification. The complexity can be reflected by the conspicuousness of IDARC characteristics. The conspicuousness could be intuitively described by the 2D Fourier spectrum of IDARC. The multichannel sequence that could induce greater spectral amplitudes has better classification accuracy in noisy environments.

This paper utilizes 2D Fourier spectrum of IDARC to evaluate the reasonability of multichannel arrangement for source localization based on CNN. In the future work, the approach should be verified by experiment data. Besides, the validation of this approach needs to be investigated for other sequence dependent neutron networks.

Acknowledgments
This work was funded by Science and Technology Innovation Foundation of National Ocean Technology Center (Z-18047-113); Natural Science Foundation of Tianjin City (18JCQNJC01200); National Key Research and Development Program of China (2016YFC1401800); National Program on Global Change and Air-Sea Interaction of China (GASI-IPOVAI-04).

References
[1] A. B. Baggeroer, W. A. Kuperman and P. N. Mikhalevsky (1993) An overview of matched field methods in ocean acoustics, IEEE J. Ocean. Eng., 18, 401-424.
[2] C. Soares and S. M. Jesus (2003) Broadband matched-field processing: Coherent and incoherent
approaches, J. Acoust. Soc. Am., 113, 2587-2598.

[3] W. Mantzel, J. Romberg and K. Sabra (2012) Compressive matched-field processing, J. Acoust. Soc. Am., 132, 90-102.

[4] A. G. Sazonov and A. I. Malekhanov (2015) Matched field signal processing in underwater sound channels (Review), Acoust. Phys., 61, 213-230.

[5] A. M. Thode (2000) Source ranging with minimal environmental information using a virtual receiver and waveguide invariant theory, J. Acoust. Soc. Am., 108 (4), 1582-1594.

[6] C. Cho, H. C. Song and W. S. Hodgkiss (2016) Robust source-range estimation using the array/waveguide invariant and a vertical array, J. Acoust. Soc. Am., 139 (1), 63-69.

[7] H. C. Song and C. Cho (2017) Array invariant-based source localization in shallow water using a sparse vertical array, J. Acoust. Soc. Am., 141 (1), 183-188.

[8] H. Niu and P. Gerstoft (2016) Source localization in underwater waveguides using machine learning, J. Acoust. Soc. Am., 140, 3232-3242.

[9] H. Niu, E. Ozanich and P. Gerstoft (2017) Source localization in an ocean waveguide using supervised machine learning, J. Acoust. Soc. Am., 142, 1176-1188.

[10] H. Niu, E. Ozanich, and P. Gerstoft (2017) Ship localization in Santa Barbara Channel using machine learning classifiers, J. Acoust. Soc. Am., 142, EL455–EL460.

[11] E. L. Ferguson, R. Ramakrishnan, S. B. Williams and C. T. Jin (2016) Deep learning approach to passive monitoring of the underwater acoustic environment, J. Acoust. Soc. Am., 140, 3351.

[12] Z. Q. Huang, J. Xu, Z. X. Gong, H. B. Wang and Y. H. Yan (2018) Source localization using deep neural networks in a shallow water environment, J. Acoust. Soc. Am., 143, 2922.

[13] R. Lefort, G. Real and A. Dremeau (2017) Direct regressions for underwater acoustic source localization in fluctuating oceans, Appl. Acoust., 116, 303-310.

[14] Y. Wang and H. Peng (2018) Underwater acoustic source localization using generalized regression neural network, J. Acoust. Soc. Am., 143, 2321.

[15] E. L. Ferguson, R. Ramakrishnan, S. B. Williams and C. T. Jin (2017) Convolutional neural networks for passive monitoring of a shallow water environment using a single sensor, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2017, New Orleans, USA, 2657-2661.