Source localization in the deep ocean using a convolutional neural network

Wenxu Liu, Yixin Yang, Mengqian Xu, Liangang Lü, Zongwei Liu, and Yang Shi

Citation: The Journal of the Acoustical Society of America **147**, EL314 (2020); doi: 10.1121/10.0001020
View online: https://doi.org/10.1121/10.0001020
View Table of Contents: https://asa.scitation.org/toc/jas/147/4
Published by the Acoustical Society of America

JASA EXPRESS LETTERS

https://doi.org/10.1121/10.0001020

Source localization in the deep ocean using a convolutional neural network

Wenxu Liu,[1] Yixin Yang,[1,a)] Mengqian Xu,[1] Liangang Lü,[2] Zongwei Liu,[2] and Yang Shi[1]

[1]*School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China*
[2]*First Institute of Oceanography, Ministry of Natural Resources of China, Qingdao 266061, China*
liuwenxu@mail.nwpu.edu.cn, yxyang@nwpu.edu.cn, xxmq0123@163.com, lvlg@fio.org.cn, liuzongwei@fio.org.cn, shiyang_eagle@mail.nwpu.edu.cn

Abstract: In deep-sea source localization, some of the existing methods only estimate the source range, while the others produce large errors in distance estimation when estimating both the range and depth. Here, a convolutional neural network-based method with high accuracy is introduced, in which the source localization problem is solved as a regression problem. The proposed neural network is trained by a normalized acoustic matrix and used to predict the source position. Experimental data from the western Pacific indicate that this method performs satisfactorily: the mean absolute percentage error of the range is 2.10%, while that of the depth is 3.08%.

[Editor: Charles C. Church] Pages: EL314–EL319

Received: 17 January 2020 **Accepted:** 16 March 2020 **Published Online:** 3 April 2020

1. Introduction

In recent years, source localization in the deep ocean has attracted wide attention. However, the traditional matched field processing (MFP)[1] is unsuitable for deep-sea source localization due to the limitation of the array aperture and the effect of environmental mismatch. Usually the deep-sea source localization problem is realized by matching sound field characteristics, such as cross-correlation of two receiver and acoustic intensity striations. However, the cross-correlation matching[2] method is suitable for two synchronous sensors only; plus, the depth estimation error is large. Acoustic intensity striations[3] (versus range and frequency) matching is applied for depth estimation of a moving source only. The normalized acoustic matrix (versus frequency and depth) matching technique[4] was proposed to estimate source range and depth, but the accuracy of the range estimation is poor because the cost function is not sensitive to the source range. The performance of model-based methods[2–4] is always limited by environmental parameter mismatches. Although the source range and depth are estimated by the model-based methods,[2,4] one of them usually has a large estimation error.

More recently, machine learning techniques[5] have been studied to solve the acoustic source localization problem.[6–14] Machine learning is used to estimate the source range[6–8] with better performance than MFP in shallow water. Residual neural networks[9] are used to estimate source range and depth. Although the method locates the broadband source with uncertain bottom parameters using the data of only one sensor, it only works when the signal-to-noise ratio (SNR) is high enough. A convolutional neural network (CNN) can be used to detect and range the broadband noise source,[10,11] but the source range is estimated only within a few hundred meters. These machine learning methods[6–13] only work in shallow water and are not suitable for deep water. Furthermore, most methods[6–8,10–12,14] estimate the source range only. However, for a submerged source, the source depth is also an interesting parameter.

There are few deep-sea source localization methods based on machine learning. The only example is that ensemble convolutional networks[14] can be used to estimate the source range in the deep ocean.

In this paper, a CNN-based source localization method aiming at estimating both the source range and depth accurately in the deep ocean, is proposed. The CNN with strong feature extraction capability is used to solve the problem that the cost function is insensitive to the source range, which leads to a large distance estimation error. The proposed neural network is trained by a normalized acoustic matrix acquired from a non-synchronous vertical array. The dataset is composed of simulated data and measured data to mitigate the impact of the

[a)]Author to whom correspondence should be addressed.

https://doi.org/10.1121/10.0001020

environmental mismatch. Then, the trained network is used to predict the source position. The proposed method is demonstrated on simulated and experimental data.

2. CNN-based localization method

The applications of CNN in underwater source ranging[10,11] demonstrate its fitting ability. A deep CNN is trained to find the connection between the normalized acoustic matrix and source position, and is used to predict the source range and depth. In this section, the normalized acoustic matrix is briefly introduced. Then, the structure of the CNN is shown.

2.1 Normalized acoustic matrix

In the deep ocean, a receiver can receive strong signals emitted by a submerged source within moderate range (30 km) when it is moored below a critical depth (the depth where the speed of sound is equal to that at the sea surface). The received signal is a mixture of multipath signals in which direct and surface-reflected waves play a leading role. The signals from the two paths interfere with each other to form a regular acoustic interference pattern, which is closely related to the source range and depth. More details are shown in Liu *et al.*[4] The acoustic interference pattern can be described by the normalized acoustic matrix versus frequency and depth, which is the training data for the proposed neural network (as illustrated in the left-hand panel of Fig. 1).

The normalized acoustic matrix versus frequency and depth is acquired by a non-synchronous vertical array. The normalized acoustic matrix $\mathbf{M}_{K\times N}$ versus frequency and depth consists of the modulus of complex pressure $P(f_k, z_n)$ at K frequencies and N depth, where f_k is the kth frequency and z_n is the nth receiver depth. Every row of the matrix is set to the range [-1,1]. The row of the matrix is a vector. The row normalization is computed by $\mathbf{A} = \{\mathbf{A} - [(A_{\max} + A_{\min})/2]\}/(A_{\max} - A_{\min})/2$, where $\mathbf{A}$ is the vector, $A_{\max}$ is the maximum of the vector, and $A_{\min}$ is the minimum of the vector. The label is a vector (d, r) composed of the source depth and source range. The final normalized acoustic matrix and label are used to train the CNN.

2.2 The CNN

A deep CNN is designed to acquire the features of the normalized acoustic matrix. Because the effects of source range and depth on the normalized acoustic matrix features are different, the CNN uses filters with different sizes of the local receptive field to learn the relationship between the normalized acoustic matrix and the source position. The size of the filter is also decided by the size of the normalized acoustic matrix. The structure of the CNN is shown in Fig. 1.

The batch normalization layer[15] is used between convolutional layer and Rectified Linear Unit (ReLU) layers to improve the network generalization ability and speed up the training of the CNN. The activation function is ReLU, and can be expressed as $\mathrm{ReLu}(x) = \max(0, x)$. The max-pooling layer is applied once, because the number of vertical array elements is small. The loss function is the half-mean-square error, and can be expressed as

$$L = \frac{1}{2M}\sum_{i=1}^{M}\left[(\hat{r}_i - r_i)^2 + (\hat{d}_i - d_i)^2\right], \tag{1}$$

where M is the number of training data, $\hat{r}$ is the network's prediction for source range, r is the target output of source range, $\hat{d}$ is the network's prediction for source depth, and d is the target output of source depth.

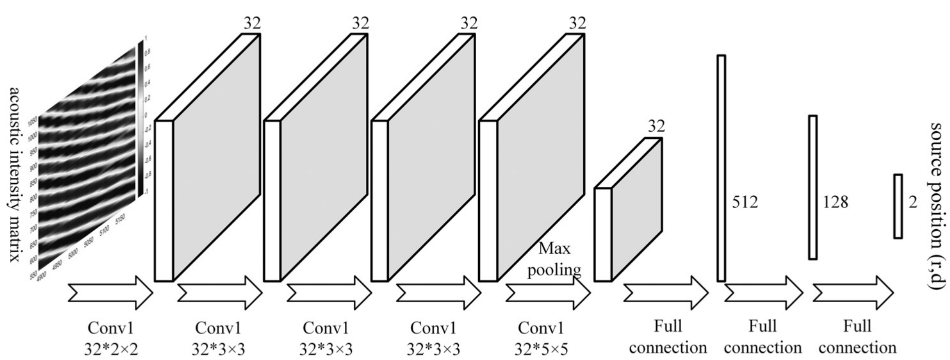


Fig. 1. Structure of the CNN.

https://doi.org/10.1121/10.0001020

3. The dataset: Simulated data and measured data

The dataset is composed of simulated data and measured data. In this section, the simulation, experiment, data processing, and dataset composition are presented.

3.1 Simulated data

Simulations are conducted to generate data. The parameters and signal used in the simulations are the same as those shown in the experiment. The speed of sound as shown in Fig. 2(a) is measured by a conductivity-temperature-depth system. The water depth is about 5300 m, and the bottom of the sea is roughly flat. The 11-element vertical array is positioned below the critical depth, with an element spacing of 33 m, and the depth of the top receiver is 4942 m. In the simulations, the water–air boundary condition is set, and a fluid half-space bottom (speed of sound: 1550 m/s; density: 1.6 g/cm^3; attenuation coefficient: 0.15 dB/λ) is considered. The source depth d is in the range of 10 to 70 m, and the source range r is from 10 to 30 km. This area is divided into grids at a range interval of 0.2 km and depth interval of 0.5 m. The BELLHOP is used to calculate the received complex pressure $P(f_k, z_n)$, where $f_k = 550 + 2k$, $k = 0, 1, \ldots, 250$ and $n = 1, 2, \ldots, 11$. Simulated data without noise are generated. To increase the amount of data, the white noise is added. The discrete Fourier transform of the uniform white noise corresponding to frequency f_k is added to the complex pressure $P(f_k, z_n)$. The SNR is defined as

$$\mathrm{SNR} = 10\lg\left(\sum_{k=1}^{K} |P_{\mathrm{signal}}(f_k)| \Big/ \sum_{k=1}^{K} |P_{\mathrm{noise}}(f_k)|\right), \tag{2}$$

where $|P_{\mathrm{signal}}(f_k)|$ is the modulus of the signal spectrum at f_k frequency, and $|P_{\mathrm{noise}}(f_k)|$ is the modulus of the noise spectrum at f_k frequency. The SNR is 5 dB. Simulated data with noise are generated. The simulated data are evenly distributed in the area of 10–30 km and 10–70 m. After normalization, the data are used to train the CNN.

3.2 Measured data in the experiment

To acquire the measured data, an experiment is performed for the western Pacific in December 2018. The experimental setup is shown in Fig. 2(a) and environmental parameters are the same as described in the simulations. A linear frequency modulation signal is transmitted by a towable transmitter whose duration is 1.8 s and frequency is 550–1050 Hz. The source depth is in the range of 45–55 m. The source moves from 25 to 30 km. A total of 92 pings of time-domain data are obtained. The normalized amplitude and spectrum of one ping of data are shown in

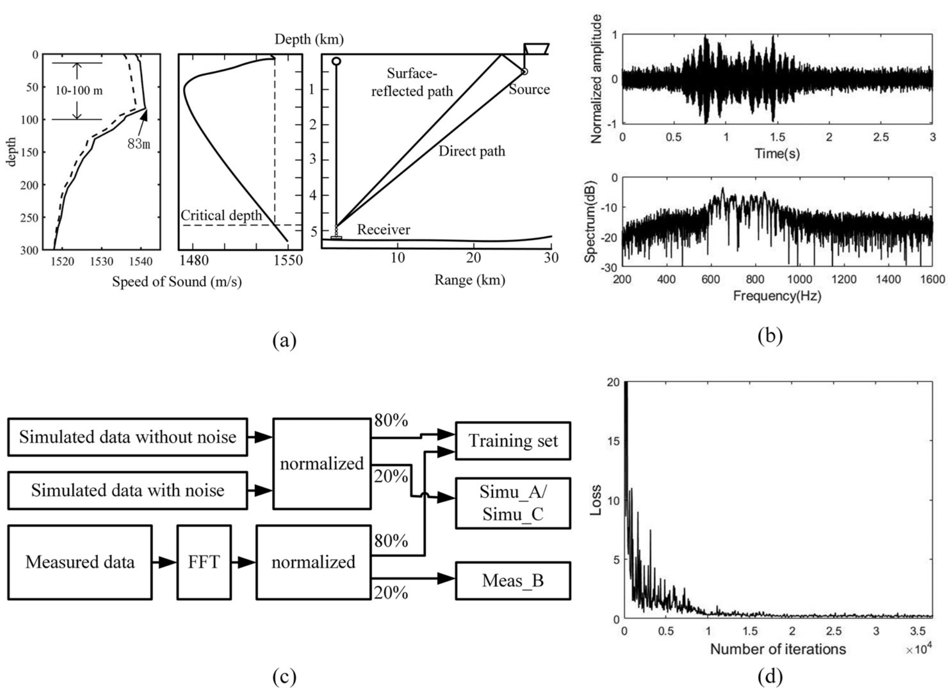


Fig. 2. (a) The solid line is a close-up view of the SSP near the surface, the dotted line is the SSP used to generate Simu_C. (left) The complete SSP (middle), submarine topography and experimental setup (right). (b) Normalized amplitude (top) and spectrum (bottom) of the received signal. (c) The dataset composition and data processing procedure. (d) Training loss.

Fig. 2(b). Different from the simulations, the time domain signals require amplitude compensation to make the energy of each received signal the same. Then, the complex pressure $P(f_k, z_n)$, where $f_k = 550 + 2k$, $k = 0, 1, \ldots, 250$, and $n = 1, 2, \ldots, 11$, is acquired by a discrete Fourier transform of the data received by different depth receivers.

3.3 Dataset composition

To demonstrate the performance of the method, we introduce a dataset containing 24 442 simulated data and 92 measured data. Eighty percent of the simulation data and 80% of the measured data are randomly selected as the training set (total of 19 626 pieces of data). The test set Simu_A is composed of the other 20% of the simulation data (4888 pieces of data), while test set Meas_B is composed of the other 20% of the experimental data (18 pieces of data). Test set Meas_B which includes measured data only, is used to compare the performance with the model-based method in Liu *et al.*[4]

To show the influence of sound speed profile (SSP) on this method, Simu_C is generated in the same way as Simu_A, except that the SSP used is different. As a matter of experience, a linearly decreasing disturbance added to the SSP within 300 m as the dotted line shown in Fig. 2(a) (left). The maximum variation of the SSP is 3 m/s in the sea surface. The dataset composition is shown in Fig. 2(c).

4. Training and results

The optimization algorithm is the Adam method.[16] The initial learning rate is 0.001. The loss of the training process shown in Fig. 2(d) proves that the CNN converges. The CNN is used to predict the result of two test sets.

The mean absolute errors (MAEs), mean absolute percentage errors (MAPEs), and root-mean-square errors (RMSEs) are used to assess the performance of the method

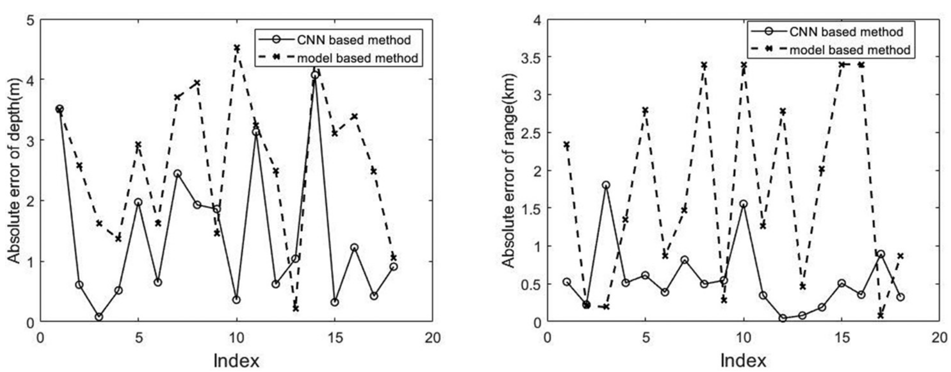
$$\mathrm{MAE}_r = \frac{1}{M}\sum_{i=1}^{M} |\hat{r}_m - r_m|, \quad \mathrm{MAE}_d = \frac{1}{M}\sum_{i=1}^{M} |\hat{d}_m - d_m|, \tag{3}$$

$$\mathrm{MAPE}_r = \frac{100}{M}\sum_{i=1}^{M} \frac{|\hat{r}_m - r_m|}{r_m}, \quad \mathrm{MAPE}_d = \frac{100}{M}\sum_{i=1}^{M} \frac{|\hat{d}_m - d_m|}{d_m}, \tag{4}$$

$$\mathrm{RMSE}_r = \sqrt{\frac{1}{M}\sum_{m=1}^{M} (\hat{r}_m - r_m)^2}, \quad \mathrm{RMSE}_d = \sqrt{\frac{1}{M}\sum_{m=1}^{M} \left(\hat{d}_m - d_m\right)^2}, \tag{5}$$

where M is the number of test data, $\hat{r}_m$ is the estimated source range, r_m is the real source range, $\hat{d}_m$ is the estimated source depth, and d_m is the real source depth.

The results of the proposed method and the model-based method in Liu *et al.*[4] for test set Meas_B, which is composed of experimental data only, are shown in Fig. 3. Figure 3(a) shows the absolute errors of the depth estimated by the two methods, while Fig. 3(b) shows the absolute errors of the range. Comparing the absolute errors of the two methods, we can see that the absolute errors of the proposed method are less than those of the model-based method.



(a) (b)

Fig. 3. (a) Absolute error of source depth estimation for Meas_B. (b) Absolute error of source range estimation for Meas_B.

Table 1. The MAEs, MAPEs, and RMSEs of the proposed method and model-based method.

Method	Proposed	Proposed	Proposed	Model-based
test set	Simu_A	Simu_C	Meas_B[a]	Meas_B[a]
MAE_r (km)	0.323	1.174	0.568	1.698
MAE_d (m)	0.571	0.995	1.428	2.646
$MAPE_r$	1.52%	6.37%	2.10%	6.41%
$MAPE_d$	1.11%	2.21%	3.08%	5.62%
$RMSE_r$ (km)	0.521	1.504	0.724	2.092
$RMSE_d$ (m)	0.777	1.222	1.842	2.896

[a]The data of Meas_B are distributed in the area of 25–30 km and 45–55 m.

The statistical performance is compared using the MAE, MAPE, and RMSE values. The MAEs, MAPEs, and RMSEs of the proposed method and the model-based method are listed in Table 1.

The MAEs, MAPEs, and RMSEs of the proposed method for Simu_A and Meas_B show that this method is robust and estimates the source range and depth accurately. The MAEs, MAPEs, and RMSEs of the proposed method for Simu_C is greater than that for Simu_A, which shows the environmental mismatch affects the accuracy of this method. Comparing the MAEs, MAPEs, and RMSEs of the proposed method with the model-based method for Meas_B, the performance of the proposed method is better. The estimation of depth is accurate in a time-varying environment when the source distance is remote, which makes it difficult to improve the accuracy of range estimation greatly. The performance improvement with respect to range estimation, however, is large. On the one hand, acoustic intensity matrices versus depth and frequency acquired by the simulations and experiment are used to train the CNN, which reduce the localization error caused by environmental mismatch. On the other hand, the CNN obtains more accurate features in acoustic intensity matrices to predict the source position.

The $MAPE_r$ of the proposed method is 2.10% and the $MAPE_d$ of the proposed method is 3.08% for test set Meas_B composed of experimental data only which are distributed in the area of 25–30 km and 45–55 m. The $MAPE_r$ 2.10% of the proposed method is also less than 7.90% of the method in Ref. 14, which estimates the source range only in the deep ocean.

5. Conclusion

In this paper, a broadband submerged source localization method based on a CNN has been studied. The CNN, trained by simulated and measured acoustic intensity matrices, predicts the source range and depth accurately. The results for test set Meas_B, composed of experimental data, demonstrate the robustness and high performance of the method, with a MAPE for range of 2.10%, and a MAPE for depth of 3.08%.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant Nos. 11974286 and 41906160) and the 111 Project under Grant No. B18041.

References and links

[1]A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**(4), 401–424 (1993).

[2]Z. Lei, K. Yang, and Y. Ma, "Passive localization in the deep ocean based on cross-correlation function matching," J. Acoust. Soc. Am. **139**(6), EL196–EL201 (2016).

[3]R. Duan, K. Yang, H. Li, Q. Yang, F. Wu, and Y. Ma, "A performance study of acoustic interference structure applications on source depth estimation in deep water," J. Acoust. Soc. Am. **145**(2), 903–916 (2019).

[4]W. Liu, Y. Yang, L. Lü, Y. Shi, and Z. Liu, "Source localization by matching sound intensity with a vertical array in the deep ocean," J. Acoust. Soc. Am. **146**(6), EL477–EL481 (2019).

[5]M. J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M. A. Roch, S. Gannot, and C. Deledalle, "Machine learning in acoustics: Theory and applications," J. Acoust. Soc. Am. **146**(5), 3590–3628 (2019).

[6]H. Niu, E. Ozanich, and P. Gerstoft, "Ship localization in Santa Barbara channel using machine learning classifiers," J. Acoust. Soc. Am. **142**(5), EL455–EL460 (2017).

[7]H. Niu, E. Reeves, and P. Gerstoft, "Source localization in an ocean waveguide using supervised machine learning," J. Acoust. Soc. Am. **142**(3), 1176–1188 (2017).

[8]J. Chi, X. Li, H. Wang, D. Gao, and P. Gerstoft, "Sound source ranging using a feed-forward neural network trained with fitting-based early stopping," J. Acoust. Soc. Am. **146**(3), EL258–EL264 (2019).

[9]H. Niu, Z. Gong, E. Ozanich, P. Gerstoft, H. Wang, and Z. Li, "Deep-learning source localization using multi-frequency magnitude-only data," J. Acoust. Soc. Am. **146**(1), 211–222 (2019).

[10]E. L. Ferguson, R. Ramakrishnan, S. B. Williams, and C. T. Jin, "Convolutional neural networks for passive monitoring of a shallow water environment using a single sensor," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2017), pp. 2657–2661.
[11]E. L. Ferguson, S. B. Williams, and C. T. Jin, "Sound source localization in a multipath environment using convolutional neural networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2018), pp. 2386–2390.
[12]W. Yun and P. Hua, "Underwater acoustic source localization using generalized regression neural network," J. Acoust. Soc. Am. **143**(4), 2321–2331 (2018).
[13]Z. Huang, J. Xu, Z. Gong, H. Wang, and Y. Yan, "Source localization using deep neural networks in a shallow water environment," J. Acoust. Soc. Am. **143**(5), 2922–2932 (2018).
[14]Y. Liu, H. Niu, and Z. Li, "Source ranging using ensemble convolutional networks in the direct zone of deep water," Chin. Phys. Lett. **36**(4), 044302 (2019).
[15]S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," preprint arXiv:1502.03167 (2015).
[16]D. Kingma and J. Ba, "Adam: A method for stochastic optimization," Computer Science (2014).