sensors

MDPI

Article

Multiple Source Localization in a Shallow Water Waveguide Exploiting Subarray Beamforming and Deep Neural Networks

Zhaoqiong Huang [1,2], Ji Xu [1,2,*], Zaixiao Gong [2,3], Haibin Wang [2,3] and Yonghong Yan [1,2,4]

1 Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China; huangzhaoqiong@hccl.ioa.ac.cn (Z.H.); yanyonghong@hccl.ioa.ac.cn (Y.Y.)
2 University of Chinese Academy of Sciences, Beijing 100049, China; gzx@mail.ioa.ac.cn (Z.G.); whb@mail.ioa.ac.cn (H.W.)
3 State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China
4 Xinjiang Key Laboratory of Minority Speech and Language Information Processing, Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi 830011, China
* Correspondence: xuji@hccl.ioa.ac.cn; Tel.: +86-135-5294-6494

Received: 18 September 2019; Accepted: 30 October 2019; Published: 2 November 2019

Abstract: Deep neural networks (DNNs) have been shown to be effective for single sound source localization in shallow water environments. However, multiple source localization is a more challenging task because of the interactions among multiple acoustic signals. This paper proposes a framework for multiple source localization on underwater horizontal arrays using deep neural networks. The two-stage DNNs are adopted to determine both the directions and ranges of multiple sources successively. A feed-forward neural network is trained for direction finding, while the long short term memory recurrent neural network is used for source ranging. Particularly, in the source ranging stage, we perform subarray beamforming to extract features of sources that are detected by the direction finding stage, because subarray beamforming can enhance the mixed signal to the desired direction while preserving the horizontal-longitudinal correlations of the acoustic field. In this way, a universal model trained in the single-source scenario can be applied to multi-source scenarios with arbitrary numbers of sources. Both simulations and experiments in a range-independent shallow water environment of SWellEx-96 Event S5 are given to demonstrate the effectiveness of the proposed method.

Keywords: multiple source localization; deep neural network; subarray beamforming; shallow water environment

1. Introduction

Multiple source localization in an ocean waveguide is a challenging task because of the interactions among multiple acoustic signals. Several multiple source localization methods have been proposed for tracking underwater targets in past decades. Matched-field processing (MFP) is a classical approach for underwater source localization by correlating the modeled field and the experimental field [1–3]. The range and depth of source are given by the global maximum in the ambiguity surface generated by MFP.

However, the model based methods usually require the environmental parameters to model the acoustic model in advance. Difficulty in obtaining complete knowledge of the real environment may lead to incorrect or inaccurate localization results. To reduce the dependence on environmental information, recently, many data-driven techniques are introduced to source localization in ocean waveguides [4–14]. In previous works, researchers applied deep neural networks (DNNs) to source localization in shallow water environments and obtained promising results [7–14]. However, these studies usually focus on single-source localization. In real-world environments, there are usually multiple sources emerging. Therefore, it is significant to solve the multi-source localization problem in real environments. For a multiple source localization task, several variants of MFP have been proposed through modified Bartlett functions [15,16], maximum likelihood (ML) estimation [17,18], maximum a posteriori (MAP) processors [19], and so forth. Besides, compressive sensing (CS) [20–22] or sparse Bayesian learning (SBL) [23] have been combined with beamforming or MFP to estimate sources' locations in multi-source scenarios. To our best knowledge, there are a few methods that apply DNNs to multiple source localization. In a multi-source scenario, sources tend to emerge in various directions. The directions of sources will be a valuable clue to discriminate multiple sources (the source direction is also represented by source azimuth angle). In this paper, we propose a DNN based method for multiple source localization on underwater horizontal arrays (UHAs).

To apply DNNs to a multiple source localization task, generally, there are two ideas in previous studies. The first idea is to train a single neural network that detects the locations of multiple sources using the mixed signals emitted from various location combinations directly [24–28]. However, training a single network from mixtures to estimate the locations of multiple sources is not an easy task, the reasons of which include—(1) It is hard to traverse all the combinations of source locations with different azimuth angles and ranges (it is supposed that the source location is determined by azimuth angle and range). To get an idea of how much training is required, we consider the two-source scenario for example. We start with training the network with 1° separation of azimuth angles from 0° to 359° (e.g., $(0°, 1°)$, $(1°, 2°)$,..., $(359°, 0°)$). Next we repeat the same procedure with 2° to 180° separations. Assuming the azimuth angle is integer, the combinations of azimuth angle are C_{360}^{2} for two-source scenario. Then we also take the range combinations into consideration, the possible training combinations will be enormous because of the exhaustive training; (2) if we do not separate the mixed signal in advance, the feature for learning is highly correlated with the source combination. Thus the estimation would fail if the test sources' location combination is mismatched with the training set, and the application will be limited. For example, in the two-source scenario with test source one at [125°, 1.2 km] and test source two at [220°, 2.5 km], if this combination does not exist in the training set, the single network (trained for two-source scenario) may fail to give an accurate estimation. Therefore, training the network suitable for various scenarios by mixtures directly is not an optimal scheme.

The second idea tries to simplify the multi-source localization task to single-source localization task. The most popular methods are based on the sparsity assumption on sound source signal [29,30]. Although simultaneous sources overlap in time, if the signal (e.g., speech signal), conforms to be sparsely distributed in the time-frequency (TF) domain, multiple sources will have different distributions in the frequency domain. Hence, this allows training using single-source data and the DNN-based single source localization methods can be conducted on each TF bin. Then, a fusion process is leveraged to integrate the localization results on all TF bins into the spatial information, such as the direction-of-arrivals (DOAs) and the number of multiple sources. However, the underwater sources usually cannot satisfy the sparsity assumption, so this idea is not suitable for our work.

To circumvent these problems, a two-stage DNN based method is proposed to determine both the azimuth angles and ranges of multiple sources successively, which includes a feed-forward neural network (FNN) for direction finding and a long short term memory recurrent neural network [31] (LSTM-RNN) for source ranging. Basically, there are three originalities of our proposed framework. First, in a feature extraction module, we design a subarray beamforming [32] based feature extractor to separate multiple sources at the level of feature, so that the multi-source localization can be simplified

to the single-source localization. Consider the horizontal-longitudinal correlations of the low-frequency acoustic field [33], the UHA is divided into several subarrays and the conventional beamforming (CBF) [34] is conducted on each subarray. The spatial correlation matrix (SCM) of the beamformed signals at all subarrays is taken as the feature. Second, since different sources are discriminated by the features, the multiple sources' ranges can be respectively estimated by the DNN model trained in the single-source scenario. Besides, the LSTN-RNN is adopted to take full advantage of long-term temporal contextual information for the current estimation. Third, an FNN-based direction finding method is presented. A FNN model with a back propagation (BP) algorithm [35] is trained to find the possible directions of sources and determine source number. Then the features of multiple sources can be extracted based on the direction candidates. With subarray beamforming and two-stage DNNs, the need to include multi-source data for training is avoided and the model trained by single-source data can be applied to the multi-source scenarios with arbitrary numbers of sources. In particular, we can localize sources that even overlap fully in the frequency domain.

The rest of the paper is organized as follows. Section 2 formulates the signal model. Section 3 describes the proposed method and each module in detail. Sections 4 and 5 give various simulations and experiments for evaluation. Finally, Section 6 concludes this work.

2. Signal Model

Consider D broadband sound sources impinge on an array of K hydrophones in a far-field scenario, the signal at frequency f_i received by the hydrophones is described as

$$\mathbf{Y}(f_i) = \sum_{d=1}^{D} S_d(f_i)\mathbf{A}(\boldsymbol{\theta}_d, f_i) + \mathbf{N}(f_i), \;\; i \in \{1, \ldots, F\}, \tag{1}$$

where $S_d(f_i)$ denotes the dth signal, $\mathbf{A}(\boldsymbol{\theta}_d, f_i)$ denotes the $K \times 1$ steering vector corresponding to the dth source, $\boldsymbol{\theta}_d$ denotes the DOA of the dth signal, $\mathbf{N}(f_i)$ denotes the noise at the hydrophones, i denotes the frequency index, and F denotes the number of frequency bins. Denote

$$\begin{aligned} \mathbf{H}(\boldsymbol{\theta}_d, f_i) &= \mathbf{A}(\boldsymbol{\theta}_d, f_i)/||\mathbf{A}(\boldsymbol{\theta}_d, f_i)||_2, \\ x_d(f_i) &= S_d(f_i)||\mathbf{A}(\boldsymbol{\theta}_d, f_i)||_2, \end{aligned} \tag{2}$$

Equation (1) can be rewritten using the matrix notation as

$$\mathbf{Y}(f_i) = \mathbf{H}(f_i)\mathbf{X}(f_i) + \mathbf{N}(f_i), \tag{3}$$

where $\mathbf{H}(f_i) = [\mathbf{H}(\boldsymbol{\theta}_1, f_i), \ldots, \mathbf{H}(\boldsymbol{\theta}_D, f_i)]$ is a $K \times D$ steering matrix defining all the potential positions, $\mathbf{H}^H(\boldsymbol{\theta}_d, f_i)\mathbf{H}(\boldsymbol{\theta}_d, f_i) = 1$, $\mathbf{X}(f_i) = [x_1(f_i), \ldots, x_D(f_i)]^T$ is a $D \times 1$ dimensional vector denoting the signal, $(\cdot)^H$ denotes the Hermitian transpose, and $(\cdot)^T$ denotes the transpose.

The DOA $\boldsymbol{\theta}_d$ is represented by the azimuth angle α_d and the grazing angle β_d,

$$\boldsymbol{\theta}_d = [\cos\alpha_d \cos\beta_d, \sin\alpha_d \cos\beta_d, \sin\beta_d]^T. \tag{4}$$

The geometrical relationship of the DOA ($\boldsymbol{\theta}$) and the azimuth angle (α) and the grazing angle (β) is shown in Figure 1. For horizontal array, the grazing angle of propagation is small in the far-field scenario ($\beta < 20°$) [36], that is, $\cos\beta \approx 1$. Therefore, the steering vector depends mainly on the azimuth angle α. For simplicity, $\boldsymbol{\theta}_d$ is approximated to $[\cos\alpha_d, \sin\alpha_d, 0]^T$ in the following process.

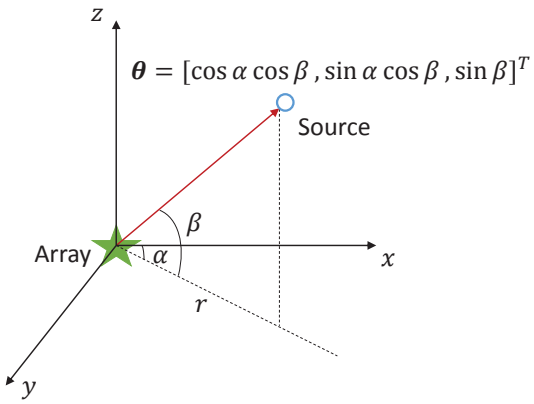

Figure 1. Geometrical relationship of direction of arrival (DOA) ($\boldsymbol{\theta}$) and the azimuth angle (α) and the grazing angle (β). A horizontal array is deployed at the xy plane. The horizontal distance between source and array is r km.

3. Proposed Method

The block diagram of the proposed method is shown in Figure 2. In the training stage, the features are extracted from the single source signal radiated from different locations by performing subarray beamforming and calculating the SCM of the beamformed signals at all subarrays. Then DNN-2 is trained to model the regression relationship between the extracted feature and the source range. In the testing stage, the azimuth angles of sources are firstly estimated by DNN-1. The features of sources are extracted based on all azimuth angle candidates at subarrays. Finally, the range of each source is inferred by feeding the feature associated with each source to DNN-2.

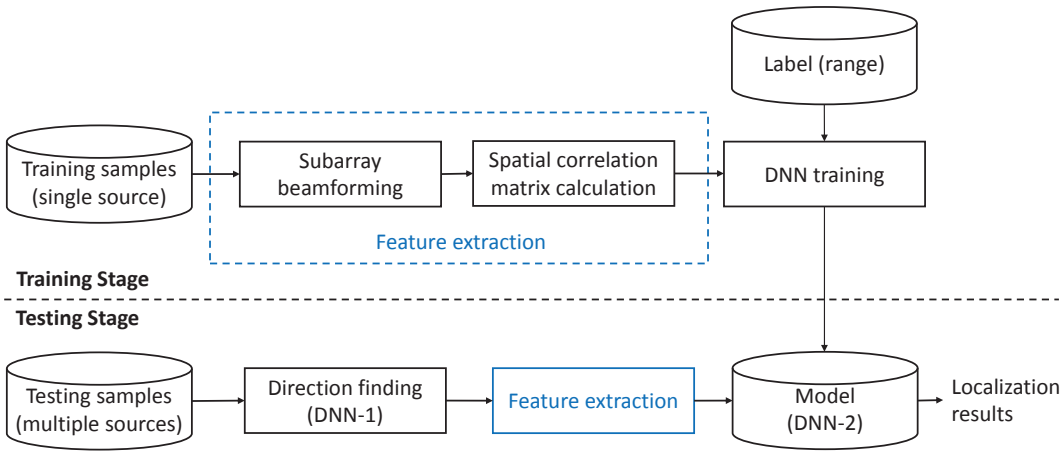

Figure 2. Block diagram of the proposed method.

3.1. Direction Finding

Rstogi et al. proposed using the hopfield network [37] in direction finding [38]. The basic idea is to use a neural network to find the best possible choice of directions present in the received signal through minimizing a quadratic cost function. Compared to the conventional neural network, DNN with a BP algorithm has a stronger capability for finding the good solutions to a difficult optimization problem. However, there are few methods that apply DNNs to direction finding in the ocean environments. In this paper, we attempt to get desirable results of sources' directions using a FNN. The configuration of FNN (i.e., DNN-1 in Figure 2) is shown in Figure 3, where the projection from the input vector $\boldsymbol{\nu}_\iota$ at the ιth layer to the output vector $\boldsymbol{\nu}_{\iota+1}$ at the $(\iota+1)$th layer is represented as

$$\boldsymbol{\nu}_{\iota+1} = \mathbf{W}_\iota \boldsymbol{\nu}_\iota + \mathbf{b}_\iota, \tag{5}$$

where $\mathbf{W}_\iota$ and $\mathbf{b}_\iota$ denote the weight and bias matrix from the ιth layer to the $(\iota+1)$th layer. The feature of DNN-1 is the FFT coefficients of the observed signal $\mathbf{Y}$. The real and imaginary part of FFT coefficients are concatenated as the input of DNN-1. Denote $\mathbf{H}(\boldsymbol{\theta}_d, f_i) = \left[1,\ e^{j2\pi f_i \tau_2},\ e^{j2\pi f_i \tau_3}, \cdots, e^{j2\pi f_i \tau_K}\right]^T$ (τ_k is the time delay between the kth hydrophone and the first hydrophone), which is the steering vector of the dth source, the cost function for the broadband case can be expressed as

$$\Lambda = \frac{1}{L \times F} \sum_{l=1}^{L} \sum_{i=1}^{F} \left|\left| \mathbf{Y}_l(f_i) - \left[\boldsymbol{\Gamma}_{f,1}\mathbf{Y}_l(f_i)\ \cdots\ \boldsymbol{\Gamma}_{f,P}\mathbf{Y}_l(f_i)\right]\mathbf{z} \right|\right|^2, \tag{6}$$

where $\boldsymbol{\Gamma}_{f,p} = \mathbf{H}(\boldsymbol{\theta}_d, f_i)\left[\mathbf{H}^H(\boldsymbol{\theta}_d, f_i)\mathbf{H}(\boldsymbol{\theta}_d, f_i)\right]\mathbf{H}^H(\boldsymbol{\theta}_d, f_i)$, L denotes the the snapshot number and $\mathbf{z} = [z_1, z_2, \cdots, z_P]^T$ $\left(z_p \in [0,1]\right)$ is the output vector of the neural network. $\boldsymbol{\Gamma}_{f,1}\mathbf{Y}_l(f_i)$ is the $K \times 1$ dimensional vector of the observed signal projected onto the steering vector $\mathbf{H}(\boldsymbol{\theta}_d, f_i)$. The cost function will be minimized by the best linear combination of the steering vectors, when convergence, the extremums in vector $\mathbf{z}$ indicate the possible sources.

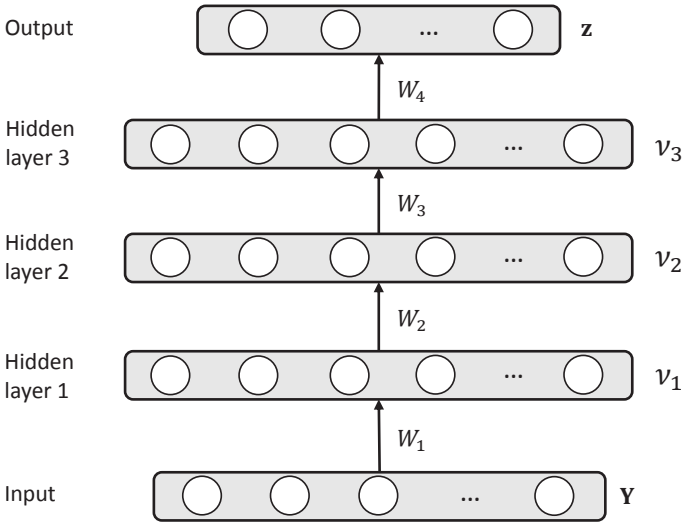

Figure 3. The architecture of FNN/DNN-1.

Each significant peak of vector $\mathbf{z}$ is identified as a sound source, the probability of which is greater than the threshold,

$$\delta = O_{avg} + \eta(O_{max} - O_{avg}), \tag{7}$$

where O_{avg} and O_{max} denotes the average and maximum of the smoothed probabilities, and the coefficient η $(0 < \eta < 1)$ is set by experiment.

Note that only in the testing stage is the FNN using BP algorithm trained to find the directions that sound sources may emerge. For each direction candidate, we extract the corresponding features, then the sources' ranges are estimated by feeding the features into DNN-2 (i.e., LSTM-RNN).

3.2. Source Ranging

To avoid the exhaustive training, we aim to train a general and flexible model that is suitable for situations with different source numbers. Thus, how to design an effective feature, which can be used for various scenarios, is a critical problem. For DNN analysis, the more similar the test set is to the training set, the better the testing result will be. However, in our task, the training set is composed by the single-source signals at different locations while only the mixture is available when testing. It is vital to extract a feature that can represent each single source information from the mixture, so that the test signal (or feature) can be matched with the training signals. Beamforming, which can

enhance the signal from the desired direction while attenuating others, is ideal to extract the individual signal component from the mixture. Nevertheless, if we perform beamforming using all sensors, the horizontal-longitudinal correlations of the acoustic field, which include the spatial information of source, will lost in the enhanced signal. Therefore, we introduce subarray beamforming to extract the individual source component, meanwhile preserving the horizontal-longitudinal correlations. The SCM of the enhanced signals at all subarrays is used as the feature.

3.2.1. Feature Extraction

Beamforming algorithms can be used to track those interested sources and null out the other sources as interference by controlling the beampattern of an array. The simplest beamforming technique is adopted in our framework, which refers to the delay-and-sum beamforming. It delays the multi-channel signals so that all versions of the source signal are time-aligned before they are summed. To preserve the horizontal-longitudinal correlations of the low-frequency acoustic field, this CBF is conducted on each subarray. The hydrophone array is divided to B subarrays, $\left\{\Omega_1, \ldots, \Omega_B\right\}$, then the signal enhanced to the dth direction at the bth subarray is obtained by applying CBF to the signals received by the hydrophones in the bth subarray,

$$\begin{aligned} g_b^d(f_i) &= \sum_{k \in \Omega_b} Y_k(f_i) e^{-j2\pi f_i \tau_{k,d}}, \\ \tau_{k,d} &= \ell_k \boldsymbol{\gamma}_k^T \boldsymbol{\theta}_d / c, \end{aligned} \tag{8}$$

where $\tau_{k,d}$ denotes the dth time delay of the kth hydrophone corresponding to the first hydrophone at the bth subarray (the first hydrophone is chosen as the reference), ℓ_k and $\boldsymbol{\gamma}_k^T$ denote the distance and the unit directional vector between the kth hydrophone and the reference hydrophone, Ω_b denotes the hydrophone index set of the bth subarray, c denotes the sound speed and $j = \sqrt{-1}$ denotes the imaginary unit. The enhanced signals of the dth source at frequency f_i obtained by all subarrays are given by $\mathbf{G}_d(f_i) = \left[g_1^d(f_i), \ldots, g_B^d(f_i)\right]^T$. The block diagram of subarray beamforming is shown in Figure 4.

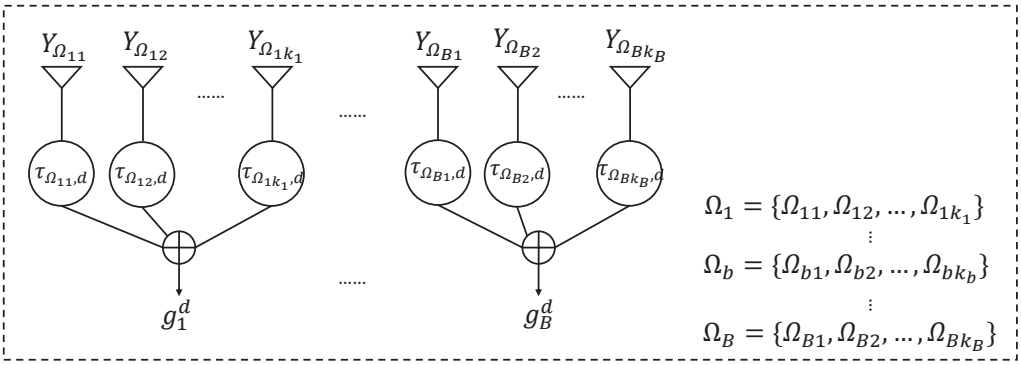

Figure 4. Block diagram of subarray beamforming.

The SCM of the signals enhanced to each source direction is used as the feature, because it contains sufficient information about the individual signal. The SCM of the dth source is calculated by

$$\mathbf{R}_d(f_i) = E[\tilde{\mathbf{G}}_d(f_i)\tilde{\mathbf{G}}_d^H(f_i)], \tag{9}$$

where $\tilde{\mathbf{G}}_d(f_i) = \mathbf{G}_d(f_i)/||\mathbf{G}_d(f_i)||$. The real and imaginary part of the upper triangular matrix of the SCM is concatenated as a $B \times (B+1)$ dimensional vector denoted by $\mathbf{u}_d$, which is used as the input feature of the neural network.

3.2.2. DNN Analysis with LSTM-RNN

DNN [39] is a data-driven technique that learns the potential patterns from the original acoustic data directly. Due to the movement of the source, we take source localization to be a regression task. In the regression problem, the target output $r \in (0, \infty)$ is a continuous range variable. For the source localization task, current range of a source is considered to be related to its adjacent locations. However, FNN, or time delay neural network [40] (TDNN), can provide only limited temporal modeling by splicing fixed frames of features in the input or hidden layers. By contrast, RNNs contain cycles that feed the network activations from a previous time step as inputs to the network to influence predictions at the current time step, so the more sufficient long-term temporal contextual information can be used. In particular, LSTM architecture [31] overcomes the gradients vanishing and exploding existing in traditional RNNs by introducing some special units called memory blocks. Therefore, we adopt LSTM-RNN to model the mapping between the feature and source range in our framework.

The deep LSTM-RNN is shown in Figure 5a, and the configuration of LSTM memory blocks is shown in Figure 5b, where the input and output vectors are denoted as $\boldsymbol{u} = (\mathbf{u}_1, \cdots, \mathbf{u}_T)$ and $\boldsymbol{v} = (\mathbf{v}_1, \cdots, \mathbf{v}_T)$. The configuration of LSTM memory blocks that unfolded across time (the yellow dashed box in Figure 5a) is shown in Figure 6. The memory block contains several self-parameterized controlling gates, i.e., input gate, output gate, and forget gate, to control the flow of information. The input gate controls the flow of input activations into the memory cell. The output gate controls the output flow of cell activations into the rest of the network. Finally, the forget gate is added to forget or reset the cell's memory adaptively.

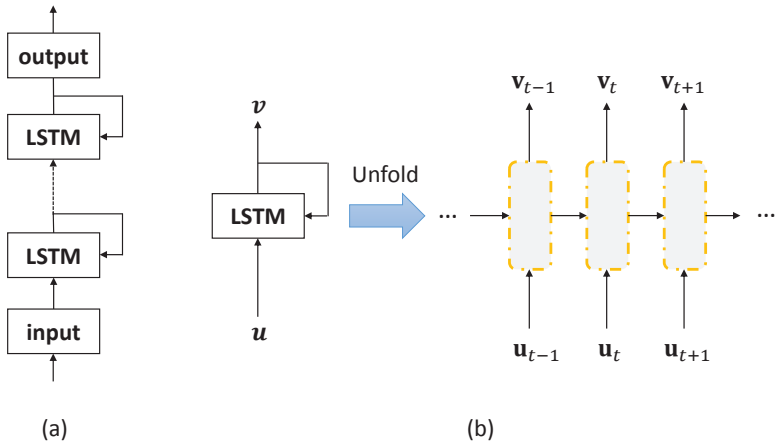

Figure 5. The configuration of LSTM-RNN. (**a**) The deep LSTM-RNN; (**b**) The configuration of LSTM memory blocks that unfolded across time.

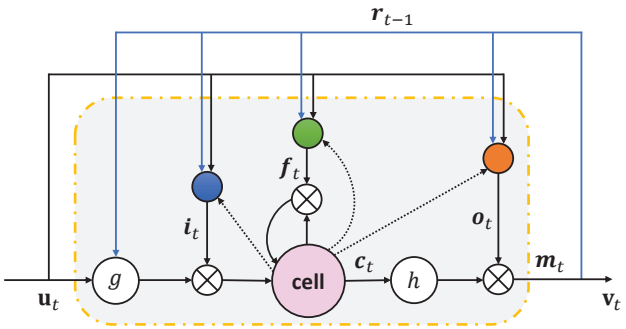

Figure 6. The configuration of LSTM memory block.

The associated computations that map the input vector to the output vector are given as follows:

$$\mathbf{i}_t = \sigma(\mathbf{W}_{iu}\mathbf{u}_t + \mathbf{W}_{im}\mathbf{m}_{t-1} + \mathbf{W}_{ic}\mathbf{c}_{t-1} + \mathbf{b}_i) \tag{10}$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_{fu}\mathbf{u}_t + \mathbf{W}_{fm}\mathbf{m}_{t-1} + \mathbf{W}_{fc}\mathbf{c}_{t-1} + \mathbf{b}_f) \tag{11}$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot g(\mathbf{W}_{cu}\mathbf{u}_t + \mathbf{W}_{cm}\mathbf{m}_{t-1} + \mathbf{b}_c) \tag{12}$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_{ou}\mathbf{u}_t + \mathbf{W}_{om}\mathbf{m}_{t-1} + \mathbf{W}_{oc}\mathbf{c}_t + \mathbf{b}_o) \tag{13}$$

$$\mathbf{m}_t = \mathbf{o}_t \odot h(c_t) \tag{14}$$

$$\mathbf{v}_t = \mathbf{m}_t \tag{15}$$

where **i**, **f**, **o**, **c**, **m** denote the input gate, forget gate, output gate, cell activation, and cell output activation vectors respectively, **W** terms denote the weight matrices, in which $\mathbf{W}_{ic}$, $\mathbf{W}_{fc}$, $\mathbf{W}_{oc}$ are diagonal weight matrices for peephole connections (the dotted lines from cell to gates in Figure 6), **b** denotes the bias matrices, σ denotes the sigmoid activation function, $\odot$ denotes the element-wise product, g and h are the cell input and cell output activation functions that are *tanh* in this paper.

The cost function is defined as the mean square error (MSE) between the estimated source range r_q and the reference source range $\hat{r}_q$, given by

$$E = \frac{1}{Q}\sum_{q=1}^{Q}(r_q - \hat{r}_q)^2, \tag{16}$$

where Q denotes the sample number. We use the truncated back propagation through time (BPTT) learning algorithm [41] to update the parameters.

3.2.3. Data Augmentation

In our framework, the two-stage DNNs are used to determine both the azimuth angles and ranges of multiple sources. In source ranging stage, we need azimuth angles that estimated by DNN-1 to perform feature extraction for DNN-2. The accuracy of estimated source range by DNN-2 is not only determined by DNN-2, but also the feature extracted based on the estimated results of DNN-1. Therefore, if the azimuth angles are inaccurately estimated by DNN-1, the features generated based on the deviant azimuth angles may lead to differences from the correct features (i.e., the result of subarray beamforming using the estimated azimuth angle α is different from that using the true azimuth angle $\hat{\alpha}$). Therefore, the error introduced by direction finding may cause the inaccurate estimation of source range. To reduce the negative effect of direction finding on source ranging and improve the generalization ability of DNN-2, we introduce some disturbances in feature extraction and the disturbed features are merged to the training set in the training stage of DNN-2. This strategy is called data augmentation [42–44] (which is widely used in speech recognition or speech enhancement). The original data, denoted as Φ, are disturbed during feature extraction stage to obtain the augmented features, denoted as Ψ. Explicitly, for each sample in Φ, we obtained the augmented feature $\boldsymbol{u}_\zeta^\kappa$ (where the superscript κ denotes the sample index in Φ) by introducing an offset angle α_ζ to the true azimuth angle $\hat{\alpha}$. The augmented beamformed signal calculated by the disturbed azimuth angle $\alpha^{'} = \hat{\alpha} + \alpha_\zeta$ is obtained by modifying Equation (8) as

$$\begin{aligned} g_b^{'}(f_i) &= \sum_{k\in\Omega} Y_k(f_i)e^{-j2\pi f_i \tau_k^{'}}, \\ \tau_k^{'} &= \ell_k \boldsymbol{\gamma}_k^T \boldsymbol{\theta}^{'}/c, \end{aligned} \tag{17}$$

where $\boldsymbol{\theta}^{'} = [\cos\alpha^{'}, \sin\alpha^{'}, 0]^T$. The augmented feature $\boldsymbol{u}_\zeta^\kappa$ is obtained by calculating the SCM of the augmented signals at all subarrays, $\mathbf{G}^{'}(f_i) = \left[g_1^{'}(f_i), \ldots, g_B^{'}(f_i)\right]^T$. The data augmentation process is detailed as Table 1 (Algorithm 1), where ϑ limits the range of angle offset and ϑ_o is the step size.

Table 1. Algorithm 1: data augmentation process.

Input: original data Φ;
Output: augmented training set Ψ;
Set $\Psi = \varnothing$;
For each sample $\mathbf{Y}^{\kappa}(f_i)$ in Φ **do**
  For offset $\alpha_{\zeta} = -\vartheta : \vartheta_o : \vartheta$ **do**
    Add α_{ζ} to the true azimuth $\hat{\alpha}$, $\alpha^{'} = \hat{\alpha} + \alpha_{\zeta}$;
    Generate the beamformed signals using Equation (17);
    Generate feature $\boldsymbol{u}_{\zeta}^{\kappa}$ using Equation (9);
    $\Psi = \Psi \bigcup \boldsymbol{u}_{\zeta}^{\kappa}$;
  End
End

4. Simulations

4.1. Acoustic Environmental Model

To investigate the performance of the proposed method, we simulated the relatively range independent SWellEx-96 Event S5 [45] environment. The sound speed profile (SSP) and geoacoustic parameters for SWellEx-96 Event S5 are shown in Figure 7. The seafloor is composed first of a 23.5 m thick sediment layer with a density of 1.76 g/cm^3 and an attenuation of 0.2 dB/kmHz. The top and bottom sound speeds are 1572.368 m/s and 1593.016 m/s. Below the sediment layer is an 800 m thick mudstone layer with a density of 2.06 g/cm^3 and an attenuation of 0.06 dB/kmHz. The top and bottom sound speeds of the mudstone layer are 1881 m/s and 3245 m/s. The geoacoustic model is completed by a halfspace with a density of 2.66 g/cm^3, an attenuation of 0.02 dB/kmHz, and a compressional sound speed of 5200 m/s.

4.2. Data Description

In the simulation, the bandwidth of signal was [50, 210] Hz and the sampling rate was 3276.8 Hz. The hydrophone array was deployed at a 213 m depth of water. We investigated two topologies of UHAs, including a horizontal circular array (HCA) and a horizontal line array (HLA) (note that our method is suitable for UHA with arbitrary topologies). The HCA was 50-element with a 250 m radius, where the hydrophones were uniformly distributed. The HLA was 27-element, the layout of which was the same as that of the HLA North of SWellEx-96 Event S5 (the details can refer to the web page http://swellex96.ucsd.edu/hla_north.htm). In fact, the line array was not strictly linear but had a certain degree of curvature. The map of source movement and the location of the hydrophone array are depicted in Figure 8. The training data included sources with azimuth angles from 0° to 180° with 5° intervals (the course equals to azimuth angle). In each azimuth angle, the source ranged from 1.0 to 5.6 km at a speed of 5 knots (2.5 m/s). The source depth was fixed to 54 m. When testing, every testing segmentation contained ten minutes (including 960 samples) and the two-source scenario included source one from [64.7°, 2.05 km] to [66.9°, 3.59 km], and source two from [115.6°, 1.95 km] to [113.6°, 3.49 km]. The three-source scenario included source one from [64.7°, 2.05 km] to [66.9°, 3.59 km], source two from [115.6°, 1.95 km] to [113.6°, 3.49 km], and source three from [173.3°, 2.00 km] to [174.9°, 3.54 km]. The training data and testing data were mutually different.

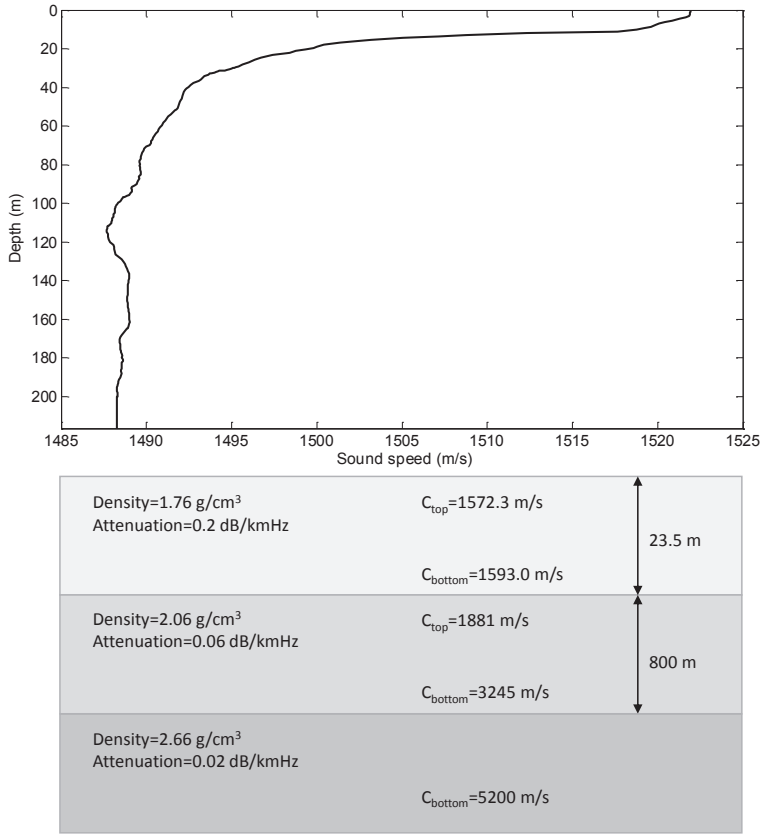

Figure 7. Waveguide with sound speed profile and geoacoustic parameters for range-independent SWellEx-96 Event S5.

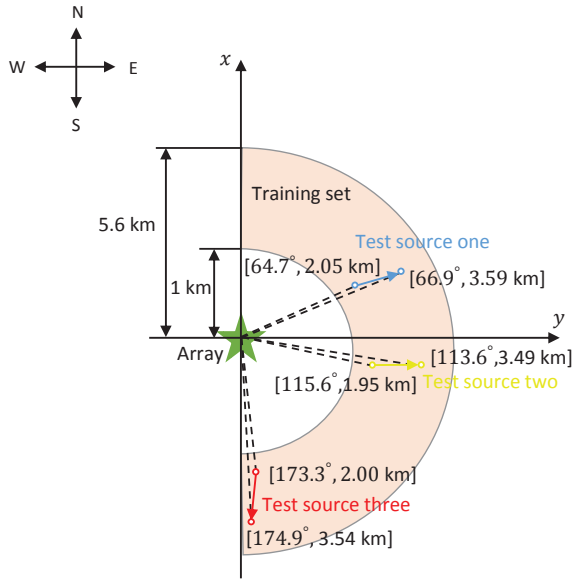

Figure 8. The map of source movement and the location of the hydrophone array in the simulation. The semi-annular orange region covers the ranges of training sources' motions. The training data included sources with azimuth angles from 0° to 180° with 5° intervals (the course equals to azimuth angle). In each azimuth angle, the source was ranging from 1.0 to 5.6 km at a speed of 5 knots (2.5 m/s). The blue, yellow, and red lines were the trajectories of test source one, two, and three. The array includes two topologies, including HCA and HLA. The HCA was 50-element with a 250 m radius, where the hydrophones were uniformly distributed. The HLA was 27-element, the layout of which is the same as the HLA North of SWellEx-96 Event S5.

The signal was transformed to the frequency domain by operating fast Fourier transformation (FFT) (Hanning windowed). The frame length was 1.25 s with 50% overlap. The bandwidth for processing was set to $[100, 200]$ Hz (with 5 Hz increment, totally 21 frequency bins). For HCA, the 50 hydrophones were divided into five subarrays uniformly, that is, $\Omega_1 = \{1, \cdots, 10\}$, $\Omega_2 = \{11, \cdots, 20\}$, $\Omega_3 = \{21, \cdots, 30\}$, $\Omega_4 = \{31, \cdots, 40\}$, and $\Omega_5 = \{41, \cdots, 50\}$. For HLA, the 27 hydrophones were divided into four subarrays, the hydrophone indexes of subarrays were $\Omega_1 = \{1, \cdots, 7\}$, $\Omega_2 = \{8, \cdots, 14\}$, $\Omega_3 = \{15, \cdots, 21\}$, and $\Omega_4 = \{22, \cdots, 27\}$. Twenty snapshots were used to calculate the SCM. Data augmentation was performed using $\vartheta = 7^\circ$ and $\vartheta_o = 0.5^\circ$, generating about 3.1×10^6 training samples.

4.3. The Configuration of DNNs

In direction finding, the configuration of FNN was 5 layers (one input layer + three hidden layers + one output layer) with 128 hidden nodes. The rectified linear units [46] (ReLU), $f(x) = max(0, x)$, was used as the activation function. The initial learning rate was 0.001 and the batch size was 6. The input of FNN was the FFT coefficients of each frame, so the input dimension of FNN were 1134 ($27 \times 2 \times 21$, real and imaginary parts were concatenated) for HLA and 2100 ($50 \times 2 \times 21$) for HCA.

In source ranging, the LSTM-RNN was three layers with 896 nodes. The activation function was ReLU. The initial learning rate was 0.001 and the batch size was 512. The input dimension of LSTM-RNN were 420 ($4 \times 5 \times 21$) for HLA and 630 ($5 \times 6 \times 21$) for HCA.

It should be mentioned that all parameters (e.g., hidden nodes, hidden layers, learning rate, and batch size) of FNN or LSTM-RNN were chosen based on experiments. The tensorflow [47] toolkit was taken for FNN and LSTM-RNN training. Adam [48] was utilized for optimization.

4.4. Metrics

4.4.1. Direction Finding

For direction finding, the detected sources were classified into two categories, namely the correctly detected sources and the incorrectly detected sources. The detection was considered to be correct if the estimated azimuth angle deviated no more than 7° from the real azimuth angle of any source. The incorrectly detected sources consisted of the imaginary sources (detected but non-existing sources) and the inaccurately detected sources. The detection correctness was mainly evaluated in terms of the positive detection rate (PDR) (i.e., the ratio of the number of correctly detected sources to the total number of sources) and the false detection rate (FDR) (i.e., the ratio of the number of incorrectly detected sources to the total number of sources). The receiver operating performance characteristics (ROC) curve gave a complete description of the relationship between PDR and FDR with the change of threshold η (0 to 0.95 with 0.05 steps). Define

$$\eta_o = \min_{\eta} |1 - \mathrm{PDR}(\eta) + \mathrm{FDR}(\eta)|, \tag{18}$$

the mean absolute error (MAE) between the true azimuth angles and the estimated azimuth angles of correctly detected sources when $\eta = \eta_o$ was combined with ROC curve to evaluate the performance in direction finding stage. The MAE between the true azimuth angles ($\hat{\alpha}$) and the estimated azimuth angles (α) is defined as

$$\mathrm{MAE}_\alpha = \frac{1}{\Xi} \sum_{\xi=1}^{\Xi} \min_{d \in \{1,\ldots,D\}} \mathcal{F}(\alpha_\xi - \hat{\alpha}_{\xi,d}), \tag{19}$$

where $\mathcal{F}(\alpha)$ is denoted as

$$\mathcal{F}(\alpha) = \min_{n} \left|\alpha + 360^\circ \times n\right|, \tag{20}$$

where n is an integer denoting the number of azimuth period, $\mathcal{F}(\alpha) \in [0, 180^\circ]$, and Ξ denotes the number of estimation results and ξ is the sample index.

4.4.2. Source Ranging

The objective evaluation metrics used for source ranging were the MAE and the mean relative error (MRE) between the estimated ranges (r) and the true ranges ($\hat{r}$),

$$\mathrm{MAE}_r = \frac{1}{\Xi}\sum_{\xi=1}^{\Xi}|r_\xi - \hat{r}_\xi|, \tag{21}$$

$$\mathrm{MRE}_r = \frac{1}{\Xi}\sum_{\xi=1}^{\Xi}\frac{|r_\xi - \hat{r}_\xi|}{\hat{r}_\xi} \times 100\%. \tag{22}$$

4.5. Simulation Results

The first simulation was conducted to investigate the performance of the proposed method under different signal-to-noise ratios (SNRs). White noise was added to the simulated signals, resulting in SNRs of 15, 5, and -5 dB. The SNR [49] reported here was defined as the SNR (at 210 Hz) at a single hydrophone when the source range was 1 km (SNR would decrease with source range increasing). Both source level (SL) and noise level (NL) were attenuated by -6 dB/Oct. The CBF [34] was chosen as the competing algorithms in direction finding. Twenty snapshots were used to calculate beamformer power of CBF. For the sake of fairness, the posterior probability of FNN was averaged over every twenty frames. The results of the two-source scenarios and three-source scenarios on HCA are summarized in Table 2. The ROC curves of two-source scenario and three-source scenario are plotted in Figures 9 and 10 (The SNR shown here is the SNR of the received signal for each source, and the SL of each source is assumed to be equal). It should be mentioned that, the number of points seen on the figures may be less than the number of points actually sampled, because (1) there are some η correspond to the same PDR and FDR and they are overlapped in the figures; (2) there are some points of CBF go out of scope because of the large FDR when η is small. From the ROC curves, although the performance degrades with the lower SNR, the FNN and CBF can detect sources effectively in general. Superficially, the three methods can give a high PDR with a low FDR by setting an appropriate threshold; however, the values of η_o of CBF are larger than FNN significantly. The smaller η_o implies the stronger ability of suppressing the interference. Thus, there are little phantom peaks of FNN than CBF, which is a good indication of its better capability of suppressing interference. When SNR decreases to -5 dB, the FDR of CBF rises and PDR decreases, which reveals the proposed method is more robust than CBF under a lower SNR. Furthermore, the estimation errors of FNN are smaller than CBF in all conditions as shown in Table 2.

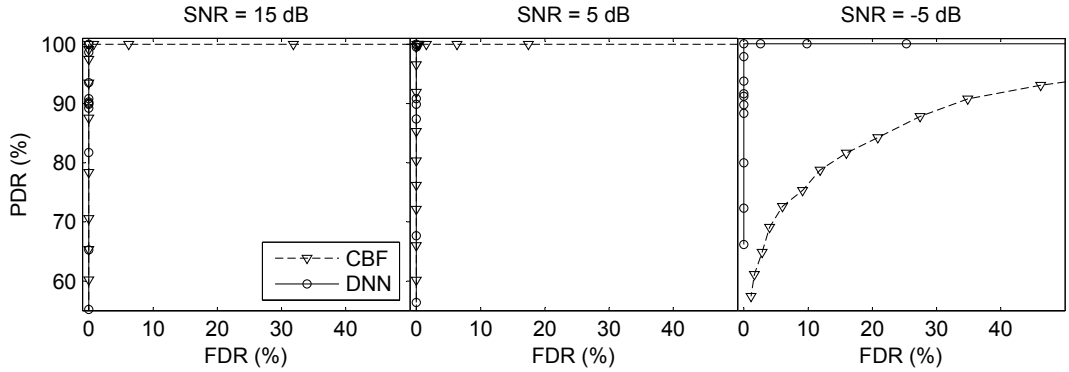

Figure 9. ROC curves of direction finding on HCA under different SNRs in the two-source scenarios.

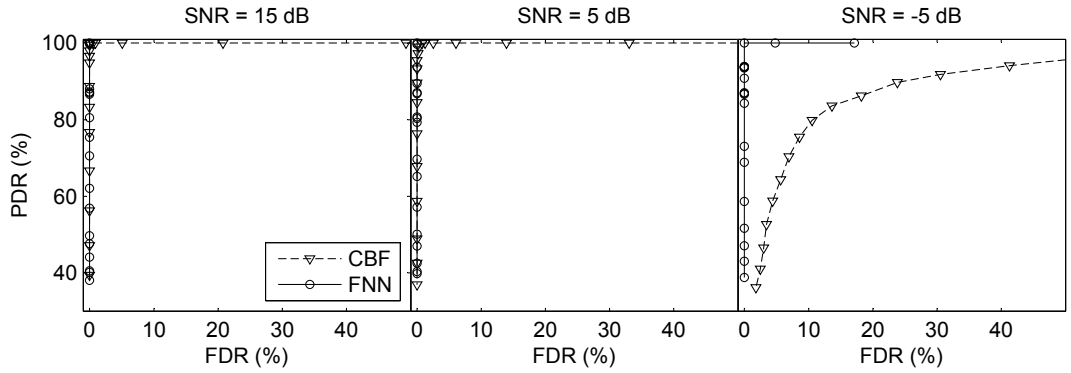

Figure 10. ROC curves of direction finding on HCA under different SNRs in the three-source scenarios.

Table 2. The performance comparison under different SNRs in the two-source and three-source scenarios using the simulated data on HCA.

	SNR (dB)	Method	η_o	MAE_α (degree)	PDR (%)	FDR (%)	MAE_r (km)	MRE_r (%)
Two sources	15	FNN+LSTM-RNN	**0.1**	**0.24**	**100.0**	**0.0**	**0.08**	**3.2**
		FNN+FNN					0.43	16.0
		CBF	0.25	0.26	**100.0**	**0.0**	—	—
	5	FNN+LSTM-RNN	**0.1**	**0.24**	**100.0**	**0.0**	**0.09**	**3.4**
		FNN+FNN					0.57	22.2
		CBF	0.45	0.28	100.0	0.05	—	—
	−5	FNN+LSTM-RNN	**0.2**	**0.25**	**100.0**	**0.0**	**0.59**	**21.3**
		FNN+FNN					0.76	28.7
		CBF	0.55	0.53	82.4	20.8	—	—
Three sources	15	FNN+LSTM-RNN	**0.1**	**0.25**	**100.0**	**0.0**	**0.18**	**7.0**
		FNN+FNN					0.66	25.6
		CBF	0.3	0.29	**100.0**	**0.0**	—	—
	5	FNN+LSTM-RNN	**0.1**	**0.25**	**100.0**	**0.0**	**0.32**	**12.0**
		FNN+FNN					0.71	27.7
		CBF	0.35	0.31	99.9	0.7	—	—
	−5	FNN+LSTM-RNN	**0.1**	**0.27**	**100.0**	**0.0**	**0.74**	**28.8**
		FNN+FNN					0.81	31.9
		CBF	0.5	0.66	83.6	13.6	—	—

For source ranging, we compared the performance of LSTM-RNN with FNN. The FNN was five layers with three hidden layers and 896 hidden nodes. From Table 2, the LSTM-RNN outperforms FNN, which demonstrates the superiority of LSTM-RNN in modeling the long-term temporal information. In addition, we may notice that the locations of the test sources may not exist in the training set. However, the proposed method can still give reliable estimates to sources' ranges, which reveals that the proposed method can localize the sources as long as the test source locations are in the region of the training set.

We also evaluated the performance on HLA under different SNRs. The results are summarized in Table 3. We can find that the proposed method also exhibits a good performance on direction finding and source ranging on HLA. Comparing Tables 2 and 3, basically, the performance of the proposed method is similar to different array topologies. Whereas the MAE_α of HLA is larger than HCA, the reason of which considers the angular resolution of HCA is constant with the change of azimuth angles while it varies for HLA. The experimental results indicate that the proposed method can be applied to the UHA with arbitrary topologies. For simplicity, the following simulations were all conducted on HCA.

Table 3. The performance comparison under different SNRs in the two-source and three-source scenarios using the simulated data on HLA.

	SNR (dB)	Method	η_o	MAE_α (degree)	PDR (%)	FDR (%)	MAE_r (km)	MRE_r (%)
Two sources	15	FNN+LSTM-RNN	**0.15**	**1.37**	100.0	1.5	**0.04**	**1.6**
		FNN+FNN					0.51	19.9
		CBF	0.6	1.79	**100.0**	**0.0**	—	—
	5	FNN+LSTM-RNN	**0.1**	**1.39**	**100.0**	**0.0**	**0.06**	**2.1**
		FNN+FNN					0.53	20.5
		CBF	0.7	1.79	100.0	2.3	—	—
	−5	FNN+LSTM-RNN	**0.1**	**1.49**	**100.0**	**0.0**	**0.67**	**25.8**
		FNN+FNN					0.71	27.7
		CBF	0.9	1.69	99.4	2.6	—	—
Three sources	15	FNN+LSTM-RNN	**0.1**	**1.52**	100.0	0.2	**0.15**	**5.8**
		FNN+FNN					1.00	38.5
		CBF	0.65	2.06	**100.0**	**0.0**	—	—
	5	FNN+LSTM-RNN	**0.1**	**1.55**	**100.0**	**0.0**	**0.22**	**8.1**
		FNN+FNN					0.98	38.8
		CBF	0.7	2.06	100.0	0.0	—	—
	−5	FNN+LSTM-RNN	**0.1**	**1.61**	**100.0**	**0.0**	**0.65**	**24.0**
		FNN+FNN					1.04	38.9
		CBF	0.9	2.00	99.6	4.7	—	—

The second simulation evaluated the performance with or without data augmentation in the two-source scenario. The SNR was set to 5 dB and the neural network was LSTM-RNN. The MAE_r and MRE_r without data augmentation are 0.56 km and 20.9%. From Table 3, with data augmentation, the MAE_r and MRE_r drop to 0.09 km and 3.4% respectively. The results demonstrate that data augmentation can improve the generalization ability of DNN model.

The third simulation was made to investigate the performance of the proposed method when the SLs of two testing sources were different, where the source with the higher SL referred to the dominant source. The SNR of the dominant source was 5 dB. Define $\Delta SL = SL_1 - SL_2$ (dB) (SL_1 corresponds to the dominant source and SL_2 corresponds to the weak source), Figure 11 compares the ROC curves of CBF and FNN when $\Delta SL = 2,\ 4,\ 6$ dB. Both methods can give high PDR with low FDR when two SLs are comparable. Nevertheless, the false detections of CBF rise faster than FNN when the difference between the two SLs increases. In addition, the MAE_r and MRE_r of source ranging are summarized in Table 4. With ΔSL increasing, the estimation error increases because the weak source is masked by the presence of the dominant source, which leads to the larger error of the weak source.

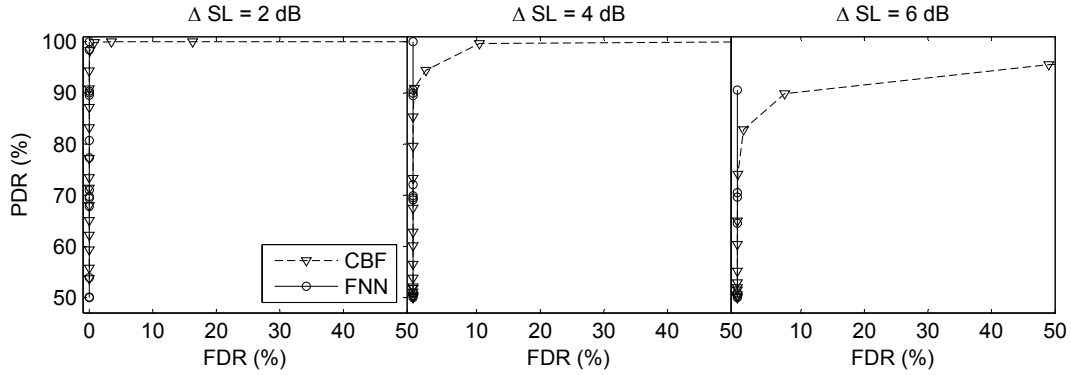

Figure 11. ROC curves of direction finding when the SLs of two testing sources are different in the two-source scenarios. The SNR of the dominant source was 5 dB. $\Delta SL = SL_1 - SL_2$ (dB) (SL_1 corresponds to the dominant source and SL_2 corresponds to the weak source).

Table 4. MAE_r and MRE_r comparison when two SLs are different on HCA.

ΔSL (dB)	MAE_r (km)	MRE_r (%)
2	0.19	6.7
4	0.27	9.8
6	0.32	12.4

The last experiment investigated the spatial resolution of the proposed method. The separations of two sources were set to 2°, 3°, 5°, 7°, and 10°. Here, the azimuth of each source was fixed, while the range of each source was from 1 km to 2.5 km. The SNR was set to 5 dB. The detection accuracies of FNN and CBF in direction finding are shown in Figure 12. Here, only when the source number and the azimuth angles of two sources are estimated correctly is the detection deemed to be correct. The accuracy is defined as the ratio of the number of accurate detections and the number of test samples. From Figure 12, generally, FNN and CBF can discriminate two widely separated sources, and the accuracy of FNN outperforms CBF. When the separation of two sources becomes smaller, FNN presents its superiority in discriminating two closely separated sources. We evaluated the performance of source ranging using LSTM-RNN. The results of source ranging are summarized in Table 5, where the MAE_r and MRE_r are calculated using the test samples with the accurate estimated azimuth angles. The results show that the separations have little influence on source ranging if the azimuth angles are estimated accurately. Note that the MAE_r and MRE_r are slightly smaller than those shown in Table 2, because the range of testing sources here are nearer than those in the first simulation.

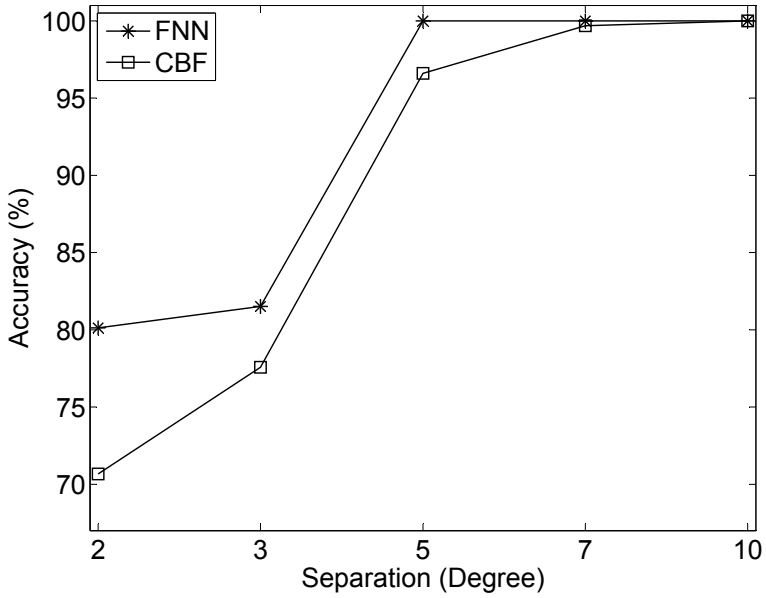

Figure 12. Detection accuracies of FNN and CBF. The detection is deemed to be correct only when the source number and the azimuths of two sources are estimated correctly. The accuracy is defined as the ratio of the number of accurate detections and the number of test samples.

Table 5. MAE_r and MRE_r comparison under different source separations on HCA.

Separation (Degree)	MAE_r (km)	MRE_r (%)
2	0.03	1.4
3	0.04	1.9
5	0.03	2.1
7	0.03	2.1
10	0.03	2.2

5. Experiments

5.1. Experimental Database

The proposed method was further evaluated by real experimental data that were recorded by HLA North of SWellEx-96 Event S5. The water depth was 213 m and the HLA North array is a 240 m aperture horizontal array deployed on the seafloor. The source ship (R/V Sproul) started its track south of the array and proceeded northward at a speed of 5 knots. The signals of the deep source were used for processing. The map of the source movement and the location of the hydrophone array were shown in Figure 13. There were fifty minute signals from J131 23:40 GMT to J132 00:30 GMT that were recorded by HLA North (Day J131 corresponds to 5/10/96). The range and azimuth angle motions between source and array were plotted in Figure 14. To imitate the multi-source signals (i.e., a snapshot generated by several sources), we combined snapshots from the same source recorded at different positions. As a result, the NL of the resultant multi-source signal was higher than that in the original recordings, that is, the SNR was reduced when increasing the source number.

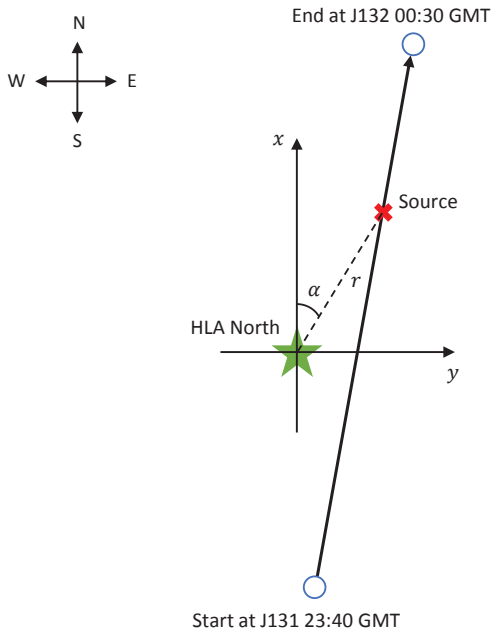

Figure 13. Map of the source movement and the location of the hydrophone array.

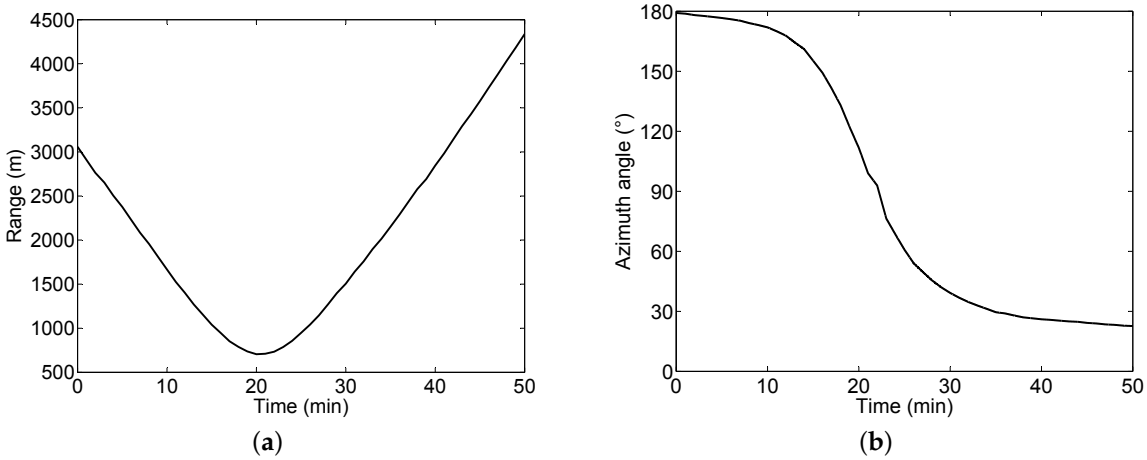

Figure 14. The ranges (**a**) and azimuth angles (**b**) between source and array from J131 23:40 GMT to J132 00:30 GMT.

The experimental data with sample rate 3276.8 Hz were transformed to frequency by 4096-point FFT (Hanning windowed). The frame length was 1.25 s and the SCMs were averaged over 20 snapshots with 50% overlap. Considering the Doppler effect, processing frequencies were selected from three frequency bins centered on each of the nominal source frequencies. Accordingly, there were $3 \times F$ processing frequency bins if we took F source frequencies into account. Referring to Doppler Shift theory, the maximum Doppler shift is $\triangle f = \pm\frac{2.5}{1500} f_i = \pm 1.7 \times 10^{-3} f_i$ (f_i is the source frequency), which corresponds to ± 0.083 to ± 0.66 Hz for the pilot tones. Similar to Section 4.2, data augmentation is used to generate the training set (refer to Algorithm 1, $\vartheta = 7^\circ$ and $\vartheta_o = 0.5^\circ$).

5.2. Experimental Results

Firstly, we investigated the performance of our proposed method using different frequency bins in the two-source scenarios. The two-source signals were the combination of snapshots from J131 23:47 GMT to J131 23:53 GMT and snapshots from J132 00:19 GMT to J132 00:25 GMT, which were six minutes in total. Three frequency bin sets were investigated, which were $\{49\ 64\ 79\ 94\ 112\ 130\ 148\ 166\ 201\ 235\ 283\ 338\ 388\}$ Hz, $\{94\ 112\ 130\ 148\ 166\ 201\ 235\ 283\ 338\ 388\}$ Hz, and $\{49\ 94\ 148\ 235\ 283\ 338\}$ Hz (i.e., 3×13, 3×10, and 3×6 frequency bins used for processing because of Doppler shift). The parameters of DNNs in direction finding and source ranging were set the same as those in the simulations, while the input dimensions were slightly different from those in the simulation because of the difference in the number of frequency bins.

In direction finding, the ROC curves are plotted in Figure 15. The results show that the proposed direction finding method outperforms CBF significantly. The FNN can detect more sources effectively while having lower false detection relative to CBF. Also, the lower threshold η_o means the strong ability to suppress interferences. As there are more phantom peaks of CBF, its FDRs are much higher than FNN. The MAE_α, MAE_r, MRE_r, the corresponding η_o, PDR and FDR are summarized in Table 6. The proposed method achieves the best performance in all conditions. Besides, the source range estimates across time are plotted in Figure 16, where the results using the three sets of frequency bins are respectively shown in Figure 16a–c. We can see that the proposed method can give reliable estimates of the range of two sources successively, although the performance degrades with reduction of the frequency bins.

Table 6. The performance comparison with different frequency bins in the two-source and three-source scenarios using the real experimental data.

	Frequency (Hz)	Method	η_o	MAE_α (degree)	PDR (%)	FDR (%)	MAE_r (km)	MRE_r (%)
Two sources	$\{49\ 64\ 79\ 94\ 112\ 130\ 148\ 166\ 201\ 235\ 283\ 338\ 388\}$	FNN+LSTM-RNN	0.1	**2.74**	**100.0**	**0.0**	**0.11**	**5.0**
		FNN+FNN					0.14	5.6
		CBF	0.3	3.49	90.1	12.3	—	—
	$\{94\ 112\ 130\ 148\ 166\ 201\ 235\ 283\ 338\ 388\}$	FNN+LSTM-RNN	0.1	**3.32**	**100.0**	**0.0**	**0.13**	**5.4**
		FNN+FNN					0.18	7.9
		CBF	0.25	3.34	89.6	15.4	—	—
	$\{49\ 94\ 148\ 235\ 283\ 338\}$	FNN+LSTM-RNN	0.2	**3.35**	**95.9**	**3.0**	**0.15**	**6.7**
		FNN+FNN					0.24	10.2
		CBF	0.3	3.44	86.4	17.1	—	—
Three sources	$\{49\ 64\ 79\ 94\ 112\ 130\ 148\ 166\ 201\ 235\ 283\ 338\ 388\}$	FNN+LSTM-RNN	0.1	**3.34**	**89.4**	**0.0**	**0.36**	**15.6**
		FNN+FNN					0.47	23.7
		CBF	0.15	3.42	79.2	22.0	—	—
	$\{94\ 112\ 130\ 148\ 166\ 201\ 235\ 283\ 338\ 388\}$	FNN+LSTM-RNN	0.1	3.84	**78.1**	**1.0**	**0.34**	**14.0**
		FNN+FNN					0.55	25.2
		CBF	0.15	**3.24**	71.2	26.8	—	—
	$\{49\ 94\ 148\ 235\ 283\ 338\}$	FNN+LSTM-RNN	0.1	3.57	**89.2**	**10.9**	**0.41**	**19.3**
		FNN+FNN					0.51	24.8
		CBF	0.15	**3.55**	82.3	22.0	—	—

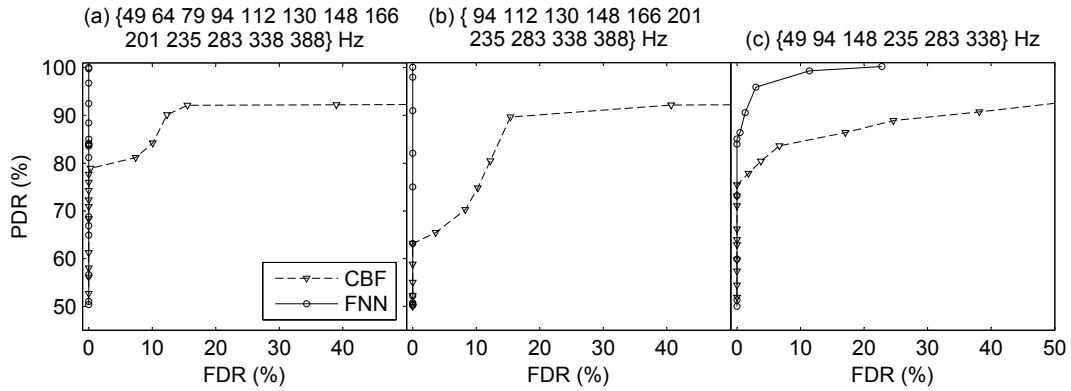

Figure 15. ROC curves of direction finding using different frequency bins in the two-source scenario using the real experimental data.

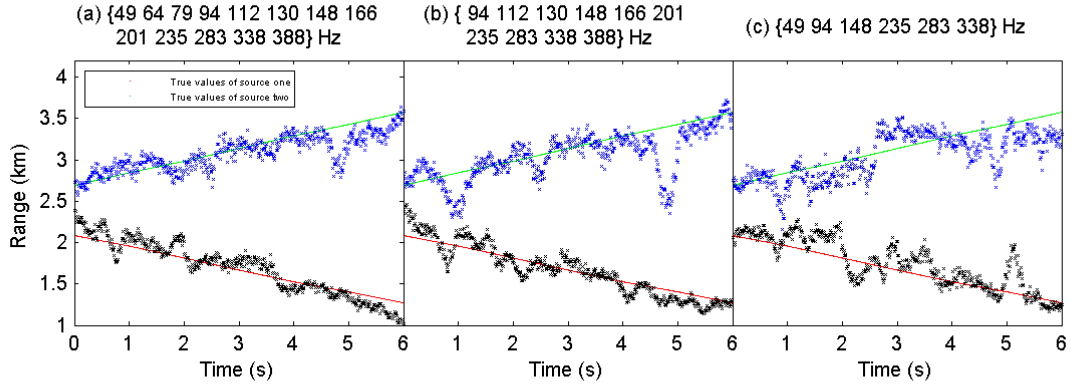

Figure 16. The source rang estimates across time using different frequency bins in the two-source scenario using the real experimental data.

To demonstrate that LSTM-RNN can make full advantage of the long-term temporal contextual information, we compared the FNN with LSTM-RNN for source ranging. The results are also shown in Table 6. It can be seen that the LSTM-RNN outperforms FNN, especially when the number of frequency bins decreases. The results reveal the superiority of LSTM-RNN on modeling the long-term information.

Next, we investigated the influence of the parameters of LSTM-RNN on the performance of source ranging. Thirteen source frequencies were used (39 bins). The hidden layers were changed from 2 to 4, the hidden nodes were set to 512, 896, and 1024, and the learning rates were chosen from 0.0005, 0.001, and 0.002. The testing results are summarized in Table 7. The best results were achieved by the network with 3 hidden layer, 896 hidden nodes, and learning rate 0.001. From the results, generally, the change in parameters has little influence on the performance of source ranging.

Finally, we evaluated the proposed method on the three-source scenario. The three-source signals contained six minutes that were combined by snapshots from J131 23:47 GMT to J131 23:53 GMT, snapshots from J132 00:07 GMT to J132 00:13 GMT and snapshots from J132 00:23 GMT to J132 00:29 GMT. The ROC curves are plotted in Figure 17 and the MAE_α, MAE_r, MRE_r, and the corresponding PDR and FDR are summarized in Table 6. The threshold η_o is the same as the two-source scenario. From the results, we can find the proposed method generally outperforms the competing methods. Also, the LSTM-RNN exhibits a more robust performance than FNN.

Table 7. MAE_r and MRE_r comparison with different parameters of LSTM-RNN using the experimental data.

Parameter			MAE_r	MRE_r
Hidden Layer	**Hidden Node**	**Learning Rate**		
3	512	0.001	0.16	6.4%
3	896	0.001	**0.11**	**5.0**%
3	1024	0.001	0.14	5.9%
3	896	0.0005	0.13	5.2%
3	896	0.002	0.13	5.2%
2	896	0.001	0.12	5.0%
4	896	0.001	0.13	5.4%

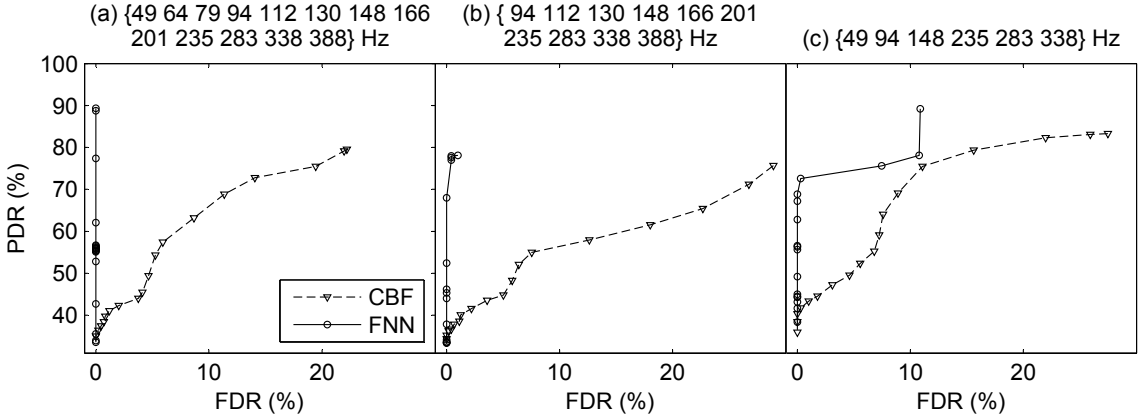

Figure 17. ROC curves of direction finding using different frequency bins in the three-source scenario using the real experimental data.

6. Conclusions

This paper presents a two-stage DNN based method for multiple source localization in a shallow water environment using UHA. We attempt to train a general and flexible model using single-source signals that is suitable for source ranging in various scenarios with different source numbers. The subarray beamforming technique is taken as the feature extractor that separate sources at the level of feature and LSTM-RNN is leveraged for source ranging. Since the subarray beamforming requires the direction information to be known beforehand, a FNN model is trained for direction finding, meanwhile determine the source number. Both the simulation and experimental results demonstrate the effectiveness and superiority of the proposed framework. As LSTM-RNN can make full use of long-term temporal contextual information for the current estimation, it is an ideal model for source ranging. Our method can localize arbitrary numbers of sources that overlap in the TF domain. In our future work, we will make further efforts to improve the robustness of the proposed method in the more complex environments with lower SNRs and more sources.

Author Contributions: Z.H., J.X., and Z.G. contributed to the idea of this paper and designed the algorithms and simulations; Z.H. was responsible for performing the experiments and dealt with the data. Z.H., J.X., Z.G., H.W., and Y.Y. analyzed the simulation and experimental results. Z.H., J.X., and Z.G. contributed with the structure, content and the paper check. All of the authors were involved in writing the paper.

Funding: This work is partially supported by the National Natural Science Foundation of China (Nos. 11590770-4 and 11434012) and the Strategic Priority Research Program of Chinese Academy of Sciences (No. XDC02050400).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DNN	Deep neural network
MFP	Matched-field processing
ML	Maximum likelihood
CS	Compressive sensing
MAP	maximum a posteriori
SBL	Sparse Bayesian learning
UHA	Underwater horizontal arrays
TF	Time-frequency
DOA	Direction of arrival
FNN	Feed-forward neural network
LSTM-RNN	Long short term memory - recurrent neural network
CBF	Conventional beamforming
SCM	Spatial correlation matrix
BP	Back propagation
TDNN	Time delay neural network
MSE	Mean square error
SSP	Sound speed profile
HCA	Horizontal circular array
HLA	Horizontal line array
FFT	Fast Fourier transformation
ReLU	Rectified linear units
PDR	Positive detection rate
FDR	False detection rate
ROC	Receiver operating performance characteristics
MAE	Mean absolute error
MRE	Mean relative error
SNR	Signal-to-noise ratio
SL	Source level
NL	Noise level

References

1. Baggeroer, A.B.; Kuperman, W.A.; Mikhalevsky, P.N. An overview of matched field methods in ocean acoustics. *IEEE J. Ocean. Eng.* **1993**, *18*, 401–424. [CrossRef]
2. Bucker, H.P. Use of calculated sound fields and matched field detection to locate sound source in shallow water. *J. Acoust. Soc. Am.* **1976**, *59*, 368–373. [CrossRef]
3. Westwood, E.K. Broadband matched-field source localization. *J. Acoust. Soc. Am.* **1992**, *91*, 2777–2789. [CrossRef]
4. Li, X.; Zhang, C.; Yan, L.; Han, S.; Guan, X. A Support Vector Learning-Based Particle Filter Scheme for Target Localization in Communication-Constrained Underwater Acoustic Sensor Networks. *Sensors* **2018**, *18*, 8. [CrossRef] [PubMed]
5. Chan, S.-C.; Lee, K.-C.; Lin, T.-N.; Fang, M.-C. Underwater positioning by kernel principal component analysis based probabilistic approach. *Appl. Acoust.* **2013**, *74*, 1153–1159. [CrossRef]
6. Lefort, R.; Real, G.; Drémeau, A. Direct regressions for underwater acoustic source localization in fluctuating oceans. *Appl. Acoust.* **2017**, *116*, 303–310. [CrossRef]
7. Niu, H.; Reeves, E.; Gerstoft, P. Source localization in an ocean waveguide using supervised machine learning. *J. Acoust. Soc. Am.* **2017**, *142*, 1176–1188. [CrossRef]
8. Niu, H.; Ozanich, E.; Gerstoft, P. Ship localization in Santa Barbara Channel using machine learning classifiers. *J. Acoust. Soc. Am.* **2017**, *142*, 455–460. [CrossRef]

9. Ferguson, E.; Ramakrishnan, R.; Williams, S.; Jin, C. Convolutional neural networks for passive monitoring of a shallow water environment using a single sensor. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 2657–2661.
10. Huang, Z.; Xu, J.; Gong, Z.; Wang, H.; Yan, Y. Source localization using deep neural networks in a shallow water environment. *J. Acoust. Soc. Am.* **2018**, *143*, 2922–2932. [CrossRef]
11. Wang, Y.; Peng, H. Underwater acoustic source localization using generalized regression neural network. *J. Acoust. Soc. Am.* **2018**, *143*, 2321–2331. [CrossRef]
12. Niu, H.; Gong, Z.; Reeves, E.; Gerstoft, P.; Wang, H.; Li, Z. Deep-learning source localization using multi-frequency magnitude-only data. *J. Acoust. Soc. Am.* **2019**, *146*, 211–222. [CrossRef] [PubMed]
13. Chi, J.; Li, X.; Wang, H.; Gao, D.; Gerstoft, P. Sound source ranging using a feed-forward neural network with fitting-based early stopping. *J. Acoust. Soc. Am.* **2019**, *146*, EL258–EL264. [CrossRef] [PubMed]
14. Wang, W.; Ni, H.; Su, L.; Hu, T.; Ren, Q.; Gerstoft, P.; Ma, L. Deep transfer learning for source ranging: Deep-sea experiment results. *J. Acoust. Soc. Am.* **2019**, *146*, EL317–EL322. [CrossRef] [PubMed]
15. Collins, M.D.; Fialkowski, L.T.; Kuperman, W.A.; Perkins, J.S. The multi-valued Bartlett procesor and source tracking. *J. Acoust. Soc. Am.* **1995**, *97*, 235–241. [CrossRef]
16. Greening, M.V.; Zakarauskas, P.; Dosso, S.E. Matched-field localization for multiple sources in an uncertain environment, with application to Arctic ambient noise. *J. Acoust. Soc. Am.* **1997**, *101*, 3525–3538. [CrossRef]
17. Mirkin A.N.; Sibul, L.H. Maximum likelihood estimation of the locations of multiple sources in an acoustic waveguide. *J. Acoust. Soc. Am.* **1994**, *95*, 877–888. [CrossRef]
18. Byun, S.-H.; Byun, G.; Sabra, K.G. Ray-based blind deconvolution of shipping sources using multiple beams separated by alternating projection. *J. Acoust. Soc. Am.* **2018**, *144*, 3525–3532. [CrossRef]
19. Michalopoulou, Z.-H. Multiple source localization using a maximum a posteriori Gibbs sampling approach. *J. Acoust. Soc. Am.* **2006**, *141*, 2627–2634. [CrossRef]
20. Gemba, K.L.; Hodgkiss, W.S.; Gerstoft, P. Adaptive and compressive matched field processing. *J. Acoust. Soc. Am.* **2017**, *141*, 92–103. [CrossRef]
21. Gerstoft, P.; Xenaki, A.; Mecklenbrauker, C.F. Multiple and single snapshot compressive beamforming. *J. Acoust. Soc. Am.* **2015**, *138*, 2003–2014. [CrossRef]
22. Li, J.; Lin, Q.; Kang, C.; Wang, K.; Yang, X. DOA Estimation for Underwater Wideband Weak Targets Based on Coherent Signal Subspace and Compressed Sensing. *Sensors* **2018**, *18*, 902. [CrossRef] [PubMed]
23. Gemba, K.L.; Nannuru, S.; Gerstoft, P.; Hodgkiss, W.S. Multi-frequency sparse Bayesian learning for robust matched field processing. *J. Acoust. Soc. Am.* **2017**, *141*, 3411–3420. [CrossRef] [PubMed]
24. El Zooghby, A.H.; Christodoulou, C.G.; Georgiopoulos, M. Performance of Radial-Basis Function Networks for Direction of Arrival Estimation with Antenna Arrays. *IEEE Trans. Antennas Propag.* **1997**, *45*, 1611–1617. [CrossRef]
25. Adavanne, S.; Politis, A.; Virtanen, T. Direction of arrival estimation for multiple sound sources using convolutional recurrent neural network. In Proceedings of the European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018.
26. Chakrabarty, S.; Habets, E.A.P. Broadband DOA estimation using convolutional neural networks trained with noise signals. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 15–18 October 2017.
27. Lo, T.; Leung, H.; Litva, J. Radial basis function neural network for direction-of-arrivals estimation. *IEEE Signal Process. Lett.* **1994**, *1*, 45–47. [CrossRef]
28. Lo, T.K.Y.; Leung, H.; Litva, J. Artificial neural network for AOA estimation in a multipath environment over the sea. *IEEE J. Ocean. Eng.* **1994**, *19*, 555–562. [CrossRef]
29. Ma, N.; May, T.; Brown, G.J. Exploiting Deep Neural Networks and Head Movements for Robust Binaural Localization of Multiple Sources in Reverberant Environments. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *25*, 2444–2453. [CrossRef]
30. Wang, Z.-Q.; Zhang, X.; Wang, D. Robust Speaker Localization Guided by Deep Learning-Based Time-Frequency Masking. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2019**, *27*, 178–188. [CrossRef]
31. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

32. Nuttall, J.; Willett, P. Adaptive-adaptive subarray narrowband beamforming. In Proceedings of the 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, Minneapolis, MN, USA, 27–30 April 1993; pp. 305–308.
33. Wang, Q.; Zhang, R. Sound spatial correlations in shallow water. *J. Acoust. Soc. Am.* **1992**, *92*, 932–938. [CrossRef]
34. Van Trees, H.L. *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV)*; Wiley-Interscience: New York, NY, USA, 2002; Chapter 1–10.
35. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
36. Byun, G.; Song, H.C.; Kim, J.S.; Park, J.S. Real-time tracking of a surface ship using a bottom-mounted horizontal array. *J. Acoust. Soc. Am.* **2018**, *144*, 2375–2382. [CrossRef] [PubMed]
37. Hopfield, J.J.; Tank, D.W. Neural computation of decisions in optimization problems. *Biol. Cybern.* **1985**, *52*, 141–152. [PubMed]
38. Rastogi, R.; Gupta, P.K.; Kumaresan, R. Array signal processing with interconnected neuron-like elements. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, TX, USA, 6–9 April 1987; pp. 2328–2331.
39. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef] [PubMed]
40. Waibel, A.; Hanazawa, T.; Hinton, G.; Shikano, K.; Lang, K.J. Phoneme recognition using time-delay neural networks. *IEEE/ACM Trans. Audio Speech Lang. Process.* **1989**, *37*, 328–339. [CrossRef]
41. Werbos, P.J. Backpropagation through time: what it does and how to do it. *Proc. IEEE* **1990**, *78*, 1550–1560. [CrossRef]
42. Ko, T.; Peddinti, V.; Povey, D.; Khudanpur, S. Audio augmentation for speech recognition. In Proceedings of the INTERSPEECH, Dresden, Germany, 6–10 September 2015.
43. Cui, X.; Goel, V.; Member, S.; Kingsbury, B. Data Augmentation for Deep Neural Network Acoustic Modeling. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 1469–1477.
44. Ko1, T.; Peddinti, V.; Povey, D.; Seltzer, M.L.; Khudanpur, S. A study on data augmentation of reverberant speech for robust speech recognition. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 5220–5224.
45. The SWellEx-96 Experiment. Available online: http://swellex96.ucsd.edu (accessed on 15 September 2019).
46. Glorot, X.; Bordes, A.; Bengio, Y. Deep Sparse Rectifier Neural Networks. In Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS), Fort Lauderdale, FL, USA, 11–13 April 2011; Volume 15, pp. 315–323.
47. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. *OSDI* **2016**, *16*, 265–283.
48. Kingma, D.; Jimmy, B. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
49. George, P.C.; Paulraj, A. Optimising the active sonar system design. *Def. Sci. J.* **1985**, *35*, 295–311. [CrossRef]