IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 23, 2020, accepted June 22, 2020, date of publication June 26, 2020, date of current version July 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3005221

Two-Dimensional DOA Estimation via Deep Ensemble Learning

WENLI ZHU[1], MIN ZHANG[1], PENGFEI LI[2], AND CHENXI WU[1]
[1]College of Electronic Engineering, National University of Defense Technology, Hefei 230037, China
[2]Luoyang Electronic Equipment Test Center of China, Luoyang 471003, China

Corresponding author: Wenli Zhu (li_3crystal@163.com)

This work was supported by the Natural Science Foundation Project of Anhui under Grant 1908085QF280.

ABSTRACT To achieve fast and accurate two-dimensional (2D) direction of arrival (DOA) estimation, a novel deep ensemble learning method is presented in this paper. First, a convolutional neural network (CNN) is employed to learn a mapping between the spatial covariance matrix of the received signals from the antenna elements and the directions of arrival. To avoid any explicit feature extraction step, the real and imaginary parts of the spatial covariance matrix are fed to the CNN. The output layer of the CNN uses three neurons, two of them are the sine and cosine values of the azimuth angle that are used to uniquely determine the azimuth angle, and the third neuron is a normalized value for representing the elevation angle. Second, to improve the prediction performance, since that a single CNN with limited training data has difficulties learning the highly complex and nonlinear mapping from the received signal to the angle of arrival, an ensemble learning method is proposed. Five different CNN networks are trained independently with different training conditions. The prediction results of each individual CNN are calculated as an average to obtain the final estimated results of the azimuth and elevation angles. Simulation results show that the processing time of the proposed deep ensemble learning method is dramatically reduced. In terms of the accuracy, it outperforms the neural network-based 2D DOA estimation and achieves performance comparable to the MUSIC algorithm.

INDEX TERMS Convolutional neural network, deep learning, ensemble learning, two-dimensional direction of arrival estimation, uniform circle array.

I. INTRODUCTION

Direction of arrival (DOA) estimation is a hot research topic. It plays an important role in array signal processing, and it has been studied in many areas such as radar, wireless communication, sonar, electronic countermeasures, etc. Two-dimensional (2D) DOA estimation obtains the azimuth and elevation angles of a target simultaneously and more accurately describes the spatial characteristics of the incident signal. Therefore, 2D DOA estimation is often required in real situations. Compared with one dimensional DOA estimation, 2D DOA estimation problem is more complicated due to the array geometry.

So far, many 2D DOA estimation algorithms have been proposed. Conventional algorithms can be mainly classified into two categories: subspace based algorithms and sparsity-based algorithms. Subspace based algorithms include multiple signal classification (MUSIC) and estimation of signal parameters via rotational invariance technique (ESPRIT). These high resolution algorithms have received consistent attention. In [1], two dimension reduction MUSIC algorithms are presented for coherently distributed sources consisting of circular and noncircular signals in a massive MIMO system. In [2], a modified MUSIC algorithm is applied for the circular and noncircular signals received by polarization sensitive array. In [3], the authors propose a MUSIC algorithm based on a Toeplitz matrix to estimate the DOAs of the coherent signals. In [4], an ESPRIT algorithm is developed to estimate the 2D DOA of mixed circular and noncircular signals, and joint diagonalization is used to solve the angle ambiguity problem. Due to extensive computations caused by the eigen-decomposition of the matrix and the spectrum peak search for all angles both in the azimuth and elevation angles, subspace based algorithms are difficult to implement in real-time. Sparsity-based algorithms such as sparse representation have also been studied [5]–[7]. In [8],

The associate editor coordinating the review of this manuscript and approving it for publication was Syed Mohammad Zafaruddin.

the authors use the multimodal joint sparse representation method to estimate the DOAs of the wideband signals. This kind of algorithm needs to discretize the range of interested angles into grid points and assume that the position of the incident signal falls on the predefined grid. However, in practice, no matter how dense the grid points are, the actual DOA cannot be located on the predetermined grid points, which leads to the off-grid problem and the deterioration of the signal recovery performance. Sparsity-based algorithms also have great computational burdens. In addition, other algorithms have also been proposed. In [9], a beamspace based method is used for the incoherently distributed sources. In [10] and [11], the authors employ the sparse array to obtain the estimated direction cosines to estimate the 2D DOAs. These algorithms outperform subspace based methods in terms of computational speed, but the computational complexity grows linearly with the dimension.

Artificial neural networks (ANNs) have been proved to be more efficient in terms of computational speed while maintaining a comparable resolution capability for 2D DOA estimation. Since these methods do not require complex mathematical operations such as matrix inversion, the important advantage of parallel operation makes it fast at estimating both azimuth and elevation angles. In [12], a linear vector quantization neural network (LVQNN) is developed for 2D DOA estimation. To reduce the training set size, the authors build two different datasets for the azimuth and elevation angles, respectively. Although the method reduces the dimension from two dimensions to one dimension, the azimuth angle is estimated based on the precisely estimated elevation angle. Thus, even at a high signal-to-noise ratio (SNR) up to 20 dB, the estimation error is still large at approximately 2°. Similarly, the LVQNN is also used in [13]. The main difference between these two papers is that the latter employs a single uniform linear array (ULA) and the spatial covariance matrix organized as a $2M^2$-dimensional vector is input to the LVQNN. In [14], two radial basis function neural networks (RBFNNs) are trained for 2D DOA estimation of two coherent sources. The model limits the azimuth and elevation angles to a small range of about (−45°, 45°). In practice, the direction of the received signal covers a wide range, and it is also omnidirectional. In [15], the authors provide a more accurate 2D DOA estimation method by developing an RBFNN-based model combining real electromagnetic sources and a simulated-based neural network. In [16], the model based on a multi-layer perceptron (MLP) for stochastic electromagnetic (EM) sources is offered. Due to the complexity of 2D DOA estimation, these above methods all consider a limited angle range of the direction of the incident signal and require a high SNR to have a good estimation performance.

Deep learning (DL) has the powerful ability to approximate a highly nonlinear function and a good error tolerance capacity. It has been used for one dimension DOA estimation [17]. In [18], a deep neural network (DNN) is trained for multi-speaker DOA estimation. A convolutional neural network (CNN) is a deep network architecture, and it is a promising regression and classification technique proposed by LeCun [19]. Two advantages, sparse connectivity and shared weights, enable CNNs have small numbers of parameters during learning. By now, it has been successfully applied in many areas, such as image recognition [20], [21] and face recognition [22], [23]. When generating a training dataset, covering all possible combinations of azimuth and elevation angles is unpractical. However, a limited amount of training data is not sufficient for a single CNN network to completely predict all unknown test examples.

To improve the generalization ability and overcome this limitation, we propose the use of ensemble learning. Ensemble learning is a machine learning technique that uses multiple base learners to increase the predictive accuracy, and the method is one of the most attractive methods for classification and regression problems. Furthermore, combining the predictions of an ensemble is often more accurate than using an individual base learner. The CNN network is chosen as the base learner. The input data are organized by the real and image parts of the elements of the spatial covariance matrix. The output layer with three neurons aims to realize the regression task. Our main contributions are as follows:

1) A CNN is first used to estimate the 2D DOA. This deep network architecture has a better generalization ability to achieve the inverse nonlinear mapping from the received signal to the angle of arrival compared to other neural networks such as the RBFNN and LVQNN. In addition, the CNN is able to obtain the omnidirectional estimation of the angle of the incident signal.
2) We design the CNN network architecture including a preprocessing to generating the input data and a post-processing to smooth the output values. The ensemble learning method further improves the accuracy of 2D DOA estimation by eliminating the random errors in the training procedure to some extent. In terms of the timeliness and accuracy, the proposed method has a better performance.

The rest of the paper is organized as follows. In section II, the data model used for narrowband far-field signals is established. Section III addresses the proposed deep ensemble learning method for efficient 2D DOA estimation, and it describes the architecture of the CNN and the corresponding input data, output labels and training procedure. Numerical examples and analyses are presented in section IV. Finally, section V summarizes our conclusion.

II. DATA MODEL

To estimate the azimuth and elevation angles simultaneously, a planar array is required for the direction-finding (DF) system. A uniform circle array (UCA) with omnidirectional elements is able to provide 360° azimuthal coverage and a 90° elevation range. Hence, we use the UCA to receive the spatial signals in this paper. It is assumed that K narrowband,

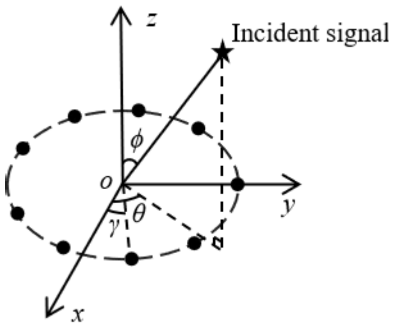

FIGURE 1. Uniform circle array.

uncorrelated incident waves impinge onto an M omnidirectional elements UCA with a radius of r. The structure of the UCA is shown in Fig. 1 and the reference point is the origin of the coordinate system. The source azimuth angle, $\theta \in [0, 2\pi]$, and the elevation angle, $\phi \in [0, 2/\pi]$, in a spherical coordinate system, are measured from the x and the z-axes, respectively. Therefore, the received signal vector obtained at the output of this antenna array can be defined as

$$\boldsymbol{x}(t) = \boldsymbol{A}s(t) + \boldsymbol{n}(t) \tag{1}$$

where $\boldsymbol{x}(t)$, $\boldsymbol{s}(t)$, and $\boldsymbol{n}(t)$ are given by

$$\boldsymbol{x}(t) = (x_1(t), x_2(t), \ldots, x_M(t))^T \tag{2}$$
$$\boldsymbol{s}(t) = (s_1(t), s_2(t), \ldots, s_K(t))^T \tag{3}$$
$$\boldsymbol{n}(t) = (n_1(t), n_2(t), \ldots, n_M(t))^T \tag{4}$$

In the above equations, the superscript T indicates the transpose of the matrix. Additionally, in (3), $\boldsymbol{s}(t)$ is a $K \times 1$ vector that denotes the source signals coming from the far-field directions with the azimuth and elevation angles $\{(\theta_1, \phi_1), (\theta_2, \phi_2), \ldots, (\theta_K, \phi_K)\}$. In (4), $\boldsymbol{n}(t)$ is an $M \times 1$ vector representing the noise signals received by the M elements. It is assumed to be complex, zero mean, Gaussian white process, and independent of the signals. $\mathbf{A}$ is an $M \times K$ array manifold matrix whose columns are steering vectors towards K different directions of arrival and it can be written as

$$\mathbf{A} = (\boldsymbol{a}(\theta_1, \phi_1), \boldsymbol{a}(\theta_2, \phi_2), \ldots, \boldsymbol{a}(\theta_K, \phi_K))^T \tag{5}$$

The steering vector of the i-th incident signal is given by

$$\boldsymbol{a}(\theta_i, \phi_i) = \begin{bmatrix} e^{j2\pi r \cos(\gamma_0 - \theta_i)\cos\phi_i/\lambda} \\ e^{j2\pi r \cos(\gamma_1 - \theta_i)\cos\phi_i/\lambda} \\ \vdots \\ e^{j2\pi r \cos(\gamma_{M-1} - \theta_i)\cos\phi_i/\lambda} \end{bmatrix} \tag{6}$$

where λ is the incident signal wavelength and γ are angular positions of the antenna array elements which can be calculated as

$$\gamma_i = \frac{2\pi i}{M}, \quad i = 0, 1, \ldots, M-1 \tag{7}$$

As long as the array geometry is formed, the array manifold is uniquely determined, which contains enough information to estimate the angle of arrival. From (1), the spatial covariance matrix $\mathbf{R}$ of the received noisy signals can be formulated as

$$\mathbf{R} = E\left\{\boldsymbol{x}(t)\boldsymbol{x}^H(t)\right\} = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \sigma^2\mathbf{I} \tag{8}$$

where $E\{\}$ is the statistical expectation operator, the superscript H denotes the complex conjugate transpose operation, σ^2 represents the noise power at the array elements, $\mathbf{I}$ is the identity matrix with dimensions $M \times M$, and $\mathbf{R}_s$ is $K \times K$ signal covariance matrix that is written as

$$\mathbf{R}_S = E\left\{\boldsymbol{s}(t)\boldsymbol{s}^H(t)\right\} \tag{9}$$

In practice, the spatial covariance matrix $\mathbf{R}$ can be only obtained via finite samples. Thus, the maximum likelihood estimation of the spatial covariance matrix is shown as follows:

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{i=1}^{N} x(i)x^H(i) \tag{10}$$

where N is the number of the snapshots used for the formation of the spatial covariance matrix. The problem is to estimate the direction of the incident signal on the basis of a set of snapshots $x(1), \ldots, x(N)$. In the following, it will be assumed that the number of signals has been determined.

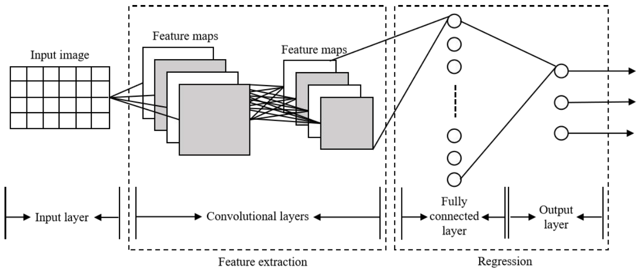

FIGURE 2. The CNN architecture for 2D DOA estimation.

III. METHOD

A. CONVOLUTIONAL NEURAL NETWORK

Convolutional neural networks are a variant of deep neural network frameworks that have one or more convolutional layers to extract the discriminative feature from the input data. The main motivation to construct a CNN model in this paper is that the reception field can obtain the phase map relationship from the input data to the output values. The feature maps are required to learn from the phase correlation between all the array elements. After all the convolutional layers, these learned features are then aggregated to the vectors by the fully connected layers for the regression task. The CNN architecture for 2D DOA estimation is shown in Fig. 2. Once the training of the CNN is accomplished, it has established an approximation of the desired input/output mapping. Hence, during the testing process, the CNN produces outputs of previously unseen inputs by interpolating between the inputs used (seen) in the training process.

1) PREPROCESSING OF THE RECEIVED SIGNALS

To efficiently generalize the CNN, preprocessing is an intelligent operation to accelerate the training time. The received signals from the array elements are preprocessed to eliminate extraneous data, and to convert the available signals into something that the neural network can efficiently use. The preprocessing requires only simple linear algebraic evaluations (matrix-matrix and matrix-vector multiplication). We choose the variant of the spatial covariance matrix **R** as the input of the network because it contains enough statistical information about the direction of the incident signal, and the cross-correlation of the signals received by different array elements removes redundant or irrelevant information.

Since the matrix **R** is conjugate-symmetrical with respect to the diagonal, the elements of its upper or lower triangular part provide enough information for 2D DOA estimation. Notice that **R** contains complex numbers and CNN networks do not support operations with complex numbers. Therefore, the input data are organized in a matrix in such a manner that real and imaginary parts of the complex matrix components are separated. We call this input representation the direction image. If an M element UCA is employed at the receiver, it implies that there are $M \times (M-1)$ neurons in the input layer. The transformation from **R** into the direction image is exemplified as follows.

Step 1: According to (10), we obtain the spatial covariance matrix **R** from the received signal. Here, R_{ij} is the element in the i-th column and j-th row of **R** estimated during the measurement.

$$\mathbf{R} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1M} \\ R_{21} & R_{22} & \cdots & R_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ R_{M1} & R_{M2} & \cdots & R_{MM} \end{bmatrix}$$

Step 2: The real part $\text{Re}(R_{\text{ij}})$ and the imaginary part $\text{Im}(R_{\text{ij}})$ of $R_{\text{ij}}(i < j)$ are organized into a single vector.

$$r = \left[\text{Re}\,(R_{12}) \;\; \text{Im}\,(R_{12}) \;\; \cdots \;\; \text{Re}\left(R_{(M-1)M}\right) \;\; \text{Im}\left(R_{(M-1)M}\right)\right]$$

Step 3: The vector $\boldsymbol{r}$ is normalized using Min-Max scaling, and then the normalized vector $\boldsymbol{r}_{\text{norm}}$ is reconverted into the $M\times(M-1)$ matrix, whose values are used during the training. The $M\times(M-1)$ matrix is regarded as the direction image. The training is performed offline by presenting input/output pairs $\{\boldsymbol{r}_{\text{norm}}, (\theta, \phi)\}$ to the network.

$$\boldsymbol{r}_{norm} = \frac{r_{ij} - \min(\boldsymbol{r})}{\max(r) - \min(\boldsymbol{r})}$$

2) OUTPUT SMOOTHING

The CNN is established to implement the regression task. For each input example, it predicts the azimuth and elevation angles of the incident signal as continuous values. The number of output neurons is equal to the amount of information to be output. In this paper, we attempt to use two different outputs to represent the azimuth angle. One is the normalized value of the azimuth angle as the training output, and the other is the sine and cosine values of the azimuth angle. For the elevation angle, a single output neuron is used.

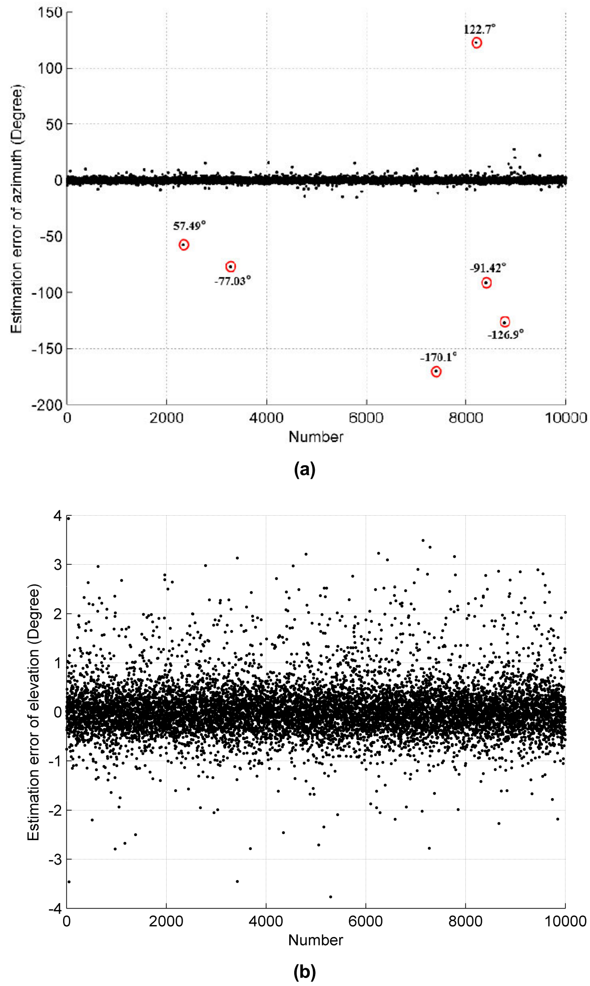

FIGURE 3. **Estimation errors of the CNN with one output for the azimuth angle: (a) estimation errors of the azimuth angle, and (b) estimation errors of the elevation angle.**

We evaluate two different outputs for representing the azimuth angle in the output layer of the CNN. The CNN architecture used in this experiment has five convolutional layers with 128 feature maps in each convolutional layer and one fully connected layer with 512 neurons. The size of the feature map in each convolutional layer is fixed at 2×2. The simulation parameters are set in Table 2, and the sizes of training and test datasets are 361501 and 10000, respectively. The results are shown in Fig. 3 and Fig. 4. The maximum absolute error of the azimuth angle is 170.1° in Fig. 3 while this value is less than 4.5° in Fig. 4.

In Table 1, test statistics of two different outputs of the CNN are given in terms of the mean absolute error and the root mean square error of the azimuth angle. For these simulations, it may be observed that the CNN trained with three output neurons has yielded more accurate source directions since the sine and cosine values can uniquely determine the angle over the full range. Hence, the output of the CNN that

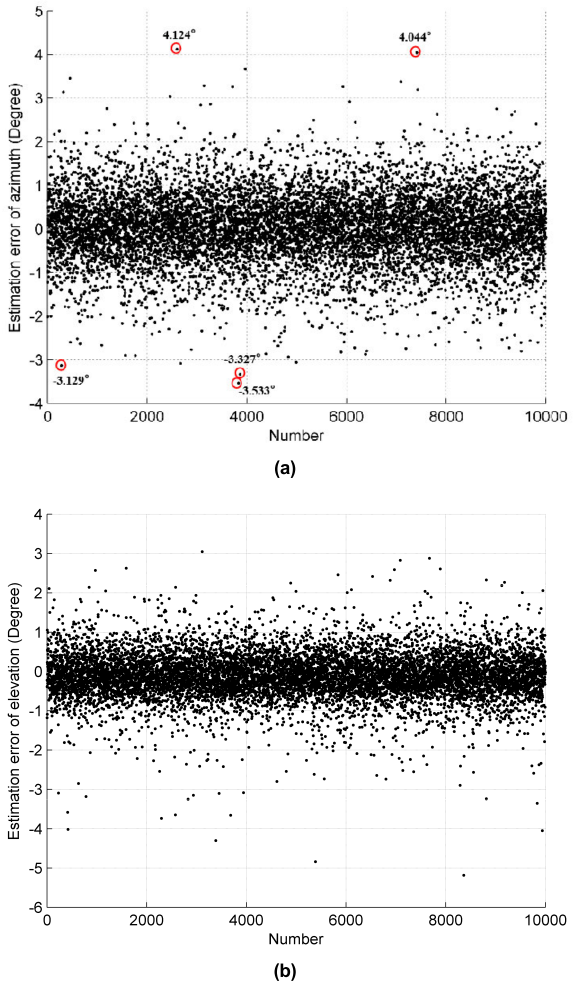

FIGURE 4. Estimation errors of the CNN with two outputs for the azimuth angle: (a) estimation errors of the azimuth angle, and (b) estimation errors of the elevation angle.

TABLE 1. The estimation results of different outputs for the azimuth angle.

CNN architecture	Number of output neurons	Mean absolute error (°)	Root mean square error (°)
CNN128-128-128-128-128(2-2-2-2-2)	2	1.53	3.25
	3	**0.54**	**0.69**

is chosen for the rest of the simulations is the one with the sine and cosine values for the azimuth angle.

3) TRAINING SETUP

The ANN is subject to a training procedure before its use in any environment. In this article, the CNN is trained by using the adaptive momentum (Adam) optimizer to minimize the mean squared error (MSE) between the network's actual output vector and the network's desired output vector with a given accuracy of 10^{-4}. During training, in order to avoid overfitting, a dropout operation with a rate of 0.5 is used at the end of the convolutional layers and after each fully connected layer. The activation functions used in the convolutional layer and the output layer are rectified linear unit (ReLU) and sigmoid activation function, respectively. Therefore, the output values are confined to the interval (0, 1).

B. RANDOM ERROR ELIMINATION BASED ON ENSEMBLE LEARNING

For 2D DOA estimation, the nonlinear mapping function from the input data to the output values is so complex that a single CNN network with a limited training dataset cannot learn well resulting in high estimation errors. On the other hand, the CNN network learns via a stochastic training algorithm which means that the network is sensitive to the training conditions, both in terms of the initial random weights and in terms of the batch size and it may learn a slightly different version of the mapping function each time in the training process, which in turn produces different predictions of the direction of the incident signal. When we prepare a trained CNN as a final model to make predictions, the uncertainties may degrade the performance. Based on this, we proposed a deep ensemble learning method to ensure that the most stable and best possible prediction is made for 2D DOA estimation.

Ensemble learning is a successful approach to reduce the generalization error of neural network models by training multiple models instead of an individual model and combining the predictions from these models in some way. The ensemble model not only reduces the prediction errors but also results in predictions that are better than those of any single network model. Ensemble learning improves the generalization capability and robustness. Techniques for ensemble learning can be grouped by the element that is varied, such as the training data, the base learner, and how predictions are combined. The most popular ensemble techniques are bagging [24], boosting [25] and stacking [26].

Since the number of possible training data combinations is enormous, a certain amount of training data is generated to train the CNNs in this paper. Several CNN networks with different configurations (e.g. the number of convolutional layers or the number of neurons in the fully connected layer), different initial random weights and different conditions (e.g. the learning rate) are trained on the same dataset. Each CNN model is then used to make a prediction and the actual predictions for the azimuth and elevation angles are calculated as the averages of the predictions. The proposed deep ensemble model for 2D DOA estimation is shown in Fig. 5. First, the signal received by the UCA is preprocessed to form the direction image. Second, the direction image is fed to several developed CNNs as the input data. Finally, each CNN output is calculated as the average to obtain the estimated azimuth and elevation angles. During the training procedure, each individual CNN is trained independently using the same training dataset. During the testing procedure, a test example is applied to all constituent CNNs simultaneously and a collective prediction is obtained based on the average result. A few different CNN structures, all having 72 input neurons and 3 output neurons, are considered and trained. The

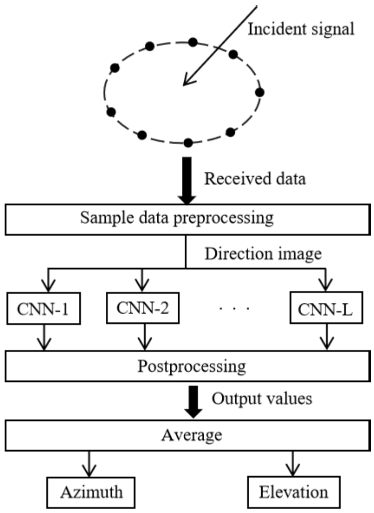

FIGURE 5. The block diagram of the deep ensemble model for 2D DOA estimation.

TABLE 2. The simulation parameters used for training and testing.

Simulation parameters	Values
Number of array elements	$M = 9$
Radius of UCA	$r = 0.75$ m
Number of snapshots	$N = 1024$
Number of narrowband signals	$K = 1$
Signal frequency	$f = 500$ MHz
Signal-to-noise ratio	SNR = 10 dB
Azimuth angle	$\theta = [0, 2\pi]$
Elevation angle	$\phi = [0, 2/\pi]$

postprocessing stage associates the output neurons' computed values with the desired angular parameters.

IV. SIMULATION RESULTS

A. SIMULATION CONDITIONS

1) SIMULATION PARAMETERS AND PERFORMANCES METRICS

Numerical simulations are carried out to validate the effectiveness of the proposed method. We use a nine elements UCA to gather data for the neural network testing. The simulation parameters, which are used to generate the training and test datasets, are set as in Table 2.

To evaluate the performance of the deep ensemble model for 2D DOA estimation, the three performances metrics of the mean absolute error (MAE), the root mean square error (RMSE), and the success rate are calculated. The formula for the mean absolute error computed between the estimated and actual 2D DOA is given by

$$MAE = \frac{1}{N_T}\sum_{i=1}^{N_T}\left(|\theta_{est} - \theta_{act}| + |\phi_{est} - \phi_{act}|\right) \quad (11)$$

where N_{T} is the total number of examples in the test set, and θ_{est} and ϕ_{est} are the estimated azimuth angle and the estimated elevation angle, respectively. θ_{act} and ϕ_{act} are the actual azimuth angle and the actual elevation angle, respectively. The root mean square error is defined as

$$RMSE = \sqrt{\frac{1}{N_T}\sum_{i=1}^{N_T}\left[(\theta_{est} - \theta_{act})^2 + (\phi_{est} - \phi_{act})^2\right]} \quad (12)$$

The success rate represents the degree of closeness of the measurements of a quantity to that quantity's true value. The expression is formulated as follows:

$$succese\ rate = \frac{num\left(|\theta_{est} - \theta_{act}| \le 1^{\circ} \cap |\phi_{est} - \phi_{act}| \le 1^{\circ}\right)}{N_T} \quad (13)$$

In our evaluation, the estimation of the direction of the arrival is considered accurate if the error between the actual and estimated 2D DOA is less than or equal to 1°.

2) TRAINING AND TEST DATASETS

The first step is to form appropriate datasets that are used for training and testing the CNN architectures. To provide the early description, a single training dataset is used for training several CNNs. According to the geometry-based analytical expression of the steering matrix **A** of the array, the training and test datasets are generated using (1) and (10).

A large training dataset is necessary to cover different combinations of the multiple signal directions for 2D DOA estimation. In the training dataset, the spacing of the training samples for all azimuth angles and elevation angles is 0.3°, and as a result, a training set with 361501 samples is obtained. Test dataset is formed in a similar manner. Both the azimuth and elevation angles are selected at random to ensure that the training and test dataset do not overlap. Thus, our test dataset consists of $N_{\mathrm{test}} = 10000$ samples. Additionally, we use noisy training samples, which mean that in all cases the network should be able to learn the noise's influence on the signal.

B. RESULTS

In this section, simulations are presented to illustrate the performance of the proposed fast 2D DOA estimation method.

1) PERFORMANCE OF THE PROPOSED METHOD

The established CNN network approximately achieves the inverse mapping from the spatial covariance matrix to the direction of the incident signal and the performance is shown in Fig. 4 and Table 1. As we can see, the RMSE of the azimuth angle is up to 0.69°. To obtain perfect performance, we utilize ensemble learning to eliminate random errors. Five CNNs are trained with different numbers of feature maps, different numbers of convolutional layers, different learning rates and different batch sizes. The details of the five CNNs are given in Table 3.

Table 4 shows the results of the five CNNs and the proposed ensemble model. From these results, we can see that

TABLE 3. The parameters of five CNN networks.

Model	Architecture	Batch size	Learning rate	Standard deviation of weight
CNN-1	128-128-128-128-128 (2-2-2-2-2)	256	0.001	0.05
CNN-2	128-128-128-128-128 (2-2-2-2-2)	128	0.01	0.1
CNN-3	128-128-128 (3-3-3)	200	0.01	0.1
CNN-4	128-128-128 (3-3-3)	200	0.001	0.05
CNN-5	128-128-128 (2-3-4)	200	0.001	0.05

TABLE 4. The statistic results for different models.

Model	MAE (°)	RMSE (°)	Success rate (%)
CNN-1	0.93	0.88	81.22
CNN-2	1.02	0.95	78.63
CNN-3	0.98	0.92	79.43
CNN-4	1.03	0.98	75.19
CNN-5	0.98	0.93	77.86
Deep ensemble learning	**0.62**	**0.59**	**94.32**

TABLE 5. The NMSEs of the azimuth and elevation angles for different models.

Model	NMSE of the azimuth angle (°)	NMSE of the elevation angle (°)
CNN-1	1	0.9996
CNN-2	1	0.9994
CNN-3	1	0.9995
CNN-4	0.9999	0.9995
CNN-5	1	0.9995
Deep ensemble learning	1	0.9998

the proposed ensemble model achieves an RMSE less than 0.6° and a success rate of more than 94%. It proves that the generalization capability of the ensemble model is better than that of any single CNN. The ensemble model that is the average of multiple well-performing CNN models achieves state-of-the-art results.

The normalized mean square errors (NMSEs) of the azimuth and elevation angles for different models are given in Table 5. If the NMSE is equal to 1, then the estimated value is no better than a straight line at matching the actual value. From Table 5, it can be seen that the NMSE of the azimuth angle is almost equal to 1 and the NMSE of the elevation angle is approximately to 1. The results show that the proposed deep ensemble learning method has an excellent predictive ability.

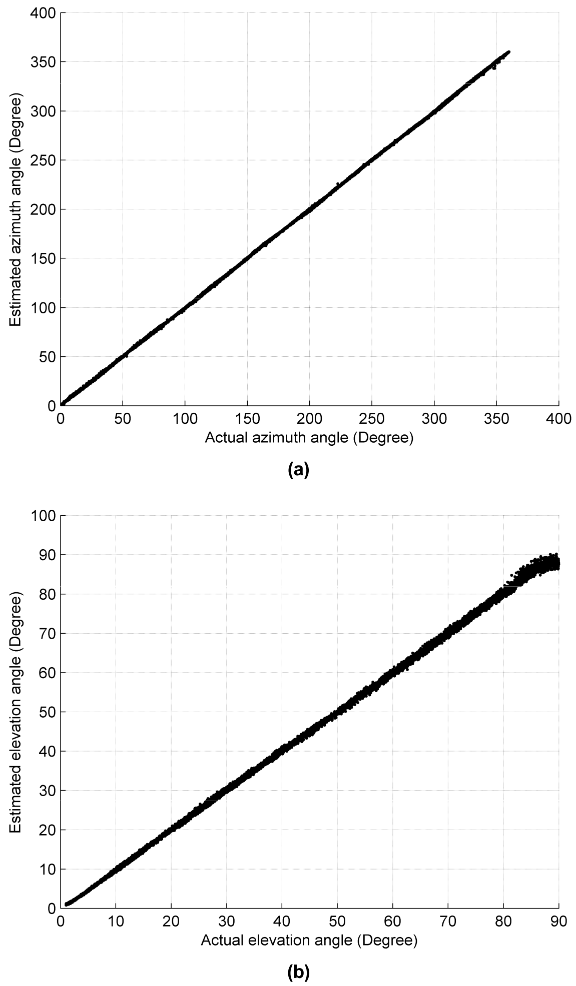

FIGURE 6. Scatter graphs of the estimation results for CNN128-128-128-128-128(2-2-2-2-2): (a) the estimated azimuth angles vs the actual azimuth angles, and (b) the estimated elevation angles vs the actual elevation angles.

To understand the simulation results more intuitively, we plot the responses of CNN-1 in the form of scatter graphs in Fig. 6. The estimation results for the azimuth and elevation angles of the proposed ensemble model are illustrated in Fig. 7. Comparing of these graphs with those from Fig. 6, we see that the thinner curves mean that the estimated angles are close to the actual angles, and the ensemble model improves the randomness and nondeterminacy of the single CNN network to a certain extent. Importantly, it can be concluded that the good generalized ensemble model performs well in response to input matrices that have not been involved in the training process. In Fig. 8, we depict the responses of the ensemble model for 100 test samples. As it can be noticed, angular positions of the incident signals estimated by the ensemble model are in good agreement with the actual ones, which proves the performance of the proposed method.

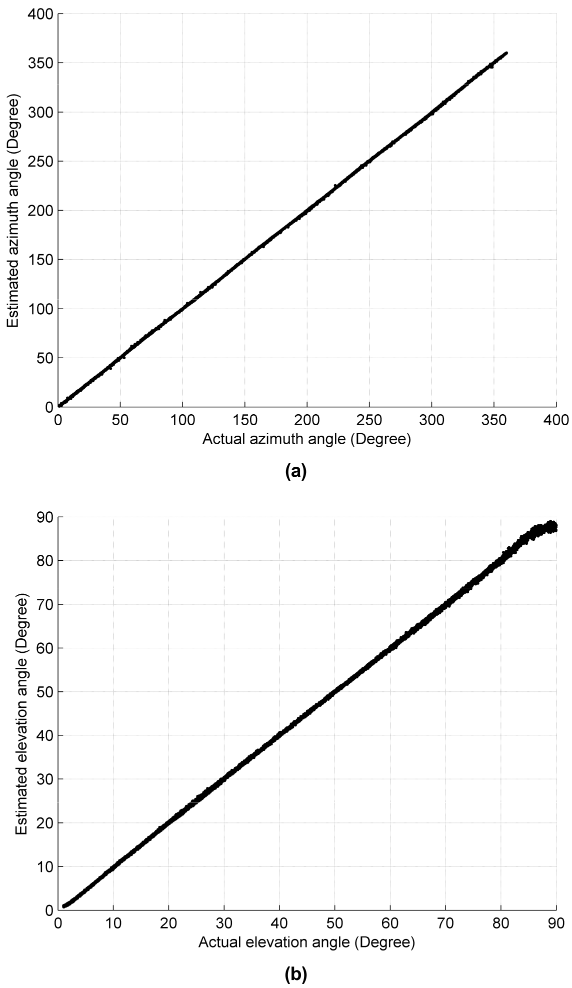

FIGURE 7. Scatter graphs of the estimation results for the ensemble model: (a) the estimated azimuth angles vs the actual azimuth angles, and (b) the estimated elevation angles vs the actual elevation angles.

2) COMPARISON EXPERIMENT

As an illustration of the proposed ensemble model's efficiency for 2D DOA estimation, a comparison of the simulation run times required to calculate the angular positions of the incident signals defined by the test dataset of it, MUSIC [27], l_1-SVD [5], LVQNN [12] and RBFNN [15] is shown in Table 6. With an Intel(R) Core(TM) i7-4790 CPU computer running at 3.6 GHz, the neural networks are implemented in Python using the TensorFlow framework, and the conventional 2D DOA estimation algorithms are realized in the MATLAB simulation environment.

The MUSIC and l_1-SVD algorithms require 53.6 s and 36.2 s to estimate 100 test signals, respectively. Meanwhile, the proposed method only needs 3.29 s for 10000 test signals. It can be observed that the ANN methods respond very fast and have higher speeds than the MUSIC and l_1-SVD algorithms. This ability qualifies them as very suitable to be applied to determine the direction of the incident signal for real-time applications. Regarding the RMSE, the ANN methods have slightly worse agreement in comparison to the conventional methods; however, the accuracy is still very good.

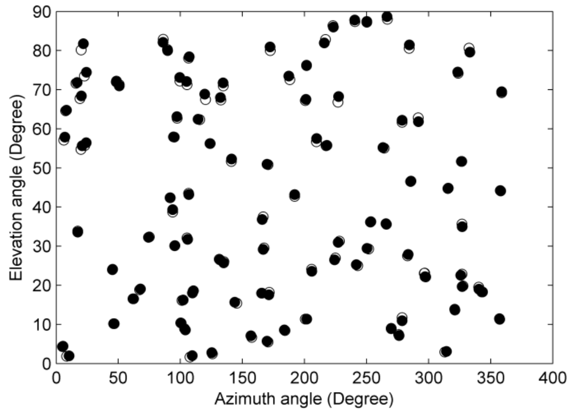

FIGURE 8. Results of the ensemble model for 100 test points (*o* - the ensemble model response, •-the actual direction of the incident signal).

TABLE 6. Comparison of different 2D DOA estimation methods.

Method	Number of test samples	Simulation run time on test samples (s)	RMSE (°)
MUSIC	100	53.6	0.61
l_1-SVD	100	36.2	0.53
LVQNN	10000	2.16	3.04
RBFNN	10000	2.03	2.36
Deep ensemble learning	10000	3.29	0.59

Although the conventional 2D DOA estimation methods have smaller RMSEs, they require much longer run times. When we have a high performance requirement for both accuracy and timeliness, the proposed deep ensemble learning method is the best at achieving fast and accurate 2D DOA estimation.

V. CONCLUSION

In this paper, we advance an efficient 2D DOA estimation method based on deep ensemble learning. The real and imaginary parts of the spatial covariance matrix are converted into a direction image to be input into the CNN. A key advantage of the proposed method over conventional 2D DOA estimation algorithms is its ability to obtain the direction of the incident signal instantaneously, making it suitable for real-time processing. Validations and experimental results also demonstrate that the ensemble model improves the generalization ability and achieves promising results.

REFERENCES

[1] L. Wan, G. Han, J. Jiang, J. J. P. C. Rodrigues, N. Feng, and T. Zhu, "DOA estimation for coherently distributed sources considering circular and noncircular signals in massive MIMO systems," *IEEE Syst. J.*, vol. 11, no. 1, pp. 41–49, Mar. 2017, doi: 10.1109/JSYST.2015.2445052.

[2] X. Wang, L. Wan, M. Huang, C. Shen, and K. Zhang, "Polarization channel estimation for circular and non-circular signals in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 1001–1016, Sep. 2019, doi: 10.1109/JSTSP.2019.2925786.

[3] Z. Zheng, Y. Huang, W.-Q. Wang, and H. C. So, "Direction-of-Arrival estimation of coherent signals via coprime array interpolation," *IEEE Signal Process. Lett.*, vol. 27, pp. 585–589, 2020, doi: 10.1109/LSP.2020.2982769.

[4] H. Chen, C. Hou, W.-P. Zhu, W. Liu, Y.-Y. Dong, Z. Peng, and Q. Wang, "ESPRIT-like two-dimensional direction finding for mixed circular and strictly noncircular sources based on joint diagonalization," *Signal Process.*, vol. 141, pp. 48–56, Dec. 2017, doi: 10.1016/j.sigpro.2017.05.024.

[5] Z. Cheng, P. Shui, H. Li, and Y. Zhao, "Two-dimensional DOA estimation algorithm with co-prime array via sparse representation," *Electron. Lett.*, vol. 51, no. 25, pp. 2084–2086, Dec. 2015, doi: 10.1049/el.2015.0293.

[6] Y.-X. Zou, B. Li, and C. H. Ritz, "Multi-source DOA estimation using an acoustic vector sensor array under a spatial sparse representation framework," *Circuits, Syst., Signal Process.*, vol. 35, no. 3, pp. 993–1020, Mar. 2016, doi: 10.1007/s00034-015-0102-9.

[7] W. Si, Y. Wang, and C. Zhang, "2D-DOA and polarization estimation using a novel sparse representation of covariance matrix with COLD array," *IEEE Access*, vol. 6, pp. 66385–66395, 2018, doi: 10.1109/ACCESS.2018.2879051.

[8] L. Wan, G. Han, L. Shu, S. Chan, and N. Feng, "PD source diagnosis and localization in industrial high-voltage insulation system via multimodal joint sparse representation," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2506–2516, Apr. 2016, doi: 10.1109/TIE.2016.2520905.

[9] Z. Zheng, W.-Q. Wang, H. Meng, H. C. So, and H. Zhang, "Efficient beamspace-based algorithm for two-dimensional DOA estimation of incoherently distributed sources in massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11776–11789, Dec. 2018, doi: 10.1109/TVT.2018.2875023.

[10] Z. Zheng, Y. Yang, W.-Q. Wang, and S. Zhang, "Two-dimensional direction estimation of multiple signals using two parallel sparse linear arrays," *Signal Process.*, vol. 143, pp. 112–121, Feb. 2018, doi: 10.1016/j.sigpro.2017.08.013.

[11] Z. Zheng and S. Mu, "Two-dimensional DOA estimation using two parallel nested arrays," *IEEE Commun. Lett.*, vol. 24, no. 3, pp. 568–571, Mar. 2020, doi: 10.1109/LCOMM.2019.2958903.

[12] J. D. Ndaw, A. Faye, and A. S. Maiga, "Decoupled 2D DOA estimation using LVQ neural networks and UCA arrays," in *Proc. IEEE Radio Antenna Days Indian Ocean (RADIO)*, St. Gilles-les-Bains, Reunion, Oct. 2016, pp. 1–2, doi: 10.1109/RADIO.2016.7771995.

[13] A. Faye, J. D. Ndaw, and A. S. Maiga, "Two-dimensional DOA estimation based on a single uniform linear array," in *Proc. 25th Telecommun. Forum (TELFOR)*, Belgrade, Serbia, Nov. 2017, pp. 1–4.

[14] M. Agatonović, Z. Stanković, N. Dońcov, B. Milovanović, and I. Milovanović, "Neural network model for 2D DOA estimation of two coherent sources," *Int. J. Reasoning-Based Intell. Syst.*, vol. 7, nos. 1–2, pp. 62–69, 2015, doi: 10.1504/ijris.2015.070914.

[15] M. Stoilkovic, Z. Stankovic, and B. Milovanovic, "A cascade-connected neural model for improved 2D DOA estimation of an EM signal," *Int. J. Numer. Model., Electron. Netw., Devices Fields*, vol. 29, no. 2, pp. 343–353, Mar. 2016, doi: 10.1002/jnm.2081.

[16] Z. Stanković, N. Dońcov, B. Milovanović, and I. Milovanović, "Efficient 2D localization of a number of mutually arbitrary positioned stochastic EM sources in far-field using neural model," *Proc. Int. Conf. Electromagn. Adv. Appl. (ICEAA)*, Verona, Italy, Sep. 2017, pp. 1391–1394, doi: 10.1109/ICEAA.2017.8065537.

[17] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018, doi: 10.1109/TVT.2018.2851783.

[18] D. Wang, Y. Zou, and W. Wang, "Learning soft mask with DNN and DNN-SVM for multi-speaker DOA estimation using an acoustic vector sensor," *J. Franklin Inst.*, vol. 355, no. 4, pp. 1692–1709, Mar. 2018, doi: 10.1016/j.jfranklin.2017.05.002.

[19] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989, doi: 10.1162/neco.1989.1.4.541.

[20] J. Fu, H. Zheng, and T. Mei, "Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 4476–4484, doi: 10.1109/CVPR.2017.476.

[21] I. Rocco, R. Arandjelovic, and J. Sivic, "Convolutional neural network architecture for geometric matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2553–2567, Nov. 2019, doi: 10.1109/TPAMI.2018.2865351.

[22] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2884–2896, Nov. 2018, doi: 10.1109/TIFS.2018.2833032.

[23] T. H. Kim, C. Yu, and S. W. Lee, "Facial expression recognition using feature additive pooling and progressive fine-tuning of CNN," *Electron. Lett.*, vol. 54, no. 23, pp. 1326–1327, 2018, doi: 10.1049/el.2018.6932.

[24] G. Collell, D. Prelec, and K. R. Patil, "A simple plug-in bagging ensemble based on threshold-moving for classifying binary and multiclass imbalanced data," *Neurocomputing*, vol. 275, pp. 330–340, Jan. 2018, doi: 10.1016/j.neucom.2017.08.035.

[25] M. Opitz, G. Waltner, H. Possegger, and H. Bischof, "BIER–boosting independent embeddings robustly," *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1. Venice, Italy, Oct. 2017, pp. 5199–5208, doi: 10.1109/ICCV.2017.555.

[26] T. Zhou, F.-L. Chung, and S. Wang, "Deep TSK fuzzy classifier with stacked generalization and triplely concise interpretability guarantee for large data," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 5, pp. 1207–1221, Oct. 2017, doi: 10.1109/TFUZZ.2016.2604003.

[27] J. J. Cai, P. Li, and G. Q. Zhao, "Two-dimensional DOA estimation with reduced-dimension MUSIC," *J. Xidian Univ.*, vol. 40, no. 3, pp. 81–86+144, 2013, doi: 10.3969/j.issn.1001-2400.2013.03.012.

WENLI ZHU received the B.S. degree from the Hefei University of Technology, Hefei, China, in 2013, and the M.S. degree from the Institute of Electronics Engineering, Hefei, in 2016. She is currently pursuing the Ph.D. degree with the National University of Defense Technology.

Her research interests include signal processing, communication security, and deep learning.

MIN ZHANG received the Ph.D. degree from Anhui University, Hefei, China. He is currently a Professor with the National University of Defense Technology. His research interests include signal processing, intelligent information processing, machine learning, and data mining.

PENGFEI LI received the Ph.D. degree from the Institute of Electronics Engineering, Hefei, China, in 2012. He is currently engaged in the test and evaluation of artificial intelligence with Luoyang Electronic Equipment Test Center of China. His research interests include signal processing, artificial intelligence, and network security.

CHENXI WU received the Ph.D. degree from the Institute of Electronics Engineering, Hefei, China, in 2017. He is currently a Lecturer with the National University of Defense Technology. His research interests include communication signal processing, compressed sensing, and space target localization and navigation.

• • •