


RESEARCH

Open Access



# Deep learning-based DOA estimation for hybrid massive MIMO receive array with overlapped subarrays

Yifan Li<sup>1\*</sup> , Baihua Shi<sup>1</sup>, Feng Shu<sup>1,2</sup>, Yaoliang Song<sup>1</sup> and Jiangzhou Wang<sup>3</sup>

\*Correspondence:  
liyifan97@foxmail.com

<sup>1</sup> School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

<sup>2</sup> School of Information and Communication Engineering, Hainan University, Haikou 570228, China

<sup>3</sup> School of Engineering, University of Kent, Canterbury CT2 7NT, UK

## Abstract

As massive MIMO is a key technology in the future sixth generation (6G), the large-scale antenna arrays are widely considered in direction-of-arrival (DOA) estimation for they can provide larger aperture and higher estimation resolution. However, the conventional fully digital architecture requires one radio-frequency (RF) chain per antenna, and this is challenging for the high hardware costs and much more power consumption caused by the large number of RF chains. Therefore, an overlapped subarray (OSA) architecture-based hybrid massive MIMO array is proposed to reduce the hardware costs, and it can also have better DOA estimation accuracy compared to non-overlapped subarray (NOSA) architecture. The simulation results also show that the accuracy of the proposed OSA architecture has 6° advantage over the NOSA architecture with signal-to-noise ratio (SNR) at 10 dB. In addition, to improve the DOA estimation resolution, a deep learning (DL)-based estimator is proposed by combining convolution denoise autoencoder (CDAE) and deep neural network (DNN), where CDAE can remove the approximation error of sample covariance matrix (SCM) and DNN is used to perform high-resolution DOA estimation. From the simulation results, CDAE-DNN can achieve the accuracy lower bound at SNR = -8 dB and the number of snapshots  $N = 100$ , this means it has better performance in poor communication situation and can save more software resources compared to conventional estimators.

**Keywords:** Direction-of-arrival (DOA) estimation, Massive MIMO, Overlapped subarray, Deep learning, Cramer–Rao lower bound (CRLB)

## 1 Introduction

Direction-of-arrival (DOA) estimation has been an important research direction in the areas of wireless communications, radar, sonar, etc., for a long time [1]. With the development of 5G, the massive MIMO system has been studied extensively. However, the realization of the traditional full digital system requires a high hardware complexity, so the hybrid analog and digital (HAD) system was considered as an alternative [2]. Then, the DOA estimation problem for the HAD system was discussed in [3], and the directional modulation design for hybrid antenna array was also considered in [4]. In addition to the common architecture, various hybrid architectures were considered in [5]

and [6], the overlapped subarrays (OSA) architecture in [7] was proved to have better beamforming performance than nonoverlapped subarrays (NOSA) architecture.

Recently, machine learning (ML) and deep learning (DL) have been applied to wireless communication and signal processing in many papers, such as [8] proposed a ML-based low-complexity method for channel state information (CSI) estimation, the neural network is used to locate the structural sound source in [9]. DL-based methods are also popular in DOA estimation. Traditional DOA estimation methods are mainly divided into two categories: parameter estimation-based methods and subspace methods [10]. The first category contains nonlinear least-square (NLS) estimator and maximum likelihood estimator, while the classical subspace methods include MUSIC, ESPRIT, root-MUSIC, etc. Compared to them, the DL-based methods have lower complexity than parameter estimation-based methods and higher accuracy than subspace methods. In [11], a deep neural network (DNN) was proposed for DOA estimation with array imperfections. [12] gave a DNN-based DOA estimation method for hybrid massive MIMO systems with uniform circular arrays (UCAs). [13] also combined DL with massive MIMO system and proposed a super-resolution channel estimation method.

Besides the ordinary DNN, there are some special neural network structures that can be used for estimation. For example, the convolution neural network (CNN) is another popular choice [14, 15], especially to improve the accuracy in the low signal-to-noise-ratio (SNR) regime [16, 17]. In [18], the complex ResNet was applied for DOA estimation in the near-field MIMO systems. A deep residual network was also proposed for underdetermined DOA estimation in [19]. And deep adaptive temporal network (DAT-Net) was also considered in [20].

The autoencoder (AE) is a kind of neural network, and it is trained to copy the input to the output. In [11], the AE was used to map the inputs into the corresponding DNN network. When the input data contains noise, we can obtain the noiseless data by denoising autoencoder (DAE) proposed by [21]. And replace the hidden layers in the DAE with the convolution layers, then the convolution denoising autoencoder (CDAE), which is widely used in the field of image processing [22].

Hybrid architecture has been popular in the research of DOA estimation with large-scale arrays, but there are also some problems that arise, such as NOSA architecture can significantly reduce the hardware costs, but at the expense of estimation performance, and fully-connected (FC) architecture requires much more phase shifters than NOSA to get better performance. Therefore, in order to achieve a better balance between DOA estimation performance and hardware costs for hybrid architectures, the OSA-based hybrid massive MIMO receive array is proposed in this work, which can be viewed as the common form of NOSA and FC. In addition, to further improve the DOA estimation accuracy, we propose a novel DL-based method, which combines CDAE and DNN, where CDAE can remove the noise of sample covariance matrix, and DNN is used to perform high-precision DOA estimation. Therefore, our main contributions are summarized as follows:

1. Different from [6, 7], the OSA architecture is implied in hybrid massive MIMO receive array to achieve a better balance between DOA estimation accuracy and hardware complexity. As the number of elements in each subarray is the same as the

NOSA architecture, the OSA architecture has more radio-frequency (RF) chains to achieve a larger virtual aperture, and it can get more accurate estimation results. The simulation results also show that the OSA has better performance than NOSA when the SNR and the number of snapshots are low. And the CRLB for the special HAD-OSA architecture is also given in this work.

- In order to solve the DOA estimation problem for HAD-OSA architecture, a DL-based method called CDAE-DNN is also proposed in this letter. To improve the accuracy of the estimator under poor situations with low SNR and small number of snapshots, the inputted sample covariance matrix (SCM) is first imported to the CDAE for clearing the approximation errors, and then the DNN is employed to perform the multi-classification task. Comparing the simulation results of the proposed CDAE-DNN, MUSIC, and CNN in [16], it is obvious that the CDAE-DNN has significant advantages over the other methods, especially when  $SNR \leq -6$  dB and  $N \leq 2000$ .

**Notation:** Matrices, vectors, and scalars are denoted by letters of bold upper case, bold lower case, and lower case, respectively. Signs  $(\cdot)^T$  and  $(\cdot)^H$  represent transpose and conjugate transpose.  $\mathbf{I}$  and  $\mathbf{0}$  denote the identity matrix and matrix filled with zeros.  $\text{Re}\{\cdot\}$  and  $\text{Im}\{\cdot\}$  represent the real part and imaginary part of a complex number.  $\text{tr}(\cdot)$  stands for trace operation, and  $\odot$  denotes Hadamard product.

## 2 System model

The diagram of a DOA estimation system with the hybrid massive MIMO receive array is shown in Fig. 1. This hybrid array is equipped with an  $M$ -antennas uniform linear array (ULA), and antennas are connected to  $K$  RF chains via the OSA architecture. As shown in Fig. 2, we make a comparison between the proposed hybrid OSA architecture and other hybrid architectures, let  $M_s$  represent the number of antennas connected to the same RF chain, and these  $M$  antennas form a subarray, then the number of repeated antennas between two adjacent subarrays is denoted by  $\Delta M_s$ . Therefore, for a hybrid array with OSA architecture, given  $M$  and  $K$ , we can get the following relationship

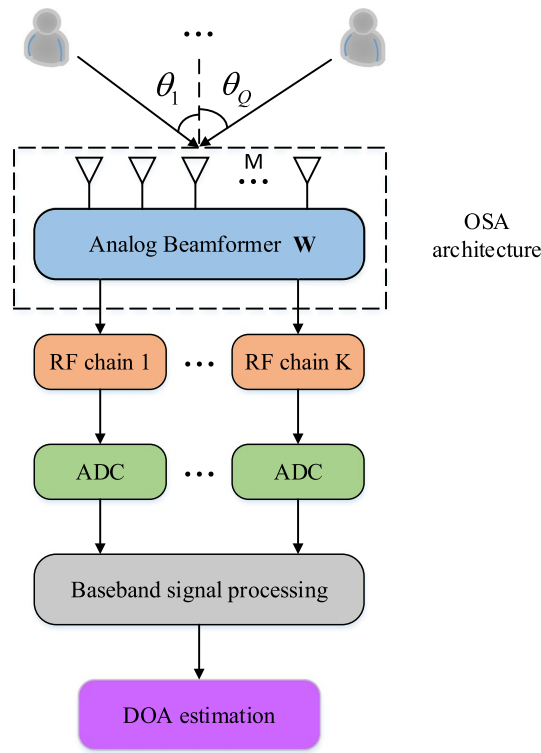
$$M = KM_s - (K - 1)\Delta M_s, \tag{1}$$

it is obviously that when  $\Delta M_s = 0$  and  $\Delta M_s = M_s$ , the OSA architecture is transformed to NOSA architecture and FC architecture, respectively.

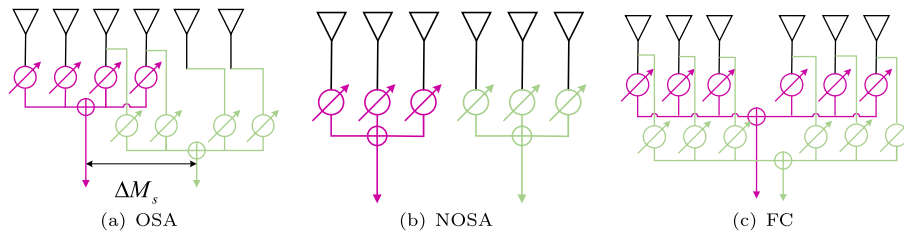
Then, back to the DOA estimation problem, we assume  $Q$  narrowband signals from different far-field sources impinge on the hybrid array with OSA architecture, and the  $q$ -th signal is denoted by  $\tilde{s}_q(t) = s_q(t)e^{j2\pi f_c t}$ , where  $s_q(t)$  is baseband signal and  $f_c$  is carrier frequency. Then, the signal received by the  $m$ -th antenna is given as [5]

$$x_m(t) = \sum_{q=1}^Q \tilde{s}_q(t)e^{-j2\pi f_c \tau_{q,m}} + v_m(t), \tag{2}$$

where  $v_m(t) \sim \mathcal{CN}(0, \sigma_v^2)$  is the additive white Gaussian noise (AWGN) term and  $\tau_{q,m}$  represents the propagation delay of the  $q$ -th signal to  $m$ -th antenna, it is expressed by



**Fig. 1** An  $M$ -antennas hybrid massive MIMO array with overlapped subarray (OSA) architecture performs DOA estimation via receiving signals from  $Q$  different sources



**Fig. 2** Examples of OSA and other hybrid array architectures, corresponding to the part of Fig. 1 framed by dashed lines, where  $K = 2$  and  $M = 6$ . **a** OSA with  $M_s = 4$  and  $\Delta M_s = 2$ , **b** non-overlapped subarray (NOSA) with  $M_s = 3$  and  $\Delta M_s = 0$ , **c** fully-connected (FC) with  $M_s = \Delta M_s = 6$

$$\tau_{q,m} = \tau_0 - \frac{(m - 1)d \sin \theta_q}{c}, \tag{3}$$

where  $\tau_0$  is the propagation delay from the signal to a reference point on the ULA,  $d$  and  $c$  denote the antenna space and light speed, respectively, and  $\theta_q$  is the DOA of the  $q$ -th signal. Then, by combining the received signals of all the  $M$  antennas, and times the analog beamforming matrix  $\mathbf{W}$ , the output signal of the  $K$  RF chains can form a  $K \times 1$  vector, which is given by

$$\mathbf{y}(t) = [y_1(t), y_2(t), \dots, y_K(t)]^T = e^{j2\pi f_c t} \mathbf{W}^H \mathbf{A}(\boldsymbol{\theta}) \mathbf{s}(t) + \mathbf{W}^H \mathbf{v}(t), \tag{4}$$

where  $y_k(t)$  is the output signal of the  $k$ -th RF chain,  $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_Q(t)]^T \in \mathbb{C}^{Q \times 1}$  and  $\mathbf{v}(t) = [v_1(t), v_2(t), \dots, v_M(t)]^T \in \mathbb{C}^{M \times 1}$ .  $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_Q)] \in \mathbb{C}^{M \times Q}$  is the array steering matrix and

$$\mathbf{a}(\theta_q) = [1, e^{j\frac{2\pi}{\lambda}d \sin \theta_q}, \dots, e^{j\frac{2\pi}{\lambda}(M-1)d \sin \theta_q}]^T, \tag{5}$$

where  $\lambda$  denotes the signal wavelength. The analog beamforming matrix  $\mathbf{W} \in \mathbb{C}^{M \times K}$  is expressed by

$$\mathbf{W} = [\mathbf{W}(1), \mathbf{W}(2), \dots, \mathbf{W}(K)], \tag{6}$$

where  $\mathbf{W}(k)$  represents the  $k$ -th column of  $\mathbf{W}$  and it is given as [7]

$$\mathbf{W}(k) = \left[ \mathbf{0}_{(k-1)(M_s-\Delta M_s)}^T \quad \mathbf{w}_k^T \quad \mathbf{0}_{(K-k)(M_s-\Delta M_s)}^T \right]^T, \tag{7}$$

where  $\mathbf{0}_L$  is an  $L \times 1$  vector filled with 0, and

$$\mathbf{w}_k = [w_{k,1}, w_{k,2}, \dots, w_{k,M_s}]^T, \tag{8}$$

where  $w_{k,m_s} = \frac{1}{\sqrt{M_s}} e^{j\alpha_{k,m_s}}$ ,  $\alpha_{k,m_s}$  is the corresponding phase of  $m_s$ th phase shifter in the  $k$ th subarray. Finally, after the down conversion and ADC operations, we can get the discrete baseband signal based on (4)

$$\mathbf{y}(n) = \mathbf{W}^H \mathbf{A}(\boldsymbol{\theta}) \mathbf{s}(n) + \mathbf{W}^H \mathbf{v}(n), \tag{9}$$

where  $\mathbf{y}(n) = \mathbf{y}(nT_s)$ ,  $1 \leq n \leq N$  and  $T_s$  denotes the sampling interval,  $\mathbf{v}(n) \sim \mathcal{CN}(\mathbf{0}, \sigma_v^2 \mathbf{I}_M)$ .

For the signal waveform  $\mathbf{s}(n)$ , there are usually two modeling approaches, the first means treating  $\mathbf{s}(n)$  as an unknown deterministic vector, and the second means considering it as a stochastic process[23]. In [24] and [5], the maximum-likelihood estimators were proposed based on the assumption of deterministic signal waveform. But in this paper, we let  $\mathbf{s}(t)$  be a zero-mean Gaussian random vector, and its covariance matrix is denoted by

$$\mathbf{C}_s = E[\mathbf{s}(n) \mathbf{s}^H(n)]. \tag{10}$$

Therefore, based on the statistical properties of  $\mathbf{s}(n)$  and  $\mathbf{v}(n)$ , we can know  $\mathbf{y}(n)$  is also a zero-mean Gaussian random vector and its covariance matrix can be expressed by

$$\mathbf{C} = E[\mathbf{y}(n) \mathbf{y}^H(n)] = \mathbf{W}^H (\mathbf{A} \mathbf{C}_s \mathbf{A}^H + \sigma_v^2 \mathbf{I}_M) \mathbf{W} \tag{11}$$

However, since the steering matrix  $\mathbf{A}$  and noise power are unknown, the covariance matrix  $\mathbf{C}$  is usually unavailable in practice, then the sample covariance matrix  $\tilde{\mathbf{C}}$  can be employed as an approximation

$$\tilde{\mathbf{C}} = \frac{1}{N} \sum_{n=1}^N \mathbf{y}(n) \mathbf{y}^H(n) = \mathbf{C} + \boldsymbol{\varepsilon} \tag{12}$$

where  $\boldsymbol{\varepsilon}$  denotes the approximation error. And if  $N \rightarrow \infty$ , we can get  $\tilde{\mathbf{C}} = \mathbf{C}$ .

### 3 Deep learning-based DOA estimation method for HAD-OSA

In this section, we propose a DNN-based DOA estimator called CDAE-DNN to improve the accuracy of DOA estimation for hybrid massive array with OSA architecture. The proposed CDAE-DNN is composed of two parts, a CDAE network is used to recover the covariance matrix of the received signals from the approximation errors, and a DNN can realize the high-resolution DOA estimation. A diagram of CDAE-DNN is shown in Fig. 3. The CDAE is composed of an encoder and a decoder, which have symmetrical convolution layers. The feature matrices extracted from SCM are sent to CDAE for denoising the approximation errors, and the input vectors of the next part are extracted from the denoised SCM. Then, the DNN-based estimator is a multilayer neural network, and each element of its output layer corresponding to a direction of the considered spatial region. Therefore, the proposed CDAE-DNN relies on the denoised input provided by CDAE and high-resolution estimator to improve the DOA estimation accuracy.

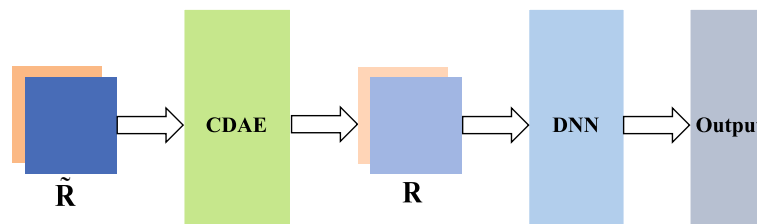
#### 3.1 Data preprocessing

To ensure the stability of the input data and improve the accuracy of the neural network model, we choose the sampled covariance matrix  $\tilde{\mathbf{C}}$  as the input feature, which is an alternative to the unavailable covariance matrix  $\mathbf{C}$ . However, the input of neural networks must be real numbers, so we extract both the real part and the imaginary part of  $\tilde{\mathbf{C}}$  and construct a  $K \times K \times 2$  tensor  $\tilde{\mathbf{R}}$ , i.e.,  $\tilde{\mathbf{R}}_{:,:,1} = \text{Re}\{\tilde{\mathbf{C}}\}$  and  $\tilde{\mathbf{R}}_{:,:,2} = \text{Im}\{\tilde{\mathbf{C}}\}$ .

Then, the label vector  $\mathbf{z} = [z_1, z_2, \dots, z_L]^T$  of input data is defined as follows. Firstly, we assume the angular region containing all the emitters is  $[-\theta_0, \theta_0]$ , and the label interval is  $\Delta\theta$  which is determined by resolution requirement. Therefore, the length of  $\mathbf{z}$  is given as  $L = \frac{2\theta_0}{\Delta\theta} + 1$ . And  $\mathbf{z}$  is a binary vector containing label 1 at the positions corresponding to the  $Q$  training angles and label 0 at the rest positions. So the training dataset can be finally expressed by  $\{(\tilde{\mathbf{R}}^{(1)}, \mathbf{z}^{(1)}), (\tilde{\mathbf{R}}^{(2)}, \mathbf{z}^{(2)}), \dots, (\tilde{\mathbf{R}}^{(T)}, \mathbf{z}^{(T)})\}$ .

#### 3.2 CDAE

The traditional autoencoder (AE) is a kind of neural network consisting of three parts: encoder, code, and decoder. The input data are first compressed to a lower dimension form, i.e., code, by the encoder, and then the decoder recovers the code to the initial form of the input data. Encoder and decoder have symmetric neural network architectures to perform the opposite operations, so the traditional autoencoders can be summarized as a two-step process



**Fig. 3** The procedure of proposed CDAE-DNN. This network is consisting of two parts, convolution denoise autoencoder (CDAE) and fully-connected deep neural network (DNN), the sample covariance matrix  $\tilde{\mathbf{R}}$  is firstly input to CDAE, then the denoised covariance matrix  $\mathbf{R}$  is sent to DNN for DOA estimation

$$\mathbf{r} = f(\tilde{\mathbf{R}}) \quad \tilde{\mathbf{R}} = g(\mathbf{r}), \tag{13}$$

where  $f(\cdot)$  and  $g(\cdot)$  denote encode and decode, respectively,  $\mathbf{r}$  represents the code.

Since the input data contains error  $\varepsilon$ , we consider using DAE technique to remove it. Different from the conventional autoencoders, DAE accepts corrupted data as input and is trained to predict the initial uncorrupted data as output. That is, the two-step process of the autoencoder is transformed to:  $\mathbf{r} = f(\tilde{\mathbf{R}})$  and  $\tilde{\mathbf{R}} = g(\mathbf{r})$ . Otherwise, because the input data are a  $K \times K \times 2$  tensor, we consider using the convolution network to implement the function of both the encoder and the decoder to improve accuracy. Next, we are going to introduce the complete procedure of CDAE.

Firstly, assuming the encoder is constructed by a  $H$ -layers convolution network, the encode function can be modified as

$$\mathbf{r} = f(\tilde{\mathbf{R}}) = f_H\left(f_{H-1}\left(\dots f_1(\tilde{\mathbf{R}})\right)\right), \tag{14}$$

and each layer contains a convolution layer, a batch normalization (BN) layer and an activation layer. For the  $H$  convolution layers, each has  $G_h$  filters  $h \in \{1, 2, \dots, H\}$ . Since the input data are 2-channel, the size of the first convolution layer is  $\kappa_1 \times \kappa_1 \times 2 \times G_1$ . And the sizes of the other  $H - 1$  convolution layers are given by  $\kappa_h \times \kappa_h \times G_h$ . Therefore, the output of the  $h$ th convolution layer can be denoted by  $\mathbf{F}_h \in \mathbb{R}^{D_h \times D_h \times G_h}$  and its  $u$ th channel,  $u \in \{1, 2, \dots, G_h\}$ , is given by

$$\mathbf{F}_{h,u} = c(\mathbf{K}_{h,u}, \mathbf{r}_{h-1}, \delta_h) + \mathbf{b}_{h,u}, \tag{15}$$

where  $c(\cdot)$  denotes convolution operation,  $\mathbf{K}_{h,u}$  represents the  $u$ th filter in the  $h$ th convolution layer.  $\mathbf{r}_{h-1}$  is the output of the corresponding layer in the encoder,  $\mathbf{r}_0 = \tilde{\mathbf{R}}$ .  $\delta_h$  denotes the stride.  $\mathbf{b}_{h,u}$  is the bias matrix of the  $u$ th filter. The activation function adopted here is RELU, so that the layer output of the encoder can be obtained as

$$\mathbf{r}_{h,u} = \text{RELU}(\mathbf{F}_{h,u}) = \max(\mathbf{0}, \mathbf{F}_{h,u}), \tag{16}$$

and  $\mathbf{r} = \{\mathbf{r}_{H,u}\}_{u=1}^{G_H}$ .

Contrary to the encoder, the decoder is required to restore the extracted feature to the form of the original input, which is an upsampling process, also called deconvolution in [25]. Similar to (14), the decode function is expressed by

$$\hat{\mathbf{R}} = g(\mathbf{r}) = g_H(g_{H-1}(\dots g_1(\mathbf{r}))), \tag{17}$$

since the structure of the decoder is symmetric with the encoder, each layer of decoder also contains convolution, BN and activation layers. And the layers' sizes are the same as that of encoder, i.e.,  $\text{size}(g_h) = \text{size}(f_{H-h+1})$ . It is obvious that in practical application the DAE cannot completely remove the noise  $\varepsilon$ , so the output of the decoder here is  $\hat{\mathbf{R}}$  rather than  $\mathbf{R}$ .

In the DAE training period, our goal is to find the optimal network parameters based on the training dataset. Thus, we choose MSE as the loss function, and it is defined as

$$\mathcal{L}(\Theta) = \frac{1}{T} \sum_{i=1}^T \left\| \hat{\mathbf{R}}^{(i)} - \mathbf{R}^{(i)} \right\|_F^2, \tag{18}$$

where  $i$  denotes the index of training data,  $\Theta$  contains all the weights and biases in the DAE network.

### 3.3 Proposed CDAE-DNN

As shown in Fig. 3, the extracted feature tensor  $\tilde{\mathbf{R}}$  is first inputted to the CDAE for eliminating the estimation error  $\boldsymbol{\varepsilon}$ . Then, the output  $\hat{\mathbf{R}}$  is inputted to a  $(H_d + 2)$ -layers DNN. The first layer is a flatten layer, which is used for transforming  $\hat{\mathbf{R}}$  into a  $2K^2 \times 1$  vector. And it is followed by  $H_d$  dense layers, each containing  $G_{h_d}$  neurons,  $h_d \in \{1, 2, \dots, H_d\}$ . We also choose RELU as the activation function for them, and to achieve regularization in the learning process, the dropout ratio is set as 20%. Therefore, the output of  $h_d$ th dense layer is given as

$$\mathbf{r}_{h_d} = \text{RELU}(\mathbf{w}_{h_d} \mathbf{r}_{h_d-1} + \mathbf{b}_{h_d}), \tag{19}$$

where  $\mathbf{w}_{h_d}$  and  $\mathbf{b}_{h_d}$  denote weight vector and bias vector, respectively. When  $h_d = 1$ ,  $\mathbf{r}_0 = \text{vec}(\hat{\mathbf{R}})$ .

The last layer of the DNN is the output layer with  $L$  neurons, and the form of the final output vector is expressed as

$$\hat{\mathbf{z}} = [\hat{z}_1, \hat{z}_2, \dots, \hat{z}_L]^T. \tag{20}$$

In order to satisfy  $0 \leq z_l \leq 1$ ,  $l \in \{1, 2, \dots, L\}$ , the activation function for this layer can use sigmoid, which is defined as

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{21}$$

Then, the  $Q$  biggest elements are selected from  $\hat{\mathbf{z}}$ , and their corresponding angles are the estimation results.

Since this is a multi-label problem and we want the final output vectors in the form of probability distributions, we decide to use the binary cross-entropy (BCE) as loss function, which is given by [16]

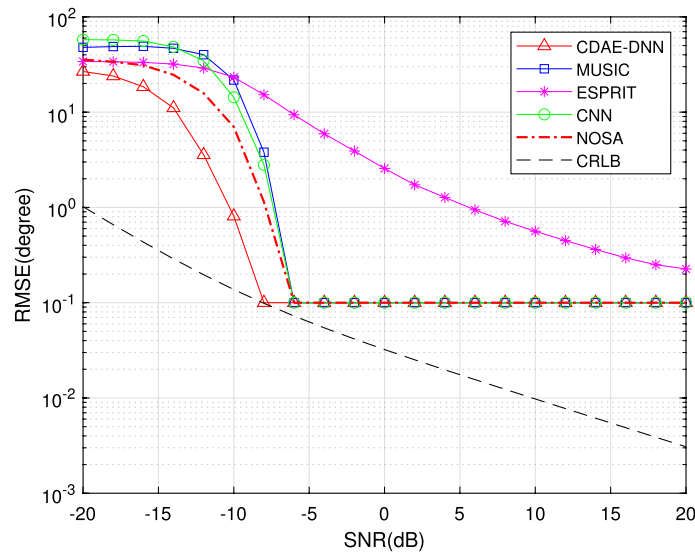
$$\mathcal{L}_{FC}(\Theta_{FC}) = -\frac{1}{L} \sum_{l=1}^L \left[ z_l^{(i)} \log(\hat{z}_l^{(i)}) + (1 - z_l^{(i)}) \log(1 - \hat{z}_l^{(i)}) \right], \tag{22}$$

where  $i \in \{1, 2, \dots, T\}$ , then the optimal weights and biases of the DNN can be obtained by minimizing it.



**Table 1** The values of simulation parameters

Parameters	Values
$M$	128
$M_s$	16
$\Delta M_s$	6
$K$	15
$Q$	1
Angular range	$[-90^\circ, 90^\circ]$
$\Delta\theta$	$1^\circ$



**Fig. 4** Comparison of DOA estimation performance between different methods and different hybrid architectures. This demonstrates the excellent performance of the proposed CDAE-DNN at low SNR

### 4 Simulation results

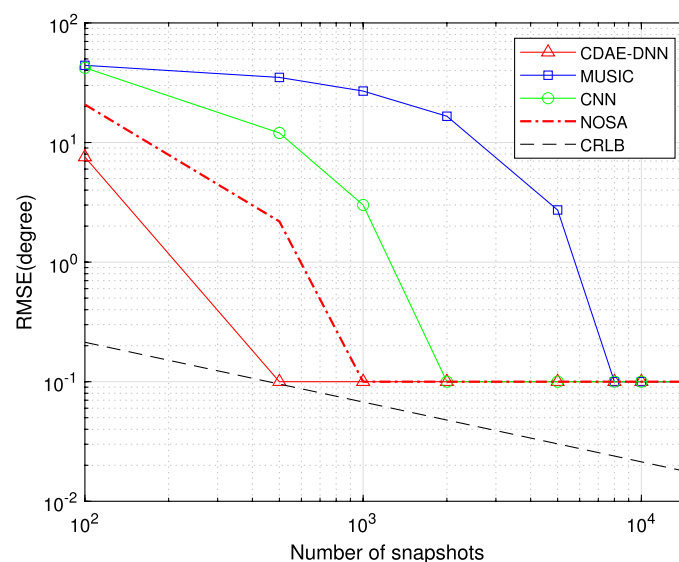
In this section, the simulation results are provided for evaluating the performance of the proposed DNN-based DOA estimator for the HAD-OSA architecture, and all the simulations related to deep learning are done based on TensorFlow and Matlab platforms. The batch size and the number of epochs are set as 1000 and 30. We also choose stochastic gradient descent (SGD) algorithm as the optimizer and learning rate is set as 0.1. And the other parameters considered in the simulation are listed in Table 1.

Figure 4 displays how the DOA estimation accuracy varies with the improvement of SNR. In this simulation, the direction of signal source is set as  $\theta = -35.9^\circ$ , the number of snapshots is  $N = 100$ , the range of SNR is -20 dB to 20 dB, and all the simulation results are averaged over 5000 Monte-Carlo experiments. Except for the proposed CDAE-DNN in this work, we also take NOSA architecture and three existing methods into consideration as benchmarks. Since predicting DOA by using DNN is essentially a multi-classification problem, and the implementation principle of the MUSIC algorithm is also based on the grid search, then there is a lower bound on the estimation accuracy of these methods when the angle to be estimated is off-grid, as shown in Fig. 4.

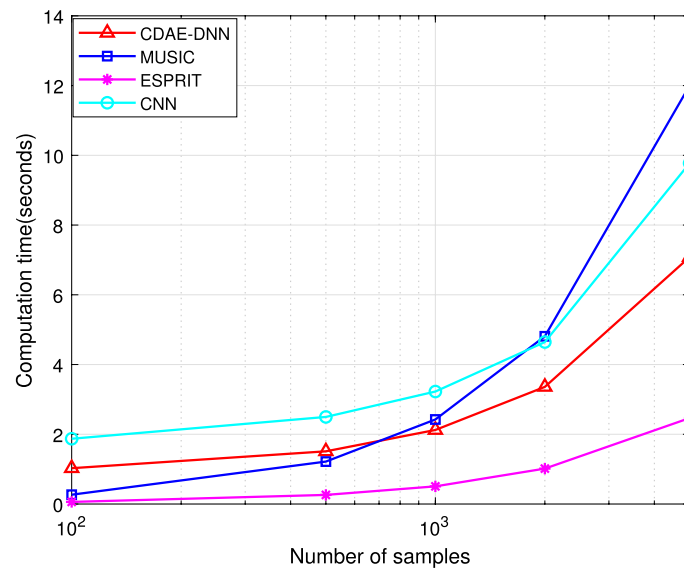
This lower bound is dependent on the grid size, which is set as  $1^\circ$  in this simulation, and hence the best estimation RMSE of grid-based methods is  $0.1^\circ$ . For the classical off-grid DOA estimation algorithms like ESPRIT, when applied to arrays with HAD structures, they will be affected by the phase ambiguity problem, and resulting in a severe decrease in estimation accuracy as shown in Fig. 4. We can see in this figure, the proposed CDAE-DNN has great advantages in the low SNR region especially when  $\text{SNR} \leq -8$  dB, for example, the rmse of CDAE-DNN at -12 dB is  $3.56^\circ$ , while that for MUSIC and CNN are  $40.11^\circ$  and  $34.65^\circ$ , respectively, so our proposed method can improve the DOA estimation accuracy more than  $30^\circ$  over traditional methods. Then, compared to NOSA architecture, the accuracy of proposed OSA architecture is approximately  $10^\circ$  higher at low SNR.

Figure 5 shows the relationship between RMSE and the number of snapshots in the environment with  $\text{SNR} = -13$  dB. The error decreases as  $N$  increases and eventually reaches the accuracy lower-bound of  $0.1^\circ$ . As can be seen in this figure, our proposed method has great performance advantages under low number of snapshots, it can achieve the lower bound at  $N = 500$ , but MUSIC need  $N = 8000$  and CNN need  $N = 2000$  to achieve that, and the rmse of MUSIC and CNN at  $N = 500$  are  $35^\circ$  and  $12^\circ$  respectively, so CDAE-DNN can save a lot of resource overhead compared with the traditional algorithms. And OSA needs only half the number of snapshots to achieve the same DOA estimation performance as NOSA.

To investigate the computational complexity of the proposed method, the relationship between the computation time and the sample number in the test set is given in Fig. 6. For two DL-based methods, CDAE-DNN and CNN, their computational complexities come from the generation of test sets as well as the processing of the neural networks. It can be seen that the computational time of the proposed method is lower than that of the CNN, the time difference is about 1 s as the number of samples less than 2000.



**Fig. 5** DOA estimation performance of different methods versus the number of snapshots. This indicates the performance advantage of the proposed CDAE-DNN with lower  $N$



**Fig. 6** Comparison of computation time between CDAE-DNN, MUSIC, CNN, and ESPRIT. This demonstrates the proposed CDAE-DNN has lower computation complexity than traditional on-grid methods

Then compared with MUSIC, the computation time of CDAE-DNN is slightly higher when the sample number less than 500. However, as the number increases to about  $10^3$ , its computation time is lower than that of MUSIC algorithm, and the largest gap can achieve 5 s as sample number grows to 5000.

## 5 Conclusion

In this paper, the OSA-based hybrid architecture is considered for DOA estimation with massive MIMO receive array, which can achieve a better balance between estimation resolution and hardware costs, and we also derive the CRLB for the OSA architecture. The simulation results show that OSA has better performance than NOSA especially when  $\text{SNR} \leq -8$  dB, and the maximum accuracy difference is  $13.6^\circ$ . Then in order to further improve the resolution of DOA estimation, the DL-based method CDAE-DNN is also proposed. This method is composed of two parts, where CDAE is used to remove the approximation error from SCM and then the denoised covariance matrix is sent to DNN for DOA estimation. In the simulation section, the proposed CDAE-DNN is compared to MUSIC and CNN, results show that CDAE-DNN is much better than other methods at  $\text{SNR} \leq -8$  dB and  $N \leq 1000$ , and the computation complexity of CDAE-DNN is also lower when the number of samples greater than  $10^3$ . The simulation results prove that CDAE-DNN has better performance under poor communication conditions and saves more computation resources.

## 6 Future works

The OSA architecture is a considerable solution for DOA estimation problems with large-scale antenna arrays, for it can bring a trade-off between estimation accuracy and hardware complexity. Especially, the OSA architecture can be generalized to two special cases, FC and NOSA, which are two most widely considered hybrid architectures, so the OSA can

provide new ideas for the studies related to hybrid arrays, rather than considering only one aspect of performance or hardware complexity. DL is an another exciting technique utilized in this work, its importance has been proved by many works, and DL methods are gradually considered for DOA estimation, but mostly ordinary DNNs and CNNs. In the future, we will try to bring more new DL techniques such as recurrent neural network (RNN), graph learning, transformer, etc., and design some new integrated methods like the CDAE-DNN proposed in this paper.

**Appendix: Derivation of CRLB for HAD-OSA**

Referring to the derivation in [10], the Fisher information matrix (FIM) related to  $\theta$  is given as

$$\mathbf{F} = \begin{bmatrix} F_{\theta_1\theta_1} & \cdots & F_{\theta_1\theta_Q} \\ \vdots & \ddots & \vdots \\ F_{\theta_Q\theta_1} & \cdots & F_{\theta_Q\theta_Q} \end{bmatrix} \tag{23}$$

and its element  $F_{\theta_p\theta_q}$  can be expressed as

$$F_{\theta_p\theta_q} = \text{tr} \left( \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_p} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_q} \right) \tag{24}$$

where  $1 \leq p, q \leq Q$ ,  $\mathbf{C}$  is the covariance matrix of received signal and

$$\begin{aligned} \frac{\partial \mathbf{C}}{\partial \theta_q} &= \frac{\partial \tilde{\mathbf{A}}}{\partial \theta_q} \mathbf{C}_s \tilde{\mathbf{A}}^H + \tilde{\mathbf{A}} \mathbf{C}_s \frac{\partial \tilde{\mathbf{A}}^H}{\partial \theta_q} \\ &= \mathbf{W}^H \mathbf{D}_q \mathbf{C}_s \tilde{\mathbf{A}}^H + \tilde{\mathbf{A}} \mathbf{C}_s \mathbf{D}_q^H \mathbf{W} \\ &= \mathbf{W}^H \mathbf{D}_q \mathbf{e}_q \mathbf{e}_q^T \mathbf{C}_s \tilde{\mathbf{A}}^H + \tilde{\mathbf{A}} \mathbf{C}_s \mathbf{e}_q \mathbf{e}_q^T \mathbf{D}_q^H \mathbf{W} \end{aligned} \tag{25}$$

where  $\tilde{\mathbf{A}} = \mathbf{W}^H \mathbf{A}$ ,  $\mathbf{e}_q$  denotes the  $q$ th column of identity matrix  $\mathbf{I}_Q$  and

$$\mathbf{D} = \sum_{q=1}^Q \mathbf{D}_q, \tag{26}$$

$$\mathbf{D}_q = [\mathbf{0}_{M \times (q-1)} \quad \mathbf{d}_q \mathbf{a}(\theta_q) \quad \mathbf{0}_{M \times (Q-q)}], \tag{27}$$

$$\mathbf{d}_q = \text{diag} \left\{ 0, j \frac{2\pi}{\lambda} d \cos \theta_q, \dots, j \frac{2\pi}{\lambda} (M-1) d \cos \theta_q \right\}, \tag{28}$$

according to the equation  $\text{tr}(\mathbf{A}^H) = \text{tr}(\mathbf{A})^*$  we can get

$$\begin{aligned}
 F_{\theta_p \theta_q} &= \text{tr} \left\{ (\mathbf{F}_1 + \mathbf{F}_1^H) \mathbf{C}^{-1} (\mathbf{F}_2 + \mathbf{F}_2^H) \mathbf{C}^{-1} \right\} \\
 &= \text{tr} \left\{ \mathbf{F}_1 \mathbf{C}^{-1} \mathbf{F}_2 \mathbf{C}^{-1} \right\} + \text{tr} \left\{ \mathbf{F}_1 \mathbf{C}^{-1} \mathbf{F}_2^H \mathbf{C}^{-1} \right\} + \\
 &\quad \text{tr} \left\{ \mathbf{F}_1^H \mathbf{C}^{-1} \mathbf{F}_2 \mathbf{C}^{-1} \right\} + \text{tr} \left\{ \mathbf{F}_1^H \mathbf{C}^{-1} \mathbf{F}_2^H \mathbf{C}^{-1} \right\} \\
 &= \text{tr} \left\{ \mathbf{F}_1 \mathbf{C}^{-1} \mathbf{F}_2 \mathbf{C}^{-1} \right\} + \text{tr} \left\{ \mathbf{F}_1 \mathbf{C}^{-1} \mathbf{F}_2 \mathbf{C}^{-1} \right\}^* + \\
 &\quad \text{tr} \left\{ \mathbf{F}_1 \mathbf{C}^{-1} \mathbf{F}_2^H \mathbf{C}^{-1} \right\} + \text{tr} \left\{ \mathbf{F}_1 \mathbf{C}^{-1} \mathbf{F}_2^H \mathbf{C}^{-1} \right\}^* \\
 &= 2\text{Re} \left\{ \text{tr} \left\{ \mathbf{F}_1 \mathbf{C}^{-1} \mathbf{F}_2 \mathbf{C}^{-1} \right\} + \text{tr} \left\{ \mathbf{F}_1 \mathbf{C}^{-1} \mathbf{F}_2^H \mathbf{C}^{-1} \right\} \right\}
 \end{aligned} \tag{29}$$

where  $\mathbf{F}_1 = \mathbf{W}^H \mathbf{D}_p \mathbf{C}_s \tilde{\mathbf{A}}^H$  and  $\mathbf{F}_2 = \mathbf{W}^H \mathbf{D}_q \mathbf{C}_s \tilde{\mathbf{A}}^H$ . Since  $\mathbf{D}_q = \mathbf{D} \mathbf{e}_q \mathbf{e}_q^T$ , then the equation above can be further transformed as

$$\begin{aligned}
 F_{\theta_p \theta_q} &= 2\text{Re} \left\{ \text{tr} \left\{ \mathbf{W}^H \mathbf{D} \mathbf{e}_p \mathbf{e}_p^T \mathbf{C}_s \tilde{\mathbf{A}}^H \mathbf{C}^{-1} \mathbf{W}^H \mathbf{D} \mathbf{e}_q \mathbf{e}_q^T \mathbf{C}_s \tilde{\mathbf{A}}^H \mathbf{C}^{-1} \right\} \right. \\
 &\quad \left. + \text{tr} \left\{ \mathbf{W}^H \mathbf{D} \mathbf{e}_p \mathbf{e}_p^T \mathbf{C}_s \tilde{\mathbf{A}}^H \mathbf{C}^{-1} \tilde{\mathbf{A}} \mathbf{C}_s \mathbf{e}_q \mathbf{e}_q^T \mathbf{D}^H \mathbf{W} \mathbf{C}^{-1} \right\} \right\} \\
 &= 2\text{Re} \left\{ \mathbf{e}_p^T \mathbf{C}_s \tilde{\mathbf{A}}^H \mathbf{C}^{-1} \mathbf{W}^H \mathbf{D} \mathbf{e}_q \mathbf{e}_q^T \mathbf{C}_s \tilde{\mathbf{A}}^H \mathbf{C}^{-1} \mathbf{W}^H \mathbf{D} \mathbf{e}_p \right. \\
 &\quad \left. + \mathbf{e}_p^T \mathbf{C}_s \tilde{\mathbf{A}}^H \mathbf{C}^{-1} \tilde{\mathbf{A}} \mathbf{C}_s \mathbf{e}_q \mathbf{e}_q^T \mathbf{D}^H \mathbf{W} \mathbf{C}^{-1} \mathbf{W}^H \mathbf{D} \mathbf{e}_p \right\}
 \end{aligned} \tag{30}$$

then by combining all the  $Q^2$  elements, the FIM of  $\boldsymbol{\theta}$  can be expressed by

$$\begin{aligned}
 \mathbf{F} &= 2\text{Re} \left\{ \left( \mathbf{C}_s \tilde{\mathbf{A}}^H \mathbf{C}^{-1} \mathbf{W}^H \mathbf{D} \right) \odot \left( \mathbf{C}_s \tilde{\mathbf{A}}^H \mathbf{C}^{-1} \mathbf{W}^H \mathbf{D} \right)^T \right. \\
 &\quad \left. + \left( \mathbf{C}_s \tilde{\mathbf{A}}^H \mathbf{C}^{-1} \tilde{\mathbf{A}} \mathbf{C}_s \right) \odot \left( \mathbf{D}^H \mathbf{W} \mathbf{C}^{-1} \mathbf{W}^H \mathbf{D} \right)^T \right\}.
 \end{aligned} \tag{31}$$

By collecting the signals at all the  $N$  snapshots, the CRLB is given as

$$\text{CRLB} = \frac{1}{N} \mathbf{F}^{-1}. \tag{32}$$

**Acknowledgements**

Not applicable.

**Author contributions**

Not applicable.

**Funding**

This work was supported in part by the National Natural Science Foundation of China (Nos. U22A2002, and 62071234), the Hainan Province Science and Technology Special Fund (ZDKJ2021022), and the Scientific Research Fund Project of Hainan University under Grant KYQD(ZR)-21008.

**Availability of data and materials**

Not applicable.

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

Received: 7 August 2023 Accepted: 20 October 2023

Published online: 30 October 2023

**References**

1. R.W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, A.M. Sayeed, An overview of signal processing techniques for millimeter wave MIMO systems. *IEEE J. Sel. Top. Signal Process.* **10**(3), 436–453 (2016)
2. F. Sohrabi, W. Yu, Hybrid digital and analog beamforming design for large-scale antenna arrays. *IEEE J. Sel. Topics Signal Process.* **10**(3), 501–513 (2016)
3. F. Shu, Y. Qin, T. Liu, L. Gui, Y. Zhang, J. Li, Z. Han, Low-complexity and high-resolution DOA estimation for hybrid analog and digital massive MIMO receive array. *IEEE Trans. Commun.* **66**(6), 2487–2501 (2018)
4. M. Li, B. Zhang, B. Zhang, W. Liu, T. Kim, C. Wang, Directional modulation design for multi-beam multiplexing based on hybrid antenna array structures. *EURASIP J. Adv. Signal Process.* **2023**(1), 1–16 (2023)
5. R. Zhang, B. Shim, W. Wu, Direction-of-arrival estimation for large antenna arrays with hybrid analog and digital architectures. *IEEE Trans. Signal Process.* **70**, 72–88 (2021)
6. N. Song, T. Yang, H. Sun, Overlapped subarray based hybrid beamforming for millimeter wave multiuser massive MIMO. *IEEE Signal Process. Lett.* **24**(5), 550–554 (2017)
7. A. Hassaniien, S.A. Vorobyov, Phased-MIMO radar a tradeoff between phased-array and MIMO radars. *IEEE Trans. Signal Process.* **58**(6), 3137–3151 (2010)
8. J. Meng, Z. Wei, Y. Zhang, B. Li, C. Zhao, Machine learning based low-complexity channel state information estimation. *EURASIP J. Adv. Signal Process.* **2023**(1), 1–11 (2023)
9. X. Huang, R. Xu, W. Yu, T. Peng, Research on structural sound source localization method by neural network. *EURASIP J. Adv. Signal Process.* **2023**(1), 54 (2023)
10. E. Tuncer, B. Friedlander, *Classical and Modern Direction-of-Arrival Estimation* (Academic Press, Boston, 2009)
11. Z.-M. Liu, C. Zhang, S.Y. Philip, Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections. *IEEE Trans. Antennas Propag.* **66**(12), 7315–7327 (2018)
12. D. Hu, Y. Zhang, L. He, J. Wu, Low-complexity deep-learning-based DOA estimation for hybrid massive MIMO systems with uniform circular arrays. *IEEE Wirel. Commun. Lett.* **9**(1), 83–86 (2019)
13. H. Huang, J. Yang, H. Huang, Y. Song, G. Gui, Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system. *IEEE Trans. Veh. Technol.* **67**(9), 8549–8560 (2018)
14. L. Wu, Z.-M. Liu, Z.-T. Huang, Deep convolution network for direction of arrival estimation with sparse prior. *IEEE Signal Process. Lett.* **26**(11), 1688–1692 (2019)
15. W. Zhu, M. Zhang, P. Li, C. Wu, Two-dimensional DOA estimation via deep ensemble learning. *IEEE Access* **8**, 124544–124552 (2020)
16. G.K. Papageorgiou, M. Sellathurai, Y.C. Eldar, Deep networks for direction-of-arrival estimation in low SNR. *IEEE Trans. Signal Process.* **69**, 3714–3729 (2021)
17. J. Yu, Y. Wang, Deep learning-based multipath DOAs estimation method for MM wave massive MIMO systems in low SNR. *IEEE Trans. Veh. Technol.* (2023)
18. Y. Cao, T. Lv, Z. Lin, P. Huang, F. Lin, Complex resnet aided DOA estimation for near-field MIMO systems. *IEEE Trans. Veh. Technol.* **69**(10), 11139–11151 (2020)
19. C. Ying, W. Xiang, H. Zhitao, Underdetermined DOA estimation via multiple time-delay covariance matrices and deep residual network. *J. Syst. Eng. Electron.* **32**(6), 1354–1363 (2021)
20. K. Yan, W. Jin, Y. Huang, P. Song, Z. Li, Deep adaptive temporal network (DAT-Net): an effective deep learning model for parameter estimation of radar multipath interference signals. *EURASIP J. Adv. Signal Process.* **2023**(1), 94 (2023)
21. P. Vincent, H. Larochelle, Y. Bengio, P.-A. Manzagol, Extracting and composing robust features with denoising autoencoders. In: *Proceedings of the 25th International Conference on Machine Learning*, pp. 1096–1103 (2008)
22. X.-J. Mao, C. Shen, Y.-B. Yang, Image restoration using convolutional auto-encoders with symmetric skip connections. *arXiv preprint arXiv:1606.08921* (2016)
23. C.E. Chen, F. Lorenzelli, R.E. Hudson, K. Yao, Stochastic maximum-likelihood DOA estimation in the presence of unknown nonuniform noise. *IEEE Trans. Signal Process.* **56**(7), 3038–3044 (2008)
24. M. Pesavento, A.B. Gershman, Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise. *IEEE Trans. Signal Process.* **49**(7), 1310–1324 (2001)
25. H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1520–1528 (2015)

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.