This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2019.2945115, IEEE Signal Processing Letters

Deep Convolution Network for Direction of Arrival Estimation with Sparse Prior

Liu-li Wu, Zhang-meng Liu, Zhi-tao Huang

***Abstract*—In this letter, a deep learning framework for direction of arrival (DOA) estimation is developed. We first show that the columns of the array covariance matrix can be formulated as under-sampled noisy linear measurements of the spatial spectrum. Then, a deep convolution network (DCN) that learns the inverse transformation from large training dataset is introduced. In contrast to traditional sparsity-inducing methods with computationally complex iterations, the proposed DCN-based framework could efficiently obtain DOA estimates in near real time. Moreover, the utilization of the sparsity prior improves DOA estimation performance compared to existing deep learning based methods. Simulation results have demonstrated the superiority of the proposed method in both DOA estimation precision and computation efficiency especially when SNR is low.**

***Index Terms*—direction of arrival estimation, deep convolution network, sparse representation**

I. Introduction

DIRECTION of arrival (DOA) estimation is an attractive field of research with many applications including wireless communications, astronomical observation, radar and sonar [1]–[8]. Considerable attention has been paid to this topic and long-lasting developments have been made.

Traditionally, DOA estimation was achieved by the so-called model-driven approaches [1]–[4], which first formulate forward parametric models from signal directions to array outputs then estimate the directions by exploiting the properties of the pre-assumed formulations. Performances of these model-driven methods depend heavily on the accuracy of the pre-formulated models. However, non-ideal conditions such as multipath interference and array imperfections make it almost impossible to formulate a completely accurate model, and consequently, result in significant performance degradation.

To attach this problem, learning-based (data-driven) [5]–[8] approaches which directly learn the nonlinear relationship between the array outputs and signal directions from training data have been proposed for direction finding and satisfactory results have been achieved. These methods make no pre-assumptions on the forward models, so they can provide robust performance against non-ideal conditions.

Though the idea of using machine learning techniques such as SVR [9], [10] and RBF [11] for direction finding dates back to the early nineties, these proposals were somehow limited in scope since large realistic datasets and computation resources were not viable at this time. Recently, DOA estimation using neural networks attracts renewed interest due to the fast developments in deep learning (DL) theory and computation capacity [5]–[7], [12]. Researchers in [13]–[15] introduce DL techniques to solve sound source localization problems with microphone arrays. These papers either deal with single-signal scenario [13] or locate sounds on very coarse grids with inter-grid spacings of 5° [14], [15] or even larger, so they can hardly be directly used for electromagnetics (EM) DOA estimation since high estimation precision and super-resolution of multiple overlapped signals are usually required. Liu [5] proposes an end-to-end deep neural network (DNN) to detects the presence of EM signals on predefined direction grids by a series of classifiers. [6] utilizes an autoencoder to extract the features of different directions in VHF radar rather than estimate the DOAs directly. [7] develops a DNN scheme for super-resolution DOA estimation in MIMO systems. These networks all exploit fully-connected layers.

Due to the fact that DOA estimation can be viewed as an application of the sparse linear inverse problem in compressive sensing (CS) [2], [4], [16], we propose a deep inverse network with convolution layers to learn the inverse transformation from array outputs to DOA spectrum.Compared with existing DL based DOA estimation schemes, this work has taken the spatial sparsity of the incident signals into consideration. The utilization of this prior information helps to facilitate network designing and training. Moreover, the sparse-connected convolution layer with shared weights across the entire receptive fields has faster learning speed comparing to the fully-connected networks [5]–[7].

II. Problem Formulation

Assume that K narrowband independent EM waves $\boldsymbol{s}(t) = [s_1(t), s_2(t), \ldots, s_K(t)]^{\mathrm{T}}$ impinge on an M-element array from directions $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_K]^{\mathrm{T}}$. The array output $\boldsymbol{x}(t) = [x_1(t), x_2(t), \ldots, x_M(t)]^{\mathrm{T}}$can be written as:

$$\boldsymbol{x}(t) = \sum_{k=1}^{K} \boldsymbol{a}(\theta_k) s_k(t) + \boldsymbol{v}(t) \tag{1}$$

where $\boldsymbol{a}(\theta_k)$ denotes the steering vector corresponding to θ_k and $\boldsymbol{v}(t)$ represents the independent complex Gaussian white noise with zero mean and variance σ_v^2 . Superscripts T and H denote transpose and Hermitian transpose, respectively.

Let $\boldsymbol{\Phi} = [\phi_1, \phi_2, \ldots, \phi_L]^{\mathrm{T}}$ be a discrete direction set sampled from the potential space of incident signals, with $\Delta\phi$ denoting the sampling interval. The true source directions $\boldsymbol{\theta}$

The authors are with State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System, National University of Defense Technology, Changsha, China (e-mail: wll_me1991@163.com, zm_liupaper@163.com,tald_paper@163.com)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2019.2945115, IEEE Signal Processing Letters

are contained in $\boldsymbol{\Phi}$ with moderately small quantization errors. Then $\boldsymbol{x}(t)$ can be reformulated in an overcomplete form:

$$\boldsymbol{x}(t) = \sum_{l=1}^{L} \boldsymbol{a}(\phi_l)\overline{s}_l(t) + \boldsymbol{v}(t) \tag{2}$$

where $\overline{s}_l(t) = s_k(t)$ if $\phi_l = \theta_k$ (with small quantization error), otherwise, $\overline{s}_l(t) = 0$. The spatial correlation matrix $\boldsymbol{R}$ is defined as:

$$\boldsymbol{R} = E[\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{H}}(t)] = \sum_{l=1}^{L} \eta_l \boldsymbol{a}(\phi_l)\boldsymbol{a}^{\mathrm{H}}(\phi_l) + \sigma_v^2 \boldsymbol{I} \tag{3}$$

where $\eta_l = E[\overline{s}_l(t)\overline{s}_l^{\mathrm{H}}(t)]$ represents the signal power on the lth direction grid. Similarly, $\boldsymbol{\eta} = [\eta_1, \eta_2, \ldots, \eta_L]^{\mathrm{T}}$ only has nonzero elements at the true source locations, i.e., the spatial spectrum $\boldsymbol{\eta}$ is sparse. In general, DOA estimation can be achieved by recovering $\boldsymbol{\eta}$ from $\boldsymbol{R}$.

Reconsidering Eq.(3), we find that the mth column of $\boldsymbol{R}$ can be written as:

$$\boldsymbol{y}_m = \boldsymbol{R}\boldsymbol{e}_m \triangleq \boldsymbol{A}_m\boldsymbol{\eta} + \sigma_v^2\boldsymbol{e}_m \tag{4}$$

where $\boldsymbol{A}_m = [\boldsymbol{a}(\phi_1)\boldsymbol{a}^{\mathrm{H}}(\phi_1)\boldsymbol{e}_m, \ldots, \boldsymbol{a}(\phi_L)\boldsymbol{a}^{\mathrm{H}}(\phi_L)\boldsymbol{e}_m]$. $\boldsymbol{e}_m$ is an $M \times 1$ vector with the mth element being 1 and others being 0. Then, a new measurement vector can be defined by vectorizing $\boldsymbol{R}$:

$$\boldsymbol{y} = [\boldsymbol{y}_1; \ldots; \boldsymbol{y}_M] = \tilde{\boldsymbol{A}}\boldsymbol{\eta} + \sigma_v^2\tilde{\boldsymbol{e}} \tag{5}$$

where $\tilde{\boldsymbol{A}} = [\boldsymbol{A}_1; \ldots; \boldsymbol{A}_M], \tilde{\boldsymbol{e}} = [\boldsymbol{e}_1; \ldots; \boldsymbol{e}_M]$.

In practical applications, the covariance matrix can only be estimated using N snapshots, $\hat{\boldsymbol{R}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}(t_n)\boldsymbol{x}^{\mathrm{H}}(t_n)$, so the estimated measurement vector $\hat{\boldsymbol{y}}$ is:

$$\hat{\boldsymbol{y}} = \boldsymbol{y} + \Delta\boldsymbol{y} = \tilde{\boldsymbol{A}}\boldsymbol{\eta} + \sigma_v^2\tilde{\boldsymbol{e}} + \Delta\boldsymbol{y} \triangleq \tilde{\boldsymbol{A}}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{6}$$

where $\Delta\boldsymbol{y}$ is the estimation error of $\boldsymbol{y}$, $\boldsymbol{\varepsilon} = \sigma_v^2\tilde{\boldsymbol{e}} + \Delta\boldsymbol{y}$. Eq.(6) indicates that recovering spatial spectrum $\boldsymbol{\eta}$ from measurement $\hat{\boldsymbol{y}}$ is a typical sparse linear inverse problem.

Many sparsity-driven methods have been developed to solve this problem [2], [4], including both sequential and iterative varieties. Despite of their successes, these methods still face great challenges, such as huge computation complexity or instability under conditions with low SNR. Due to the powerful capacity of DL, we propose the following deep convolutional network (DCN) to learn the inverse mapping from $\hat{\boldsymbol{y}}$ to $\boldsymbol{\eta}$.

III. Deep Convolutional Network for DOA Estimation

In this section, we propose a DCN framework where DL is integrated into super-resolution DOA estimation.

As shown in Fig.1(a), we design an L-dimension input layer, which takes the spectrum proxy $\breve{\boldsymbol{\eta}} = \tilde{\boldsymbol{A}}^{\mathrm{H}}\hat{\boldsymbol{y}}$ as the input. Then, four convolution layers are used to learn the features of the spectrum and reconstruct it in the output layer. The number of hidden layers is chosen to reach the trade-off between the nonlinear expressivity (improves with deeper networks) and overfitting risk (aggravates with more network parameters) of the network. For comparison, a counterpart DCN (Fig.1(b)) with tanh activation function and a fully-connected DNN (Fig.1(c)) are also displayed.

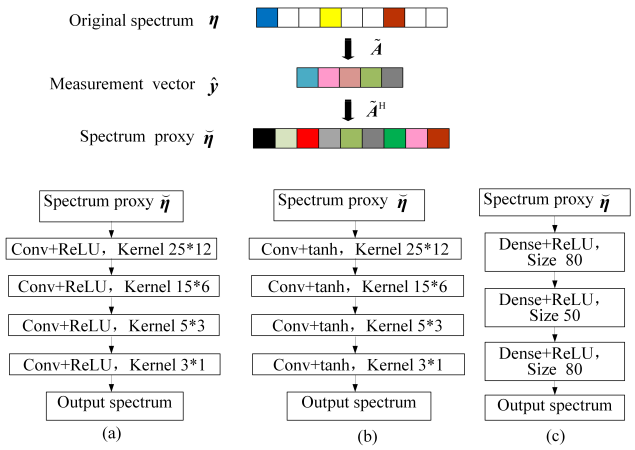


Fig. 1. Network structures of (a) proposed DCN (b) the counterpart DCN with tanh activation function (c) DNN with fully-connected layers

For an input $\breve{\boldsymbol{\eta}}$, the output of the kth layer in DCN is

$$\boldsymbol{c} = \mathcal{P}(\mathrm{ReLU}(\boldsymbol{W}^k * \boldsymbol{c}^k + \boldsymbol{b}^k)), k = 1, 2, 3, 4 \tag{7}$$

where $\boldsymbol{c}^0 = \breve{\boldsymbol{\eta}}$. The reconstructed spectrum in the output layer is $\hat{\boldsymbol{\eta}} = \boldsymbol{c}^4$. $\boldsymbol{W}^k$ and $\boldsymbol{b}^k$ represents the convolution kernel and bias corresponding to the kth layer. $*$ is the convolution operation, and ReLU denotes the activation function [17]. The padding operator $\mathcal{P}(\bullet)$ takes the output of $\mathrm{ReLU}(\bullet)$ to the original input size by zero-padding it at the borders.

The goal of network training is minimizing the mean squared error (MSE) between the reconstructed spectrum $\hat{\boldsymbol{\eta}}$ and the original spectrum $\boldsymbol{\eta}$ over the training dataset $\mathcal{D}_{train} = \{(\hat{\boldsymbol{y}}^1, \boldsymbol{\eta}^1), (\hat{\boldsymbol{y}}^2, \boldsymbol{\eta}^2), \ldots, (\hat{\boldsymbol{y}}^D, \boldsymbol{\eta}^D)\}$, i.e.

$$\{\hat{\boldsymbol{W}}^k, \hat{\boldsymbol{b}}^k\}_{k=1}^4 = \underset{\{\boldsymbol{W}^k, \boldsymbol{b}^k\}_{k=1}^4}{argmin} \frac{1}{2}\sum_{d=1}^{D} \|\hat{\boldsymbol{\eta}}^d - \boldsymbol{\eta}^d\|^2 \tag{8}$$

Once trained, the network can be used to predict $\boldsymbol{\eta}$ that corresponds to a new measurement vector $\hat{\boldsymbol{y}}$ in near real time.

IV. Experimental Results

In this section we describe the implementation of DCN[1] and compare its performance to the DL-based DOA estimation method. We choose [5] as a benchmark since its background is similar with ours and this will make a fair comparison[2]. The sparsity-inducing method based on sparse Bayesian learning (SBL) [4] is also simulated to indicate that the proposed DCN closely approximates the solution produced by CS recovery algorithms yet is hundreds of times faster in run time.

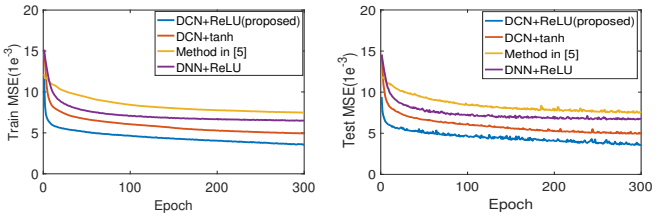
A. Experiment Setup

An 8-element uniform linear array (ULA) with half-wavelength inter-element spacing is used to estimate directions

[1]Our Python- and Matlab- based implementation can be downloaded from https://github.com/liuliWu/Deep-Convolution-Network-for-Direction-of-Arrival-stimation-with-Sparse-Prior.

[2]The code is download from https://github.com/LiuzmNUDT/DNN-DOA.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2019.2945115, IEEE Signal Processing Letters



(a) Training MSE (b) Validation MSE

Fig. 2. Training and validation MSE of all networks

of signals impinging from the spatial scope of [-60°, 60°]. This potential space is sampled with interval $\Delta\phi = 1°$, thus $L = 121$. The measurement vectors are obtained from N=256 snapshots. For the training set, $K = 2$ sources and a set of angular separations $\{1°, 2°, \ldots, 39°, 40°\}$ are chosen. For each angular separation $\Delta\varphi$, the directions of the first and second signal are uniformly generated in the range of [-60°, 60°$-\Delta\varphi$] and [-60°$+\Delta\varphi$, 60°] with a step of 1°, respectively. The SNR of both signals is randomly distributed between [-10dB, 0dB] and 10 groups of snapshots are collected for each direction setting with random noise. Finally, a total of 40200 measurement vectors are collected in the sample dataset. Our recent implementation is based on Keras [18] and Adam [19] optimizer is used. The measurement vectors are used for training with mini-batch size of 64 and the order of the vectors is shuffled during each of the 300 training epochs. The experiment platform is a PC with one Intel i7-8550U CPU.

B. Recovery MSE during training and validation

The overall training and validation MSEs of each network are analyzed with respect to the number of epochs by first randomly dividing the training samples into 80% for training and 20% for validation. This process is applied 10 times and the averaged results are reported in Fig.2. In order to illustrate the superiority of the proposed sparsely-connected convolutional layer with ReLU activation, the counterpart networks in Fig.1(b) and Fig.1(c) are also trained for comparison.

Fig.2 shows that the validation accuracy curves level off after about 150 epochs, and remain roughly constant thereafter, confirming that there is no overfitting in any of the deep architectures implemented. Moreover, it may be observed that the network with the highest performance is the proposed DCN with ReLU activation. Interestingly, we find that network with convolution layers converges faster than the counterpart with fully-connected layers and ReLU activation also increases the learning speed. As a result, the preponderance of the proposed network benefits from both the convolution operation and the the nonlinear ReLU activation.

C. Spatial spectra and DOA estimation result

This subsection gives an intuitive view of the reconstructed spectra and DOA estimation results of DCN and compares the results with two baseline methods [4], [5].

Firstly, two 0 dB narrowband signals in the far field impinge on this array from directions of 14.7° and 28.3° (off the presumed grid), respectively. This angular distance and the

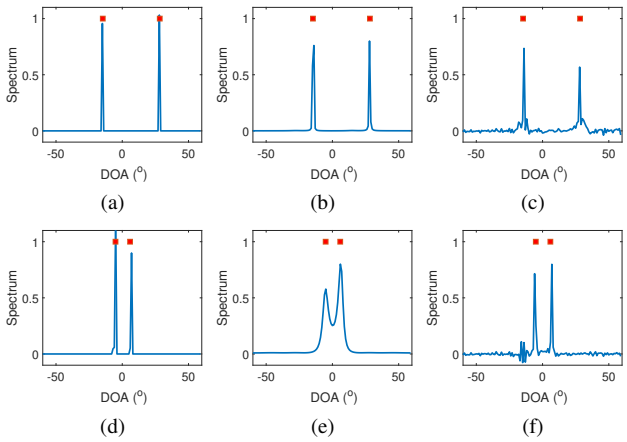


Fig. 3. Reconstructed spatial spectrum of two signals. First row: 14.7° and 28.3°. Second row: -5.1° and 5.9° (a)(d) Proposed DCN (b)(e) SBL (c)(f) Method in [5]

TABLE I
MODEL COMPLEXITY AND AVERAGED RECONSTRUCTION TIME

	DCN	Method in [5]	SBL
Total params	1801	32466	\
Train time	19min11.40s	69min12.57s	\
Test time	0.0023s	0.01320s	2.84423s

directions of the signals are not contained in the training set. The reconstructed spectrums averaged by 10 tests are shown in Fig.3(a)-(c), where the red squares indicate the true signal locations. Then we set the directions of the two signals to be -5.1° and 5.9° with the corresponding results shown in Fig.3(d)-(f).

Fig.3 shows that both the proposed DCN and SBL yield zeros on the spectrum grids with no incident signals and sharp peaks at true directions, while method in [5] gets a spectrum with many burrs which will deteriorate DOA estimation results in low SNR scenario. Moreover, when the two signals get closer, SBL has significant sidelobe while the proposed DCN still remains good recovery results. These results verified the satisfactory recovery performance and super-resolution superiority of our proposed method.

To assess the computation complexity of the proposed DCN scheme, we record the time needs for very methods in TABLE I (averaged by 1000 tests). The result shows that the proposed DCN scheme has a run time that is a tiny fraction of the current SBL algorithm. Moreover, it is more efficient than [5] too since multiple (6 in that paper) DNNs needed to be trained in [5]. This superiority makes DCN especially suitable for applications where real time DOA estimation is needed.

Then, Fig.4 displays the DOA estimates and the corresponding estimation errors of two 0dB sources with a set of angular separations $\{10.4°, 15.2°, 20.7°, 25.9°\}$ in the sector [-55°, 55°]. The final DOA estimates are obtained via amplitude interpolation within the two most significant peaks of the reconstructed spectra. It may be observed that the DOA estimates of DCN well match their true values and most of the estimation errors are smaller than 2.5°. SBL and method in [5] fail to recover the spatial spectrum sometimes since they are not robust enough when two signals are close to each other.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSP.2019.2945115, IEEE Signal Processing Letters

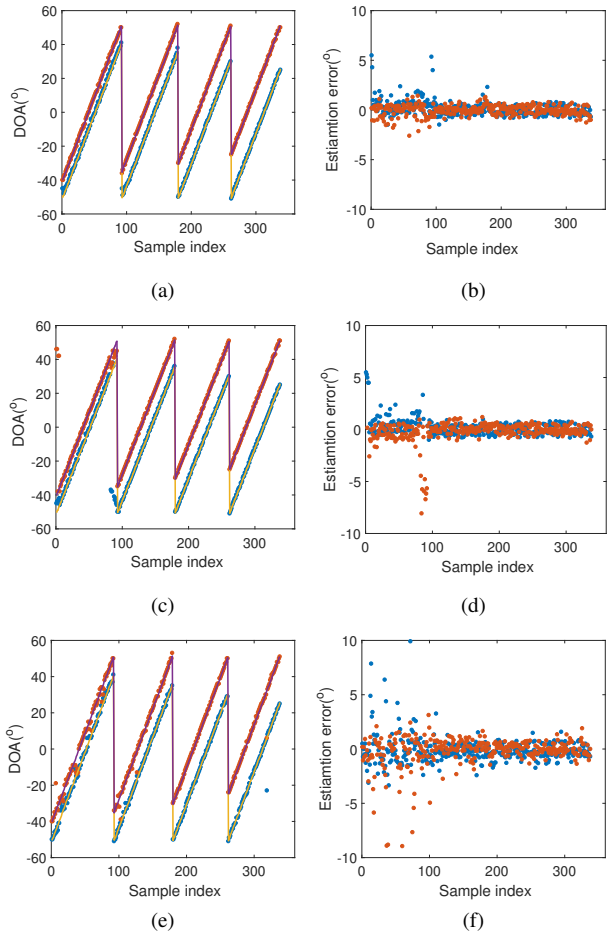


Fig. 4. DOA estimates and the corresponding errors of off-grid signals (a)(b) proposed DCN (c)(d) SBL (e)(f) Method in [5]

Finally, we show how the proposed method behaves when the testing data contains different numbers of signals as the training data. The networks have been trained with array outputs in two-signal scenarios, but tested under one-signal scenario with DOA of -14.7° and three-signal scenario with DOAs of -34.7°, 5.3°, 20.3°. The proposed DCN-based method still performs satisfyingly in these scenarios, as shown in Fig.5(a) and 5(d). SBL works well too since it needs no training so no mismatching exists. Similarly, method in [5] can also recover the spectrum but having many burrs.

D. Statistic performance

In this subsection, we use average root-mean-square-error (RMSE) of all incident signals to evaluate the statistic performance of all methods. The RMSE is defined as follows:

$$RMSE = \sqrt{\frac{1}{HK}\sum_{h=1}^{H}\|\hat{\boldsymbol{\theta}}^h - \boldsymbol{\theta}\|^2} \tag{9}$$

where $\hat{\boldsymbol{\theta}}^h$ is the estimation results in the hth test, $\boldsymbol{\theta}$ is the true signal direction set, H is times of the Monte-Carlo simulations and K is the number of signals.

First, two signals located in -10.35° and 4.65° and SNR within [-10dB, 12dB] are considered. 600 independent simulations are carried out for each SNR and the RMSE versus SNR

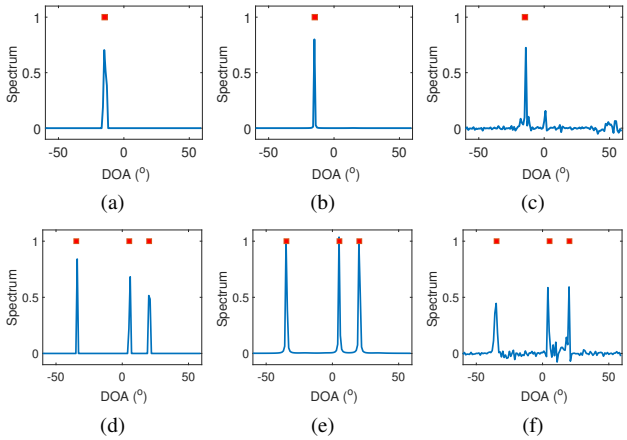


Fig. 5. Spatial spectra in one-signal and three-signal scenarios with DNN trained with two-signal dataset. First row: one-signal scenario. Second row: three-signal scenario. (a)(d) Proposed DCN (b)(e) SBL (c)(f) Method in [5]

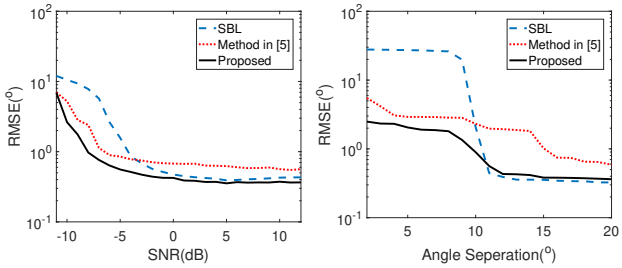


Fig. 6. Statistic Performance

curves are drawn in Fig.6(a). Then, two signals with SNR=0dB and angle separations within [2°,20°] are considered. For each angle separation $\Delta\varphi$, the signal directions are $0° + \delta$ and $\Delta\varphi + \delta$, respectively. δ is a random variable uniformly distributed in (0°,0.2°). The results are shown in Fig.6(b). We can see that the estimation precision predominance of DCN over methods in [4], [5] is significant when the SNR is low or inter-signal angle separation is small. In scenarios with high SNR and large inter-signal angle separation, DCN still offers DOA estimation performance comparable to SBL, but with much fewer computation consumption.

V. Conclusion

In this letter, an efficient DCN based spatial spectra recovery algorithm is proposed and applied to EM DOA estimation. We first convert the DOA estimation problem into a sparse linear inverse problem by introducing spatially overcomplete formulation. Then the structure designing and training procedure of DCN are introduced. Compared with conventional iterative-based sparse recovery algorithms, the DCN-based method just needs feedforward calculations, which makes it possible to achieve real time direction finding. Also, the learning and generalization capability of the convolution layers and ReLU activation helps it achieve competitive or even better DOA estimation performance when the SNR is low or inter-signal angle separation is small. Simulations and experiments have verified the superiority of the proposed method clearly.

REFERENCES

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[2] J. Dai, X. Xu, and D. Zhao, "Direction-of-arrival estimation via real-valued sparse representation," *IEEE Antennas and Wireless Propagation Letters*, vol. 12, pp. 376–379, 2013.

[3] J. Lin, X. Ma, S. Yan, and C. Hao, "Time-frequency multi-invariance esprit for doa estimation," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 770–773, 2016.

[4] Z. Liu, Z. Huang, and Y. Zhou, "Sparsity-inducing direction finding for narrowband and wideband signals based on array covariance vectors," *IEEE Transactions on Wireless Communications*, vol. 12, no. 8, pp. 1–12, 2013.

[5] Z. Liu, C. Zhang, and P. S. Yu, "Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections," *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 12, pp. 7315–7327, 2018.

[6] B. Chen, H. Xiang, M. Yang, and C. Li, "Altitude measurement based on characteristics reversal by deep neural network for vhf radar," *IET Radar, Sonar and Navigation*, vol. 13, no. 1, pp. 98–103, 2019.

[7] H. Huang, J. Yang, Y. Song, H. Huang, and G. Gui, "Deep learning for super-resolution channel estimation and doa estimation based massive mimo system," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 1–1, 2018.

[8] L. Wu and Z. Huang, "Coherent svr learning for wideband direction-of-arrival estimation," *IEEE Signal Processing Letters*, vol. 26, no. 4, pp. 642–646, 2019.

[9] A. Randazzo, M. A. Abou-Khousa, M. Pastorino, and R. Zoughi, "Direction of arrival estimation based on support vector regression: Experimental validation and comparison with music," *IEEE Antennas and Wireless Propagation Letters*, vol. 6, pp. 379–382, 2007.

[10] M. Pastorino and A. Randazzo, "A smart antenna system for direction of arrival estimation based on a support vector regression," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 7, pp. 2161–2168, 2005.

[11] T. Lo, H. Leung, and J. Litva, "Radial basis function neural network for direction-of-arrivals estimation," *IEEE Signal Processing Letters*, vol. 1, no. 2, pp. 45–47, 1994.

[12] H. He, C. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmwave massive mimo systems," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 852–855, 2018.

[13] X. Xiao, S. Zhao, X. Zhong, D. L. Jones, E. S. Chng, and H. Li, "A learning-based approach to direction of arrival estimation in noisy and reverberant environments," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 2814–2818.

[14] S. Chakrabarty and E. A. P. Habets, "Broadband DOA estimation using convolutional neural networks trained with noise signals," 2017. [Online]. Available: http://arxiv.org/abs/1705.00919

[15] R. Takeda and K. Komatani, "Discriminative multiple sound source localization based on deep neural networks using independent location model," in *2016 IEEE Spoken Language Technology Workshop (SLT)*, 2016, pp. 603–609.

[16] T. M. Beck A, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 1, no. 2, pp. 183–202, 2009.

[17] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," vol. 15, 01 2010.

[18] F. Chollet, "Keras," 2015. [Online]. Available: http://keras.io/

[19] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 2014.