A Toeplitz Prior-Based Deep Learning Framework for DOA Estimation With Unknown Mutual Coupling

Zhuoqian Jiang[†‡], Jingmin Xin[†‡], Weiliang Zuo[†‡♮], Nanning Zheng[†‡], and Akira Sano[§]

[†] Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China

[‡] National Engineering Laboratory for Visual Information Processing and Applications, Xi'an Jiaotong University, Xi'an 710049, China

[♮] Shunan Academy of Artificial Intelligence, Ningbo, Zhejiang 315000, China

[§] Department of System Design Engineering, Keio University, Yokohama 223-8522, Japan

***Abstract*—In this paper, we explore the problem of direction-of-arrival (DOA) estimation with unknown mutual coupling using a deep learning (DL) framework which is based on Toeplitz prior. First, for source number estimation, we model it as a multi-label classification task and build a source number detection network (SNDN) to learn relevant information in the real sample covariance matrix. Next, taking full advantage of the Toeplitz structure, an ideal covariance reconstruction network (ICRN) is proposed to recover the ideal covariance matrix free from mutual coupling and noise interference. Furthermore, we design a database to store the parameters of ICRN after training on different numbers of sources, and its role is to load the corresponding parameters for ICRN according to the detection results of SNDN. Finally, the DOAs can be easily estimated from the restored covariance matrix by the MUSIC. The simulation results show our proposed approach not only outperforms the existing classical methods, but in some cases its DOA estimation accuracy can even exceed the Cramér-Rao Lower Bound.**

***Index Terms*—direction-of-arrival (DOA) estimation, mutual coupling, deep learning (DL), Toeplitz structure**

I. Introduction

Direction-of-arrival (DOA) estimation is a significant problem in various areas of array signal processing, such as radar, sonar, mobile and wireless communication (e.g., [1]–[4]). In the past decades, many high-resolution algorithms like MUSIC [2] and ESPRIT [3] have been proposed to achieve DOA estimation. However, these methods rely heavily on an accurate characterization of the array for desirable performance. In practical application, unknown mutual coupling inevitably exists between array elements, which will seriously deteriorate the performance of the above algorithms [4].

To reduce or even eliminate the mutual coupling effect, a large number of classical algorithms have been suggested, which can be roughly divided into three categories. The first one makes full use of the symmetric Toeplitz structure of the mutual coupling matrix for the uniform linear arrays (ULAs) (e.g., [5]–[7]). By only using the output signal of the middle subarray of ULAs, accurate DOA positioning can be achieved without compensating for mutual coupling. Nevertheless, there is no doubt that using the information of the middle subarray will reduce the array aperture, which will lead to limited applicability of such algorithms. In addition, there are reduction-based algorithms that utilize all the information of the array (e.g., [8], [9]), but in some cases the appearance of false peaks greatly reduces the positioning accuracy of the algorithm. The above two types of algorithms are mainly based on the nature of the subspace to achieve positioning, while finally there is another type of algorithm based on sparse reconstruction, which has achieved great performance (e.g., [10], [11]). However, this method usually requires the application of second-order statistics or higher-order statistics, which has relatively high computational complexity. In general, the above traditional algorithms have their own irreparable defects. In comparison, deep learning (DL) has shown outstanding positioning advantages in many harsh environments (e.g., [12]–[17]). To the best of our knowledge, the application of DL for multi-DOA estimation under unknown mutual coupling has not been well studied.

Therefore, in this paper, we propose a Toeplitz prior-based DL framework for multi-DOA estimation in unknown mutual coupling. First, a **s**ource **n**umber **d**etection **n**etwork (SNDN) is constructed, whose input and label are the sample covariance matrix and source number respectively. Then, based on the Toeplitz matrix structure, an **i**deal **c**ovariance **r**econstruction **n**etwork (ICRN) is designed, which can be used to restore an ideal covariance matrix without mutual coupling effect and noise interference. It is worth noting that in order to improve the performance, the parameters of ICRN are determined from the pre-training database according to the estimation results of SNDN. Finally, we employ MUSIC to estimate DOA from the recovered covariance matrix.

II. Signal Model

Consider K narrowband far-field sources impinging on a uniform linear array (ULA). The ULA is assumed to have M sensors with spacing d. The K signals, $\{s_k(n)\}$, arrive

This work was supported in part by the National key research and development program of China 2021ZD0113603 and the Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2022JQ-640).

at the array from different directions, $\{\theta_k\}$, with respect to the normal line of the array. Taking the mutual coupling into consideration, the received data at discrete time n can be approximated as

$$\boldsymbol{x}(n) = \boldsymbol{CAs}(n) + \boldsymbol{w}(n) \tag{1}$$

where $\boldsymbol{x}(n)$, $\boldsymbol{s}(n)$ and $\boldsymbol{w}(n)$ are the vectors of the array's output, the incident signals and the additive noises, which are given by $\boldsymbol{x}(n) = [x_1(n), x_2(n), \cdots, x_M(n)]^T$, $\boldsymbol{s}(n) = [s_1(n), s_2(n), \cdots, s_K(n)]^T$, and $\boldsymbol{w}(n) = [w_1(n), w_2(n), \cdots, w_M(n)]^T$ respectively. In additional, the matrix $\boldsymbol{C}$ stands for the mutual coupling effect, which has a banded symmetric Toeplitz structure, i.e., $\boldsymbol{C} = toeplitz(\boldsymbol{c})$, where $\boldsymbol{c} = [1, c_1, c_2, \cdots, c_P, 0, \cdots, 0]$ with $0 < |c_1|, |c_2|, \cdots, |c_P| < 1$. The array response matrix $\boldsymbol{A}$ is defined by $\boldsymbol{A} = [\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \cdots, \boldsymbol{a}(\theta_K)]$, and $\boldsymbol{a}(\theta_k)$ is the array steering vectors which can be expressed as

$$\boldsymbol{a}(\theta_k) = [1, e^{-j2\pi d sin\theta_k/\lambda}, \cdots, e^{-j2\pi(M-1)d sin\theta_k/\lambda}]^T \tag{2}$$

where $(\cdot)^T$ denotes the transposition, and the distance d between adjacent sensors is the same with half the carrier wavelength λ.

From (1), the array covariance matrix is given by

$$\boldsymbol{R_C} = E\{\boldsymbol{x}(n)\boldsymbol{x}^H(n)\} = \boldsymbol{C}\underbrace{\boldsymbol{AR}_s\boldsymbol{A}^H}_{\boldsymbol{R}}\boldsymbol{C}^H + \sigma^2\boldsymbol{I}_M \tag{3}$$

where $E\{\cdot\}$ represents the statistical expectation, $(\cdot)^H$ denotes the Hermitian transposition, $\boldsymbol{R}_s \triangleq E\{\boldsymbol{s}(n)\boldsymbol{s}^H(n)\}$ represents the covariance matrix of the incident source, σ^2 denotes the noise power and $\boldsymbol{I}_M$ stands for the $M \times M$ identify matrix. Note that $\boldsymbol{R}$ denotes the ideal array covariance matrix, which is a Hermitian matrix with Toeplitz structure.

In this letter, we concentrate on the multi-DOA $\{\theta_k\}_{k=1}^K$ estimation from the finite array data $\{\boldsymbol{x}(n)\}_{n=1}^N$ with noise and mutual coupling. For this purpose, we design to reconstruct $\boldsymbol{R}$ from the sample covariance matrix $\hat{\boldsymbol{R}}_{\boldsymbol{C}}$ by the DL algorithms, which is given by $\hat{\boldsymbol{R}}_{\boldsymbol{C}} = 1/N\sum_{n=1}^N \boldsymbol{x}(n)\boldsymbol{x}^H(n)$. However, it is relatively difficult to directly reconstruct $\boldsymbol{R}$ with $M \times M$ variables for the DL methods, because a complex network and a large amount of training data are required to achieve satisfactory estimation performance. Luckily, taking advantage of the Toeplitz structure of $\boldsymbol{R}$ mentioned above, we observe that $\boldsymbol{R}$ can be simply represented by its first row $\boldsymbol{r}$, whose size is only $M \times 1$. Therefore, we propose to learn the mapping relationship between $\hat{\boldsymbol{R}}_{\boldsymbol{C}}$ and $\boldsymbol{r}$ by the DL algorithms. After obtaining $\boldsymbol{r}$, accurate DOA estimation can be easily realized by MUSIC.

III. Proposed Method

As shown in Fig. 1, the proposed architecture takes the N snapshots of array data as input, which is then transformed into a real-value sample covariance matrix $\bar{\boldsymbol{R}}_{\boldsymbol{C}}$. The ICRN is used to recover the first row real-value data $\hat{\boldsymbol{r}}$ of the ideal covariance matrix from $\bar{\boldsymbol{R}}_{\boldsymbol{C}}$ containing mutual coupling and noise information. It should be noted that the parameters

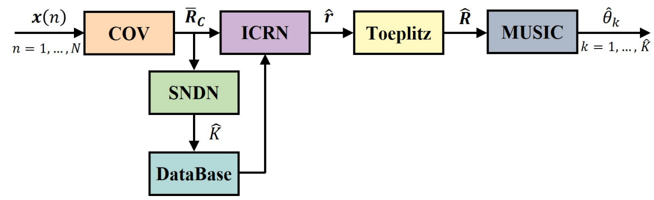


Fig. 1. Description of the proposed architecture for multi-DOA estimation under the mutual coupling.

of the ICRN are loaded from the database, which stores the corresponding network weight information of the ICRN trained under different numbers of sources. In additional, the SNDN mainly undertakes the task of estimating the number of sources $\hat{K}$ in the sample data, which is used to assist the parameter selection of the ICRN. Next, according to the nature of the Toeplitz structure, the $\hat{\boldsymbol{r}}$ obtained from the ICRN is converted into a complete ideal covariance matrix $\hat{\boldsymbol{R}}$. At last, multi-DOA can be estimated from $\hat{\boldsymbol{R}}$ via MUSIC.

A. Feature Selection

Because the sample covariance matrix $\hat{\boldsymbol{R}}_{\boldsymbol{C}}$ contains all the second-order statistical information of the received signal, we consider using it without changing its matrix structure. Naturally, we retain the real and imaginary parts of $\hat{\boldsymbol{R}}_{\boldsymbol{C}}$ to form a $M \times M \times 2$ real-valued matrix $\bar{\boldsymbol{R}}_{\boldsymbol{C}}$ as the input of ICRN and SNDN, whose first and second "channels" are given by $\bar{\boldsymbol{R}}_{\boldsymbol{C}:,:,1} = \mathrm{Re}\{\hat{\boldsymbol{R}}_{\boldsymbol{C}}\}$ and $\bar{\boldsymbol{R}}_{\boldsymbol{C}:,:,2} = \mathrm{Im}\{\hat{\boldsymbol{R}}_{\boldsymbol{C}}\}$, respectively.

B. Design for SNDN

Obviously, it is crucial to obtain the number of sources accurately before estimating DOA for most methods. Due to this reason, we design SNDN to realize target number detection. The SNDN receives $\bar{\boldsymbol{R}}_{\boldsymbol{C}}$ as input, and its label is a one-hot encoding of target number. Specifically, if the target number of a sample is K, the corresponding label of SNDN is the $(K+1)$th column of the $M \times M$ identity matrix. Of course, the label design of SNDN includes the case of no signal, that is, when the label is the first column of the identity matrix.

Fig. 2 shows the structure of SNDN, which contains a total of 8 layers after the input. Easily, SNDN can be represented by a non-linear mapping function $\mathcal{F}_{SNDN} : \mathbb{R}^{M\times M\times 2} \rightarrow \mathbb{R}^M$. In particular, we have

$$\mathcal{F}_{SNDN}(\bar{\boldsymbol{R}}_{\boldsymbol{C}}) = f_8(f_7(\cdots f_1(\bar{\boldsymbol{R}}_{\boldsymbol{C}}))) = \hat{\boldsymbol{u}} \tag{4}$$

where $\{f_i(\cdot)\}_{i=1,2,3,4}$ all represent the same structure: a 2D convolutional layer followed by a batch normalization (BN) layer and a rectified linear unit (ReLU) layer. For each 2D convolutional layer, 128 filters of size 2×2 and the same padding strategy are adopted to achieve multi-dimensional feature acquisition while avoiding the loss of edge information. Then, $\{f_i(\cdot)\}_{i=5,6,7}$ respectively denote a dense layer with 512, 256 and 128 neurons. Based on the settings for SNDN labels, we can treat source number detection as a multi-class classification problem with the number of classes M. Correspondingly, the last layer $f_8(\cdot)$ is designed as a dense layer

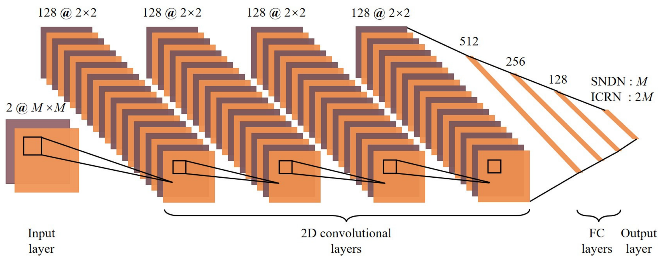


Fig. 2. The structures of SNDN and ICRN. Each 2D convolutional layer is followed by the batch normalization and ReLU layers.

with M neurons followed by a softmax activation function. The output $\hat{\boldsymbol{u}}$ of SNDN is denoted as $\hat{\boldsymbol{u}} = [\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_M]^T$, where any element $\hat{u}_j$ represents the possibility of the target quantity being $j-1$.

The loss function of SNDN is the categorical cross entropy, which is defined as follows

$$\mathcal{L}_{SNDN}\left(\boldsymbol{u}, \hat{\boldsymbol{u}}\right) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{M} u_j(t)\log[\hat{u}_j(t)] \tag{5}$$

where T is the total number of samples, $\hat{u}_j(t)$ represents the j-th element of SNDN prediction result for the t-th sample, while the corresponding label is defined as $u_j(t)$.

C. Design for Database

In fact, training neural networks separately for different numbers of sources is beneficial to improve the estimation accuracy and enhance the generalization performance of the algorithm [17]. Therefore, we build a database to store the network parameters pre-trained by ICRN under different target numbers, and more importantly, it also stores the mapping relationship between the number of sources and the corresponding trained network parameters. Specifically, the prediction results of the SDNN about the number of sources are input into the database, which will directly determine the selection of ICRN parameters in the database.

D. Design for ICRN

In order to achieve decoupling while denoising, we design the ICRN to learn the relevant information $\bar{\boldsymbol{r}}$ of the ideal array covariance matrix from $\bar{\boldsymbol{R}}_{\boldsymbol{C}}$. The label $\bar{\boldsymbol{r}}$ of the ICRN is represented as $\bar{\boldsymbol{r}} = [\text{Re}\{r_1\}, \text{Re}\{r_2\}, \cdots, \text{Re}\{r_M\}, \text{Im}\{r_1\}, \text{Im}\{r_2\}, \cdots, \text{Im}\{r_M\}]^T$, whose size is $2M\times 1$, where $\boldsymbol{r} = [r_1, r_2, \cdots, r_M]^T$ is the first row of corresponding ideal array covariance matrix $\boldsymbol{R}$. For convenience, $\boldsymbol{R}$ is calculated by the array manifold matrix $\boldsymbol{A}$, that is, $\boldsymbol{R} = \boldsymbol{A}\boldsymbol{A}^H$. Therefore, we can use a non-linear function $\mathcal{G}_{ICRN} : \mathbb{R}^{M\times M\times 2} \rightarrow \mathbb{R}^{2M}$ to generalize ICRN. Specifically, it is expressed by

$$\mathcal{G}_{ICRN}(\bar{\boldsymbol{R}}_{\boldsymbol{C}}) = g_8(g_7(\cdots g_1(\bar{\boldsymbol{R}}_{\boldsymbol{C}}))) = \hat{\boldsymbol{r}}. \tag{6}$$

As shown in Fig. 2, the architecture of ICRN is composed of 8 layers, which is similar to that of SNDN. We can see that the difference between ICRN and SNDN is the label and their respective last layer. The last layer $g_8(\cdot)$ for ICRN is a dense layer with $2M$ neurons for regression task. Furthermore, mean square error (MSE) is used as the loss function of ICRN, which is given by

$$\mathcal{L}_{ICRN}\left(\bar{\boldsymbol{r}}, \hat{\boldsymbol{r}}\right) = \frac{1}{T}\sum_{t=1}^{T}\left[\bar{\boldsymbol{r}}(t) - \hat{\boldsymbol{r}}(t)\right]^2 \tag{7}$$

where $\bar{\boldsymbol{r}}(t)$ and $\hat{\boldsymbol{r}}(t)$ represent the label and actual output of the ICRN for the t-th sample, respectively. T is still the total sample size.

IV. Numerical Simulations

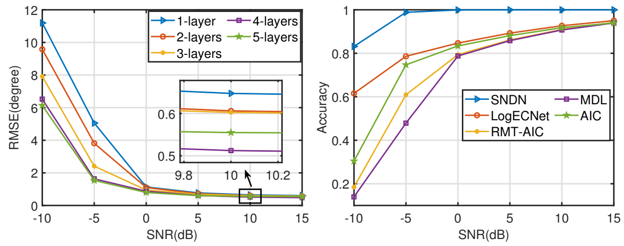
A. Simulation Settings

Theoretically, for different environment and array configuration, DOA estimation can be easily achieved by fine-tuning the relevant parameters in our network and then updating the database after pre-training. For simplicity, the proposed architecture is evaluated on a ULA of $M = 7$ elements in Gaussian white noise, and the number of mutual coupling coefficients is $P = 2$ with $c_1 = 0.5663 + 0.4114i$ and $c_2 = 0.2898 - 0.0776i$. The datasets of SNDN is set as follows: The number of snapshots N is fixed at 200, the number of sources K is first randomly generated 0.1 million times over $[0, M-1]$, then the DOA of each source is randomly generated between $[-60°, 60°]$. For every time implementation, we generate 10 samples with the SNR randomly distributed in $[-10$ dB,15 dB$]$. Therefore, there are a total of 1 million samples in the datasets of SNDN, of which the training data and validation data account for 80% and 20% respectively.

In order to obtain the ICRN parameters for locating different source numbers, we set up datasets for the $M-1$ situations where the number of sources exists, respectively. For the k-th situation, we assume that there are k sources randomly distributed in the spatial scope of $[-60°, 60°]$, where $1 \leq k \leq M-1$. For each situation, we perform 60000 times random distributions of DOA. Then, 30 samples are generated for each time execution, where the SNR of samples is randomly distributed from -10 dB to 15 dB, and the snapshot N is 200. Thus, we generate 1.8 million samples for the datasets of ICRN in each situation.[1] Furthermore, 80% and 20% of the datasets are chosen for training and validation data, respectively. It should be noted that for each situation, we train ICRN separately and store its corresponding parameters in the database.

After generating the datasets, considering the tradeoff between the expression ability and generalization performance of the network, we set the training parameters for SNDN and ICRN: The adaptive moment estimation (ADAM) optimizer is employed to determine the optimal parameters with an initial learning rate of 0.001 [18], and the learning rate will be halved after every 10 epochs. The maximum number of epoch for training is fixed at 30, and the order of the samples is shuffled during every epoch. In addition, a mini-batch size of 64 samples is used for network training. In the following, we will present several simulations to verify the effectiveness of

[1]Due to space constraints, the impact of datasets size on the algorithm is not discussed in this paper, and we will investigate it in our future work.



(a) (b)

Fig. 3. (a) Convolutional layers' impact on DOA estimation, where $K = 4$. (b) The accuracy of source number detection versus the SNR.

our proposed architecture, and the simulations experimental results are all based on 1000 independent random trials. Furthermore, all test data are excluded from the proposed algorithm's training and validation datasets.

B. Impact of Convolutional Layers

Before evaluating the parameters of interest, we study the impacts of convolutional layers on the ICRN network performance. For comparison, we train different ICRN networks with 1 to 5 convolutional layers. Fig. 3(a) shows the root mean square errors (RMSEs) of the estimated DOAs achieved by different ICRN networks in terms of the SNR. We can see that at small SNRs, as the number of convolutional layers increases from 1 to 5, the estimation performance improves due to the enhanced expression power of the networks [19]. However, when the SNR is greater than 5 dB, it is worth noting that the ICRN network consisting of four convolutional layers outperform those of five layers. The reason may be that too many network parameters lead to the phenomenon of overfitting [20]. Therefore, the proposed ICRN with four convolutional layers is based on a comprehensive trade-off between the network of expressive ability and overfitting risk.

C. Source Number Detection

For the source number detection problem, we adopt the LogECNet [21], the RMT-AIC [22], the MDL and the AIC estimators [23] as benchmarks. Fig. 3(b) depicts the accuracy performance of all estimators versus the SNR, where the number of sources in the test datasets is distributed in $[1, 4]$. It can be seen from the figure that the proposed SNDN significantly outperforms all other estimators, especially in the low SNR region. Besides, even when the SNR reaches 15 dB, the accuracy of the LogECNet and classical estimators can only reach 0.95. The reason may be that these estimators are all based on the eigenvalues of the covariance matrix with mutual coupling effects, which do not eliminate the influence of mutual coupling effects on the estimators.

D. DOA Estimation

In order to verify the superiority of the proposed ICRN algorithm, the performance of the proposed method is compared with the R-RARE method in [9], the RARE method in [8], the MID-ARRAY method in [5], the MUSIC method in [2] and the CRB in [4]. For the sake of fairness, it is assumed that the number of sources is accurately known in advance, and all algorithms are based on the MUSIC method for DOA estimation, where the grid spacing of spectral peak search is fixed at $\Delta\theta = 0.1°$. In addition, it should be mentioned that for the R-RARE and RARE algorithms, they can only be used to estimate DOAs when $K_{RARE} \leq M - P - 1$ [9], while for the MID-ARRAY method, the DOAs can only be estimated when $K_{MID} \leq M - 2P - 1$ [5].

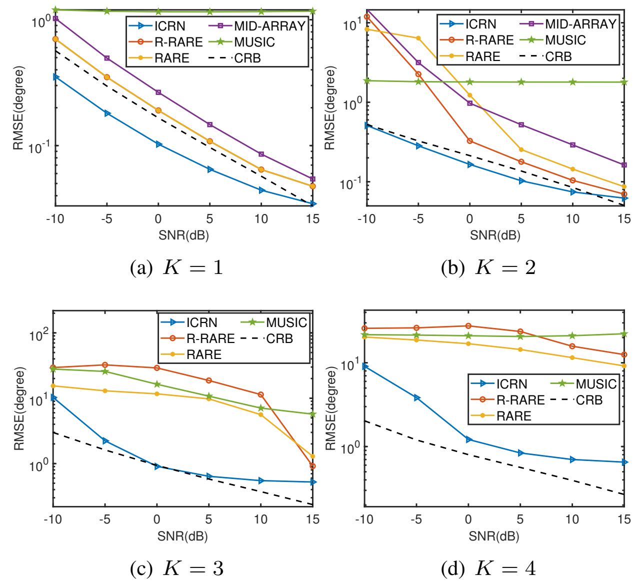
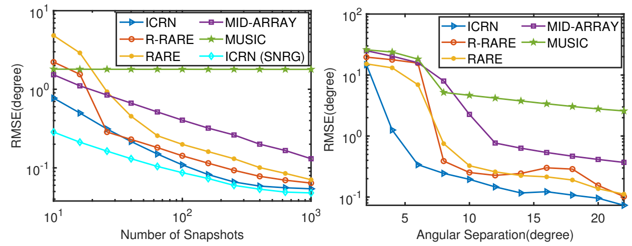


Fig. 4. Performance of DOA estimation with different source numbers

The performance of estimated DOA in terms of SNR for different number of sources is shown in Fig. 4. Under the experimental conditions designed in this paper, the maximum number of sources that can be estimated by the R-RARE/RARE and MID-ARRAY algorithms is 4 and 2, respectively. Therefore, we do not show the performance of the MID-ARRAY algorithm in Fig. 4(c) and (d). We can see that compared to other classical algorithms, our proposed ICRN algorithm performs best in the DOAs estimation with different source numbers involved in Fig. 4. It is worth noting that in Fig. 4(a) and (b), the ICRN even beats the CRB when the SNR is small. The possible reason for this is that the CRB represents the best accuracy of the unbiased estimator, while the ICRN is a biased estimator, which is not constrained by the Cramér-Rao Lower Bound. Furthermore, we can also observe that in Fig. 4(c) and (d), as the number of sources increases, the performance of ICRN compared to CRB gradually deteriorates. One possible reason is that the accuracy of ICRN as a biased estimator is limited by the size of the training dataset [15]. Combining the results of previous simulation, it can be known easily that the proposed ICRN algorithm has effectively solved the problem of array aperture reduction when some existing algorithms reduce or eliminate the mutual coupling effect.

Fig. 5(a) displays the RMSEs of the estimated DOAs with



(a) (b)

Fig. 5. RMSEs of the estimated DOAs versus (a) the number of snapshots and (b) the angular separation, where $K = 2$.

respect to the number of snapshots. It should be noted that the simulation conditions in Fig. 5(a) are: the SNR of the test sample is fixed at 10 dB, and there are two signal sources. In this figure, both the ICRN and the ICRN (SNRG) are implemented according to the structure of Fig. 2, the only difference is their training data. In particular, the ICRN is trained under the dataset as described in section IV-A with a fixed snapshots N of 200 and randomly generated SNR. In contrast, the SNR of the training dataset for the ICRN (SNRG) is fixed at 10 dB, and its **s**napshot **n**umbers are **r**andomly **g**enerated (SNRG) within $[10, 1000]$, and other simulation settings are consistent with those of the ICRN. We can find that even if the ICRN is not trained for different snapshot numbers, it still has the competitive performance compared with other model-based algorithms. This shows that the ICRN has good generalization performance under different snapshot number scenarios. Of course, it is natural that the ICRN (SNRG) trained with different number of snapshots performs better than the ICRN.

Finally, we explore the effect of the angular separation on the DOA estimation performance in Fig. 5(b), where the SNR is fixed at 10dB, and the number of snapshots is set at 200. In this experiment, two sources located in $-20.65°$ and $-20.65°+\Delta\theta$ are considered, where $\Delta\theta$ is varied from $2°$ to $22°$ with $\Delta\theta = 2°$. We can see from the figure that no matter the angular separation is large or small, the DOA estimation accuracy of our proposed ICRN algorithm is always higher than other algorithms. In addition, it is worth noting that the proposed ICRN method is more advantageous over some small angular intervals such as $4°$ to $6°$.

V. Conclusion

In this paper, a deep learning based framework is proposed to estimate DOA with unknown mutual coupling, which mainly consists of the SNDN network responsible for source number detection and the ICRN network responsible for ideal covariance matrix recovery. The proposed SNDN can achieve end-to-end output by learning the source number information in the real sample covariance matrix. The proposed ICRN network is designed based on the Toeplitz matrix structure, and its parameters are loaded from a database containing network weight parameters pre-trained for different source numbers. Numerical results show that, compared with classical algorithms, the proposed approach can significantly improve the accuracy of both source number detection and DOA estimation.

References

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

[2] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[3] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.

[4] B. Friedlander and A. J. Weiss, "Direction finding in the presence of mutual coupling," *IEEE Trans. Antennas Propag.*, vol. 39, no. 3, pp. 273–284, 1991.

[5] Z. Ye and C. Liu, "On the resiliency of MUSIC direction finding against antenna sensor coupling," *IEEE Trans. Antennas Propag.*, vol. 56, no. 2, pp. 371–380, 2008.

[6] Z. Ye, J. Dai, X. Xu and X. Wu, "DOA estimation for uniform linear array with mutual coupling," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 45, no. 1, pp. 280–288, 2009.

[7] J. Dai and Z. Ye, "Spatial smoothing for DOA estimation of coherent signals in the presence of unknown mutual coupling," *Signal Process.*, vol. 5, no. 4, pp. 418–425, 2011.

[8] M. Lin and L. Yang, "Blind calibration and DOA estimation with uniform circular arrays in the presence of mutual coupling," *IEEE Antennas Wireless Propag. Lett.*, vol. 5, pp. 315–318, 2006.

[9] J. Dai, X. Bao, N. Hu, C. Chang and W. Xu, "A recursive RARE algorithm for DOA estimation with unknown mutual coupling," *IEEE Antennas Wireless Propag. Lett.*, vol. 13, pp. 1593–1596, 2014.

[10] Q. Wang, T. Dou, H. Chen, W. Yan and W. Liu, "Effective block sparse representation algorithm for DOA estimation with unknown mutual coupling," *IEEE Commun. Lett.*, vol. 21, no. 12, pp. 2622–2625, 2017.

[11] X. Zhang, T. Jiang, Y. Li and Y. Zakharov, "A novel block sparse reconstruction method for DOA estimation with unknown mutual coupling," *IEEE Commun. Lett.*, vol. 23, no. 10, pp. 1845–1848, 2019.

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

[13] C. Zhou, Y. Gu, X. Fan, Z. Shi, G. Mao and Y. D. Zhang, "Direction-of-Arrival estimation for coprime array via virtual array interpolation," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5956–5971, 2018.

[14] A. Barthelme and W. Utschick, "Doa estimation using neural network-based covariance matrix reconstruction," *IEEE Signal Process. Lett.*, vol. 28, pp. 783–787, 2021.

[15] G. K. Papageorgiou, M. Sellathurai and Y. C. Eldar, "Deep networks for direction-of-arrival estimation in low snr," *IEEE Trans. Signal Process.*, vol. 69, pp. 3714–3729, 2021.

[16] Z. Jiang, J. Xin, W. Zuo, N. Zheng and A. Sano, "Deep residual learning based localization of near-field sources in unknown spatially colored noise fields," in *Proc. IEEE Eur. Signal Process. Conf. (EUSIPCO)*, pp. 1741–1745, 2022.

[17] X. Wu, X. Yang, X. Jia and F. Tian, "A gridless DOA estimation method based on convolutional neural network with toeplitz prior," *IEEE Signal Process. Lett.*, vol. 29, pp. 1247–1251, 2022.

[18] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations.*, vol. 12, pp. 1–9, 2014.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[20] S. Ioffe, C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv: 1502.03167, 2015.

[21] D. T. Hoang and K. Lee, "Deep learning-aided coherent direction-of-arrival estimation with the FTMR algorithm," *IEEE Trans. Signal Process.*, vol. 70, pp. 1118–1130, 2022.

[22] R. R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2625–2638, 2008.

[23] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, 1985.