Robust DOA Estimation Using Deep Complex-Valued Convolutional Networks with Sparse Prior

2023 6th International Conference on Information Communication and Signal Processing (ICICSP) | 979-8-3503-3999-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICICSP59554.2023.10390873

Shulin Hu
National Key Laboratory of Radar Signal Processing, Xidian University
Xi'an, China, 710071
shulinhu@stu.xidian.edu.cn

Cao Zeng*
National Key Laboratory of Radar Signal Processing, Xidian University
Xi'an, China, 710071
czeng@mail.xidian.edu.cn

Minti Liu
National Key Laboratory of Radar Signal Processing, Xidian University
Xi'an, China, 710071
mintiliu@163.com

Haihong Tao
National Key Laboratory of Radar Signal Processing, Xidian University
Xi'an, China, 710071
hhtao@xidian.edu.cn

Shihua Zhao
National Key Laboratory of Radar Signal Processing, Xidian University
Xi'an, China, 710071
zshihua99@163.com

Yu Liu
National Key Laboratory of Radar Signal Processing, Xidian University
Xi'an, China, 710071
349319623@qq.com

***Abstract*—Mathematically speaking, the direction of arrival (DOA) estimation methods with data-driven deep learning generally exhibit enhanced robustness under unknown clutter and noise scenarios. However, it is worthily noticed that the existing methods mostly suffer from signal model mismatch, due to relying on real-valued operations ignoring actual received signals are complex values. To this regard, a novel DOA estimation method with sparse prior based on deep complex-valued convolutional neural network (CV-CNN) is proposed. Specifically, the complex signal features of covariance matrix firstly are extracted by directly utilizing many complex-valued (CV) layers mainly referring to CV-Conv2d, CV-CELU, CV-FC, CV-BatchNorm2d and CV-Dropout operations. Then, the modified loss function based on sparsity prior constraint is formulated to improve estimation accuracy. Notably, the training process is innovatively divided into two stages, i.e., pre-training and fine-tuning, to further optimize the model parameters. Finally, simulation results indicate that the proposed CV-CNN method possesses superior performance than that of the existing real-valued deep networks on estimation accuracy and robustness under the low SNR or limited snapshots scenarios.**

Keywords—direction of arrival estimation, deep complex-valued networks, sparse representation

I. INTRODUCTION

Direction-of-arrival (DOA) estimation is the process of estimating the incoming directions of multiple signals received by a sensor array. Over the past few decades, model-driven DOA estimation methods have been long-lasting developments, such as beamforming [1]-[3], subspace -based methods [4]-[7], and maximum likelihood methods [8]-[9]. In fact the mentioned-above methods commonly is not suitable for real scenario, such as array imperfections, interference, array coupling and so on. Notably, the performance of these methods maybe significantly deteriorate or even fail under low signal-to-noise ratio (SNR) and limited snapshots situations.

To address the shortcomings associated with model-based methods, researchers have developed data-driven DOA methods based on deep learning. For instance, a fully connected deep neural network (DNN) DOA estimation framework was proposed in [10], which leverages the real and imaginary parts of the upper-triangular elements of the covariance matrix to implement precise estimate. This framework integrates autoencoders and parallel multilayer classifiers to recover the spatial spectrum on a pre-defined grid, but the spectrum resolution is relatively low under limited snapshots and low SNR, as reported in [11]. To this end, through changing the deep network structure, in [11], the 1-dimensional (1D) deep convolutional networks is suggested to improve resolution , and simulation results showed that it is capacity of providing comparable estimation accuracy and low computation cost than that of the sparse-induced methods. Nevertheless, the 1D convolutional networks exhibit a restricted capacity to extract global details from the covariance matrix. Deeply, in [12], the 2D convolutional networks is adopted to directly extract features from three channels, including the real, imaginary, and phase parts of the covariance matrix. This method is formulated as a binary classification problem on a pre-defined spatial grids and demonstrates exceptional performance in low SNR conditions. However, this method does not make full use of the relevant information between channels, and the data utilization rate is not high. Since the priori assumptions regarding the geometry of sensor array is no constraint, the DOA estimation based on deep neural network methods exhibit robust performance under non-ideal conditions [13].

Mathematically speaking, the sparse-induced methods [14]-[16] are formulated as an inverse problem of sparse representation in the spatial domain of array-sampled signals. Inspired by [17], the inverse mapping from the received signals to the sparse spatial spectrum can be learned by deep complex-valued (CV) neural networks. To the best of our knowledge, most of the existing deep learning methods for DOA estimation prefer to real-valued networks, and few studies are complex-valued networks. Therefore, joint the sparse-induced idea and complex-valued deep learning networks, we propose a DOA estimation method using a 2D deep complex-valued convolutional neural networks (CV-CNN) with sparse prior,

which directly learns the inverse mapping from the complex-valued covariance matrix of array received data to the spatial spectrum.

Notations: $\mathbf{X}$, $\mathbf{x}$, x represent a matrix, a vector and a scalar, respectively. $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, and $\mathrm{E}[\cdot]$ represent the transpose, conjugate transpose, and expectation operation, respectively.

II. Signal Model

Assume that K narrow-band independent source signals $\mathbf{s}_k(t)=\{s_k(t)\}_{k=1}^{K}$, in far-field scene, impinge onto an M-element array from directions $\boldsymbol{\theta}=\{\theta_k\}_{k=1}^{K}$. The array received data $\mathbf{x}(t)=[x_1(t),x_2(t),...,x_M(t)]^{\mathrm{T}}$ can be defined as:

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

where $\boldsymbol{a}(\theta_k)$ denotes the steering vector corresponding to θ_k and $\mathbf{n}(t)$ is additive zero-mean Gaussian noise vector with variance σ_n^2.

Given that the potential space is divided into a discrete sampled set $\tilde{\boldsymbol{\theta}}=\{\theta_l\}_{l=1}^{L}$, sampled by L grids with a sampling interval of $\Delta\theta$. Then the array received data also can be formulated in a overcomplete form:

$$\mathbf{x}(t)=\sum_{l}^{L}\boldsymbol{a}(\theta_l)s_l(t)+\mathbf{n}(t) \tag{2}$$

where $s_l(t)$ has a non-zero value only if $\theta_l\in\boldsymbol{\theta}$. Assuming the source signals and noise are uncorrelated, then, the covariance matrix of the received data can be defined as follows:

$$\begin{aligned}\mathbf{R}&=\mathrm{E}\left[\mathbf{x}(t)\mathbf{x}^{\mathrm{H}}(t)\right]\\&=\sum_{l}^{L}\zeta_l\boldsymbol{a}(\theta_l)\boldsymbol{a}^{\mathrm{H}}(\theta_l)+\sigma_n^2\mathbf{I}_M\\&=\left[\mathbf{A}_1,\mathbf{A}_2,\cdots,\mathbf{A}_L\right]\mathbf{P}+\sigma_n^2\mathbf{I}_M\end{aligned} \tag{3}$$

where $\zeta_l=E\left[s_l(t)s_l^{\mathrm{H}}(t)\right]$, $l=1,2,\cdots,L$ represents the signal power corresponding to direction θ_l, and $\mathbf{P}=diag(\zeta_1,\zeta_2,\cdots,\zeta_L)$ also has k non-zero elements when the source located at the true directions; $\mathbf{A}_l=\boldsymbol{a}(\theta_l)\boldsymbol{a}^{\mathrm{H}}(\theta_l)$, and $\mathbf{I}_M\in\mathbb{C}^{M\times M}$ is an identity matrix.

In practice, the covariance matrix $\mathbf{R}\in\mathbb{C}^{M\times M}$ is obtained by finite snapshots samples, i.e.,

$$\hat{\mathbf{R}}=\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}(t)\mathbf{x}^{\mathrm{H}}(t) \tag{4}$$

where T is snapshots number.

According to (3) and (4), DOA estimation can be achieved by recovering the position of non-zero elements in the $\mathbf{P}$ from $\hat{\mathbf{R}}$. Based on the powerful nonlinear mapping capability of deep learning, we propose to use deep complex-valued convolutional networks to learn the inverse mapping from $\hat{\mathbf{R}}$ to $\mathbf{P}$, i.e., to recover the spatial spectrum of the source signals.

III. Proposed Method

The proposed CV-CNN networks architecture is shown in Fig. 1. In particular, the complex-valued covariance matrix of received signals is firstly feed to networks, the complex-valued convolutional layers are utilized to performs feature extraction, and the map results of covariance matrix expresses multiple feature subspaces. Afterwards, the features extracted by complex-valued convolutional layers are remapped by a series of fully connected layers to the spectrum space, the recover spectrum meets the assumption of sparsity.

A. Complex-Valued Convolutional Layer (CV-Conv2d)

Consider that the complex-valued covariance matrix of signals, the complex-valued convolutional layers align well with this attribute, and enabling the extraction of the phase information of the covariance matrix. Therefore, we developed the CV-Conv2d by following the approach in [17], assuming that the convolutional kernel weight matrix $\mathbf{W}=\mathbf{A}+i\mathbf{B}$ is convolved with a complex-valued vector $\mathbf{h}=\mathbf{x}+i\mathbf{y}$, and the matrix representation of this operation is:

$$\begin{bmatrix}R(\mathbf{W}*\mathbf{h})\\I(\mathbf{W}*\mathbf{h})\end{bmatrix}=\begin{bmatrix}\mathbf{A}&-\mathbf{B}\\\mathbf{B}&\mathbf{A}\end{bmatrix}*\begin{bmatrix}\mathbf{x}\\\mathbf{y}\end{bmatrix} \tag{5}$$

where $R(\cdot)$ and $I(\cdot)$ denote respectively the real and imaginary parts of a complex number, and $*$ represents a real-valued convolution operation.

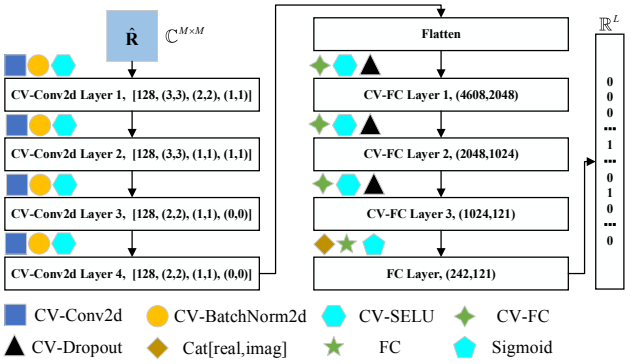

Fig. 1. The architecture of the proposed CV-CNN for DOA estimation.

As shown in Fig. 1, the CV-CNN networks consists of four layers of CV-Conv2d, and the parameters of each layer are indicated in the figure and correspond to [*filters*, *kernel size*, *stride*, *padding*]. Following CV-Conv2d operations, in order to improve the efficiency of training, we use 2D CV Batch Normalization (CV-BatchNorm2d) to normalize the output of the convolutional layer, so that the mean of the output is 0, the covariance is 1, and the correlation coefficient is 0. Instead of merely normalizing the real and imaginary components separately, CV-BatchNorm2d scales the zero-centered data by the square root of the variance for each of the two components, the details contents for implementation process can be found in [17]. Additionally, a complex-valued Scaled Exponential Linear Unit (CV-SELU) [18] is applied element-wise to the

real and imaginary parts of the variables from the previous layer, respectively. CV-SELU can be formulated as follows:

$$\mathrm{SELU}(x)=\lambda\begin{cases}x & \text{if } x>0\\ \alpha\left(e^{x}-1\right) & \text{otherwise}\end{cases} \tag{6}$$

$$\text{CV-SELU}\left(x\right)=\mathrm{SELU}\left(R\left(x\right)\right)+i\mathrm{SELU}\left(I\left(x\right)\right) \tag{7}$$

where $\lambda=1.0507$ and $\alpha=1.6733$ are two super parameters.

Therefore, the output of the *k*-th CV-Conv2d layer can be formulated as:

$$f_k\left(\mathbf{X}\right)=\text{CV-SELU}\left(\mathrm{BN}\left(\mathbf{W}_k*\mathbf{X}+\mathbf{b}_k\right)\right) \tag{8}$$

where $\mathbf{W}_k\in\mathbb{C}^{filter\times kernel\ size}$ and $\mathbf{b}_k\in\mathbb{C}^{fliter}$ denote the CV weight and bias vector of the *k*-th CV-Conv2d layer, respectively. $\mathrm{BN}(\cdot)$ represents the CV-BatchNorm2d.

The initial CV-Conv2d layer accepts the covariance matrix as input, and is followed by four layers of CV-Conv2d operations. The covariance matrix is represented as multiple feature subspaces, where the number of subspaces is dependent on the number of filters. We subsequently extract the source direction information within these subspaces by using complex-valued fully connected layers, and ultimately reconstruct the spatial spectrum.

B. Complex-Valued Fully Connected Layer CV-FC)

In this part, we flatten the features extracted from each filter on the CV-Conv2d layer, i.e., each subspace, to connect it with a series of complex-valued fully connected layers. As shown in Fig. 1, we used 3 layers of CV-FC, and each layer corresponds to a set of parameters [*input features*, *output features*]. Thus, the *k*-th CV-FC layer can be formulated as:

$$\mathbf{net}_k=\mathbf{W}_{k,k-1}\mathbf{h}_{k-1}+\mathbf{b}_k \tag{9}$$

$$\begin{bmatrix}R\left(\mathbf{W}_{k,k-1}\mathbf{h}_{k-1}\right)\\ I\left(\mathbf{W}_{k,k-1}\mathbf{h}_{k-1}\right)\end{bmatrix}=\begin{bmatrix}R\left(\mathbf{W}_{k,k-1}\right) & -I\left(\mathbf{W}_{k,k-1}\right)\\ I\left(\mathbf{W}_{k,k-1}\right) & R\left(\mathbf{W}_{k,k-1}\right)\end{bmatrix}\begin{bmatrix}R\left(\mathbf{h}_{k-1}\right)\\ I\left(\mathbf{h}_{k-1}\right)\end{bmatrix} \tag{10}$$

$$\mathbf{h}_k=\text{CV-SELU}\left(\mathbf{net}_k\right) \tag{11}$$

where $\mathbf{W}_{k,k-1}\in\mathbb{C}^{n\times m}$, $\mathbf{h}_{k-1}\in\mathbb{C}^{m}$, $\mathbf{b}_k\in\mathbb{C}^{n}$, and $\mathbf{net}_k\in\mathbb{C}^{n}$ represent the complex weight, input, bias and output between the *k*-1th and *k*-th CV-FC layer, respectively. *m* and *n* represent the *input features* and *output features* respectively.

To mitigate the risk of overfitting in the training stages, we incorporated CV Dropout operations (CV-Dropout) into each CV-FC layer, which randomly assigns a value of 0 to specific elements within the weight matrix $\mathbf{W}_{k,k-1}$ with a probability of *p*. The CV-Dropout conducts a joint dropout operation on both real and imaginary components, ensuring that these components may be set to 0 simultaneously.

C. Output Layer

To map the complex-valued features back to real space, we concatenate (Cat) the real and imaginary parts of the output from the final CV-FC layer, resulting in a real-valued vector with double the dimension. Then, we feed it into a real-valued Fully-Connected layer (FC) to map it to *L* pre-defined grids. i.e., the final output dimension of the whole networks is $\mathbb{R}^{L}$.

Additionally, the Sigmoid activation function is used to map the output vector to a probability space, so we determine whether a certain pre-defined grid is the source direction based on the vector. The output layer can be formulated as:

$$\hat{\mathbf{y}}=\mathrm{Sigmoid}\left(\tilde{\mathbf{W}}\cdot\tilde{\mathbf{h}}\right) \tag{12}$$

where $\tilde{\mathbf{h}}=\left\{Real\left(\mathbf{h}_3\right),Imag\left(\mathbf{h}_3\right)\right\}\in\mathbb{R}^{2c}$ is a fusion of real and imaginary components generated by the final CV-FC layer with a dimension 2*c*. $\tilde{\mathbf{W}}\in\mathbb{R}^{L\times 2c}$ and $\hat{\mathbf{y}}\in\mathbb{R}^{L}$ denote the weight matrix of the FC layer, the output probability vector, respectively.

D. Training Strategy

The CV-CNN model can be trained using synthetic data. Specifically, considering the potential space to be [-60°,60°], which is divided into *L* evenly spaced discrete grids with a resolution of $\Delta\theta$, i.e., $\tilde{\boldsymbol{\theta}}=\left\{-\frac{L}{2}\Delta\theta,\frac{L}{2}\Delta\theta\right\}$. In the training dataset, the directions of each sample are selected as a binomial combination of the set $\tilde{\boldsymbol{\theta}}$. For the generalization ability of the model under various SNR levels, samples with identical angle-pair will be repeatedly generated within a set of uniformly SNR levels $\left[SNR_{\min},SNR_{\max}\right]$ with a step of $\Delta\vartheta$. On this basis, *G* samples are generated by setting two sources $\left\{\left(\theta_1^{(l)},\theta_2^{(l)}\right)\middle|l=1,2,\cdots,G\right\}\subseteq\tilde{\boldsymbol{\theta}}$ with different angle-pairs. Then we obtain the expected covariance matrix $\mathbf{R}$ for each sample by (3), and the sampled covariance matrix $\hat{\mathbf{R}}$ obtained by (4) sampling *T* snapshots, respectively.

The spatial spectrum recovery problem is modeled as a multi-label classification task by the CV-CNN, where binary classification is performed on each grid. The desired output $\mathbf{y}$ of the CV-CNN is only 1 at the grid corresponding to the true directions, and otherwise 0, i.e.,

$$\mathbf{y}\{\theta_l\}=\begin{cases}1 & \text{if } \theta_l\in\boldsymbol{\theta}\\ 0 & \text{otherwise}\end{cases}\quad l=1,2,\cdots,L \tag{13}$$

Specifically, the proposed CV-CNN model undergoes pre-training on a training set consisting of expected covariance matrices, followed by fine-tuning on a sampled covariance matrices training set using a small learning rate. For the supervised learning, the loss function of the model is represented as the cross-entropy loss between the estimated spatial spectrum $\hat{\mathbf{y}}$ and the label $\mathbf{y}$. Additionally, as previously mentioned, we take into account the sparsity of the sources, which is implemented by incorporating the L_1-norm regularization of the FC layer weights to moderate model overfitting. The goal of model training is to minimize the loss function, which can be formalized as follow:

$$\varpi^{*}=\arg\min_{\varpi}\frac{1}{G}\left\{\sum_{i=1}^{G}\left(L\left(\hat{\mathbf{y}}_i;\mathbf{y}_i\right)+\gamma\sum_{j=1,k=1}^{L,2c}\tilde{\mathbf{W}}_{j,k}\right)\right\} \quad (14)$$

where ϖ, $\tilde{\mathbf{W}}_{j,k}$, $L\left(\hat{\mathbf{y}}_i;\mathbf{y}_i\right)$, and γ represent the learnable parameters of the model, the (j, k) element of the weight matrix $\tilde{\mathbf{W}}$, the cross-entropy between $\hat{\mathbf{y}}_i$ and $\mathbf{y}_i$, and the scaling factor of the sparsity constraint, respectively.

IV. Simulation Results

To illustrate effectiveness and advantage of the proposed 2D CV-CNN method, in this section, we benchmark its robustness and estimation accuracy to conduct comparisons with different methods under low SNR or limited snapshots. These method mainly involve traditional model-driven methods: MUSIC [4], R-MUSIC [5], ESPRIT [7], and method based on real-valued 2D convolutional networks (CNN) [12]. In addition, we also consider the estimated Cramer-Rao Low Bound for strictly uncorrelated sources (CRLB_{uncr}) in [19] as a reference. All traditional methods were implemented in MATLAB2022b and the simulation experiments were run on a laptop with 2.9GHz Intel Core i7 processor and 32Gb memory. The CV-CNN and CNN networks are deployed on a an NVIDIA GeForce RTX 3080 GPU based on the PyTorch.

A. Simulation Settings

Consider a 16-element uniform linear array (ULA) with half-wavelength inter-element spacing, the directions of the sources to be estimated are located within the spatial range [-60°,60°], and the pre-defined spatial sampling has a resolution of $\Delta\theta=1°$, i.e., $\tilde{\boldsymbol{\theta}}=\{-60°, -59°,\cdots, 60°\}$ and $L = 121$. Setting the SNR interval is [-20dB, 20dB], the step length of SNR $\Delta\vartheta=5$, and the number of snapshots T=400. At each SNR level, all angle-pairs of the elements in $\tilde{\boldsymbol{\theta}}$ constitute the set of source directions, thus the training dataset contains $\binom{2}{121}$=7260 pairs of angles in total, covering both large and small angular intervals.

In the training stage, the overall data set is shuffled and randomly partitioned into two subsets with a ratio of 8 to 2, i.e., the training set and the validation set. The training set consists of a total of 7260×7×0.8×2= 81312 samples, consisting of the expected covariance matrix in (3) and the sampled covariance matrix in (4). The parameters of CV-CNN are updated using Adam optimizer. We employ an innovative segmented training approach, consisting of two stages, the first stage is the pre-training stage, in which 200 epochs are used on the expected covariance matrix, using a learning rate of μ_1=10^{-4}, setting the mini-batch size to 32, and γ=5×10^{-5}. Following this is a fine-tuning stage, where 100 epochs are trained using the sampled covariance matrix with a small learning rate of μ_2=5×10^{-5}, the mini-batch size is set to 64 and γ=10^{-5}. During both stages, the learning rate will decrease by a factor of 0.7 if the validation set loss does not reduce over five consecutive epochs, and the probability of the dropout is set to p=0.3. Additionally, the crucial parameters for the Adam optimizer are set to β_1=0.85, β_2=0.95.

For a fair comparison, the grid resolution was set to the same value for both MUSIC and CNN methods as well. ESPRIT used the maximum subarray overlap as well as the row weighting technique with the maximum possible weight, and the total least-squares (TLS) criterion was employed. The performance measure metric for comparison across scenarios is defined as the Root Mean Square Error (RMSE), which can be formulated as:

$$\mathrm{RMSE}=\sqrt{\frac{1}{T_{\mathrm{mc}}K}\sum_{k=1}^{K}\sum_{t=1}^{T_{\mathrm{mc}}}\left(\theta_k^{(t)}-\hat{\theta}_k^{(t)}\right)^2} \quad (15)$$

where $T_{\mathrm{mc}}=1000$, K, $\theta_k^{(t)}$, $\hat{\theta}_k^{(t)}$ represent the times of the Monte-Carlo simulations, number of signals, true DOA of the k-th source signal at the t-th testing example, and estima-ted DOA of the k-th source signal at the t-th testing example respectively.

B. Spatial Spectra Estimation Result

The number of snapshots is 400, and SNR is 0dB. Firstly, consider the two-signals scenario with a small angle separation i.e., $\left(\theta_1=30.8°,\theta_2=33.2°\right)$, and large angle separation i.e., $\left(\theta_1=-20.2°,\theta_2=30.0°\right)$, it can be clearly observed that the testing angle-pairs deviate from the divided spatial grid set $\tilde{\boldsymbol{\theta}}$. The reconstructed spatial spectrum is shown in Fig. 2(a) and (b). Both signals are well separated, and CV-CNN can accurately reconstruct the correct spatial spectrum even in scenarios where the interval of two signals less than the Rayleigh resolution θ_{3dB} ≈6.4°, demonstrating it has super-resolution capability. The error of DOA estimation from the reconstructed spatial spectrum mainly comes from the quantization error of the discrete grids.

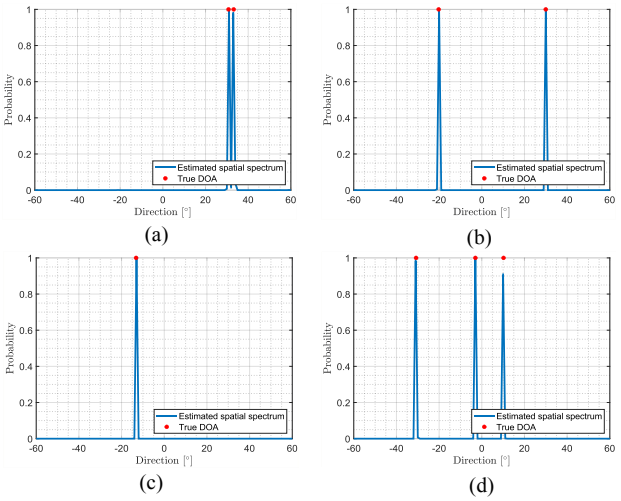

Fig. 2. Spatial estimation results for different scenarios. (a) Small angle separation with $\theta_1=30.8°$ amd $\theta_2=33.2°$, (b) Large angle separation with $\theta_1=-20.2°$ and $\theta_2=30.0°$, (c) One-signal scenario with $\theta=-13.2°$, (d) Three-signal scenario with $\theta_1=-30.8°$, $\theta_2=-3.0°$, and $\theta_3=10.2°$.

Generalization to the other number of sources is a basic requirement. Therefore, spatial spectral estimation of single-

signal scenario with $\theta = -13.2°$ and three-signal scenario with $\theta_1 = -30.8°$, $\theta_2 = -3.0°$, and $\theta_3 = 10.2°$ is performed separately, and the reconstructed spatial spectrum is shown in Fig. 2(c) and (d). In both scenarios, the spectral peaks emerge in the accurate true source direction with a probability close to 1, effectively distinguishing between the various DOAs. Furthermore, the results approach 0 in non-source directions without burr artifacts, verifying that the CV-CNN learns the mapping relationship between covariance matrix and source directions instead of solely fitting the training data, thereby exhibiting exceptional generalization capabilities for various numbers of sources.

C. Statistic Performance in Different Scenarios

1) RMSE varies with SNR: In this part, we estimate the DOA for two sources in directions $\theta_1 = 30.1°$ and $\theta_2 = 32.3°$. The number of snapshots is set to T=200 and the SNR is considered to vary uniformly at 5 dB intervals from -20 dB to 10 dB, i.e., $\Delta\vartheta = 5\text{dB}$. At each SNR level, 1000 Monte Carlo trials are performed and the variation of RMSE with SNR is showed in Fig. 3. In addition, the MUSIC, R-MUSIC, ESPRIT, and CNN methods are carraied out under the same conditions, and we also computed the CRLB_{uncr} in these scenarios to serve as a reference for a lower bound of estimation RMSE.

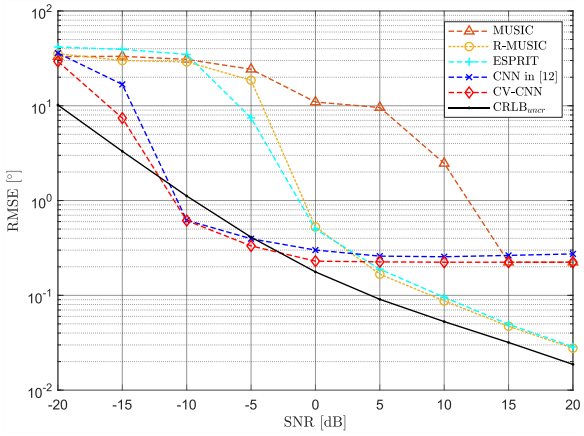

Fig. 3. RMSE versus *SNR* when T=200.

The results indicate that the RMSE of the proposed CV-CNN method is close to that of grid-based methods such as MUSIC and CNN under high SNR conditions, which is due to the presence of quantization errors in discrete grids, and the only way to further improve the performance is by increasing the resolution of the grid. However, the performance of CV-CNN and CNN, as deep learning-based methods, is demonstrated to be superior to traditional parameterization methods at low or moderate SNR levels. In particular, CV-CNN outperforms CNN methods at all SNR levels, which we attribute to the superiority of complex-valued convolution over real-valued convolution in complex-valued signal processing.

2) RMSE varies with Number of Snapshots: We fix the SNR to be 0dB, and the two directions are $\theta_1 = -10.3°$ and $\theta_2 = -7.6°$. The set {20,50,100,200,300,400,500,600,700, 800, 900, 1000} represents the number of snapshots used for covariance matrix estimation in each scenario.

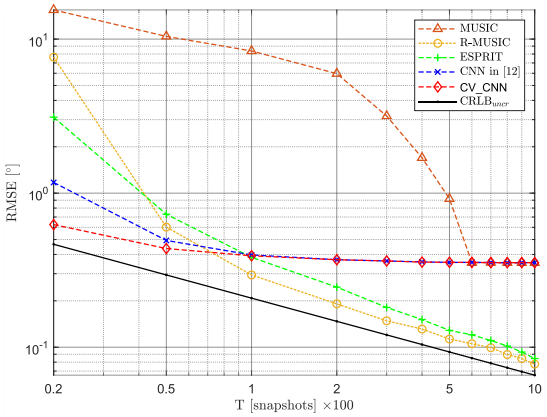

Fig. 4. RMSE versus the number of snapshots T at SNR=0dB.

Fig. 4 illustrates the variation of RMSE with respect to the number of snapshots T for two sources under the off-grid angle at SNR=0dB. The proposed CV-CNN method exhibits improved robustness for a small number of snapshots in comparison to the performance of traditional parametric methods, which are limited by grid resolution and achieve a grid-based estimation floor for a large number of snapshots. Additionally, compared to the CNN method, the proposed method enhances DOA estimation performance, particularly in scenarios with fewer than 100 snapshots.

V. Conclusion

In this work, we propose a robust DOA estimation method using modified 2D CV-CNN with data-driven manner under low SNR or limited snapshots. The 2D CV-CNN can well match the complex-valued received signal model and incorporate supplementary phase components to directly reconstruct the spatial spectrum, account for more accurate estimate than that of real-valued network on pre-defined grid. Through using the sparse constraint into the loss function and employing a two-stages training strategy, it proceed to improve the performance of DOA estimation. Numerical simulations results corroborate the superiority of our proposed method on super-resolution than that of three subspace-likewise methods and state-the-of-art CNN network method.

Apart from the above research findings, it should be noted that our future interest will be focusing on the off-grid DOA method to alleviate the quantization error, laying a theoretical foundation for the development of robust DOA. Inspired by the DOA estimated based on off-grid sparse Bayesian learning, the regression full connection is designed to estimate the first-order information parameters of the spatial spectrum and then used to fine-tune the current results in the future work.

References

[1] R. G. Pridham and R. A. Mucci, "A novel approach to digital beamforming," *J. Acoust. Soc. Amer.*, vol. 63, pp. 425-434, Feb. 1978.

[2] H. Steyskal, "Digital beamforming antennas: An introduction," *Microwave J.*, vol. 30, no. 1, pp. 107-124, Jan. 1987.

[3] H. Steyskal and J. F. Rose, "Digital beamforming for radar systems", *Microw. J.*, vol. 32, no. 1, pp. 121-136, Jan. 1989.

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

[5] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 8, pp. 336-339, 1983.

[6] T.J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 4, pp. 806–811, Aug. 1985.

[7] 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, Jul. 1989.

[8] A. G. Jaffer, "Maximum likelihood direction finding of stochastic sources: A separable solution," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1988, pp. 2893–2896.

[9] M. A. Ihedrane and S. Bri, "Direction of arrival estimation using MUSIC ESPRIT and maximum-likelihood algorithms for antenna arrays", *Walailak J. Sci. Tech.*, vol. 13, no. 6, pp. 491-502, 2015.

[10] 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, Dec. 2018.

[11] L. Wu, Z. Liu, and Z. Huang, "Deep convolution network for direction of arrival estimation with sparse prior," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1688–1692, Nov. 2019.

[12] 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.

[13] X. Gao, A. Liu and Y. Xiong, "Robust DOA Estimation Based on Deep Neural Networks in Presence of Array Phase Errors," *SSPD*, pp. 1-5, Sep. 2022.

[14] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[15] B. Zheng et al., "Joint sparse recovery for signals of spark-level sparsity and MMV Tail- ℓ_{21} minimization", *IEEE Signal Process. Lett.*, vol. 28, pp. 1130-1134, 2021.

[16] Z. Yang, L. Xie and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38-43, Jan. 2013.

[17] C. Trabelsi et al., "Deep complex networks, " 2017, [online] Available: https://arxiv.org/abs/1705.09792.

[18] G. Klambauer, T. Unterthiner, A. Mayr and S. Hochreiter, "Self-normalizing neural networks", *Proc. NeurIPS*, pp. 972-981, 2017.

[19] M. Jansson, B. Goransson, and B. Ottersten, "A subspace method for direction ofarrival estimation of uncorrelated emitter signals," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 945–956, Apr. 1999.