Multi-DOA Estimation Based on the KR Image Tensor and Improved CNN

Ye Yuan
National University of Defense Technology
Shuang Wu (ws02114006@163.com)
State Key Laboratory of Complex Electromagnetic Environment Elects on Electronics and Information System, National University of Defense Technology, Changsha, 410073 China
Yong Yang
Chinese Academy of Sciences
Naichang Yuan
National University of Defense Technology

Research Article

Keywords: Khatri-Rao (KR), convolutional neural network (CNN), curriculum learning scheme (CLS)

Posted Date: January 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-140235/v1

Version of Record: A version of this preprint was published at Scientific Reports on March 18th, 2021. See the published version at https://doi.org/10.1038/s41598-021-85864-5.

Multi-DOA estimation based on the KR image tensor and improved CNN

Ye Yuan[1], Shuang Wu[1,*], Yong Yang[2], and Naichang Yuan[1]

[1]State Key Laboratory of Complex Electromagnetic Environment Elects on Electronics and Information System, National University of Defense Technology, Changsha, 410073 China
[2]Innovation Academy for Microsatellites, Chinese Academy of Sciences, Shanghai, 200120, China
[*]ws02114006@163.com

ABSTRACT

In this paper, we propose an improved convolutional neural network (CNN) to solve the multi-DOA estimation problem. We use Khatri-Rao (KR) product to obtain the KR image tensor of covariance matrix and use the proposed estimation CNN to process the tensor. In order to increase the generalization of the proposed CNN and adapt the multi-label classification problem, we use the curriculum learning scheme (CLS) and partial label strategy (PLS) to develop an efficient training procedure. We implement several experiments to demonstrate the satisfying performance of the proposed estimation method. The simulation results show that our proposed method can finish the high resolution multi-DOA estimation use only a few sensors. Furthermore, the proposed method can obtain high estimation accuracy under low SNR situations and use fewer snapshots.

Introduction

Direction-of-Arrival (DOA) estimation technologies are widely used in the modern radar and communication systems[1–5]. In the past decades, researchers have presented many super-resolution estimation methods. Two of the most classic methods (subspace based methods) are known as the multiple signal classification (MUSIC) algorithm and estimating signal parameter via rotational invariance techniques (ESPRIT) algorithm[6,7].

Many articles have proposed numerous variants of MUSIC and ESPRIT methods[8–10], some improvements of the estimation performance have already been made. Researchers always combine the MUSIC method with other technologies to enhance the estimation performance, such as particle swarm optimization based MUSIC (PSO-MUSIC) method used for tracking sound source[11] and Low-degree root-MUSIC method used for fast estimation[12]. For the improvements of ESPRIT method, it is important to enhance the accuracy of the method and improve the estimation speed[13,14].

As an interesting mathematical tool, the Khatri-Rao (KR) product has been introduced into the DOA estimation field. Researchers have proposed many efficient KR product based methods, such as the KR subspace approach used for quasi-stationary signals[15] and the KR product based MUSIC algorithm used for Gaussian signals[16]. Researchers also find that the KR product can enhance the degrees of freedom (DOF) for arrays. Based on this conception, the so-called nested array has been proposed both in the 1-D[16] and 2-D[17,18] estimation fields. Although these KR subspace based methods have been well developed and many varietal algorithms have been proposed[19,20], the defects of these methods are obvious. Firstly, the estimation performance of subspace based methods relies on the prior information, and the prior information is hard to obtain in reality[21]. Secondly, the decomposition operation is computationally complex. For MUSIC based method, the spectrum searching operation also cost plenty of computing resources, therefore, the subspace based methods is not applicable for systems which require reliable timeliness. Therefore, it is necessary to find some alternative methods with independence of prior information to solve the DOA estimation problem.

Prior articles have proved that DOA estimation problem is a kind of convexity problem[22,23], therefore, researchers have brought some efficient convex optimization algorithms or machine learning methods into the DOA estimation field, such as radial basis function neural network (RBF-NN)[24], support vector regression (SVR)[25], and neural network[26]. These machine learning based methods are also used for dealing with special signal forms, such as the sparse Bayesian learning (SBL) method for coherent sources[27] and the modified quantum genetic algorithm (QGA) for wideband signals[28]. These methods can automatically learn the probability density function (PDF) of DOA[29], therefore, the estimation method based on machine learning need less prior knowledge from the signal[30]. But the training and estimation processes become more complex, and the accuracy of the estimating results depends on the parameters of methods. These parameters are derived by some verification experiments and they are hard to obtain[31]. Therefore, it is necessary to find another way to develop DOA estimation methods with more convenient operations.

With the rapid development of artificial intelligent (AI) technologies[32], some AI based estimation methods have been proposed in recent years. Researchers introduce the deep neural network (DNN) into the DOA estimation field and develop a DNN group to do the uniform linear array (ULA) estimation task[33]. Despite that the proposed method does not rely on the prior information and shows robustness with the array imperfections, it has some defects which need to be overcome. Firstly, while estimating DOA information by using the proposed DNN group, the DOAs of two adjacent receipt signals will need at least 10° interval. This means that the resolution of the proposed DNN group is limited, and the accuracy of estimating results will decline if DOAs of signals are getting too closer. Secondly, the ULA used in this article contains 10 sensors, which means the array aperture is pretty large. Thirdly, these DNN based methods use the vector of array observation data or vectorized covariance matrix as the input tensor of the network[34]. This operation ignores the structure of covariance matrix and can reduce the quantity of information. In order to make full use of the information hidden in the covariance matrix, researchers also propose some convolution neural network (CNN) based methods, such as the deep CNN (DCNN) used for estimating multi-DOA with sparse prior[35], the DCNN used for estimating DOAs of multi-speaker with noise signals[36], and the de-multipath CNN models used for achieving multipath DOA estimation[37]. Although these CNN based methods use the convolutional layers to extract the deep features of covariance matrix, the pre-processing operation proposed in these articles[35–37] can complicate the estimation problem and cause time-consuming.

In this paper, we propose a novel AI based method to estimate multi-DOA information directly from the covariance matrix. We transform the estimation problem into an image classification problem by using the image tensor of covariance matrix. We discuss the image feature of the covariance matrix obtained by the KR product and design a simple CNN with special training technologies. The rest of this paper is organized as follows. In method section, we explain the basic conceptions of the proposed estimation method and present the image tensor obtained by the KR product. We also show the proposed training scheme and discuss the details of the training technologies in method section. In result section, we propose several experiments to show the satisfying performance of the proposed estimation CNN. At last, we conclude the main work of this paper in discussion section.

Methods

KR product and image tensor

We assume that K individual Gaussian signals spatially distributed at the direction of $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_K]$. All signals simultaneously impinge on an ULA with M sensors. The array observation data $\boldsymbol{x}(n)$ can be expressed as Eq. (1):

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

where n denotes the n-th snapshot. $\boldsymbol{s}(n)$ (resp. $\boldsymbol{v}(n)$) denotes the signal vector (resp. noise vector). $\boldsymbol{A} = [\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \cdots, \boldsymbol{a}(\theta_K)]$ represents the array manifold, where $\boldsymbol{a}(\theta_k)$ can be expressed as Eq. (2):

$$\boldsymbol{a}(\theta_k) = \left[1, e^{-\mathrm{j}2\pi\frac{d}{\lambda}\sin\theta_k}, \cdots, e^{-\mathrm{j}2\pi\frac{(M-1)d}{\lambda}\sin\theta_k}\right]^T, \tag{2}$$

where λ denotes the wavelength of signals. We set the inter-sensor space of ULA $d = \lambda/2$. The covariance matrix $\boldsymbol{R}$ of observation data can be calculated as Eq. (3):

$$\boldsymbol{R} = E\{\boldsymbol{x}\boldsymbol{x}^H\} = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}(n)\boldsymbol{x}^H(n) = \boldsymbol{A}\boldsymbol{R}_s\boldsymbol{A}^H + {\sigma_v}^2\boldsymbol{I}_M. \tag{3}$$

where N denotes the number of snapshots. $\boldsymbol{R}_s = \mathrm{diag}(\boldsymbol{\sigma})$ represents the power matrix of signals ($\boldsymbol{\sigma} = \left[{\sigma_1}^2, {\sigma_2}^2, \cdots, {\sigma_K}^2\right]^T$). ${\sigma_v}^2$ denotes the power of white noise and $\boldsymbol{I}_M$ denotes the $M \times M$ unit matrix. The vectorized $\boldsymbol{R}$ can be regarded as signals impinge on a new array whose manifold is given by $\hat{\boldsymbol{A}} = \boldsymbol{A}^* \odot \boldsymbol{A}$[15].

$$\boldsymbol{y} = \mathrm{vec}(\boldsymbol{R}) = \mathrm{vec}\left[\sum_{k=1}^{K}{\sigma_k}^2\left(\boldsymbol{a}(\theta_k)\boldsymbol{a}^H(\theta_k)\right)\right] + {\sigma_v}^2\vec{\boldsymbol{1}}_M = (\boldsymbol{A}^* \odot \boldsymbol{A})\boldsymbol{\sigma} + {\sigma_v}^2\vec{\boldsymbol{1}}_M, \tag{4}$$

where $\odot$ denotes the KR product. $\vec{\boldsymbol{1}}_M = \left[\boldsymbol{e}_1^T, \boldsymbol{e}_2^T, \cdots, \boldsymbol{e}_M^T\right]^T$ and $\boldsymbol{e}_m$ denotes a column vector with M elements. The m-th element of $\boldsymbol{e}_m$ equals 1 while others equal 0.

The KR product can also generate several identical elements in $\boldsymbol{y}$[16], therefore, we remove the redundant elements and obtain a new vector $\hat{\boldsymbol{y}}$. The new covariance matrix $\hat{\boldsymbol{R}}$ can be expressed as Eq. (5):

$$\hat{\boldsymbol{R}} = \hat{\boldsymbol{y}}\hat{\boldsymbol{y}}^H = \hat{\boldsymbol{A}}\boldsymbol{\sigma}\boldsymbol{\sigma}^H\hat{\boldsymbol{A}}^H + {\sigma_v}^4\hat{\boldsymbol{e}}\hat{\boldsymbol{e}}^H + {\sigma_v}^2\hat{\boldsymbol{A}}\boldsymbol{\sigma}\hat{\boldsymbol{e}}^H + {\sigma_v}^2\hat{\boldsymbol{e}}\boldsymbol{\sigma}^H\hat{\boldsymbol{A}}^H, \tag{5}$$

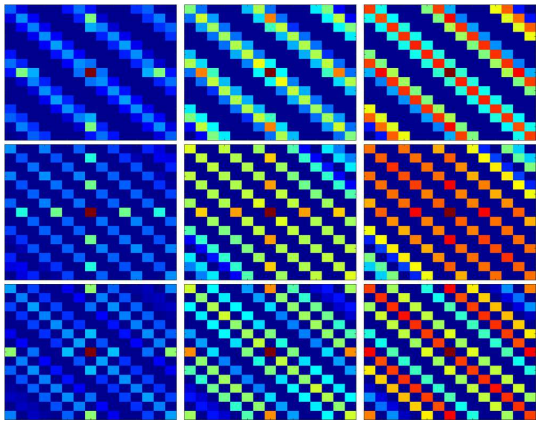

Figure 1. Image feature of matrix $\mathrm{Re}\{\hat{\boldsymbol{R}}\}$. From top to bottom, $\theta = 20°$, $40°$, and $60°$. From left to right, the SNR equals -5dB, 0dB, and 5dB.

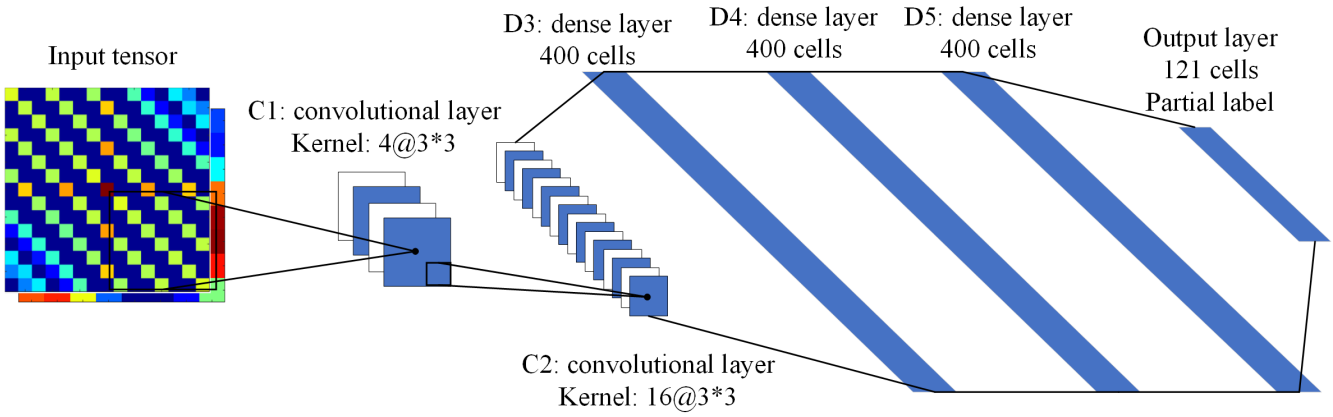

Figure 2. Framework of the proposed simple estimation CNN.

where $\hat{\boldsymbol{e}}$ denotes a vector whose $(M(M-1)/2+1)$-th element equals 1 while others equal 0.

We reshape $\hat{\boldsymbol{R}}$ into an image tensor $\mathscr{R}$ where the real part matrix $\mathrm{Re}\{\hat{\boldsymbol{R}}\}$ and imaginary part matrix $\mathrm{Im}\{\hat{\boldsymbol{R}}\}$ are two individual channels of $\mathscr{R}$. We propose a method to extract DOA information from the image feature of $\mathscr{R}$, and the advantages of this method can be summarized as follows:

- The use of tensor $\mathscr{R}$ can be processed by the CNN, which is an efficient estimation tool with more robustness[38].
- The KR product can extend the DOF of array, which allows the array to estimate more DOAs[17,18]. Furthermore, KR product can ensure the size of $\mathscr{R}$ and help the network obtain more information.
- We investigate the image feature of $\mathscr{R}$ (shown in Figure 1, here we set $M = 8$) and find out that the main structure of $\mathscr{R}$ only corresponding to θ. The estimation environment can hardly influence the image feature of $\mathscr{R}$. Therefore, it is possible to develop some estimation methods with high performance under harsh environment.

Based on these conceptions, we design an improved estimation CNN to deal with the multi-DOA estimation problem with the proposed KR image tensor $\mathscr{R}$.

Curriculum learning and partial label

Figure 2 shows the framework of the proposed CNN. It contains 6 layers: two convolutional layers, three dense layers, and one output layer. Two convolutional layers contain 4 and 16 individual feature maps, respectively. We set the kernel size as 4×4. Each dense layer contains 400 cells. For the output layer, we divide the angle space of θ (from $-60°$ to $60°$) by $1°$ and use 121 output cells to represent the spectrum of θ, therefore, the output layer contains 121 categories. We use the ReLU function as the activations of the first five layers while the output layer uses Sigmoid function. As a consequence, the proposed CNN transform the multi-DOA estimation problem into a multi-label classification problem of image tensor.

Algorithm 1 Training scheme of the proposed estimation CNN

Require: Dataset of $2@M \times M$ image tensor $\mathscr{R}$;
Ensure: Prediction of θ spectrum;
Initiate dataset with Partial Labels;
Roughly train the network with all samples;
Divide the samples into B sub-datasets according to SNR;
for B total generations of training **do**
  repeat
    Train the network with current generation of dataset;
    Predict and update the missing labels;
  until meet the stopping criteria
  Mix the dataset with training samples in the next generation;
end for

In order to increase the generalization of network, we introduce the curriculum learning scheme (CLS)[39] into CNN. We also use the partial label strategy (PLS)[40] to achieve efficient multi-label classification. The proposed learning scheme is shown in algorithm 1. The whole training scheme shown in algorithm 1 make the proposed estimation CNN become a so-called CurriculumNet[39], and the training details are shown as follows. At first, we roughly train the network with all training samples and obtain a basic estimation CNN. Then we generate B clusters of samples according to the SNR information and use them as the sub-datasets. Here we fix the range of SNR at $[-10\text{dB}, 20\text{dB}]$ and divide it into 6 segments. The training samples is also divided into 6 subsets and sorted from the sample subsets with higher SNR to the sample subsets with lower SNR. We denote the sorted subsets as $\mathscr{D}_i, i = 1, 2, \cdots, 6$ and the training samples in the current generation as $\mathscr{D}_C$. During each generation of training, we continuously mix the subsets together as the training samples: $\mathscr{D}_C = \mathscr{D}_C \cup \mathscr{D}_i$, where the initial $\mathscr{D}_C = \emptyset$.

To deal with the multi-DOA estimation problem, we also use the PLS to decrease the scale of dataset space and obtain multi-classification results with high performance. We denote the training data as $\mathscr{D} = \left\{ \left(\mathscr{R}^{(1)}, \boldsymbol{z}^{(1)} \right), \left(\mathscr{R}^{(2)}, \boldsymbol{z}^{(2)} \right), \cdots, \left(\mathscr{R}^{(S)}, \boldsymbol{z}^{(S)} \right) \right\}$, where $\mathscr{R}^{(s)}$ (resp. $\boldsymbol{z}^{(s)}$) denotes the s-th sample (resp. label vector). S denotes the number of training data. $\boldsymbol{z}^{(s)} = \left[z_1^{(s)}, z_2^{(s)}, \cdots, z_C^{(s)} \right]$ where $C = 121$ represents the number of output categories. We denote $z_c^{(s)} = 1$ (resp. -1 and 0) means the present (resp. absent and unknown) of this label. We also define a hyperparameter $p_2\%$ to randomly choose which labels should be set as present labels.

The partial labels is collaborated with partial-BCE loss[40]. We denote the s-th output of estimation network as $\boldsymbol{w}^{(s)} = \left[w_1^{(s)}, w_2^{(s)}, \cdots, w_C^{(s)} \right]$. The partial-BCE loss function $\mathscr{L}$ can be expressed as Eq. (6):

$$\mathscr{L} = -\frac{g(p_2)}{CS} \sum_{s=1}^{S} \sum_{c=1}^{C} \left[\mathbf{1}_{\left[z_c^{(s)} = 1 \right]} \log \left(\frac{1}{1 + \exp(-w_c^{(s)})} \right) + \mathbf{1}_{\left[z_c^{(s)} = -1 \right]} \log \left(\frac{-w_c^{(s)}}{1 + \exp(-w_c^{(s)})} \right) \right], \tag{6}$$

where $\mathbf{1}_{[\bullet]}$ denotes a binary detector of z_c. $g(\bullet)$ is a normalization probability distribution corresponding to the proportion of present labels.

$$g(p_2) = \alpha p_2^{\gamma} + \beta, \tag{7}$$

where α, β, and γ are hyperparameters which control the value of g. We fix the value of γ at 3 and set $g(0) = 5$, then we can obtain that $\alpha = -4$ and $\beta = 5$.

In order to recover the missing labels, we implement the Bayesian uncertainty strategy (BUS)[41] after each generation of training. We define a score vector $\boldsymbol{v}^{(s)} = \left[v_1^{(s)}, v_2^{(s)}, \cdots, v_C^{(s)} \right]$ to represent the prediction of label vector. We also define a hyperparameter ρ as the detection threshold. The objective function of prediction can be expressed as Eq. (8):

$$\min_{\mathscr{F} \in \mathscr{R}^{16@3\times3}, \boldsymbol{v} \in \{0,1\}^{S \times C}} J(\mathscr{F}, \boldsymbol{v}) = \kappa \|\mathscr{F}\|^2 + G(\boldsymbol{v}; \boldsymbol{\rho}) + \frac{1}{SC} \sum_{s=1}^{S} \sum_{c=1}^{C} v_c^{(s)} \mathscr{L}_c \left(w_c^{(s)}, z_c^{(s)} \right), \tag{8}$$

where κ is a positive value which controls the weight of $\|\mathscr{F}\|$. $\mathscr{L}_c$ denotes the loss of the c-th category. $G(\bullet)$ denotes the curriculum learning scheme mentioned before. The label vector $\boldsymbol{v}$ can be predicted as Eq. (9):

$$v_c^{(s)} = \infty \left[\mathscr{U}(\mathscr{R}^{(s)})_c \leq \rho \right], \tag{9}$$

where $\mathscr{U}(\bullet)_c$ denotes the Bayesian uncertainty[41] of the c-th category in the s-th sample.

Results

In order to verify the effectiveness of our method, we propose several experiments. We set $M = 4$, therefore, the number of virtual sensors equals to 7 according to Eq. (4) and (5). We set the learning rate of the proposed estimation CNN as 0.01 and use the stochastic gradient descent (SGD) optimizer to train the network. During training the network, the SNR of samples is randomly selected in the range of $[-10\text{dB}, 20\text{dB}]$, and the number of snapshots N is fixed at 512.

Training performance

we use the Micro-F1 metrics $\mathscr{M}_{\text{mF1}}$ proposed in[42] to evaluate the accuracy of training and validation. We change the dropout rate of label from 0% to 75% with the step of 25% and record the accuracy performance of the proposed estimation CNN.

$$\mathscr{M}_{\text{mF1}} = \frac{2\sum_{c=1}^{C}\sum_{s=1}^{S}\mathbf{1}[z_c^{(s)} = \hat{z}_c^{(s)}]}{\sum_{c=1}^{C}\sum_{s=1}^{S} z_c^{(s)} + \sum_{c=1}^{C}\sum_{s=1}^{S}\hat{z}_c^{(s)}}. \tag{10}$$

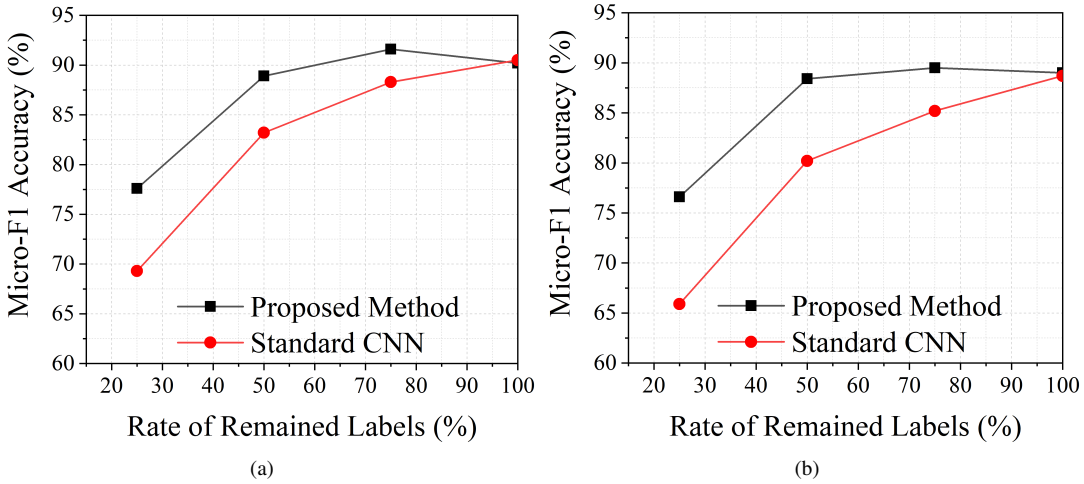

Figure 3. Micro-F1 accuracy comparison of training with PLS and standard training scheme. (a) training accuracy, (b) validation accuracy.

Table 1. Micro-F1 Accuracy Versus Different Number of Samples

Number of samples S	20000	40000	60000	80000	100000
Standard CNN (%)	52.33	61.26	71.54	82.27	89.64
Proposed Method (%)	76.46	85.61	89.93	90.12	89.91

Figure 3 shows the $\mathscr{M}_{\text{mF1}}$ score versus different value of $p_2\%$. It is implied that the standard training procedure cannot estimate DOAs precisely when labels are missing, while the proposed learning strategy have a good performance even when samples missing half of labels. When $p_2\% = 75\%$, the proposed network shows the best performance in both training and validation accuracy. We also investigate the relationship between S and $\mathscr{M}_{\text{mF1}}$. The results in Table 1 shows that $\mathscr{M}_{\text{mF1}}$ of the standard training scheme declines quickly when S become small. While the proposed partial label strategy can significantly reduce the required scale of training dataset.

After analyzing the relationship between accuracy and training samples, we decide to fix the number of samples S at 60000 and set the rate of remained labels as 75%. Figure 4 shows the training and validation performance of proposed training scheme and standard training scheme, here we implement 100 independent times of training for each sub-dataset. The total training time is 600. The $\mathscr{M}_{\text{mF1}}$ comparison results illustrate that standard training scheme can cause serious over-fitting problem, while CLS technology can overcome this problem and greatly improve the generalization of the proposed estimation network.

By using the PLS and CLS training technologies, the final $\mathscr{M}_{\text{mF1}}$ validation accuracy of the proposed estimation network can exceed 90%, which means the proposed estimation CNN is well trained. After training the network, we can implement some performance test corresponding to the stability of the proposed estimation CNN.

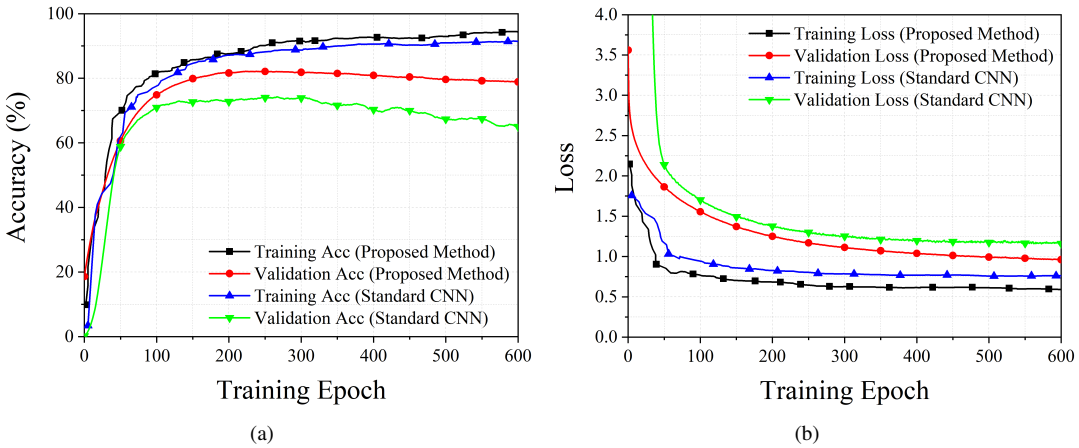

Figure 4. Training performance of the proposed method and standard training scheme. (a) training and validation accuracy, (b) training and validation loss.

Estimation performance

The angle spectrum obtained from the proposed estimation CNN is a discrete spectrum, therefore, we test the estimation performance of integer θ at first. Assuming 5 incoherent signals impinging at the direction of $\boldsymbol{\theta} = [-47°, -33°, 2°, 25°, 45°]$. The SNR (resp. number of snapshots N) is fixed at 10dB (resp. 512). Figure 5(a) shows the estimation spectrum. It is illustrated that the proposed estimation network can obtain a MUSIC-liked angle spectrum, and can estimate more signals than the number of physical sensors. We also randomly changing the SNR from -10dB to 10dB and implement 100 Monte Carlo experiments. The record of estimation error shown in Figure 5(b) implies that DOA may drop in the neighboring grid with the probability only around 5% . Both figures prove that the proposed estimation CNN can obtain high estimation accuracy for integer θ.

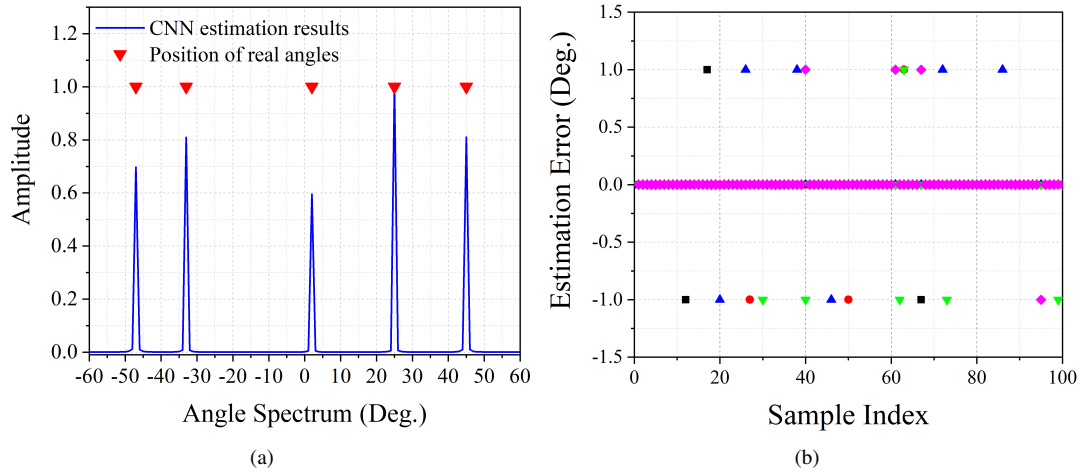

Figure 5. Estimation results of integer θ. (a), angle spectrum obtained by the proposed estimation CNN, (b) estimation error for 100 Monte Carlo experiments.

In order to estimate non-integer θ and achieve high-resolution results, we take the method of amplitude interpolation. The estimation of non-integer θ can be calculated as Eq. (11).

$$\theta_I = z_I + \frac{5}{4}\sum_{j=1}^{3} j(z_{I+j} - z_{I-j}), \tag{11}$$

where I denotes the I-th peak of the output spectrum. We set 5 non-integer signals to test the estimation performance of the proposed estimation network, and the estimation results (under SNR= 0dB situation) is shown in Figure 6. It is implied that the

estimation error is fluctuating around 0°, and the maximum error is no more than 0.5°. Figure 6 also illustrates that the proposed estimation method can cover the whole range of θ and achieve a high accuracy performance for estimating non-integer θ.

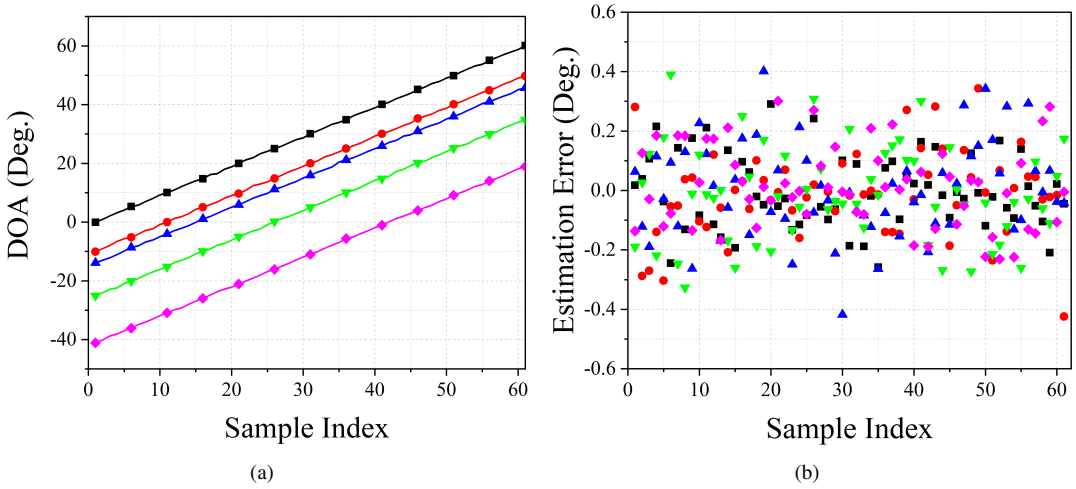

Figure 6. Estimation results of non-integer θ. (a), estimation results of each sample, (b) estimation error of each sample.

Stability of the proposed estimation CNN

This experiment investigates the stability of the proposed estimation CNN. We compare the proposed method with other methods, including MUSIC algorithm[6], RBF-NN[24], DNN group[33], and standard CNN (without CLS and PLS). The results are obtained based on 200 times independent Monte Carlo experiments. Here we use the root mean squared error (RMSE) to evaluate the performance of all methods, which can be calculated as Eq. (12).

$$\mathrm{RMSE} = \sqrt{\frac{1}{PK}\sum_{p=1}^{P}\sum_{k=1}^{K}\left(\theta_{kp}-\hat{\theta}_{kp}\right)^2}, \tag{12}$$

where P denotes the number of Monte Carlo experiments. $\hat{\theta}_{kp}$ (resp. θ_{kp}) is the k-th real angle (resp. estimated angle by the proposed estimation network).

First we investigate the relationship between RMSE and SNR. We fix the N at 512 and change the SNR from -10dB to 20dB. According to Figure 6(a), the RMSE response of RBF-NN method has the worst performance, the value of RMSE become lower than 1° when SNR is higher than 2dB. The curves of DNN group and MUSIC algorithm have similar performance when SNR is higher than 0dB, while MUSIC algorithm performs better than DNN group in the case of low SNR situation. The RMSE of standard CNN is pretty low when SNR is higher than 4dB, but the curve rises sharply as the SNR decreases. Compared with these four methods, the RMSE of proposed estimation CNN has the best performance, and the RMSE is still lower than 1° when SNR$= -10$dB.

We also investigate the relationship between RMSE and number of snapshots N. Figure 6(b) shows the RMSE results versus N, here we fix the value of SNR at 10dB. The RMSE responses of these methods (except RBF-NN method) reveal an inverse proportional function character. For the proposed estimation CNN, the RMSE remains low when N is greater than 16. This implies that the proposed estimation CNN can finish the estimation task precisely by using only a few of snapshots. However, when N is set to be 4, the RMSE rises to around 1°, which depicts that the proposed estimation CNN cannot apply to single snapshot estimation processing.

Resolution of the proposed estimation CNN

In order to test the resolution of the proposed estimation CNN, we investigate the relationship between estimation performance and angle difference $\Delta\theta$. Figure 8(a) shows the estimation results of two 0dB signals with a set of angular separations $\{3°, 4°, 5°, 6°\}$. It is obvious that the proposed estimation CNN can distinguish two adjacent DOAs in the sector of $[-60°, 60°]$. Figure 8(b) shows the RMSE results versus angle difference $\Delta\theta$. Here we fix the SNR (resp. number of snapshots N) at 10dB (resp. 512). The DNN group[33] can only estimate two signals with $\Delta\theta \geq 10°$, therefore, we do not show its performance in Figure 8(b). When $\Delta\theta \geq 3°$, the proposed method can obtain higher estimation accuracy than other methods. When $\Delta\theta < 3°$

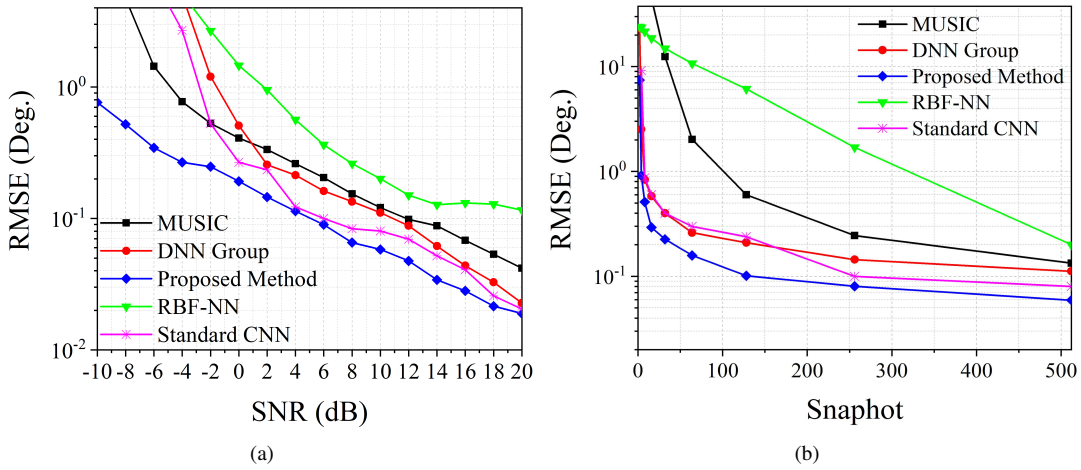

Figure 7. RMSE versus number of snapshots N, SNR equals to 10dB.

only the MUSIC algorithm and the proposed method can obtain valid estimation results. However, the precision of our proposed method is slightly lower than the MUSIC algorithm when $\Delta\theta$ become small.

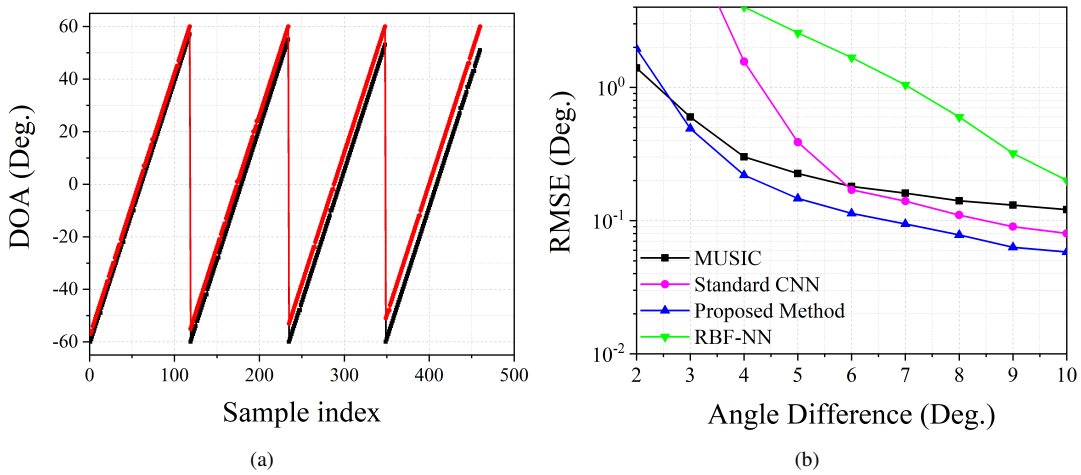

Figure 8. RMSE versus angle difference $\Delta\theta$, SNR equals to 10dB and number of snapshots $N = 512$.

Discussion

The experiments mentioned above demonstrate the satisfying performance of the proposed method. The results in Figure 3, Figure 4, and Table 1 illustrate that CLS and PLS are suitable for solving the DOA estimation problem. These two training technologies can improve the generalization of the proposed estimation CNN and reduce the scale of input and output space. Figure 5 and Figure 6 prove the validity of the proposed estimation method. By using the KR product, the DOF of array has been improved, which allows us to estimate more sources than the number of physical sensors. Figure 7 shows that our proposed method has better stability than prior estimation methods, while Figure 8 illustrate the high-resolution performance of the proposed estimation CNN.

The proposed estimation CNN performs better than the subspace based estimation methods because it is a data-driven method, and the classification results can be hardly affected by the background information (like SNR and number of snapshots). The proposed estimation CNN only use the deep features hidden in $\mathscr{R}$ to do the estimation task, and these features only rely on the DOA information. The inessential information can be ignored by the proposed estimation CNN while subspace based methods and machine learning based methods can be influenced by these information.

In conclusion, an improved CNN used for multi-DOA estimation is proposed in this paper. We discuss the image feature of tensor $\mathscr{R}$ obtained by the KR product. Based on the conceptions proposed in method section, we use an efficient training scheme for the proposed CNN. By adding the CLS and PLS into the network, the generalization of the CNN has been greatly improved. The simulation results of all experiments show that our proposed estimation method has high estimation performance and can tolerate the harsh estimation environment.

References

1. Qiang, X. *et al.* Adaptive doa estimation with low complexity for wideband signals of massive mimo systems. *Signal Process.* **176**, 107702 (2020).
2. Amirsoleimani, S. & Olfat, A. Wideband modal orthogonality: A new approach for broadband doa estimation. *Signal Process.* **176**, 107696 (2020).
3. Piccinni, G., Torelli, F. & Avitabile, G. Innovative doa estimation algorithm based on lyapunov theory. *IEEE Transactions on Circuits Syst. II: Express Briefs* **67**, 2219–2223 (2020).
4. Xu, Z., Li, H. & Yang, K. A modified differential beamforming and its application for doa estimation of low frequency underwater signal. *IEEE Sensors J.* **20**, 8890–8902 (2020).
5. Li, S. *et al.* Covariance matrix reconstruction for doa estimation in hybrid massive mimo systems. *IEEE Wirel. Commun. Lett.* **9**, 1196–1200 (2020).
6. Schmidt, R. O. A signal subspace approach to multiple emitter location and spectral estimation (1982).
7. Roy, R. & Kailath, T. Esprit-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoust. Speech, Signal Process.* **37**, 984–995 (1989).
8. Jeong, J.-S., Sakaguchi, K., Takada, J.-i. & Araki, K. Performance of music and esprit for joint estimation of doa and angular spread in slow fading environment. *IEICE Transactions on Commun.* **85**, 972–977 (2002).
9. McCloud, M. L. & Scharf, L. L. A new subspace identification algorithm for high-resolution doa estimation. *IEEE Transactions on Antennas Propag.* **50**, 1382–1390 (2002).
10. Liu, M., Cao, H. & Wu, Y. Improved subspace-based method for 2d doa estimation with l-shaped array. *Electron. Lett.* **56**, 402–405 (2020).
11. Omiya, K. & Suyama, K. Sound source tracking via two microphones based on music using pso. In *2016 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 1–6 (IEEE, 2016).
12. Yan, F., Jin, M., Zhou, H., Wang, J. & Liu, S. Low-degree root-music algorithm for fast doa estimation based on variable substitution technique. *Sci. China Inf. Sci.* **63**, 159206 (2020).
13. Fayad, Y., Wang, C., Cao, Q. & Hafez, E. D. S. A developed esprit algorithm for doa estimation. *Frequenz* **69**, 263–269 (2015).
14. Potts, D. & Tasche, M. Fast esprit algorithms based on partial singular value decompositions. *Appl. Numer. Math.* **88**, 31–45 (2015).
15. Ma, W.-K., Hsieh, T.-H. & Chi, C.-Y. Doa estimation of quasi-stationary signals with less sensors than sources and unknown spatial noise covariance: A khatri–rao subspace approach. *IEEE Transactions on Signal Process.* **58**, 2168–2180 (2009).
16. Pal, P. & Vaidyanathan, P. P. Nested arrays: A novel approach to array processing with enhanced degrees of freedom. *IEEE Transactions on Signal Process.* **58**, 4167–4181 (2010).
17. Pal, P. & Vaidyanathan, P. Nested arrays in two dimensions, part i: Geometrical considerations. *IEEE Transactions on Signal Process.* **60**, 4694–4705 (2012).
18. Pal, P. & Vaidyanathan, P. Nested arrays in two dimensions, part ii: Application in two dimensional array processing. *IEEE Transactions on Signal Process.* **60**, 4706–4718 (2012).
19. Qin, S., Zhang, Y. D. & Amin, M. G. Generalized coprime array configurations for direction-of-arrival estimation. *IEEE Transactions on Signal Process.* **63**, 1377–1390 (2015).
20. Ren, S., Dong, W., Li, X., Wang, W. & Li, X. Extended nested arrays for consecutive virtual aperture enhancement. *IEEE Signal Process. Lett.* **27**, 575–579 (2020).
21. Li, P., Zhang, M. & Zhong, Z. Integrative wideband doa estimation method based on sparse representation. *Electron. letters* **47**, 1251–1252 (2011).

22. Stoica, P. & Sharman, K. C. Maximum likelihood methods for direction-of-arrival estimation. *IEEE Transactions on Acoust. Speech, Signal Process.* **38**, 1132–1143 (1990).

23. Liu, L. & Wei, P. Off-grid doa estimation based on analysis of the convexity of maximum likelihood function. *IEICE Transactions on Fundamentals Electron. Commun. Comput. Sci.* **E98A** (2014).

24. Guo, W., Qiu, T., Tang, H. & Zhang, W. Performance of rbf neural networks for array processing in impulsive noise environment. *Digit. Signal Process.* **18**, 168–178 (2008).

25. Wu, L. & Huang, Z. Coherent svr learning for wideband direction-of-arrival estimation. *IEEE Signal Process. Lett.* 1–1 (2019).

26. Agatonovic, M., Stankovic, Z., Doncov, N., Sit, L. & Zwick. Application of artificial neural networks for efficient high-resolution 2d doa estimation. *Radioengineering* **21**, 1178–1186 (2012).

27. Liu, Z.-M., Huang, Z.-T. & Feng, D.-W. Direction-of-arrival estimation for coherent sources via sparse bayesian learning. *Int. J. Antennas Propag.* (2014).

28. Liu, F., Shaoqian, L. I., Liang, M. & Laizhao, H. U. Wideband signal doa estimation based on modified quantum genetic algorithm. *Ieice Transactions on Fundamentals Electron. Commun. Comput. Sci.* **89**, 648–653 (2006).

29. Lam, C. J. & Singer, A. C. A recursive filter approach to adaptive bayesian beamforming for unknown doa. In *Sensor Array and Multichannel Signal Processing Workshop, 2008. SAM 2008. 5th IEEE* (2008).

30. Hu, N., Sun, B., Wang, J. & Yang, J. Covariance-based doa estimation for wideband signals using joint sparse bayesian learning. In *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)* (2017).

31. Gerstoft, P. *et al.* Multisnapshot sparse bayesian learning for doa. *IEEE Signal Process. Lett.* (2016).

32. Hu, D., Zhang, Y., He, L. & Wu, J. Low-complexity deep-learning-based doa estimation for hybrid massive mimo systems with uniform circular arrays. *IEEE Wirel. Commun. Lett.* **9**, 83–86 (2019).

33. Liu, Z.-M., Zhang, C. & Philip, S. Y. Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections. *IEEE Transactions on Antennas Propag.* **66**, 7315–7327 (2018).

34. Guo, Y., Zhang, Z., Huang, Y. & Zhang, P. Doa estimation method based on cascaded neural network for two closely spaced sources. *IEEE Signal Process. Lett.* **27**, 570–574 (2020).

35. Wu, L., Liu, Z.-M. & Huang, Z.-T. Deep convolution network for direction of arrival estimation with sparse prior. *IEEE Signal Process. Lett.* **26**, 1688–1692 (2019).

36. Chakrabarty, S. & Habets, E. A. Multi-speaker doa estimation using deep convolutional networks trained with noise signals. *IEEE J. Sel. Top. Signal Process.* **13**, 8–21 (2019).

37. Xiang, H., Chen, B., Yang, T. & Liu, D. Improved de-multipath neural network models with self-paced feature-to-feature learning for doa estimation in multipath environment. *IEEE Transactions on Veh. Technol.* **69**, 5068–5078 (2020).

38. Goodfellow, I., Bengio, Y., Courville, A. & Bengio, Y. *Deep learning* (MIT press Cambridge, 2016).

39. Guo, S. *et al.* Curriculumnet: Weakly supervised learning from large-scale web images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 135–150 (2018).

40. Durand, T., Mehrasa, N. & Mori, G. Learning a deep convnet for multi-label classification with partial labels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 647–657 (2019).

41. Kendall, A. & Gal, Y. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems*, 5574–5584 (2017).

42. Tang, L., Rajan, S. & Narayanan, V. K. Large scale multi-label classification via metalabeler. In *Proceedings of the 18th International Conference on World Wide Web*, 211–220 (2009).

Author contributions statement

Y.Yuan developed the DOA estimation algorithm and conceived the experiments, S.Wu conducted the experiments and analyzed the results. Y. Yong and N. Yuan revised the manuscript.

Competing interests

The author declares no competing interests.

Figures

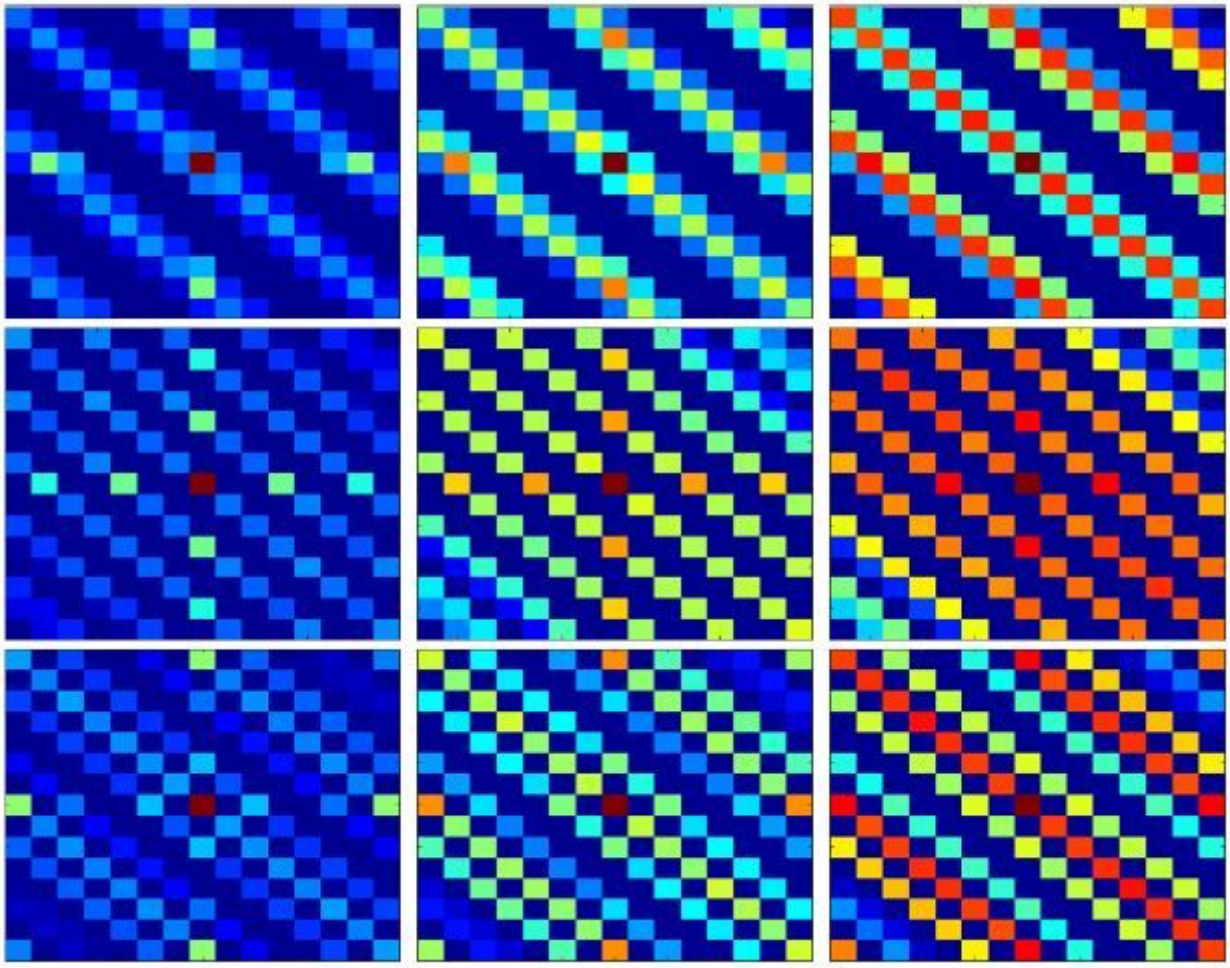

Figure 1

Image feature of matrix Refˆ RRg. From top to bottom, q = 20°, 40°, and 60°. From left to right, the SNR equals -5dB, 0dB, and 5dB.

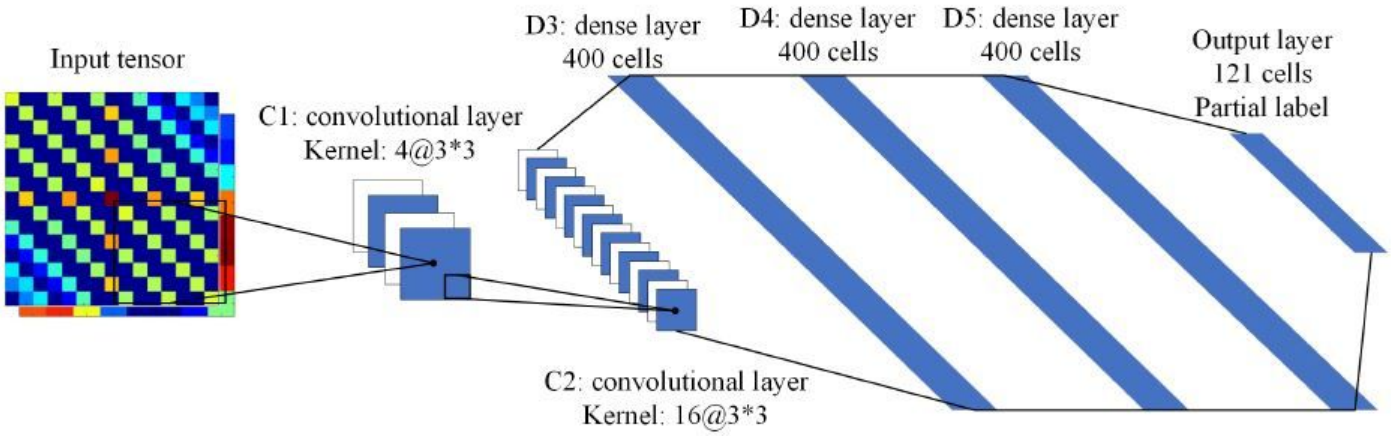

Figure 2

Framework of the proposed simple estimation CNN.

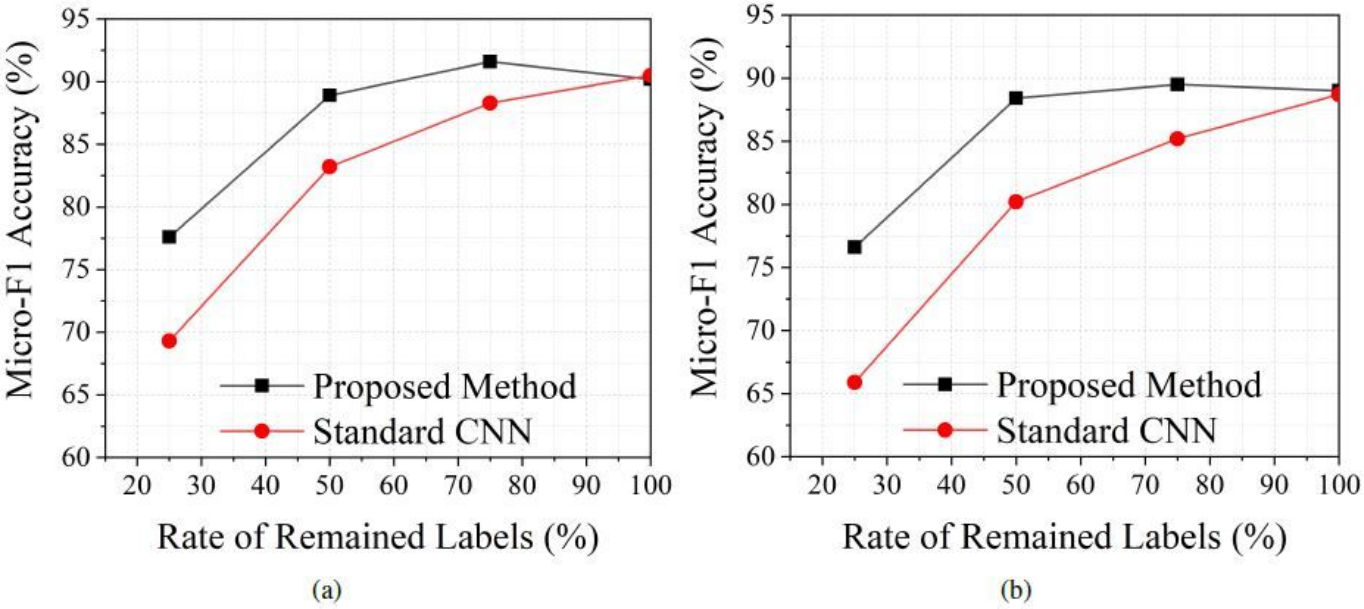

Figure 3

Micro-F1 accuracy comparison of training with PLS and standard training scheme. (a) training accuracy, (b) validation accuracy.

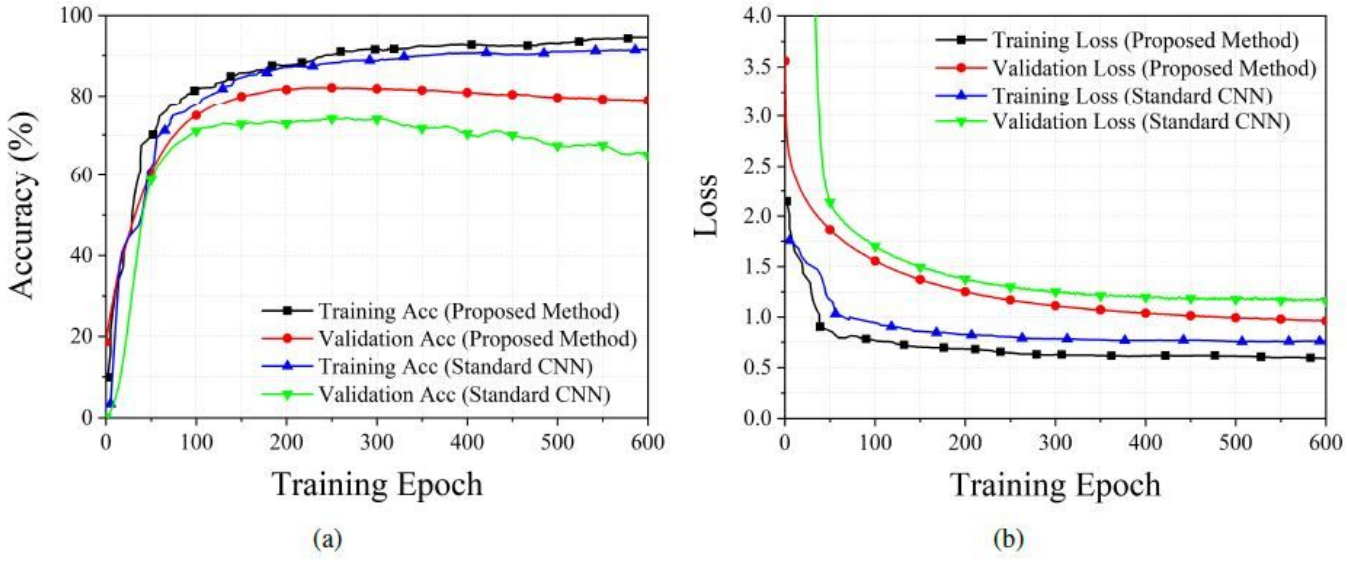

Figure 4

Training performance of the proposed method and standard training scheme. (a) training and validation accuracy, (b) training and validation loss.

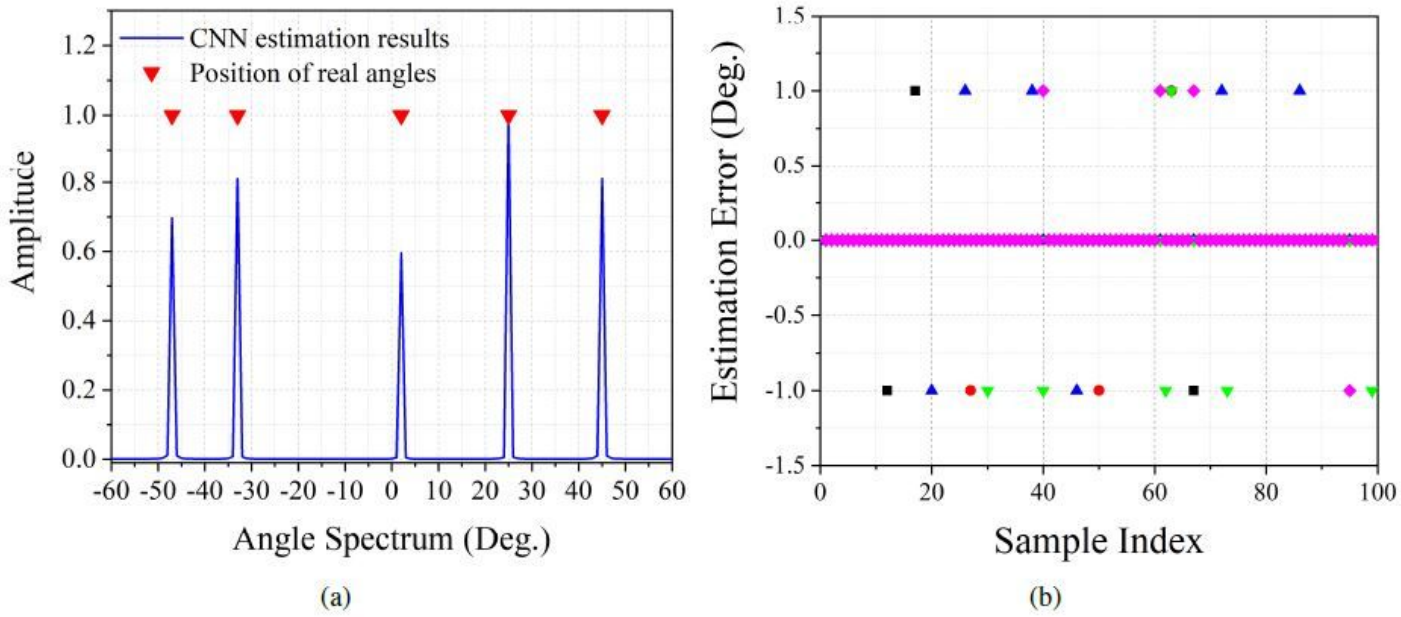

Figure 5

Estimation results of integer q. (a), angle spectrum obtained by the proposed estimation CNN, (b) estimation error for 100 Monte Carlo experiments.

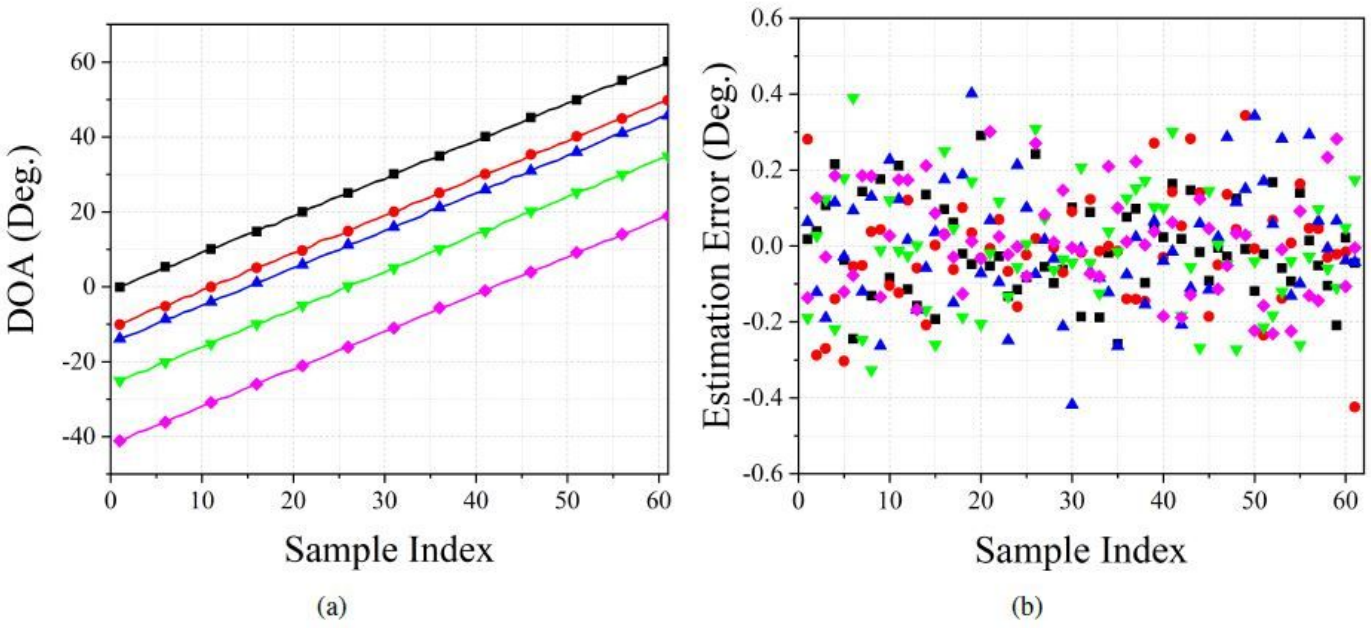

Figure 6

Estimation results of non-integer q. (a), estimation results of each sample, (b) estimation error of each sample.

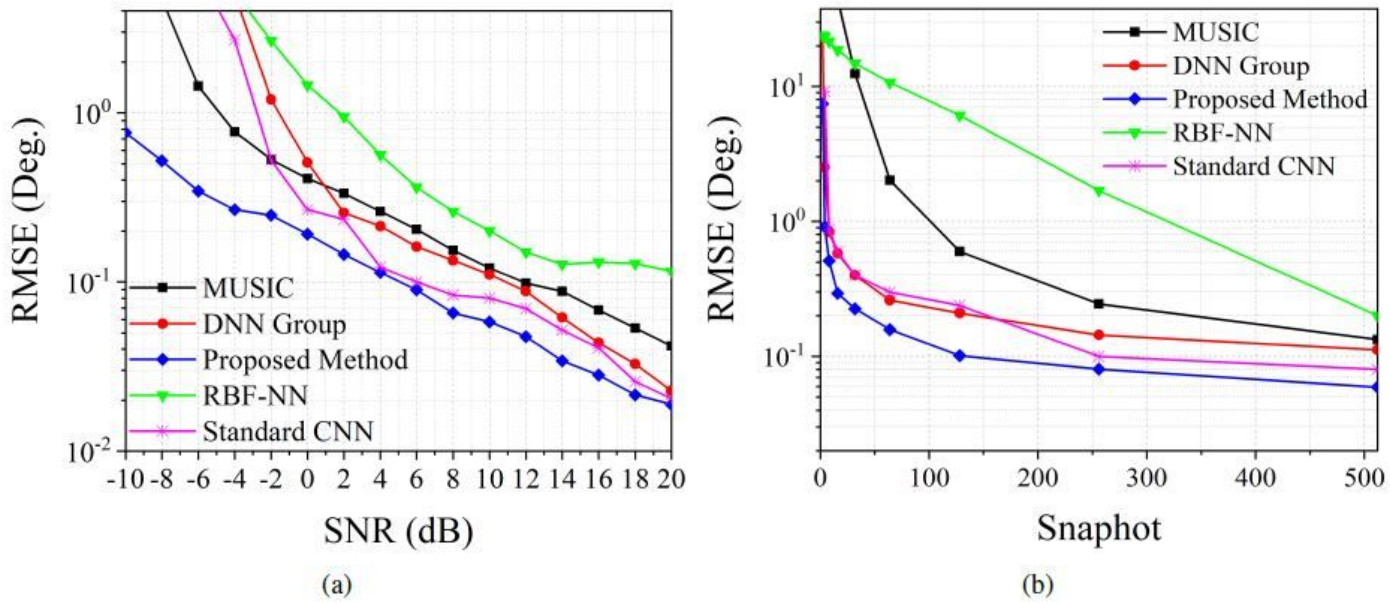

Figure 7

RMSE versus number of snapshots N, SNR equals to 10dB.

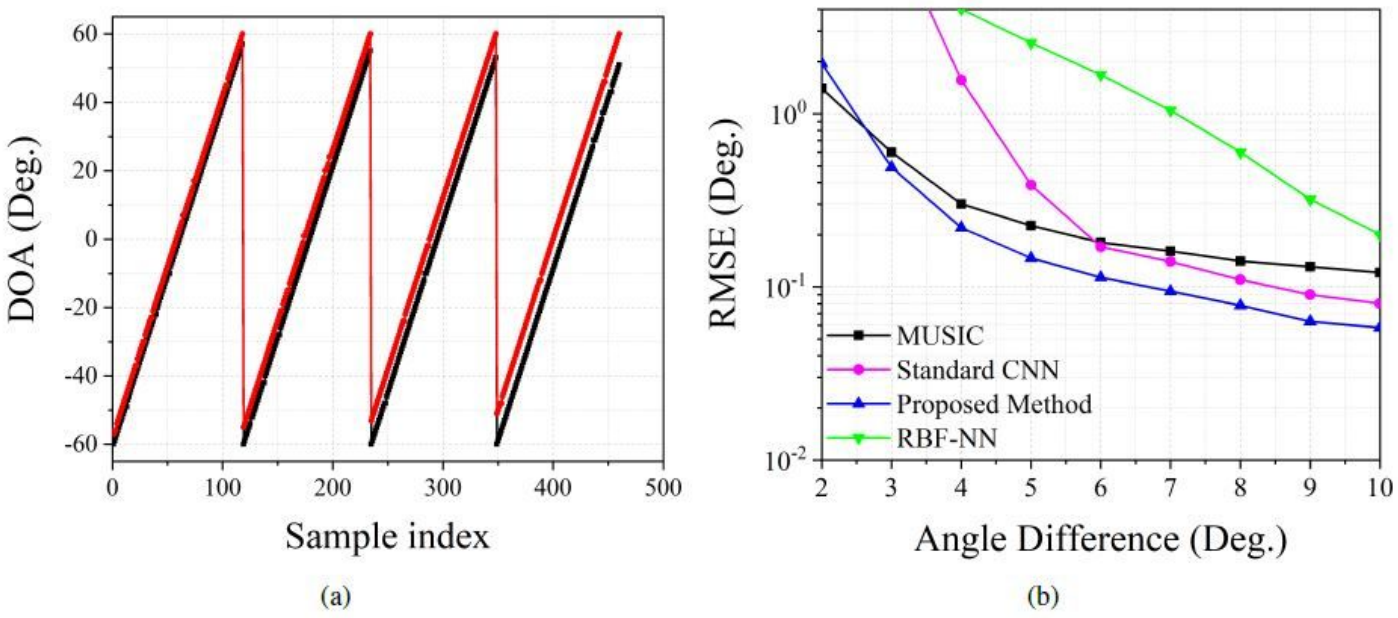

Figure 8

RMSE versus angle difference Dq, SNR equals to 10dB and number of snapshots N = 512.