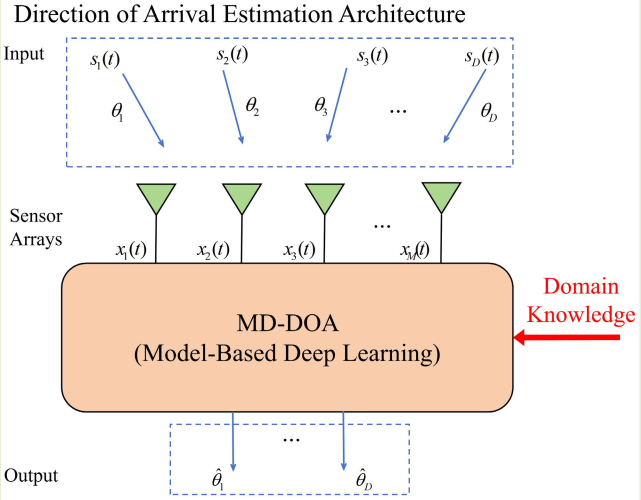
MD-DOA: A Model-Based Deep Learning DOA Estimation Architecture

Xiaoxuan Xu and Qinghua Huang



***Abstract*—Direction-of-arrival (DOA) estimation is widely used in the field of array signal processing. The model-based (MB) algorithms rely on domain knowledge and assumptions, facing limitations in estimating coherent sources and running on a few snapshots and so on. In contrast, deep learning approaches can learn from data, offering a promising alternative for DOA estimation. In this article, a novel end-to-end MB deep learning DOA estimation architecture (MD-DOA) is proposed to estimate the DOAs of multiple narrowband signals captured by a uniform linear array (ULA). Specifically, the multibranch convolutional recurrent neural network with a residual link (MBCR2net) is developed to extract multiscale features and learn correlation in received temporal signals. Subsequently, the weighted noise subspace network (WNSnet) is proposed to learn a more representative noise subspace from the one obtained by eigenvalue decomposition (EVD), developing the more precise subspace division. The matrix reshape process (MRP) then generates the pseudo covariance matrix (PCM) and captures the correlation in the weighted noise subspace. Notably, EVD and MRP are the MB modules to preserve the interpretability. Finally, the PCM-based DOA-finding network (PDFnet) estimates the desired DOAs. MD-DOA integrates the MB and data-driven (DD) advantages. It inherits the overall framework of the subspace-based methods while using the network to augment the covariance matrix estimation, subspace division, and peak-finding process. Our proposed architecture can operate successfully in the presence of array mismatch, low signal-to-noise ratios (SNRs), and a few snapshots. It is also applicable to real-world measurements and demonstrates superior performance compared with other existing algorithms in this field.**

***Index Terms*—Direction-of-arrival (DOA) estimation, end to end, model-based (MB) deep learning, multibranch, weighted noise subspace.**

I. Introduction

DIRECTION-OF-ARRIVAL (DOA) estimation aims to estimate the arrival angle of each signal relative to the array reference element. It has found extensive applications in diverse areas including hearing aids, autonomous driving, wireless communications, and underwater acoustic engineering [1], [2], [3], [4], [5], [6], [7]. Over the years, there have been considerable advancements in the DOA estimation algorithms.

A. Related Works and Motivation

Conventional beamforming (BF) can be seen as the application of Fourier-based spectral analysis for DOA estimation, but it has low angular resolution limited by the physical aperture [8]. The minimum variance distortionless response (MVDR) beamformer enhances the ability to suppress noise and improves the resolution of DOA estimation, but it is still confined by the Rayleigh limit [9]. On the other hand, the subspace-based approaches, commonly referred to as super-resolution algorithms, have been applied widely. In general, such methods can be divided into three steps, as shown in Fig. 1, i.e., the empirical covariance matrix is created from the input signals; then, the signal and noise subspaces are obtained after subspace decomposition; finally, the DOAs are estimated by using the subspace-based methods [10]. Essentially, the performance of subspace-based algorithms significantly depends on the estimation accuracy of the signal subspace or noise one. For the noise subspace-based algorithms, a partial noise subspace, which can be formed by a part of eigenvectors corresponding to small eigenvalues, was proposed [11]. It is more representative than the original noise subspace and obtains better performance. Therefore, how to select the eigenvectors and form the partial noise subspace is important. The multisignal classification (MUSIC) algorithm, which utilizes the noise

Manuscript received 1 April 2024; revised 29 April 2024; accepted 29 April 2024. Date of publication 8 May 2024; date of current version 14 June 2024. This work was supported by the National Natural Science Foundation of China under Grant 61571279. The associate editor coordinating the review of this article and approving it for publication was Prof. Seongwook Lee. *(Corresponding author: Qinghua Huang.)*

The authors are with the School of Communication and Information Technology, Shanghai University, Shanghai 200444, China (e-mail: xiaoxuan@shu.edu.cn; qinghua@shu.edu.cn).

Digital Object Identifier 10.1109/JSEN.2024.3396337

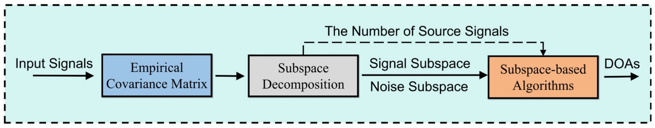

The Number of Source Signals
Input Signals
Empirical Covariance Matrix
Subspace Decomposition
Signal Subspace
Noise Subspace
Subspace-based Algorithms
DOAs

Fig. 1. Subspace-based methods for DOA estimation.

subspace to construct a spatial spectrum function, has gained popularity in the subspace-based algorithms [12]. It estimates the DOAs by finding the spectral peaks. However, the process is a time-consuming task, so the root-MUSIC algorithm was proposed [13]. The root-MUSIC algorithm offers a solution to simplify the peak-finding process by substituting the search for spectral peaks with the polynomial roots. All of the above methods discussed are model-based (MB), which rely on classical statistical model and additional domain knowledge [14]. As a result, they have some certain limitations, such as their inability to handle scenarios with a low signal-to-noise ratio (SNR) and to accurately estimate the DOAs for coherent signals [1].

With the development of deep learning, there has been a growing interest in the methods, which can learn from data [15], [16], [17]. These data-driven (DD) methods are able to capture nonlinear relationship between the input and output data, and they have been demonstrated as effective means for DOA estimation [18].

In most instances, deep learning techniques for DOA estimation are extensively employed as multilabel classification tasks. Different neural network architectures, such as convolutional recurrent neural networks (CRNNs) [19], U-Net [20], convolutional neural network (CNN) [21], and ResNets [22], were proposed to predict a probability grid of DOAs. Furthermore, the ResNet and multiscale convolutional techniques were fused to achieve multimodal DOA estimation by using audio and video data [23]. The methods mentioned above are classification-based models. However, they face the challenge due to the grid mismatch problem resulting from the discretization of spatial direction [24]. In response to this challenge, a set of regression-based techniques have been proposed for gridless DOA estimation, which are gaining popularity as effective solutions [25], [26], [27], [28].

Although neural networks can handle array imperfections without depending on specific models, they are highly parametrized [14], [29]. Consequently, they bring a substantial computational burden and encounter hardware limitations [30]. Furthermore, compared with the MB approaches, they are black-box and lack interpretability.

Recently, the hybrid MB deep learning architecture has been proposed for DOA estimation, which combines the benefits of MB and DD methods. The hybrid algorithms can inherit some structures of the MB approaches to provide interpretability, while dealing with model mismatch by learning from the data [14]. Notably, the emerging hybrid framework has garnered attention as a critical way for DOA estimation [31], [32]. In [33] and [34], CNNs and feedforward neural networks were employed to estimate MUSIC-like spectra and obtained DOAs through peak finding. However, the spectra were regarded as the training label in these techniques, and they were not trained to create DOAs directly. Shmuel et al. [35], [36] proposed trainable plug-and-play autoencoder DNN-aided model to learn how to divide the observations into distinguishable subspaces, and then, they employed the subspace-based techniques for DOA estimation. In addition, deep neural networks (DNNs) were used to estimate noiseless spatial covariance matrices (SCMs) [37], [38]. These matrices were then utilized for gridless DOA estimation with postprocessing steps, such as root-MUSIC. Although the networks were used to make the roots estimated by the root-MUSIC as close to the unit circle as possible, some errors still exist. The mentioned methods are MB deep learning algorithms [33], [34], [35], [36], [37], [38]. However, they are not designed to directly provide DOAs but are acquired through postprocessing using subspace-based methods, such as MUSIC or root-MUSIC. However, inherent problems in the subspace-based approaches still remain. For example, the time-consuming peak-finding process and the inability to ensure the location of the roots on the unit circle. Hence, DOAs can be better restored by directly obtaining through a network-based mapping. An end-to-end hybrid model recovered DOAs from the covariance matrix by combining a recurrent neural network (RNN) with additional DNNs [39], [40]. However, this model fails to make full use of neural networks to capture multiscale features and lacks feature fusion and reuse operations, resulting in poor model performance.

Remarkably, the problem of effectively learning angular features from signal sources by deep learning techniques is a topic of significant interest. In computer vision, researchers have introduced the concept of multiscale convolution to acquire features utilizing convolutional kernels with different sizes [41], [42], [43], [44], [45]. The residual connection operations were used in the networks to enhance the performance and overcome overfitting problems [46], [47], [48], [49], [50]. Meanwhile, the pyramid networks and attention mechanisms have been employed to improve the performance of the models [51], [52], [53]. The GoogLe2Net, integrating these ideas for the first time, was proposed for computer vision tasks [54]. However, how to better integrate these ideas in DOA estimation is still an open question. The DNN estimators in DOA estimation, such as those in [18] and [55], usually fail to capture higher level correlations hidden in the input signals as well as multiscale features.

To deal with the above problems, we aim to create a MB deep learning architecture to improve the performance of DOA estimation by merging feature extraction, subspace selection, and DOA-finding tasks while maintaining the model interpretability.

B. Contributions

In this article, a novel end-to-end regression architecture is proposed to estimate DOAs of multiple narrowband signals with a uniform linear array (ULA). First, a novel neural network is proposed to extract multiscale features from roughness to detail. Subsequently, the weighted noise subspace is estimated by the weighted selector based on the noise subspace generated by the eigenvalue decomposition (EVD), and the pseudo covariance matrix (PCM) is obtained by the matrix reshape process (MRP). Finally, MRP is fed into the DOA detector for the estimation. Extensive simulations demonstrate significant improvements in DOA estimation for both coherent and incoherent sources in the presence of array mismatch, low SNRs, and a few snapshots. The proposed architecture can be successfully applied to real-world measured data, and we also analyze the inference time of it. The main contributions of this work can be summarized as follows.

1) The multibranch convolutional recurrent neural network with a residual link (MBCR2net) is proposed to improve the ability to extract and utilize multiscale features and obtain the estimated covariance matrix. It first extracts multiscale features by convolutional kernels with different sizes and branches. Moreover, the information hidden in these multiscale features is exchanged by the feature re-extraction (FR) module. The outputs from a residual link (Reslink) and multiple branches are further fused. Finally, a gate recurrent unit (GRU) is added to capture long-term dependencies and learn correlation in received temporal signals.
2) The weighted noise subspace network (WNSnet), which takes the noise eigenvectors as inputs, is designed to learn weighted noise subspace, which is more representative and enhances the effectiveness of subspace division.
3) A PCM is obtained by MRP to capture correlations and rich statistical information in the weighted noise subspace. It acts as the input of the PCM-based DOA-finding network (PDFnet), which learned along with MBCR2net and WNSnet.

The remainder of this article is organized as follows. Section II explains the system model for DOA estimation in a ULA. In Section III, the proposed DOA estimation architecture is introduced in detail, including three main neural networks. Section IV shows the simulation results and discussions to demonstrate the performance of the proposed architecture. Finally, Section V gives the conclusion.

Notations: $||\cdot||$ represents the Euclidian norm. $[\cdot]^{\mathrm{H}}$ and $[\cdot]^{\mathrm{T}}$ denote the conjugate transpose and the transpose, respectively. $\mathrm{E}[\cdot]$ and $\mathrm{diag}\{\cdot\}$ mean the expectation and the diagonal operator, respectively. The symbol $\mathbb{C}$ represents the complex field, and $\mathbb{R}$ represents the real field. Throughout this article, lowercase, boldface lowercase, and boldface capital letters denote scalars, vectors, and matrices, e.g., x, $\mathbf{x}$, and $\mathbf{X}$, respectively.

II. Signal Model and Preliminaries

In this section, we describe the system model and preliminaries. To this end, the signal model is presented in Section II-A. The definition of the weighted noise subspace is then briefly explained in Section II-B, and the DOA estimation problem is given in Section II-C.

A. Signal Model

In this section, we describe the signal model to estimate DOAs of multiple narrowband signals with ULA, and more details can be found in [40], [56], and [57]. The array model used by the system is depicted in Fig. 2. Assume that D far-field narrowband sources from different directions impinge on a ULA. The ULA is composed of M antenna elements with a spacing of half a wavelength as d. Let $s_i(t)$ be the incident signal and θ_i be the incident angle. At each time instance t from T snapshots (i.e., $t \in \{1, \ldots, T\}$), the signal sources are represented as $\mathbf{s(t)} = [s_1(t), \ldots, s_D(t)]^{\mathrm{T}} \in \mathbb{C}^{D\times 1}$, and the corresponding DOAs are denoted as $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_D]$. The multidimensional observations collected by the ULA are expressed as $\mathbf{x(t)} = [x_1(t), \ldots, x_M(t)]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}$. Therefore, the signal model at t is

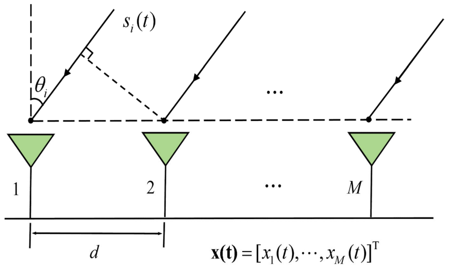


Fig. 2. Uniform linear array model.

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

where $\mathbf{n(t)} \in \mathbb{C}^{M\times 1}$ is the additive Gaussian white noise with the variance σ_{N}^2. The matrix $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_D)] \in \mathbb{C}^{M\times D}$ consists of the steering vectors $\mathbf{a}(\theta_i) \triangleq [1, e^{-j\pi \sin\theta_i}, \ldots, e^{-j\pi(M-1)\sin\theta_i}]^{\mathrm{T}}$, $(i \in \{1, \ldots, D\})$. The steering vector represents the response from the incident signal to each of the array elements.

When the received signal consists of T snapshots, referred to as $\mathbf{X} = [\mathbf{x(1)}, \ldots, \mathbf{x(T)}] \in \mathbb{C}^{M\times T}$, the model is commonly represented as

$$\mathbf{X} = \mathbf{A}\,(\boldsymbol{\theta})\,\mathbf{S} + \mathbf{N} \tag{2}$$

where $\mathbf{S} = [\mathbf{s(1)}, \ldots, \mathbf{s(T)}] \in \mathbb{C}^{D\times T}$ is the signal matrix, while the noise matrix is represented by $\mathbf{N} \in \mathbb{C}^{M\times T}$. According to the model, the covariance matrix of the received signals can be expressed as

$$\mathbf{R}_{\mathbf{X}} = \mathrm{E}\left[\mathbf{X}\mathbf{X}^{\mathrm{H}}\right] = \mathbf{A}\,(\boldsymbol{\theta})\,\mathbf{R}_{\mathbf{S}}\mathbf{A}^{\mathrm{H}}\,(\boldsymbol{\theta}) + \sigma_{\mathrm{N}}^2\mathbf{I}_M \tag{3}$$

where $\mathbf{R}_{\mathbf{S}} = \mathrm{E}[\mathbf{S}\mathbf{S}^{\mathrm{H}}]$ is the covariance matrix of the signal sources.

B. Weighted Noise Subspace

For traditional methods, the EVD of the covariance matrix from the received signals is used to obtain both the signal subspace and the noise subspace [11], [56], i.e.,

$$\mathbf{R}_{\mathbf{X}} = \mathbf{U}_{\mathbf{S}}\boldsymbol{\Lambda}_{\mathbf{S}}\mathbf{U}_{\mathbf{S}}^{\mathrm{H}} + \sigma_{\mathrm{N}}^2\mathbf{U}_{\mathbf{N}}\mathbf{U}_{\mathbf{N}}^{\mathrm{H}} \tag{4}$$

where $\boldsymbol{\Lambda}_{\mathbf{S}}$ is a diagonal matrix with its diagonal elements denoting the D largest eigenvalues. The signal subspace $\mathbf{U}_{\mathbf{S}}$ consists of the eigenvectors that correspond to the D largest eigenvalues, while the noise subspace $\mathbf{U}_{\mathbf{N}}$ consists of the eigenvectors corresponding to the other $(M–D)$ eigenvalues. The two subspaces are orthogonal to each other. Let $\mathbf{u}_i$

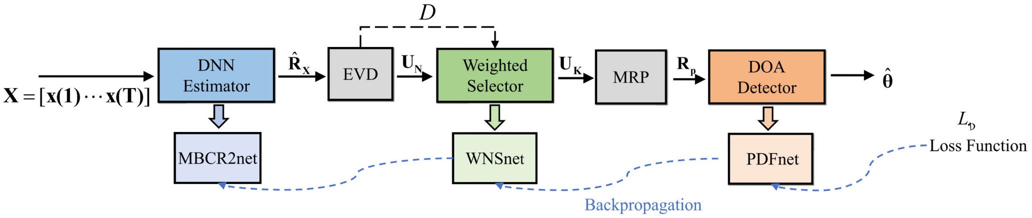

D
$\mathbf{X} = [\mathbf{x}(1) \cdots \mathbf{x}(T)]$
DNN Estimator
$\hat{\mathbf{R}}_{\mathbf{X}}$
EVD
$\mathbf{U}_{\mathrm{N}}$
Weighted Selector
$\mathbf{U}_{\mathrm{K}}$
MRP
$\mathbf{R}_{\mathrm{P}}$
DOA Detector
$\hat{\boldsymbol{\theta}}$
MBCR2net
WNSnet
PDFnet
$L_{\mathcal{D}}$
Loss Function
Backpropagation

Fig. 3. Architecture of the proposed MD-DOA.

($i = 1, \ldots, M$) denote the eigenvectors of $\mathbf{R}_{\mathbf{X}}$. Then, the signal and noise subspaces are defined as $\mathbf{U}_{\mathbf{S}} = [\mathbf{u}_1, \ldots, \mathbf{u}_D]$ and $\mathbf{U}_{\mathbf{N}} = [\mathbf{u}_{D+1}, \ldots, \mathbf{u}_M]$.

Remarkably, a weighted noise subspace, consisting of a subset of eigenvectors corresponding to the smaller $(M-D)$ eigenvalues, performs better than $\mathbf{U}_{\mathbf{N}}$ for DOA estimation [11]. It can be obtained by a Boolean vector defined as $\mathbf{h} = [, \ldots, 0, \ldots, 1, \ldots,] \in \mathbb{R}^{1\times(M-D)}$, where the number of 1 is denoted as K. Subsequently, the weighted matrix $\mathbf{W}$ is formed as $\mathbf{W} = \mathrm{diag}\{\mathbf{h}\}$. The weighted noise subspace formed mathematically can be denoted as $\mathbf{U}_{\mathbf{K}} = \mathbf{U}_{\mathbf{N}}\mathbf{W}$, with the matrix $\mathbf{W}$ providing the weighted operation on the noise eigenvectors.

C. Problem Formulation

We aim to leverage the available domain knowledge and data to develop an architecture that accurately determines the DOAs based on the received signals, finding the mapping between the received signal and DOAs. The dataset consists of L pairs of received signals and their corresponding DOAs, i.e., $\mathcal{D} = \{(\mathbf{X}_1, \boldsymbol{\theta}_1), (\mathbf{X}_2, \boldsymbol{\theta}_2), \ldots, (\mathbf{X}_l, \boldsymbol{\theta}_l), \ldots, (\mathbf{X}_L, \boldsymbol{\theta}_L)\}$.

The proposed architecture can exploit the neural networks to reliably estimate the DOAs and overcome the limitations of the MB methods, which performs poorly in the presence of array mismatch, low SNRs, and limited snapshots.

III. Proposed Architecture for DOA Estimation

In this section, the architecture called MB deep learning DOA estimation architecture (MD-DOA) is proposed, which is a hybrid DOA estimation architecture, as shown in Fig. 3. It consists of three parts: the MBCR2net as the DNN estimator, the WNSnet for weighted selector, and the PDFnet as the DOA detector. The gray blocks represent the MB modules.

The MBCR2net is employed to extract features and gets the estimated covariance matrix. Subsequently, the weighted selector forms the weighted noise subspace from $\mathbf{U}_{\mathbf{N}}$ obtained by EVD. Finally, DOA is estimated by the DOA detector. In the following, Section III-A describes the architecture of MBCR2net, while Sections III-B and III-C give the weighted selector and the DOA detector, respectively. The subsequent Section III-D concludes the training strategy and loss function.

A. DNN Estimator

In order to fuse the information more efficiently and acquire multiscale features, a DNN estimator called MBCR2net is proposed for MD-DOA, as illustrated in Fig. 4(a). The MBCR2net combines the Reslink, multiscale feature extraction, and feature reuse, which plays a central role in capturing the intricate nonlinear relationship between the signals and the covariance matrix.

As a preprocessing step, given that the observed signals are in the form of complex values, the real and imaginary parts of $\mathbf{X}$ are stacked to create a real value input denoted as $\mathbf{Y} \in \mathbb{R}^{T\times 2M}$. Furthermore, to ensure the match of feature channels between the branches of convolutional layers and the Reslink channels to the final output, a convolutional block consisting of the convolutional layer, the maximal pooling, the batch normalization, and ReLU [58] activation function is employed. This structure is symbolically represented as $\mathbf{C}_i$, i.e.,

$$\mathbf{C}_i\,(\alpha) = \mathrm{ReLU}\,(\mathrm{BN}\,(\mathrm{MaxPool}\,(\mathrm{Conv}_{i\times i}\,(\alpha)))) \tag{5}$$

where let α be the input features of the convolutional block, and $\mathrm{Conv}_{i\times i}$ denotes the convolutional layer with a kernel size of $i \times i$. $\mathrm{BN}(\cdot)$, $\mathrm{ReLU}(\cdot)$, and $\mathrm{MaxPool}(\cdot)$ denote the batch normalization [59], activation function, and the maximal pooling [60], respectively. MBCR2net exploits the convolutional blocks with different sizes to extract multiscale features and enhance network performance. The structure uses common convolutional kernel sizes, such as 1×1, 3×3, and 5×5, so that the network can extract input features from coarse to fine. For each branch of the network, the convolutional block, which the kernel size of convolutional layer is set to 1×1, allows the model to capture stronger nonlinear features within the same receptive field.

Moreover, to improve the performance, we have introduced the FR module between different branches. As illustrated in Fig. 4(b), the FR module adopts the attention mechanism on the input feature layers to pay more attention to the useful information [47], [61]. The information thus processed can flow to the next branch, which further enhances the information extraction capability. Meanwhile, the FR module also realizes cross-branch information reuse. Denote the output

$$\begin{cases} f_1\,(\mathbf{y}_i) = \sigma\,(\mathrm{BN}\,(\mathrm{Conv}_{2\times 2}\,(\mathrm{Concat}\,(\mathrm{AvgPool}\,(\mathbf{y}_i)\,,\,\mathrm{AvgPool}\,(\mathbf{y}_i))))) \\ f_2\,(\mathbf{y}_i) = \sigma\,(\mathrm{Conv}_{1\times 1}\,(f_1\,(\mathbf{y}_i))) \\ \mathrm{FR}\,(\mathbf{y}_i) = \mathrm{Multi}\,(\mathbf{y}_i,\, f_2\,(\mathbf{y}_i)\,,\, f_2\,(\mathbf{y}_i)) \end{cases} \tag{6}$$

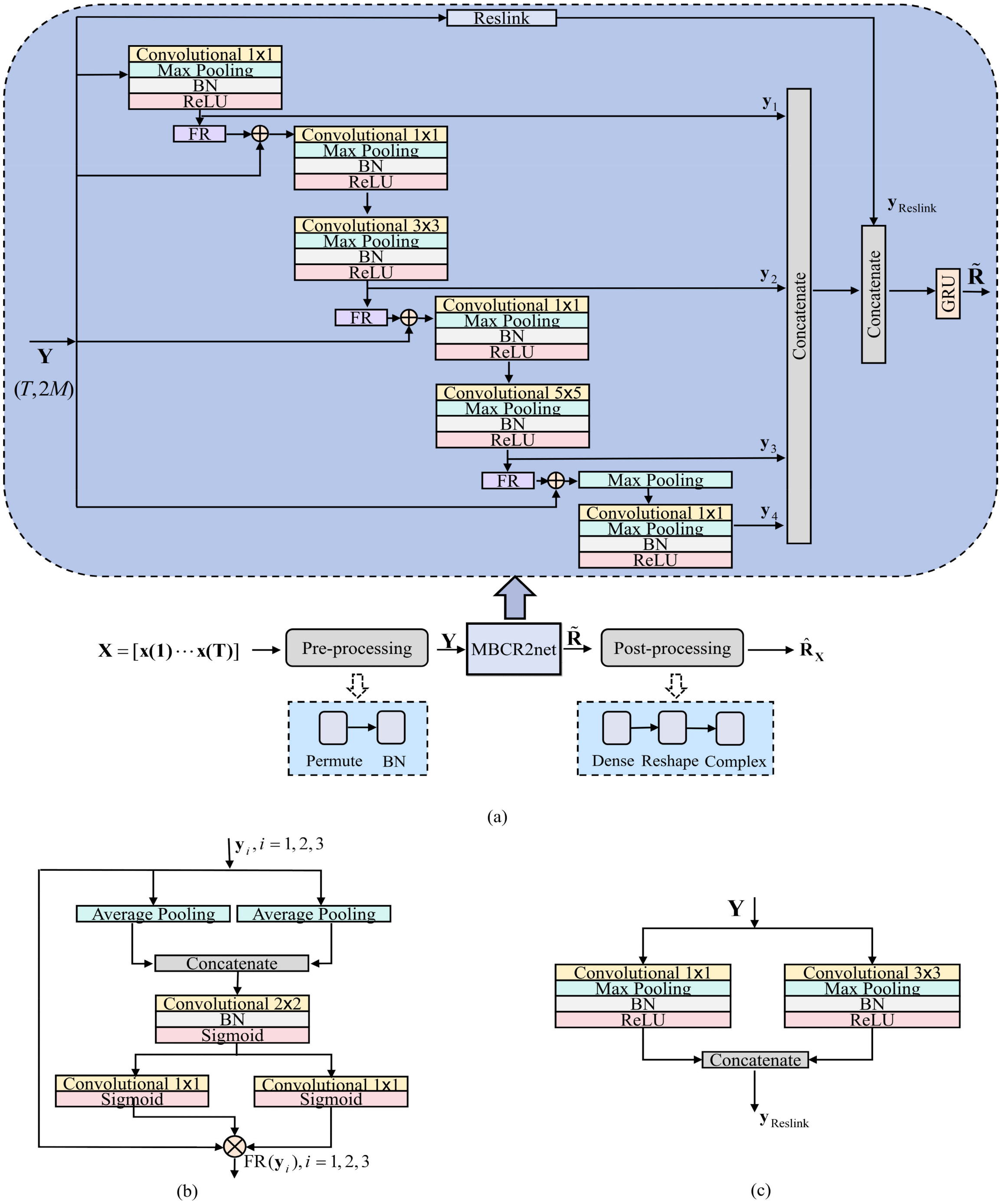


Fig. 4. Illustration of different modules in the DNN estimator. (a) Structure of the proposed MBCR2net for DNN estimator. (b) FR module from MBCR2net. (c) Reslink module from MBCR2net.

features of each branch as $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$, then the FR module can be expressed as in (6), shown at the bottom of the previous page, where $i = 1$–3. $\sigma(\cdot)$, AvgPool$(\cdot)$, Multi$(\cdot)$, and Concat$(\cdot)$ denote the sigmoid function, the average pooling [60], the multiply operation, and the concatenation operation, respectively. The FR module uses the attention mechanism to obtain more valuable features for transmission to the subsequent branch, thereby enabling cross-branch information reuse and further enhancing performance.

Furthermore, a Reslink is added into the proposed network to help reduce the difficulty of network optimization, which is presented in Fig. 4(c). The Reslink employs two convolutional blocks to establish the Reslink. The feature $\mathbf{y}_{\text{Reslink}} \in \mathbb{R}^{2T \times 2M}$ from Reslink module can be expressed as

$$\mathbf{y}_{\text{Reslink}} = \text{Reslink}\left(\mathbf{Y}\right) = \text{Concat}\left(\mathbf{C}_1\left(\mathbf{Y}\right), \mathbf{C}_3\left(\mathbf{Y}\right)\right) \qquad (7)$$

where $\mathbf{Y}$ with $T \times 2M$ is used as the input of MBCR2net. The above operations exploit the translation invariance of CNNs to extract high-level features of the signals. However, considering that the input signals contain temporal information, FR module presents a limitation in capturing long-distance dependencies, as it requires the computation of attention weights for the entire feature map. To solve this problem, we add a GRU layer to merge the temporal information, which helps to learn correlation in received temporal signals.

$\mathbf{Y}$ is transformed into a real matrix $\tilde{\mathbf{R}} \in \mathbb{R}^{2M\times M}$ by MBCR2net, and the continuous nonlinear transformation process can be defined as

$$\begin{cases} \tilde{\mathbf{R}} = \mathrm{GRU}\left(\mathrm{Concat}\left(\mathbf{y}_{\mathrm{Reslink}}, \mathrm{Concat}\left(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_4\right)\right)\right) \\ \mathbf{y}_1 = \mathbf{C}_1\left(\mathbf{Y}\right) \\ \mathbf{y}_2 = \mathbf{C}_3\left(\mathbf{C}_1\left(\mathrm{FR}\left(\mathbf{y}_1\right) + \mathbf{Y}\right)\right) \\ \mathbf{y}_3 = \mathbf{C}_5\left(\mathbf{C}_1\left(\mathrm{FR}\left(\mathbf{y}_2\right) + \mathbf{Y}\right)\right) \\ \mathbf{y}_4 = \mathbf{C}_1\left(\mathrm{MaxPool}\left(\mathrm{FR}\left(\mathbf{y}_3\right) + \mathbf{Y}\right)\right) \end{cases} \tag{8}$$

where $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{y}_3$, and $\mathbf{y}_4$ are the outputs of the four branches, respectively, and GRU$(\cdot)$ denotes the gated recurrent unit layer [62]. Besides, the addition operator is crucial in DNN estimator, and information reusing in FR modules of different branches is mainly realized through the addition operation, and the addition operator merges the features previously obtained through the preprocessing step with the FR-processed features. The output $\tilde{\mathbf{R}}$ is learned from data and acts similar to the covariance matrix of MB methods. As a postprocessing step, $\tilde{\mathbf{R}}$ is reshaped into a complex matrix to obtain the estimated covariance matrix $\hat{\mathbf{R}}_{\mathbf{X}} \in \mathbb{C}^{M\times M}$. Furthermore, the MBCR2net enhances the ability to extract features and realizes cross-branch information reuse.

B. Weighted Selector

The effectiveness of the weighted noise subspace generally surpasses that of the conventional noise subspace, as supported by previous research [11]. The conventional noise subspace $\mathbf{U}_{\mathbf{N}}$ is obtained by using EVD for $\hat{\mathbf{R}}_{\mathbf{X}}$. However, in scenarios with low SNRs and other complicated factors, it is difficult to distinguish between the signal and noise eigenvectors. When the number of sources is unknown, a mapping of the eigenvalues to the set of probabilities containing the selection of the corresponding eigenvectors as noise eigenvectors is established, allowing to learn a suitable noise subspace [40]. However, the weighting of all eigenvalues by this method may result in signal eigenvectors mixed in the noise subspace, leading to inaccurate noise subspace division. Therefore, noise eigenvectors are adopted as the network inputs, which is a better choice. In order to promote the performance of DOA estimation, the WNSnet is used to form weighted noise subspace, as shown in Fig. 5. A multilayer perceptron (MLP) network is used to achieve the clustering process and obtain a probability set denoted as $\mathbb{p}$. As an improvement, the mean-based self-attention module (MSAM), which automatically learns the weights of each noise eigenvector, is added to obtain a set of thresholds as $\mathbb{z}$. These thresholds are then compared with the probability values in the set $\mathbb{p}$ to generate an indicator $\mathbf{h}$. The indicator represents a Boolean vector and is used to create the weighted diagonal matrix $\mathbf{W}$. Finally, $\mathbf{W}$ is utilized to obtain the weighted noise subspace as $\mathbf{U}_{\mathbf{K}}$. Overall, weighted selector offers an effective method to choose a weighted noise subspace, enhancing flexibility and robustness in the process of forming the weighted noise subspace.

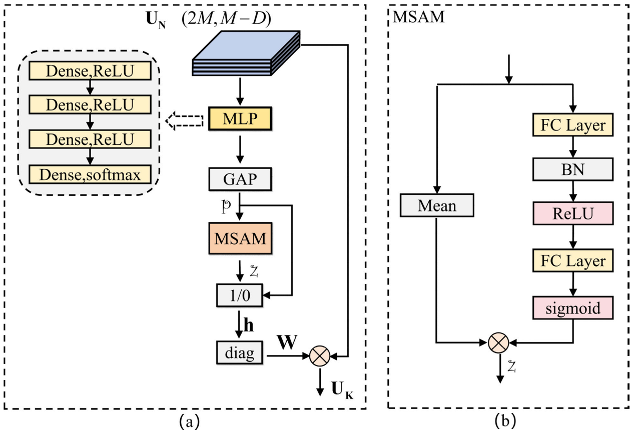


Fig. 5. Illustration of different structures in the modules to determine the weighted noise subspace. (a) Proposed WNSnet. (b) MSAM from WNSnet.

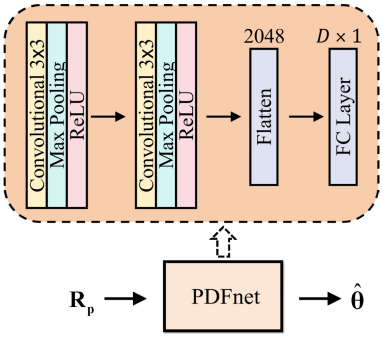


Fig. 6. Structure of the DOA detector.

C. DOA Detector

When the weighted noise subspace is inputted to the DOA detector, the MRP is required to create the PCM. Define the PCM as $\mathbf{R}_{\mathbf{p}} = \mathbf{U}_{\mathbf{N}}\mathbf{U}_{\mathbf{N}}^{\mathrm{H}}$, which has a similar mathematical form to the signal covariance matrix. Remarkably, the MRP module captures correlations and rich statistical information in the weighted noise subspace. Simultaneously, it improves the understanding of complex relationships, thereby enhancing the learning capability of the model.

To overcome the limitations of MB approaches, PDFnet is designed as a deep learning-aided DOA detector for the end-to-end regression learning to replace the postprocessing methods, and it is shown in Fig. 6. PDFnet is based on a CNN, which obtains DOAs from the PCM. The CNN contains two convolutional layers with a kernel size of 3, as well as a ReLU activation function, and we use the dense layer of linear activation to get $\hat{\boldsymbol{\theta}}$. Due to the fact that it is able to learn from data, the PDFnet can be an incredibly valuable tool for end-to-end MD-DOA architecture training. Furthermore, it can learn jointly with MBCR2net and weighted selector for promoting DOA restoration.

D. Training Strategy and Loss Function

DOA estimation is modeled as a multiple regression task, trained in a supervised setting. The training process uses a small-batch gradient descent algorithm, specifically using the Adam optimizer for the network optimization [63]. Simultaneously, PDFnet can be learned along with MBCR2net and WNSnet, sharing a loss function that enables gradient-based end-to-end optimization across the MD-DOA architecture and enhancing the learning capability of the neural network. The differentiability of this architecture enables the computation of the gradient of $\hat{\boldsymbol{\theta}}$ with respect to $\hat{\mathbf{R}}_{\mathbf{X}}$ and backpropagation throughout the entire structure. As described in Section II-C, our training set consists of the dataset $\mathfrak{D}$, which each sample contains the signals and the corresponding DOA.

To assess the performance, the loss function is computed by comparing the DOA recovered from the MD-DOA architecture with the true DOA. In regression problems, it is customary to employ root mean square error (RMSE) and mean square error (MSE) as loss functions. Nevertheless, the calculation of the difference between angles may not accurately reflect the true error due to the periodic nature of the DOA. Therefore, we use the root mean square periodic error (RMSPE) loss to overcome the periodicity problem by calculating the RMSE of the elementwise error over the periodic π range [64]. In order to compare all possible combinations of predicted and true angles, the replacement invariant matrix $\mathbf{P}$ is introduced. So, the order of DOA estimation has no impact on its performance. The loss function is denoted as [35], [36], [39], [40]

$$\mathrm{RMSPE}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\right) = \min_{\mathbf{P}} \left(\frac{1}{D} \left\| \mathrm{mod}_{\pi} \left(\boldsymbol{\theta} - \mathbf{P}\hat{\boldsymbol{\theta}} \right) \right\|^{2} \right)^{\frac{1}{2}} \tag{9}$$

where mod_{π} denotes the modulo operation with the angle range of π. The loss over the entire dataset $\mathfrak{D}$ is then denoted as

$$L_{\mathfrak{D}} = \frac{1}{L} \sum_{l=1}^{L} \mathrm{RMSPE}\left(\boldsymbol{\theta}_l, \hat{\boldsymbol{\theta}}_l\right). \tag{10}$$

IV. Simulation

In this section, the performance of the proposed MD-DOA architecture in different cases is evaluated to verify the validity of the proposed method. Initially, the experimental simulation settings are outlined in Section IV-A. In Section IV-B, to evaluate the effectiveness of our modules, a series of ablation experiments are conducted to determine the proposed architecture. In addition, we verify the DOA estimation performance of the proposed architecture in Section IV-C. Finally, in Section IV-D, the feasibility and inference time of the MD-DOA architecture are analyzed with respect to the real-world measured data.

A. Simulation Settings

Based on the signal model in Section II-A, we conduct a simulation to measure $\mathbf{x(t)}$. The simulation involves the ULA with $M = 8$ half-wavelength spaced array elements. The ULA is impinged by D signals, and the DOA for each signal is randomly generated from a uniform distribution $\mathbf{U}(-\pi/2, \pi/2)$. For all t, both the signal source $\mathbf{s(t)}$ and the noise signal $\mathbf{n(t)}$ are randomly drawn from a complex Gaussian distribution $\mathbf{CN}(0,1)$. Besides, we create synthetic datasets by simulations in different scenarios for facilitating the evaluation. In the coherent case, all signals have the same amplitude and phase. Unless otherwise noted, the model parameters are based on the settings specified in Table I.

TABLE I
Simulation Parameters

Symbol	Parameter Description	Value
M	Number of array elements	8
SNR	Signal-to-noise ratio	10 dB
batch	Batch size	16
r	Learning rate	10^{-3}

We employ an Adam optimizer with a learning rate of 0.001 and a batch size of 16 to update the model parameters during the training process. The dataset is divided into a training set (90%) and a validation set (10%) consistent with the contrasting approaches [40]. All the experiments are implemented based on Win11 by Keras, using Tensorflow as the backend.

B. Ablation Experiments

In this section, the effectiveness of our designs is demonstrated by ablation experiments. The aim is to assess whether the integration of the proposed modules in the MD-DOA architecture improves the performance of DOA estimation. To evaluate the impact of each module on the performance of the proposed method, the ablation experiments are split into three parts. First, we compare the different DOA detectors. Next, the influence of the DNN estimator is validated to ensure that key features are extracted from the input signals to support subsequent analysis and estimation. Finally, the experiments are conducted to verify the effectiveness of the weighted selector.

1) Comparison of the DOA Detectors: The final goal of DOA estimation is to get accurate localization of the signal sources. The DOA detector often plays a decisive role in the success of DOA estimation. Therefore, the first task is to establish the structure for DOA detector. As discussed in Section I-A, the postprocessing steps by subspace-based methods like MUSIC or root-MUSIC still suffer from the inherent issues, including the time-consuming peak-finding process and the inability to guarantee the location of roots on the unit circle. Thus, we choose to achieve the DOA recovery by establishing a mapping relationship to the angle through the network, as mentioned in the Augmented MUSIC algorithm [39]. To deeply validate the impact of different DOA detectors on the estimation performance, a series of comparison experiments were conducted using different DOA detectors based on the augmented MUSIC algorithm. Additionally, the effectiveness of the MRP module is experimentally validated. All the results are shown in Fig. 7. In Fig. 7, the abbreviation "MLP" represents the DOA detector structure utilized in the augmented MUSIC algorithm, while "PDFnet" denotes our

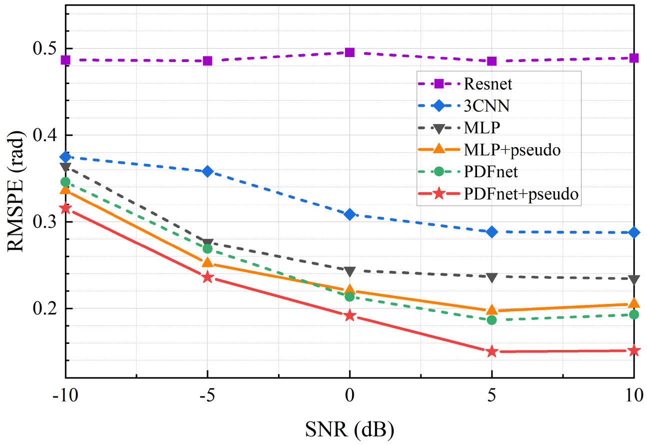


Fig. 7. Comparison of the different DOA detectors.

TABLE II
Combinations of Convolutional Kernel Sizes

Combinations of Convolutional Kernel Sizes	RMSPE [rad]
3, 3	**0.1513**
5, 5	0.1738
3, 5	0.1879
5, 3	0.1850
7, 7	0.1798

TABLE III
Effect of the Activation Function

Activation Functions	RMSPE [rad]
ReLU	**0.1513**
GeLU	0.1892
LeakyReLU	0.1882
PReLU	0.1916
SeLU	0.2198

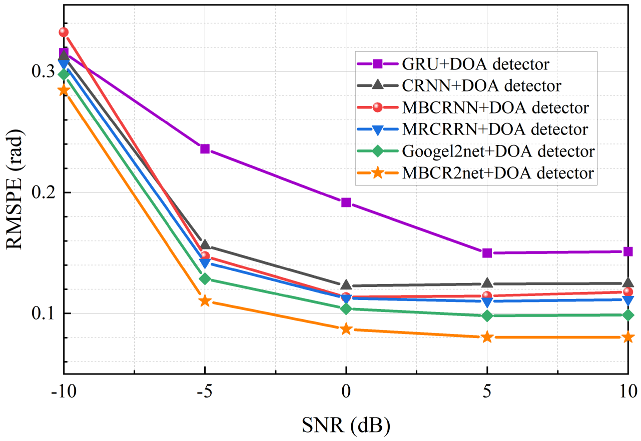


Fig. 8. Influence of the DNN estimator.

proposed network. “3CNN” refers to the network structure consisting of three convolutional layers and three max pooling layers. “Resnet” utilizes a single basic Resnet block. Methods tagged with “pseudo” introduce the MRP module. We train the algorithms by using a dataset containing a mixture of SNRs ranging from -10 to 10 dB when localizing $D = 5$ narrowband coherent signal sources with $T = 200$ snapshots.

According to Fig. 7, it is apparent that the complex neural network structure does not necessarily improve performance, as seen in the cases of Resnet and 3CNN, where performance even declined. Meanwhile, the PCM significantly reduces the RMSPE values of MLP and PDFnet, signifying an enhancement in the learning capability and performance of the DOA estimation. Finally, after experimental validation, PDFnet is identified as the optimal structure for the DOA detector. The structure effectively combines the advantages of CNN and the PCM, specifically translation invariance and correlation information extraction, resulting in a more dependable DOA estimation.

We also conducted ablation experiments on the combinations of convolutional kernel sizes and the type of activation functions to determine the optimal parameters for the PDFnet, as detailed in Tables II and III. There are five coherent sources, 200 snapshots, and SNR $=$ 10 dB. The best results are indicated in bold in the tables.

In Table II, “3, 3” means that the kernel size of both the first convolutional layer and the second one is 3. The other numbers also indicate the same meaning. The results show that “3, 3” and ReLU activation function are the optimal parameters for the PDFnet.

2) Influence of the DNN Estimator: As highlighted in Section I-A, the DNN estimator plays a vital role in DOA estimation. An effective DNN estimator has the ability to acquire multiscale features from the received signals and fuse the information more efficiently. To deeply study the impact of different DNN estimators on DOA estimation, we conduct the experiments when localizing $D = 5$ narrowband coherent source signals with $T = 200$ snapshots and SNRs varying in the range of -10 to 10 dB, as shown in Fig. 8. Our proposed network is MBCR2net, and Google2net denotes the DNN estimator mentioned in the work [54]. GRU represents the DNN estimator in the Augmented MUSIC algorithm [39]. CRNN, MBCRNN, and MBCRRN denote the CRNN, the multibranch CRNN, and the multibranch network with branches composed of residual blocks, respectively. They represent the gradual introduction of convolutional operations, multibranch operations, and residual blocks to the DNN estimator of the Augmented MUSIC algorithm.

Based on the results in Fig. 8, it is evident that the RMSPE values of CRNN, MBCRNN, and MBCRRN are lower compared with that of GRU. We can conclude that the introduction of convolution, multibranch operations, and residual blocks in the DNN estimator of the Augmented MUSIC algorithm contributes to the enhanced performance of DOA estimation. The results further validate the importance of these operations in DNN estimator, which are able to generate more high-level feature representations while effectively preserving signal information. In our experiments, it is noticed that MBCR2net performed the best as a DNN estimator, relying on the excellent performance in extracting and reusing multiscale features from the received signals. Based on the experimental results, MBCR2net is chosen as our DNN estimator.

TABLE IV
INFLUENCE OF THE FR MODULE

Method	RMSPE [rad]
FR	**0.0856**
$\mathbf{C}_1$	0.1038
CG	0.0949
MCG	0.0985
CBAM	0.1047

TABLE V
INFLUENCE OF THE RESLINK MODULE

Method	RMSPE [rad]
FR+Reslink	**0.0834**
FR+$\mathbf{C}_1$	0.0856
FR+HCSM	0.0879
FR+BiGRU	0.0918
FR+Light-Attention	0.1117

In MBCR2net, the FR and Reslink modules are added to enhance the feature reuse and help reduce the difficulty of network optimization. To validate the effectiveness of these modules, the experiments are divided into two parts. For FR module, we compare four methods: $\mathbf{C}_1$, FR, CBAM [65], channelwise gate (CG), and multigroup CG (MCG) [47]. These experiments aim to assess the effect of different feature reuse mechanisms, particularly in terms of feature refinement and cross-branch information transfer within the network. In the case of the Reslink module, we compare the performance of four methods: $\mathbf{C}_1$, Reslink, hybrid contextual semantic module (HCSM) [66], and bidirectional gated recurrent unit (BiGRU) [67]. The "Light-Attention" refers to the lightweight attention in [68]. These experiments aim to explore the effects of different Reslinks on network optimization. The results are detailed in Tables IV and V, using a dataset containing five coherent sources, 200 snapshots, and SNR = 0 dB.

The data presented in Table IV demonstrate that the proposed FR module exhibits superior performance compared with other methods, exhibiting a substantial 17.53% improvement over $\mathbf{C}_1$. Similarly, Table V shows the effectiveness of the Reslink module, and there is a 2.57% performance enhancement over $\mathbf{C}_1$. These results suggest that the FR and Reslink modules demonstrate significant enhancements in the performance of MBCR2net, surpassing other approaches for feature reuse and Reslinks.

3) Effectiveness of the Weighted Selector: In order to get a more accurate division of the noise subspace, we introduce a weighted selector to get the weighted noise subspace. In this part, the effectiveness of the weighted selector (referred to as "WS") has been verified, and the results are depicted in Figs. 9 and 10. First, the RMSPE of the weighted selector versus varying number of signal sources is shown in Fig. 9. Second, the corresponding dimensions of the weighted noise subspace are given in Fig. 10.

For the simulations, the parameters are set as follows: $T = 200$ and SNR = 10 dB. To conveniently demonstrate the change in the dimension of the weighted noise subspace, the number of array elements is kept 12, and the number of coherent narrowband signal sources varies from 3 to 7. For example, the dimension of the noise subspace in the conventional definition is 5 for ULA with 12 array elements when $D = 7$. However, when the weighted selector is used in the model, the corresponding dimension of the weighted noise subspace is 3, and the value of RMSPE has become lower. The figure clearly demonstrates that utilizing a weighted noise subspace, compared with the traditional noise subspace, enhances the estimation performance.

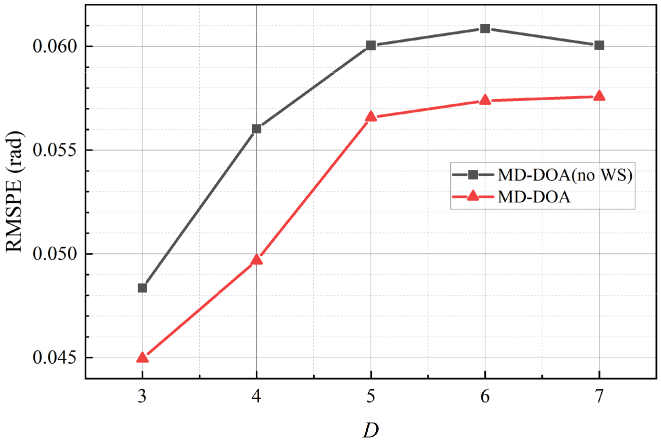


Fig. 9. Effectiveness of the weighted selector.

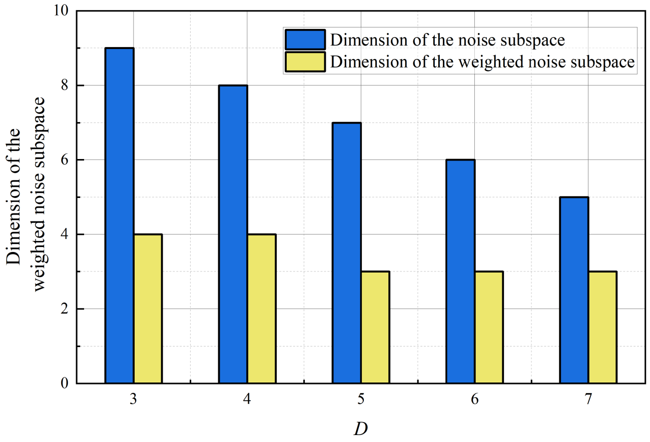


Fig. 10. Dimension of the weighted noise subspace.

C. Performance of the Proposed Architecture in DOA Estimation

In order to fully evaluate the overall performance of the proposed architecture, we conduct a series of comparison experiments. The compared techniques include the following approaches: MB methods including BF [8] and MUSIC [12], purely DD Deep MUSIC [33], as well as hybrid algorithms Subspacenet [36] and Augmented MUSIC [39].

Additionally, the Cramér–Rao lower bound (CRB) is provided as the benchmark [69]. Minor adjustments are made to accommodate differences in settings and ensure successful model training. These comparative experiments aim to thoroughly assess the efficacy, robustness, and adaptability of our

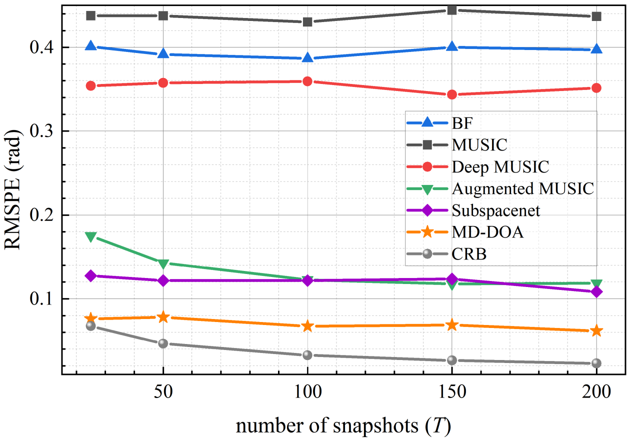


Fig. 11. DOA estimation of $D = 5$ coherent signals.

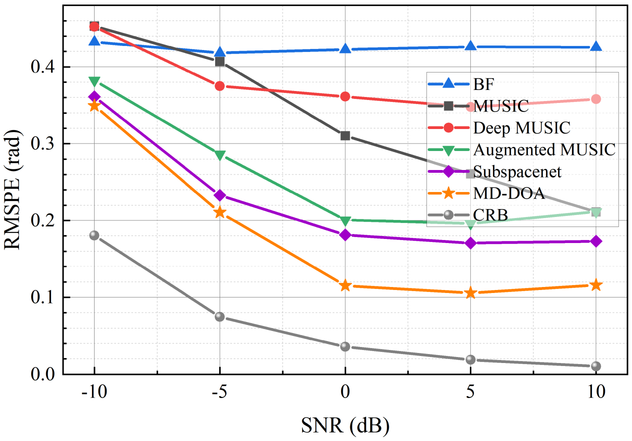


Fig. 13. DOA estimation of $D = 5$ signals with different SNRs.

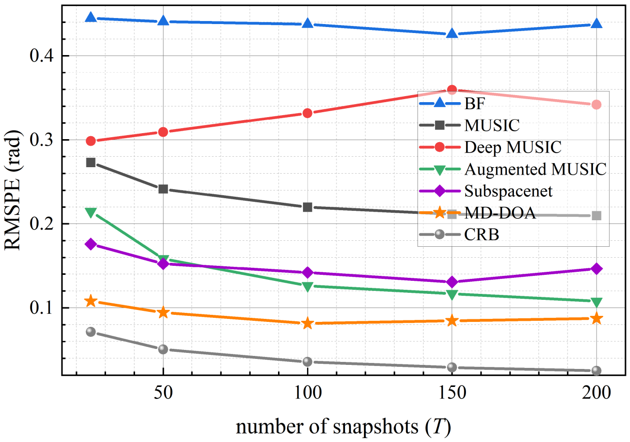


Fig. 12. DOA estimation of $D = 5$ incoherent signals.

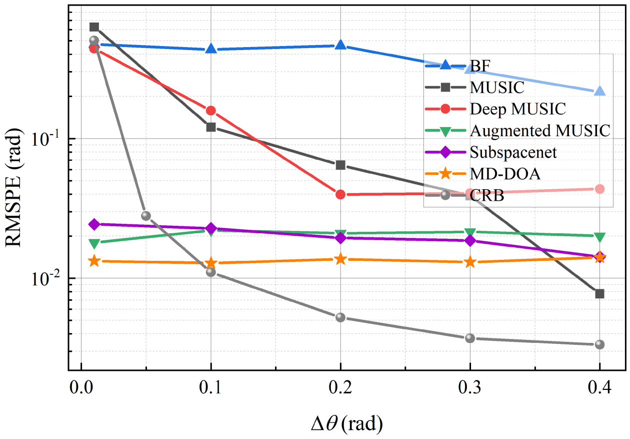


Fig. 14. DOA estimation of $D = 2$ sources with different angle separation $\Delta\theta$.

proposed architecture by comparing their performance with those of the existing methods.

1) Performance of the Proposed Architecture in Coherent and Incoherent Sources: By varying the number of snapshots T, the precision of the proposed algorithm is evaluated in localizing $D = 5$ sources. Here, a fixed SNR is kept 10 dB for training and testing. The RMSPE versus varying number of snapshots for both coherent and incoherent sources is illustrated in Figs. 11 and 12.

The data show that the performance of the MB algorithm is severely degraded when localizing coherent signals. Deep learning methods accomplish in learning the covariance matrix from the data and successfully estimating coherent signals. The hybrid algorithms possess the advantages of both MB and DD methods, as detailed in Section I-A, and demonstrate better performance. Our proposed MD-DOA architecture is closer to the CRB in the coherent and incoherent sources, significantly outperforming all the compared methods and possessing robustness. Notably, the hybrid algorithms effectively exploit the correlation information between signals to estimate coherent sources, leading to enhanced feature extraction. Consequently, the hybrid algorithms exhibit advantage in estimating coherent sources.

2) Performance of the Proposed Architecture Under Different SNRs: To compare the performance of the algorithms under different SNRs, Fig. 13 illustrates the RMSPE when there are five incoherent sources and 100 snapshots. We train the algorithms by using the dataset mixed with SNRs from −10 to 10 dB.

We can clearly observe that the MD-DOA architecture consistently presents the lower RMSPE compared to other algorithms under different SNRs. It is worth highlighting that Deep MUSIC limits adaptability to varying SNRs and performs worse than MUSIC. In particular, the MD-DOA architecture exhibits a constant and low error with no significant fluctuations under positive SNR settings demonstrating excellent robustness. It is noteworthy that the MD-DOA architecture shows better performance compared with other hybrid algorithms, such as Augmented MUSIC and Subspacenet. It maybe attributes to the efficiently extraction and reuse of the features provided by the MD-DOA architecture.

3) Resolution of the Proposed Architecture: The resolution of the proposed architecture is evaluated when locating $D = 2$ and $T = 200$ incoherent sources. These experiments are set up with two sources and the distance between them is denoted by $\Delta\theta$. The results are shown in Fig. 14.

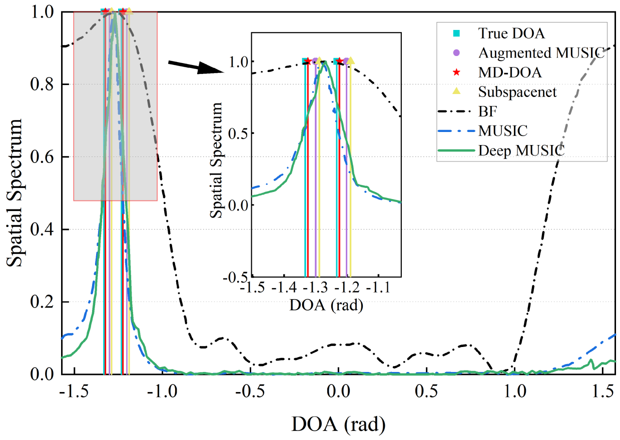


Fig. 15. Estimated spatial spectra of two coherent signals.

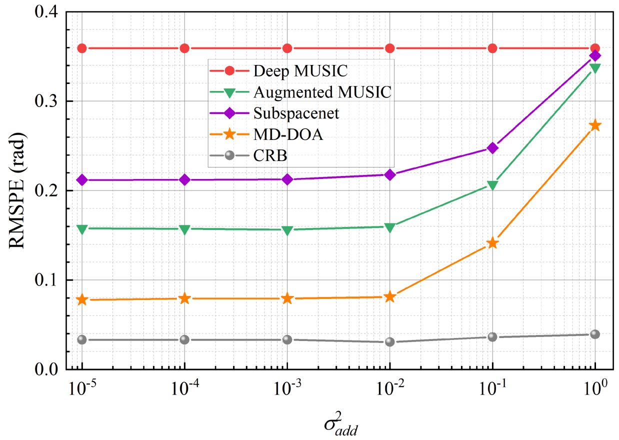


Fig. 16. DOA estimation with mismatch in the array geometry.

It shows that the MUSIC algorithm is effective when the distance is 0.4 rad, but it begins to break down when $\Delta\theta$ decreases gradually. Meanwhile, the BF and Deep MUSIC perform worse as $\Delta\theta$ decreases. The results imply that the MB approaches and the purely DD approaches have some difficulties in processing the DOA of neighboring signals due to the fact that the DOA estimation may be confused when the sources are close to each other. However, MB deep learning methods are robust and exhibit consistently low error levels under different $\Delta\theta$ conditions, and they even defeat CRB when the signal sources are very close. The possible reason is that DOA estimation with DD methods is biased [21], [70]. Among them, the MD-DOA architecture has better resolution.

4) Estimated Spatial Spectra: The estimated spatial spectra of the two coherent signals with $\Delta\theta = 0.1$ rad are shown in Fig. 15. We set the SNR to -5 dB and kept all other conditions constant. Our method can estimate DOAs more accurately, outperforming other algorithms.

5) Effectiveness of the Proposed Architecture in Array Mismatch: In practical signal processing environments, ULAs are often encounter unavoidable errors that deviate from the ideal state. Therefore, the calibration of the array shape becomes an important consideration for DOA estimation. To deeply study the effectiveness of the proposed architecture in learning from the data when dealing with array mismatch, we consider the following scenario: complex Gaussian noise with zero-mean and variance σ^2_{add} is added to each element of the steering vector $\mathbf{a}(\theta)$. The experimental setup includes the consideration of coherent signal sources at $D = 5$ and $T = 100$, as shown in Fig. 16. The primary focus of this experiment is to assess the ability of algorithms to recover DOAs by learning from the array mismatch data, so we only consider methods related to deep learning for comparison.

From Fig. 16, it is evident that the performance of the MD-DOA is the closest to the CRB. In such scenarios, our proposed architecture can learn well from array mismatch data to recover DOAs and demonstrates the capability to achieve more precise DOA estimation. It is worth noting that these results show the robustness and effectiveness of hybrid methods in array mismatch, especially in enhancing the accuracy of DOA estimation. This is important for practical applications where complex environments need to be addressed.

TABLE VI
RESULTS OF DOA ESTIMATION METHODS FOR REAL-WORLD MEASUREMENTS

Method	RMSPE [rad]	Method	RMSPE [rad]
MD-DOA	**0.0146**	Deep MUSIC	0.5171
MUSIC	0.6274	Augmented MUSIC	0.0180
BF	0.6479	Subspacenet	0.0186

D. Feasibility and Inference Time of the Proposed Architecture for Real-World Measurements

1) Real-World Measurement Settings: The real-world dataset, called DOA ESTIMATION DATASET, is available on IEEE DataPort [71], which is from Duy-Thai Nguyen (Academy of Military Science and Technology). The transmitter part of the array consists of one antenna, which transmits the signals continuously at a carrier frequency of 1090 MHz. The receiver part consists of four Vivaldi antennas placed at a distance of half a wavelength (i.e., 13.76 cm) with a tolerance of about 0.2 cm. The transmitted antenna is positioned 3.0 m away from the received antennas, which ensures that the sources are in the far field with respect to the four-element received array. The signals are acquired with a sampling rate of 20 MHz, and the number of samples in one recording is 2000. The measurement process is performed sequentially for angles from $-60°$ to $60°$ with a step of $1°$. Each angle receives 1000 signals, so the total number of received signals is 121 000, presented in a 4×2000 complex number format, and 10% of the signals for each angle is randomly selected for experimentation. The other experimental conditions are the same as Section IV-A. Additional comprehensive information regarding this dataset can be found in [71].

2) Feasibility of the Proposed Architecture: In this part, we demonstrate the performance of the proposed architecture to process data in the real-world environment, and the results are shown in Table VI.

From the results, it can be seen that the MB deep learning algorithms can effectively operate in the real-world

TABLE VII
INFERENCE TIME

Method	RMSPE [rad]	Time [ms]
MD-DOA	**0.0184**	**27.50**
Augmented MUSIC	0.0275	61.25
Subspacenet	0.0421	90.17
MUSIC	0.6545	2.24
BF	0.6283	2.70
Deep MUSIC	0.6021	33.39

environment, with the MD-DOA exhibiting the best performance. Conversely, the MB and Deep MUSIC algorithms perform worse than the MB deep learning algorithms, potentially due to their inability to overcome factors, such as complex noise in the environment.

3) Inference Time of the Proposed Architecture: In order to assess the computational complexity of all methods, the inference time for each method is recorded in Table VII. It shows the average time required to perform 100 independent inferences for a single signal with $T = 3$. The computing platform is 13th Gen Intel[1] Core[2] i7-13700 KF, 32-GB RAM, and a single NVIDIA GeForce GTX 4090 GPU.

In Table VII, the proposed MD-DOA architecture performs better than other algorithms and has the shortest inference time among the deep learning algorithms. However, due to the high complexity of the algorithm itself, the proposed MD-DOA architecture requires a longer inference time compared with MUSIC and BF algorithms.

4) Discussion: In order to quickly and efficiently reduce the MD-DOA inference time, making it comparable to the inference time of traditional methods, we adopted post training quantization (PQ) [72]. The result shows that there is a slight reduction in the performance of MD-DOA. However, the MD-DOA with PQ is still far superior to the traditional algorithms and has a significant time advantage over other deep learning algorithms. To maintain performance, we changed the training strategy by doubling the size of the training dataset, which proved to be effective. The RMSPE value had just a 0.0067 increase compared to MD-DOA, and the inference time is 1.68 ms.

However, how to find a better balance between the DOA estimation performance and inference time is an open question, which will be further explored in future research. To further tradeoff between the performance of DOA estimation and inference time, we should also consider other factors that determine the inference time of the algorithm, such as the sampling rate, the number of sampling points, and the level of hardware. In addition to the methods mentioned above, we consider continuing to optimize the algorithm structure in future research. For example, instead of MBCR2net, we can design a more lightweight network, thus reducing redundant computations, or use knowledge distillation to further optimize the algorithm, and using a kernel optimized for flexible inference may also be a better choice.

V. Conclusion

In this article, a novel end-to-end MB deep learning architecture called MD-DOA is proposed, integrating the advantages of MB and DD approaches. MD-DOA inherits the overall framework of the subspace-based algorithm, enhancing the covariance matrix estimation, noise subspace division, and peak-finding process with networks. First, the architecture improves the ability to learn correlation features between received temporal signals by employing multiscale, Reslink, and feature reuse techniques. Subsequently, the WNSnet helps to obtain a more accurate division of the subspace, and the MRP module captures the correlations in the weighted noise subspace. Finally, PDFnet is used for DOA estimation, thereby avoiding the inherent limitations of subspace algorithms. Notably, EVD and MRP serve as MB modules, ensuring interpretability. Experimental results show that the architecture outperforms similar techniques and exhibits robustness in the presence of coherent sources, low SNRs, and array mismatch. For real-world measurements, it can also run successfully.

It is important to note in Section IV-C that the MD-DOA architecture can acquire knowledge from the data and overcome array mismatch. Moreover, exploring the tradeoff between DOA estimation performance and inference time is just a process of trial and error. We make Section IV-D.4 as a good example to show that the balance between performance and inference time is possible to achieve.

Acknowledgment

The authors would like to thank the editor and anonymous reviewers for their valuable comments.

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, Jul. 1996, doi: 10.1109/79.526899.

[2] D. Fejgin and S. Doclo, "Assisted RTF-vector-based binaural direction of arrival estimation exploiting a calibrated external microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5, doi: 10.1109/ICASSP49357.2023.10095634.

[3] M. Farmani, M. S. Pedersen, Z.-H. Tan, and J. Jensen, "Informed sound source localization using relative transfer functions for hearing aid applications," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 3, pp. 611–623, Mar. 2017, doi: 10.1109/TASLP.2017.2651373.

[4] I. Bilik, O. Longman, S. Villeval, and J. Tabrikian, "The rise of radar for autonomous vehicles: Signal processing solutions and future research directions," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 20–31, Sep. 2019, doi: 10.1109/MSP.2019.2926573.

[5] M. Nouri, H. Behroozi, A. Jafarieh, and N. K. Mallat, "DOA estimation based on gridless fuzzy active learning under unknown mutual coupling and nonuniform noise: Experimental verification," *IEEE Sensors J.*, vol. 23, no. 14, pp. 15713–15720, Jul. 2023, doi: 10.1109/JSEN.2023.3266354.

[6] A. Gil-Martinez, M. Poveda-Garcia, J. A. Lopez-Pastor, J. C. Sanchez-Aarnoutse, and J. L. Gomez-Tornero, "Wi-Fi direction finding with frequency-scanned antenna and channel-hopping scheme," *IEEE Sensors J.*, vol. 22, no. 6, pp. 5210–5222, Mar. 2022, doi: 10.1109/JSEN.2021.3122232.

[7] L. Jia et al., "Direction of arrival estimation for single microelectromechanical systems vector hydrophone using modified wavelet packet de-noising," *IEEE Sensors J.*, vol. 23, no. 12, pp. 13165–13174, Jun. 2023, doi: 10.1109/JSEN.2023.3269857.

[8] M. S. Bartlett, "Smoothing periodograms from time-series with continuous spectra," *Nature*, vol. 161, no. 4096, pp. 686–687, May 1948, doi: 10.1038/161686a0.

[1] Registered trademark.

[2] Trademarked.

[9] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969, doi: 10.1109/PROC.1969.7278.

[10] E. Mabande, H. Sun, K. Kowalczyk, and W. Kellermann, "Comparison of subspace-based and steered beamformer-based reflection localization methods," in *Proc. 19th Eur. Signal Process. Conf.*, Barcelona, Spain, Aug. 2011, pp. 146–150.

[11] K. Xu, M. Xing, Y. Cui, and G. Tian, "How to determine an optimal noise subspace?" *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–4, 2023, doi: 10.1109/LGRS.2023.3238334.

[12] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986, doi: 10.1109/TAP.1986.1143830.

[13] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 8, Sep. 1983, pp. 336–339, doi: 10.1109/ICASSP.1983.1172124.

[14] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, "Model-based deep learning," *Proc. IEEE*, vol. 111, no. 5, pp. 465–499, May 2023, doi: 10.1109/JPROC.2023.3247480.

[15] S. Ge, K. Li, and S. N. B. M. Rum, "Deep learning approach in DOA estimation: A systematic literature review," *Mobile Inf. Syst.*, vol. 2021, pp. 1–14, Sep. 2021, doi: 10.1155/2021/6392875.

[16] W. Nie, X. Zhang, J. Xu, L. Guo, and Y. Yan, "Adaptive direction-of-arrival estimation using deep neural network in marine acoustic environment," *IEEE Sensors J.*, vol. 23, no. 13, pp. 15093–15105, Jul. 2023, doi: 10.1109/JSEN.2023.3274309.

[17] T. Cheng, B. Wang, Z. Wang, R. Dong, and B. Cai, "Lightweight CNNs-based interleaved sparse array design of phased-MIMO radar," *IEEE Sensors J.*, vol. 21, no. 12, pp. 13200–13214, Jun. 2021, doi: 10.1109/JSEN.2021.3069972.

[18] P.-A. Grumiaux, S. Kitić, L. Girin, and A. Guérin, "A survey of sound source localization with deep learning methods," *J. Acoust. Soc. Amer.*, vol. 152, no. 1, pp. 107–151, Jul. 2022, doi: 10.1121/10.0011809.

[19] L. Perotin, R. Serizel, E. Vincent, and A. Guérin, "CRNN-based multiple DoA estimation using acoustic intensity features for ambisonics recordings," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 1, pp. 22–33, Mar. 2019, doi: 10.1109/JSTSP.2019.2900164.

[20] H. Hammer, S. E. Chazan, J. Goldberger, and S. Gannot, "Dynamically localizing multiple speakers based on the time-frequency domain," *EURASIP J. Audio, Speech, Music Process.*, vol. 2021, no. 1, p. 16, Dec. 2021, doi: 10.1186/s13636-021-00203-w.

[21] 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, doi: 10.1109/TSP.2021.3089927.

[22] M. L. Lima de Oliveira and M. J. G. Bekooij, "ResNet applied for a single-snapshot DOA estimation," in *Proc. IEEE Radar Conf.* New York, NY, USA: IEEE, Mar. 2022, pp. 1–6, doi: 10.1109/RadarConf2248738.2022.9763905.

[23] Q. Wang et al., "Deep learning based audio-visual multi-speaker DOA estimation using permutation-free loss function," in *Proc. 13th Int. Symp. Chin. Spoken Lang. Process. (ISCSLP)*, Singapore, Dec. 2022, pp. 250–254, doi: 10.1109/ISCSLP57327.2022.10037995.

[24] Z. Tang, J. D. Kanu, K. Hogan, and D. Manocha, "Regression and classification for direction-of-arrival estimation with convolutional recurrent neural networks," in *Proc. Interspeech*, Sep. 2019, pp. 654–658, doi: 10.21437/interspeech.2019-1111.

[25] J. Cong, X. Wang, M. Huang, and L. Wan, "Robust DOA estimation method for MIMO radar via deep neural networks," *IEEE Sensors J.*, vol. 21, no. 6, pp. 7498–7507, Mar. 2021, doi: 10.1109/JSEN.2020.3046291.

[26] S. Adavanne, A. Politis, and T. Virtanen, "Differentiable tracking-based training of deep learning sound source localizers," in *Proc. IEEE Workshop Appl. Signal Process. to Audio Acoust. (WASPAA)*, New Paltz, NY, USA, Oct. 2021, pp. 211–215, doi: 10.1109/WASPAA52581.2021.9632773.

[27] X. Lan, H. Zhai, and Y. Wang, "A novel DOA estimation of closely spaced sources using attention mechanism with conformal arrays," *IEEE Access*, vol. 11, pp. 44010–44018, 2023, doi: 10.1109/ACCESS.2023.3272617.

[28] Y. Zhang, J. Tao, Y. Huang, L. Lan, J. Liu, and X. Guan, "Gridless DOA estimation for automotive millimeter-wave radar with a novel space-time network," in *Proc. IEEE Radar Conf. (RadarConf23)*, San Antonio, TX, USA, May 2023, pp. 1–6, doi: 10.1109/RadarConf2351548.2023.10149707.

[29] N. Shlezinger, Y. C. Eldar, and S. P. Boyd, "Model-based deep learning: On the intersection of deep learning and optimization," *IEEE Access*, vol. 10, pp. 115384–115398, 2022, doi: 10.1109/ACCESS.2022.3218802.

[30] M. Lin, Y. Tian, X. Zhang, and Y. Huang, "Parameter estimation of frequency-hopping signal in UCA based on deep learning and spatial time–frequency distribution," *IEEE Sensors J.*, vol. 23, no. 7, pp. 7460–7474, Apr. 2023, doi: 10.1109/JSEN.2023.3247623.

[31] A. Shultzman, E. Azar, M. R. D. Rodrigues, and Y. C. Eldar, "Generalization and estimation error bounds for model-based neural networks," 2023, *arXiv:2304.09802*.

[32] M. Pesavento, M. Trinh-Hoang, and M. Viberg, "Three more decades in array signal processing research: An optimization and structure exploitation perspective," *IEEE Signal Process. Mag.*, vol. 40, no. 4, pp. 92–106, Jun. 2023, doi: 10.1109/MSP.2023.3255558.

[33] A. M. Elbir, "DeepMUSIC: Multiple signal classification via deep learning," *IEEE Sensors Lett.*, vol. 4, no. 4, pp. 1–4, Apr. 2020, doi: 10.1109/LSENS.2020.2980384.

[34] 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, doi: 10.1109/TSP.2022.3144033.

[35] D. H. Shmuel, J. P. Merkofer, G. Revach, R. J. G. van Sloun, and N. Shlezinger, "Deep root music algorithm for data-driven doa estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5, doi: 10.1109/ICASSP49357.2023.10096504.

[36] D. H. Shmuel, J. P. Merkofer, G. Revach, R. J. G. van Sloun, and N. Shlezinger, "SubspaceNet: Deep learning-aided subspace methods for DoA estimation," 2023, *arXiv:2306.02271*.

[37] 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, doi: 10.1109/LSP.2022.3176211.

[38] Z.-W. Tan, Y. Liu, and A. W. H. Khong, "A dilated inception convolutional neural network for gridless DOA estimation under low SNR scenarios," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Chiang Mai, Thailand, Nov. 2022, pp. 760–764, doi: 10.23919/APSIPAASC55919.2022.9980340.

[39] J. P. Merkofer, G. Revach, N. Shlezinger, and R. J. G. Van Sloun, "Deep augmented music algorithm for data-driven Doa estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Singapore, 2022, pp. 3598–3602, doi: 10.1109/ICASSP43922.2022.9746637.

[40] J. P. Merkofer, G. Revach, N. Shlezinger, T. Routtenberg, and R. J. G. van Sloun, "DA-MUSIC: Data-driven DoA estimation via deep augmented MUSIC algorithm," *IEEE Trans. Veh. Technol.*, vol. 73, no. 2, pp. 2771–2785, Feb. 2024, doi: 10.1109/TVT.2023.3320360.

[41] S. Yang, G. Lin, Q. Jiang, and W. Lin, "A dilated inception network for visual saliency prediction," *IEEE Trans. Multimedia*, vol. 22, no. 8, pp. 2163–2176, Aug. 2020, doi: 10.1109/TMM.2019.2947352.

[42] T. Mittal, P. Agrawal, E. Pahwa, and A. Makwana, "NFResNet: Multi-scale and U-shaped networks for deblurring," 2022, *arXiv:2212.05909*.

[43] M. Afifi, K. G. Derpanis, B. Ommer, and M. S. Brown, "Learning multi-scale photo exposure correction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 9153–9163, doi: 10.1109/CVPR46437.2021.00904.

[44] L. Qi et al., "Multi-scale aligned distillation for low-resolution detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 14438–14448, doi: 10.1109/CVPR46437.2021.01421.

[45] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[46] B. Zhang, W. Sun, M. Ji, and K. Meng, "ResNect: An accurate and efficient backbone network for text detection model," in *Proc. 18th Int. Conf. Mobility, Sens. Netw. (MSN)*, Guangzhou, China, Dec. 2022, pp. 469–476, doi: 10.1109/MSN57253.2022.00081.

[47] X. Li, X. Wu, H. Lu, X. Liu, and H. Meng, "Channel-wise gated Res2Net: Towards robust detection of synthetic speech attacks," 2021, *arXiv:2107.08803*.

[48] L. Wang, B. Yeoh, and J. W. Ng, "Synthetic voice detection and audio splicing detection using SE-Res2Net-conformer architecture," in *Proc. 13th Int. Symp. Chin. Spoken Lang. Process. (ISCSLP)*, Singapore, Singapore, Dec. 2022, pp. 115–119, doi: 10.1109/ISCSLP57327.2022.10037999.

[49] R. Ramachandra and H. Li, "Finger-NestNet: Interpretable fingerphoto verification on smartphone using deep nested residual network," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Waikoloa, HI, USA, Jan. 2023, pp. 693–700, doi: 10.1109/WACVW58289.2023.00076.

[50] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2Net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, Feb. 2021, doi: 10.1109/TPAMI.2019.2938758.

[51] X. Zhang et al., "RFAConv: Innovating spatial attention and standard convolutional operation," 2023, *arXiv:2304.03198*.

[52] X. Li, J. Pan, J. Tang, and J. Dong, "DLGSANet: Lightweight dynamic local and global self-attention networks for image super-resolution," 2023, *arXiv:2301.02031*.

[53] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 936–944, doi: 10.1109/CVPR.2017.106.

[54] Y. He, "GoogLe2Net: Going transverse with convolutions," 2023, *arXiv:2301.00424*.

[55] M.-Y. You, "A unified two-dimensional direction finding approach for sensor arrays based on deep neural networks with inhomogeneous angle and frequency partition," *IEEE Sensors J.*, vol. 22, no. 7, pp. 6840–6850, Apr. 2022, doi: 10.1109/JSEN.2022.3151768.

[56] Y. Wang, *Theories and Algorithms of Spatial Spectrum Estimation*. Beijing, China: Tsinghua Univ. Press, 2004.

[57] T. E. Tuncer, T. K. Yasar, and B. Friedlander, "Narrowband and wideband DOA estimation for uniform and nonuniform linear arrays," in *Classical and Modern Direction-of-Arrival Estimation*, vol. 4. Boston, MA, USA: Academic, 2009, pp. 125–160.

[58] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, Jun. 2010, pp. 807–814. [Online]. Available: https://icml.cc/Conferences/2010/papers/432.pdf

[59] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[60] H. Gholamalinezhad and H. Khosravi, "Pooling methods in deep neural networks, a review," 2020, *arXiv:2009.07485*.

[61] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 13708–13717, doi: 10.1109/CVPR46437.2021.01350.

[62] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

[63] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[64] T. Routtenberg and J. Tabrikian, "Bayesian parameter estimation using periodic cost functions," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1229–1240, Mar. 2012, doi: 10.1109/TSP.2011.2173680.

[65] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "CBAM: Convolutional block attention module," 2018, *arXiv:1807.06521*.

[66] L. Liu, J. Chang, Z. Liu, and P. Zhang, "Hybrid contextual semantic network for accurate segmentation and detection of small-size stroke lesions from MRI," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 8, pp. 4062–4073, Aug. 2023, doi: 10.1109/JBHI.2023.3273771.

[67] S. Mekruksavanich, P. Jantawong, N. Hnoohom, and A. Jitpattanakul, "A novel deep BiGRU-ResNet model for human activity recognition using smartphone sensors," in *Proc. 19th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, Bangkok, Thailand, Jun. 2022, pp. 1–5, doi: 10.1109/JCSSE54890.2022.9836276.

[68] S. Liu, S. Miao, S. Liu, B. Li, W. Hu, and Y.-D. Zhang, "Circle-Net: An unsupervised lightweight-attention cyclic network for hyperspectral and multispectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 4499–4515, Apr. 2023, doi: 10.1109/JSTARS.2023.3271359.

[69] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.

[70] Y. Zhang, Y. Huang, J. Tao, S. Tang, H. C. So, and W. Hong, "A two-stage multi-layer perceptron for high-resolution DOA estimation," *IEEE Trans. Veh. Technol.*, pp. 1–16, Feb. 2024, doi: 10.1109/TVT.2024.3368451.

[71] D.-T. Nguyen, Nov. 9, 2023, "DOA estimation dataset," IEEE Dataport, doi: 10.21227/fgbv-st44.

[72] R. Krishnamoorthi, "Quantizing deep convolutional networks for efficient inference: A whitepaper," 2018, *arXiv:1806.08342*.

Xiaoxuan Xu received the B.S. degree in communication engineering from Shanghai University, Shanghai, China, in 2022, where she is currently pursuing the M.S. degree with the School of Communication and Information Engineering.

Her research interests include array signal processing, direction of arrival, and deep learning.

Qinghua Huang received the B.S. degree from the School of Control Science and Control Engineering, Shandong University, Jinan, China, in 2001, the M.S. degree from the Institute of Pattern Recognition, Shandong University, in 2004, and the Ph.D. degree from the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China.

In 2014, she spent one year as a Visiting Scholar with the University of Maryland, College Park, MD, USA. She is currently an Associate Professor with the School of Communication and Information Engineering, Shanghai University, Shanghai, China. Her research interests include array signal processing, statistical signal processing, and deep learning.