STNet: A Space-Time Network Solution for Gridless DOA Estimation With Small Snapshots for Automotive Radar System

Yanjun Zhang, Yan Huang, *Member, IEEE*, Jun Tao, *Senior Member, IEEE*, Cai Wen, *Member, IEEE*, Yu Han, Guisheng Liao, *Senior Member, IEEE*, and Wei Hong, *Fellow, IEEE*

***Abstract*— In order to play the key role of automotive millimeter wave radar in intelligent vehicle systems, direction-of-arrival (DOA) estimation is an essential problem to be solved. For practical intelligent driving applications, DOA estimation requires both real-time performance and high accuracy. Due to unique advantages, deep learning (DL) based methods have attracted more attention. Most of the existing DL-based methods require a large number of snapshots, but only a few snapshots can be guaranteed in practical applications. Moreover, they usually model DOA estimation as a multi-label classification task. The output represents the position of signal DOA on the discrete grid, and the resolution will be limited by the grid. In this paper, a new space-time Network (STNet) is proposed, which models DOA estimation as a regression task to achieve the effect of gridless estimation. We design a space correlation extraction module (SCEM) and a time correlation extraction module (TCEM), using the covariance matrix of the received signal and the original received signal as inputs respectively, treat them as different types of data. In these two modules, skip connection dense blocks (SCDBs) and long short-term memory (LSTM) networks are adopted to process two different forms of data. Through such processing, we retain sufficient information, obtain more features for the regression task, and ensure the estimation effect of using a small number of snapshots. The experimental results indicate that the STNet shows obvious performance gain in the case of small snapshots, achieves gridless estimation effect, and demonstrates excellent adaptability in situations where target DOAs are closely positioned.**

Manuscript received 12 January 2023; revised 8 November 2023 and 20 March 2024; accepted 5 May 2024. Date of publication 27 May 2024; date of current version 2 July 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62271142, Grant U20B2039, and Grant 62271138; in part by Shanghai Academy of Space Technology Innovation Funding under Grant SAST2021-043; and in part by Zhishan Young Scholar Funding of Southeast University. The Associate Editor for this article was Y. Zhang. *(Corresponding author: Yan Huang.)*

Yanjun Zhang is with the State Key Laboratory of Millimeter Waves, School of Information Science and Engineering, Southeast University, Nanjing 211100, China.

Yan Huang and Wei Hong are with the State Key Laboratory of Millimeter Waves, School of Information Science and Engineering, Southeast University, Nanjing 211100, China, and also with the Purple Mountain Laboratory, Nanjing 211100, China (e-mail: yellowstone0636@hotmail.com).

Jun Tao is with the Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education, School of Information Science and Engineering, Southeast University, Nanjing 210096, China.

Cai Wen is with the School of Information Science and Technology, Northwest University, Xi'an 710127, China.

Yu Han is with the School of Transportation, Southeast University, Nanjing 210096, China.

Guisheng Liao is with the National Key Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China.

Digital Object Identifier 10.1109/TITS.2024.3400888

***Index Terms*— Gridless direction-of-arrival (DOA) estimation, deep learning (DL), regression task, small snapshots.**

I. Introduction

AUTOMATIC driving technology is an indispensable part of the future intelligent transportation system. Current researches believe that this process requires the participation and fusion of multiple sensors. As an all-weather sensor with remote detection capability, millimeter wave radar has prominent advantages and will play a pivotal role in it. In order to achieve fast and accurate target detection through automotive radar, real-time and super resolution direction of arrival (DOA) estimation algorithm is required, which is also an imperative problem in array signal processing, especially for the foresight automotive radar [1]. For example, for a long-distance target detection, positioning, imaging, and other applications, there is a high demand for the accuracy of DOA estimation. Similarly, in the automotive radar system, vehicles often drive at high speed, which also requires for the real-time performance of DOA estimation. In addition, only a small number of snapshots are available for coherent processing in practical applications. Therefore, it is critical for many real-world scenarios to find out a low computational complexity method that can use small snapshots to achieve super-resolution DOA estimation.

A large amount of different DOA estimation algorithms have been proposed in the past years. The conventional beamforming developed from Fourier spectrum analysis is an early DOA estimation method [2], and its angular resolution is limited by Rayleigh limit [3]. As a result, it cannot achieve the super-resolution estimation. Subsequently, subspace-based methods have attracted more attention since 1980s [4], [5]. The most representative methods are multiple signal classification (MUSIC) [4] method and signal parameters via rotational invariance (ESPRIT) [5] method. They can break through the Rayleigh limit to achieve super-resolution estimation, but they also have several limitations. Generally, these kinds of algorithms are model-based methods, which usually require a large number of sampling snapshots to achieve an accurate estimation. Also, their computational burden is quite high, which is not in line with practical applications [6].

DOA estimation based on the sparse representation in the whole angle space is also a hot topic in recent years [7]

to realize the super-resolution performance. These sparse representation based methods can be generally classified into three categories: on-grid, off-grid, and gridless methods [8]. On-grid methods usually use a predefined discrete grid as a dictionary to recover the impinging sparse signals in the angular space from the received signals, such as l_1-SVD [9], sparse bayesian learning (SBL) [10], and sparse iterative covariance-based estiamation (SPICE) [11]. All the above methods assume that the DOAs of the incident signals are located at the determined discrete grids, however, the gap between true signal DOAs and their nearest grid point always exists in practice, and the estimation accuracy is directly related to the grid resolution. In this context, the off-grid methods further consider the off-grid error in the model to narrow the gap and better DOA estimation performance is achieved by the rectification of the dictionary. A sparse total least-squares (STLS) [12] approach is introduced to solve the perturbations of the regression matrix, which has a significant impact on DOA estimation problem. However, this method assumes that the matrix perturbation caused by the basis mismatch is Gaussian, which is not suitable for practical DOA estimation [13]. In [14], a model based on first-order Taylor expansion is proposed to solve the problem of off-grid DOA estimation, where the off-grid distance is assumed to satisfy uniform prior. A strategy using noise subspace fitting (NSF) is proposed in [15], the dictionary rectification model is extended from first-order to second-order Taylor approximation to achieve higher modeling accuracy. The aforementioned off-grid methods still need a relatively fine grid to ensure a small error so as to get a good estimation effect, which will lead to an increase in computational complexity. Therefore, grid evolution methods are proposed later [16], [17], and they take the pre-defined discrete grid as an adjustable parameter, which improves the estimation performance when using coarse grid. Subsequently, gridless methods appeared, these methods operate in the continuous domain directly so that they can avoid the grid mismatch problem. The representative methods are commonly based on the atomic norm [18], [19], [20], [21], [22], which achieve better performance at the expense of very high computational complexity. These sparse representation based schemes rely on the established models and usually need to make assumptions about the prior information which is difficult to obtain in practical applications. When the model has errors or the assumed prior information is inappropriate, it may lead to inaccurate estimation results. Moreover, methods like atomic norm minimizing (ANM) require iterative optimization of numerous parameters, resulting in extended computation time and difficulties in achieving real-time estimation.

Machine learning has also been introduced into the DOA estimation problem and has become an important research direction, among which support vector classification (SVC) [23] and support vector regression (SVR) [24], [25], [26] are representative. Generally speaking, SVC completes DOA estimation through classification task, while SVR performs regression task, which can realize gridless estimation. Recently, the application of deep learning (DL) method [27] in DOA estimation has become a research hotspot and shown unique advantages which are different from previous methods. As a data-driven algorithm, it does not need to rely on the pre-established model, which avoids the errors and limitations caused by the model. At the same time, after the network training is completed, only simple calculation is required when performing the estimation task, which is more suitable for practical scenarios. Finally, the DL method has strong adaptability to data defects, which show better stability in the case of low signal to noise ratio (SNR) and array imperfections.

In [28], a deep neural network is proposed to estimated signal DOAs, assuming that the required resolution is 1°. Each output value is defined to express the probability that a signal is incident at the corresponding angle, however, the results show that the success rate of accurate DOA estimation is quite low for the close-DOA scenario. Similarly, a DNN framework, composed of multi-task autoencoder (AE) and multiple parallel classifiers is employed in [29] for DOA estimation under different array imperfections. The author in [30] propose a deep convolutional neural network (CNN) for DOA estimation in low SNRs with the utilization of the sparsity prior. But due to the adoption of one-dimensional convolution filter, the experimental results show no significant performance gain [31]. Another CNN structure is introduced in [31] and improve the performance of DOA estimation at low SNRs. However, for the ideal covariance matrix is used in training, its performance will be seriously affected when the number of snapshots decreases. Also, it is difficult to be competent for DOA estimation in the case of few snapshots. In [32], a DeepMUSIC framework with multiple CNNs is proposed, each CNN was used to learn the MUSIC spectrum of the corresponding angular subregion. Although the framework exhibits less computational complexity in comparison with MUSIC, the results show that its estimation accuracy is limited by the MUSIC algorithm to a great extent, so it is difficult to achieve a significant improvement.

Existing typical DL-based methods model DOA estimation as a multi-label classification task [6], [28], [31]. The network output is usually the probability that the target DOAs fall on a series of predetermined discrete angles. Therefore, it is still a grid-based method in essence, and there will be a grid mismatch problem [33]. In the latest research, some methods have also been proposed to solve this problem. In [33], a gridless DL-based DOA estimation method named covariance reconstruction network (CRN) is proposed, which uses the Toeplitz structure of the noise-free array covariance matrix and learn the mapping between the received signal covariance matrix and the noise-free covariance matrix elements through neural network. However, the subsequent processing is relatively complex, Toeplitz matrix reconstruction is required and the DOA estimation result is obtained by Root-MUSIC algorithm or Vandermonde decomposition theorem. In addition, previous DOA estimation methods basically use the covariance matrix of the received signal as the input, which cannot retain all the useful information between different snapshots and can hardly achieve the optimal estimation effect. Furthermore, many existing methods still use a lot of snapshots in the training and testing stages, which is detrimental to the realization of

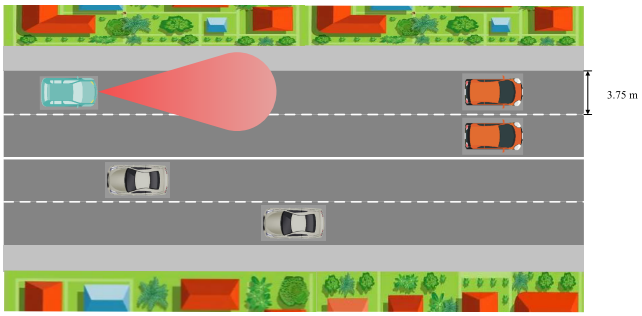


Fig. 1. Automotive radar in intelligent driving.

real-time DOA estimation. Considering that the intelligent vehicle system needs to quickly and accurately distinguish distant objects through millimeter wave radar, so as to achieve target detection, tracking and imaging, as shown in Fig. 1, the great desideratum is a method that can achieve gridless high-precision DOA estimation with small snapshots.

Therefore, in this paper, a new regression framework is proposed to seek a gridless DOA estimation with small snapshots. To mine more useful information for DOA estimation, the proposed network takes two different forms of data as inputs, namely, the elements of covariance matrix and the original sampled received signal. They are respectively sent to two proposed network modules to obtain space and time correlation features. The space correlation is extracted from the covariance matrix by fully-connected (FC) layers, and skip connection is added to enhance the resolution of close sources. The time correlation is extracted after processing the original received signal by modified LSTM layers, and the correlation between each snapshot of the signal is preserved. Then the extracted two parts of features are fused by concatenate operation. In order to avoid grid offset, the whole network performs a regression task, and then the final output results are exactly the DOA estimations through several FC layers. Numerous experimental results show that the proposed method can achieve better estimation results when using small snapshots, such as 10 snapshots, compared with the other state-of-the-art methods. The contributions of this paper are summarized as follows

(1) We model DOA estimation as a novel regression task instead of a multi-label classification problem. The output of the network directly indicates the incident angle of the signals and is regarded as a continuous variable, realizing a gridless estimation.

(2) The space-time network (STNet) framework is designed to combine both space and time features of different sources impinged on the array, yielding a better estimation result compared with the existing state-of-the-art methods under small snapshots.

(3) A space correlation extraction module (SCEM) is proposed to use the estimated covariance matrix as the input, highlighting the spatial wave-path difference caused by the signal incidence angles and different array element positions. Dense layers with skip connections are used to improve the resolution of small angle spaced targets by retaining the high-resolution features in the initial covariance matrix.

(4) A time correlation extraction module (TCEM) is also proposed, whose data input is the original received signal, which are correlated at a series of snapshots on array elements. The long short-term memory (LSTM) network structure is used to process this data, so as to utilize the real and imaginary parts as well as the correlation between data at different sampling times as much as possible, retaining past information.

The rest of paper is organized as follows: in Section II, the signal model is presented. In Section III, we introduce the proposed network framework for DOA estimation, including the specific motivations, data preprocessing, and the architecture of the proposed framework. In Section IV, the training approach and computational complexity are discussed in detail. Section V presents the experimental results for multiple circumstances compared with other state-of-the-art methods, including test results using real data, and in Section VI the whole work is summarized.

Notations: Throughout the paper the following notations are adopted: χ denotes a set and $|\chi|$ is its cardinality; The imaginary unit is j (so that $\mathrm{j}^2 = -1$). Matrices are denoted by boldface capital letters, e.g., $\mathbf{A}$; vectors are denoted by boldface lowercase letters, e.g., $\mathbf{a}$; and scalars are denoted by the lowercase letters, e.g., a. The conjugate transpose of a matrix is $(\cdot)^H$ and its transpose is $(\cdot)^T$. The $N \times N$ identity matrix is I_N. The symbol $E[\cdot]$ is the expectation operator.

II. Signal Model

We assume that K far-field narrowband sources impinge onto an N-element array from K directions, then the received signal at time t is

$$\mathbf{y}(t) = \mathbf{A}(\theta)\mathbf{s}(t) + \mathbf{n}(t), \quad t = 1, \ldots, L, \tag{1}$$

where L is the number of snapshots, the sources and array manifold matrix are, respectively, formulated as

$$\mathbf{s}(t) = [s_1(t), s_2(t), \cdots, s_K(t)]^T, \tag{2}$$

$$\mathbf{A}(\theta) = [\mathbf{a}(\theta_1), \cdots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{N\times K}, \tag{3}$$

and $\mathbf{n}(t)$ denotes the noise. Furthermore, we consider a uniform linear array (ULA) configuration for simplicity, so the steering vectors can be expressed as

$$\mathbf{a}(\theta_k) = [1, e^{\mathrm{j}\frac{2\pi d}{\lambda}\sin(\theta_k)}, \ldots, e^{\mathrm{j}\frac{2\pi d}{\lambda}\sin(\theta_k)(N-1)}]^T, \tag{4}$$

where d is the array element spacing and λ is the wavelength. The $N \times N$ covariance matrix of received signal $\mathbf{y}(t)$ is expressed as

$$\mathbf{R}_y \triangleq E[\mathbf{y}(t)\mathbf{y}^H(t)] = \mathbf{A}(\theta)\mathbf{R}_s\mathbf{A}^H(\theta) + \mathbf{R}_N, \tag{5}$$

where $E[\cdot]$ denotes the ensemble average. Herein, we assume that all the impinging signals have different DOAs and are uncorrelated signals in Gaussian distribution. Also, the additive noise values are independent and identically distributed white Gaussian, which are uncorrelated with signals, i.e., $E[\mathbf{s}(t)\mathbf{n}^H(t)] = E[\mathbf{s}^H(t)\mathbf{n}(t)] = 0$. The signal covariance matrix $\mathbf{R}_s$ and noise covariance matrix $\mathbf{R}_N$ are given by

$$\mathbf{R}_s = E[\mathbf{s}(t)\mathbf{s}^H(t)], \tag{6}$$

$$\mathbf{R}_N = E[\mathbf{n}(t)\mathbf{n}^H(t)] = \sigma_n^2\mathbf{I}_N. \tag{7}$$

By fully considering the above assumptions, the covariance matrix of the received signal can be rewritten as

$$\mathbf{R}_y \triangleq E[\mathbf{y}(t)\mathbf{y}^H(t)] = \mathbf{A}(\theta)\mathbf{R}_s\mathbf{A}^H(\theta) + \sigma_n^2\mathbf{I}_N. \tag{8}$$

However, in practical applications, we are unable to obtain this ideal covariance matrix in (8). Commonly, the covariance matrix is required to be estimated with the received data of L snapshots

$$\hat{\mathbf{R}}_y = \frac{1}{L}\sum_{t=1}^{L}\mathbf{y}(t)\mathbf{y}^H(t). \tag{9}$$

As mentioned before, we hope to estimate the unknown parameter θ, i.e., the DOAs of the signals, according to the received signal $\mathbf{y}(t)$. Although the received signal contains all the available DOA information, it is difficult to process it directly. Therefore, in the traditional methods, the covariance matrix which is easier to process is usually used for DOA estimation. Most of the existing DL-based methods also use this idea. In the STNet, we take them as different forms of information, both as part of the network input, and process them in different ways to obtain different features.

III. DOA Estimation of Small Snapshots With Deep Learning

In this part, we model DOA estimation as a regression task, and a new framework including a multiple-input DL network is proposed. In Section A, we analyze the previous DOA estimation algorithms and derive the motivation of this paper. Then the input and output forms and data preprocessing methods are introduced in Section B. Finally, Section C is devoted to the description of the network architecture and the implementation process of DOA estimation. The SCDBs perform the feature extraction from the covariance matrix. Meanwhile, LSTM layers extract features directly from the received data. Subsequently, the FC layers infer the DOA estimates through the combination of the two outputs.

A. Algorithm Review and Motivation

1) Space Correlation Extraction Module (SCEM): To estimate DOAs from multiple sources, one simple idea is to estimate the phase difference of the received signals on multiple adjacent channels. As is well known, it extracts the delay difference of the source onto different elements, which is corresponding to its DOA. In the previous algorithms, such as MUSIC, ESPRIT, and other derivative methods, super-resolution DOA estimation is usually realized based on the covariance matrix of the received signal. The covariance matrix can be formulated as

$$\mathbf{R}_y = \begin{bmatrix} r_{1,1} & r_{2,1}^* & \cdots & r_{N,1}^* \\ r_{2,1} & r_{2,2} & \cdots & r_{N,2}^* \\ \vdots & \vdots & \ddots & \vdots \\ r_{N,1} & r_{N,2} & \cdots & r_{N,N} \end{bmatrix}, \tag{10}$$

where the superscript * is the conjugate operation, $r_{p,q}$ denotes the covariance between the p-th channel and the q-th channel. As a second-order statistic, the covariance matrix highlights the relationship between each array element, that is, the wave path difference caused by the different spatial positions of the array elements and the angle of signal incidence, which is a kind of spatial correlation information. The traditional algorithms conduct eigenvalue decomposition (EVD) to construct the signal subspace and noise subspace and carry out DOA estimation on this basis

$$\mathbf{R}_y = \mathbf{U}\Lambda\mathbf{U}^H, \tag{11}$$

where Λ is a diagonal matrix composed of the eigenvalues of $\mathbf{R}_y$ in the descending order and $\mathbf{U} = [\mathbf{U_S}; \mathbf{U_N}]$ is an $N \times N$ eigenvector matrix. In theory, the first K column vectors correspond to the signal subspace by $\mathbf{U_S}$, and the remaining $N - K$ column vectors are the noise subspace eigenvectors as $\mathbf{U_N} \in {}^{N\times(N-K)}$. In practical simulations, it also works by choosing less than $N - K$ columns as the noise subspace for DOA estimation. Then the spatial spectrum of MUSIC method can be expressed as

$$\mathrm{P_{MUSIC}}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{U_N}\mathbf{U_N}^H\mathbf{a}(\theta)}. \tag{12}$$

The DOA estimate $\hat{\theta}$ is obtained from the peaks of $\mathrm{P_{MUSIC}}(\theta)$, which can be approximately formulated as

$$\hat{\theta} = \arg\max_{\theta\in\theta^c} \ \mathrm{P_{MUSIC}}(\theta), \tag{13}$$

where θ^c denotes the corresponding region of one source angle. The covariance matrix of the received signal contains more information about the spatial dimension that is easy to extract, and it is also the most commonly used input form of previous DL methods. The powerful nonlinear mapping ability of neural network is used to realize the transformation from the covariance matrix to the spatial spectrum [30], [32], or realize end-to-end DOA estimation with the covariance matrix [28], [31].

Without generality, take MUSIC method as an example, the EVD operation is a nonlinear projection to extract the noise subspace vectors from the original covariance matrix, i.e., $\mathbf{U_N} = \mathcal{F}_1(\mathbf{R}_y)$. Then the spatial spectrum presents another linear projection and the search of maximum peaks likes the ReLU operation in nerual networks, which is formulated as $\hat{\theta} = \mathcal{F}_2(\mathbf{U_N})$. The aforementioned steps are quite similar to a multiple-layer neural network, which is capable to realize an end-to-end nonlinear fitting. Therefore, we designed a SCEM network to extract the features of spatial dimensions from the covariance matrix and use it as one of the basis for DOA estimation.

2) Time Correlation Extraction Module (TCEM): In another perspective, it is also a typical idea to directly process the original received signal and to obtain the information of the directions. One classic method is digital beamforming (DBF), which directly designs a weight vector on each element to realize the selection of the direction angle through spatial domain filtering. Its output can be simply expressed as

$$\mathbf{h}(t) = \mathbf{w}^H\mathbf{y}(t) = \mathbf{w}^H\mathbf{A}(\theta)\mathbf{s}(t) + \mathbf{w}^H\mathbf{n}(t), \tag{14}$$

where $\mathbf{w}$ is the weight vector, it increases the signal in the corresponding direction and suppresses the signal and noise in

the other directions. By directly processing the spatial angle information in the received signal, all the information of the antenna array unit signal is saved, and the incident angle of the signal can be quickly estimated. Because of its low computational complexity, it has been widely used in practice. However, the defect of DBF algorithm is that the angular resolution is limited by the array aperture, so it can not achieve super-resolution estimation.

To obtain the super-resolution estimation results via the received signal directly, the compressed sensing (CS) methods are proposed in recent years. Specifically, the CS methods directly utilize the sparse characteristics of the impinging sources in the spatial domain to figure out the sparse representation of the received signal with an over-complete basis composed of samples from the array manifold. These methods are usually based on a set of determined sampling girds $\{\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_{N_\theta}\}$ of the angle search space and the objective function is [9]

$$\min \|\mathbf{Y} - \mathbf{AS}\|_F^2 + \lambda \left\| \mathbf{s}^{(\ell_2)} \right\|_1, \tag{15}$$

where $\mathbf{Y}$ is the received signal matrix, $\mathbf{A} = [\mathbf{a}(\tilde{\theta}_1), \mathbf{a}(\tilde{\theta}_2), \ldots, \mathbf{a}(\tilde{\theta}_{N_\theta})]$ is an over-complete representation in terms of all possible source locations, $\mathbf{S}$ is the sparse signal matrix to be recovered and $\left\| \mathbf{s}^{(\ell_2)} \right\|_1$ imposes a prior that signal $\mathbf{s}(t)$ has sparsity in the spatial dimension and is not sparse in time dimension. To be specific,

$$\mathbf{s}^{(\ell_2)} = [s_1^{(\ell_2)}, \cdots, s_{N_\theta}^{(\ell_2)}], \tag{16}$$

$$s_i^{(\ell_2)} = \|[s_i(t_1), s_i(t_2), \ldots, s_i(L)]\|_2, \tag{17}$$

it is done by first computing the ℓ_2 norm of all time-sample of a particular spatial index of $\mathbf{s}$ and use the ℓ_1 norm of $\mathbf{s}^{(\ell_2)}$. The sparse prior of the original signal is used to construct the cost function, and obtain the sparse form of the original signal through a series of parameter transformations. Finally, the direction information is obtained from the peaks of $\mathbf{S}$.

Actually, the form of (15) could be transformed to a question that learning feature $\mathbf{S}$ from input $\mathbf{Y}$ based on the representation basis $\mathbf{A}$. Then $\mathbf{AS}$ clearly defines the form of the feature representation and the regularization term $\left\| \mathbf{s}^{(\ell_2)} \right\|_1$ incorporates the prior knowledge. The problem could be solved by a similar class of iterative algorithms [34].

$$\mathbf{z}^{k+1} = \mathcal{N}(\mathcal{L}_1(\mathbf{Y}) + \mathcal{L}_2(\mathbf{z}^k)), \tag{18}$$

where $\mathbf{z}^k$ denotes the intermediate output of the $k-$th iteration, $\mathcal{L}_1, \mathcal{L}_2$ are linear (or convolutional) operators and $\mathcal{N}$ is a simple nonlinear operator. Therefore, (15) could be expressed by a recursive system then unfolded and truncated to k iterations, to construct a $(k+1)$-layer feed-forward network. Without any further tuning, the resulting architecture will output a k-iteration approximation of the exact solution $\mathbf{S}$. Therefore, the neural network can also realize the transformation from received signals to the sparse form of signal $\mathbf{s(t)}$ and avoid the complex calculation and parameter tuning, but this may require a lot of snapshots, and a large number of $\mathbf{S}$ needs to be calculated in advance as training samples.

Most previous solutions based on received signals rarely consider the information correlation between different snapshots. In practical scenarios, the arrival angle is approximately constant within a small number of snapshots, which is the basis for multi-snapshot DOA estimation. Then the reflections in different snapshots are actually correlated, which is also the basis of a precise estimation of covariance matrix in Eq. (9). For the original received signal

$$\mathbf{Y} = \begin{bmatrix} y_1(1) & y_1(2) & \cdots & y_1(L) \\ y_2(1) & y_2(2) & \cdots & y_2(L) \\ \vdots & \vdots & \ddots & \vdots \\ y_N(1) & y_N(2) & \cdots & y_N(L) \end{bmatrix}, \tag{19}$$

each row represents the sampled signal on a single array element, which only has time correlation, and there is no correlation information between each array element. Therefore, we design a TCEM, which mainly uses the time dimension information in the received signal to obtain the features containing direction information and then use it as another input basis for the DOA estimation.

3) Space-Time Network (STNet) for Regression Task: In the whole STNet framework, the first thing to do is to extract the space- and time-correlation features of interest from the original data. The two modules mentioned above have completed this part of the task. After using SCEM and TCEM to extract features from the original received signal and covariance matrix, two different output features are obtained. Then, these two output features are concatenated and DOA estimation is performed subsequently. Most previous approaches treat DOA estimation as a multi-label classification task and utilize multiple FC layers as classifiers to ultimately obtain the estimated values on the discrete grids predetermined across the entire angle search range. At present, estimation accuracy will be constrained by grid resolution, and the output scale will be relatively large if high resolution is desired. To avoid the grid mismatch problem caused by multi-label classification task modeling, we approach the DOA estimation problem as a regression task. This means that the final outputs directly correspond to the estimated DOAs. Thus, only a small-scale output is needed in this context rather than a large-scale one, reducing the computational complexity simultaneously. Finally, our goal is to combine two DOA estimation concepts by simultaneously processing two different data structures. This will result in more precise and reliable DOA estimation under small snapshots.

B. Data Preprocessing and Labeling

DOA prediction is modeled as a regression task, so the length of the output vector directly corresponds to the number of sources K, and the output value is equal to the angle value of the sources. First, the covariance matrix of the received signal is also a common input of DL-based DOA estimation networks. It contains different forms of DOA information from the original sampling data, so we take the preprocessed covariance matrix as the first input. Since it is a Hermitian matrix and all elements except the main diagonal are complex,

the input form of SCEM is formulated as

$$\mathbf{x}_1 = [r_{1,1}, \ldots r_{N,N}, \Re(r_{2,1}), \Im(r_{2,1}), \Re(r_{3,1}), \Im(r_{3,1}), \Re(r_{3,2}), \Im(r_{3,2}), \ldots, \Re(r_{N,N-1}), \Im(r_{N,N-1})]^T, \quad (20)$$

where $r_{i,j}, i, j \in \{1, 2, \ldots, N\}$ denotes the $(i, j)-$th element of the covariance matrix. In addition, in order to retain all useful information, we directly take the original sampling data as the second input. For the complex received signal $\mathbf{y}(t), t = 1, \ldots, L$, we need to preprocess it as a real-value vector. Then input $\mathbf{x}_2$ of TCEM can be expressed as

$$\mathbf{x}_2 = [\Re(\mathbf{y}(1)), \Im(\mathbf{y}(1)), \ldots, \Re(\mathbf{y}(L)), \Im(\mathbf{y}(L))]^T, \quad (21)$$

where $\Re(\cdot)$ and $\Im(\cdot)$ represent the real and imaginary parts, respectively. Therefore, the inputs of the two parts are a $N^2 \times 1$ vector and a $2L \times N$ tensor, respectively. And the total input is a collection of D data points defined as $\chi = \{\mathbf{x}_1^1, \mathbf{x}_2^1, \mathbf{x}_1^2, \mathbf{x}_2^2, \ldots, \mathbf{x}_1^D, \mathbf{x}_2^D\}$.

Next, in order to achieve gridless estimation, the whole problem is modeled as a regression task in our method, which is a critical improvement compared with previous DL methods. As a result, the output directly represents the DOAs of the target signals. For example, we assume that two targets are incident on the array from directions θ_1 and θ_2 respectively, then the output is $\mathbf{z} = [\theta_1, \theta_2]^T$. Hence, the $i-$th training sample in the dataset is with the form $(\mathbf{x}_1^i, \mathbf{x}_2^i, \mathbf{z}^i)$ and the whole dataset can be expressed as $\mathcal{D} = \{(\mathbf{x}_1^1, \mathbf{x}_2^1, \mathbf{z}^1), (\mathbf{x}_1^2, \mathbf{x}_2^2, \mathbf{z}^2), \cdots, (\mathbf{x}_1^D, \mathbf{x}_2^D, \mathbf{z}^D)\}$. Next, we aim to extract the information about signal DOAs from the input through the DL method and output the estimation results.

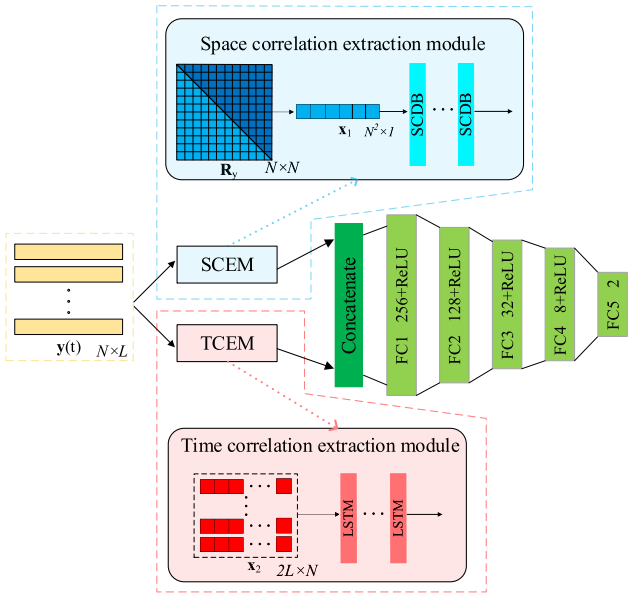


Fig. 2. The STNet framework.

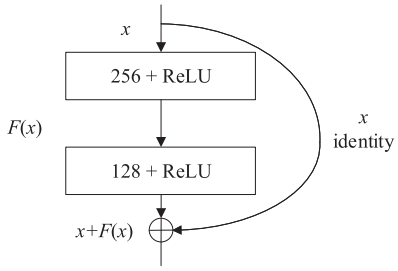


Fig. 3. The skip connection dense block.

C. The Proposed Network Architecture

The general approximation theorem indicates that a single-layer feedforward neural network with enough neurons can approximate any continuous function on a compact subset of $\mathbb{R}^n$ [35]. In practical applications, we usually use different networks to process different forms of data. First, the input of the first part is a vector, which is composed of the elements in the covariance matrix of the received signal. The DOA information in it is not as complete as that in the original signal, but it is easier to extract. We use the SCDBs to process this part of the data. Secondly, the input of the second part is the multi-snapshot signal, so it can be regarded as a time-related sequence. Therefore, we use the LSTM layers to process this part of the data. Finally, the features extracted from the two networks are combined, followed by several FC layers, and the final output directly represents the estimation result of the target DOAs. The data processing flow and network structure are illustrated in Fig. 2.

1) The SCEM Architecture: In this module, the input is a vector containing the elements of covariance matrix. Previous studies have shown that a simple FC network can extract features from the input vector and realize high-precision DOA estimations [28]. But an important problem is that it is difficult to distinguish targets with close DOAs. The skip connection mechanism can retain the shallow features with a higher resolution and effectively reduce the gradient disappearance, which is often used in image recognition and image super-resolution problems [36], [37], [38]. Therefore, we use the dense block with skip connection to process this part of data. Specifically, by retaining the clear angle information in the shallow layer, the subsequent deep features have a stronger ability to distinguish targets with similar angles, and improve the super-resolution estimation performance of the network. In general, the input is followed by three SCDBs with its specific structure shown in Fig. 3. Each SCDB contains two layers of networks with 256 and 128 neurons respectively, using the ReLU activation function, i.e, $f_{ReLU}(u) = \max(0, u)$. The function $f_1(f_2(\cdot))$ represents the SCDB, then the output of the first part is expressed as

$$\mathbf{w}_1 = f_1(f_2(\mathbf{v}_2)), \quad (22)$$

where we have

$$\mathbf{v}_2 = \text{concatenate}(\mathbf{v}_1, f_1(f_2(\mathbf{v}_1))), \quad (23)$$

$$\mathbf{v}_1 = \text{concatenate}(\mathbf{x}_1, f_1(f_2(\mathbf{x}_1))). \quad (24)$$

2) The TCEM Architecture: In this module, the input is the discrete sampled received signal at different time samples, which can be regarded as a time series. For a time series of received signal $\mathbf{y}(t)$, it can be proved that there is a strong correlation between two different time samples especially

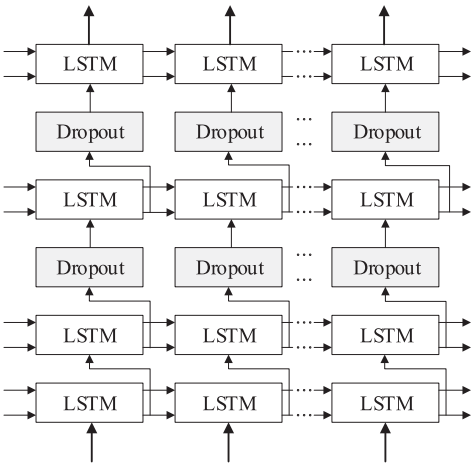


Fig. 4. The LSTM layers.

under high input signal-to-noise ratio (SNR) [39], [40]. The correlation coefficient can be expressed as

$$\rho = \frac{Cov(y(t_i), y(t_j))}{\sigma_{y(t_i)}\sigma_{y(t_j)}}, \tag{25}$$

meanwhile, the number of columns represents the time step and the number of rows represents the feature dimension. Also, it is worth noting that it does not have sparsity in the time dimension, and any two time points are related to each other. To retain and utilize the correlation of information between all time points as much as possible, a appropriate choice is to use LSTM layers on processing the received data. At this point, the second input is followed by four LSTM layers, whose dimensions are 32, 64, 64, and 32 respectively, and two Dropout layers are added, as illustrated in Fig. 4. For each LSTM layer, the *tanh* activation function is used, i.e, $f_{\tanh}(u) = (e^u - e^{-u})/(e^u + e^{-u})$. Then, the output of the TCEM is denoted as

$$\mathbf{w}_2 = f_3(f_4(f_5(f_6(\mathbf{x}_2)))), \tag{26}$$

where $f_i(\cdot), i = 3, 4, 5, 6$ represent the LSTM layers, respectively.

3) DOA Estimation Through Regression: After the features, extracted from the two modules of the network, are obtained, they are integrated and passed through several FC layers to obtain the final output, that is,

$$\widehat{\mathbf{z}} = f_7(\cdots(f_{12}(\mathbf{v}_3))), \tag{27}$$

where the functions $f_i(\cdot), i = 7, 8, 9, 10, 11, 12$ represent the FC layers and $\mathbf{v}_3 = \text{concatenate}(\mathbf{w}_1, \mathbf{w}_2)$.

The proposed network is trained in a supervised way on dataset $\mathcal{D}$, the pair $(\mathbf{x}_1^i, \mathbf{x}_2^i, \mathbf{z}^i)$ is used as the inputs and the expected output to train the network. In particular, since the adopted approach is a regression task, the mean square error (MSE) between the actual output and the expected one $\mathbf{z}^i = [\mathbf{z}^i(1), \cdots, \mathbf{z}^i(K)]^T$ is used as the loss function

$$L_{MSE}\left(\widehat{\mathbf{z}^i}, \mathbf{z}^i\right) = \frac{1}{K}\sum_{k=1}^{K}\left(\widehat{\mathbf{z}^i}(k) - \mathbf{z}^i(k)\right). \tag{28}$$

IV. Algorithm Analysis

In this section, the training strategy of the network framework is described in detail. In Section A, the parameter and setting training process of the proposed network are described; In Section B, we analyze and compare the computational complexities of the STNet and other methods.

A. Training Analysis

In this subsection, we consider the training strategy of the proposed network in two cases. One is the case of two sources, i.e., $K = 2$, which is also the main research object in many related literature. In the other case, when the number of targets is greater than two ($K = 3, 4$), for the proposed regression network framework, only the number of neurons in the output layer needs to be changed.

1) DOA Estimation of Two Targets: Herein, we consider that the angle search range is $-60°$ to $60°$, and the number of incident sources is 2, i.e., $K = 2$.

Since the STNet is intended to achieve gridless DOA estimation, the signal angles should not be limited to fixed discrete grid points, but take values in the continuous domain within the angle search range. In this way, numerous cases need to be considered, and in fact, we can not achieve infinite precision. Therefore, appropriate strategies should be adopted to generate training dataset. For signal DOAs, we use *float32* data type, which is sufficient for experimental requirements and practical applications. Meanwhile, in order to reduce the possible angle pairs and compress the dataset, we restrict the angle interval $\Delta\theta$ between two targets. The range of angle interval $\Delta\theta$ is $[1°, 10°)$, because we pay more attention to the estimation effect of small-interval targets, and every $0.01°$ is regarded as one case, so there are 900 different angle intervals in total. For each angle interval, we randomly select 500 angle pairs, where the first signal DOA θ_1 locates in the region of $[-60°, 60° - \Delta\theta]$ and the second signal DOA range is $\theta_2 = \theta_1 + \Delta\theta$. Hence, there are 450,000 samples for one specific signal-to-noise ratio (SNR). For the SNR, the setting of increasing from 0 dB to 30 dB in a 5 dB step is adopted, that is, 0, 5, 10, 15, 20, 25, 30 dB. This setting can effectively reduce the amount of data, and because of the stability of the neural network, it can also maintain good performance in the case of untrained SNR situation. By considering low SNR conditions can not significantly improve the estimation effect. Thus, we obtain a training dataset with $D = 450,000 \times 7 = 3,150,000$ samples.

During the training phase of the network, the dataset is randomly divided into training set (90%) and validation set (10%). The batch size is set to 256 and the network is trained for 100 epochs. We employ Adam optimizer [41] with an initial learning rate of 0.0001 for the optimization of the network's parameters. In addition, the learning rate decreases by a factor of 0.7 every 10 training epochs. The network is implemented in Keras using Tensorflow as the backend. The experiment platform is Ubuntu running on a server with an NVIDIA GeForce RTX 3090 GPU.

2) DOA Estimation of Multiple Targets: Similar to the two-target case, we describe our training approach for multiple

TABLE I
COMPARISON OF COMPUTATIONAL COMPLEXITY VIA DIFFERENT METHODS

Algorithms	Complexity
STNet	$\mathcal{O}\left(N^2L\right)+\mathcal{O}\left(N_{STNet}\right)$
MUSIC	$\mathcal{O}\left(N^2L+N^3+J(N-K)(N+1)\right)$
ANM	$\mathcal{O}\left(J_a(N+1)^3+N^2+N^3\right)$
ESPRIT	$\mathcal{O}\left(N^2L+N^3+2K^2(N-1)+3K^3\right)$
DNN [28]	$\mathcal{O}\left(N^2L\right)+\mathcal{O}\left(N_{DNN}\right)+\mathcal{O}\left(J(N-K)(N+1)\right)$
CNN [31]	$\mathcal{O}\left(N^2L\right)+\mathcal{O}\left(N_{CNN}\right)+\mathcal{O}\left(J(N-K)(N+1)\right)$
CRN [33]	$\mathcal{O}\left(N^2L\right)+\mathcal{O}\left(N_{CRN}\right)+\mathcal{O}(N^3+N^2K)$

targets, i.e., the number of targets $K = 3, 4$. At this time, we randomly select the corresponding number of values within the angle search range of $-60°$ to $60°$ as the incident direction of the target. Therefore, we can get the angle combination $(\theta_1, \cdots, \theta_K)$, $K = 3, 4$ to generate data and use it as a label $\mathbf{z}$. Among them, each $\theta_i, i = 1, \cdots, K$ is a *float32* data type, and 1,000,000 angle combinations are randomly generated in each SNR case, yielding $D = 7,000,000$ samples in the whole dataset. Similarly, 90% of them are used as training set and 10% as validation set. The batch size was set to 128 and the network was trained for 100 epochs. The rest of the settings are the same as before.

B. Computational Complexity

The calculation of deep learning methods mainly focuses on the offline training stage. The computational complexity of the trained model in practical application is quite low, and only a small amount of computation is required, which presents its unique advantage. However, its limitation is that it usually needs to be trained for specific array models or scenes. Once the array model is replaced or the application scenario changes, it often needs to be retrained. Classic methods are usually model-driven solutions, with rigorous physical models and mathematical derivations, which are universal for different models and application scenarios. The consequent problem is that each estimation requires complex calculations, and is more vulnerable to data defects. Table I summarizes the approximate computational complexity of each method, where J represents the accuracy of DOA peak search, J_a represents the number of iterations, and $\mathcal{O}\left(N_{STNet}\right)$, $\mathcal{O}\left(N_{DNN}\right)$, $\mathcal{O}\left(N_{CNN}\right)$ and $\mathcal{O}\left(N_{CRN}\right)$ represent the complexity of STNet, DNN, CNN and CRN, respectively. Compared to other DL methods, the proposed STNet avoids peak search or post-processing, reducing computational complexity. However, when it comes to DL methods, the computational complexity is difficult to directly reflect through expressions and is closely related to the number of parameters. Typically, runtime serves as a more intuitive representation.

In Table II, we compare the calculation time of various DOA estimation methods. The experiment was repeated 10 times, each method performs 1000 operations in each experiment on the same equipment and averages the results to obtain the time for single estimation. The results indicate that DL methods have shorter run time compared to classic methods.

TABLE II
MODEL COMPLEXITY AND AVERAGED COMPUTING TIME

	STNet	MUSIC	ANM	ESPRIT	NN [28]	CNN [31]	CRN [33]
Total params	462,906	/	/	/	107,971	15,609,721	5,610,264
Test time /s	$5.02e^{-4}$	0.065	0.273	$6.32e^{-3}$	$2.87e^{-4}$	$6.72e^{-3}$	$5.68e^{-3}$

Furthermore, the STNet is more efficient than CNN in [31] and CRN in [33], the parameters that need to be trained are greatly reduced, and the estimation results can be obtained faster. This superiority makes the STNet especially suitable for applications where real-time DOA estimation is needed. It is worth noting that the CNN framework used for comparison is identical to that in [31]. However, the assumed number of array elements is $N = 12$, which is less than the case in [31]. Due to the reduction of the input data matrix, the convolutional layer parameters are reduced. Although the model structure remains unchanged, the total number of parameters is smaller compared to the model in [31].

V. EXPERIMENTAL RESULTS

In this section, we provide the simulation results under different experimental conditions, where we evaluate the DOA estimation performance of the STNet under various settings. In both training and testing experiment phases, the array model considered is an equivalent $N = 12$ element ULA, and the array element spacing is equal to half wavelength with the theoretical resolution being about 9.55°. The state-of-the-art methods we use for comparison mainly include MUSIC [4], ANM [19], [21], [22], ESPRIT [5], SVR [24], [26], DNN [28], CNN [31] and CRN [33], where the MUSIC and ESPRIT methods are traditional subspace-based methods, ANM is a gridless estimation method based on compressed sensing, SVR is a gridless method based on machine learning, DNN and CNN are DL-based methods and CRN is a gridless DL-based method. Among them, the MUSIC method has a grid resolution of 1°. For deep learning methods, i.e., NN and CNN methods, their frameworks are fully preserved, and the output also corresponds to a grid resolution of 1°, consistent with the original [28], [31]. The training data generation process followed the method described in [31], which took into account all possible DOA combinations. For each combination, 10 sets of samples were obtained to ensure an adequate number of training samples. Additionally, the angular region and SNR settings of the training samples are consistent with the proposed method, while the other training strategies remain unchanged. For all DL methods, the received signals sampled by 10 time instances are used in the training phase, i.e., $L = 10$.

A. DOA Estimation Performance

In this series of experiments, we have adopted the received signals of 10 snapshots, and, like many other experimental settings in the literature, the incident angle of the source is accurate to 0.01°, which is enough to meet almost all application needs.

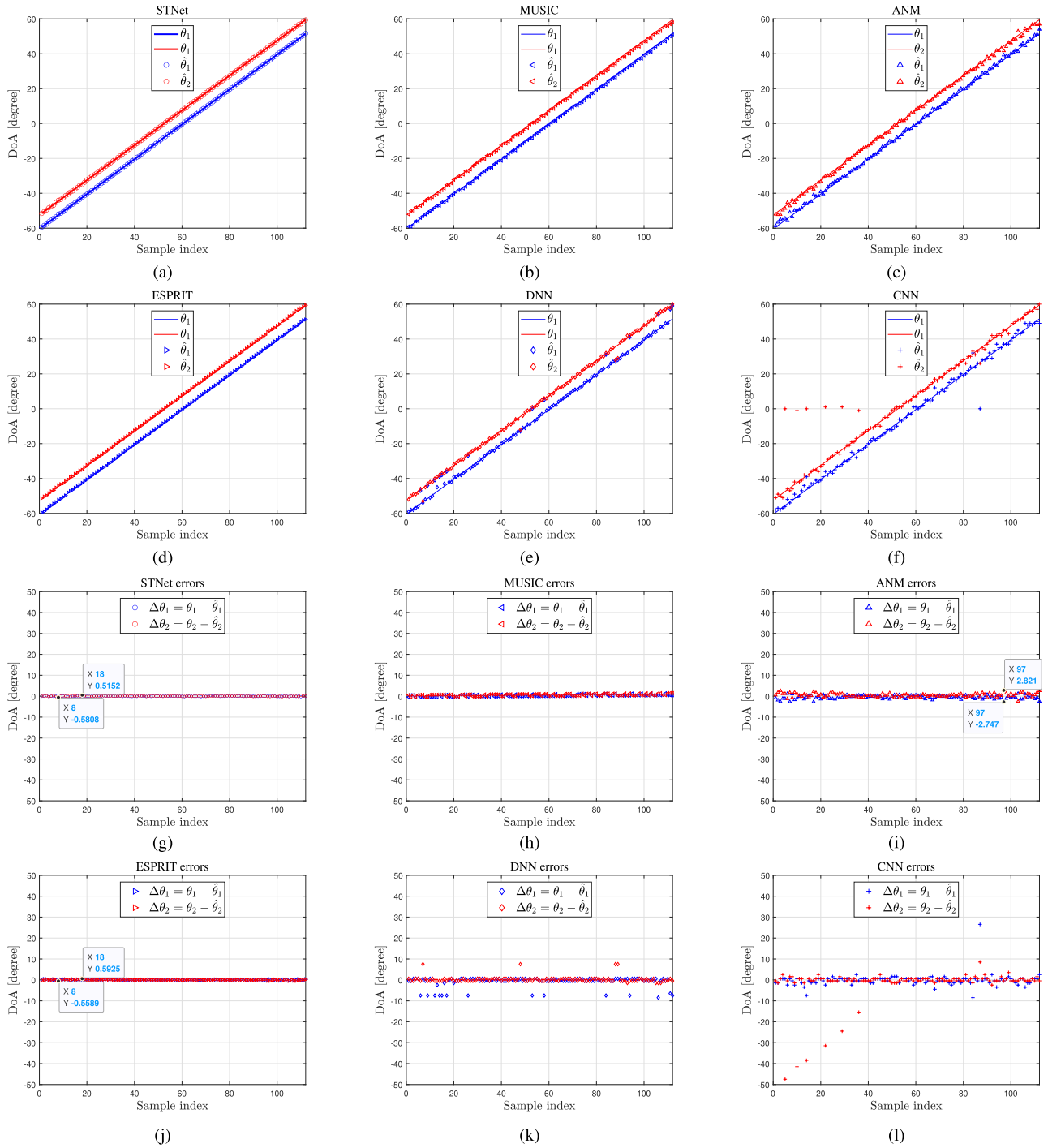


Fig. 5. DOA estimation performance on angles $\theta_1, \theta_2 \in [-60°, 60°]$ at 17 dB SNR with interval $\Delta\theta = 8°$. The results of: (a) proposed, (b) MUSIC, (c) ANM, (d) ESPRIT, (e) DNN and (f) CNN. The DOA estimation errors of the: (g) Proposed, (h) MUSIC, (i) ANM, (j) ESPRIT, (k) DNN and (l) CNN.

First, we fix the arrival angle interval of the two incident sources to 8°, in which the angle of the first source changes from $-59.47°$ to $51.53°$ in the step of 1°, so the angle of the second source changes from $-51.47°$ to $59.53°$, with a total of 112 test sample pairs. The DOA estimation results of STNet, MUSIC, ANM, ESPRIT, DNN, and CNN under SNR = 17 dB for these samples are shown in Fig. 5 (a)-(f), respectively, with the corresponding errors shown in Fig. 5 (g)-(l). It can be obviously seen that the proposed STNet outperforms the other DL-based methods. In the case of small snapshots, the DNN method sometimes can not accurately distinguish two targets, and CNN method performs similar. Since CNN method adopts the theoretical covariance matrix in (8) at the training stage, the difference between the test data and the training data increases significantly when the number of snapshots is small. And its performance decays obviously, and it is even unable to make estimates on some samples. In terms of errors, the performance of the STNet can achieve the effect similar to ESPRIT method and better than ANM method, which shows a better gridless solution. The errors of the STNet

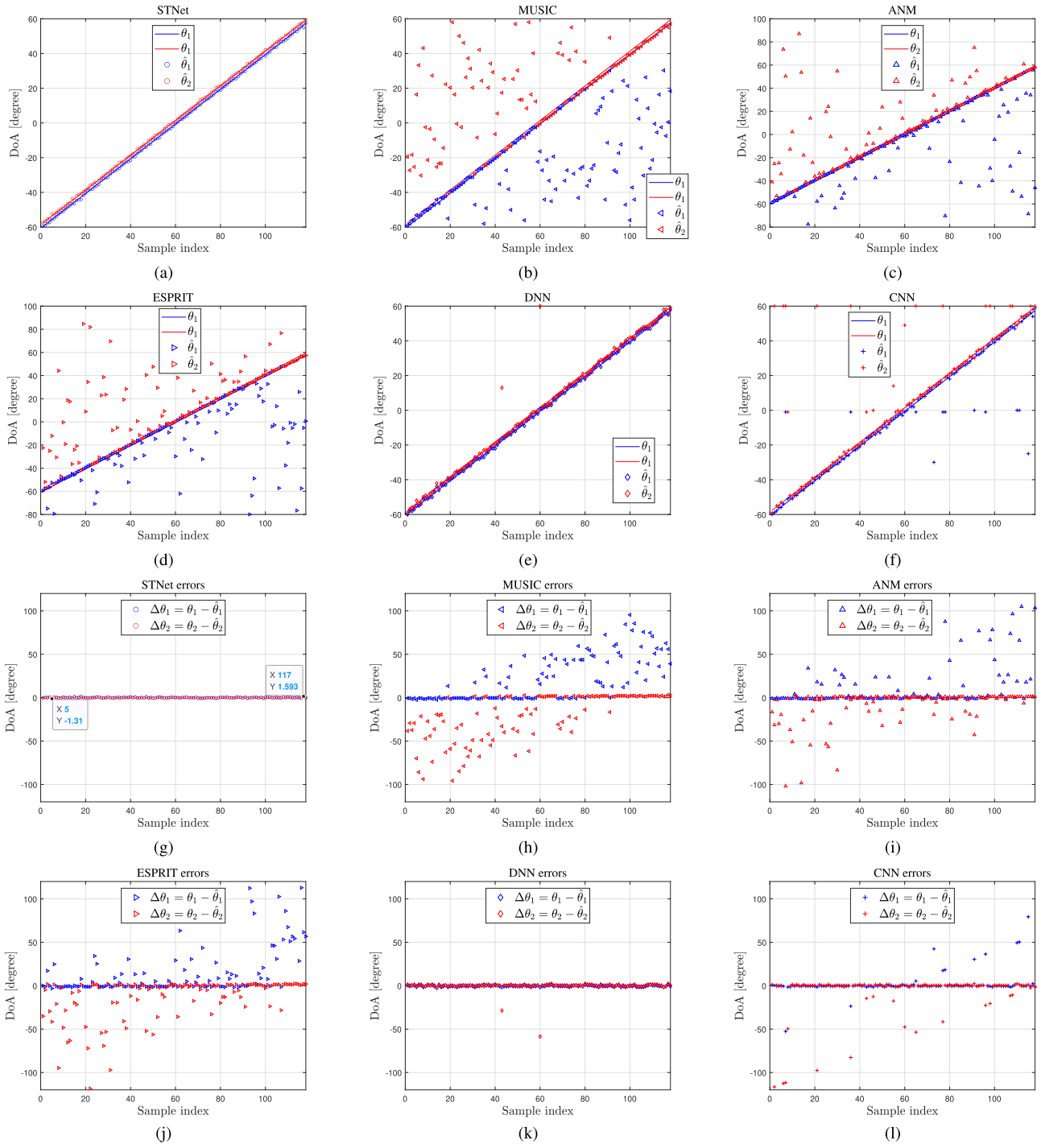


Fig. 6. DOA estimation performance on angles $\theta_1, \theta_2 \in [-60°, 60°]$ at 5 dB SNR with interval $\Delta\theta = 2°$. The results of: (a) proposed, (b) MUSIC, (c) ANM, (d) ESPRIT, (e) DNN and (f) CNN. The DOA estimation errors of the: (g) proposed, (h) MUSIC, (i) ANM, (j) ESPRIT, (k) DNN and (l) CNN.

is within the interval $[-0.5808°, 0.5152°]$, while errors of ESPRIT method and ANM method are between the interval $[-0.5589°, 0.5925°]$ and $[-2.747°, 2.821°]$, respectively.

Next, to test and compare the stability and robustness of the proposed STNet under closer interval, the angle interval between the two sources is reduced to 2°, and the SNR is reduced to 5 dB. The incident angle of the first source increases from $-59.63°$ to $57.37°$ in steps of 1°, while the incident angle of the second source increases from $-57.63°$ to $59.37°$, so there is a total of 118 test sample pairs. The DOA estimates of each method are presented in Fig. 6(a)-(f) respectively and Fig. 6(g)-(l) demonstrate the corresponding errors. Compared with other methods, the DL-based methods show better stability when the incident angles of two targets are very close with a low SNR. For the angle pairs tested, there are a large number of cases that cannot be estimated accurately by traditional methods. And when the incident direction of the signal near the edge of the angular region where the estimation is challenging, we can clearly observe that MUSIC, ANM, and ESPRIT will be more likely to produce large estimation

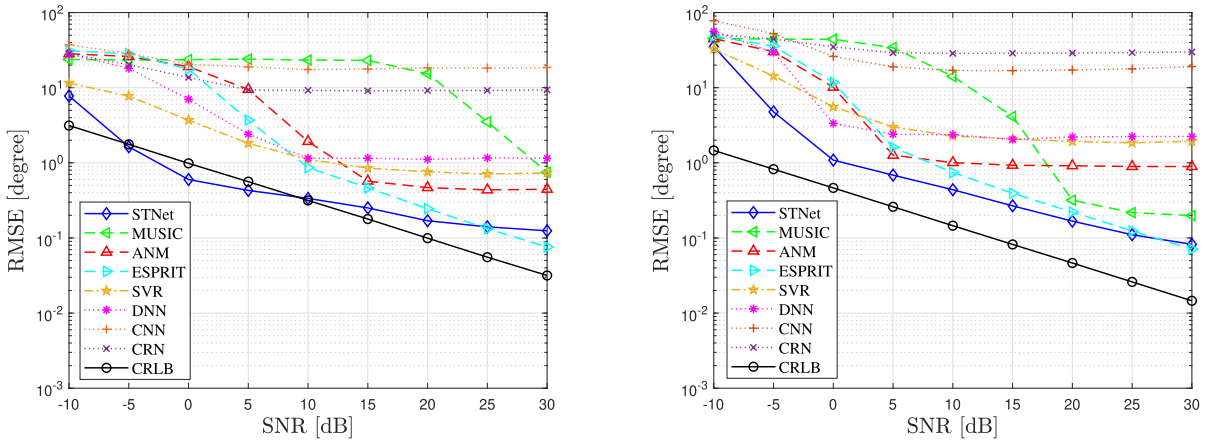


Fig. 7. The RMSE versus the SNR in the DOA estimation of two sources at (a) $\theta_1 = 2.89°$ and $\theta_2 = 5.44°$; (b) $\theta_1 = -58.77°$ and $\theta_2 = -50.23°$.

errors. For DNN and CNN methods, it is still impossible to estimate some sample points accurately because of the small number of snapshots. At this point, the STNet shows robust performance, retains the strong adaptability of the DL method to data defects, and controls the estimation error within a small range.

B. Statistical Performance

In this part of the experiment, we use Monte Carlo method to analyze the statistical performance of various DOA estimation algorithms, whereas the sample covariance is estimated by using $L = 10$ snapshots. Here, the root mean square error (RMSE) is selected as the evaluation index, which is defined as

$$\text{RMSE} = \sqrt{\frac{1}{KM}\sum_{k=1}^{K}\sum_{m=1}^{M}(\hat{\theta}_k^{(m)} - \theta_k^{(m)})^2}, \tag{29}$$

where K is the number of sources and M is total number of test samples. Herein, $\hat{\theta}_k^{(m)}$, $\theta_k^{(m)}$ represent the actual DOAs and estimated results, respectively.

1) RMSE Versus SNR: We first test the DOA estimation performance under different SNR conditions, the RMSE is calculated over 1,000 simulations at each SNR level. In the first scenario, we consider that two signals are incident on the array from $\theta_1 = 2.89°$ and $\theta_2 = 5.44°$, respectively. In the second scenario, we assume that the incident direction of the two signals are close to the boundaries of the angular region, i.e., $\theta_1 = -50.23°$ and $\theta_2 = -58.77°$. The RMSE results of both scenarios are shown in Fig. 7(a) and (b), respectively. In addition, the Cramé-Rao lower bound (CRLB) [42] curve is provided as a benchmark. Basically, when the target DOAs locate at the boundary of the angle search range, it greatly increases the difficulty of DOA estimation, especially in the low-SNR regime. In both scenarios, it can be seen that the STNet has robust performance compared with other methods, especially under the low-SNR conditions. The CNN method is restricted by the few snapshots due to the use of the real covariance matrix, even as the SNR increases the performance has no obvious improvements. The CRN method is similar, the Toeplitz matrix cannot be recovered well with only a few snapshots, so it shows a large estimation error, especially at the edge of the angle search range. The performance of DNN method is much better, and its RMSE is greater than 1 degree when estimating two off-grid sources in high SNRs. Moreover, as an on-grid method, its performance has an upper limit and does not improve under high SNRs. In contrast, STNet introduces the TCEM module, extracts the features of the original sampled received signal through the LSTM layers and uses them for subsequent DOA estimation, supplementing the integrity of information, so it has better performance than DNN under the test conditions shown in the figure. For the ANM method, the improvement of SNR can enhance its resolution for targets with small angular interval. The SVR method is similar to it, but the performance of SVR is better in low SNRs. In the high-SNR regime, the ANM method demonstrates enhanced effect, the RMSE can be reduced to about 0.4 degree, but it cannot improve its estimation performance in the boundary of the angle search range. For the STNet, when the SNR is between -10 dB to 20 dB, the smallest error is realized among all the methods in both cases, including the traditional methods and DL based methods. Although the subsequent performance improvement with the increase of SNR is not as large as that of the ESPRIT method, it does not have an obvious upper limit as other grid-based methods, and implements a gridless estimation. Finally, in both cases, compared with the CRLB curve, all other methods use small snapshots for DOA estimation show a significant difference compared to the CRLB. Only the proposed STNet method has the smallest error in most cases and has the closest performance to the CRLB benchmark. And in Fig. 7(a), when SNR is between -5 dB and 5 dB, STNet even surpasses CRLB. The possible reason is that it is a data-driven end-to-end DL-based estimator and is not unbiased.

2) RMSE Versus Angle Interval $\Delta\theta$: In this setup, we evaluate the performance of the proposed framework for two sources with different angle intervals. The setting of the angle interval is $\Delta\theta = [1°, 1.2°, 1.5°, 2°, 4°, 6°, 8°]$. For each angle interval, there are 1,000 angle pair samples randomly generated, where the DOA of the first source θ_1 is within the region $[-60°, 60° - \Delta\theta]$ and the DOA of the second source is $\theta_2 = \theta_1 + \Delta\theta$. In other words, our test sample covers the entire angular region. The experiment was conducted under SNR $=$ 20 dB, and the RMSE curve versus the angular interval is

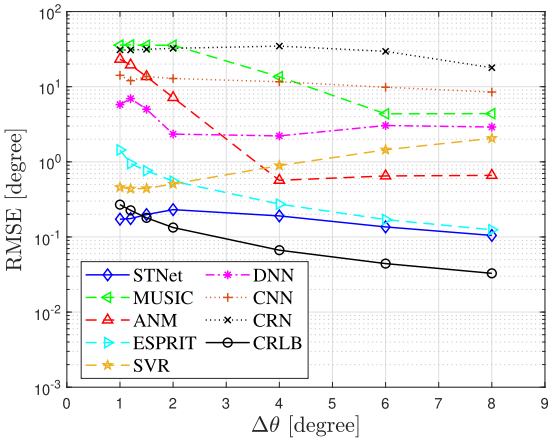


Fig. 8. The RMSE versus the angular interval $\Delta\theta$ in the DOA estimation of two sources when SNR = 20 dB.

illustrated in Fig. 8. As can be seen, for close angle pairs ($1° < \Delta\theta < 2°$), the STNet is able to resolve the angle estimations, as opposed to the rest of the methods, which cannot estimate the DOAs in a high precision. Since the SCDB module is used to retain the low-level features in the input that can be used to better distinguish the adjacent targets, the change of the angular interval between the targets can hardly affect the performance of STNet. Among several non-DL methods, MUSIC method performs worst when the target angle interval is within [1°, 4°], while the other two gridless methods, ANM and ESPRIT, also show significant performance degradation. However, the SVR method shows relatively good performance in the case of small angle interval. Among other DL-based methods, DNN performs the best. CNN and CRN still suffer from serious performance degradation in the case of small snapshots due to the use of real covariance matrix and Toeplitz structure, respectively. The angle interval change between targets has little effect on the DOA estimation performance of the STNet. The STNet outperforms the CRLB benchmark in cases of small angle intervals, due to the characteristics of deep learning methods. In general, the STNet performs the most accurate statistical estimation for each tested angle pair, particularly for the DOA estimation of targets within small angle intervals, achieving superior super-resolution results.

C. Performance of Multi-Target DOA Estimation

In the previous two-target situation, the advantages of the STNet over other DL methods have been fully explained. In practical applications, it may be occurred to face with DOA estimation of multiple targets. Neither the DNN method in [28] nor the CNN method in [31] have studied the DOA estimation cases of more than three targets. In this context, we only use the traditional MUSIC method [4], ANM method [19], [21], [22], and ESPRIT method [5] for a comparison.

1) DOA Estimation of Three Targets: First, Fig. 9 shows the DOA estimates and the corresponding estimation errors of three 20 dB sources in the sector $[-60°, 60°]$. Among them, the first and second target interval is fixed at 10.17°, and the second and third target interval is fixed at 20.6°. The DOA of the first target varies from −59.47° to 28.53°, so a total of 89 samples are considered. It can be observed from Fig. 9 that the estimation performance of the STNet is slightly worse than that of ESPRIT method. But the error difference is not so large that it can make a relatively stable and correct estimation. Then, we reduce both the SNR and the angular interval between targets. The angle interval between the first and second targets is reduced to 2.33°, the interval between the second and third targets is reduced to 4.57°, and the SNR is reduced to 5 dB. The DOA estimation results and the corresponding errors of the aforementioned methods are shown in Fig. 10. We can clearly observe that the STNet shows significant advantages over the traditional method at this time. It realizes a robust performance for all the samples while the other three methods cannot estimate DOA accurately for almost half of the samples. Therefore, when the difference between target DOAs are small and the SNR is low, the STNet still outperform the other methods.

Similarly, the experimental results of subsequent statistical characteristics further prove this, as shown in Fig. 11. First, three signals located in −30.26°, 0.89° and 30.77° are considered, the RMSE versus SNR curves are drawn in Fig. 11(a). Then, the signal DOAs are set to −2.26°, 0.89° and 3.77°, the same experiment is carried out and result is shown in Fig. 11(b). It indicates that when the target DOAs are far apart, the STNet has certain advantages over MUSIC method and it has similar performance with ANM method. Once the signal DOAs become very close, we can see that the estimation precision predominance of the STNet outperforms other methods, especially against MUSIC and ANM. Due to the inherent characteristics of deep learning methods, the estimation of STNet is not unbiased, and a limited number of training samples also suppress its performance [31]. Although it implements gridless estimation, its improvement in RMSE under high SNRs is restricted, particularly when compared to ESPRIT. However, what is more notable is that in most cases, STNet has already surpassed other methods and exhibits a significantly superior ability to resolve small angle interval targets. Only when the SNR is ideal can ESPRIT method be close to the CRLB benchmark. In contrast, regardless of the target angle interval, the proposed STNet has more advantages in low SNR situations and is closer to CRLB. That is to say, in more challenging situations with low SNR and small angle intervals, STNet is the optim choice. Therefore, the STNet can also have a good performance in DOA estimation for three targets, and retains its resolution advantage for small angle-spaced targets.

2) DOA Estimation of Four Targets: Finally, we test the DOA estimation performance of the proposed model in the case of four sources, where the interval between DOAs of four signals is set as 10°, 20°, and 30°, respectively. The results in Fig. 12 show that the STNet can also achieve good estimation performance in this case. Hence, only by changing the length of the output, it is quite convenient for the proposed STNet to estimate DOAs of different numbers of targets.

D. DOA Estimation Experiment With Real Data

In this section, we train and test the proposed STNet by measured data collected by Texas Instrument (TI)'s AWR2243

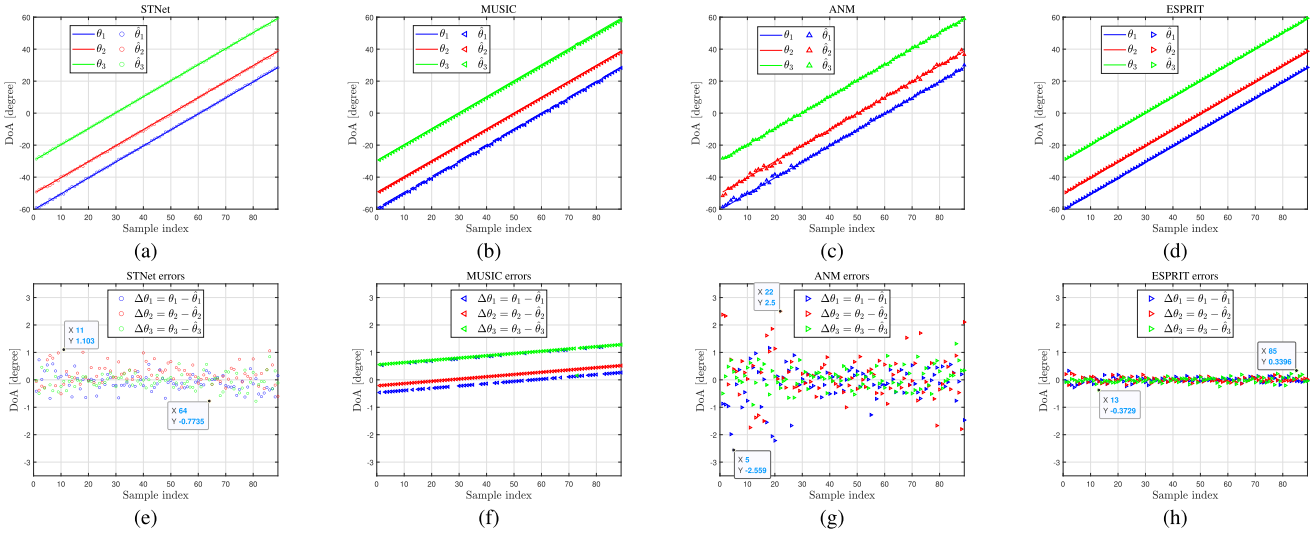


Fig. 9. DOA estimation performance of three targets with large angle intervals and SNR being 20 dB.

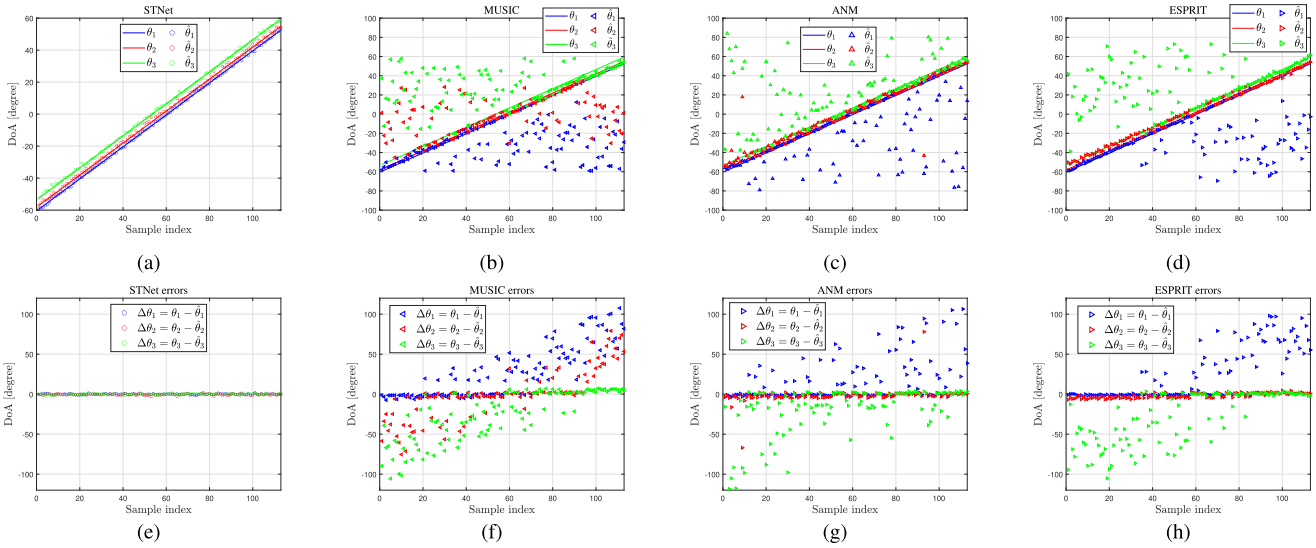


Fig. 10. DOA estimation performance of three targets with small angle intervals at 5 dB SNR.

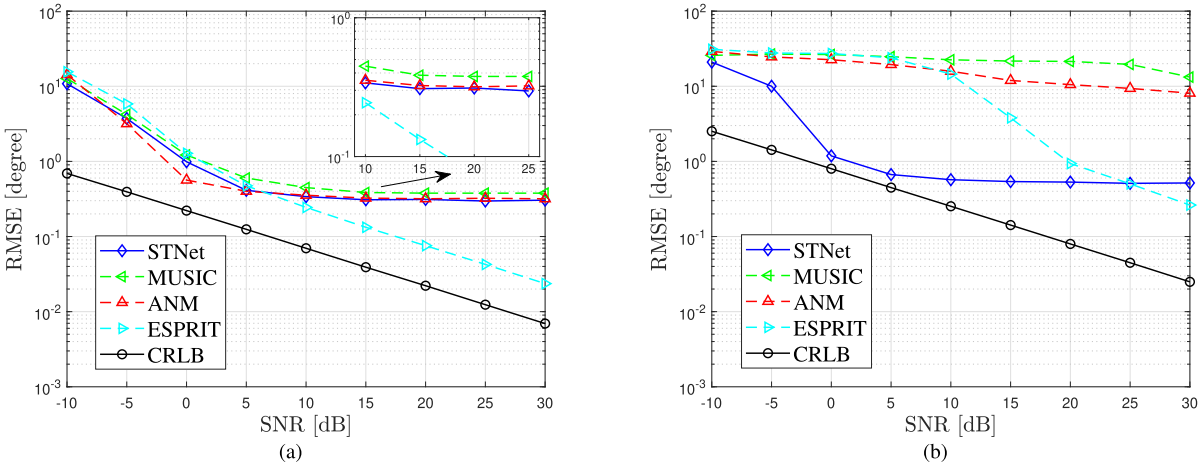


Fig. 11. The RMSE versus the SNR in the DOA estimation of three sources.

evaluation board. We adopted 3 Tx and 4 Rx for all collected data such that the resulting virtual array consists of $N = 12$ elements. The angle resolution is approximated to $\theta_{\text{res}} = \frac{\lambda}{Nd\cos\theta} = 9.5°$ for boresight $\theta = 0°$.

1) Experiment Results for the Indoor Environment: The first experiment was conducted in a microwave anechoic chamber, the equipment and environment are shown in Fig. 13. The distance between the evaluation board and the target plane is

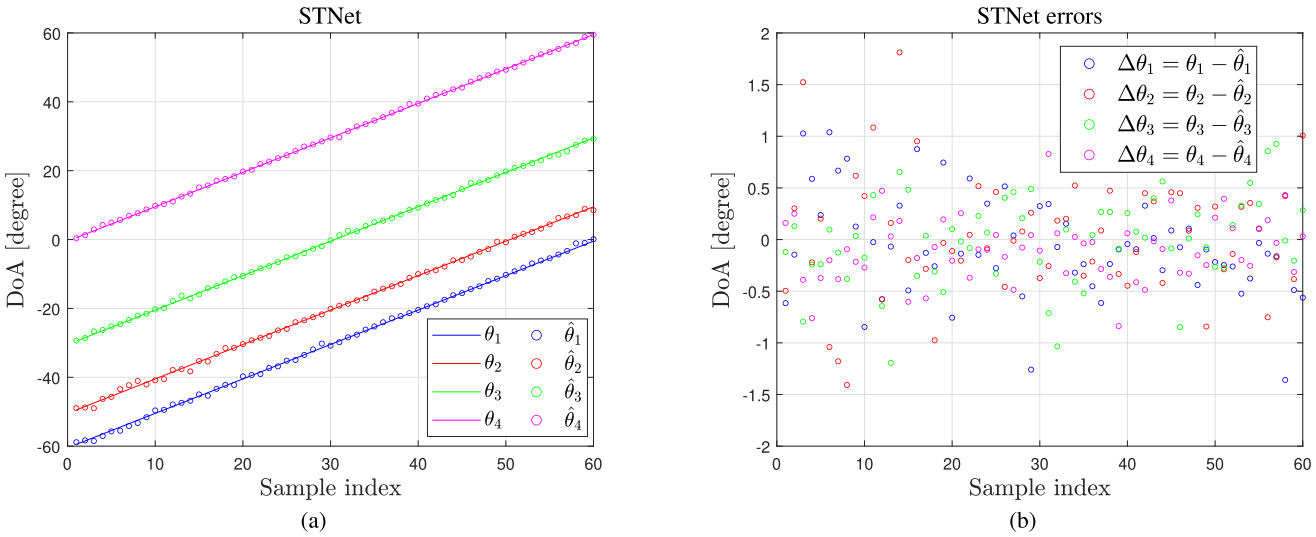


Fig. 12. DOA estimation performance of four targets.

TABLE III
RMSE ON REAL DATA

Target spacing	7 cm	15 cm	25 cm	35 cm
STNet	0.0077	0.0078	0.0087	0.0070
DNN	1.6459	1.4324	2.4106	3.2872
CRN	14.88	12.63	10.64	9.62

430 cm, the targets remained at a fixed distance and moved within a certain range on the plane. The moving range for the targets is ±100cm, where the angle corresponds to 0 cm is 0°. The two targets move in the range at four different distances, namely 7 cm, 15 cm, 25 cm and 35 cm. In these four cases, when the two targets are exactly located on both sides of the 0 cm position, i.e. ±3.5 cm, ±7.5 cm, ±12.5 cm, ±17.5 cm, the corresponding angles are ±0.47°, ±1°, ±1.67°, ±2.33°. Therefore, their angle intervals are 0.94°, 2°, 3.34°, 4.66°, but there will be slight changes in the angle interval as the two targets move. Therefore, the angle of the target will not always be on a series of fixed grid points. The other three DL-based methods, DNN in [28], CNN in [31] and CRN in [33], are also trained on the real data. However, the CNN has a serious gradient explosion phenomenon in the training process, and can not get useful models, so it will not participate in the subsequent comparison. In addition, for the CRN method, the ideal covariance matrix is constructed based on the true angle of the target.

We have obtained 336, 312, 288, 240 samples for each situation, and 240 groups of 10 snapshot data are collected for each samples. Finally, 75% of the data is used for training and 25% for testing, the RMSE of DNN and the proposed STNet on the test set is shown in Table III. As can be seen, the error of the proposed STNet in the test set is significantly smaller than that of the DNN method, and the RMSE in each case is less than 0.01°.

After that, we test STNet with two cases that are not included in the data set, and present the estimation of DNN. In addition, we also present the estimation results of DBF and MUSIC methods as a comparison. First of all, the two targets are placed at the position of −10 cm and 10 cm, and the corresponding angles are −1.33° and 1.33°, respectively. The estimated results of each method are shown in Fig. 14(a). After that, the targets are placed at the positions of −20 cm and 20 cm, and the corresponding angles are −2.67° and 2.67°. The DOA estimation performance of each method is shown in Fig. 14(b). Both cases are not included in the data samples collected for training, and the angle interval of the targets is smaller than the resolution of the array. It can be seen from the figure that the traditional DBF and MUSIC methods have a spectrum peak between the two targets, but it is difficult to distinguish the two closely spaced targets and make accurate angle estimation. Similarly, the DNN method is also difficult to obtain accurate DOA estimation results by using 10 snapshot data, and the peaks of the two targets in the probability spectrum show great differences. Moreover, because DNN is a grid-based method, it also leads to larger estimation errors. Observing the CRN model, it is challenging to make accurate estimates with a limited number of snapshots, and the resolution ability is low for targets with narrow intervals. For

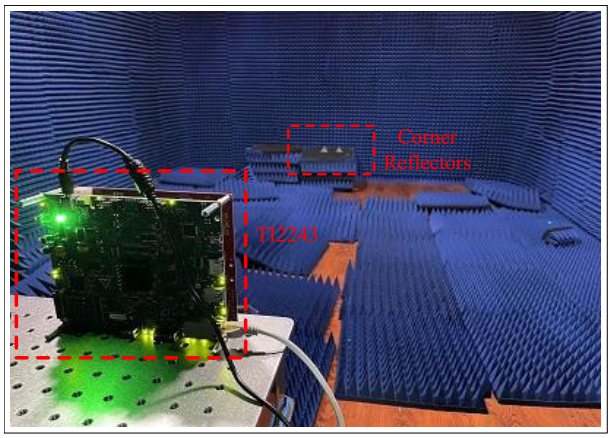


Fig. 13. Experimental equipment and environment.

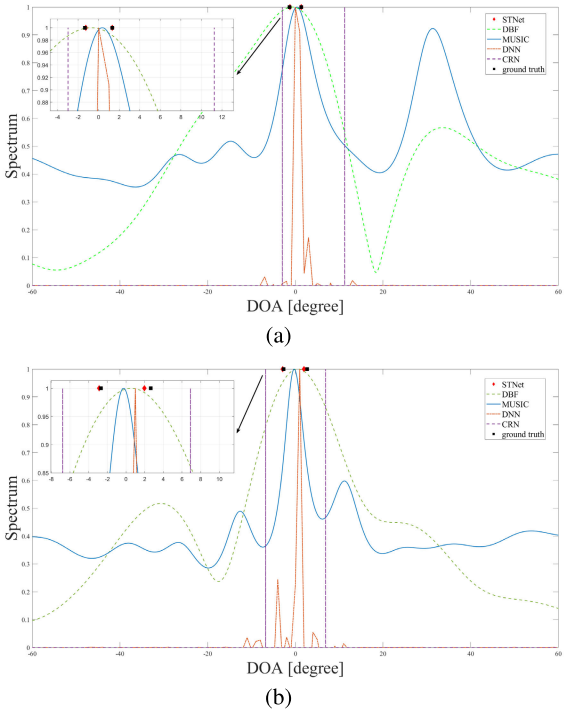


Fig. 14. The estimation results of two targets with real data.

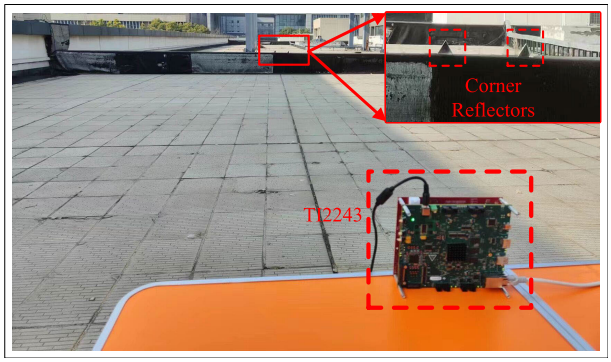


Fig. 15. Experimental equipment and environment.

the proposed STNet, the super-resolution DOA estimation of the targets is achieved in the case of small snapshots. It can distinguish the targets with small angle intervals and obtain the DOA estimation result very close to the ground truth, which is far superior to other methods.

2) Experiment Results for the Outdoor Environment: The second experiment was conducted in an outdoor environment, the experimental environment, equipment, and target corner reflectors are shown in Fig. 15. In this outdoor scene, there is increased clutter reflection and the targets are positioned on a wall, leading to a significantly lower SNR in comparison to indoor scenes. Here, the two targets move within a range of $\pm 10°$, with five different angular intervals maintained, namely 0.5°, 1°, 2°, 4°, 7°. For these five scenarios, we collected 30, 30, 30, 20, and 20 different angle pairs, respectively, and each angle pair case included 1,056 data samples with a length of 10 snapshots. All of the collected samples were used for

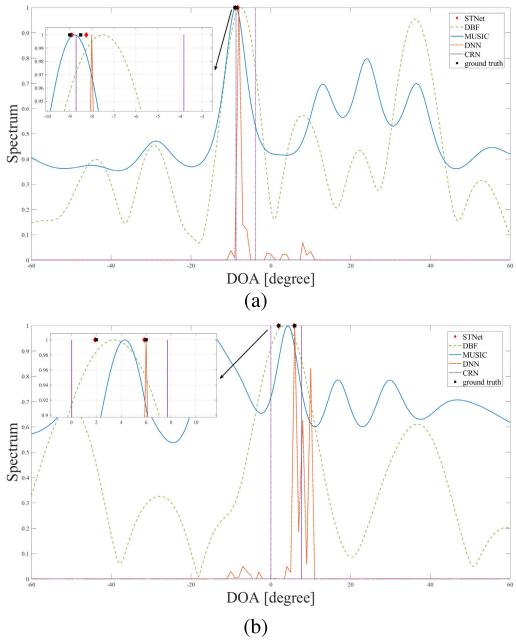


Fig. 16. The estimation results of two targets with real data.

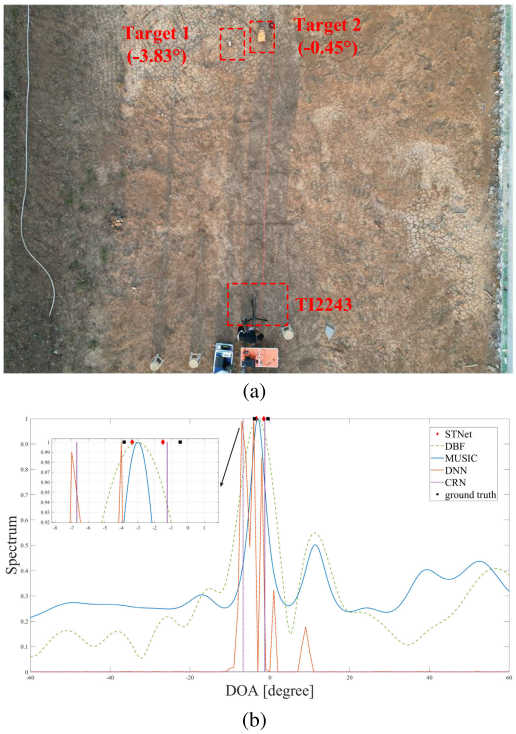


Fig. 17. DOA estimation for outdoor environment. (a) Experimental scene. (b) Estimated results.

training the STNet, and additional angle pairs not included in the training set were collected as test samples for evaluation.

The estimation results of the proposed STNet for two different angle pairs are shown in Fig. 16(a) and Fig. 16(b). These two angle pairs are not included in the training set, namely $(-9°, -8.5°)$ and $(2°, 6°)$. The estimates of DBF, MUSIC, and DNN methods are also presented for comparison. The results demonstrate that the estimated spectra of the traditional DBF and MUSIC methods only exhibit a single

peak near the DOAs of the two targets, making it difficult to distinguish between them. The DNN method also fails to differentiate between the two targets with a 0.5° interval and exhibits significant errors in the second test sample with a 4° interval. The CRN method has improved the performance of DOA estimation in this scenario. However, when two targets are very close together, only one of them can be estimated, resulting in significant errors. When the distance between two targets is large, the estimation error is also around 2°, which is not satisfactory. In contrast, the proposed STNet maintains excellent estimation performance in outdoor environments, even when the angle interval between the two targets is very small, with minimal estimation errors. Subsequently, for each different angle interval, we selected an angle pair not included in the training set and collected 350 sets of samples for RMSE testing. According to the test results, STNet performs DOA estimation at five different angle intervals of 0.5°, 1°, 2°, 4°, 7°, with RMSE being 0.62°, 0.61°, 0.97°, 0.86°, 1.17°, respectively.

Subsequently, the model underwent testing in an alternate outdoor environment. The objective was to determine the DOAs of two groups of bricks. These bricks have lower reflection energy, indicating a worse SNR condition. The entire scene can be seen in Fig. 17(a), while the estimation results for each method are displayed in Fig. 17(b). Among them, both the DBF and MUSIC methods fail to distinguish between the two targets. The two highest peaks of the DNN method estimation result correspond to $-7°$ and $-4°$, the estimated CRN method results are $-6.69°$ and $-1.24°$, while the angle estimation results of STNet are $-3.34°$ and $-1.47°$. Overall, in practical scenarios, STNet can achieve excellent DOA estimation performance with a small number of snapshots and demonstrates good robustness in low SNR situations.

VI. Conclusion

In this paper, we proposed a new deep learning network framework for super-resolution DOA estimation in the case of small snapshots. We used the original sampled signal and the estimated covariance matrix of the received signal as the input, the SCEM and TCEM modules were designed to process the data of these two parts respectively. Furthermore, we proposed the SCDB structure to improve the network's resolution of small interval targets, and use the LSTM structure to extract time correlation. After that, the features extracted by the two modules were processed through a series of FC layers, the STNet modeled DOA estimation as a regression problem, which can achieve the effect of gridless estimation. Therefore, different features are obtained and combined for DOA estimation. On the one hand, it could provide more information for DOA estimation with small snapshots, and on the other hand, it could improve the accuracy of regression tasks by providing more information. Experimental results showed that, compared with the previous DL methods, the STNet achieved the gridless estimation effect, and the estimation accuracy was not limited by the grid resolution. It showed a more advantageous estimation effect in the case of small snapshots. Meanwhile, it retained the advantages of DL-based methods, has stronger adaptability and stability than conventional methods, and could still maintain good estimation performance in the case of very close signal DOAs and low SNRs.

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.

[2] E. Aboutanios and A. Hassanien, "Low-cost beamforming-based DOA estimation with model order determination," in *Proc. IEEE 11th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2020, pp. 1–5.

[3] 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.

[4] R. O. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, 1981.

[5] R. Roy, A. Paulraj, and T. Kailath, "Estimation of signal parameters via rotational invariance techniques–ESPRIT," in *Proc. MILCOM IEEE Mil. Commun. Conf., Commun.-Comput., Teamed 90's*, Oct. 1986, p. 41.

[6] C. Ying, W. Xiang, and H. Zhitao, "Underdetermined DOA estimation via multiple time-delay covariance matrices and deep residual network," *J. Syst. Engr. Electron*, vol. 32, no. 6, pp. 1354–1363, Dec. 2021.

[7] B. Li, Y. Zou, and Y. Zhu, "Direction estimation under compressive sensing framework: A review and experimental results," in *Proc. IEEE Int. Conf. Inf. Autom.*, Jun. 2011, pp. 63–68.

[8] Y. Ling, H. Gao, S. Zhou, L. Yang, and F. Ren, "Robust sparse Bayesian learning-based off-grid DOA estimation method for vehicle localization," *Sensors*, vol. 20, no. 1, p. 302, Jan. 2020.

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

[10] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, Jul. 2007.

[11] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, Feb. 2011.

[12] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.

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

[14] A. Gretsistas and M. D. Plumbley, "An alternating descent algorithm for the off-grid DOA estimation problem with sparsity constraints," in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 874–878.

[15] H. Duan, Z. Qian, and Y. Wang, "Off-grid DOA estimation based on noise subspace fitting," in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2015, pp. 675–678.

[16] J. Dai, X. Bao, W. Xu, and C. Chang, "Root sparse Bayesian learning for off-grid DOA estimation," *IEEE Signal Process. Lett.*, vol. 24, no. 1, pp. 46–50, Jan. 2017.

[17] Q. Wang, Z. Zhao, Z. Chen, and Z. Nie, "Grid evolution method for DOA estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2374–2383, May 2018.

[18] Z. Chen, P. Chen, Z. Guo, Y. Zhang, and X. Wang, "A RIS-based vehicle DOA estimation method with integrated sensing and communication system," *IEEE Trans. Intell. Transp. Syst.*, early access, 2024.

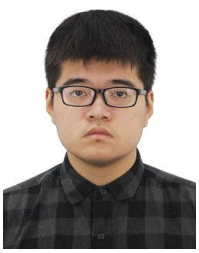
[19] P. Chen, Z. Chen, Z. Cao, and X. Wang, "A new atomic norm for DOA estimation with gain-phase errors," *IEEE Trans. Signal Process.*, vol. 68, pp. 4293–4306, 2020.

[20] Y. Jiang, D. Li, X. Wu, and W.-P. Zhu, "A gridless wideband DOA estimation based on atomic norm minimization," in *Proc. IEEE 11th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2020, pp. 1–5.

[21] Y. Zhang, G. Zhang, and H. Leung, "Atomic norm minimization methods for continuous DOA estimation in colored noise," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Sep. 2019, pp. 1–5.

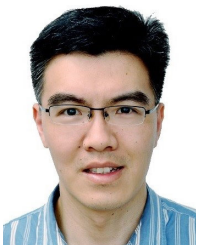
[22] M. Wagner, Y. Park, and P. Gerstoft, "Gridless DOA estimation and root-MUSIC for non-uniform linear arrays," *IEEE Trans. Signal Process.*, vol. 69, pp. 2144–2157, 2021.

[23] D. Jin-Xiang, F. Xi-An, and M. Yan, "DOA esitmation based on support vector machine—Large scale multiclass classification problem," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Sep. 2011, pp. 1–4.
[24] M. Pastorino and A. Randazzo, "A smart antenna system for direction of arrival estimation based on a support vector regression," *IEEE Trans. Antennas Propag.*, vol. 53, no. 7, pp. 2161–2168, Jul. 2005.
[25] A. El Gonnouni, M. Martinez-Ramon, J. L. Rojo-Alvarez, G. Camps-Valls, A. R. Figueiras-Vidal, and C. G. Christodoulou, "A support vector machine MUSIC algorithm," *IEEE Trans. Antennas Propag.*, vol. 60, no. 10, pp. 4901–4910, Oct. 2012.
[26] M. I. Hasan and M. Saquib, "Low complexity single source 2-D DOA estimation based on reduced dimension SVR," in *Proc. IEEE 22nd Annu. Wireless Microw. Technol. Conf. (WAMICON)*, Apr. 2022, pp. 1–4.
[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 7553.
[28] Y. Kase, T. Nishimura, T. Ohgane, Y. Ogawa, D. Kitayama, and Y. Kishiyama, "DOA estimation of two targets with deep learning," in *Proc. 15th Workshop Positioning, Navigat. Commun. (WPNC)*, Oct. 2018, pp. 1–5.
[29] Z.-M. Liu, C. Zhang, and S. Y. Philip, "Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections," *IEEE Trans. Antennas Propag.*, vol. 66, no. 12, pp. 7315–7327, Dec. 2018.
[30] L. Wu, Z.-M. Liu, and Z.-T. Huang, "Deep convolution network for direction of arrival estimation with sparse prior," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1688–1692, Nov. 2019.
[31] 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.
[32] A. M. Elbir, "DeepMUSIC: Multiple signal classification via deep learning," *IEEE Sensors Lett.*, vol. 4, no. 4, pp. 1–4, Apr. 2020.
[33] 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.
[34] Z. Wang, Y. Fu, and T. S. Huang, "Deep learning through sparse and low-rank modeling," in *Computer Vision and Pattern Recognition*. New York, NY, USA: Academic, 2019, pp. 1–7.
[35] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, Jan. 1989.
[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[37] F. Wang et al., "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6450–6458.
[38] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4809–4817.
[39] N. Bershad and A. Rockmore, "On estimating signal-to-noise ratio using the sample correlation coefficient (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 1, pp. 112–113, Jan. 1974.
[40] H. Sadreazami, M. Bolic, and S. Rajan, "On the use of ultra wideband radar and stacked LSTM-RNN for at home fall detection," in *Proc. IEEE Life Sci. Conf. (LSC)*, Oct. 2018, pp. 255–258.
[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, New York, NY, USA: Academic, 2014, pp. 1–11.
[42] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer–Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.



Yanjun Zhang received the B.S. degree in electronic information science and technology from Shandong University. He is currently pursuing the Ph.D. degree in electronic and information engineering with the State Key Laboratory of Millimeter Waves, Southeast University. His research interests include signal processing, millimeter wave radar, target detection, and machine learning.

Yan Huang (Member, IEEE) received the B.S. degree in electrical engineering and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2013 and 2018, respectively. He was a Visiting Ph.D. Student with the Electrical and Computer Engineering Department, University of Florida, from September 2016 to July 2017, and the Electrical and Systems Engineering Department, Washington University in St. Louis, from July 2017 to August 2018. He is currently an Associate Professor with the State Key Laboratory of Millimeter Waves, Southeast University. His research interests include machine learning, synthetic aperture radar, image processing, and remote sensing.



Jun Tao (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the Department of Radio Engineering, Southeast University, Nanjing, China, in 2001 and 2004, respectively, and the Ph.D. degree in electrical engineering from the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO, USA, in 2010. From 2004 to 2006, he was a System Design Engineer with Realsil Microelectronics Inc., (a subsidiary of Realtek), Suzhou, China. From 2011 to 2015, he was a Senior System Engineer with Qualcomm Inc., Boulder, CO, USA, working on the baseband algorithm and architecture design for the UMTS/LTE modem. Since April 2016, he has been with the School of Information Science and Engineering, Southeast University, as a Full Professor. His research interests include wireless cellular communications, underwater acoustic communications, and localization and tracking, including channel modeling and estimation, turbo equalization, adaptive filtering, Bayesian inference, and machine learning.

Cai Wen (Member, IEEE) received the B.E. degree from the School of Electronic Engineering, Xidian University, Xi'an, China, in 2009, and the Ph.D. degree from the National Laboratory of Radar Signal Processing, Xidian University, in 2014. He was a Research Scientist with China Norinco Group, Beijing, China, from January 2015 to October 2016. Since November 2016, he has been with the School of Information Science and Technology, Northwest University, Xi'an, where he is currently an Associate Professor. In November 2019, he became a full-time Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada. His current research interests include sensor array signal processing, multiple input multiple output (MIMO) radar signal processing, integrated radar and communication, and mathematical optimization.

Yu Han received the Ph.D. degree from Delft University of Technology in 2017. He is currently an Associate Professor with the School of Transportation, Southeast University. He has authored numerous articles. His research expertise involves traffic flow theory, traffic simulation, and road traffic control. He is interested in employing interdisciplinary approaches to manage and control traffic flows, including techniques, such as model predictive control, reinforcement learning, and parallel learning that integrate traffic domain knowledge with data-driven learning. He also served as a reviewer for prestigious journals within the fields of traffic and transportation, including *Transportation Research Part C* and IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.

Guisheng Liao (Senior Member, IEEE) was born in Guilin, Guangxi, China, in 1963. He received the B.S. degree in mathematics from Guangxi University, Guangxi, in 1985, and the M.S. degree in computer software and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 1990 and 1992, respectively.

He has been the First Dean of Hangzhou Institute of Technology, Xidian University, since 2021, where he was the Dean of the School of Electronic Engineering, from 2013 to 2021. He was a Senior Visiting Scholar with The Chinese University of Hong Kong, from 1999 to 2000. He is currently a Full Professor with the National Key Laboratory of Radar Signal Processing. His research interests include array signal processing, space-time adaptive processing, radar waveform design, airborne/space surveillance, and warning radar systems. He won the National Science Fund for Distinguished Young Scholars in 2008.

Wei Hong (Fellow, IEEE) received the B.S. degree from the University of Information Engineering, Zhengzhou, China, in 1982, and the M.S. and Ph.D. degrees from Southeast University, Nanjing, China, in 1985 and 1988, respectively, all in radio engineering.

Since 1988, he has been with the State Key Laboratory of Millimeter Waves, where he has been the Director, since 2003. He is currently a Professor with the School of Information Science and Engineering, Southeast University. In 1993 and from 1995 to 1998, he was a short-term Visiting Scholar with the University of California at Berkeley and the University of California at Santa Cruz, Santa Cruz, respectively. He has been engaged in numerical methods for electromagnetic problems, millimeter wave theory and technology, antennas, and RF technology for wireless communications. He has authored and coauthored over 300 technical publications and two books. He is a fellow of CIE, the Vice President of the CIE Microwave Society and Antenna Society, and the Chair of the IEEE MTT-S/AP-S/EMC-S Joint Nanjing Chapter, and was an elected IEEE MTT-S AdCom Member, from 2014 to 2016. He was twice awarded the National Natural Prizes and thrice awarded the First-Class Science and Technology Progress Prizes issued by the Ministry of Education of China and Jiangsu Province Government. Besides, he also received the Foundations for China Distinguished Young Investigators and for "Innovation Group" issued by NSF of China. He served as an Associate Editor for IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES from 2007 to 2010.