Advancing Subspace Representation Learning for DOA Estimation Using Sparse Arrays

Ruonan Yang[†], Yunqiao Hu[†], Shunqiao Sun[†], and Yimin D. Zhang[§]

[†]Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL 35487
[§]Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122

***Abstract*—We present a Siamese neural network (SNN) framework for direction-of-arrival (DOA) estimation that learns robust subspace embeddings directly from sample covariance matrices (SCMs). Classical subspace methods such as MUSIC and ESPRIT degrade under limited snapshots, low signal-to-noise ratio (SNR), and mutual coupling, while existing deep neural network (DNN) approaches often require large training datasets and have limited generalization ability. The proposed SNN consists of two convolutional encoder branches with shared parameters and is trained on pairs of SCMs labeled as similar (same DOAs under varying SNR, reflection coefficients, or mutual coupling) or dissimilar otherwise. A joint loss combining contrastive embedding subspaces and angle estimation encourages compact clustering of embedding representations associated with identical DOAs. During inference, the model uses a single branch of the convolutional encoder, after which a polynomial root-finding is applied to obtain the final DOA estimates. Simulation results show that the proposed method outperforms classical subspace techniques and existing DNN-based estimators, achieving improved accuracy across a range of SNR and mutual coupling conditions while requiring substantially fewer training samples.**

***Index Terms*—Siamese neural network, Direction-of-arrival (DOA) estimation, Contrastive learning, Sparse arrays, Deep neural network (DNN)**

I. Introduction

Direction-of-arrival (DOA) estimation is a fundamental task in array signal processing and is essential for accurate source localization. Classical subspace-based methods such as MUSIC [1] and ESPRIT [2] exploit the eigenstructure of the sample covariance matrix (SCM) for uniform linear arrays (ULAs). To reduce hardware cost and mitigate mutual coupling, sparse linear arrays (SLAs) are often employed by selecting a subset of ULA elements, achieving the same aperture with a much smaller number of nonuniformly spaced sensors. Regardless of whether grid-based or gridless DOA estimators are used, the quality of the signal and noise subspaces derived from the SCM is critical, as these subspaces encode the array manifold and the underlying DOA information. However, their performance can degrade severely under practical challenges such as low signal-to-noise ratio (SNR), limited snapshots, mutual coupling effects, and random sensor failures [3], [4], [5].

Most existing SCM-based DOA estimation methods for sparse arrays rely on covariance reconstruction or covariance fitting. For grid-based approaches, [6] proposed a compressed covariance sensing framework that designs samplers and reconstructs the full covariance matrix from SLA measurements. For co-array based reconstruction, [7] applied spatial smoothing to the nested co-array to obtain a virtually filled ULA covariance matrix, yielding many more degrees of freedom than the available physical sensors. For gridless approaches, [8] introduced a structured covariance recovery method via atomic norm minimization, solvable using semidefinite programming (SDP). A related gridless method for SLAs was presented in [9], which estimates a structured covariance matrix using the Bures–Wasserstein distance and solves it via SDP and gradient descent. Furthermore, [9] proposed an irregular root-MUSIC algorithm based on irregular Vandermonde decomposition to extract DOAs from the reconstructed noise subspace.

Recent advances in deep learning have shown strong potential for DOA estimation [10], where networks learn complex representations directly from data without explicit covariance reconstruction. In [11], a purely data-driven convolutional neural network (CNN) was proposed to estimate DOAs from SCMs, outperforming classical subspace methods in low-SNR scenarios. Hybrid model-based data-driven approaches such as DA-MUSIC [12, 13] replace the SCM construction and peak search stages of MUSIC with neural network modules, yielding improved accuracy for broadband and coherent sources. More recently, SubspaceNet [14] demonstrated that learning enhanced covariance representations via an autoencoder enables classical algorithms such as MUSIC and root-MUSIC to remain effective under challenging conditions including coherence, low SNR, array mismatches, and limited snapshots.

However, DNN-based methods typically require extremely large training datasets to capture variations in SNR, snapshot numbers, and target reflection coefficients. Many existing approaches also depend on distance metrics between covariance matrices, which add complexity to the learning process. Moreover, as highlighted in [15], learning representations of the signal and noise subspaces is more essential than learning the SCM itself, since these subspaces are directly tied to subspace-based DOA estimation algorithms, whereas accurately estimating the true covariance matrix is inherently more difficult under diverse operating conditions.

Motivated by these challenges, we propose a Siamese neural network (SNN) framework [16], [17] specifically for subspace representation learning in DOA estimation. Comprising two identical encoder branches, the SNN is optimized to discern similarity relationships within data pairs. During training (see Figures 1 and 2), the weight-sharing encoders process pairs of SCMs that correspond to the same DOAs but distinct reflection coefficients and noise conditions. These inputs are subsequently mapped into their corresponding embedded subspaces via the encoder network.

The contrastive loss function [18] is applied to learn the embedding space representations, ensuring that similar pairs are close in the learned representation while dissimilar pairs are pushed apart. Additionally, a downstream DOA estimation block is used to estimate the DOAs from the subspaces obtained from the learned representation. This DOA estimation block can be either a traditional subspace algorithm, such as root-MUSIC, or a neural network. The training loss function combines the contrastive loss with the root-mean-square error (RMSE) for DOA estimation.

The work of R. Yang, Y. Hu and S. Sun was supported in part by U.S. National Science Foundation (NSF) under Grants CCF-2153386 and ECCS-2033433. The work of Y. D. Zhang was supported in part by NSF under Grant ECCS-2236023.

Simulation results for ULA and SLA under a range of diverse SNR and snapshot conditions validate the effectiveness of our proposed approach. Compared to traditional subspace methods and conventional deep neural network (DNN) models, the proposed method achieves significantly lower RMSE in the estimated DOAs and requires substantially fewer training samples, while maintaining robust performance across a wide range of signal scenarios.

II. Problem Formulation

We consider the problem of DOA estimation for far-field narrowband signals impinging on an M-element linear array. The sensor locations are denoted as $\mathbf{d} = [d_1, d_2, \cdots, d_M]^T$, where $(\cdot)^T$ denotes transpose. Assume that K incoherent signals arrive from directions $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_K]^T$. Then, the observed data of the M-element array are modeled as:

$$\mathbf{y}(t) = \mathbf{C}_{\text{cpl}}\mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t), \quad t = 1, 2, \cdots, T, \tag{1}$$

where $\mathbf{s}(t) \in \mathbb{C}^{K\times 1}$ contains the source signals, and $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), ..., \mathbf{a}(\theta_K)] \in \mathbb{C}^{M\times K}$ is the array steering matrix with steering vector

$$\mathbf{a}(\theta_k) = [e^{-j\frac{2\pi}{\lambda}d_1 \sin\theta_k}, ..., e^{-j\frac{2\pi}{\lambda}d_M \sin\theta_k}]^T \in \mathbb{C}^{M\times 1}. \tag{2}$$

The noise vector $\mathbf{n}(t) \sim \mathcal{CN}(\mathbf{0}_M, \sigma_n^2\mathbf{I}_M)$ represents complex additive white Gaussian noise that is independent and identically distributed (i.i.d.) and has zero mean and variance σ_n^2 per array element. Here, T denotes the number of temporal snapshots.

To incorporate array mutual coupling effects, we adopt a commonly used Toeplitz-like coupling matrix [19]:

$$\mathbf{C}_{\text{cpl}} = \begin{bmatrix} 1 & c & c^2 & \cdots & c^{M-1} \\ c & 1 & c & \cdots & c^{M-2} \\ c^2 & c & 1 & \cdots & c^{M-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c^{M-1} & c^{M-2} & c^{M-3} & \cdots & 1 \end{bmatrix},$$

where c is the coupling coefficient. When adjacent sensors satisfy $\Delta d_i = |d_i - d_{i-1}| = \lambda/2$, the array becomes a ULA; non-uniform element spacings yield a non-uniform linear array (NULA), and selecting only a small subset of ULA elements forms an SLA.

A. Uniform Linear Arrays

Stacking T-snapshot measurements from (1) as $\mathbf{Y} = [\mathbf{y}(1), \cdots, \mathbf{y}(T)] \in \mathbb{C}^{M\times T}$, the sample covariance matrix is constructed as:

$$\hat{\mathbf{R}} = \frac{1}{T}\mathbf{Y}\mathbf{Y}^H \in \mathbb{C}^{M\times M}, \tag{3}$$

where $(\cdot)^H$ denotes conjugate transpose. For subspace-based DOA estimation methods such as MUSIC, root-MUSIC, and ESPRIT, eigenvalue decomposition (EVD) is applied on the sample covariance matrix $\hat{\mathbf{R}}$ as:

$$\hat{\mathbf{R}} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^H, \tag{4}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues in descending order, and the columns of $\mathbf{E}$ are the corresponding eigenvectors. The eigenvectors associated with the largest $\hat{K}$ eigenvalues form the signal subspace $\mathbf{E}_s \in \mathbb{C}^{M\times\hat{K}}$, while the remaining $(M - \hat{K})$ eigenvectors span the noise subspace $\mathbf{E}_n \in \mathbb{C}^{M\times(M-\hat{K})}$. $\hat{K}$ is typically estimated using model order selection criteria such as the Akaike information criterion (AIC), minimum description length (MDL), or Bayesian information criterion (BIC) [20].

Since the steering vector corresponding to a true DOA lies in the signal subspace in the noiseless case, and the signal subspace is orthogonal to the noise subspace, the DOAs can be estimated by solving:

$$\hat{\theta} = \arg\min_{\theta} \left\|\mathbf{a}^H(\theta)\mathbf{E}_n\right\|. \tag{5}$$

The MUSIC algorithm replaces (5) with a pseudospectrum peak search:

$$p_{\text{MUSIC}}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\,\mathbf{E}_n\mathbf{E}_n^H\,\mathbf{a}(\theta)}, \tag{6}$$

and the DOAs correspond to the angles that maximize $p_{\text{MUSIC}}(\theta)$. To avoid peak searching on discrete angle grids, the root-MUSIC algorithm reformulates (5) via polynomial root finding. Let $\mathbf{P}_n = \mathbf{E}_n\mathbf{E}_n^H$ denote the noise subspace projection matrix. The root-MUSIC polynomial is constructed as:

$$\begin{aligned} G(z) &= [z^{-d_1}, ..., z^{-d_M}]\mathbf{P}_n[z^{d_1}, ..., z^{d_M}]^T \\ &= \sum_{m=1}^{M}\sum_{n=1}^{M} p_{m,n} z^{d_m - d_n}, \end{aligned} \tag{7}$$

where $z = e^{j\pi\sin(\theta)}$ and $p_{m,n}$ denotes the (m, n)-th entry of $\mathbf{P}_n$. The $\hat{K}$ roots of $G(z)$ that lie closest to the unit circle correspond to the DOAs of signal. Each valid root z_k yields DOA estimate:

$$\widehat{\theta}_k = \arcsin\left(\frac{\lambda}{2\pi d}\angle z_k\right), \quad k = 1, \ldots, \hat{K}. \tag{8}$$

B. Sparse Linear Arrays

For an SLA, the sparse-array measurements are obtained by applying a diagonal matrix

$$\mathbf{D} = \text{diag}(\mathbf{1}_{d_i}), \forall i \in \{1, 2, ..., M\} \tag{9}$$

to the ULA outputs, where the binary mask is defined as

$$\mathbf{1}_{d_i} = \begin{cases} 1, & d_i \in \mathcal{D}, \\ 0, & d_i \notin \mathcal{D}, \end{cases} \tag{10}$$

and $\mathcal{D}$ denotes the set of active sensor positions. This sparse activation pattern preserves the physical aperture of the original ULA. The resulting SLA snapshots are

$$\mathbf{x}(t) = \mathbf{D}\,\mathbf{y}(t), \quad t = 1, 2, \ldots, T, \tag{11}$$

which zeroes out the inactive channels while keeping the ambient dimension M unchanged. Consequently, the SCM becomes

$$\hat{\mathbf{R}}_x = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}(t)\mathbf{x}^H(t) = \mathbf{D}\,\hat{\mathbf{R}}\,\mathbf{D}^H + \sigma_n^2\mathbf{I}_M \in \mathbb{C}^{M\times M}, \tag{12}$$

where the rows and columns corresponding to inactive sensors take zero values.

For the ULA, the sensor positions $d_1, \cdots, d_M$ are equally spaced, so the resulting polynomial $G(z)$ has a Vandermonde structure. In contrast, for a SLA, the element positions are non-uniform, and $G(z)$ becomes a sparse polynomial [21] rather than a Vandermonde polynomial. As a consequence, the $\hat{K}$ roots of interest lie on or close to the unit circle, but they are not guaranteed to be the $\hat{K}$ roots that are closest to the unit circle. As a result, the standard ULA root-selection strategy does not apply to SLAs. A more robust strategy is to locate the desired roots by directly solving the sparse polynomial minimization problem [22]:

$$\min_{z} G(z) = \sum_{m=1}^{M}\sum_{n=1}^{M} p_{m,n}\, z^{\,d_m - d_n}, \tag{13}$$

where the coefficients $p_{m,n}$ are determined by the entries of the noise-subspace projection matrix. To implement (13), a coarse MUSIC spectrum search is first performed to identify the feasible DOA regions, followed by constrained minimization within these regions to reliably extract the roots of interest [22].

Since $\hat{\mathbf{R}}_x$ in (12) is a SCM with missing entries, several works reconstruct it into a "hole-free" covariance matrix [9], [23], [24], [25]. The completed matrix enables the use of gridless ULA-based algorithms such as root-MUSIC and ESPRIT for high-resolution DOA estimation. In particular, [24, 25] introduce neural network-based methods that learn the covariance completion directly for SLA based DOA estimation.

III. Algorithm Framework

A. Architecture

The overall architecture is shown in Fig. 1. The encoder takes the sample covariance matrix $\hat{\mathbf{R}}$ as input and outputs a subspace representation $\tilde{\mathbf{U}}$. This representation is then fed into a DOA estimation block to produce the final angle estimates. During training, two identical encoder branches with shared weights are used to form a Siamese structure, which enables the contrastive loss described in Section III-B(2). During inference, only a single encoder branch is used, as illustrated in Fig. 2, and the Siamese branch is no longer required.

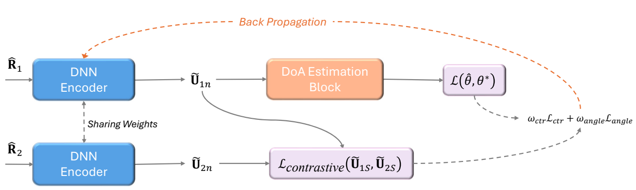


Fig. 1: Training scheme.

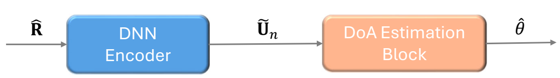


Fig. 2: Inference scheme.

1) Encoder: As illustrated in Fig. 3, the encoder extracts a subspace representation from the sample covariance matrix $\hat{\mathbf{R}}$. The input matrix is treated as a two channel representation (real and imaginary), which preserves its spatial structure, including Hermitian symmetry, coupling-induced correlations, and the sparsity patterns introduced by SLA configurations.

The front-end consists of three convolutional blocks, each containing a two-dimensional (2D) convolution layer, a 2D batch normalization layer, and a ReLU activation function, which collectively capture local spatial dependencies within $\hat{\mathbf{R}}$. Although the SLA covariance matrix contains zero-filled entries due to inactive sensors, the convolutional layers learn to extract robust features and implicitly compensate for the structural distortions caused by mutual coupling or missing elements. A max-pooling layer further aggregates spatial information, and an adaptive average pooling layer normalizes the feature size across different array configurations. The resulting feature map is then flattened into a compact representation.

Following the convolutional front-end, a lightweight multi-layer perceptron (MLP) serves as the subspace regression head. The three

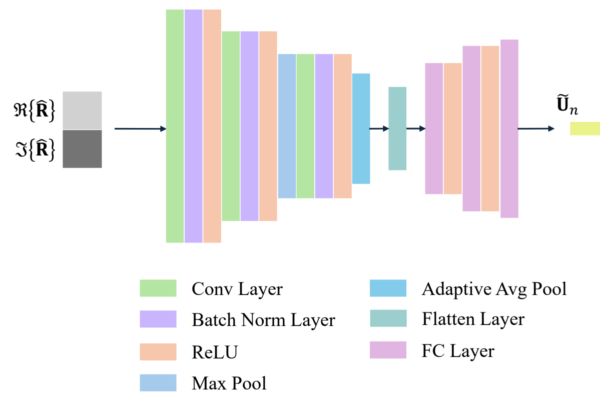


Fig. 3: Network encoder architecture.

fully connected layers map the flattened features to the real and imaginary components of the estimated noise subspace basis $\tilde{\mathbf{U}}_n$, producing an embedding suitable for subspace-based DOA estimation.

Overall, the encoder architecture follows the sequence shown in Fig. 3 and contains a total of $697,104$ trainable parameters.

2) DOA Estimation Block: The DOA estimation block maps the learned subspace representation $\tilde{\mathbf{U}}_n$ from the encoder to angle estimates that supervise the subspace learning. Given $\tilde{\mathbf{U}}_n$, the block recovers the DOA estimates $\hat{\boldsymbol{\theta}}$, and the resulting angle regression loss is backpropagated through the DOA estimator and the encoder. In this way, the network is trained end-to-end so that the latent subspace is optimized for DOA estimation.

In this work, root-MUSIC is employed as the final DOA estimation module. From the latent subspace estimate $\tilde{\mathbf{U}}_n$, we construct the projection matrix $\hat{\mathbf{P}}_n = \tilde{\mathbf{U}}_n\tilde{\mathbf{U}}_n^H$ and form the root-MUSIC polynomial $G(z)$ as in (7), after which the selected roots are mapped to angle estimates using (8). Unlike approaches that explicitly reconstruct a hole-free covariance matrix, our framework learns a latent subspace directly from the masked SLA covariance matrix and optimizes it through the root-MUSIC objective. A key aspect is the interpretation of the encoder output for SLA inputs. Because an SLA contains missing entries and lacks the Vandermonde structure required to form a valid root-MUSIC polynomial, the standard noise subspace extracted from an SLA SCM cannot be directly applied for root-MUSIC. Instead, the encoder learns to produce a subspace-like latent embedding $\tilde{\mathbf{U}}_n$ whose dimension is fixed to the true noise-subspace dimension corresponding to the full ULA aperture. This embedding does not need to coincide with the physical noise subspace of the SLA; rather, it serves as a learned surrogate that restores the structural properties required by root-MUSIC. As a result, a single root-MUSIC head can be applied consistently to both ULA and SLA inputs, enabling gridless DOA estimation even when the original measurement covariance matrix is masked or structurally distorted.

B. Loss Function Design and Training Procedure

1) Loss function design: This total loss consists of two parts: a contrastive loss and an angle regression loss. The two loss terms are combined using weights w_{angle} and w_{ctr}, given as:

$$\mathcal{L}_{\text{total}} = w_{\text{ctr}}\,\mathcal{L}_{\text{ctr}} + w_{\text{angle}}\,\mathcal{L}_{\text{angle}}. \tag{14}$$

Contrastive loss. As shown in Fig. 1, the network takes as input a pair of sample covariance matrices $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$. During training, we randomly form input pairs either from samples sharing the same ordered DOA label (similar) or from samples with different labels (dissimilar), and assign the pair label $z = 1$ or $z = 0$, respectively. Each covariance matrix is passed through the encoder to produce its respective learned noise-subspace representation, $\tilde{\mathbf{U}}_{n_1}$ or $\tilde{\mathbf{U}}_{n_2}$.

To quantify the similarity between two noise subspaces, we adopt a Grassmannian projector metric, which compares the orthogonal projection matrices associated with the subspaces [15, 26]. For a subspace basis $\mathbf{U}$, the corresponding projection matrix is

$$\mathcal{P}(\mathbf{U}) = \mathbf{U}(\mathbf{U}^H\mathbf{U})^{-1}\mathbf{U}^H, \tag{15}$$

and the Grassmann distance between $\tilde{\mathbf{U}}_{n_1}$ and $\tilde{\mathbf{U}}_{n_2}$ is defined as

$$d = \frac{1}{\sqrt{2}} \left\| \mathcal{P}(\tilde{\mathbf{U}}_{n_1}) - \mathcal{P}(\tilde{\mathbf{U}}_{n_2}) \right\|_{\mathrm{F}}, \tag{16}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. Using this distance, we employ a contrastive loss to learn the embedding space: similar pairs are encouraged to have small Grassmann distance, while dissimilar pairs are enforced to be separated by at least a margin m. For a labeled pair $z \in \{1, 0\}$ (similar/dissimilar) and a mini-batch of B pairs, the contrastive loss is mathematically expressed as

$$\mathcal{L}_{\text{ctr}} = \frac{1}{B} \sum_{l=1}^{B} \Big[z_l\, d_l^2 + (1 - z_l) \left(\max(0,\, m - d_l) \right)^2 \Big]. \tag{17}$$

When the two subspaces in the j-th pair are similar, the distance d_j is small, and the contrastive loss becomes small. On the other hand, when the two subspaces are dissimilar, the distance d_j is large, and the contrastive loss becomes large. Therefore, the contrastive loss give large punishment on dissimilar pairs and small punishment on similar pairs.

Angle loss. To directly supervise the DOA regression, we use an angle-based loss that measures the wrapped angular difference between the predicted angles and the ground truth. Since the ordering of the K targets is arbitrary, we evaluate the error under all possible permutations of the target indices and select the one yielding the smallest total mismatch. For predicted angles $\hat{\theta}_{b,i}$ and ground-truth angles $\theta^{\star}_{b,i}$, the wrapped difference is defined as

$$\mathrm{wrap}(\Delta\theta) = \big((\Delta\theta + \pi) \bmod 2\pi \big) - \pi, \tag{18}$$

which maps angular errors to the interval $[-\pi, \pi]$. The permutation-invariant angle loss over a mini-batch of size B is then given as:

$$\mathcal{L}_{\text{angle}} = \frac{1}{B} \sum_{b=1}^{B} \min_{\pi \in \mathcal{P}_K} \left(\frac{1}{K} \sum_{i=1}^{K} \left(\mathrm{wrap}(\hat{\theta}_{b,i} - \theta^{\star}_{b,\pi(i)}) \right)^2 \right), \tag{19}$$

where $\mathcal{P}_K$ denotes the set of all permutations of $\{1, 2, \ldots, K\}$.

2) Training datasets generation: We simulate signals using a 10-element ULA with half-wavelength inter-element spacing and include Toeplitz mutual coupling with the coefficient $c = 0.316$. The field of view is $\phi_{\text{FOV}} = [-60^\circ, 60^\circ]$, discretized at 1° intervals to form a directional grid $G \in \mathbb{R}^{1 \times P}$ with $P = 121$ possible DOAs. We consider $K = 2$ distinct sources with a minimum angular separation of $\Delta\theta_{\min} = 1^\circ$. Over this grid, we construct all ordered two-target pairs (θ_1, θ_2) satisfying $|\theta_2 - \theta_1| \geq 1^\circ$, yielding $121 \times 120 = 14,520$ unique angular combinations. To avoid locking the data strictly to the discrete grid, the actual target angles are jittered within a small range (up to 1°) around their nominal grid locations.

For each unique ordered combination, we assign a label and generate 40 independent signal-pair realizations. In each realization, an independent complex reflection coefficient s_k is drawn for each target, with $|s_k|$ uniformly distributed between 0.3 and 1, and a single SNR level is randomly selected from $\{0, 5, 10, 15, 20\}$ dB and applied to the mixture. The resulting noisy snapshots are used to form the corresponding covariance matrices. In the Siamese training setup, one realization in each pair is fed to the main branch and the other to the Siamese branch, while all realizations associated with the same ordered DOA pair share the same label.

Then we generate the SLA data from the same 10-element ULA manifold by activating the sensor indices $[1, 2, 4, 6, 9, 10]$, resulting in sparse snapshots whose sample covariance remains $\hat{\mathbf{R}}_x \in \mathbb{C}^{10 \times 10}$.

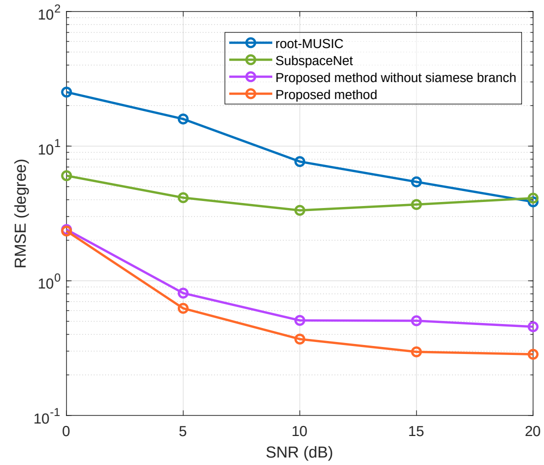


Fig. 4: DOA RMSE versus SNR for the ULA, comparing baseline methods and the proposed model with and without the Siamese branch.

IV. Numerical Evaluations

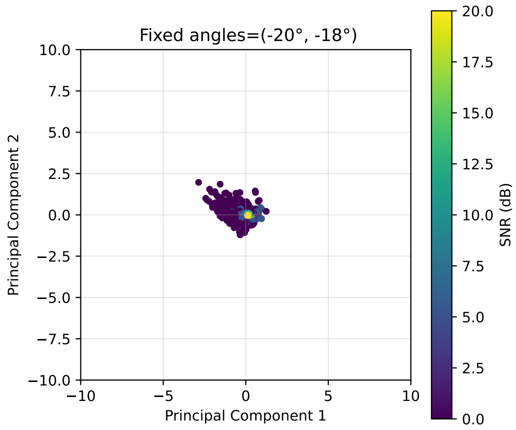
For evaluation, we generate ULA and SLA test datasets following the same simulation settings as in Section III-B(1). Performance is quantified by the RMSE results of the estimated DOAs.

We implement the proposed model and the SubspaceNet [14] baseline in PyTorch. All models are trained using the Adam optimizer with a learning rate of 1×10^{-4} for 200 epochs and a batch size of 512. A ReduceLROnPlateau scheduler [27] is employed to adaptively decrease the learning rate when the validation loss stops improving. For our proposed method, the loss weights are set to $w_{\text{angle}} = w_{\text{ctr}} = 1$. For the contrastive loss, the margin is fixed to $m = 0.8$ in all experiments. All training is performed on a single NVIDIA Quadro RTX 4000 GPU.

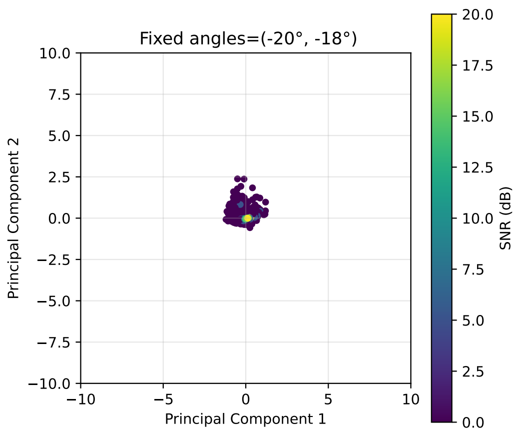
A. ULA Results

In the ULA scenario, we compare four methods: the proposed model, SubspaceNet, the classical root-MUSIC algorithm, and an ablation version of the proposed model without the Siamese branch. The RMSE results are shown in Fig. 4. Across all SNR conditions, the proposed method achieves the lowest RMSE; moreover, the Siamese branch provides noticeable improvements, particularly in the high-SNR regime.

To further examine the Siamese network's ability to learn stable subspace representations, Fig. 5 visualizes the Principal Component Analysis (PCA) projection of learned subspace embeddings for a fixed DOA combination $(-20^\circ, -18^\circ)$ under different SNR levels. The visualization is generated using 8,000 signals that share the same DOA pair but differ in noise levels and reflection coefficients. Each color corresponds to a different SNR. Without using the Siamese branch, the embeddings spread widely as SNR varies. In



(a)



(b)

Fig. 5: PCA result of proposed method in the ULA scenario: (a) without Siamese branch, and (b) with Siamese branch.

contrast, when the Siamese branch is used, the embeddings collapse into a much tighter cluster, indicating that the contrastive Siamese training enforces more discriminative and SNR-invariant subspace representations.

B. SLA Results

In the SLA scenario, we evaluate the proposed method with and without using the Siamese branch. The RMSE performance in Fig. 6 shows that both versions of the proposed model achieve high DOA estimation accuracy despite the array sparsity. The Siamese branch provides additional improvements, particularly in the high-SNR regime, indicating that it helps the network learn more SNR-invariant subspace representations under irregular array configurations.

To further analyze the properties of the learned subspace, Fig. 7 compares the MUSIC spectra corresponding to three different subspaces. We randomly select two representative test samples from the SLA dataset, and for each case we obtain: (i) the latent subspace produced by our proposed model, (ii) the subspace obtained by applying singular value decomposition (SVD) to the same SLA covariance matrix, and (iii) the corresponding ULA subspace used as a reference.

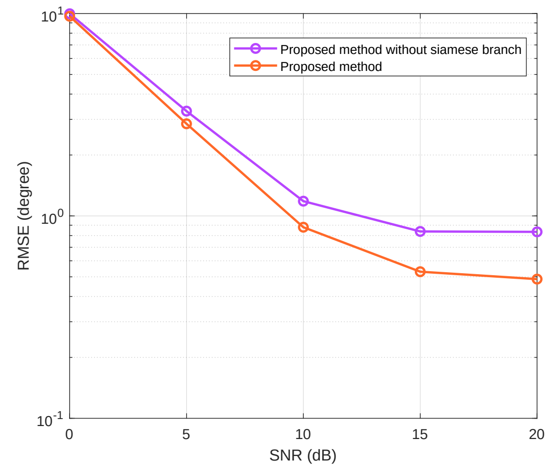


Fig. 6: DOA RMSE results versus SNR for the SLA, comparing models with and without using the Siamese branch.

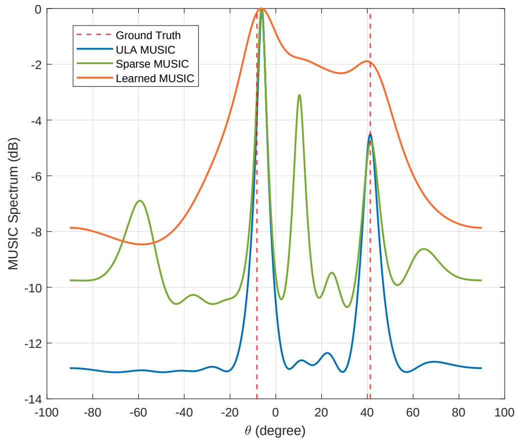
At SNR = 0 dB (Fig. 7a), the SLA-SVD spectrum exhibits high sidelobes, distorted main lobes, and noticeable peak shifts, making it difficult to correctly locate the DOAs. At SNR = 10 dB (Fig. 7b), these issues persist, with high residual lobes and amplitude inconsistencies still present. In contrast, the spectrum produced by the learned subspace consistently suppresses sidelobes and yields sharp, well-localized peaks that align closely with the ground truth at both SNR levels. The close match to the ULA reference spectrum indicates that the learned subspace behaves like a virtual ULA-like manifold, preserving the correct peak geometry even under sparse spatial sampling and mutual coupling. Although the model output in the SLA case is a latent subspace rather than a physically interpretable noise subspace, the angle-regression supervision guides it to approximate the underlying true subspace sufficiently well to recover the correct peak locations in the MUSIC spectrum.

Fig. 8 provides additional insight through a PCA projection of the learned embeddings. Here we visualize $8,000$ SLA samples that share the same DOA pair $(-20^\circ, -8^\circ)$ but differ in SNR and reflection coefficients. Without using the Siamese branch, the embeddings spread widely in the PCA space as the noise level varies. In contrast, with the Siamese branch, the embeddings collapse into a significantly tighter and more coherent cluster. The horizontally elongated shape of the resulting cluster arises from the anisotropic geometry of the SLA subspace manifold for this particular DOA pair, where most of the variance is concentrated along a single principal direction. Nevertheless, the Siamese branch clearly enhances compactness and robustness, and its benefits are even more pronounced in the SLA setting than in the ULA case.

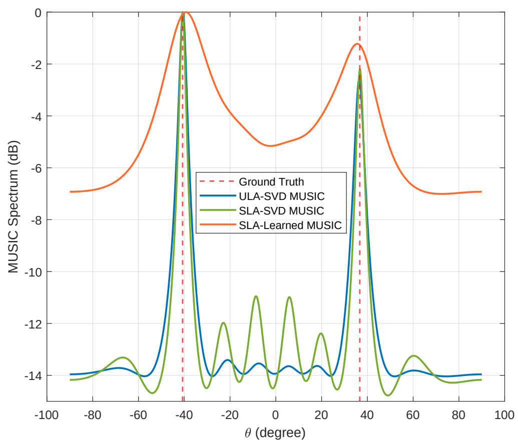
C. Separation

To further evaluate the separation capability of the proposed method, we consider a two-sources scenario with fixed angular spacing $\Delta\theta \in \{1^\circ, 2^\circ, 4^\circ, 6^\circ, 8^\circ, 10^\circ\}$. A test sample is counted as a pass if the DOA estimation errors of both sources are within $\delta = 1^\circ$. The pass rate is defined as the fraction of passed samples, and the RMSE is computed only over those passed samples.

Fig. 9a shows the results at low SNR (0 dB). The pass rate increases from about 6% at $\Delta\theta = 1^\circ$ to roughly 42% at $\Delta\theta = 10^\circ$, while the RMSE evaluated over the passed samples decreases mildly from approximately 0.59° to 0.51°. This indicates that, under severe noise,

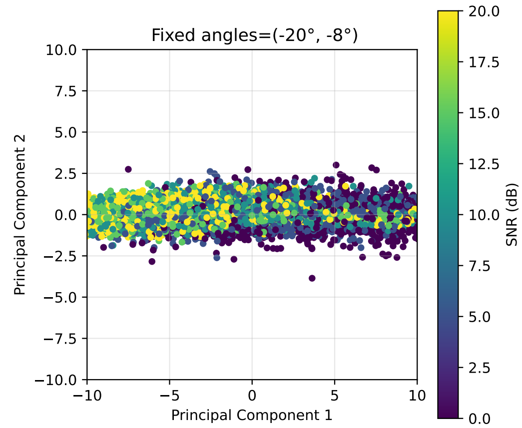


(a)

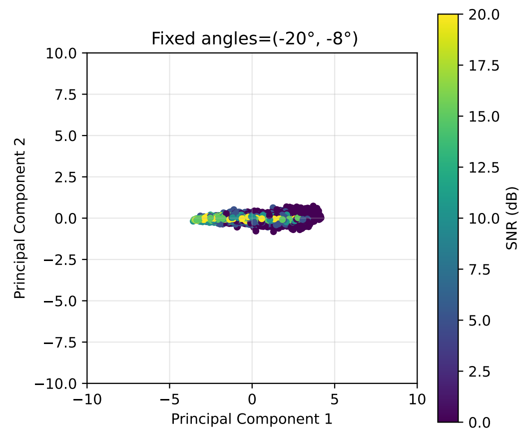


(b)

Fig. 7: MUSIC spectra comparison for two representative SLA test cases: (a) SNR = 0 dB and (b) SNR = 10 dB.



(a)



(b)

Fig. 8: PCA result of proposed method in the SLA scenario: (a) without Siamese branch, and (b) with Siamese branch.

the model can reliably resolve the two sources only when they are moderately separated, and many closely spaced cases fail to meet the $\delta = 1^\circ$ requirement.

Fig. 9b reports the same experiment at high SNR (20 dB). The pass rate reaches about 77% at $\Delta\theta = 1^\circ$, exceeds 90% for $\Delta\theta \geq 4^\circ$, and approaches 99% at $\Delta\theta = 10^\circ$. Meanwhile, the RMSE computed over the passed samples decreases from roughly 0.50° to about 0.24° as the separation increases. These results show that, under favorable SNR conditions, the proposed model can resolve two sources separated by as little as 1°, and provides highly accurate angle estimates for the successfully resolved cases.

V. Conclusion

In this paper, we proposed an SNN for subspace-based DOA estimation with SLAs. By combining a Siamese encoder with a root-MUSIC layer, the proposed method learns SNR-invariant latent subspace representations that remain robust under array sparsity and mutual coupling. Numerical results demonstrated its clear improvements over classical root-MUSIC and the SubspaceNet baseline in the ULA case, and enhanced performance under SLA configurations, with the Siamese branch yielding especially clear gains at high SNR. A ULA separation test als confirms that the model can resolve two sources down to 1° spacing at high SNR with high pass rates and low erros. PCA analyses of the learned embeddings further confirmed that the contrastive training yields compact and discriminative subspace clusters, highlighting the effectiveness of subspace representation learning for challenging sparse-array scenarios.

References

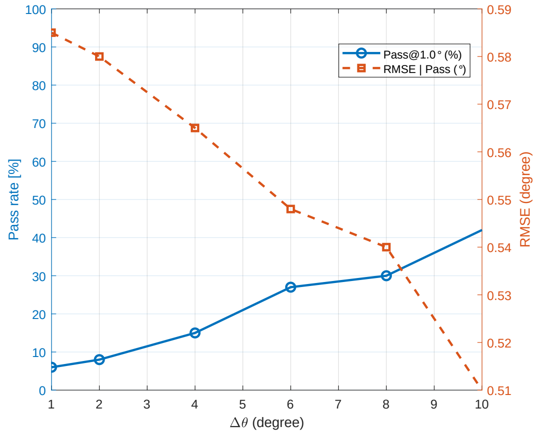
[1] R. O. Schmidt, *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation.* Ph.D. Dissertation, Stanford University, 1982.

[2] R. Roy and T. Kailath, "ESPRIT – Estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

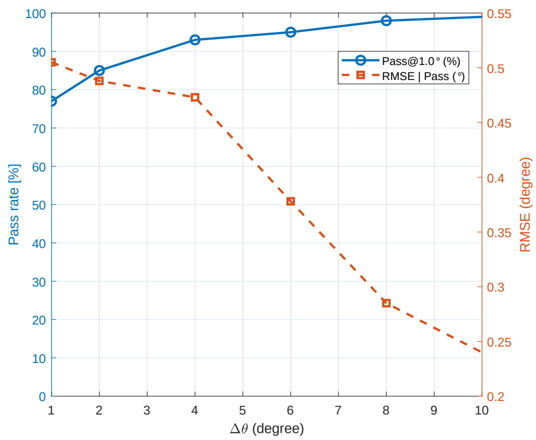
[3] S. Sun, A. P. Petropulu, and H. V. Poor, "MIMO radar for advanced driver-assistance systems and autonomous driving: Advantages and challenges," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 98–117, 2020.

[4] R. Zheng, S. Sun, H. Liu, H. Chen, M. Soltanalian, and J. Li, "Antenna failure resilience: Deep learning-enabled robust DOA estimation with single snapshot sparse arrays," in *2024 58th Asilomar Conference on Signals, Systems, and Computers*, 2024.

[5] S. Sun and Y. D. Zhang, "4D automotive radar sensing for autonomous vehicles: A sparsity-oriented approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 879–891, 2021.



(a)



(b)

Fig. 9: Pass rate and RMSE versus source separation $\Delta\theta$ for the two-source ULA under (a) SNR = 0 dB and (b) SNR = 20 dB.

[6] D. Romero, D. D. Ariananda, Z. Tian, and G. Leus, "Compressive covariance sensing: Structure-based compressive sensing beyond sparsity," *IEEE Signal Processing Magazine*, vol. 33, no. 1, pp. 78–93, 2016.

[7] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167–4181, 2010.

[8] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1257–1269, 2016.

[9] M. Wang, Z. Zhang, and A. Nehorai, "Grid-less DOA estimation using sparse linear arrays based on Wasserstein distance," *IEEE Signal Processing Letters*, vol. 26, no. 6, pp. 838–842, 2019.

[10] S. Ge, K. Li, and S. N. B. M. Rum, "Deep learning approach in DOA estimation: A systematic literature review," *Mobile Information Systems*, vol. 2021, no. 1, p. 6392875, 2021.

[11] G. K. Papageorgiou, M. Sellathurai, and Y. C. Eldar, "Deep networks for direction-of-arrival estimation in low SNR," *IEEE Transactions on Signal Processing*, vol. 69, pp. 3714–3729, 2021.

[12] J. P. Merkofer, G. Revach, N. Shlezinger, and R. J. G. van Sloun, "Deep augmented MUSIC algorithm for data-driven DOA estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 3598–3602.

[13] 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 Transactions on Vehicular Technology*, vol. 73, no. 2, pp. 2771–2785, 2024.

[14] 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," *IEEE Transactions on Vehicular Technology*, vol. 74, no. 3, pp. 4962–4976, 2025.

[15] K.-L. Chen and B. D. Rao, "Subspace representation learning for sparse linear arrays to localize more sources than sensors: A deep learning methodology," *IEEE Transactions on Signal Processing*, vol. 73, pp. 1293–1308, 2025.

[16] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a 'Siamese' time delay neural network," in *Advances in Neural Information Processing Systems (NIPS)*, 1993, pp. 737–744.

[17] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *Proceedings of the 32nd International Conference on Machine Learning (ICML) Deep Learning Workshop*, 2015.

[18] F. Wang and H. Liu, "Understanding the behaviour of contrastive loss," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 2495–2504.

[19] X. Xu, Z. Ye, and Y. Zhang, "DOA estimation for mixed signals in the presence of mutual coupling," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3523–3532, 2009.

[20] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, 2004.

[21] D. S. Roche, "What can (and can't) we do with sparse polynomials?" in *Proceedings of the 2018 ACM International Symposium on Symbolic and Algebraic Computation*, 2018, pp. 25–30.

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

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

[24] A. Barthelme and W. Utschick, "DOA estimation using neural network-based covariance matrix reconstruction," *IEEE Signal Processing Letters*, vol. 28, pp. 783–787, 2021.

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

[26] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303–353, 1998.

[27] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, *Dive into Deep Learning*. Cambridge University Press, 2023.