Deep Learning-Based Direction-of-Arrival Estimation with Covariance Reconstruction

Ahmed Manavi Alam, Cemre Omer Ayna, Sabyasachi Biswas
John T. Rogers, John E. Ball, and Ali C. Gurbuz
Dept. of Electrical and Computer Engineering, Mississippi State University, MS, USA
{aa2863, ca1389, sb3682, jtr242}@msstate.edu, {jeball, gurbuz}@ece.msstate.edu

***Abstract*—Accurately determining Direction of Arrival (DoA) is pivotal for various applications such as wireless communication, radar, and sensor arrays, where precise spatial localization is crucial in enhancing system performance and overall efficiency. Low signal-to-noise ratio (SNR) and limited number of snapshots pose formidable challenges to accurate DoA estimation. Both conventional model-based techniques and recent deep learning (DL) based DoA estimation models that map sample covariance matrices to DoA spectrum estimations struggle in such environments. In this study, we introduce a comprehensive DL framework that leverages sample covariance as input to predict the corresponding DoA jointly with the estimation of the true covariance matrix. The proposed architecture comprises two main components that employ Convolutional Neural Networks (CNN). The first part focuses on covariance reconstruction, aligning with the true covariance of a specific sample, and the second part applies multi-label classification for the DOA estimation step. Distinct from employing only Binary Cross-Entropy (BCE) loss for the previous on-grid CNN approaches, our study implements a holistic training strategy incorporating three individual loss terms into one novel combined loss function. The proposed overall framework integrates the Mean Squared Error (MSE) loss for the true covariance matrix reconstruction, to enhance model performance, particularly in low SNR and snapshot number scenarios, coupled with the BCE and MSE losses for angle estimation. This strategic combination demonstrates improved robustness and performance compared to existing CNN-based approaches.**

***Index Terms*—Direction of Arrival (DOA), array signal processing, covariance reconstruction, deep learning**

I. Introduction

The problem of precise determination of Direction of Arrival (DoA) holds critical importance across various research domains such as wireless communication, radar systems, sensor arrays, and many others [1], [2]. Achieving spatial localization accuracy, a key facet of a successful DoA estimation significantly bolsters system performance and overall operational efficiency. With that in mind, the task of DoA estimation presents formidable challenges, particularly in real-world scenarios marked by multipath propagation, ambient noise, low Signal-to-Noise Ratio (SNR), and a constrained number of available snapshots [3], [4] specifically for fast-changing environments.

Model-based DoA estimation techniques have played a pivotal role in array signal processing for localizing sources in sensor array applications. The Multiple Signal Classification (MUSIC) algorithm leverages the eigen-structure of the signal's covariance matrix to estimate DoA with high resolution [5], [6]. Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) exploits the invariance of the signal subspace under a unitary transformation, providing accurate estimates even in low SNR scenarios [7]–[9]. The Minimum Variance Distortionless Response (MVDR) algorithm employs a spatial filter to minimize the received power while preserving the signal of interest, enhancing robustness in noisy environments [10]. Conventional beamformers, such as delay-and-sum, offer simplicity but are limited in resolving closely spaced sources [11], [12]. Despite their efficacy, these model-based techniques face challenges, including sensitivity to model mismatch, limited performance in the presence of strong noise or interference, low number of snapshots, or correlated sources, and suffer high computational complexity, especially for subspace techniques. Striking a balance between computational cost and accuracy under such challenges remains an ongoing concern in deploying these methods in real-world applications.

Stated existing challenges for conventional DoA estimation approaches prompted many researchers to explore deep learning (DL) based techniques for DoA estimation due to their ability to handle complex functional mappings [13]–[15]. DL-based techniques operate as non-linear mapping functions that learn the directional characteristics of array systems [13], [16]–[18]. By constructing training datasets with labeled DoAs, these methods establish a relation between observed array data and signal directions, allowing for the estimation of signal directions in test data for the same array geometry. Deep learning-based approaches offer advantages such as inherent adaptability, the ability to handle complex, non-linear relationships, and rapid execution without specific parameter optimization. They showcase improved performance in low SNR scenarios and with fewer snapshots. Various implementations, including end-to-end deep neural networks (DNNs) for massive MIMO and convolutional neural networks (CNNs) for low SNR scenarios, showcase the versatility and accuracy of deep learning in DoA estimation, making them well-suited for real-time applications [13], [19].

The model in [13] utilizes a conventional CNN architecture that takes the sample covariance matrix as input and implements a multi-label classification scheme with a binary entropy loss to estimate an output vector of DoAs with probability of a target is present at each angle grid. These models are trained

2024 IEEE Radar Conference (RadarConf24) | 979-8-3503-2920-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/RADARCONF2458775.2024.10548988

with true covariance matrices but tested with the sample covariance. Although sample covariance is a good estimate of the true covariance when the number of snapshots gets closer to infinity, for a limited number of snapshot cases and under low SNR regime the input sample covariance matrix is not a good estimate of the true covariance leading to significant performance loss under such conditions.

In this study, we developed an improved joint covariance matrix reconstruction and DoA estimation approach utilizing two different CNN architectures backpropagated with a novel combined loss term. The proposed comprehensive DL framework is tailored to provide enhanced performance compared to current CNN-based approaches in different SNR levels and diverse snapshot scenarios. Unlike existing DL models, proposed architecture uses sample covariance indirectly to predict the corresponding DoA through a covariance matrix reconstruction step. Covariance matrix estimation is first introduced in [20] The proposed architecture comprises two sub-networks as its main components: the first sub-network focuses on learning the reconstruction of the true covariance matrix from sample covariance during training, and the second part is an angle estimation network that employs a CNN-based on-grid approach for multi-label classification of the DoA problem. To enhance the model approximation performance, the final loss function is defined as the weighted combination of three separate loss functions during training. The first loss function is essentially a mean square error (MSE) that is computed by comparing the estimated and true covariance matrix in the covariance reconstruction sub-network. In the angle estimation network, the second and third loss functions are integrated metrics of the final estimation; these are BCE loss for on-grid multi-label classification and MSE loss to gauge the disparity between the estimated angles and the corresponding ground truth angles. Utilizing only the classification loss for angle estimation is not enough since an erroneous classification that is closer or far to the correct angle is punished similarly. Adding MSE loss in angle estimates forces the network to make better decisions. Notably, our study diverges from the common practice of employing only binary-cross entropy (BCE) loss for on-grid CNN approaches as done in [13]. Instead, we adopt a holistic training strategy that incorporates joint learning of the true covariance matrix and DoA estimation. This strategic combination not only improves the overall robustness of our model but also addresses specific challenges posed by adverse conditions, such as low SNR and a limited number of snapshots.

The paper is organized as follows: summarizing the DoA signal model in Section II, discussing the overall framework in Section III, analyzing performance in Section IV, and concluding and discussing future directions in Section V.

II. Signal Model

The success of a DoA estimation algorithm depends on the knowledge of the received and transmitted signal and the behavior of the noise through the transmission. In this section, we lay out the mathematical foundation of the signal model used in this study. The data generation scheme and the numerical characteristics of the generated data will be expressed together with the model architecture in the next section.

In our model, we are assuming to work with an M element-long uniform linear antenna array (ULA) operating in the narrow band and reading in total K number of distinct and uncorrelated signal sources at a given time period. Each individual transmitted signal $\mathbf{s}_k$ will contribute to the receiver's stimulation depending on its arrival angle θ_k along with the captured noise η.

A received signal in this configuration with respect to the sampling time $t \in \{1, ..., T\}$ will be described as;

$$\mathbf{x}(t) = \sum_{k=1}^{K} \mathbf{a}(\theta_k)\mathbf{s}_k(t) + \eta(\mathrm{t}) \tag{1}$$

where $\mathbf{a}(\theta_k) = [1, e^{jd2\pi/\lambda * sin(\theta_k)}, ..., e^{jd2\pi(M-1)/\lambda * sin(\theta_k)}]$ is the steering vector of each antenna element, defined by the uniform distance between antennas d and the wavelength of the incoming signal λ, and it provides the information on the phase shift in received signal for each antenna. Now, we define $A = [a(\theta_1), ..., a(\theta_K)]$ as the compact form of the steering vectors in a $M \times K$ matrix form, and $s(t)$ as the vector form of all received targets.

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \eta(t) \tag{2}$$

In this model, actual signals and additive noise are considered to be independent. In many real-life scenarios, the noise can be modeled with Gaussian distribution with zero mean and an unknown noise power; $\eta(\mathrm{t}) \sim \mathcal{N}(0, \sigma^2)$.

Subspace techniques for the DoA estimation, such as MUSIC and ESPRIT, employ separation of signal and noise subspaces by analyzing the covariance matrix of the received signal. The true covariance matrix can be stated as:

$$\mathbf{R_x} = \mathbb{E}[\mathbf{x}(t)\mathbf{x}^H(t)] = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \sigma^2\mathrm{I}_M \tag{3}$$

where $\mathbf{R}_s \in \mathbb{R}^{K \times K}$ is the signal covariance matrix, a diagonal matrix for the uncorrelated signal case with non-zero elements corresponding to the power of the actual target signals. In real-life scenarios, we lack the true covariance matrix information and instead employ sample covariance matrix as an estimation of the actual covariance matrix.

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

If the number of snapshots is adequately high, the sample covariance matrix $\tilde{\mathbf{R}}_x$ converges to the true covariance matrix $\mathbf{R}_x$. We formulated the synthetic data generation according to the provided signal model and employed the opportunity of access to the actual true covariance matrix that the synthetic data gives us by embedding the estimation of true covariance matrix $\mathbf{R}_x$ from the sample covariance matrix $\tilde{\mathbf{R}}_x$ into our model's design. Next section details this process.

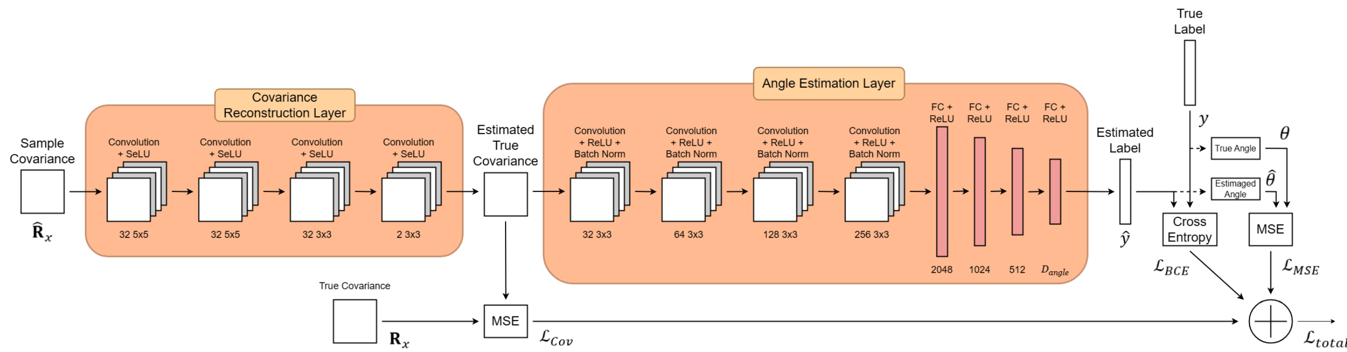


Fig. 1: The proposed joint covariance reconstruction and angle estimation model with novel combined loss terms.

III. Proposed Method

A. Data Preprocessing and Labeling

The synthetic data used in this study is created in MATLAB by following the signal model described in Section II. The simulated data includes all SNR and snapshot values determined to be used in the tests. The tested SNR values range between -20 dB and 20 dB with 5 dB increments. The tested snapshot values include $N \in \{50, 100, 150, 200, 300, 400, 500, 600\}$. The models in the evaluation were trained with the whole dataset and tested separately over an independently generated dataset according to a specific test SNR and snapshot number value.

The training dataset includes all possible angles for targets between $20°$ and $160°$ in the two-target case with the minimum angular distance between targets defined as $\Delta\theta = 1°$. The grid scale is also selected as $1°$, therefore the on-grid configuration has in total $D_{angle} = 141$ angle values. The targets are assumed to be noncoherent and are affected mainly by the simulated thermal noise from the system which follows the Gaussian distribution.

B. Overall Framework

The full proposed DoA estimation model is depicted in Fig. 1. This architecture is a sequential combination of two sub-networks that are defined to solve separate objectives. In that respect, the overall model accepts the sample covariance matrix of the received signal as input of the first sub-network and recovers an estimate of the true covariance matrix as output. Then, the reconstructed covariance matrix is fed into the second sub-network for DoA estimation. This network returns a fully connected layer output vector, wherein each element represents the probability that there is an estimated target in the corresponding angle.

The key component of the overall network is to employ true covariance directly where the sample covariance matrix estimation is poor due to the low number of snapshots. The initial covariance estimation network mitigates this problem by cleaning the sample covariance to retrieve a better estimate of the true covariance than the bare sample covariance. The covariance reconstruction layer learns its objective alongside the DoA estimation network during training by including the covariance loss L_{Cov} into the final loss term L_{total}.

The second important novelty of the proposed model is the joint loss term that not only includes the classification loss, Binary Cross-Entropy (BCE) loss for multi-target multilayer classifications, based on the on-grid angle model but also employs Mean Squared Error (MSE) based on the numerical angle values converted from the final output. The BCE loss is defined for each possible angle location as in (5). The MSE loss on the other hand was defined according to the two-target case and directly utilizes the correct angle values θ_1 and θ_2.

$$\mathrm{L}_{BCE} = \frac{1}{D_{angle}} \sum_{i=1}^{D_{angle}} \hat{\mathbf{y}}_i p(\mathbf{y}_i) + (1-\hat{\mathbf{y}}_i)(1-p(\mathbf{y}_i)) \quad (5)$$

$$\mathrm{L}_{MSE} = \frac{1}{2}\sum_{i=1}^{2}(\hat{\theta}_i - \theta_i)^2 \quad (6)$$

C. Covariance Estimation

The covariance estimation sub-network is a concatenation of convolutional layers with Scaled Exponential Linear Unit (SELU) activation functions. There are four convolutional blocks in total with 1 stride and equal padding, the first two of them having 32 5×5 kernels, the third one having 32 5×5 kernels, and the last one having 2 5×5 kernels. Both sample covariance and true covariance matrices are originally complex numbers, and in our model they are represented as images having two channels.

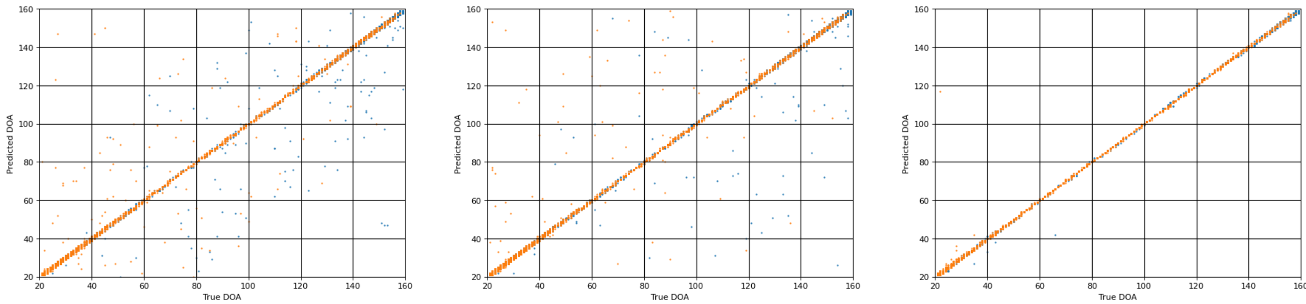
D. CNN-Based Angle Estimation

The angle estimation sub-network receives the true covariance matrix from the initial covariance reconstruction network and passes it through a sequence of convolutional blocks composed of a convolutional layer, a Residual Linear Unit (ReLU), and a Batch Normalization operation.

E. Training the Overall Framework

The overall model combines all three loss terms in the training process, as shown in (7).

$$\mathrm{L}_{total} = \mathrm{L}_{Cov} + \mathrm{L}_{BCE} + \alpha\mathrm{L}_{MSE} \quad (7)$$



(a) (b) (c)

Fig. 2: Scatter plot for the estimated target angles for 200 snapshots and 0 dB SNR for (a) Base model (b) Base CNN + Covariance Estimation (c) Base CNN + Covariance Estimation + Angular MSE Loss

The strength of the MSE loss on the overall training is controlled by a weight term α that is defined heuristically. In our experiments, the α value was chosen as 0.1. The kernel size is kept 3×3 with 1 stride for all blocks, and the kernel number changes to 32, 64, 128, and 256 from the first block to the last block. The last convolutional block is followed by a sequence of fully-connected layers. The final layer has the number of neurons equal to the amount of the total angle grids D_{angle}.

IV. Performance Analysis

A. Evaluation of Covariance Estimation Layer

In the proposed architecture, the first issue is to make sure that the covariance estimation network works properly; therefore the first analysis looks at whether the estimated true covariance is close to the actual true covariance.

Fig. 3 shows three random samples from the test dataset selected from simulated test data with different snapshot and noise power values. The top figures are the sample covariances, and they clearly demonstrate the effect of the low snapshot number on the poor estimation; the noise in sample covariance dominates the matrix as information is low due to the inadequate number of snapshots. The middle figures are the actual true covariances of the test signals. The bottom figures are the reconstructed covariance matrices, which are closer to the actual true covariance matrices. The diagonal entries seem to be more apparent in the reconstructed images, but the entries are still more distinctly recognizable.

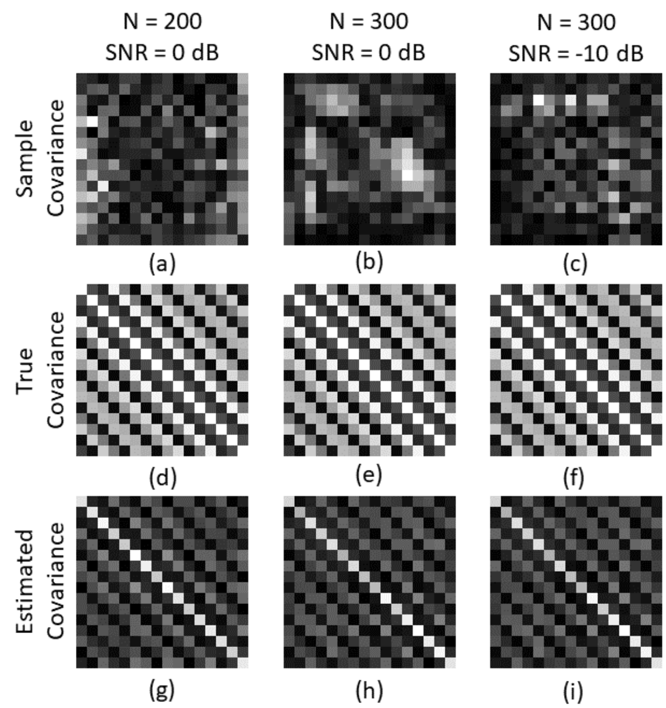


Fig. 3: Covariance estimation results for three random examples. Sample covariance matrices for a) N=200 0 dB SNR case, b) N=300 0 dB SNR case, c) N=300 -10 dB SNR case are shown above. The middle figures are the true covariance matrices, and the bottom figures are the estimated true covariances of the respective column.

B. Evaluation Across SNR

This analysis is performed to evaluate the proposed model's performance with respect to the noise power. The snapshot value throughout the analysis was kept constant at $N = 200$. Accuracy and RMSE values of the proposed model and the compared methods with respect to the SNR values in tests can be found in Fig. 4 and 5.

First comment is on the increase of performance in the CNN model with the addition of the covariance estimation block. Both models converge to a very high accuracy for high-SNR cases, for the sample covariance approaches to the true covariance as the number of snapshots increases. However, in low-SNR cases, retrieving the true covariance from the sample covariance input with the estimation block results in a significant increase in the accuracy; with the boost of the additional MSE loss the final accuracy score increases to 1.92% for 0 dB, 5.56% for -5 dB, and 6.99% for -10 dB SNR.

The second point is the comparison of the CNN model with the traditional algorithms in accuracy and RMSE metrics. The proposed model approaches to the error rate of MUSIC in all SNR scenarios, and surpasses it above 0 dB SNR rates. On the other hand, the accuracy for the proposed model is higher than the conventional methods in both low-SNR and high-SNR. One explanation for these results might be understood by looking at the scatter plots in Fig. 2. The misses of the base CNN model is not guaranteed to be around the true angle, while misses of the traditional source separation methods usually approach closer to the true angle. The MSE loss term in our case helps to direct the estimations to converge to the true targets even in misses.

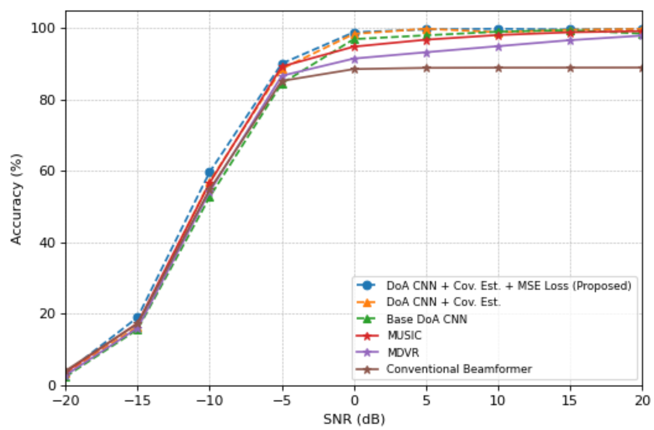


Fig. 4: Accuracy of the compared methods with respect to SNR values, for 200 snapshots.

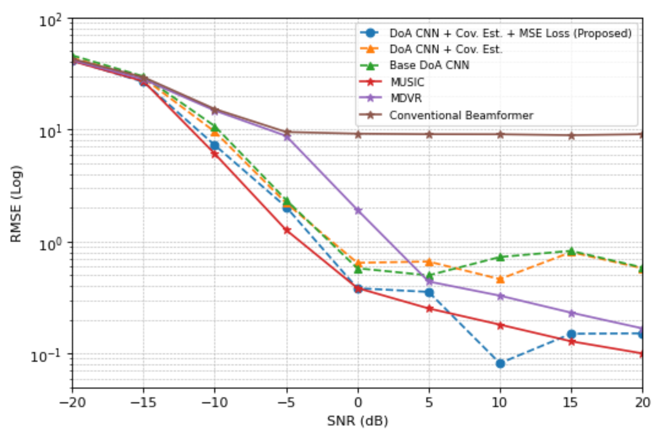


Fig. 5: RMSE loss of the compared methods with respect to SNR values, for 200 snapshots.

C. Evaluation Across Snapshot Number

The performance analysis of the proposed model over the snapshot values show a cleaner picture for the potential of the covariance estimation. The SNR value throughout the analysis was kept constant at 0 dB. Fig. 6 shows the accuracy with respect to the number of snapshots and shows that the proposed model outperforms the conventional beamformer and MUSIC along with the base CNN model.The accuracy difference between the base CNN and the proposed model is 0.92% for 200 snapshots, 1.4% for 150 snapshots, and 3.81% for 100 snapshots.

The comparison of RMSE error with respect to the number of snapshots is presented in Fig. 7. The proposed model for all snapshot values provides a better result compared to both the conventional methods and the base network and closely surpasses the error for MUSIC. The inclusion of the covariance estimation network seems to help DoA estimation for lower snapshot values.

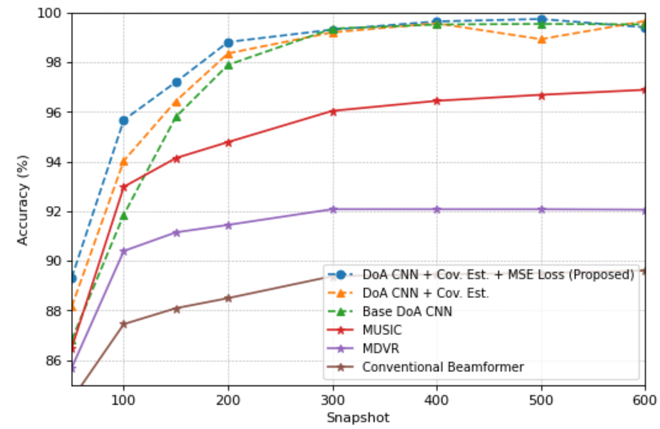


Fig. 6: Accuracy of the compared methods with respect to the number of snapshot for 0 dB SNR

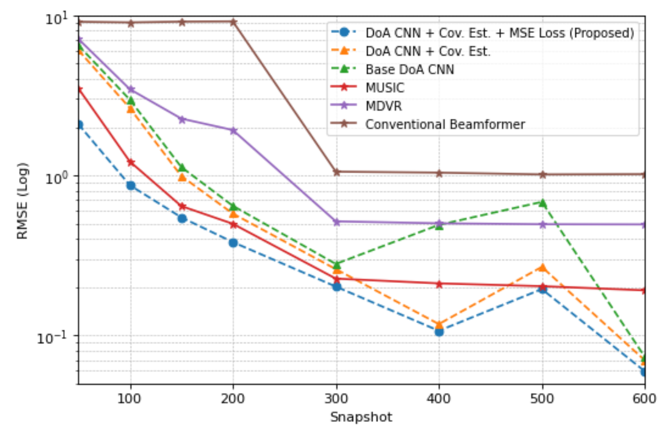


Fig. 7: RMSE loss of the compared methods with respect to the number of snapshot for 0 dB SNR

V. Conclusion

In this study, we developed a neural network model for the direction of arrival estimation problem that incorporates both covariance reconstruction and combined weighted loss functions to boost the DoA estimation performance in low SNR and low snapshot conditions. The proposed model includes a covariance estimation module for recovering the true covariance matrix from the sample covariance and a comprehensive loss term that minimizes the DoA errors in multiple respects. Achieved results indicate that the proposed covariance reconstruction scheme and novel loss term enhance the general DoA estimation performance of the deep learning architecture under various SNR and snapshot conditions. A further study will include extensive analysis and comparison of the proposed model under various more realistic scenarios.

Acknowledgment

The presented study was funded in part by the National Science Foundation (NSF) Awards #2047771.

References

[1] S. Chandran, *Advances in Direction-of-arrival Estimation*. Artech House Boston, 2006.

[2] Z. Chen, G. Gokeda, and Y. Yu, *Introduction to Direction-of-arrival Estimation*. Artech House, 2010.

[3] A. B. Gershman, M. Rübsamen, and M. Pesavento, "One-and two-dimensional direction-of-arrival estimation: An overview of search-free techniques," *Signal Processing*, vol. 90, no. 5, pp. 1338–1349, 2010.

[4] Y. Gu, Y. D. Zhang, and N. A. Goodman, "Optimized compressive sensing-based direction-of-arrival estimation in massive mimo," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 3181–3185.

[5] H. Abeida and J.-P. Delmas, "Music-like estimation of direction of arrival for noncircular sources," *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp. 2678–2690, 2006.

[6] X. Zhang, L. Xu, L. Xu, and D. Xu, "Direction of departure (dod) and direction of arrival (doa) estimation in mimo radar with reduced-dimension music," *IEEE communications letters*, vol. 14, no. 12, pp. 1161–1163, 2010.

[7] B. Ottersten and T. Kailath, "Direction-of-arrival estimation for wide-band signals using the esprit algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 2, pp. 317–327, 1990.

[8] B. D. Rao and K. Hari, "Performance analysis of esprit and tam in determining the direction of arrival of plane waves in noise," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 37, no. 12, pp. 1990–1995, 1989.

[9] J. Pan, M. Sun, Y. Wang, and X. Zhang, "An enhanced spatial smoothing technique with esprit algorithm for direction of arrival estimation in coherent scenarios," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3635–3643, 2020.

[10] L. Tzafri and A. J. Weiss, "High-resolution direct position determination using mvdr," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6449–6461, 2016.

[11] J. Benesty, J. Chen, and Y. Huang, "Conventional beamforming techniques," *Microphone array signal processing*, pp. 39–65, 2008.

[12] V. Krishnaveni, T. Kesavamurthy, and B. Aparna, "Beamforming for direction-of-arrival (doa) estimation-a survey," *International Journal of Computer Applications*, vol. 61, no. 11, 2013.

[13] 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.

[14] A. M. Elbir, "Deepmusic: Multiple signal classification via deep learning," *IEEE Sensors Letters*, vol. 4, no. 4, pp. 1–4, 2020.

[15] Y. Kase, T. Nishimura, T. Ohgane, Y. Ogawa, D. Kitayama, and Y. Kishiyama, "Doa estimation of two targets with deep learning," in *2018 15th Workshop on Positioning, Navigation and Communications (WPNC)*. IEEE, 2018, pp. 1–5.

[16] Z.-M. Liu, C. Zhang, and S. Y. Philip, "Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections," *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 12, pp. 7315–7327, 2018.

[17] L. Wu, Z.-M. Liu, and Z.-T. Huang, "Deep convolution network for direction of arrival estimation with sparse prior," *IEEE Signal Processing Letters*, vol. 26, no. 11, pp. 1688–1692, 2019.

[18] E. Ozanich, P. Gerstoft, and H. Niu, "A feedforward neural network for direction-of-arrival estimation," *The journal of the acoustical society of America*, vol. 147, no. 3, pp. 2035–2048, 2020.

[19] M. L. Bencheikh and Y. Wang, "Joint dod-doa estimation using combined esprit-music approach in mimo radar," *Electronics letters*, vol. 46, no. 15, p. 1, 2010.

[20] J. T. Rogers, J. E. Ball, and A. C. Gurbuz, "Data-driven covariance estimation," in *2022 IEEE International Symposium on Phased Array Systems & Technology (PAST)*. IEEE, 2022, pp. 1–5.