Neural Network-Based DOA Estimation in the Presence of Non-Gaussian Interference

S. Feintuch, J. Tabrikian, *Fellow, IEEE*, I. Bilik, *Senior Member, IEEE*, and H. Permuter, *Senior Member, IEEE*

Abstract—This work addresses the problem of direction- of-arrival (DOA) estimation in the presence of non-Gaussian, heavy-tailed, and spatially-colored interference. Conventionally, the interference is considered to be Gaussian-distributed and spatially white. However, in practice, this assumption is not guaranteed, which results in degraded DOA estimation performance. Maximum likelihood DOA estimation in the presence of non-Gaussian and spatially colored interference is computationally complex and not practical. Therefore, this work proposes a neural network (NN) based DOA estimation approach for spatial spectrum estimation in multi-source scenarios with *a-priori* unknown number of sources in the presence of non-Gaussian spatially-colored interference. The proposed approach utilizes a single NN instance for simultaneous source enumeration and DOA estimation. It is shown via simulations that the proposed approach significantly outperforms conventional and NN-based approaches in terms of probability of resolution, estimation accuracy, and source enumeration accuracy in conditions of low SIR, small sample support, and when the angular separation between the source DOAs and the spatially-colored interference is small.

Index Terms—Array Processing, DOA Estimation, Source Enumeration, Spatially-Colored Interference, Non-Gaussian Interference, Neural Networks, Deep Learning, Machine Learning, MVDR, MDL, AIC, Radar.

I. INTRODUCTION

Direction-of-arrival (DOA) estimation using a sensor array is required in multiple applications, such as radar, sonar, ultrasonic, wireless communications, and medical imaging [1]. In real-world applications, the signal received at the sensor array is a superposition of signals from the sources of interest, interference, and receiver thermal noise. In radars, the received signal consists of a target echo, clutter, and thermal noise. In multiple scenarios, the radar clutter has a spatially-colored, heavy-tailed non-Gaussian distribution [2], which can significantly degrade the performance of conventional estimators.

Minimum-variance-distortionless-response (MVDR) [3], is a conventional adaptive beamforming approach for DOA estimation. MVDR estimates the spatial spectrum and obtains the source DOAs via a one-dimensional peak search on a predefined grid. The estimation of signal parameters using rotational invariance techniques (ESPRIT) [4], multiple signal classification (MUSIC) [5], and root-MUSIC (R-MUSIC) [6] are additional widely used DOA estimation approaches. These approaches involve received signal autocorrelation matrix processing, which conventionally is performed via the sample autocorrelation matrix estimation [3]–[6]. However, the

Stefan Feintuch, Joseph Tabrikian, Igal Bilik, and Haim H. Permuter are with the School of Electrical and Computer Engineering, Ben Gurion University of the Negev, Beer Sheva, Israel. (e-mails: stefanfe@post.bgu.ac.il, joseph@bgu.ac.il, bilik@bgu.ac.il, haimp@bgu.ac.il)

performance of the sample autocorrelation matrix estimator degrades in small sample support or non-Gaussian scenarios. Furthermore, these methods use the second-order statistics only and omit the higher-order statistics on non-Gaussian-distributed interference. In addition, ESPRIT, MUSIC, and R-MUSIC approaches require *a-priori* knowledge of the number of sources (or targets), which limits their practical use.

The problem of DOA estimation in the presence of non-Gaussian interference is of great practical interest. The maximum likelihood estimator (MLE) for DOA estimation in the presence of non-Gaussian interference does not have a closed-form analytical solution [7], [8]. Multiple model-based DOA estimation approaches have been intensively studied in the literature [7]–[18].

Robust covariance matrix-based DOA estimation and source enumeration methods have been studied in the literature. For complex elliptically symmetric (CES) distributed data, the authors in [9] showed that a scatter matrix-based beamformer is consistent, and the semiparametric lower bound and Slepian-Bangs formula for DOA estimation were derived in [10]. In [11], a generalized covariance-based (GC) approach for the covariance matrix estimation in scenarios with impulsive alpha-stable noise was proposed for MUSIC DOA estimation. However, these methods consider a specific family of distributions, such as the CES or alpha-stable, and are therefore, limited in the case of model mismatch. In [12], a probability measure transform (MT) based covariance matrix estimator was proposed for MUSIC-based DOA estimation and minimum descriptive length (MDL) based source enumeration. The MT-based covariance estimator was also adopted for robust MVDR beamformer [13]. These methods are usually based on setting a parameter that determines the tradeoff between the level of robustness and performance.

The problem of DOA estimation in the presence of a mixture of spatially-white K-distributed and Gaussian-distributed noise under a deterministic and unknown (conditional) source model was studied in [7]. An iterative MLE-based approach for the conditional and joint likelihood of interference distribution's parameters was derived in [14], [15]. This approach was further extended in [16] to marginal likelihood function. However, this approach is computationally complex due to numerical integral evaluation that involves a $2M$ dimensional grid search for M targets [8]. Therefore, [8] proposed a kernel minimum error entropy-based adaptive estimator and a novel criterion to reduce the estimator's computational complexity. The expectation-maximization (EM) with a partial relaxation-based DOA estimation algorithm under the conditional model assumption was proposed in [17]. In [18] a sparse Bayesian learning (SBL) approach for outlier rejection of impulsive

and spatially-white interference was proposed. This EM-based approach does not require *a-priori* knowledge of the number of sources and was shown to resolve highly-correlated and coherent sources. However, none of these model-based DOA estimation approaches considered an *a-priori* unknown number of sources and spatially-colored interference and therefore are limited for real-world applications. Although source enumeration methods, such as MDL and Akaike information criterion (AIC) [19] can be used, they assume signal Gaussianity, and can therefore be inaccurate in non-Gaussian scenarios.

Deep learning and machine learning approaches were recently adopted for radar signal processing. Three types of NN-based DOA estimation approaches have been introduced in literature [20]. The first approach assumes *a-priori* known number of sources, and uses a NN, which is optimized to output a vector of the estimated DOAs [21]–[27]. The second approach does not assume *a-priori* known number of sources and uses a NN for source enumeration [25]–[31]. The third approach uses a NN to estimate source presence probability at each DOA on a predefined angular grid and obtains the source DOAs via a peak search [32]–[41]. However, all these approaches have not addressed non-Gaussian and spatially-colored interference [20]–[41].

The cases of non-Gaussian and/or spatially-colored interference have been addressed using machine learning-based approaches. For massive MIMO cognitive radar, a reinforcement learning-based approach for multi-target detection under heavy-tailed spatially-colored interference was proposed in [42]. In [43], authors addressed the MIMO radar target detection under non-Gaussian spatially-colored interference by using a CNN architecture that is optimized according to a novel loss. A radial-basis-function (RBF) NN [44] and a convolutional neural network (CNN) [45] architectures were proposed for DOA estimation in the presence of non-Gaussian impulsive noise. In [46], a CNN-based architecture that includes denoising NN, source enumeration NN, and DOA estimation sub-NNs, was introduced. However, [44]–[46] consider spatially-white noise and are suboptimal for scenarios with spatially-colored interference.

This work addresses the problem of DOA estimation of *a-priori* unknown number of sources in the presence of non-Gaussian, heavy-tailed, spatially-colored interference at a low signal-to-interference ratio (SIR) and small sample size. The contribution of this work include:

1) A novel NN-based processing mechanism is used for array processing within non-Gaussian spatially-colored interference. The proposed NN architecture utilizes the structure of information within the set of received complex snapshots.
2) The proposed NN is optimized to output an interference-mitigated spatial spectrum, and is used for simultaneous source enumeration and DOA estimation of sources within non-Gaussian spatially-colored interference.

The proposed approach outperforms conventional adaptive beamforming and competing straightforward NN-based methods in terms of probability of resolution and estimation accuracy in scenarios with non-Gaussian spatially-colored interference. In addition, the proposed approach outperforms conventional source enumeration techniques in scenarios characterized by non-Gaussian spatially-colored interference.

The following notations are used throughout the paper. Roman boldface lower-case and upper-case letters represent vectors and matrices, respectively while Italic letters stand for scalars. $\mathbf{I}_N$ is the identity matrix of size $N \times N$ and $\mathbf{1}_N$ is a column vector of length N whose entries are equal to one. $\mathbb{E}(\cdot)$, $(\cdot)^T$, and $(\cdot)^H$ are the expectation, transpose, and Hermitian transpose operators, respectively. $\text{Vec}(\cdot)$, $\text{diag}(\cdot)$, and $|\cdot|$ stand for the vectorization, diagonalization, and absolute value operators, respectively. $[\mathbf{a}]_n$ and $[\mathbf{A}]_{n,m}$ are the n-th and n,m-th elements of the vector $\mathbf{a}$ and the matrix $\mathbf{A}$, respectively.

The remainder of this paper is organized as follows. The addressed problem is stated in Section II. Section III introduces the proposed NN-based DOA estimation approach. The proposed approach is evaluated via simulations in Section IV. Our conclusions are summarized in Section V.

II. PROBLEM DEFINITION

This work considers the problem of DOA estimation using an array of L receiving elements and M distinct and unknown sources with DOAs, $\mathbf{\Theta} = \{\theta_1, \ldots, \theta_M\}$. The measurements contain K spatial snapshots, $\{\mathbf{x}_k\}_{k=1}^K$:

$$\mathbf{x}_k = \mathbf{A}(\mathbf{\Theta})\,\mathbf{s}_k + \sigma_c \mathbf{c}_k + \mathbf{n}_k \,, \qquad (1)$$
$$= \sum_{m=1}^{M} \mathbf{a}(\theta_m)\,s_{k,m} + \sigma_c \mathbf{c}_k + \mathbf{n}_k \,, \ k = 1, \ldots, K \,,$$

where $\mathbf{A}(\mathbf{\Theta}) = \begin{bmatrix} \mathbf{a}(\theta_1) & \cdots & \mathbf{a}(\theta_M) \end{bmatrix}$, with $\mathbf{a}(\theta_m) \in \mathbb{C}^L$ denoting the steering vector for source at direction θ_m, and $\mathbf{s}_k \triangleq \begin{bmatrix} s_{k,1} & \cdots & s_{k,M} \end{bmatrix}^T$ is the source signal vector. We assume an unconditional model [47], where $\{\mathbf{s}_k\} \overset{i.i.d.}{\sim} \mathcal{CN}(\mathbf{0}_M, \text{diag}(\sigma_1^2, \ldots, \sigma_M^2))$, is temporally uncorrelated between pulses. The targets are assumed to be spatially distinct. The receiver thermal noise, denoted by $\mathbf{n}_k$, is considered to be complex Gaussian-distributed $\{\mathbf{n}_k\} \overset{i.i.d.}{\sim} \mathcal{CN}(\mathbf{0}_L, \sigma_n^2 \mathbf{I}_L)$. The heavy-tailed non-Gaussian and spatially-colored interference is modeled by the interference amplitude σ_c, and the interference component $\mathbf{c}_k \in \mathbb{C}^L$. The considered compound-Gaussian distributed interference, $\{\mathbf{c}_k\} \overset{i.i.d.}{\sim} \mathcal{K}(\nu, \theta_c)$ represents a non-Gaussian interference with angular spread around an unknown direction θ_c, such that $\mathbf{c} \sim \mathcal{K}(\nu, \theta_c)$ implies

$$\mathbf{c} = \sqrt{\tau}\mathbf{z} \,, \qquad (2)$$
$$\tau \perp\!\!\!\perp \mathbf{z}, \ \tau \sim \Gamma(\nu, \nu) \,, \ \mathbf{z} \sim \mathcal{CN}(\mathbf{0}_L, \mathbf{M}_{\theta_c}) \,.$$

The compound-Gaussian statistical model is conventionally used in the literature to model heavy-tailed non-Gaussian interference [7], [8], [14], [16], [43], [48]. The texture component, $\tau \in \mathbb{R}_+$, determines the heavy-tailed behavior and is characterized by, ν. The speckle component, $\mathbf{z} \in \mathbb{C}^L$, determines the spatial distribution of the interference and is characterized by the covariance matrix, $\mathbf{M}_{\theta_c}$. The spatial covariance matrix of the interference upholds:

$$\mathbb{E}\left[\sigma_c^2 \mathbf{c}\mathbf{c}^H\right] = \sigma_c^2 \mathbb{E}[\tau]\,\mathbb{E}\left[\mathbf{z}\mathbf{z}^H\right] = \sigma_c^2 \mathbf{M}_{\theta_c} \,, \qquad (3)$$

where $\mathbf{M}_{\theta_c}$ can be modeled as [14]–[16], [43], [48]:

$$[\mathbf{M}_{\theta_c}]_{m,l} = \rho^{|m-l|} e^{j(m-l)\pi \sin \theta_c} \ . \tag{4}$$

The model in (3) and (4), represents the spatial interference, characterized by ρ, with a spread around the interference DOA, θ_c.

III. THE PROPOSED DAFC-BASED NEURAL NETWORK

The proposed approach generalizes the NN architecture that was introduced for linear-frequency-modulated (LFM) radar target detection in the range-Doppler domain [49]. In the following, the data pre-processing and the proposed NN-based processing mechanism are introduced in Subsections III-A and III-B. The proposed NN architecture and loss function are detailed in Subsections III-C and III-D, respectively.

A. Pre-Processing

The input matrix, $\mathbf{X} \in \mathbb{C}^{L \times K}$ is constructed from the set of K snapshots in (1), $\{\mathbf{x}_k\}$:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_K \end{bmatrix} \ , \tag{5}$$

where the k-th column of $\mathbf{X}$ contains the k-th snapshot. The variation between the columns of $\mathbf{X}$ is induced by the statistical characteristics of the source signal $\mathbf{s}_k$, interference signal $\mathbf{c}_k$, and thermal noise $\mathbf{n}_k$. Therefore, each column in $\mathbf{X}$ can be interpreted as a complex "feature" vector containing essential information for DOA estimation. The set of columns in $\mathbf{X}$ can be interpreted as "realizations" of that feature.

The complex-valued matrix, $\mathbf{X}$, is converted into real-valued representation needed for the NN-based processing. To keep consistency with [49], we apply a transpose operator to the input matrix, such that the snapshots are stacked in rows. The output of the pre-processing denoted by $\mathbf{Z}_0 \in \mathbb{C}^{K \times 2L}$, is:

$$\mathbf{Z}_0 = \begin{bmatrix} \mathrm{Re}\{\mathbf{X}^T\}, & \mathrm{Im}\{\mathbf{X}^T\} \end{bmatrix} \ . \tag{6}$$

B. Dimensional Alternating Fully-Connected

The dimensional alternating fully-connected (DAFC) block was introduced to process measurements in a form similar to the model in Section II [49]. Fig. 1 schematically shows the DAFC mechanism.

For arbitrary dimensions D_1, D_2, D_3, the formulation of a general fully-connected (FC) layer applied to each row in a given matrix $\mathbf{Z} \in \mathbb{R}^{D_1 \times D_2}$ can be represented by the transform $\mathcal{F}(\cdot)$:

$$\mathcal{F} : \mathbb{R}^{D_1 \times D_2} \rightarrow \mathbb{R}^{D_1 \times D_3} \ , \tag{7}$$
$$\mathcal{F}(\mathbf{Z}) \triangleq h\left(\mathbf{Z}\mathbf{W} + \mathbf{1}_{D_1}\mathbf{b}^T\right) \ .$$

This matrix-to-matrix transformation is characterized by the "learnable" weight matrix, $\mathbf{W} \in \mathbb{R}^{D_2 \times D_3}$, the bias vector, $\mathbf{b} \in \mathbb{R}^{D_3}$, and a scalar element-wise activation function, $h(\cdot)$.

Let $\mathcal{F}_r(\cdot)$ and $\mathcal{F}_c(\cdot)$ be two separate, and not necessarily identical instances of $\mathcal{F}(\cdot)$ from (7), and $\mathbf{Z}_{in}$ be an arbitrary input matrix. The DAFC mechanism is formulated by the following operations:

Dimensional Alternating Fully Connected

- **Input:** $\mathbf{Z}_{in} \in \mathbb{R}^{H \times W}$
  $$\mathcal{F}_r : \mathbb{R}^{H \times W} \rightarrow \mathbb{R}^{H \times W'}$$
  $$\mathcal{F}_c : \mathbb{R}^{W' \times H} \rightarrow \mathbb{R}^{W' \times H'}$$

  1) Apply a single FC layer to each row in $\mathbf{Z}_{in}$:
  $$\mathbf{Z}_r = \mathcal{F}_r(\mathbf{Z}_{in})$$

  2) Apply a single FC layer to each column in $\mathbf{Z}_r$:
  $$\mathbf{Z}_c = \mathcal{F}_c\left(\mathbf{Z}_r^T\right)$$

  3) Transpose to keep orientation:
  $$\mathbf{Z}_{out} = \mathbf{Z}_c^T$$

- **Output:** $\mathbf{Z}_{out} \triangleq \mathcal{S}(\mathbf{Z}) \in \mathbb{R}^{H' \times W'}$

In the following, three DAFC design principles are detailed.
1) Structured transformation
The input to the first DAFC block is the pre-processed, $\mathbf{Z}_0$, given in (6). Therefore, the first FC layer, $\mathcal{F}_r$, of the first DAFC block extracts spatial-related features from each row in $\mathbf{Z}_0$. The second FC layer, $\mathcal{F}_c$, of the first DAFC block, introduces an interaction between transformed rows. This implies that a) $\mathcal{F}_r$ performs "spatial-feature" extraction by transforming the pre-processed i.i.d. snapshots (the rows of $\mathbf{Z}_0$) to a high-dimensional feature space, and b) the $\mathcal{F}_c$ performs a nonlinear transformation of the extracted features (the columns of $\mathcal{F}_r(\mathbf{Z}_0)$) from each snapshot. In this way, the DAFC utilizes both spatial and statistical information. In addition, it can exploit high-order statistics-related features. Thus, the DAFC mechanism can contribute to estimating the source DOAs and mitigating the interference when incorporated into a NN.
2) Sparsity
Conventional DOA estimation considers the input data as the collection of measurement vectors (the snapshots $\{\mathbf{x}_k\}$) in a matrix form. One straightforward approach to processing the input data using a NN is to reshape it and process it via an FC-based architecture. In this way, each neuron in the layer's output interacts with every neuron in the input. On the other hand, the DAFC block transforms the data using a structured transformation, which is significantly sparser in terms of learnable parameters compared to the straightforward FC-based approach.

This parameter reduction can be observed in the following typical case. Consider an input matrix $\mathbf{Z}_1 \in \mathbb{R}^{D_1 \times D_1}$, which is transformed to an output matrix $\mathbf{Z}_2 \in \mathbb{R}^{D_2 \times D_2}$. The number of learnable parameters in the FC- and the proposed DAFC-based approaches is of the order of $\mathcal{O}\left(D_1^2 D_2^2\right)$, and $\mathcal{O}\left(D_1 D_2\right)$, respectively. Notice that the DAFC-based transformation complexity grows linearly with the number of learnable parameters compared to the quadratic complexity growth of the straightforward, FC-based approach.

The contribution of learnable parameters dimension reduction is twofold. First, the conventional NN optimization is gradient-based [50]. Therefore, a significant reduction in the learnable parameter dimension reduces the degrees of freedom in the optimizable parameter space and improves the gradient-based learning algorithm convergence rate. Second, reduction
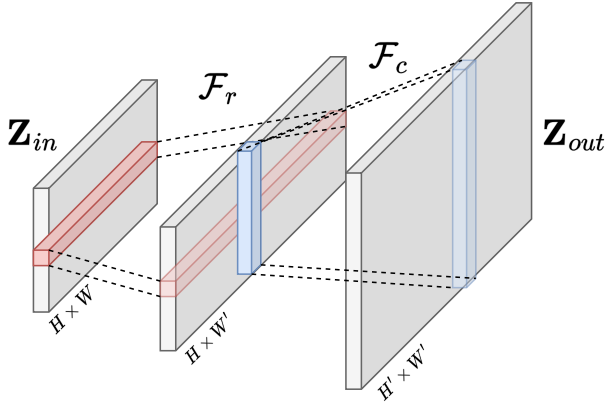
Figure 1: The DAFC mechanism concept. Each row of dimension W in $\mathbf{Z}_{in}$, represented by the red color, is transformed by $\mathcal{F}_r$ to a row of dimension W' in the middle matrix, represented by the transparent red color. Next, each column of dimension H in the middle matrix, represented by the blue color, is transformed by $\mathcal{F}_c$ to a column of dimension H' in $\mathbf{Z}_{out}$, represented by the transparent blue color.

in the learnable parameter dimension can be interpreted as increasing the "inductive bias" of the NN model [51], which conventionally contributes to the NN statistical efficiency and generalization ability, thus, reducing the NNs tendency to overfit the training data.

3) Nonlinearity

The proposed DAFC considers an additional degree of nonlinearity compared to the straightforward FC-based approach. A straightforward matrix-to-matrix approach includes an interaction of every neuron in the output matrix with every neuron in the input matrix, followed by an element-wise nonlinear activation function. On the other hand, the proposed DAFC consists of two degrees of nonlinearity, in $\mathcal{F}_r$ and $\mathcal{F}_c$. Although the weight matrices applied as part of $\mathcal{F}_r$ and $\mathcal{F}_c$ are of lower dimension than the weight matrix used in the straightforward approach, the extra degree of nonlinearity can increase the NN's capacity [50]. Therefore, a NN architecture with the proposed DAFC is capable of learning a more abstract and rich transformation of the input data.

C. NN Architecture

The continuous DOA space is discretized into a d-dimensional grid: $\boldsymbol{\phi} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_d \end{bmatrix}^T$. This implies that the entire field-of-view (FOV) is partitioned into d DOAs, $\{\phi_i\}_{i=1}^d$, determined by the selected grid resolution, $\Delta\phi \triangleq \phi_{i+1} - \phi_i$. The proposed NN is designed to represent a mapping from the input set of snapshots, $\{\mathbf{x}_k\}$ given in (1), into the probability of source present in the DOAs $\{\phi_i\}_{i=1}^d$. The proposed NN architecture is formulated as follows:

$$\mathbf{Z}_0 = \mathcal{P}\left(\mathbf{X}\right) , \tag{8}$$
$$\mathbf{z}_{\text{vec}} = \text{Vec}\left(\mathcal{S}_6\left(\ldots \mathcal{S}_1\left(\mathbf{Z}_0\right)\right)\right) ,$$
$$\hat{\mathbf{y}} = \mathcal{G}_3\left(\mathcal{G}_2\left(\mathcal{G}_1\left(\mathbf{z}_{\text{vec}}\right)\right)\right) ,$$

Operator	Output Dimension	Activation	# Parameters
$\mathcal{P}$	$K \times 2L$	-	-
$\mathcal{S}_1$	64×256	tanh-ReLu	9,536
$\mathcal{S}_2$	128×512	tanh-ReLu	139,904
$\mathcal{S}_3$	256×1024	tanh-ReLu	558,336
$\mathcal{S}_4$	64×512	tanh-ReLu	541,248
$\mathcal{S}_5$	16×256	tanh-ReLu	132,368
$\mathcal{S}_6$	4×128	tanh-ReLu	32,964
vec	512	-	-
$\mathcal{G}_1$	1024	tanh	525,312
$\mathcal{G}_2$	256	tanh	262,400
$\mathcal{G}_3$	d	sigmoid	31,097

Table I: Specification of the proposed NN architecture for $K = 16$, $L = 16$, $d = 121$. "tanh-ReLu" activation stands for tanh in $\mathcal{F}_r$ and ReLU in $\mathcal{F}_c$ of each DAFC block. The number of total learnable parameters is $2,233,165$.

where $\mathbf{Z}_0$ is the output of the pre-processing procedure, denoted as $\mathcal{P}\left(\cdot\right)$ and detailed in Section III-A, and $\mathbf{X}$ is the input matrix in (5).

In the next stage, six DAFC instances, represented by $\mathcal{S}_1\left(\cdot\right), \ldots, \mathcal{S}_6\left(\cdot\right)$, of different dimensions with tanh activation for the row transform ($\mathcal{F}_r$ in Section III-B) and ReLu activation for the column transform ($\mathcal{F}_c$ in Section III-B), are used to generate the vectorized signal $\mathbf{z}_{\text{vec}}$. Our experiments showed that this configuration of row and column activation functions provides the best performance. At the last stage, the signal, $\mathbf{z}_{\text{vec}}$, is processed by three FC layers, where the first two use tanh activation, and the final (output) layer of equal size to the DOA grid dimension, d, uses sigmoid activation function to output $\hat{\mathbf{y}} \in [0, 1]^d$. Thus, $\{[\hat{\mathbf{y}}]_i\}_{i=1}^d$ represent the estimated probabilities of a source presence at $\{\phi_i\}_{i=1}^d$. Table I and Fig. 2 summarize the parameters and architecutre of the proposed NN-based approach.

The estimated source DOAs are extracted from the spatial spectrum via *peak_search* and applying 0.5 threshold:

$$\{i_1, \ldots, i_{\hat{N}}\} = peak\_search\left(\{[\hat{\mathbf{y}}]_i\}_{i=1}^d\right) \tag{9}$$
$$\hat{\boldsymbol{\Theta}} = \left\{\phi_{i_n} : [\hat{\mathbf{y}}]_{i_n} > 0.5\right\}_{n=1}^{\hat{N}} .$$

Namely, the set of estimated DOAs, $\hat{\boldsymbol{\Theta}}$, consists of the grid points corresponding to the peaks of $\hat{\mathbf{y}}$ that exceed the 0.5 threshold. The number of peaks that exceed this threshold is used for source enumeration, and therefore the proposed NN can be utilized as a source enumeration method as well.
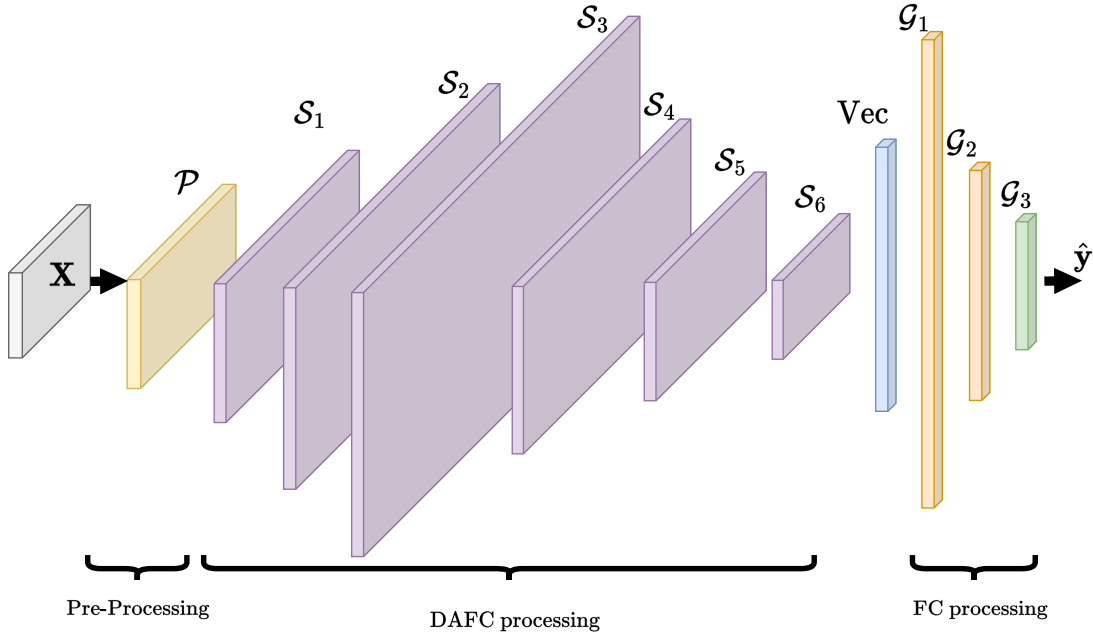
The dimensionality of the hidden layers in the proposed

Figure 2: Proposed NN architecture. The pre-processing $\mathcal{P}$ is described in Section III-A and appears in yellow. The purple matrices denote the concatenation of DAFC blocks, which is detailed in Section III-B. The blue vector represents a vectorization of the last DAFC output, and the orange vectors stands for FC layers with tanh activation function. The last green vector is the output of the last FC layer, which consists of sigmoid activation function and yields the estimated spatial spectrum $\hat{\mathbf{y}}$.

NN architecture expands in the first layers and then reduces. This trend resembles the NN architecture presented in [49] and characterizes both the DAFC-based and FC-based processing stages. This expansion-reduction structure can be explained by a) the early NN stages need to learn an expressive and meaningful transformation of the input data by mapping it to a higher dimensional representation and b) the late stages need to extract significant features from the early mappings, and are therefore limited in dimensionality. In addition, the late stages are adjacent to the output vector and therefore need to be of similar dimension.

D. Loss Function

The label used for the supervised learning process, $\mathbf{y} \in \{0,1\}^d$, is defined as a sparse binary vector with the value 1, at the grid points that correspond to the source DOAs, and 0, otherwise. In practice, the DOAs in Θ do not precisely correspond to the grid points. Therefore, for each DOA in Θ, the nearest grid point in $\{\phi_i\}_{i=1}^d$ is selected as the representative grid point in the label. Each training example is determined by the input-label pair, $(\mathbf{X}, \mathbf{y})$. Using the NN feed-forward in (8), $\mathbf{X}$ is used to generate the output spatial spectrum, $\hat{\mathbf{y}}$, $\hat{\mathbf{y}}$ is considered as the estimated label.

The loss function, $\mathcal{L}$, is a weighted mean of the binary cross entropy (BCE) loss computed at each grid point:

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}, t) = \frac{1}{d} \sum_{i=1}^{d} w_i^{(t)} \mathrm{BCE}\left([\mathbf{y}]_i, [\hat{\mathbf{y}}]_i\right), \quad (10)$$

$$\mathrm{BCE}(y, \hat{y}) = -y \log(\hat{y}) - (1-y) \log(1-\hat{y}),$$

where $w_i^{(t)}$ represents the loss weight of the i-th grid point at the t-th epoch. The loss value for equally-weighted BCEs evaluated per grid point ($w_i^{(t)} = 1$ in (10)) does not significantly increase in the case of a large error in source/interference estimated probability, due to the sparsity of the label $\mathbf{y}$. This forces the NN convergence into a sub-optimal solution that is prone to "miss" the sources. Therefore, the loss weights, $\{w_i^{(t)}\}_{i=1}^d$, are introduced to "focus" the penalty on source/interference grid points.

The loss weight of the i-th grid point, $w_i^{(t)}$, is determined by the presence of source or interference in the corresponding label entry $[\mathbf{y}]_i$. This relation is defined using the epoch and label dependent factors $e_0^{(t)}, e_1^{(t)}$, according to:

$$w_i^{(t)} = \begin{cases} 1/e_1^{(t)}, & \text{if } \phi_i \text{ contains source or interference} \\ 1/e_0^{(t)}, & \text{else} \end{cases}. \quad (11)$$

For $t = 0$, the factor $e_1^{(0)}$ is determined by the fraction of label grid points that contain source or interference out of the total label grid points in the training set, and $e_0^{(0)}$ is the corresponding complement. For subsequent epochs, the factors are updated according to a predefined schedule, similarly to a predefined learning rate schedule. The loss weights are updated N_w times with spacing of Δt epochs during training. The update values are determined by updating $e_0^{(t)}, e_1^{(t)}$, according to the following decaying rule:

$$e_q^{(t)} = (1 - \beta^{(l)}) e_q^{(l\Delta t)} + \beta^{(l)}, \ l\Delta t \le t < (l+1)\Delta t \quad (12)$$
$$q = 0, 1, \ l = 1, \ldots, N_w,$$

where l is the loss weight update iteration, and $\{\beta^{(l)}\}_{l=1}^{N_w}$ represent the loss weight update factors which uphold, $0 \leq \beta^{(l)} \leq 1$. Note that for $N_w \Delta t \leq t$, the weight factor remains $e_i^{(N_w \Delta t)}$ during the rest of the training stage. Notice that as $\beta^{(l)} \to 1$, the corresponding loss weights will tend to be equally distributed across the grid points, i.e., $e_1^{(t)} \approx e_0^{(t)}$. In this case, an erroneously estimated probability for source/interference containing grid point is equally weighted to a neither-containing grid point. On the other hand, as $\beta^{(l)} \to 0$, the corresponding factors will uphold $e_1^{(t)} \ll e_0^{(t)}$, yielding a significantly larger contribution of source/interference containing grid points to the loss value. The rule in (12) enables a "transition of focus" throughout the training. That is, during the early epochs $\beta^{(l)} \to 0$, which contributes more weight to the source/interference containing areas in the estimated label $\hat{\mathbf{y}}$ (i.e., the estimated spatial spectrum) to focus the NN to being correct for source/interference. During the later epochs, $\beta^{(l)}$ is incrementally increased, which relaxes the focus on source/interference from early epochs. Thus, reducing erroneously estimated sources in areas that do not contain source/interference (i.e. "false-alarms").

IV. PERFORMANCE EVALUATION

This section evaluates the performance of the proposed DAFC-based NN approach and compares it to the conventional approaches, summarized in Subsection IV-A1. The data for all considered scenarios is simulated using the measurement model from Section II.

A. Setup & Training

This work considers a uniform linear array (ULA) with half-wavelength-spaced L elements. Each simulated example consists of the input-label pair, $(\mathbf{X}, \mathbf{y})$, where the input $\mathbf{X}$ is defined in (5), and the label $\mathbf{y}$ is defined in Section III-D. The simulation configurations are detailed in Table II. The performance of the proposed approach is evaluated using a **single NN instance**. Therefore, a single NN model is used for various signal-to-interference ratios (SIRs), signal-to-noise ratios (SNRs), interference-to-noise ratios (INRs), DOAs, interference distribution, and the number of sources for joint DOA estimation and source enumeration. The following definitions for the m-th source are used in all experiments:

$$\text{INR} = \frac{\mathbb{E}\left[\|\mathbf{c}\|^2\right]}{\mathbb{E}\left[\|\mathbf{n}\|^2\right]} = \sigma_c^2/\sigma_n^2 , \tag{13}$$

$$\text{SNR}_m = \frac{\mathbb{E}\left[\|\mathbf{a}(\theta_m)s_m\|^2\right]}{\mathbb{E}\left[\|\mathbf{n}\|^2\right]} = \sigma_m^2/\sigma_n^2 , \tag{14}$$

$$\text{SIR}_m = \frac{\mathbb{E}\left[\|\mathbf{a}(\theta_m)s_m\|^2\right]}{\mathbb{E}\left[\|\mathbf{c}\|^2\right]} = \sigma_m^2/\sigma_c^2 . \tag{15}$$

The NN optimization for all evaluated architectures is performed using the loss function in (10) and Adam optimizer [52] with a learning rate of 10^{-3}, and a plateau learning rate scheduler with a decay of 0.905. The set of loss weight update factors, $\{\beta^{(l)}\}_{l=1}^{N_w}$, in (12) is chosen as the evenly-spaced logarithmic scale between 10^{-5} and 10^{-2} with $N_w = 6$, that is $\{10^{-5}, 7.25 \cdot 10^{-5}, 5.25 \cdot 10^{-4}, 3.8 \cdot 10^{-3}, 2.78 \cdot 10^{-2}, 0.2\}$. The chosen batch size is 512, the number of epochs is 500, and early stopping is applied according to the last 200 epochs.

Notation	Description	Value
M_{max}	Maximal number of sources	4
L	Number of sensors	16
K	Number of snapshots	16
d	Angular grid dimension	121
$\Delta\phi$	Angular grid resolution	$1°$
FOV	Field of view	$[-60°, 60°]$
σ_n^2	Thermal noise power	1

Table II: Simulation Configurations.

1) DOA Estimation Approaches: This subsection briefly summarizes the conventional DOA estimation approaches. The performance of the proposed approach is compared to the conventional MVDR, CNN, and FC-based NN. All the NN-based approaches were implemented using similar number of layers and learnable parameters. In addition, the FC-based NN and CNN were optimized using the same learning algorithm and configurations.

(a) Conventional Adaptive Beamforming
The MVDR [3] estimator is based on adaptive beamforming, and it is the maximum likelihood estimator in the presence of unknown Gaussian interference [53]. The MVDR estimates DOAs by a peak search on the MVDR spectrum:

$$P_{MVDR}(\phi) = \frac{1}{\mathbf{a}^H(\phi)\hat{\mathbf{R}}_x^{-1}\mathbf{a}(\phi)} , \tag{16}$$

where $\hat{\mathbf{R}}_x = \frac{1}{K}\sum_{k=1}^{K} \mathbf{x}_k \mathbf{x}_k^H$ is the sample covariance matrix estimator. Notice that the MVDR spectrum utilizes only second-order statistics of the received signal $\mathbf{x}_k$. For Gaussian-only interference (i.e. $\mathbf{c}_k = 0$ in (1)), the second-order statistics contains the entire statistical information. However, for non-Gaussian interference, information from higher-order statistics is needed.

(b) CNN Architecture
We consider a CNN-based DOA estimation approach using a CNN architecture that is similar to the architecture provided in [38]. The input to the CNN of dimension $L \times L \times 3$ consists of the real, imaginary, and angle parts of $\hat{\mathbf{R}}_x$. The CNN architecture consists of 4 consecutive CNN blocks, such that each block contains a convolutional layer, a batch normalization layer, and a ReLu activation. The convolutional layers consist of $[128, 256, 256, 128]$ filters. Kernel sizes of 3×3 for the first block and 2×2 for the following three blocks are used. Similarly to [38], 2×2 strides are used for the first block and 1×1 for the following three blocks. Next, a flatten layer is used to vectorize the hidden tensor, and 3 FC layers of dimensions 1024, 512, 256 are used with a ReLu activation and Dropout of 30%. Finally, the output layer is identical to the proposed DAFC-based NN as detailed in Subsection III-C. The considered loss function is identical to the proposed DAFC-based approach in (10). The number of trainable parameters in the considered CNN

architecture accounts for $3,315,449$. Notice that the CNN-based architecture utilizes the information within the sample covariance matrix and therefore, is limited to second-order statistics only.

(c) FC Architecture

A straightforward implementation of an FC-based architecture, as mentioned in Subsection III-B, was implemented. The data matrix, $\mathbf{X}$, is vectorized, and the real and imaginary parts of the values were concatenated to obtain a $2KL$-dimension input vector. The selected hidden layers are of sizes: $[512, 512, 1024, 1024, 512, 256]$ where each hidden layer is followed by a tanh activation function. The output layer is identical to the proposed DAFC-based NN approach as detailed in Subsection III-C. The considered loss function is (10), and the number of trainable parameters in the FC-based NN accounts for $2,787,449$. Notice that the FC-based NN architecture utilizes all the measurements by interacting with all samples in the input data. However, this processing is not specifically tailored to the structure of information within the measurements. On the other hand, the proposed DAFC-based NN utilizes the information structure to process the input data. Therefore, for the considered DOA estimation problem, the "inductive bias" [51] for this approach is improper and can result in under-fitted NN architecture.

2) Performance Evaluation Metrics: This subsection discusses the criteria for the performance evaluation of the proposed DOA estimation approach. In this work, similarly to [38], the DOA estimation accuracy of a set of sources is evaluated by the Hausdorff distance between sets. The Hausdorff distance, d_H between the sets, $\mathcal{A}$, and $\mathcal{B}$, is defined as:

$$d_H(\mathcal{A}, \mathcal{B}) = \max\{d(\mathcal{A}, \mathcal{B}), d(\mathcal{B}, \mathcal{A})\} , \quad (17)$$
$$d(\mathcal{A}, \mathcal{B}) = \sup\{\inf\{|\alpha - \beta| : \beta \in \mathcal{B}\} : \alpha \in \mathcal{A}\} .$$

Notice that $d(\mathcal{A}, \mathcal{B}) \neq d(\mathcal{B}, \mathcal{A})$. Let $\boldsymbol{\Theta} = \{\theta_m\}_{m=1}^{M}$ and $\hat{\boldsymbol{\Theta}} = \{\hat{\theta}_m\}_{m=1}^{\hat{M}}$ be the sets of true and estimated DOAs, respectively. The estimation error is obtained by evaluating the Hausdorff distance, $d_H(\boldsymbol{\Theta}, \hat{\boldsymbol{\Theta}})$. We define the root mean squared distance (RMSD) for an arbitrary set of N examples (e.g., test set), $\{\mathbf{X}^{(n)}, \mathbf{y}^{(n)}\}_{n=1}^{N}$, with the corresponding true and estimated DOAs, $\{\boldsymbol{\Theta}^{(n)}, \hat{\boldsymbol{\Theta}}^{(n)}\}_{n=1}^{N}$ as:

$$\text{RMSD} \triangleq \sqrt{\frac{1}{N}\sum_{n=1}^{N} d_H^2\left(\boldsymbol{\Theta}^{(n)}, \hat{\boldsymbol{\Theta}}^{(n)}\right)} . \quad (18)$$

Angular resolution is one of the key criteria for DOA estimation performance. The probability of resolution is commonly used as a performance evaluation metric for angular resolution. In the considered problem, resolution between two sources and between source and interference are used for performance evaluation. For an arbitrary example with M sources, the resolution event A_{res} is defined as:

$$A_{res}\left(\boldsymbol{\Theta}, \hat{\boldsymbol{\Theta}}\right) \triangleq \begin{cases} 1, & \bigcap_{m=1}^{M} \xi_m \leq 2° \text{ and } |\hat{\boldsymbol{\Theta}}| \geq M \\ 0, & \text{else} \end{cases} , \quad (19)$$

$$\xi_m \triangleq \min_{\hat{\theta} \in \hat{\boldsymbol{\Theta}}} |\theta_m - \hat{\theta}|, \ m = 1, \ldots, M .$$

For example, a scene with M sources is considered successfully resolved if for each true DOA a) there exists a close-enough estimated DOA, $\hat{\theta} \in \hat{\boldsymbol{\Theta}}$, that is at most $2°$ apart, and b) there exists at least M DOA estimations. According to (18), the probability of resolution, can be defined as:

$$P_{res} = \frac{1}{N}\sum_{n=1}^{N} A_{res}\left(\boldsymbol{\Theta}^{(n)}, \hat{\boldsymbol{\Theta}}^{(n)}\right) . \quad (20)$$

3) Data Sets: This subsection describes the structure and formation of *Training & Test sets*.

(a) Training Set

The considered training set contains $N_{train} = 10,000$ examples re-generated at each epoch. For each example, i.e. an input-label pair $(\mathbf{X}, \mathbf{y})$, the number of DOA sources, M, is generated from uniform and $i.i.d.$ distribution, $\{1, \ldots, M_{max}\}$. The training set contains 10% of interference-free examples and 90% of interference-containing. Out of the interference-containing examples, 90% generated such that the source DOAs, $\{\theta_m\}_{m=1}^{M}$, and the interference's DOA, θ_c, are distributed uniformly over the simulated FOV. The remaining 10% are generated such that θ_c is distributed uniformly over the FOV, and the source DOAs, $\{\theta_m\}_{m=1}^{M}$, are distributed uniformly over the interval $[\theta_c - 8°, \theta_c + 8°]$. This data set formation enables to "focus" the NN training on the challenging scenarios where the source and interference DOAs are closely spaced. The generalization capabilities of the proposed NN to variations in interference statistics are achieved via the interference angular spread parameter, ρ, from the uniform distribution, U $([0.7, 0.95])$, and the interference spikiness parameter, ν, from the uniform distribution, U $([0.1, 1.5])$. The INR for each interference-containing example and $\{\text{SIR}_m\}_{m=1}^{M}$ or $\{\text{SNR}_m\}_{m=1}^{M}$ are drawn independently according to Table III.

(b) Test Set

The test set consists of $N_{test} = 20,000$ examples. The results are obtained by averaging the evaluated performance over 50 independent test set realizations. Considering the low-snapshot support regime, the number of snapshots is set to $K = 16$, except for experiment (c) in IV-B2. Considering heavy-tailed interference, the spikiness parameter is set to $\nu = 0.2$. The INR is set to INR $= 5 \ dB$, and the interference angular spread parameter is set to $\rho = 0.9$. The signal amplitude was set to be identical for all sources, $\sigma_1 = \cdots = \sigma_m$, except for experiment (b) in IV-B2.

B. Experiments

1) Single Source Within Interference: In this scenario, the ability to resolve a single source from interference is evaluated. Let $M = 1$ with $\theta_1 = 0.55°$, and $\theta_c = \theta_1 + \Delta\theta_c$ such that $\Delta\theta_c$ is the angular separation between the single source and interference. The $0.55°$ offset is considered to impose a realistic off-grid condition. Fig. 3 shows the RMSD and probability of resolution for all evaluated approaches.

Fig. 3a shows that the FC-based NN approach does not manage to resolve the single source from the interference for all evaluated angular separations. This result supports the
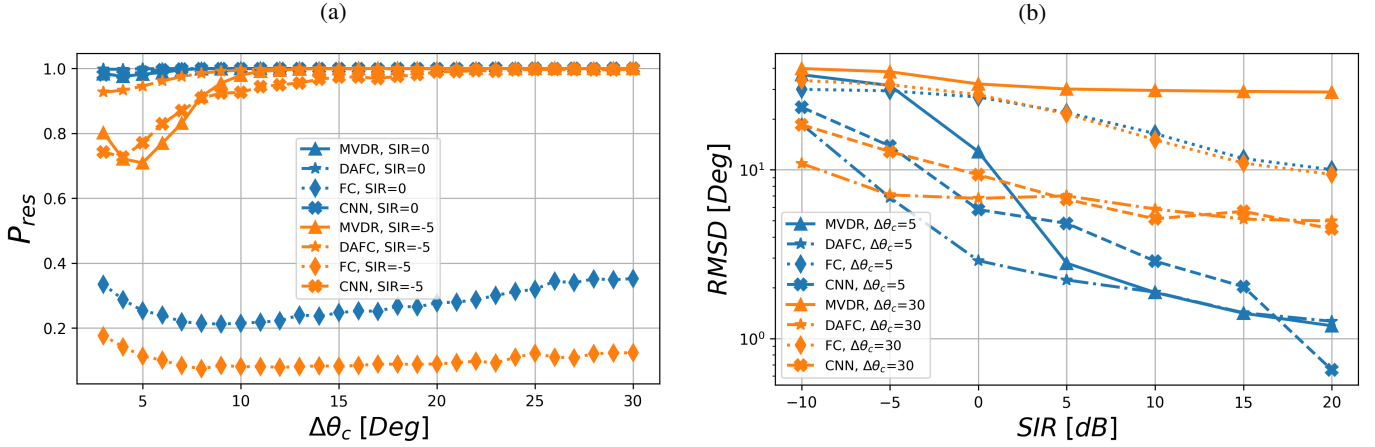
Figure 3: Scenario with a single source at $\theta_1 = 0.55°$ and interference located at $\theta_c = \theta_1 + \Delta\theta_c$. (a) probability of resolution and (b) RMSD.

Notation	Description	Value
ρ	Interference angular spread parameter	$\sim U\left([0.7, 0.95]\right)$
ν	Interference spikiness parameter	$\sim U\left([0.1, 1.5]\right)$
INR	INR	$\sim U\left([0, 10]\right)\ [dB]$
SIR_m	SIR of m-th source	$\sim U\left([-10, 10]\right)\ [dB]$
SNR_m	SNR of m-th source	$\sim U\left([-10, 10]\right)\ [dB]$

Table III: Training set parameters. SNR_m distribution applies to interference-free examples.

under-fitting limitation of the FC-based NN approach for the DOA estimation, which can be explained by the architecture that processes the input data as-is, without any structured transformation or model-based pre-processing.

The MVDR and CNN performance in terms of the resolution are similar since both rely only on second-order statistics, which is sufficient in scenarios with widely separated sources and interference. Fig. 3a shows that the proposed DAFC-based NN approach outperforms all other considered approaches in low angular separation scenarios. This can be explained by the fact that the DAFC uses the high-order statistics needed for the resolution of closely spaced sources and interference.

Fig. 3b shows the RMSD of all considered DOA estimation approaches. The proposed DAFC-based NN approach outperforms the other tested approaches in low SIR. At high SIR and small angular separation, $\Delta\theta_c = 5°$, the interference is negligible with respect to the strong source signal, and therefore, the DAFC-based, CNN, and MVDR approaches obtain similar performance. For large angular separation, $\Delta\theta_c = 30°$, the source and the interference are sufficiently separated, and therefore, DOA estimation errors are mainly induced by the interference DOA, θ_c. The MVDR spectrum contains a peak at $\theta_c = 30.55°$, and therefore, MVDR's

$RMSD = 30°$ is approximately constant. The NNs are trained to output a 0-probability for the interference, therefore, the NN-based approaches: FC, CNN, and DAFC achieve a smaller DOA estimation error. The DAFC-based NN and CNN utilize structured transformations, which better fit the input data, and therefore, they outperform the FC-based NN approach in terms of RMSD.

2) Resolving Two Sources from Interference: This subsection evaluates the performance of the tested DOA estimation approaches in scenarios with two sources within AWGN and interference.

(a) Resolution of Equal-Strength Sources
In the following experiment, the resolution between two equal-power sources, $M = 2$, with $\theta_1 = -\frac{\Delta\theta}{2} + 0.55°$, and $\theta_2 = \frac{\Delta\theta}{2} + 0.55°$, is evaluated. The off-grid additional $0.55°$ offset to the $\Delta\theta$ angular separation between the sources represents the practical scenario. The interference at $\theta_c = 0.55°$ influences the two sources similarly. Fig. 4 shows the probability of resolution of the tested approaches in scenarios with (a) the AWGN only and (b) spatially-colored interference.

The FC-based NN approach does not resolve the two targets in both evaluated scenarios. Subplot (a) in Fig. 4 shows that the proposed DAFC-based NN approach outperforms the MVDR and the CNN at low-SNR and small angular separation scenarios due to its generalization ability to spatially-white interference. Subplot (b) in Fig. 4 shows that at low SIR of $SIR = -5\ dB$, the performances of MVDR and CNN significantly degrade compared to the proposed DAFC-based NN approach. Comparing subplots in Fig. 4, notice that at $SIR = -5\ dB$, the MVDR fails to resolve the sources with angular separation $\Delta\theta < 20°$ due to the presence of the heavy-tailed spatially-colored interference in the proximity of the sources. However, the proposed DAFC-based NN approach mitigates this interference and resolves the sources, and hence, outperforms other tested approaches at both $SIR = 0\ dB$ and $SIR = -5\ dB$.

Subplot (b) in Fig. 4 shows the non-monotonic trend of CNN and MVDR performance at $4° < \Delta\theta < 18°$ and
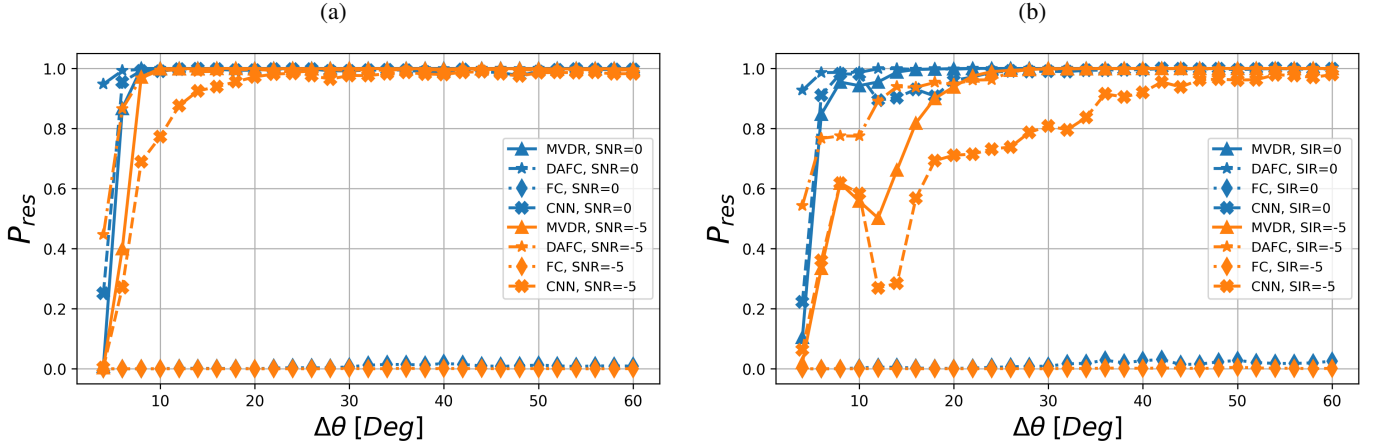
Figure 4: Probability of resolution for two sources located at $\theta_{1,2} = \theta_c \pm \Delta\theta/2$, and interference located at $\theta_c = 0.55°$. (a) AWGN-only scenario and (b) interference-containing scenario.
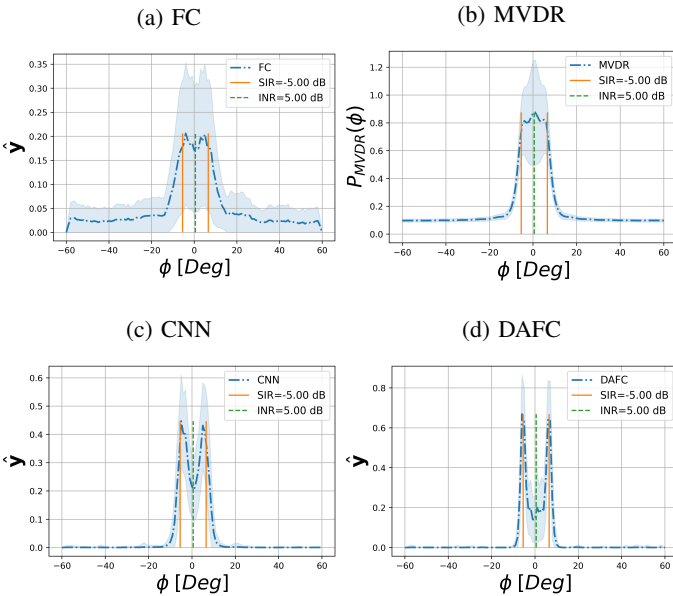

Figure 5: Spatial spectrum, two sources with SIR $= -5\ dB$ located at $\theta_{1,2} = \theta_c \pm \Delta\theta/2$ with $\Delta\theta = 12°$ and $\theta_c = 0.55°$. The dashed blue lines represent the mean spatial spectrum, and the color fill represents the standard deviation around the mean obtained from $2,000$ i.i.d. examples. The solid vertical orange lines represent the true source DOAs, and the dashed vertical green line represents the interference DOA.

SIR $= -5\ dB$. For $4° < \Delta\theta < 8°$ the sources are closer to the peak of the interference's lobe and are therefore less mitigated by it. As $\Delta\theta$ initially increases, $8° < \Delta\theta < 12°$, the sources reach DOAs which are in the proximity of the interference lobe's "nulls" which explains the reduction in resolution, and as $\Delta\theta$ further increases, $16° < \Delta\theta$, the sources are sufficiently separated from the interference such that the resolution increases. As a result, MVDR and CNN-based approaches that use second-order statistics only, can not

resolve the sources in the vicinity of a stronger interference.

Fig. 5 shows the average spatial spectrum of all tested approaches for $\Delta\theta = 12°$ and SIR $= -5\ dB$. The average spatial spectrum of the FC-based NN approach does not show two prominent peaks, which results in its poor probability of resolution in Fig. 4. The MVDR "bell-shaped" spatial spectrum does not contain the two prominent peaks at $\theta_{1,2}$ since the interference "masks" the two sources. The CNN and proposed DAFC-based NN approaches show two peaks at the average spatial spectrum. The peaks at the CNN's average spatial spectrum are lower, resulting in a low-resolution probability. The average spatial spectrum of the proposed DAFC-based NN approach contains two high peaks, resulting in a superior probability of resolution in Fig. 4.

(b) Resolution of Unequal-Power Sources
Fig. 6 shows the probability of resolution in a scenario with two sources, $M = 2$, at $\theta_1 = -\Delta\theta/2 + 0.55°$, and $\theta_2 = +\Delta\theta/2 + 0.55°$ with interference located between the sources at $\theta_c = 0.55°$. The signal strength of the second source is set to $SIR_1 = SIR_2 + 10\ dB$. Comparing Fig. 6 to Fig. 4b, the competing methods show similar trends, except the degradation of the CNN's probability of resolution for the SIR $= 0\ dB$ case. On the other hand, the proposed DAFC-based NN approach outperforms other tested approaches in terms of the probability of resolution. Therefore, Fig. 6 demonstrates the generalization ability of the proposed DAFC-based NN approach to a variance between source strengths.

(c) Effect of the Number of Snapshots on the Resolution
This experiment investigates the influence of the number of snapshots, K, on the ability to resolve two proximate sources from heavy-tailed spatially-colored interference. The equal-strength resolution scenario is repeated using $K = 4,\ 8,\ 16,\ 32,\ 64$ with different instances of NN training for each K value. Fig. 7 shows the probability of resolution for two equal-strength sources at $\theta_{1,2} = \theta_c \pm \Delta\theta/2$ for $\Delta\theta = 12°$ and $\theta_c = 0.55°$.

The FC-based NN approach fails to resolve the two sources. For SIR $= 0\ dB$, the MVDR, CNN, and DAFC-based NN
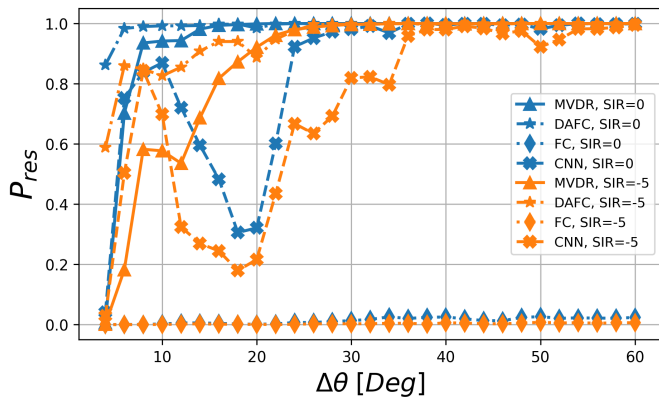
Figure 6: Probability of resolution for two sources located at $\theta_{1,2} = \theta_c \pm \Delta\theta/2$, and interference located at $\theta_c = 0.55°$. The SIR in the legend represents the SIR of the first source, SIR_1. The SIR of the second source is set to $\text{SIR}_2 = \text{SIR}_1 + 10\ dB$.

approaches achieve a monotonic increasing probability of resolution with increasing K. The proposed DAFC-based NN approach slightly outperforms other tested approaches. At low SIR of SIR $= -5\ dB$, the proposed DAFC-based NN approach significantly outperforms the other tested approaches. This can be explained by the fact that increasing K increases the probability for outliers to be present in the input data matrix, $\mathbf{X}$. Therefore, the estimated autocorrelation matrix, $\hat{\mathbf{R}}_x$, is more likely to be biased by the interference-related outliers, which results in interference "masking" the sources. The proposed DAFC-based NN approach is immune to these outliers and successfully exploits the information from the additional snapshots to improve the probability of resolution.
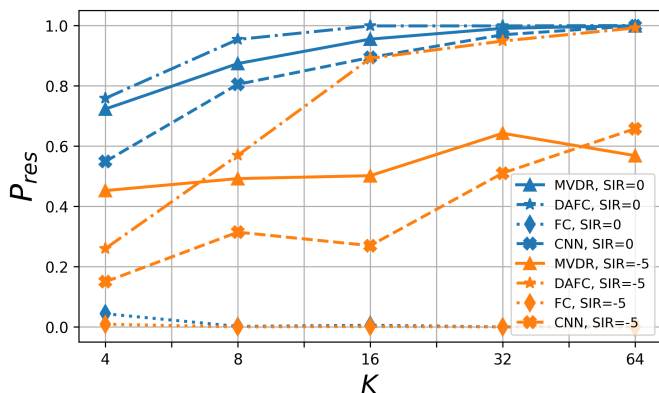

Figure 7: Probability of resolution for two sources located at $\theta_{1,2} = \theta_c \pm \Delta\theta/2$ with $\Delta\theta = 12°$, and interference located at $\theta_c = 0.55°$, as a function of the number of snapshots, K.

Figs. 4, 5, 6, and 7 show the ability of the proposed DAFC-based NN approach to utilize the information structure of the input data by exploiting the higher-order statistics and performing the domain-fitted transformation in order to provide superior resolution ability in the case of proximate heavy-tailed spatially-colored interference, low SIR and small sample size.

3) Multiple Source Localization: The performances of the tested DOA estimation approaches are evaluated and compared in a multi-source scenario. Four sources, $(M = 4)$ were simulated with angular separation, $\Delta\theta$: $\{\theta_1, \theta_2, \theta_3, \theta_4\} = \theta_c + \{-2\Delta\theta, -\Delta\theta, \Delta\theta, 2\Delta\theta\}$, where $\theta_c = 0.51°$ represents a realistic off-grid condition. The RMSD of evaluated methods is depicted in Fig. 8. The proposed DAFC-based NN approach outperforms the other tested approaches at low SIR (SIR $< 0\ dB$) for large and small angular separations. For high SIR and low angular separation, $\Delta\theta = 5°$, the MVDR achieves the lowest RMSD. The reason is that for this case, the interference is negligible with respect to the lobe of the strong source in the MVDR's spectrum. However, at high angular separation, $\Delta\theta = 20°$, the proposed DAFC-based NN approach significantly outperforms the other tested approaches. This is explained by Fig. 9, that shows the spectrum of the tested DOA estimation approaches. Notice that the proposed DAFC-based NN mitigates interference, while the spectra of other tested approaches contain high peaks at the interference DOA, θ_c. These peaks increase the Hausdorff distance in (17), increasing the RMSD of other tested approaches in Fig. 8.
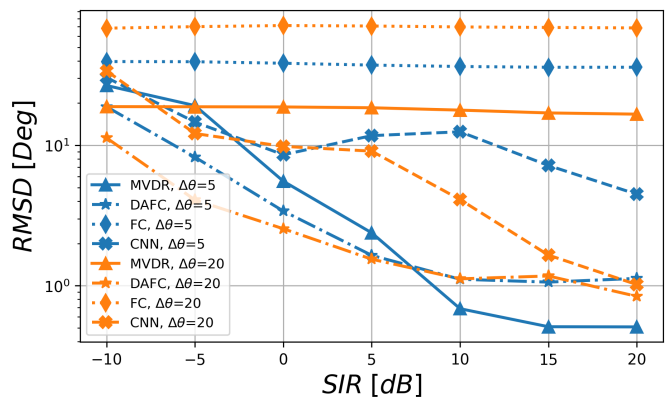

Figure 8: RMSD in scenarios with $M = 4$ sources located at $\{\theta_1, \theta_2, \theta_3, \theta_4\} = \theta_c + \{-2\Delta\theta, -\Delta\theta, \Delta\theta, 2\Delta\theta\}$, where $\theta_c = 0.51°$.

4) Multiple Source Enumeration: The source enumeration performance is evaluated in this experiment. The DOAs of the sources are selected from the set of following values: $\{10.51°, -9.49°, -19.49°, 10.51°\}$ such that for M sources, the DOAs are selected to be the first M DOAs. The interference is located at $\theta_c = 0.51°$. The proposed DAFC-based NN approach is compared to the MDL and AIC [19]. Fig. 10 shows the source enumeration confusion matrices for the MDL, AIC, and the proposed DAFC-based NN with SIR $= 0\ dB$.

Figs. 10a, 10b show that in both the MDL and the AIC, the predicted number of sources has a constant bias for each true M due to the spatially-colored interference. Fig. 10c shows the source enumeration performance of the proposed DAFC-based NN approach in the presence of spatially colored interference. The DAFC-based NN identifies the interference and does not count it as one of the sources by outputting a low probability for angular grid points near θ_c, resulting in a better source enumeration performance.
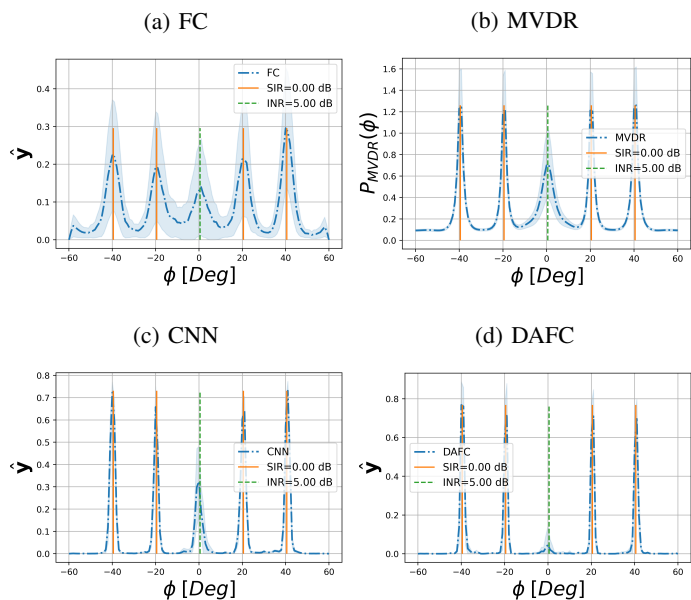
Figure 9: Spatial spectrum, four sources with SIR $= 0$ dB located at $\{\theta_1, \theta_2, \theta_3, \theta_4\} = \theta_c + \{-2\Delta\theta, -\Delta\theta, \Delta\theta, 2\Delta\theta\}$, where $\theta_c = 0.51°$ and $\Delta\theta = 20°$. The dashed blue lines represent the mean spatial spectrum, and the color fill represents the standard deviation around the mean obtained from $2,000$ i.i.d. examples. The solid vertical orange lines represent the true source DOAs and the dashed vertical green line represents the interference DOA.

5) Loss Weights: This experiment evaluates the effect of the loss weight update factors, $\{\beta^{(l)}\}_{l=1}^{N_w}$, introduced in (12), on the confidence level in the spatial spectrum. Let $\widetilde{\mathbf{B}}$ denote the set of $\{\beta^{(l)}\}_{l=1}^{N_w}$ values used in the proposed approach. The loss weights, $\{w_i^{(t)}\}_{i=1}^d$, are defined by the factors $e_0^{(t)}, e_1^{(t)}$ according to (11), and are introduced to provide a trade-off between the penalty obtained on source/interference and the penalty obtained for the rest of the output spatial spectrum.

For comparison, we set $\mathbf{B}_0 = \{10^{-6}, 3.98 \cdot 10^{-6}, 1.58 \cdot 10^{-5}, 6.31 \cdot 10^{-5}, 2.51 \cdot 10^{-4}, 10^{-3}\}$, and $\mathbf{B}_1 = \{10^{-3}, 3.98 \cdot 10-3, 0.0158, 0.063, 0.25, 0.1\}$ as two sets of loss weight update factors. For $\mathbf{B}_0$, the loss weight update factors are closer to 0, hence the loss weights emphasize the source/interference, since $e_1^{(t)} \ll e_0^{(t)}$ which, according to (11), translates to larger $w_i^{(t)}$ for source/interference grid points. For $\mathbf{B}_1$ the values are closer to 1, hence the loss weights are more equally distributed among grid points, since $e_1^{(t)} \approx e_0^{(t)}$. The experiment in IV-B1 is repeated here for the DAFC-based NN approach with the two additional $\mathbf{B}_0, \mathbf{B}_1$ values mentioned above.

Let $\hat{p}_1$ represent the probability assigned for the source-containing grid point in the estimated label $\hat{\mathbf{y}}$. Let $\hat{p}_0$ represent the maximum over probabilities assigned for non-source grid points in $\hat{\mathbf{y}}$, excluding a 5-grid point guard interval around the source. Fig. 11 shows $\hat{p}_1$ and $\hat{p}_0$ for various angular separations between the source and interference for SIR $= -5$ dB. For $\mathbf{B}_0$, the source's contribution to the loss value is substantially higher, which results in a higher probability for the
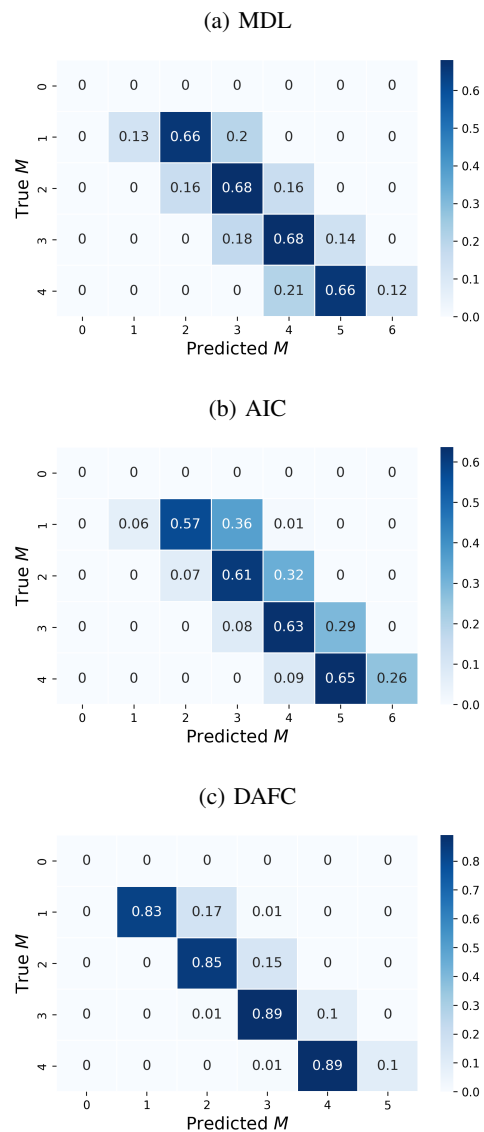

Figure 10: Confusion matrix for source enumeration, SIR $= 0$ dB, sources located at $\{10.51°, -9.49°, -19.49°, 10.51°\}$. (a) MDL, (b) AIC, (c) proposed DAFC-based NN.

source-containing grid point. However, this results in a higher probability obtained for non-source grid points, since their contribution to the loss value is negligible compared to the source-containing grid point, increasing "false-alarm" peaks in the spatial spectrum, subsequently increasing the estimation error. Correspondingly, for $\mathbf{B}_1$ the source's contribution to the loss value is less significant, which results in low probability assigned for the source-containing grid points, as well as low probability for non-source grid points.

V. CONCLUSION

This work addresses the problem of DOA estimation and source enumeration of an unknown number of sources within heavy-tailed, non-Gaussian, and spatially colored interference. A novel DAFC-based NN approach is proposed for this
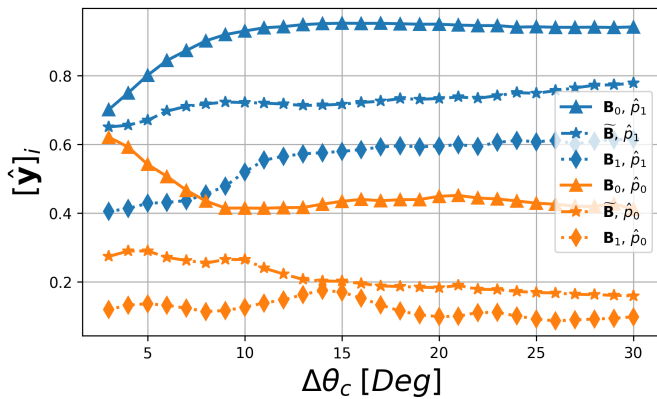
Figure 11: Loss weight update factor impact on probability levels obtained in the DAFC-based NN's spatial spectrum, single target at $\theta_1 = 0.55°$ with interference at $\theta_c = \theta_1 + \Delta\theta_c$, SIR $= -5\ dB$. $\hat{p}_1$ represents the probability obtained for source-containing grid points. $\hat{p}_0$ represents the probability obtained for non-source grid points.

problem. The DAFC mechanism applies a structured transformation capable of exploiting the interference non-Gaussianity for its mitigation while retaining a low complexity of learnable parameters. The proposed DAFC-based NN approach is optimized to provide an interference-mitigated spatial spectrum using a loss weight scheduling routine, performing DOA estimation and source enumeration using a unified NN.

The performance of the proposed approach is compared to MVDR, CNN-based, and FC-based approaches. Simulations showed the superiority of the proposed DAFC-based NN approach in terms of probability of resolution and estimation accuracy, evaluated by RMSD, especially in weak signal power, small number of snapshots, and near-interference scenarios. The source enumeration performance of the proposed DAFC-based NN approach was compared to the MDL and AIC. It was shown that in the considered scenarios, the proposed approach outperforms the MDL and the AIC in the source enumeration accuracy.

REFERENCES

[1] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. John Wiley & Sons, 2004.

[2] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Transactions on signal processing*, vol. 60, no. 11, pp. 5597–5625, 2012.

[3] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[4] 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.

[5] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[6] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *ICASSP'83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 8. IEEE, 1983, pp. 336–339.

[7] O. Besson, Y. Abramovich, and B. Johnson, "Direction-of-Arrival estimation in a mixture of K-distributed and Gaussian noise," *Signal Processing*, vol. 128, pp. 512–520, 2016.

[8] U. K. Singh, R. Mitra, V. Bhatia, and A. K. Mishra, "Kernel minimum error entropy based estimator for mimo radar in non-Gaussian clutter," *IEEE Access*, vol. 9, pp. 125 320–125 330, 2021.

[9] E. Ollila and V. Koivunen, "Influence function and asymptotic efficiency of scatter matrix based array processors: Case MVDR beamformer," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 247–259, 2008.

[10] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir, and M. Rangaswamy, "Semiparametric CRB and Slepian-Bangs formulas for complex elliptically symmetric distributions," *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5352–5364, 2019.

[11] S. Luan, M. Zhao, Y. Gao, Z. Zhang, and T. Qiu, "Generalized covariance for non-Gaussian signal processing and GC-MUSIC under Alpha-stable distributed noise," *Digital Signal Processing*, vol. 110, p. 102923, 2021.

[12] K. Todros and A. O. Hero, "Robust multiple signal classification via probability measure transformation," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1156–1170, 2015.

[13] N. Yazdi and K. Todros, "Measure-transformed MVDR beamforming," *IEEE Signal Processing Letters*, vol. 27, pp. 1959–1963, 2020.

[14] X. Zhang, M. N. El Korso, and M. Pesavento, "Maximum likelihood and maximum a posteriori Direction-of-Arrival estimation in the presence of SIRP noise," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 3081–3085.

[15] ——, "MIMO radar target localization and performance evaluation under SIRP clutter," *Signal Processing*, vol. 130, pp. 217–232, 2017.

[16] B. Meriaux, X. Zhang, M. N. El Korso, and M. Pesavento, "Iterative marginal maximum likelihood DOD and DOA estimation for MIMO radar in the presence of SIRP clutter," *Signal Processing*, vol. 155, pp. 384–390, 2019.

[17] M. Trinh-Hoang, M. N. El Korso, and M. Pesavento, "A partially-relaxed robust DOA estimator under non-Gaussian low-rank interference and noise," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 4365–4369.

[18] J. Dai and H. C. So, "Sparse Bayesian learning approach for outlier-resistant direction-of-arrival estimation," *IEEE Transactions on Signal Processing*, vol. 66, no. 3, pp. 744–756, 2017.

[19] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[20] J. Fuchs, M. Gardill, M. Lübke, A. Dubey, and F. Lurz, "A machine learning perspective on automotive radar direction of arrival estimation," *IEEE Access*, 2022.

[21] J. Fuchs, R. Weigel, and M. Gardill, "Single-snapshot direction-of-arrival estimation of multiple targets using a multi-layer perceptron," in *2019 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*. IEEE, 2019, pp. 1–4.

[22] A. H. El Zooghby, C. G. Christodoulou, and M. Georgiopoulos, "Performance of radial-basis function networks for direction of arrival estimation with antenna arrays," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 11, pp. 1611–1617, 1997.

[23] B. Milovanovic, M. Agatonovic, Z. Stankovic, N. Doncov, and M. Sarevska, "Application of neural networks in spatial signal processing," in *11th Symposium on Neural Network Applications in Electrical Engineering*. IEEE, 2012, pp. 5–14.

[24] G. Ofek, J. Tabrikian, and M. Aladjem, "A modular neural network for direction-of-arrival estimation of two sources," *Neurocomputing*, vol. 74, no. 17, pp. 3092–3102, 2011.

[25] A. Barthelme and W. Utschick, "A machine learning approach to DoA estimation and model order selection for antenna arrays with subarray sampling," *IEEE Transactions on Signal Processing*, vol. 69, pp. 3075–3087, 2021.

[26] O. Bialer, N. Garnett, and T. Tirer, "Performance advantages of deep neural networks for angle of arrival estimation," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 3907–3911.

[27] J. Cong, X. Wang, X. Huang, and L. Wan, "Robust DOA estimation method for MIMO radar via deep neural networks," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 7498–7507, 2020.

[28] M. Gardill, J. Fuchs, C. Frank, and R. Weigel, "A multi-layer perceptron applied to number of target indication for direction-of-arrival estimation in automotive radar sensors," in *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2018, pp. 1–6.

[29] J. Fuchs, R. Weigel, and M. Gardill, "Model order estimation using a multi-layer perceptron for direction-of-arrival estimation in automotive

radar sensors," in *2019 IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet)*. IEEE, 2019, pp. 1–3.

[30] J. Rogers, J. E. Ball, and A. C. Gurbuz, "Estimating the number of sources via deep learning," in *2019 IEEE Radar Conference (Radar-Conf)*. IEEE, 2019, pp. 1–5.

[31] ——, "Robust estimation of the number of coherent radar signal sources using deep learning," *IET Radar, Sonar & Navigation*, vol. 15, no. 5, pp. 431–440, 2021.

[32] 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.

[33] E. Ozanich, P. Gerstoft, and H. Niu, "A deep network for single-snapshot direction of arrival estimation," in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2019, pp. 1–6.

[34] M. Gall, M. Gardill, T. Horn, and J. Fuchs, "Spectrum-based single-snapshot super-resolution direction-of-arrival estimation using deep learning," in *2020 German Microwave Conference (GeMiC)*. IEEE, 2020, pp. 184–187.

[35] M. Gall, M. Gardill, J. Fuchs, and T. Horn, "Learning representations for neural networks applied to spectrum-based direction-of-arrival estimation for automotive radar," in *2020 IEEE/MTT-S International Microwave Symposium (IMS)*. IEEE, 2020, pp. 1031–1034.

[36] A. M. Ahmed, O. Eissa, and A. Sezgin, "Deep autoencoders for DOA estimation of coherent sources using imperfect antenna array," in *2020 Third International Workshop on Mobile Terahertz Systems (IWMTS)*. IEEE, 2020, pp. 1–5.

[37] G. K. Papageorgiou and M. Sellathurai, "Direction-of-arrival estimation in the low-SNR regime via a denoising autoencoder," in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2020, pp. 1–5.

[38] 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.

[39] 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.

[40] Y. Yao, H. Lei, and W. He, "A-CRNN-based method for coherent DOA estimation with unknown source number," *Sensors*, vol. 20, no. 8, p. 2296, 2020.

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

[42] A. M. Ahmed, A. A. Ahmad, S. Fortunati, A. Sezgin, M. S. Greco, and F. Gini, "A reinforcement learning based approach for multitarget detection in massive MIMO radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 5, pp. 2622–2636, 2021.

[43] D. Luo, Z. Ye, B. Si, and J. Zhu, "Deep MIMO radar target detector in Gaussian clutter," *IET Radar, Sonar & Navigation*, 2022.

[44] W. Guo, T. Qiu, H. Tang, and W. Zhang, "Performance of RBF neural networks for array processing in impulsive noise environment," *Digital Signal Processing*, vol. 18, no. 2, pp. 168–178, 2008.

[45] D. Chen and Y. H. Joo, "A novel approach to 3D-DOA estimation of stationary EM signals using convolutional neural networks," *Sensors*, vol. 20, no. 10, p. 2761, 2020.

[46] ——, "Multisource DOA estimation in impulsive noise environments using convolutional neural networks," *International Journal of Antennas and Propagation*, vol. 2022, 2022.

[47] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1783–1795, 1990.

[48] M. Viberg, P. Stoica, and B. Ottersten, "Maximum likelihood array processing in spatially correlated noise fields using parameterized signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 996–1004, 1997.

[49] S. Feintuch, H. Permuter, I. Bilik, and J. Tabrikian, "Neural network-based multi-target detection within correlated heavy-tailed clutter," *Submitted to IEEE Transactions on Aerospace and Electronic Systems*, 2022.

[50] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[51] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From theory to algorithms*. Cambridge university press, 2014.

[52] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. [Online]. Available: https://doi.org/10.48550/arXiv.1412.6980

[53] K. Harmanci, J. Tabrikian, and J. L. Krolik, "Relationships between adaptive minimum variance beamforming and optimal source localization," *IEEE Transactions on Signal Processing*, vol. 48, no. 1, pp. 1–12, 2000.