IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 November 2022, accepted 19 November 2022, date of publication 24 November 2022,
date of current version 30 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3224608

RESEARCH ARTICLE

Deep Learning-Aided Signal Enumeration for Lens Antenna Array

DAI TRONG HOANG AND KYUNGCHUN LEE, (Senior Member, IEEE)
Research Center for Electrical and Information Technology, Department of Electrical and Information Engineering, Seoul National University of Science and Technology, Seoul 01811, South Korea

Corresponding author: Kyungchun Lee (kclee@seoultech.ac.kr)

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2019R1A6A1A03032119, and in part by the NRF Grant through the Korean Government [Ministry of Science and ICT (MSIT)] under Grant NRF-2022R1A2C1006566.

ABSTRACT This work investigates a data-driven approach to detect the number of incoming signals for a lens antenna array (LAA). First, the energy-focusing property of an electromagnetic (EM) lens is utilized to generate an input spectrum, which can be used to enumerate both the multipath and independent signals. Next, we present the deep learning (DL)-assisted sharp peak recognition method referred to as the power spectrum-based convolutional neural network (PSCNet). Unlike classical techniques, such as constant false alarm rate (CFAR) detection, this data-driven detector can count received signals adaptively based on the LAA power spectrum without requiring any initial configurations. In addition, the PSCNet outperforms other state-of-the-art subspace-based detectors, even under challenging conditions, such as a low signal-to-noise ratio (SNR), a small observation size, and angular ambiguity. For the training phase, we propose a pretrained-model reusing strategy and an input pre-processing approach referred to as the power spectrum shortening (PSS) to alleviate the training burden and achieve lower complexity compared to fully retraining all isolated networks. The simulation results demonstrate that our proposed sharp peak-recognition algorithm not only accomplishes the improved signal enumeration performance but also requires lower computational resources than other subspace-based approaches.

INDEX TERMS Signal enumeration, lens antenna array (LAA), convolutional neural network (CNN), signal power spectrum.

I. INTRODUCTION

The lens antenna array (LAA) is among the most promising technologies to reduce the hardware cost and signal-processing complexity for massive multiple-input multiple-output (MIMO) systems and millimeter wave communications based on the power-focusing property of the electromagnetic (EM) lens [1]. The energy of the received signal is concentrated on a subset of array elements, which can then be distinguished and connected to a limited number of radio frequency (RF) chains. Therefore, it is crucial to identify the beams from the transmitter and accurately estimate the direction of arrival (DoA). Recently, various DoA-estimation strategies have been reported [2], [3]. A three-step algorithm based on a class of multiple signal classification (MUSIC) called the root-MUSIC was proposed to estimate the DoA in a scenario of entirely uncorrelated signals [2]. The study [3] investigated path number detection and simultaneously presented a signal direction estimator considering the power leakage phenomenon of an LAA. However, the coherence of the incoming signals has been overlooked, and significant difficulties occur when they are highly correlated. A preprocessing stage is crucial in handling the coherent received signals because the reflected components are combined into the direct ones, resulting in the rank deficiency of the signal subspace. Spatial smoothing techniques, such as forward/backward spatial smoothing (FBSS) algorithms [4], [5], [6], and the full-row Toeplitz matrices reconstruction

The associate editor coordinating the review of this manuscript and approving it for publication was Yafei Hou.

(FTMR) [7], [8], [9] in the rotational invariance technique (ESPRIT)-like family [10], [11], [12], are effective preprocessing tools to circumvent the problems of coherent signals. However, it is worth emphasizing that these approaches are applied when employing a linear antenna array, whereas no studies on LAA-assisted direction estimation have been conducted. For example, each complex-valued element of the steering matrix of a uniform linear array (ULA) can be expressed as an exponential form. On the other hand, the array response of an antenna in an LAA follows a "sinc" function [1], which cannot form the Vandermonde structures of the modified steering matrices appearing in the closed-form expressions of the FBSS or FTMR's output matrices. As a result, the LAA's smoothed covariance matrices have a full rank, so that the signal and noise eigenvalues cannot be distinguished. In other words, the spatial smoothing methods fail to decorrelate the coherent signals received by an LAA.

Most of the existing DoA-related works [1], [2], [3], [4], [7], [11], [12] rely on the subspace-based array processing algorithms, whose common drawback is that the signal number must be known as a prerequisite. Compared to the subspace-based approaches, the compressed sensing (CS) method [13] solves the issue of coherent DoA estimation without a priori knowledge of the number of signals. This method can be used to reconstruct the spatial spectrum based on the sparsity of the non-zero entries in the angle space of the array signals. In the CS sparse signal recovery variants, the orthogonal matching pursuit (OMP) has a noticeably lower computational complexity and a higher recovery speed [14]; thus, it is an appropriate choice for both signal enumeration and direction approximation [15], [16]. However, the OMP is susceptible to noisy signals and multipath distortion. As a result, combining the CS-assisted estimation with a sharp peak detector based on the recovered spectrum may result in an enormous detection error probability of signal number detection.

The signal-enumeration problem has been investigated for several signal-processing and wireless communications applications. The most commonly used techniques are the gap measurements of two consecutive eigenvalues, such as the Akaike information criterion (AIC) [17] and minimum description length (MDL) [18]. In addition, the second-order statistic of the eigenvalues (SORTE) [19] and eigenvalue ratio (ER) [20] are employed to strengthen the detection capability of conventional methods. Furthermore, two deep learning (DL)-aided detectors called the eigenvalue-based classification network (ECNet) and regression network (ERNet) [21] have been introduced to overcome the severe performance degradation in some challenging situations, including a low signal-to-noise ratio (SNR) and limited number of snapshots. However, the difference in the LAA's physical structure from that of the classical ULA hinders the implementation of these algorithms and other supporting dimensionality-reduction methods [8], [9] in a coherent scenario. Moreover, the eigenvalue decomposition (EVD) with a complex-valued matrix creates a substantial computational burden, which generally contributes about half of the total complexity of these subspace-based algorithms. To alleviate the computational demands of the aforementioned data-preprocessing step, the signal power distribution across the LAA elements can be utilized as an alternative solution for the EVD. This feature can then be combined with a simple sharp peak recognition mechanism like the constant false alarm rate (CFAR) scheme [22] to count the number of incoming signals. However, the CFAR detector needs to be configured with several parameters, such as the number of training cells, number of guard cells, and false alarm probability for operation. It is noted that different values of configuration parameters affect the performance of the CFAR detector. Last but not least, the angle ambiguity affects the signal enumeration performance of the spectrum-assisted and subspace-fitting algorithms. Precisely speaking, the peaks representing neighboring antennas in an EF spectrum may be merged into one, which hinders the prediction of a peak-finding algorithm, such as CFAR. In addition, angle ambiguity results in the degeneration of the steering vectors, each of which is approximated as a linear combination of the others [23]. This undesirable phenomenon causes the failure of the eigenvalue-based schemes. These problems motivate the proposal of an efficient signal enumeration algorithm for the LAA in this work. The contributions of this work are as follows:

- The power spectrum-based convolutional neural network (PSCNet), which utilizes the energy-focusing (EF) property of an LAA, is proposed to obtain high signal-enumeration accuracy without assigning values to any configuration parameters. The simulation results show that the PSCNet not only outperforms the eigenvalue gap-based detectors and CFAR, even in harsh conditions, such as a highly contaminated environment or a limited sample size, but also improves the enumeration capability in the case of manifold ambiguity.
- The input power spectrum is analyzed to show its adaptation to both the noncoherent and coherent scenarios. In addition, it has been mathematically proven that the proposed PSCNet has lower computational complexity than the prior state-of-the-art subspace-based algorithms, and OMP-assited path number detector.
- The power spectrum shortening (PSS) technique, which eliminates some abundant elements in the original spectrum, is employed as an input conversion method so that a transferred model can be reused to efficiently predict the number of signals. Additionally, a model reusing strategy is proposed to save on training time and computational resources as well as achieve better performance compared to the traditional single-task learning model.

Notations: In this paper, lowercase, boldface lowercase, and boldface capital letters denote scalars, vectors, and matrices, respectively. The superscripts $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ denote the conjugate, transpose, and conjugate transpose of a complex vector or matrix, respectively. The (m, n)th (or mth) element of a matrix $\boldsymbol{A}$ (or vector $\boldsymbol{a}$) is denoted by $a_{m,n}$ (or a_m). The expression $\|\boldsymbol{A}\|_0$ ($\|\boldsymbol{a}\|_0$) indicates the ℓ_0-norm that is the

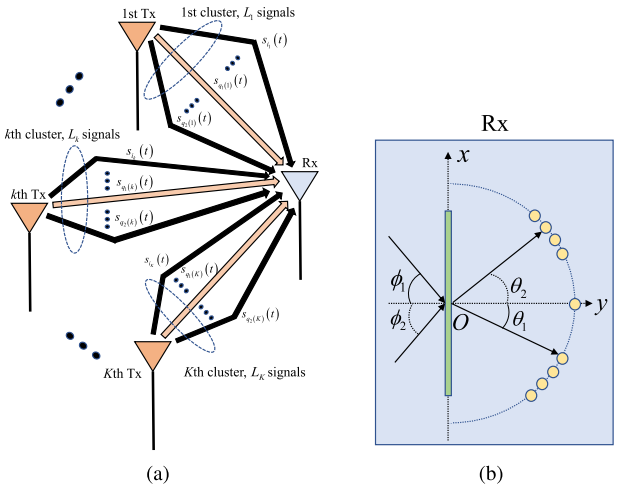

FIGURE 1. Descriptions of (a) system model and (b) typical symmetric LAA.

number of non-zero elements in matrix $\boldsymbol{A}$ (vector $\boldsymbol{a}$). The notation $E\{\cdot\}$ represents the expectation of a random variable. Finally, the operator $(*)$ denotes convolution.

II. SYSTEM MODEL

As shown in Fig. 1(a), there are P far-field signals from K transmit (Tx) sources impinging upon a receive (Rx) LAA. For the kth cluster, let L_k be the number of signals, and let $q_1(k) = \sum_{l=1}^{k} L_{l-1} + 1$, and $q_2(k) = \sum_{l=1}^{k} L_l$ be indices of the first and last signals, respectively. It is obvious that $P = \sum_{k=1}^{K} L_k$ and $L_0 = 0$. In some applications such as airborne radar [24] and global navigation satellite systems [25], multipath echo is modeled as multiple incident signals that are coherent with the direct one. Hence, we assume that $\beta_i = \vartheta_i e^{j\varphi_i}$ represents the complex attenuation coefficient of $s_i(t)$, which is the replica of the first signal in the kth cluster, $s_{q_1(k)}(t)$, whereas ϑ_i and φ_i denote the fading amplitude and phase difference of $s_i(t)$ and $s_{q_1(k)}(t)$, respectively.[1] Then the ith coherent signal can be written as

$$s_i(t) = \beta_i s_{q_1(k)}(t), \quad i \in [q_1(k), q_2(k)]. \tag{1}$$

Fig. 1(b) illustrates an LAA system with a $D_y \times D_z$ EM lens, which is placed on the y–z plane. The thickness of this lens is negligible compared with the focal length F, and its center is at the origin. Assume P far-field narrowband signals impinge upon an LAA composed of $M_0 = 2M + 1$ identical omnidirectional sensors from distinct directions ϕ_i, where $i = 1, 2, \ldots, P$. Let the central element of the LAA be the reference. Then, the array elements are placed based on the angular resolution of the lens as follows [1]:

$$\sin\theta_m = \frac{m\lambda}{D_y}, \tag{2}$$

where m represents the index of the sensor such that $m \in [-M, M]$, λ is the carrier wavelength, and θ_m denotes the angle of sensor m. According to [1], the array response of antenna m to a signal from direction ϕ is given as

$$a_m(\phi) = \sqrt{\alpha}\,\mathrm{sinc}\left(m - \frac{D_y}{\lambda}\sin\phi\right), \tag{3}$$

where $\alpha = D_y D_z / \lambda^2$ is the effective aperture, and $\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$.

The received signal $x_m(t)$ at the mth sensor at time t over T uniquely spaced time snapshots can be expressed as

$$\begin{aligned} x_m(t) &= \sum_{i=1}^{P} s_i(t)\, a_m(\phi_i) + n_m(t) \\ &= \sum_{k=1}^{K} s_{q_1(k)}(t) \sum_{i=q_1(k)}^{q_2(k)} \beta_i a_m(\phi_i) + n_m(t), \end{aligned} \tag{4}$$

where $n_m(t)$ represents the zero-mean Gaussian noise at the mth sensor. The noise is independent and identically distributed (i.i.d) with a variance σ_n^2. It is assumed that the signals and noise are mutually independent. The entire received signal in (4) can be rewritten in vector notation as

$$\begin{aligned} \boldsymbol{x}(t) &= [x_{-M}(t), \ldots, x_0(t), \ldots, x_M(t)]^T \\ &= \boldsymbol{A}(\boldsymbol{\Phi})\,\boldsymbol{s}(t) + \boldsymbol{n}(t), \end{aligned} \tag{5}$$

where $\boldsymbol{s}(t) = [s_1(t), \ldots, s_P(t)]^T$ denotes the source signal, $\boldsymbol{\Phi} = \{\phi_i\}_{i=1}^{P}$ contains all signal directions, $\boldsymbol{A}(\boldsymbol{\Phi}) = [\boldsymbol{a}(\phi_1), \boldsymbol{a}(\phi_2), \ldots, \boldsymbol{a}(\phi_P)]$ is the steering matrix with each column vector $\boldsymbol{a}(\phi_i) = [a_{-M}(\phi_i), \ldots, a_0(\phi_i), \ldots, a_M(\phi_i)]^T$, and $\boldsymbol{n}(t) = [n_{-M}(t), \ldots, n_0(t), \ldots, n_M(t)]^T$ denotes the white Gaussian noise vector.

III. SIGNAL-ENUMERATION ALGORITHM

A. INPUT FEATURE SELECTION

From the received signal $\boldsymbol{x}(t)$, we define $\boldsymbol{r} = [r_{-M}, \ldots, r_0, \ldots, r_M] \in \mathbb{R}^{M_0}$ as the vector of signal power distribution. Then, the mth element of $\boldsymbol{r}$ can be expressed as

$$\begin{aligned} r_m &= E\left\{x_m(t)\, x_m^*(t)\right\} \\ &= E\left\{\left[\sum_{k=1}^{K} s_{q_1(k)}(t) \sum_{i=q_1(k)}^{q_2(k)} \beta_i a_m(\phi_i) + n_m(t)\right] \times \left[\sum_{k=1}^{K} s_{q_1(k)}^*(t) \sum_{l=q_1(k)}^{q_2(k)} \beta_l^* a_m(\phi_l) + n_m^*(t)\right]\right\} \\ &= \sum_{k=1}^{K} P_k \sum_{i=q_1(k)}^{q_2(k)} \beta_i a_m(\phi_i) \sum_{l=q_1(k)}^{q_2(k)} \beta_l^* a_m(\phi_l) + \sigma^2 \\ &= \sum_{i=1}^{P} a_m(\phi_i)\, d_{i,m} + \sigma^2, \end{aligned} \tag{6}$$

[1] Without loss of generality, we assume that $\vartheta_{q_1(k)} = 1$, and $\varphi_{q_1(k)} = 0$. For noncoherent sources, only one incoming signal appears in each cluster. In other words, $L_k = 1$ for $k \in [1, K]$ and $\beta_i = 1$ for $i \in [1, P]$.

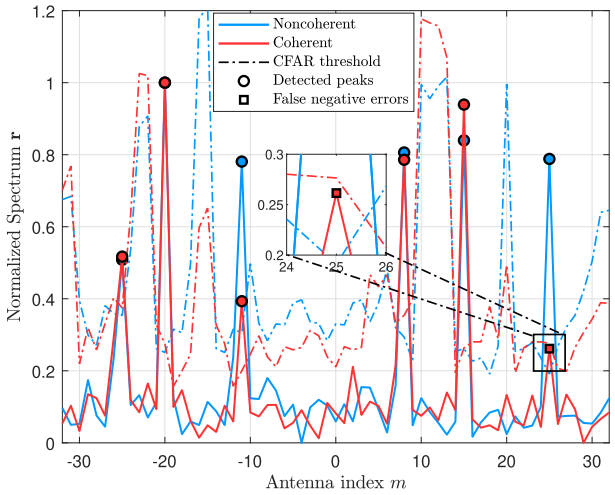

FIGURE 2. Signal power distributed across LAA elements for SNR = 0 dB, $P = 6$, and $M_0 = 65$ in both noncoherent and coherent environments.

where $P_k = E\left\{\left|s_{q_1(k)}(t)\right|^2\right\}$ denotes the signal power of the direct component in the kth cluster, and the pseudosignal $d_{i,m}$ is given as

$$d_{i,m} = P_k \beta_i \sum_{l=q_1(k)}^{q_2(k)} \beta_l^* a_m(\phi_l), \quad k \in [1, K], \quad i \in [q_1(k), q_2(k)]. \tag{7}$$

From (3) and (7), it is noted that any value of m and ϕ_i satisfying $\left|m - \frac{D_y}{\lambda} \sin\phi_i\right| \approx 0$ yields $a_m(\phi_i) \approx \sqrt{\alpha}$ and $d_{i,m} \approx P_k |\beta_i|^2 \sqrt{\alpha}$. In other words, a signal arriving from a particular direction ϕ_i is focused by the EM lens on certain antenna elements, whose energy is $r_m \approx \alpha P_k |\beta_i|^2 + \sigma^2$, whereas those of the unexcited ones are approximately zero. Fig. 2 visualizes the power distribution of an $M_0 = 65$-element LAA system in the case of $P = 6$ signals from directions $\boldsymbol{\Phi} = \{-58°, -42°, -22°, 15°, 30°, 56°\}$. From the figure, the normalized spectrum $\boldsymbol{r}$ provides sharp peaks representing the indices of the antenna elements capturing the maximum energy of the incoming signals. The conspicuous peak recognition based on spectrum $\boldsymbol{r}$ does not rely on any subspace-based signal analysis; therefore, it can be utilized to predict the source number in both coherent and noncoherent environments. In addition, a classical sharp peak detector, the CFAR scheme [22], can be applied in this situation. However, the peak representing the multipath signal is possibly ignored by the CFAR detector owing to its power degradation, as illustrated by the square marker in Fig. 2.

As mentioned in Section I, the OMP can be employed to recover a sparse signal and subsequently reconstruct the spatial spectrum. In the OMP algorithm, the number of grid points $\widetilde{P}$ can be computed as $\widetilde{P} = (\phi_{\max} - \phi_{\min})/\rho$, where ρ denotes the resolution of DoA in the range of $[\phi_{\min}, \phi_{\max})$. Fig. 3 depicts the spectrum constructed by the OMP algorithm with different values of ρ in the case of $P = 6$ signals from the same aforementioned directions. We note that the signal with

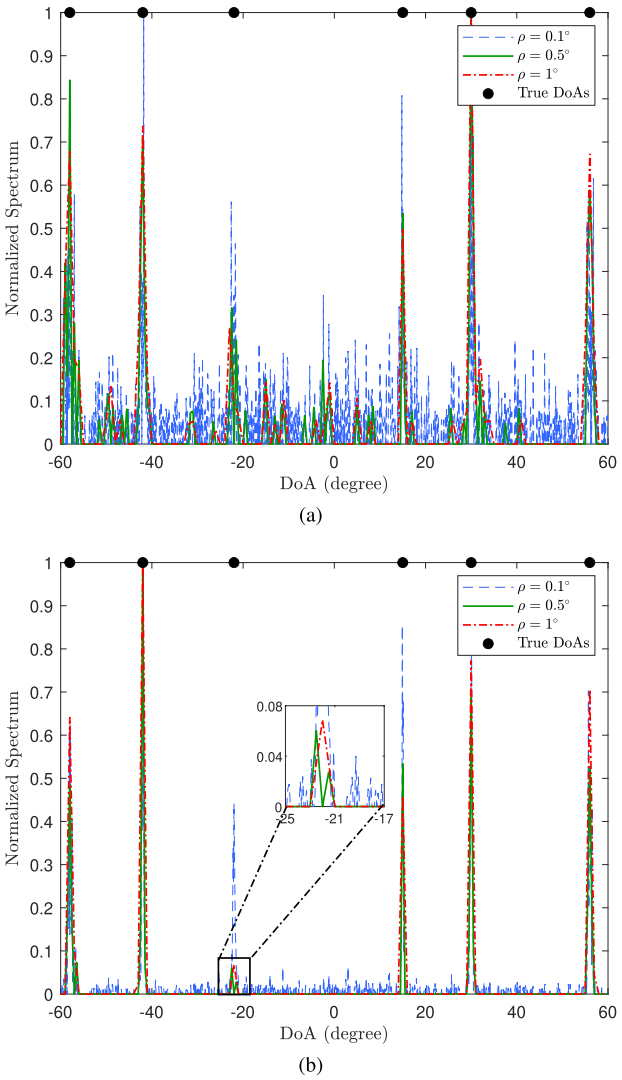

FIGURE 3. Spatial spectrum recovered by the OMP algorithm for (a) SNR = 0 dB, and (b) SNR = 10 dB.

an arrival angle of $\phi = -22°$ is a multipath signal. As shown in Fig. 3(a), the recovered spectra contain the noise elements in all direction samples. Therefore, the features obtained by the OMP-based recovery cannot be used for DoA estimation and signal enumeration in a highly contaminated scenario. In a high-SNR region, although the normalized spectrum provided by the OMP has conspicuous peaks that can be utilized for a subsequent peak recognition process, it is still a saw-tooth waveform, as shown in Fig. 3(b). Furthermore, the peaks representing the direction $\phi = -22°$ in the cases of $\rho = 0.5°$ and $\rho = 1°$ are not clearly visible due to the power degradation of multipath propagation, and are likely to be considered as the noise elements. Thus, the OMP-based spatial spectrum reconstruction is not an efficient feature engineering process for signal enumeration.

B. PROPOSED SIGNAL ENUMERATOR

As can be seen in Fig. 4, the structure of our proposed PSCNet comprises L_C layers of one-dimensional (1D) convolutions,

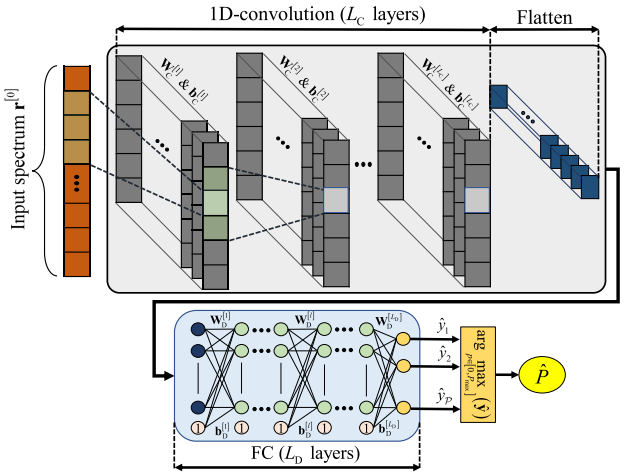

FIGURE 4. Description of the proposed PSCNet.

a flattening layer, and L_{D} fully connected (FC) layers. Let $\boldsymbol{r}^{[0]}$ and $\overline{M_0}$ be the input array of PSCNet and its number of elements, respectively. The estimated output is obtained via consecutive nonlinear transformations of $\boldsymbol{r}^{[0]}$ as

$$\widehat{\boldsymbol{y}} = \mathcal{T}^{[L_{\mathrm{C}}+L_{\mathrm{D}}]}\left(\ldots\left(\mathcal{T}^{[1]}\left(\boldsymbol{r}^{[0]};\boldsymbol{\eta}^{[1]}\right);\ldots\right),\boldsymbol{\eta}^{[L_{\mathrm{C}}+L_{\mathrm{D}}]}\right), \tag{8}$$

where $\mathcal{T}^{[l]}(\cdot)$, $\boldsymbol{\eta}^{[l]}$, and $\boldsymbol{r}^{[l]}$ are the transformation function, learnable parameters, and output in the lth layer of the model, respectively. In the lth convolutional layer, we assume that the number of feature maps is $\mathcal{C}$, whereas a bank of filters $\boldsymbol{W}_{\mathrm{C}}^{[l]} = \left\{\boldsymbol{W}_{\mathrm{C}}^{[l,c]}\right\}_{c=1}^{\mathcal{C}}$, the element of which is $\boldsymbol{W}_{\mathrm{C}}^{[l,c]} = \left\{\boldsymbol{W}_{\mathrm{C}}^{[l,c,c']}\right\}_{c'=1}^{\mathcal{C}}$, and bias vectors $\boldsymbol{b}_{\mathrm{C}}^{[l]} = \left\{\boldsymbol{b}_{\mathrm{C}}^{[l,c]}\right\}_{c=1}^{\mathcal{C}}$ are employed. Then, the cth channel of this convolution result is calculated by

$$\begin{aligned}&\mathcal{T}_{\mathrm{C}}^{[l,c]}\left(\boldsymbol{r}^{[l-1]};\boldsymbol{\eta}_{\mathrm{C}}^{[l]}\right)\\&\quad=\mathcal{T}^{[l]}\left(\boldsymbol{r}^{[l-1]};\boldsymbol{\eta}^{[l]}\right)\\&\quad=\boldsymbol{\psi}^{[l]}\left(\sum_{c'=1}^{\mathcal{C}}\boldsymbol{W}_{\mathrm{C}}^{[l,c,c']}*\boldsymbol{r}^{[l-1,c']}+\boldsymbol{b}_{\mathrm{C}}^{[l,c]}\right),\\&\qquad l\in[1,L_{\mathrm{C}}],\quad c\in[1,\mathcal{C}],\end{aligned} \tag{9}$$

where $\boldsymbol{W}_{\mathrm{C}}^{[l,c,c']} \in \mathbb{R}^{h\times 1}$ denotes the cth weighting kernel sliding along the c'th component of previous feature map $\boldsymbol{r}^{[l-1,c']} \in \mathbb{R}^{\overline{M_0}\times 1}$, and $\boldsymbol{b}_{\mathrm{C}}^{[l,c]} \in \overline{M_0}\times 1$ is the cth bias element of $\boldsymbol{b}_{\mathrm{C}}^{[l]}$. In addition, $\boldsymbol{\eta}_{\mathrm{C}} = \left\{\boldsymbol{\eta}^{[l]}\right\}_{l=1}^{L_{\mathrm{C}}} = \left\{\boldsymbol{W}_{\mathrm{C}}^{[l]},\boldsymbol{b}_{\mathrm{C}}^{[l]}\right\}_{l=1}^{L_{\mathrm{C}}}$ contains the trained parameters, and $\boldsymbol{\psi}^{[l]}(\cdot)$ represents an activation function at the output of layer l, which is determined by a scaled exponential linear unit (SELU) because of its self-normalizing property and faster training convergence than other commonly used functions [26]. Specifically, the SELU function is given by

$$\begin{aligned}\boldsymbol{\psi}^{[l]}(z)=\mathrm{SELU}(z)=\zeta_1\cdot\begin{cases}z, & \text{if } z>0,\\ \zeta_2\left(e^z-1\right), & \text{if } z\le 0,,\end{cases}\\ l\in[1,L_{\mathrm{C}}+L_{\mathrm{D}}-1],\end{aligned} \tag{10}$$

where ζ_1 and ζ_2 are two fixed parameters. Furthermore, we utilize a padding method and single-stride convolutions to avoid rapid downsampling on the feature spatial sizes. After feature extraction, the data is flattened for conversion into a 1D vector, $\widetilde{\boldsymbol{r}} \in \mathbb{R}^{\mathcal{C}h}$, to feed the subsequent FC layers.

The FC layers include an input layer, $(L_{\mathrm{D}}-2)$ hidden layers, and a single node in the output layer. Let Λ_l be the number of neurons in the lth layer of the FC network. $\boldsymbol{\eta}_{\mathrm{D}} = \left\{\boldsymbol{\eta}^{[l]}\right\}_{l=L_{\mathrm{C}}+1}^{L_{\mathrm{C}}+L_{\mathrm{D}}} = \left\{\boldsymbol{W}_{\mathrm{D}}^{[l]},\boldsymbol{b}_{\mathrm{D}}^{[l]}\right\}_{l=1}^{L_{\mathrm{D}}}$ contains the weight matrix $\boldsymbol{W}_{\mathrm{D}}^{[l]} \in \mathbb{R}^{\Lambda_l\times\Lambda_{l-1}}$ and bias vector $\boldsymbol{b}_{\mathrm{D}}^{[l]} \in \mathbb{R}^{\Lambda_l\times 1}$ between the $(l-1)$th and lth layers, whereas $\widetilde{\boldsymbol{r}}^{[l]} \in \mathbb{R}^{\Lambda_l\times 1}$ represents the output of the lth layer. Then, the transformation function $\mathcal{T}_{\mathrm{D}}^{[l]}(\cdot)$ utilized in the FC layers is defined as

$$\begin{aligned}\mathcal{T}_{\mathrm{D}}^{[l]}\left(\widetilde{\boldsymbol{r}}^{[l-1]};\boldsymbol{\eta}_{\mathrm{D}}^{[l]}\right)&=\mathcal{T}^{[l+L_{\mathrm{C}}]}\left(\boldsymbol{r}^{[l+L_{\mathrm{C}}-1]};\boldsymbol{\eta}^{[l+L_{\mathrm{C}}]}\right)\\&=\boldsymbol{\psi}^{[l+L_{\mathrm{C}}]}\left(\boldsymbol{W}_{\mathrm{D}}^{[l]}\widetilde{\boldsymbol{r}}^{[l-1]}+\boldsymbol{b}_{\mathrm{D}}^{[l]}\right),\\&\qquad l\in[1,L_{\mathrm{D}}].\end{aligned} \tag{11}$$

We assume that the maximum number of signals is $\mathcal{P}$ and note that the proposed convolutional neural network (CNN) is designed to handle multiclass classification problems. Therefore, the softmax activation function is employed to map $\mathcal{P}$ outputs $\widetilde{y}_i$ in the output layer to $\mathcal{P}$ probabilities $\widehat{y}_i$, which is calculated as

$$\widehat{y}_i=\boldsymbol{\psi}^{[L_{\mathrm{C}}+L_{\mathrm{D}}]}\left(\widetilde{y}_i\right)=\frac{e^{\widetilde{y}_i}}{\sum_{i=1}^{\mathcal{P}}e^{\widetilde{y}_i}},\qquad i\in[1,\mathcal{P}]. \tag{12}$$

For the training session, the adaptive moment estimation (Adam) [27] optimizer is employed to minimize the categorical cross-entropy loss function $\mathcal{L}(\cdot)$ given by

$$\mathcal{L}(\boldsymbol{\eta})=-\frac{1}{\mathcal{B}}\sum_{b=1}^{\mathcal{B}}\sum_{i=1}^{\mathcal{P}}y_i(\boldsymbol{b})\log\left[\widehat{y}_i(\boldsymbol{b})\right], \tag{13}$$

where $\mathcal{B}$ denotes the number of samples in each batch $\boldsymbol{\Omega} = \left\{\boldsymbol{r}^{[0]}(\boldsymbol{b}),\boldsymbol{y}(\boldsymbol{b})\right\}_{\boldsymbol{b}=1}^{\mathcal{B}}$, which is stochastically selected from the training set. In addition, $\boldsymbol{r}(\boldsymbol{b})$ and $\boldsymbol{y}(\boldsymbol{b})$ are the input spectrum and one-hot output vector at the bth data point, respectively.

C. TRAINING STRATEGIES

First, the elements of the input array $\boldsymbol{r}^{[0]}$ feeding the PSCNet has to be considered deliberately. It is worth emphasizing that a peak in $\boldsymbol{r}$ is only generated if there is a signal direction ϕ satisfying $\frac{D_y}{\lambda}\sin(\phi)\approx m$, for $m\in[-M,M]$. This yields a constraint that M must be large enough to cover all values of $\frac{D_y}{\lambda}\sin(\phi)$. Thus, to accurately detect the signal from any direction, the array must satisfy

$$M\ge\left\lceil\frac{D_y}{\lambda}\right\rceil. \tag{14}$$

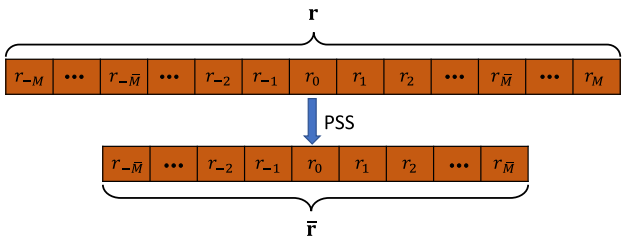

FIGURE 5. Description of the PSS technique.

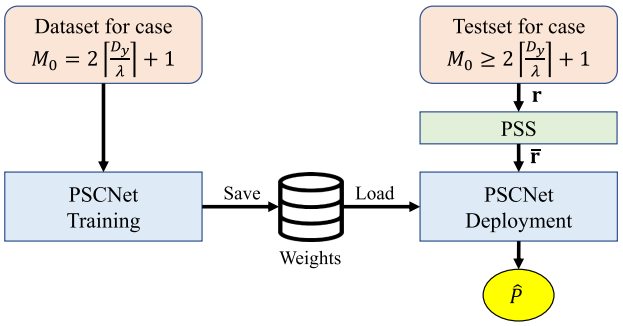

FIGURE 6. Pretrained-model reusing strategy in the scenario of $M \geq D_y/\lambda$.

We also note that the data points for the offline training phase are generated without any assumptions about the angle difference χ of the two received signals. Therefore, our proposed PSCNet can learn to determine the signal number based on the width of the peaks, which are possibly formed by two closely adjacent directions.

It is also noted that the peaks representing the energy of signals only appear in the antenna indices $m \in \left[-\left\lceil \frac{D_y}{\lambda} \right\rceil, \left\lceil \frac{D_y}{\lambda} \right\rceil\right]$. However, the spectrum outside this range corresponds to the noise power, which is the abundant information for counting incoming signals. From this perspective, we propose a feature engineering method referred to as the power spectrum shortening (PSS) that saves computational resources for the latter forward propagation of PSCNet. As shown in Fig. 5, this approach cuts off some components from two heads of the spectrum $\boldsymbol{r}$, then the output of the PSS method is $\bar{\boldsymbol{r}} = \left[\boldsymbol{r}_{-\overline{M}}, \ldots, r_0, \ldots, r_{\overline{M}}\right]$, which has half of the $\bar{\boldsymbol{r}}$'s size of $\overline{M}$. From (14), the value of $\overline{M}$ can be calculated as

$$\overline{M} = \min\left\{M, \left\lceil \frac{D_y}{\lambda} \right\rceil\right\}. \tag{15}$$

In a specific scenario of $\left\{T, \text{SNR}, D_y/\lambda\right\}$, the spectrums formed by the $\overline{M}_0$ central elements of $\boldsymbol{r}$ are similar, regardless of the number of antennas $M_0 \geq \overline{M}_0$. Thus, we propose the application of a pretrained-model of an $\overline{M}_0$-antenna system to enumerate signals with an LAA with M_0 sensors. With the help of the PSS technique, this strategy can overcome the limitations of time and computational resources of independent training for an M_0-antenna LAA. As shown in Fig. 6, the PSS technique is included in this strategy to pre-process raw data.

Algorithm 1 Proposed PSCNet-Based Signal Enumeration

Input: $\{\boldsymbol{x}(t)\}_{t=1}^{T}$, M_0, T, SNR, and D_y/λ.
Output: $\widehat{P}$.

1: Initialize $\bar{\boldsymbol{r}}$ as an empty array.
2: Compute $\overline{M} = \min\left\{\left\lfloor \frac{M_0}{2} \right\rfloor, \left\lceil \frac{D_y}{\lambda} \right\rceil\right\}$.
3: **for** $m = -\overline{M}$ to $\overline{M}$ **do**
4: Compute $r_m = \frac{1}{T}\sum_{t=1}^{T} x_m(t)\, x_m^*(t)$.
5: Append r_m to $\bar{\boldsymbol{r}}$.
6: **end for**
7: Load the trained parameters $\boldsymbol{\eta}$ for the PSCNet based on $\left\{\overline{M}_0, T, \text{SNR}, D_y/\lambda\right\}$, where $\overline{M}_0 = 2\overline{M} + 1$.
8: Obtain $\widehat{\boldsymbol{y}}$ using (8).
9: Predict $\widehat{P} = \arg\max_{i \in [1,\mathcal{P}]} (\widehat{\boldsymbol{y}})$.

Subsequently, the transferred model can utilize this desired information to predict the number of incoming signals.

In the training phase, the noise power can be randomly generated, thus, the model has to adapt to a wide SNR range of the environment. This means that PSCNet needs an enormous number of hidden layers and learning parameters, which requires significantly high complexity and execution time in both the training and testing phases. To efficiently train PSCNet, we consider dividing the entire SNR range into smaller ones $\left[\text{SNR} - \Delta/2, \text{SNR} + \Delta/2\right]$, and independently training multiple networks in those intervals. Here, SNR is the medium value of the SNR range for which a CNN model is trained, and $\Delta \geq 0$ represents the SNR interval. For example, SNR $= 0$ dB and $\Delta = 4$ dB indicate that the SNR values of the incoming signals are stochastically chosen in the range of $[-2, 2]$ dB. Although this strategy requires additional memory to store multiple sets of trained parameters, we can reduce the training convergence time and achieve a higher prediction accuracy. Notably, the proposed model should also be trained with different numbers of snapshots, and ratios of lens' size to the carrier wave length, which significantly influence the formation of the power spectrum. Consequently, we consider training PSCNet for consistent situations regarding the SNR value, the number of snapshots, and the D_y–λ ratio, i.e., SNR, T, and D_y/λ.

D. PRACTICAL DEPLOYMENT AND COMPUTATIONAL COMPLEXITY ANALYSIS

Algorithm 1 summarizes the practical deployment of the PSCNet. Step 2 involves computing half of the input size $\overline{M}$ based on the LAA's parameters. Subsequently, steps 3–6 are iterated to construct the power spectrum $\bar{\boldsymbol{r}}$. The signal energy is computed practically as $r_m = \frac{1}{T}\sum_{t=1}^{T} x_m(t)\, x_m^*(t)$ instead of using (6). In step 7, the pretrained parameters are loaded into the PSCNet based on information on the environment's SNR, the number of snapshots, T, the array configuration parameters, $\overline{M}_0$ and D_y/λ. Steps 8 and 9 indicate the feedforward

propagation of the PSCNet to predict the one-hot vector, $\widehat{\boldsymbol{y}}$, and number of signals, $\widehat{P}$, respectively.

The computational complexity in our study is measured by counting the number of floating-point operations (FLOPs) [28]. In Algorithm 1, the complexity mainly relies on the iterations of steps 3–6 and step 8. Because the computation of r_m yields $(2T-1)$ operations, the total number of operations in steps 3–6 can be expressed as

$$\mathcal{O}_{\text{Steps } 3\text{-}6} = (2T-1)\,\overline{M_0}. \tag{16}$$

According to (8), the computational cost of the feedforward propagation of PSCNet comes from the successive convolutions and matrix multiplication in the FC layers. In the first layer, because the spectrum $\boldsymbol{r}$ is fed as a single channel, the multiplications and bias additions of the first convolution cost $2\overline{M_0}h\mathcal{C}$ FLOPs. For the subsequent convolutions, the padding technique keeps the output size in one channel unchanged; therefore, we have $\boldsymbol{r}^{[l]} \in \mathbb{R}^{\overline{M_0}\times 1\times \mathcal{C}}$. According to (9), the complexity of the lth convolution with $l \in [2, L_{\text{C}}]$ is $2\overline{M_0}h\mathcal{C}^2$. Regarding the propagation in the lth FC layer, the computational load includes $\Lambda_l\,(2\Lambda_{l-1}-1)$ multiplications and additions between the current weight matrix and the output of the previous layer and Λ_l additions with the bias vector. Then, the number of FLOPs in the lth FC layer is $2\Lambda_{l-1}\Lambda_l$. As a result, the total complexity of step 8 is given by

$$\mathcal{O}_{\text{Step } 8} = 2\overline{M_0}h\mathcal{C}\,(L_{\text{C}}\mathcal{C}+1) + 2\sum_{l=1}^{L_{\text{D}}}\Lambda_{l-1}\Lambda_l. \tag{17}$$

IV. SIMULATION RESULTS AND ANALYSIS

In our simulations,[2] an LAA system with half-wavelength element spacing is used to receive a maximum of $\mathcal{P}=8$ far-field signals coming from directions that are stochastically chosen in a range of $\phi \in [-60°, 60°)$. The received signals are assumed to be captured over a finite number of spaced time snapshots $T \in [10, 100]$. We also examine the performance of antenna arrays with the various numbers of sensing elements M_0 and different D_y–λ ratios, i.e., $M_0 \in [53, 117]$ and $D_y/\lambda \in [10, 90]$. Unless otherwise stated, we assume that $M_0 = 65$ antennas, $T = 40$ snapshots, SNR $= 0$ dB, and $D_y/\lambda = 30$. Without loss of generality, we assume that the normalized aperture $\alpha = 1$, and a carrier frequency $f_c = 1$ GHz is used to transmit the signals. In addition, the number of incoming signals P and the number of clusters K follow the i.i.d. uniform distribution within $[1, \mathcal{P}]$ and $[1, P]$, respectively. The attenuation coefficient β_i in the kth cluster is generated as $\beta_i \sim \mathcal{CN}(0, 0.2)$, for $i \in (q_1(k), q_2(k)]$. For the convolutions of PSCNet, the parameters of feature extracting kernels are set as $\mathcal{C}=8$ and $h=3$. The dataset is created for both the training and testing phases by subsequently following steps 1–6 of Algorithm 1. PSCNet should be trained for consistent cases regarding the SNR value, number of snapshots, and D_y–λ ratio, whereas the related parameters can be randomly chosen within their ranges. Moreover, the output $\boldsymbol{y} \in \mathbb{N}^{\mathcal{P}}$ of the deep network is a one-hot vector, whose pth element is determined as $\boldsymbol{y}_p = \Upsilon_P(p)$, where $\Upsilon_P(i)$ can be defined as:

$$\Upsilon_P(p) = \begin{cases} 1 & \text{if} \quad p = P \\ 0 & \text{if} \quad p \neq P \end{cases} \tag{18}$$

For a specific situation of $\left(\text{SNR}, T, D_y/\lambda\right)$, a collection of $\mathcal{B}_0 = 5\times 10^4$ data-label pairs $(\bar{\boldsymbol{r}}, \boldsymbol{y})$ is generated. In addition, $\mathcal{B}_1 = 80\%\mathcal{B}_0$ data points are chosen for training, whereas the remaining $\mathcal{B}_2 = 20\%\mathcal{B}_0$ samples are used for testing. Finally, the detection error probability (DEP) metric with $\mathcal{B}_2$ trials is employed to evaluate the signal enumeration performance of the examined detectors, which is expressed as

$$\text{DEP} = \frac{1}{2\mathcal{B}_2}\sum_{\boldsymbol{b}=1}^{\mathcal{B}_2}\|\boldsymbol{y}(\boldsymbol{b}) - \widetilde{\boldsymbol{y}}(\boldsymbol{b})\|_0, \tag{19}$$

where $\widetilde{\boldsymbol{y}}$ denotes the categorical encoded vector, whose pth element is computed as $\widetilde{\boldsymbol{y}}_p = \Upsilon_{\widehat{P}}(p)$. It is noted that the term $\|\boldsymbol{y}(\boldsymbol{b}) - \widetilde{\boldsymbol{y}}(\boldsymbol{b})\|_0$ in (19) is equal to 2 when the $\boldsymbol{b}$th sample is misclassified; thus, the sum of these terms divided by 2 becomes the number of misdetected cases.

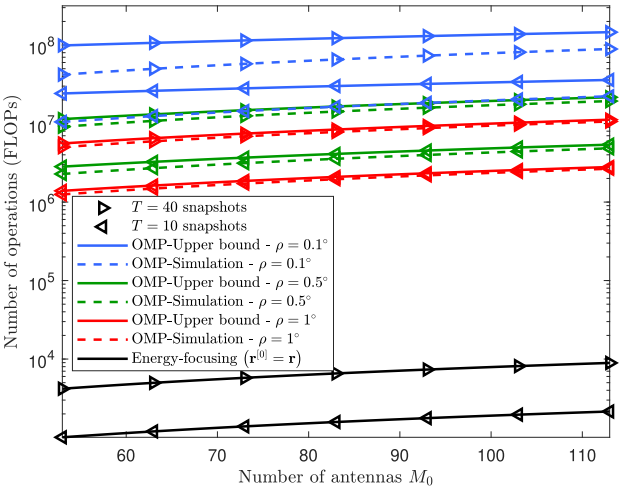

FIGURE 7. Comparison of computational complexity between OMP-based and EF-based spectrum reconstruction for SNR = 10 dB.

Fig. 7 illustrates the complexity versus M_0 of the EF property-based spectrum formation and the OMP algorithm. We note that an error-tolerant factor, i.e., $\varepsilon = 10^{-6}$, is used for early termination of signal recovery in one snapshot, then the number of FLOPs of the OMP, i.e., $\mathcal{O}_{\text{OMP}}$, can be obtained by the Monte Carlo simulations. In case that there is no termination condition, a signal is recovered after $\mathcal{P}$ iterations, and the upper bound of OMP complexity, i.e., $\mathcal{O}_{\text{OMP,max}}$, can be computed. As shown in Fig. 7, the computational load of the EF method is substantially lower than those of the OMP scheme, regardless of the values of ρ. It is noted

[2]For more details, please refer to our simulation code in https://github.com/daihoang25/LAA_SignalEnum.git.

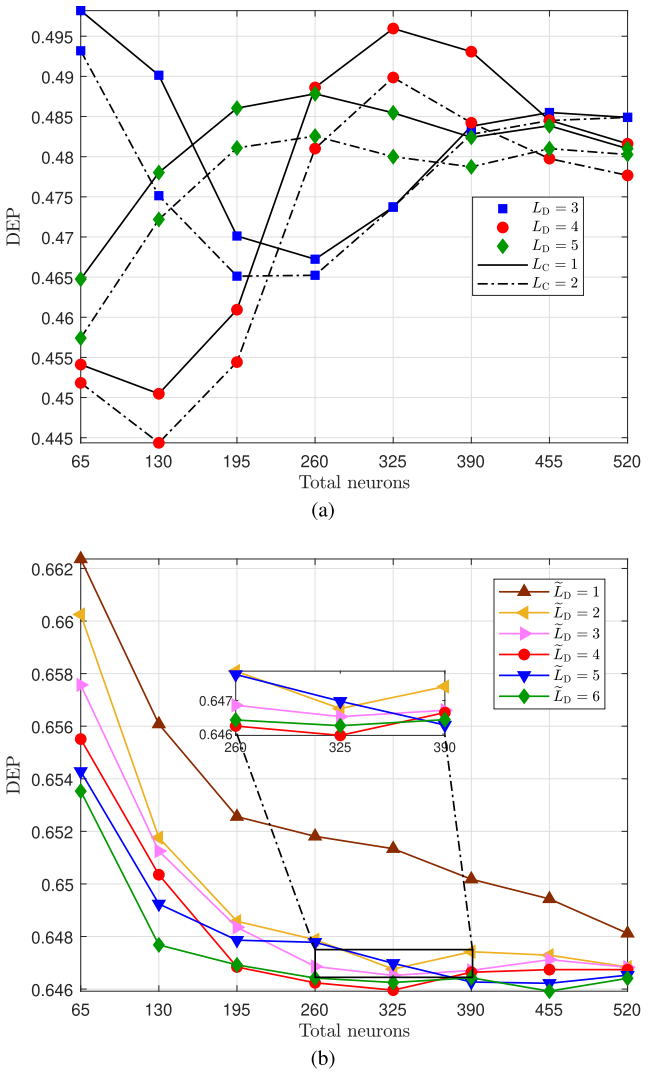

FIGURE 8. DEP validation of (a) PSCNet and (b) ECNet with different hidden configurations for $D_y/\lambda = 30$ and $M_0 = 61$ antennas.

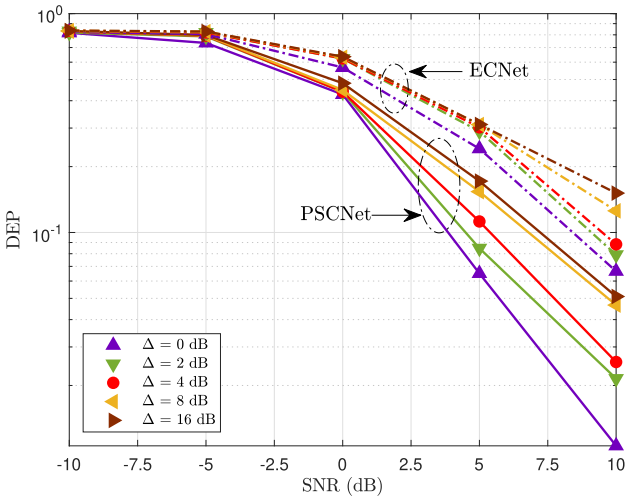

FIGURE 9. DEP validation of PSCNet and ECNet for various Δ.

that the matching steps and solving least square problems occupy most of the OMP complexity [13], and these steps yield $\mathcal{O}_{\text{OMP,max}} = O\left(M_0\left(\widetilde{P}+\mathcal{P}^3\right)T+\widetilde{P}^2T\right) \geq \mathcal{O}_{\text{OMP}} = O\left(M_0\widetilde{P}T+\widetilde{P}^2T\right)$ which is substantially larger than the complexity of the EF-based method, shown in (16). In addition, Fig. 7 shows that the complexity of the EF scheme with $T = 40$ snapshots still remains lower than those of the OMP technique with $T = 10$ snapshots. Therefore, the CS-based feature selection is excluded from the latter simulations, based on the aforementioned analysis.

ECNet and ERNet [21] are state-of-the-art data-driven methods to handle the signal-enumeration problem. It is noted from [21] that ECNet achieves better performance than ERNet. In addition, the enumeration mechanisms based on ECNet and ERNet are similar, so their computational complexities are approximately equal. Therefore, the ECNet is the only DL-based benchmark to compare with our proposed PSCNet in this paper. Fig. 8 shows the DEP in the validation stage of our proposed PSCNet and ECNet [21] with different numbers of hidden layers and neurons. In Fig. 8(a), the CNN networks with two convolutional layers provide more reliable enumeration results than those with one layer. In addition, too many hidden nodes in the FC layers lead to an overfitting phenomenon that worsens the performance of the PSCNet. It is observed that the PSCNet with $L_C = 2$ and $L_D = 4$ with $2M_0$ total neurons offers the highest accuracy; thus, this configuration is applied for subsequent simulations. For the ECNet, let $\widetilde{L}_D$ be the number of neurons. As shown in Fig. 8(b), the ECNet with a total of $5M_0$ neurons distributed into $\widetilde{L}_D = 4$ layers achieves the lowest DEP. Consequently, these parameters are utilized for the ECNet in the following experiments.

The DEP performance based on our proposed CNN and ECNet for $\Delta \in \{0, 2, 4, 8, 16\}$ dB is illustrated in Fig. 9. It is observed that the detection performance is significantly enhanced as Δ decreases. It is worth noting that training a deep network with signals having the same SNR values, i.e., $\Delta = 0$ dB, is impractical. Thus, the dataset is generated with the assumption of $\Delta = 4$ dB for the latter simulations.

Fig. 10 presents the DEP performance of our proposed PSCNet in different cases of D_y/λ with two deployment methods. The first one involves applying a pretrained network of an $\overline{M_0}$-antenna system with the PSS technique subsequently. The other solution is to employ an independent training model of an LAA with M_0 sensing elements. It is apparent that the DEP of PSCNet is enormous when the number of sensing elements M_0 is limited. For example, a 53-antenna LAA with $D_y/\lambda = 50$ achieves only 30% of the accuracy of signal enumeration in both coherent and noncoherent environments. In this case, half of the array size $M = 26$ violates the constraint (14), leading to high misdetection probabilities. Furthermore, in both Fig. 10(a) and Fig. 10(b), it is observed that the PSS process helps retain the error probability when the number of array elements increases. In contrast, the single-task learning models, which utilize all the elements

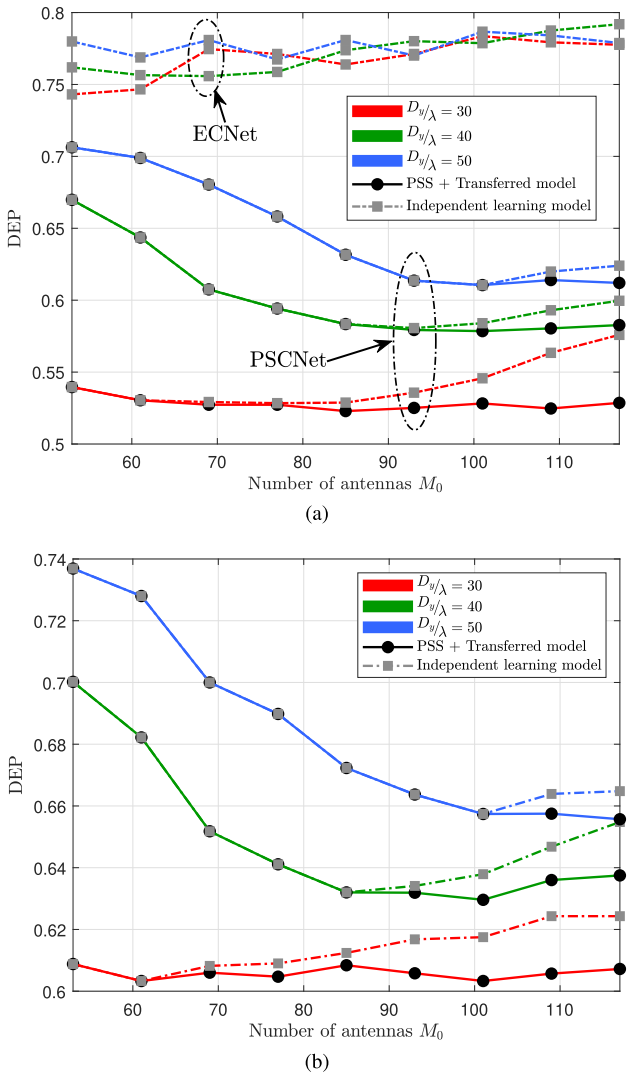

FIGURE 10. DEP versus number of antennas of PSCNet and ECNet with different values of D_y/λ in (a) a noncoherent environment and (b) a coherent environment.

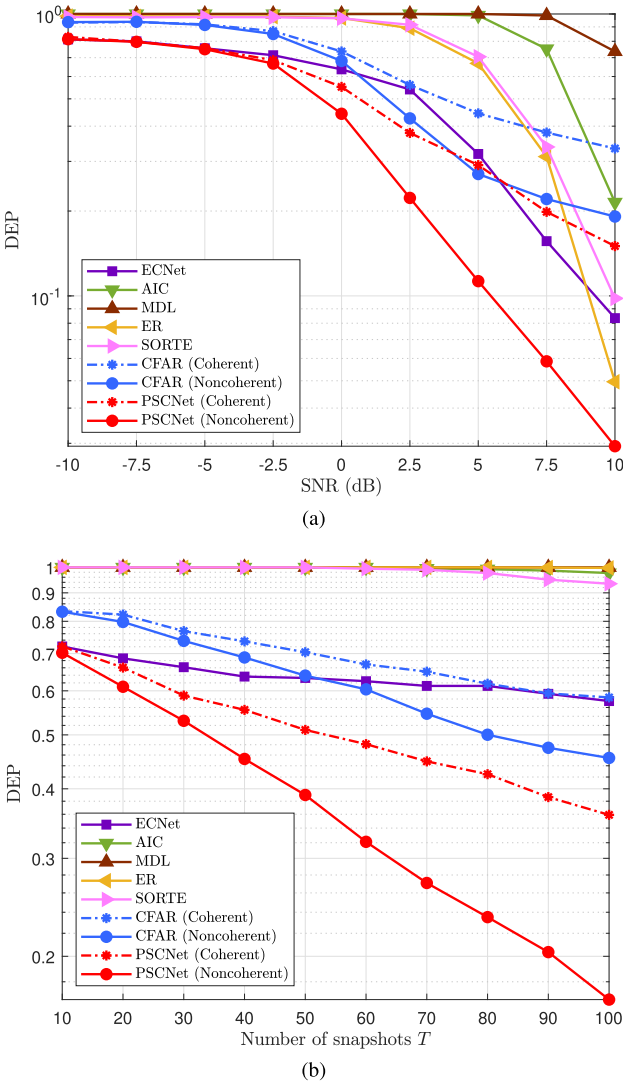

FIGURE 11. DEP versus (a) SNR for $T = 40$ snapshots, and (b) number of snapshots T for SNR = 0 dB.

of the power spectrum $\boldsymbol{r}$, have higher error probabilities than those of the counterpart. In Fig. 10(a), the upward trend also occurs in the DEP of ECNet as M_0 increases.

Fig. 11 compares the signal-enumeration performance of the PSCNet with that of six other detection schemes, including AIC [17], MDL [18], ER [20], SORTE [19], ECNet [21], and CFAR [22]. We note that the signal power spectrum can be constructed in both the coherent and noncoherent scenarios with an LAA, but the eigenvalues cannot; therefore, the PSCNet and CFAR detectors are the only methods that can predict the number of coherent signals. In Fig. 11(a), it is observed that the PSCNet outperforms the other compared enumerators. For example, for SNR = 5 dB, the fully uncorrelated signal misdetection of PSCNet is approximately 11%, whereas those of CFAR, ECNet, and SORTE are approximately 27%, 32%, and 71%, respectively. Fig. 11(b) also shows that the performance gain of the proposed method significantly increases as the number of snapshots T increases in the environment where SNR $=$ 0 dB. Additionally, the amplitudes of attenuation factors associated with the coherent signals are typically weaker than those of direct ones [29]. This yields $|\beta_i|^2 < |\beta_{q_1(k)}|^2$; thus, the peak representing the energy of a multipath signal impinging upon the mth antenna, i.e., $r_m \approx P_k|\beta_i|^2\sqrt{\alpha}+\sigma^2$, can be misclassified as a noise element due to its power degradation. This phenomenon makes the detection performance in a coherent environment worse than that in a noncoherent case, as shown in Fig. 11.

In Fig. 12, we compare the error probability of PSCNet, ECNet, and CFAR for various values of angular separation $\chi > 0$. In this simulation, we assume that two signals are coming from directions of ϕ_1, and $\phi_2 = \phi_1 + \chi$ in each data point, whereas other angular difference satisfies $|\phi_m - \phi_n| > \chi$, for $1 \le m < n \le P$, and $(m, n) \ne (1, 2)$. It is apparent that the PSCNet outperforms other schemes in both the high and low-SNR regions. Specifically, for SNR $=$ 10 dB, the error probability for detecting fully independent signals of

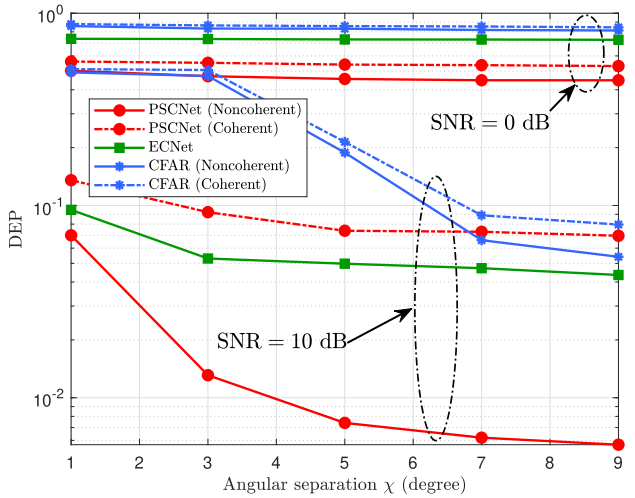

FIGURE 12. DEP versus angular separation for two DoAs of ϕ, and $(\phi + \chi)$.

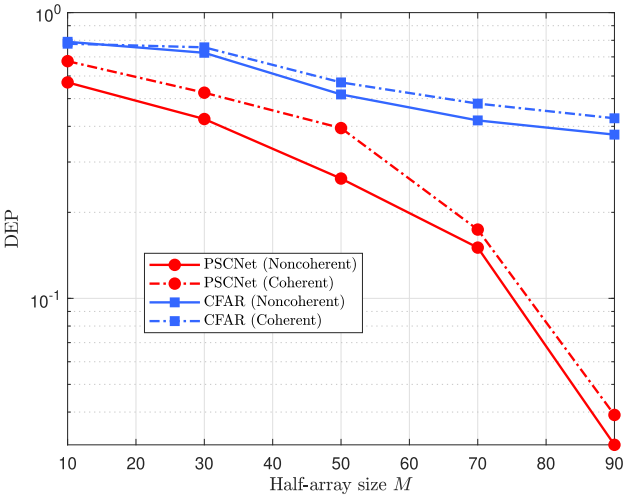

FIGURE 14. DEP versus half of the array size $M = \lceil D_y/\lambda \rceil$.

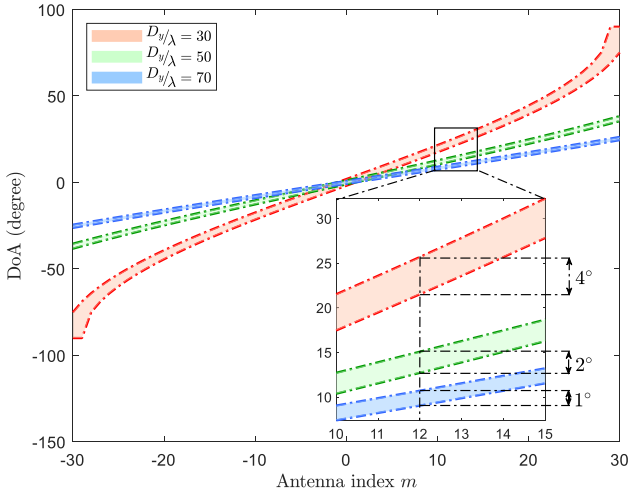

FIGURE 13. DoA distribution on antennas for various D_y/λ.

PSCNet is lower than 1% at $\chi = 9$, whereas those of ECNet and CFAR are 4.4% and 5.4%, respectively. In addition, our proposed PSCNet resolves the angular ambiguity problem better than CFAR by utilizing the same input feature, which is the EF power spectrum. It is worth emphasizing that the condition of having conspicuous peaks representing two closest DoAs is given as $|\sin(\phi + \chi) - \sin(\phi)| \geq 2\left(\frac{D_y}{\lambda}\right)^{-1}$, which means $\chi \geq \sin^{-1}\left(\sin(\phi) + 2\left(\frac{D_y}{\lambda}\right)^{-1}\right) - \phi$. For example, for $\phi = 58°$, the CFAR may miscount the signals unless χ satisfies $\chi \geq 8.2°$. On the contrary, the PSCNet can learn to predict the signal number based on the peak width; thus, it provides a better performance.

To further analyze the ambiguity effect of two closely incoming waves, we examine the distribution of signal directions on the mth antenna for different values of the D_y–λ ratio, as shown in Fig. 13. A far-field signal from the direction ϕ concentrating the most energy on the mth antenna satisfies $m - 1 < {D_y}/{\lambda} \sin\phi < m + 1$. Then the condition of ϕ is given by

$$\sin^{-1}\left(\frac{m-1}{D_y/\lambda}\right) < \phi < \sin^{-1}\left(\frac{m+1}{D_y/\lambda}\right). \tag{20}$$

In Fig. 13, the upper and lower limits in each case of the D_y–λ ratio are plotted by computing the right- and left-hand sides of (20), respectively. In addition, a shaded area between two bounds (indicated by dotted dashed lines) illustrates the possible DoAs of a signal arriving at the mth antenna. It is shown in Fig. 13 that the shaded area becomes more narrow as the value of ${D_y}/{\lambda}$ increases. For example, in a case of ${D_y}/{\lambda} = 30$, the antenna with $m = 12$ can receive a signal from direction $\phi \in [21°, 25°]$, which results in an ambiguity error of 4°. However, this error is reduced to only 1° when D_y–λ ratio is 70. Therefore, it is worth claiming that the DoA difference on an LAA's element can be reduced by expanding the normalized lens dimension ${D_y}/{\lambda}$.

Fig. 14 indicates the DEP performance when the D_y–λ ratio increases. We assume that $M = \left\lceil {D_y}/{\lambda} \right\rceil$ without loss of generality. In Fig. 14, it is shown that the DEP of the PSCNet considerably decreases for both the coherent and noncoherent scenarios when the number of antennas increases. Particularly, the increase of M yields a narrow DoA difference, improving the probability of accurately enumerating the incoming signals. In addition, the PSCNet also outperforms the CFAR detector. The explanation for this phenomenon is similar to the analysis in Fig. 12.

Fig. 15 compares the computational complexity between the feedforward propagation of ECNet and PSCNet with different input features. It is clear that the PSCNet-based signal enumeration has relatively lower complexity than the ECNet. According to [8] and [9], the eigenvalue decomposition in the ECNet-based scheme requires the complexity of $O\left(M_0^3\right)$ as M_0 increases. In contrast, for the proposed methods, the numbers of the operations are approximately $O(M_0)$ and

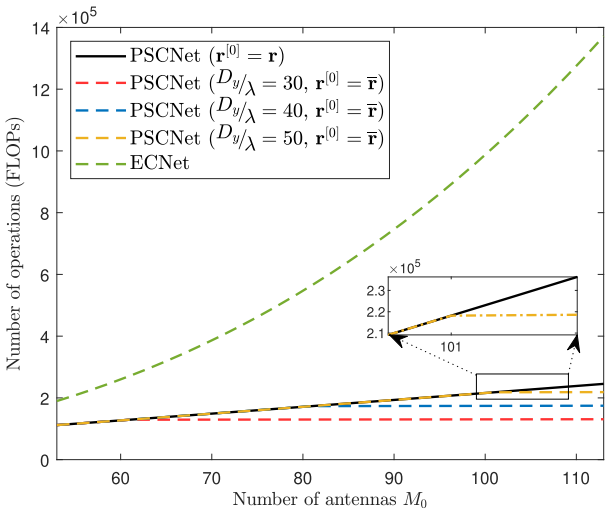

FIGURE 15. Computational complexity of PSCNet with different input features compared with ECNet for $T = 40$ snapshots.

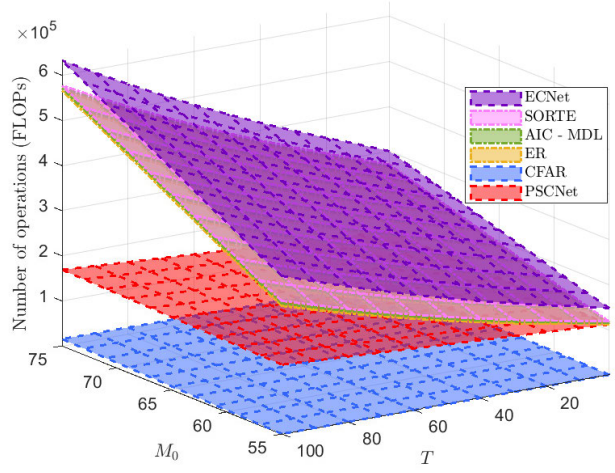

FIGURE 16. Complexity comparison between signal enumeration algorithms.

$O\left(\overline{M_0}\right)$ for the independent learning and transferred model, respectively. Furthermore, it can be observed that the PSS technique and reused models contribute to computational resource saving because the size of the input spectrum $\boldsymbol{r}^{[0]}$ remains unchanged as the number of antennas increases.

Fig. 16 illustrates the complexity of the PSCNet-based signal-enumeration algorithm compared with those of the prior schemes. It is shown that the computational load of the proposed method is substantially lower than those of subspace-based schemes as M_0 and T increase. In the eigenvalue-aided enumeration, the computation of the signal covariance matrix and eigenvalue decomposition process occupy the most complexity, which is $O\left(M_0^3 + M_0^2 T\right)$ [8], [9], whereas the maximum computational cost of the proposed method, according to (16) and (17), is $\mathcal{O}_{\text{PSCNet}} = O\left(M_0 T\right)$ when M_0 and T are large enough. Furthermore, it is noted that our scheme requires more computational resources than the CFAR method owing to the use of CNN. Although the feedforward propagation of the PSCNet is more computationally expensive than determining the CFAR threshold, it yields reliable detection performance.

V. CONCLUSION

In this paper, we have presented a novel CNN-based signal-enumeration scheme for a symmetric LAA. The power spectrum, which is formed based on the energy distributed across antenna elements, is used in both the noncoherent and coherent scenarios. The proposed PSCNet employs several 1D convolutional layers to extract the feature maps from the input spectrum and generate a one-hot vector representing the number of impinging signals. Furthermore, the PSS technique maintains the size of the input feature so that the reused model can be applied regardless of the number of antennas. This saves on the computational resources in practical deployment and reduces the time for training. Compared with its predecessors [17], [18], [19], [20], [21], the proposed detector provides more reliable performance with substantially lower computational complexity. Lastly, based on the numerical simulation and analysis results, we can conclude that a large-scale LAA system with an extensive normalized lens dimension can achieve a more reliable signal enumeration performance.

REFERENCES

[1] Y. Zeng and R. Zhang, "Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1557–1571, Apr. 2016.
[2] Z. Ni, Y. Luo, M. Motani, and C. Li, "DoA estimation for lens antenna array via root-MUSIC, outlier detection, and clustering," *IEEE Access*, vol. 8, pp. 199187–199196, 2020.
[3] F. Dong, W. Wang, Z. Huang, and P. Huang, "High-resolution angle-of-arrival and channel estimation for mmWave massive MIMO systems with lens antenna array," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 12963–12973, Nov. 2020.
[4] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 4, pp. 806–811, Apr. 1985.
[5] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Trans. Acoust., Speech Signal Process.*, vol. ASSP-37, no. 1, pp. 8–15, Jan. 1989.
[6] W. Du and R. L. Kirlin, "Improved spatial smoothing techniques for DOA estimation of coherent signals," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1208–1210, May 1991.
[7] F.-M. Han and X.-D. Zhang, "An ESPRIT-like algorithm for coherent DOA estimation," *IEEE Antennas Wireless Propag. Lett.*, vol. 4, pp. 443–446, 2005.
[8] D. T. Hoang and K. Lee, "Deep learning-aided coherent direction-of-arrival estimation with the FTMR algorithm," *IEEE Trans. Signal Process.*, vol. 70, pp. 1118–1130, 2022.
[9] D. T. Hoang and K. Lee, "Coherent signal enumeration based on deep learning and the FTMR algorithm," in *Proc. IEEE Int. Conf. Commun.*, Seoul, South Korea, May 2022, pp. 5098–5103.
[10] C. Qian, L. Huang, W. J. Zeng, and H. C. So, "Direction-of-arrival estimation for coherent signals without knowledge of source number," *IEEE Sensors J.*, vol. 14, no. 9, pp. 3267–3273, Sep. 2014.
[11] W. Zhang, Y. Han, M. Jin, and X.-S. Li, "An improved ESPRIT-like algorithm for coherent signals DOA estimation," *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 339–343, Nov. 2019.
[12] W. Zhang, Y. Han, M. Jin, and X. Qiao, "Multiple-Toeplitz matrices reconstruction algorithm for DOA estimation of coherent signals," *IEEE Access*, vol. 7, pp. 49504–49512, 2019.

[13] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Jan. 2007.
[14] Y. Arjoune, N. Kaabouch, H. El Ghazi, and A. Tamtaoui, "Compressive sensing: Performance comparison of sparse recovery algorithms," in *Proc. IEEE 7th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2017, pp. 1–7.
[15] M. Lin, M. Xu, X. Wan, H. Liu, Z. Wu, J. Liu, B. Deng, D. Guan, and S. Zha, "Single sensor to estimate DOA with programmable metasurface," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 10187–10197, Jun. 2021.
[16] K. Aghababaiyan, V. Shah-Mansouri, and B. Maham, "High-precision OMP-based direction of arrival estimation scheme for hybrid non-uniform array," *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 354–357, Feb. 2020.
[17] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.
[18] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.
[19] Z. He, A. Cichocki, S. Xie, and K. Choi, "Detecting the number of clusters in n-way probabilistic clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2006–2021, Nov. 2010.
[20] H. Sun, J. Guo, and L. Fang, "Improved singular value decomposition (TopSVD) for source number estimation of low SNR in blind source separation," *IEEE Access*, vol. 5, pp. 26460–26465, 2017.
[21] Y. Yang, F. Gao, C. Qian, and G. Liao, "Model-aided deep neural network for source number detection," *IEEE Signal Process. Lett.*, vol. 27, pp. 91–95, 2020.
[22] M. A. Richards, *Fundamentals of Radar Signal Processing*. New York, NY, USA: McGraw-Hill, 2005.
[23] Y. I. Abramovich, N. K. Spencer, and A. Y. Gorokhov, "Resolving manifold ambiguities in direction-of-arrival estimation for nonuniform linear antenna arrays," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2629–2643, Oct. 1999.
[24] D. Bonacci, F. Vincent, and B. Gigleux, "Robust DoA estimation in case of multipath environment for a sense and avoid airborne radar," *IET Radar, Sonar Navigat.*, vol. 11, no. 5, pp. 797–801, May 2017.
[25] X. Chen and Y. Morton, "Iterative subspace alternating projection method for GNSS multipath DOA estimation," *IET Radar, Sonar Navigat.*, vol. 10, no. 7, pp. 1260–1269, Aug. 2016.
[26] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," *Adv. Neural Inform. Proc. Sys.*, vol. 30, pp. 971–980, Dec. 2017.
[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, Dec. 2014, pp. 1–15.
[28] R. Hunger, *Floating Point Operations in Matrix-Vector Calculus*. Munich, Germany: Munich University of Technology, Institute for Circuit Theory and Signal Processing, 2005.
[29] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 3679–3684.

DAI TRONG HOANG received the B.S. degree in control engineering and automation from the Ho Chi Minh City University of Technology (HCMUT), Ho Chi Minh City, Vietnam, in 2019, and the M.S. degree in electrical and information engineering from the Seoul National University of Science and Technology, Seoul, South Korea, in 2021. He is currently a Research Scientist with the Research Center for Electrical and Information Technology, Seoul National University of Science and Technology. His research interests include the areas of applied machine learning, wireless communication, signal processing, and optimization.

KYUNGCHUN LEE (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, in 2000, 2002, and 2007, respectively. From April 2007 to June 2008, he was a Postdoctoral Researcher at the University of Southampton, U.K. From July 2008 to August 2010, he was at Samsung Electronics, Suwon, South Korea. Since September 2010, he has been with the Seoul National University of Science and Technology, South Korea. In 2017, he was a Visiting Assistant Professor at North Carolina State University, Raleigh, NC, USA. His research interests include wireless communications and applied machine learning. He received the Best Paper Awards at the IEEE International Conference on Communications (ICC) and IEEE Wireless Communications and Networking Conference (WCNC), in 2009 and 2020, respectively.