A Dilated Inception Convolutional Neural Network for Gridless DOA Estimation Under Low SNR Scenarios

Zhi-Wei Tan, Yuan Liu, Andy W. H. Khong
Nanyang Technological University, School of Electrical and Electronic Engineering, Singapore
E-mail: zhiwei001@e.ntu.edu.sg, {yuan.liu, andykhong}@ntu.edu.sg

***Abstract*—This paper addresses the direction-of-arrival (DOA) estimation-based source localization problem by using the convolutional neural network (CNN) and root-MUltiple SIgnal Classification (MUSIC) technique. Existing grid-less neural network-based approach employs a LeNet-based CNN, where its network complexity depends on the number of sensors. To overcome this issue, we propose a LeDIM-net CNN that works for a uniform linear array with an arbitrary number of sensors. The proposed LeDIM-net architecture maintains spatial resolution throughout the network while exploiting non-local spatial information. Simulation results demonstrate the effectiveness of the proposed LeDIM-net over the existing grid-less LeNet-based approach and root-MUSIC at low SNRs for arrays with different sensors by maintaining the same network complexity.**

***Index Terms*—Direction-of-arrival (DOA) estimation, convolution neural network, deep learning, array signal processing, gridless DOA estimation**

I. Introduction

Direction-of-arrival (DOA) estimation is important for radar [1], sonar [2], and acoustics [3] applications. Subspace methods such as multiple signal classification (MUSIC), Root-MUSIC, and estimation of signal parameters via rotational invariance techniques (ESPRIT) rely on long snapshots and high signal-to-noise ratios (SNRs) to achieve reasonable performance [4–9]. These methods decompose spatial information into the noise and signal subspaces before exploiting them for DOA estimation. With sufficient snapshots, although these approaches are effective under high SNR conditions, DOA accuracy often deteriorates under low SNRs. In addition, compressive sensing-based methods [10–13] have been proposed to achieve high-resolution DOA estimation.

Approaches based on deep learning have gained increasing attention as they achieve satisfactory performance for target localization applications, especially under low SNR scenarios [14]. In general, the DOA estimation process is formulated as a multi-label classification task [15, 16], where a deep neural network (DNN) estimates probabilities of source directions within a pre-defined grid of a specified angular resolution. This approach allows the DNN to estimate multiple sources within a given grid resolution. However, such approaches result in performance degradation due to grid mismatch.

In recent years, gridless-based neural network approaches have been proposed to alleviate the grid mismatch problem [14]. These approaches employ DNN for the estimation of the noise-free spatial covariance matrix (SCM). A gridless DOA estimate is then achieved by employing post-processing modules such as Root-MUSIC described in [14]. However, such a regression-based approach is inefficient since the number of parameters of a noise-free SCM scales quadratically with the number of sensors resulting in overfitting issues for the DNN. In [14], the Toeplitz property is employed to reduce the number of unknown parameters to the same order of magnitude as the number of sensors for SCM estimation. While the estimation accuracy of the SCM has been improved, the LeNet-based convolution neural network (CNN) [17] employed in [14] is designed specifically for an array with four sensor elements. One of the most parameter-efficient approaches that cater to the extension of the CNN to an arbitrary number of sensor elements is proposed in [18]. However, this technique still requires a significant number of neural network parameters with the increasing number of sensor elements.

To this end, we propose LeDIM-net — a CNN for reconstructing the noise-free spatial covariance matrix for gridless DOA estimation. The proposed LeDIM-net extracts non-local spatial information via dilated inception modules (DIMs). Each of these modules comprises convolution operations of various dilation rates leading to a large receptive field at earlier layers of the network. The use of DIM also reduces the number of convolutional layers required to achieve a similar receptive field, allowing the proposed LeDIM-net to be extended to a uniform linear array (ULA) with arbitrary sensor elements without increasing the neural network parameter complexity. Furthermore, the proposed LeDIM-net architecture maintains the spatial resolution throughout the network, allowing it to operate on an array of an arbitrary number of sensors. Simulation results show the efficacy of the proposed LeDIM-net architecture, which outperforms the approach in [14] and Root-MUSIC, especially in low SNRs scenarios for a uniform linear array (ULA) with different numbers of sensor elements while maintaining the same

This research was conducted under project WP6 within the Delta-NTU Corporate Lab with funding support from A*STAR under its IAF-ICP programme (Grant no: I2201E0013) and Delta Electronics Inc.

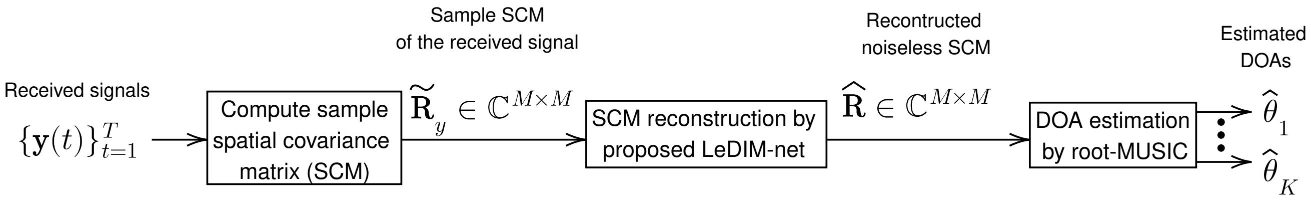
Received signals
$\{\mathbf{y}(t)\}_{t=1}^{T}$
Compute sample spatial covariance matrix (SCM)
Sample SCM of the received signal
$\widetilde{\mathbf{R}}_y \in \mathbb{C}^{M\times M}$
SCM reconstruction by proposed LeDIM-net
Recontructed noiseless SCM
$\widehat{\mathbf{R}} \in \mathbb{C}^{M\times M}$
DOA estimation by root-MUSIC
Estimated DOAs
$\widehat{\theta}_1$
$\widehat{\theta}_K$

Fig. 1. The framework of the gridless DOA estimator with the proposed LeDIM-net and root-MUSIC.

network parameter complexity.

II. Signal Model

We assume K far-field narrow-band uncorrelated source signals impinging onto an M-element ULA. The $M \times 1$ received signal vector is given by

$$\mathbf{y}(t) = \mathbf{A}(r, \boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t), \tag{1}$$

where $t = 1, \dots, T$ denotes the time index with T being the number of snapshots, $\mathbf{s}(t) = [s_1(t), \dots s_K(t)]^{\mathsf{T}} \in \mathbb{C}^{K\times 1}$ denotes the signal vector, $\mathbf{n}(t) \in \mathbb{C}^{M\times L}$ is the Gaussian white noise, and $(\cdot)^{\mathsf{T}}$ denotes the transpose operator. Here, the $M \times K$ array steering matrix

$$\mathbf{A}(r, \boldsymbol{\theta}) = [\mathbf{a}(r, \theta_1), \dots, \mathbf{a}(r, \theta_K)] \tag{2}$$

comprises K columns of of the array steering vector such that each vector $\mathbf{a}(r, \theta_k) = [1, e^{j\pi r \sin\theta_k} \dots, e^{j\pi(M-1)r\sin\theta_k}]^{\mathsf{T}} \in \mathbb{C}^{M\times 1}$ models the phase of the kth source signal arriving from θ_k. The parameter r defines the number of half-wavelengths between array sensors and $\boldsymbol{\theta} = [\theta_1, \dots, \theta_K]^{\mathsf{T}} \in \mathbb{R}^{K\times 1}$ is the true DOA vector.

Arising from the above, the spatial covariance matrix (SCM) of the received signal can be expressed as

$$\begin{aligned}\mathbf{R}_y &= \mathbb{E}\left\{\mathbf{y}(t)\mathbf{y}^{\mathsf{H}}(t)\right\}\\ &= \mathbb{E}\left\{\mathbf{A}(r, \boldsymbol{\theta})\mathbf{s}(t)\mathbf{s}^{\mathsf{H}}(t)\mathbf{A}^{\mathsf{H}}(r, \boldsymbol{\theta})\right\} + \mathbb{E}\left\{\mathbf{n}(t)\mathbf{n}^{\mathsf{H}}(t)\right\}\\ &= \mathbf{A}(r, \boldsymbol{\theta})\mathbf{R}_s\mathbf{A}^{\mathsf{H}}(r, \boldsymbol{\theta}) + \sigma_N^2\mathbf{I}_M,\end{aligned} \tag{3}$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator, $\mathbf{R}_s = \mathbb{E}\left\{\mathbf{s}(t)\mathbf{s}^{\mathsf{H}}(t)\right\}$ is the source covariance matrix, σ_N^2 is the noise power, $\mathbf{I}_M \in \mathbb{R}^{M\times M}$ is an identity matrix, and $(\cdot)^{\mathsf{H}}$ is the conjugate transpose operator. Here, we note that the noise-free covariance matrix $\mathbf{R} = \mathbf{A}(r, \boldsymbol{\theta})\mathbf{R}_s\mathbf{A}^{\mathsf{H}}(r, \boldsymbol{\theta}) \approxeq \mathbf{A}(r, \boldsymbol{\theta})\mathbf{A}^{\mathsf{H}}(r, \boldsymbol{\theta})$ encompasses DOA information of the sources. In general, due to the limited number of snapshots in practical applications, the sample SCM is computed via

$$\widetilde{\mathbf{R}}_y = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}(t)\mathbf{y}^{\mathsf{H}}(t). \tag{4}$$

Given the received signals $\{\mathbf{y}(t)\}_{t=1}^{T}$ and r, the goal of the proposed gridless DOA estimation approach is to estimate the source DOAs $\widehat{\boldsymbol{\theta}} = [\widehat{\theta}_1, \dots, \widehat{\theta}_K]^{\mathsf{T}} \in \mathbb{R}^{K\times 1}$.

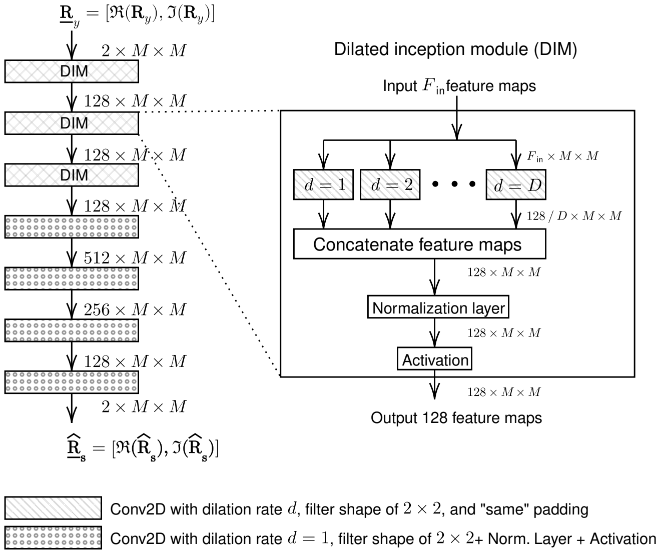
$\underline{\mathbf{R}}_y = [\Re(\mathbf{R}_y), \Im(\mathbf{R}_y)]$
$2 \times M \times M$
DIM
$128 \times M \times M$
DIM
$128 \times M \times M$
DIM
$128 \times M \times M$
$512 \times M \times M$
$256 \times M \times M$
$128 \times M \times M$
$2 \times M \times M$
$\widehat{\underline{\mathbf{R}}}_s = [\Re(\widehat{\mathbf{R}}_s), \Im(\widehat{\mathbf{R}}_s)]$
Dilated inception module (DIM)
Input F_{in}feature maps
$F_{\text{in}} \times M \times M$
$d = 1$
$d = 2$
$d = D$
$128 / D \times M \times M$
Concatenate feature maps
$128 \times M \times M$
Normalization layer
$128 \times M \times M$
Activation
$128 \times M \times M$
Output 128 feature maps
Conv2D with dilation rate d, filter shape of 2×2, and "same" padding
Conv2D with dilation rate $d = 1$, filter shape of 2×2+ Norm. Layer + Activation

Fig. 2. The architecture of the proposed LeDIM-net for estimating the noise-free covariance matrix in low SNR. The LeDIM-net comprises three dilated inception modules (DIM), followed by four convolution modules. The dimension of the features maps is denoted on the right of each arrow, and each output feature map has a shape of $M \times M$.

III. The Proposed LeDIM-net Architecture

The proposed CNN-based gridless DOA estimation framework is shown in Fig. 1. A new covariance reconstruction neural network based on CNN is first employed to reconstruct the noise-free SCM from $\widetilde{\mathbf{R}}_y$. With the estimated noise-free SCM $\widehat{\mathbf{R}}$, the Root-MUSIC is employed to achieve source DOAs. To cater for the ability to extend the number of sensor elements without increasing the neural network parameters, we proposed the LeDIM-net CNN—a dilated inception module-based method as shown in Fig. 2. The proposed LeDIM-net consists of seven layers expressed as

$$\widehat{\underline{\mathbf{R}}} = f_7(\dots f_1(\widetilde{\underline{\mathbf{R}}}_y)), \tag{5}$$

where $\widetilde{\underline{\mathbf{R}}}_y = \left[\Re\left(\widetilde{\mathbf{R}}_y\right), \Im\left(\widetilde{\mathbf{R}}_y\right)\right] \in \mathbb{R}^{2\times M\times M}$ and $\widehat{\underline{\mathbf{R}}} = \left[\Re\left(\widehat{\mathbf{R}}\right), \Im\left(\widehat{\mathbf{R}}\right)\right] \in \mathbb{R}^{2\times M\times M}$ are the real-imaginary composites of the sample SCM and the reconstructed noise-free SCM, respectively. Here, $\Re(\cdot)$ and $\Im(\cdot)$ are operators that extract the

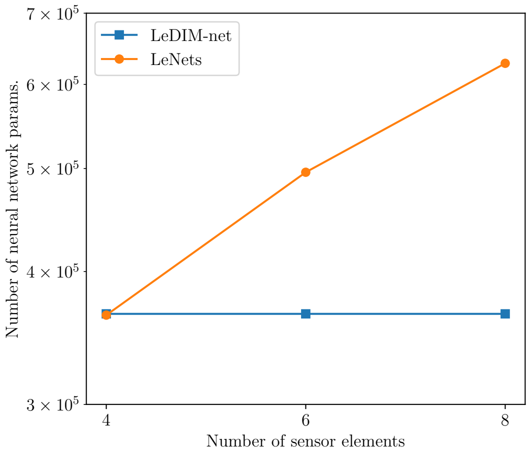

Fig. 3. Variation of the number of network parameters with the number of sensor elements.

real and imaginary components, respectively.

In (5), each of the first three functions $\{f_l(\cdot)\}_{l=1,2,3}$ represents a DIM [19] with 128 two-dimensional filters of shape 2×2 and while each of the following three functions $\{f_l(\cdot)\}_{l=4,5,6}$ represents a series of convolution blocks with 512, 256, and 128 filter of shape 1×1. The last function $f_7(\cdot)$ represents a convolution layer with two filters of shape 1×1. It is important to highlight that the network parameter complexity of the LeNet-based CNN in [14, 18] increases significantly with M, while the proposed LeDIM-net in (5) is not dependent on M. Specifically, the number of network parameters in each function $\{f_l(\cdot)\}_{l=1,\ldots,7}$ are 1408, 65920, 65920, 67072, 131840, 33152, and 258, respectively.

The Root-MUSIC algorithm [5] is subsequently employed to achieve DOA estimation without pre-defined grids via

$$\widehat{\boldsymbol{\theta}} = f_{\mathsf{R-MUSIC}}(\widehat{\underline{\mathbf{R}}}, r), \tag{6}$$

where $\widehat{\boldsymbol{\theta}} = [\widehat{\theta}_1, \ldots, \widehat{\theta}_K]^{\mathsf{T}}$ is the estimated DOA vector and $f_{\mathsf{R-MUSIC}}(\cdot)$ denotes the Root-MUSIC function.

Similar to the CNN in [14], the proposed approach employs the convolution operation along the spatial dimensions. By doing so, the network can leverage parameter sharing and equivalent representation across different sensors resulting in more efficient parameter learning than conventional fully-connected layers. In contrast to the LeNet-based CNN in [14], where each convolution layer extracts adjacent spatial features of its input feature maps, the dilation property of the convolution filter [19–22] within the DIM enables the extraction of non-local spatial information in the earlier layers of LeDIM-net without increasing the network parameter complexity. Specifically, the proposed approach can achieve a receptive field of lD elements at its lth layer compared to that of $l{+}1$ elements in LeNet-based CNN [18]. Hence, the proposed LeDIM-net can avoid the need of appending additional convolution layers to the CNN in [18] to obtain a larger receptive field. While the network complexity increases in terms of floating point operations, we note that the network parameter complexity does not increase with M.

The proposed architecture can easily be extended to M channels since each function $\{f_l(\cdot)\}_{l=1,\ldots,7}$ in (5) generates output feature maps with the same spatial resolution as its input feature maps, i.e., $f_1(\cdot) : \mathbb{R}^{2\times M\times M} \to \mathbb{R}^{128\times M\times M}$. This is achieved by padding the input features with zeros before the convolution operation in $\{f_l(\cdot)\}_{l=1,2,3}$. With this structure, the spatial resolution is preserved throughout the neural network, and the architecture of the proposed LeDIM-net can therefore operate with $\widetilde{\underline{\mathbf{R}}}_y$ for arbitrary M.

To train the parameters of the LeDIM-net, we employ a mean squared error (MSE) as the loss function via

$$\mathcal{L}_{\mathsf{MSE}}(\widehat{\underline{\mathbf{R}}}, \underline{\mathbf{R}}) = \frac{1}{2M^2}\sum_{q=1}^{2}\sum_{j=1}^{M}\sum_{i=1}^{M}\left(\widehat{\underline{\mathbf{R}}}(q,j,i) - \underline{\mathbf{R}}(q,j,i)^2\right),$$

where $\mathcal{L}_{\mathsf{MSE}}(\cdot)$ is the MSE function, $\underline{\mathbf{R}} = [\Re\left(\mathbf{A}(r,\boldsymbol{\theta})\mathbf{A}^{\mathsf{H}}(r,\boldsymbol{\theta})\right), \Im\left(\mathbf{A}(r,\boldsymbol{\theta})\mathbf{A}^{\mathsf{H}}(r,\boldsymbol{\theta})\right)] \in \mathbb{R}^{2\times M\times M}$, q, j, and i are indices for the real-imaginary composite, row, and columns of the covariance matrices, respectively. It is worth noting that we do not impose the Toeplitz property when estimating the noise-free SCM by using the proposed LeDIM-net. This allows the proposed neural network to have the potential to be extended to other array configurations that cannot leverage the Toeplitz property.

IV. Simulation Results

A. Training and testing dataset

To train the proposed LeDIM-net, we simulate a dataset $\mathcal{D}_M$ with $500,000$ training, and $50,000$ validation data points for each $M \in [4,6,8]$. A ULA with $r = 0.5$ wavelength is employed. Each data point in the dataset has $T = 256$ snapshots and $K = 2$ angles uniformly generated from $[-60°, 60°]$ with SNR uniformly generated from -15 dB to 5 dB. For each data point, the angle between sources in the datasets is at least $\frac{2}{3} \times \theta_{\text{3dB}}$ apart, where θ_{3dB} denotes the halfpower beamwidth. For testing, we generate $10,000$ testing samples for each $M \in [4,6,8]$, and scale its SNR$\in [-12,-9,-6,-3,0]$ dB.

B. Baselines and training hyperparameters

Due to the limitation of LeNet-based CNN employed in [14], which has been designed explicitly for $M = 4$, we added two additional convolutional blocks resulting in $M = 6$ and 132k additional network parameters (36% increase), and four more resulting in $M = 8$ and 264k additional network parameters (72% increase). This extension is similar to the architecture described in [18]. For clarity, we denote the LeNets following its number of layers as LeNet-7, LeNet-9, and LeNet-11 which are trained using $\mathcal{D}_4, \mathcal{D}_6$ and $\mathcal{D}_8$, respectively. For the proposed LeDIM-net, we employ $D = 4$, batch normalization layer [23], and Leaky ReLU as activation [24]. One LeDIM-net is trained for each $\mathcal{D}_M$ without modification to its architecture.

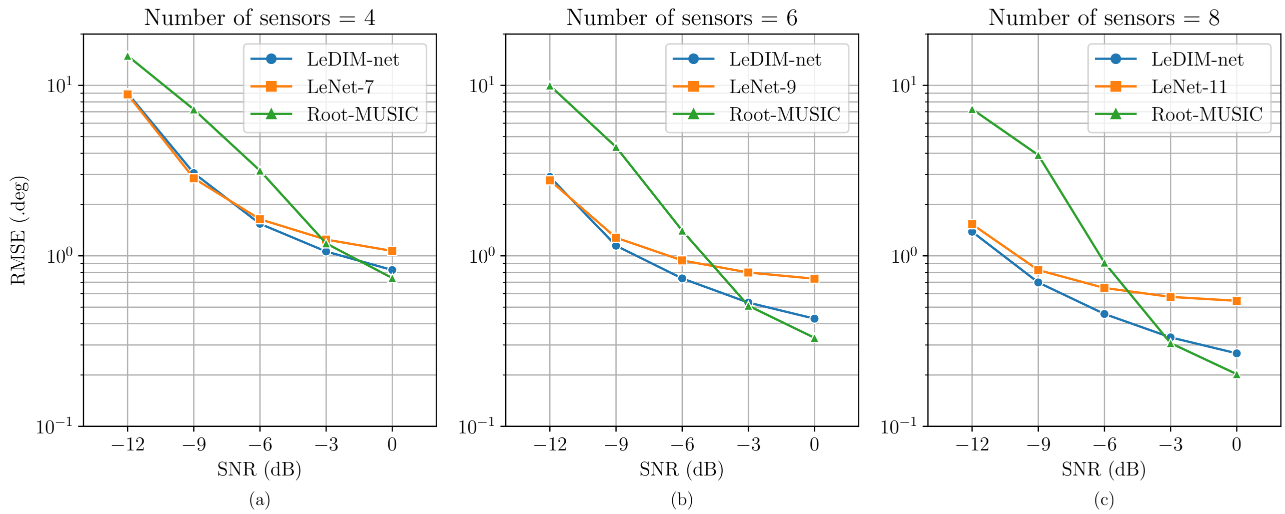

Fig. 4. The RMSE-SNR performance curve of the proposed LeDIM-net, LeNets, and Root-MUSIC for (a) $M = 4$, (b) $M = 6$, and (c) $M = 8$.

For updating the LeDIM-net and LeNets parameters, we employed Adam [25] with an initial learning rate of 0.001. The learning rate is subsequently halved after the validation root mean squared error (RMSE) of its DOA plateaus for five epochs. We set a batch size of 256 and trained each network for fifty epochs. The loss objective of LeNets is the MSE between the estimated and the actual row vector of the noise-free SCM [14]. As such, for a fairer comparison between LeNets and LeDIM-net, instead of using validation loss to select the best parameter from each model for testing, we used the lowest validation RMSE of its estimated DOA. It is important to note that both frameworks employ the same input space and Root-MUSIC to achieve source DOAs.

C. DOA estimation

The test performance of the gridless framework with the proposed LeDIM-net is compared with Root-MUSIC and the LeNets with Toeplitz property [14]. The RMSE results are plotted in Fig. 4 for different SNRs. The number of network parameters associated with each network is plotted in Fig. 3. The proposed LeDIM-net generally outperforms the LeNets for $M \in [4, 6, 8]$ and SNR $\in [-12, -9, -6, -3, 0]$ dB despite employing the same number of parameters. This result highlight that the proposed approach can be extended to different M without an increase in network parameters. Notably, both LeNets and LeDIM-net suffer from modestly lower performance than that of Root-MUSIC at SNRs of -3 dB and 0 dB. This result is consistent with that in [14], and is attributed to artefacts being introduced to the estimation of the noise-free SCM at higher SNRs.

D. Effectiveness of dilated inception modules

To compare the efficacy of the dilation in the proposed LeDIM-net, we perform an ablation study where the standard convolution is employed, i.e., $D = 1$. We then compared its performance with that of $D = 4$. Here, it is important to note that LeDIM-net with $D = 4$ and $D = 1$ requires the same number of network parameters. Fig. 5 illustrates the test RMSE results. Since LeDIM-net with $D = 4$ can capture non-local spatial information and a larger receptive field at each layer, it achieves a lower RMSE than LeDIM-net with $D = 1$. It is useful to note that LeDIM-net with $D = 1$ outperforms Root-MUSIC by 21.7%, 34.1%, and 40.4% in terms of RMSE for 4, 6, and 8 sensors elements, respectively.

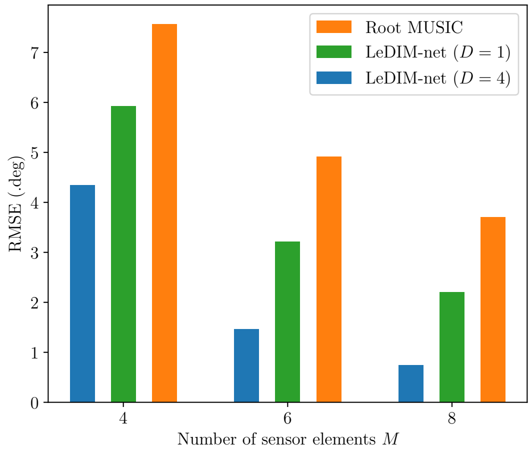

Fig. 5. The RMSE performance curve of the proposed LeDIM-net (D=1, D=4), and Root-MUSIC for (a) $M = 4$, (b) $M = 6$, and (c) $M = 8$.

V. Conclusion

We propose a LeDIM-net-based gridless DOA estimation algorithm that can be extended for a ULA with different sensor elements without increasing the neural network parameter complexity. Specifically, the proposed LeDIM-net in the framework maintains the spatial resolution throughout its network to extend to a different number of sensor elements.

Further, to avoid increasing the network parameters with an increasing number of sensor elements, the proposed LeDIM-net employs a range of dilation at earlier layers to exploit non-local spatial information. Simulation results indicate that the proposed approach outperforms existing LeNet-based CNN and Root-MUSIC algorithms.

References

[1] Y. Liu, Z.-W. Tan, A. W. H. Khong, and H. Liu, "Joint source localization and association through overcomplete representation under multipath propagation environment," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2022, pp. 5123–5127.

[2] W. Shi, J. Huang, and Y. Hou, "Fast DOA estimation algorithm for MIMO sonar based on ant colony optimization," *Journal of Systems Engineering and Electronics*, vol. 23, no. 2, pp. 173–178, 2012.

[3] K. Wu, V. G. Reju, A. W. H. Khong, and S. T. Goh, "Swarm intelligence based particle filter for alternating talker localization and tracking using microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 6, pp. 1384–1397, 2017.

[4] R. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph.D. dissertation, Stanford University, 1982.

[5] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 8, 1983, pp. 336–339.

[6] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Ant. Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[7] R. Roy and T. Kailath, "ESPRIT-Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.

[8] V. V. Reddy, B. P. Ng, and A. W. H. Khong, "Insights into MUSIC-like algorithm," *IEEE Trans. Signal Process.*, vol. 61, pp. 2551–2556, 2013.

[9] Y. Liu, H. Liu, L. Wang, and G. Bi, "Target localization in high-coherence multipath environment based on low-rank decomposition and sparse representation," *IEEE Trans. Geosci. Remote Sens.,*, vol. 58, no. 9, pp. 6197–6209, 2020.

[10] Y. Liu, B. Jiu, H. Liu, L. Zhang, and Y. Zhao, "Compressive sensing for very high frequency radar with application to low-angle target tracking under multipath interference," in *Proc. Int. Workshop Compressed Sens. Theory Appl. Radar, Sonar Remote Sens.*, 2016, pp. 188–192.

[11] Y. Liu, H. Liu, X.-G. Xia, L. Wang, and G. Bi, "Target localization in multipath propagation environment using dictionary-based sparse representation," *IEEE Access*, vol. 7, pp. 150 583–150 597, 2019.

[12] Y. Liu, X.-G. Xia, H. Liu, A. H. T. Nguyen, and A. W. H. Khong, "Iterative implementation method for robust target localization in a mixed interference environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.

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

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

[15] X. Xiao, S. Zhao, X. Zhong, D. L. Jones, E. S. Chng, and H. Li, "A learning-based approach to direction of arrival estimation in noisy and reverberant environments," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2015, pp. 2814–2818.

[16] S. Chakrabarty and E. A. P. Habets, "Broadband DOA estimation using convolutional neural networks trained with noise signals," in *Proc. IEEE Workshop App. Signal Process. Audio and Acoust.*, 2017, pp. 136–140.

[17] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[18] S. Chakrabarty and E. A. P. Habets, "Multi-speaker DOA estimation using deep convolutional networks trained with noise signals," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 1, pp. 8–21, 2019.

[19] S. Yang, G. Lin, Q. Jiang, and W. Lin, "A dilated inception network for visual saliency prediction," *IEEE Trans. Multimed.*, vol. 22, pp. 2163–2176, 2020.

[20] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *Proc. Int. Conf. Learn. Represent.*, 2016.

[21] Z.-W. Tan, A. H. T. Nguyen, and A. W. H. Khong, "An efficient dilated convolutional neural network for UAV noise reduction at low input SNR," in *Proc. Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2019, pp. 1885–1892.

[22] Z.-W. Tan, A. H. T. Nguyen, Y. Liu, and A. W. H. Khong, "Multichannel noise reduction using dilated multichannel U-net and pre-trained single-channel network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2022, pp. 266–270.

[23] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Proc. Int. Conf. Mach. Learn.*, pp. 448–456, 2015.

[24] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn.*, 2013.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 127–142.