Deep Residual Learning Based Localization of Near-Field Sources in Unknown Spatially Colored Noise Fields

Zhuoqian Jiang[†‡], Jingmin Xin[†‡], Weiliang Zuo[†‡♮], Nanning Zheng[†‡], and Akira Sano[§]
[†] Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China
[‡] National Engineering Laboratory for Visual Information Processing and Applications, Xi'an Jiaotong University, Xi'an 710049, China
[♮] Shunan Academy of Artificial Intelligence, Ningbo, Zhejiang 315000, China
[§] Department of System Design Engineering, Keio University, Yokohama 223-8522, Japan

***Abstract*—In this paper, we explore the problem of near-field source localization in an unknown spatially colored noise environment using an end-to-end neural network which is based on deep residual learning. Specifically, the proposed approach uses the multi-dimensional information of the array covariance as input, and finally directly outputs the location information of the near-field sources through the regression structure. The architecture of deep neural network is well designed taking into account the trade-off between the expression ability and computational complexity. In addition, benefiting from the method of generating training data that combines the degree of separation to traverse the spatial location, the proposed approach has a robust performance for different location parameter separation. The simulation results demonstrate that the proposed approach outperforms the existing model-driven methods under various conditions, especially for the adverse scenes with low SNRs, small number of snapshots, or correlated sources.**

***Index Terms*—near-field source localization, deep residual learning, unknown spatially colored noise, regression structure**

I. Introduction

Localization of near-filed sources plays a significant role in various areas such as radar, sonar, wireless communications and speech processing (e.g., [1]–[4]). In past decades, a large number of model-driven algorithms with the near-field assumption have emerged, among which the maximum likelihood estimation (MLE) method [2], the two-dimensional multiple signal classification (2DMUSIC) method [3] and the generalized ESPRIT based method (GESPRIT) [4] are very popular. However, performances of these methods will be significantly deteriorated in practical application.

In comparison, deep learning (data-driven) algorithms can learn the complex nonlinear relationship between the location parameters and the array outputs accurately, and have obvious advantages in localization problems, including no prior knowledge of source statistical assumptions and robustness to practical systems (e.g., [5]–[7]). At present, deep learning has achieved great success in the field of source localization, and its related algorithms can be roughly divided into three categories. The first one converts the localization problem into a classification model by discretizing the spatial location area (e.g., [7]–[10]). However, the resolution of most algorithms is mostly 1° [7], [9] or even larger [8], [10], which is too low a resolution to meet the requirements of high accuracy and super-resolution of multiple overlapping signals. Apart from these classification-based neural networks, estimating the discrete spectrum by the regression neural network, and then combining the peaks to obtain the position parameters is a highly accurate algorithm (e.g., [11] and [12]). Unfortunately, it is very difficult to extend the algorithm to solve the near-field source localization, because it requires estimation of a highly refined two-dimensional spectrum. Finally, some algorithms for directly predicting location parameters using the regression neural network have been studied in [13]–[15]. It should be noted that most of the algorithms mentioned above study the localization in a simple Gaussian noise environment. To the best of our knowledge, the localization of near-field sources in unknown spatially colored noise environment has not been well studied in the literature of array processing.

Motivated by the above discussion, we propose an end-to-end deep neural network for near-field source localization, which is on the basis of deep residual learning [16] and regression model. Distinguished from previous methods which mainly focus on the localization in a simple noise environment, we concentrate on the localization in unknown spatially colored noise environment. The network we designed is trained on multi-channel data, which are formed by complex-valued data of array covariance matrix, and can predict the DOA and range of near-field sources at the same time.

II. Signal Model

Consider K narrowband near-field signals $\{s_k(n)\}$ impinging on a uniform linear array (ULA). The ULA is assumed to have M sensors with spacing d. The received noisy signal $\{x_m(n)\}$ at the m-th sensor can be approximated as

$$x_m(n) = \sum_{k=1}^{K} s_k(n) e^{j\tau_{m,k}} + w_m(n) \tag{1}$$

This work was supported in part by the China Postdoctoral Science Foundation Funded Project under Grant 2019M663727 and the Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2022JQ-640).

for $m = 1, \cdots, M$, where $w_m(n)$ is the additive noise, and $\tau_{m,k}$ is the phase delay due to the time delay between the reference sensor and the m-th sensor for the signal $\{s_k(n)\}$ from the k-th near-field source, which is given by [17]

$$\tau_{m,k} = \frac{2\pi}{\lambda}\left(\sqrt{r_k^2 + m^2d^2 - 2r_k m d sin\theta_k} - r_k\right) \quad (2)$$

where λ denotes the wavelength of signals, θ_k and r_k are the DOA and range of the k-th signal source. For the k-th near-field signal source, r_k is in the Fresnel region (i.e., $r_k \in (0.62(D^3/\lambda)^{1/2}, 2D^2/\lambda)$, where D is the array aperture [18]). Then the received data $\{x_m(n)\}$ can be rewritten compactly by using vector-matrix notation as

$$\begin{aligned} \boldsymbol{x}(n) &= [x_1(n), x_2(n), \cdots, x_M(n)]^T \\ &= \sum_{k=1}^{K} \boldsymbol{a}(\theta_k, r_k)s_k(n) + \boldsymbol{w}(n) \\ &= \boldsymbol{A}\boldsymbol{s}(n) + \boldsymbol{w}(n) \end{aligned} \quad (3)$$

where $(\cdot)^T$ denotes the transposition, $\boldsymbol{s}(n)$ and $\boldsymbol{w}(n)$ are the vectors of the incident signals and the additive noises given by $\boldsymbol{s}(n) = [s_1(n), s_2(n), \cdots, s_K(n)]^T$, and $\boldsymbol{w}(n) = [w_1(n), w_2(n), \cdots, w_M(n)]^T$ respectively, while $\boldsymbol{A}$ is the array response matrix defined by $\boldsymbol{A} \triangleq [\boldsymbol{a}(\theta_1, r_1), \boldsymbol{a}(\theta_2, r_2), \cdots, \boldsymbol{a}(\theta_K, r_K)]$, and $\boldsymbol{a}(\theta_k, r_k)$ is the array steering vectors which can be expressed as

$$\boldsymbol{a}(\theta_k, r_k) = [e^{j\tau_{1,k}}, e^{j\tau_{2,k}}, \cdots, e^{j\tau_{M,k}}]^T. \quad (4)$$

From (3), the array covariance matrix is given by

$$\boldsymbol{R} = E\{\boldsymbol{x}(n)\boldsymbol{x}^H(n)\} = \boldsymbol{A}\boldsymbol{R}_s\boldsymbol{A}^H + \boldsymbol{Q} \quad (5)$$

where $E\{\cdot\}$ represents the statistical expectation, $(\cdot)^T$ denotes the Hermitian transposition, $\boldsymbol{R}_s \triangleq E\{\boldsymbol{s}(n)\boldsymbol{s}^H(n)\}$, and $\boldsymbol{Q} \triangleq E\{\boldsymbol{w}(n)\boldsymbol{w}^H(n)\}$ represent the covariance matrix of the incident source signal and the noise, respectively. In practical applications, the covariance matrix can only be estimated using N snapshots, $\hat{\boldsymbol{R}} = 1/N \sum_{n=1}^{N} \boldsymbol{x}(n)\boldsymbol{x}^H(n)$, which is an unbiased estimator of $\boldsymbol{R}$.

In this letter, we assume that the number of near-field sources K is known, for the number can be estimated by the existing number detection techniques in advance. Cases such as the number of sources being unknown or being estimated incorrectly are beyond the scope of this letter and will be investigated in our future work. We concentrate on the estimation of location parameters $\{\theta_k\}_{k=1}^K$ and $\{r_k\}_{k=1}^K$ of multiple near-field sources from the finite noisy array data $\{\boldsymbol{x}(n)\}_{n=1}^N$. For this purpose, we design an estimation framework based on deep residual learning, which is fed by the array covariance $\hat{\boldsymbol{R}}$, and gives the DOAs and ranges at the output. For simplicity, the proposed method is called *DRN*.

III. Proposed Method

A. Feature Selection and Labeling

While most methods use the feature structure that extracts only part of the information of the covariance matrix [7], [14]. We retain all the information of the covariance matrix, including the real, imaginary, and the angular values of the covariance matrix $\hat{\boldsymbol{R}}$, since it provides superior feature extraction performance in the input as well as achieving satisfactory estimation accuracy [9], [12]. The input data X of the *DRN* is an $M \times M \times 3$ real-valued matrix, whose first and second "channels" are given by $X_{:,:,1} = \mathrm{Re}\{\hat{\boldsymbol{R}}\}$ and $X_{:,:,2} = \mathrm{Im}\{\hat{\boldsymbol{R}}\}$, respectively. Whereas the third "channel" corresponds to phase entries, which is defined as $X_{:,:,3} = \angle\{\hat{\boldsymbol{R}}\}$. We set the label as the location parameter set y (i.e., $y = \{\theta_k, r_k\}_{k=1}^K$) to directly perform regression estimation. Thus, the i-th training sample is composed of pairs in the form $(X_{(i)}, y_{(i)})$ leading to the training datasets $T = \{(X_{(1)}, y_{(1)}), (X_{(2)}, y_{(2)}), \cdots, (X_{(T)}, y_{(T)})\}$.

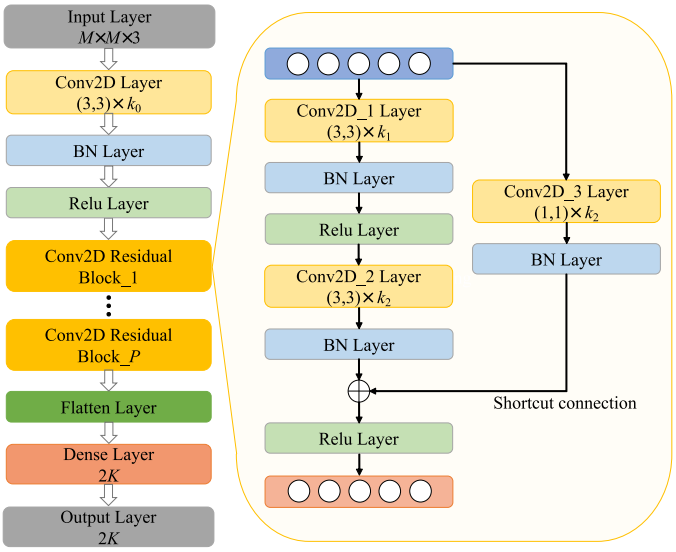

Fig. 1. Architecture of the proposed *DRN*. The network is shown on the left side, and the right side represents the Conv2D residual block structure.

B. The Proposed DRN's Architecture

As shown in Fig. 1, the proposed *DRN* takes the multi-channel information of the covariance matrix as the input. The first three layers after the input are the convolution layer, the batch normalization (BN) layer and the rectified linear unit (ReLU) layer respectively, which can initially extract the location features of the sources. For this convolution layer, small filter of size 3×3 and "same" padding method are used to capture detailed features while avoiding losing edge information. In addition, k_0 represents the number of its convolution kernels. Then, in order to improve the positioning accuracy of the network structure and solve the problem of gradient disappearance caused by deep network, we design a residual architecture consisting of P blocks in the middle layers [16]. The right part of Fig. 1 shows the details of the residual block, where k_1 and k_2 represent kernel numbers in each convolution layer with "same" padding. Therefore, the block output can be denoted as

$$\boldsymbol{Y}_p = \alpha(\mathcal{F}_p(\boldsymbol{Y}_{p-1}) + \mathcal{F}_s(\boldsymbol{Y}_{p-1})) \quad (6)$$

for $p = 1, \cdots, P$, where $\boldsymbol{Y}_0$ represents the input of the entire residual architecture, $\boldsymbol{Y}_p$ represents the output of the p-th residual block, α is the activation function, and $\mathcal{F}_s$ plays the role of the shortcut connection, which is simply used to match the dimension for layer addition. And $\mathcal{F}_p$ can help realize the identity mapping when Y_{p-1} is approximately optimal. Finally, the residual block is followed by a flatten layer, which converts the multi-dimensional data into one-dimensional data. The results after converting are input to the fully connected layer to realize the regression prediction capability of the network.

The goal of our network is to minimize the difference between the predicted parameters and the true parameters. The network uses the mean absolute error (MAE) as the loss function, and its specific form is as follows:

$$\mathcal{L}\left(\hat{y}, y\right) = \frac{1}{T}\sum_{i=1}^{T} \left\|\hat{y}_i\text{-}y_i\right\|_1 \tag{7}$$

where T is the total number of samples, $\hat{y}_i$ and y_i represent the true and predicted parameters for the i-th sample respectively.

IV. Numerical Simulations

A. Simulation Settings

A 9-element uniform linear array with quarter-wavelength inter-element spacing is used to evaluate the proposed *DRN* algorithm. In theory, multi-sources localization can be achieved by adjusting the number of residual blocks and various parameters in our network, while for the sake of simplicity, we take $K = 2$ as an example. Assume that two sources impinging on the ULA come from the spatial scope of $[-30^\circ, 30^\circ]$ from the center of the array within the range of 2λ to 4λ, which are in the Fresnel region of the array aperture $(1.75\lambda < r < 8.0\lambda)$. For the training datasets, a set of angular separations $\{1^\circ, 2^\circ, \cdots, 18^\circ\}$ and a set of range separations $\{0.2\lambda, 0.4\lambda, \cdots, 1.0\lambda\}$ are chosen. For each angular separation $\Delta\varphi$, the DOAs of the first and second source are uniformly generated in the range of $[-30^\circ, 30^\circ - \Delta\varphi]$ and $[-30^\circ + \Delta\varphi, 30^\circ]$ with a step of 1°. Similarly, for each range separation $\Delta\psi$, the range of $[2\lambda, 4\lambda - \Delta\psi]$ and $[2\lambda + \Delta\psi, 4\lambda]$ with a step of 0.2λ are uniformly selected as the ranges of two sources, respectively. 50 examples are generated to sample each combination of DOA and range. Thus, a total of 1.854 million samples are collected in the datasets, where the training and validation data account for 80% and 20%, respectively. The SNR of every sample is randomly distributed between [-10dB, 20dB], or the number of snapshots is randomly generated between [10, 1000]. And the various datasets for the number of snapshots are of the same size as for SNR. Furthermore, the sources spread in a spatially colored noise environment, and its covariance matrix is given by [20]

$$\boldsymbol{Q}_{m,l} = \sigma_n^2 \exp\{-0.5(m-l)^2\} \tag{8}$$

After generating the training datasets, considering the trade-off between the expression ability and complexity of the network, we set two residual blocks with $(k_1,\ k_2)$ chosen as (64, 64) and (32, 32), respectively. We select $k_0 = 128$ at the same time. Then, the Adam optimizer is employed to determine the optimal neural parameters with an initial learning rate of 0.001 [19], and we reduce the learning rate by the factor of 0.2 after each 10 epochs. Additionally we set the maximum number of epochs for training to 30. The samples are used for training with mini-batch size of 128 and the order of the samples is shuffled during each epoch.

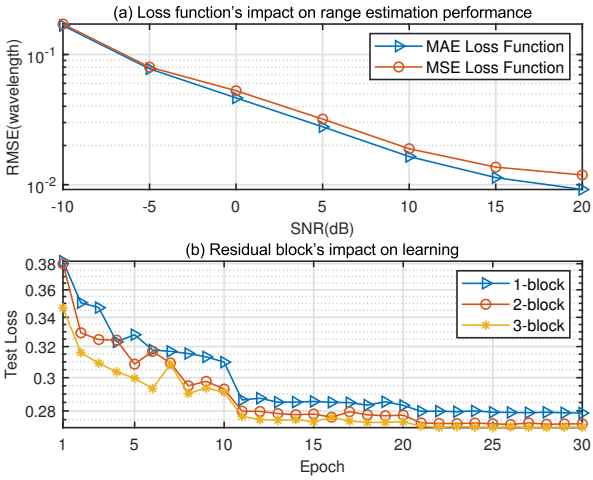

Fig. 2. The impact of (a) Loss function and (b) residual block

Based on the above simulation conditions, we investigate the effect of some settings. As shown in Fig. 2(a), it shows the range estimation performance of different loss functions (i.e. MAE and MSE) under the network structure of two residual blocks. We can see that as the SNR increases, the performance of the MSE-based network is getting worse and worse than that of MAE-based. The reason is that the MAE is more sensitive to small errors than the MSE [15], which can lead to better estimation performance. In addition, the MAE loss of networks based on different numbers of residual blocks are shown in Fig. 2(b). It is seen that the test performance improves with the number of residual blocks increasing from 1 to 3. Note that although the 2-block network performs slightly worse than the 3-block network, it has a lower complexity. Therefore, considering the accuracy and real-time performance of the algorithm, the parameters set in the simulation are appropriate.

In order to verify the superiority of the proposed *DRN* algorithm, the performance of the proposed method is compared with the 2DMUSIC method in [3], the GESPRIT method in [4] and the WLPM method in [21]. For the 2DMUSIC, the angular and range grid spacings $\Delta\theta$ and Δr are fixed at $\Delta\theta = 0.09^\circ$ and $\Delta r = 0.09\lambda$, respectively. The results are all based on 1000 independent trials.

B. DOA and Range Estimation Performance

Figure. 3 shows the RMSEs of the estimated DOAs and ranges in terms of the SNR. In this simulation, two near-field sources are located at $(8.1^\circ, 2.19\lambda)$ and $(13.5^\circ, 2.66\lambda)$ respectively, which are not included in the training datasets. The number of snapshots is fixed at 200, and the SNR is

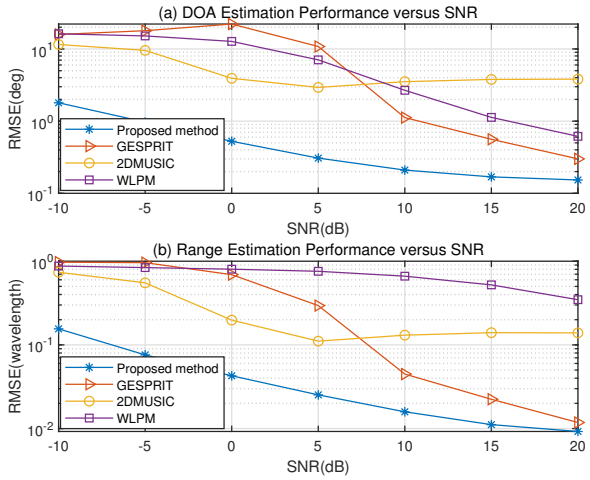

Fig. 3. RMSEs of the estimated (a) DOAs and (b) ranges versus the SNR

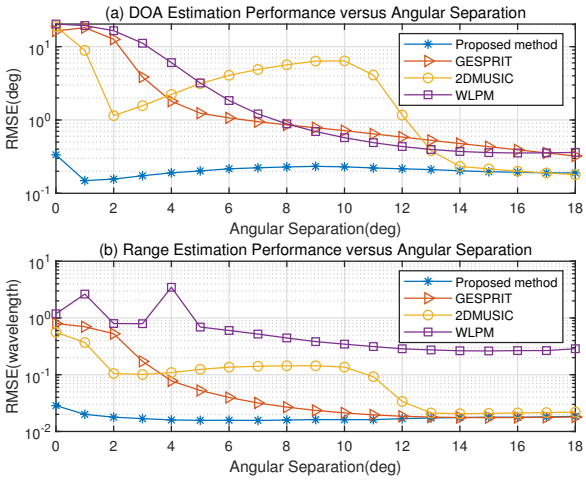

Fig. 5. RMSEs of the estimated (a) DOAs and (b) ranges versus the angular separation

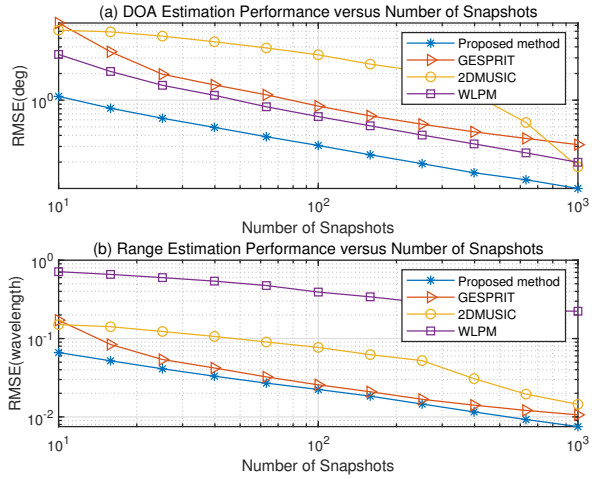

Fig. 4. RMSEs of the estimated (a) DOAs and (b) ranges versus the number of snapshots

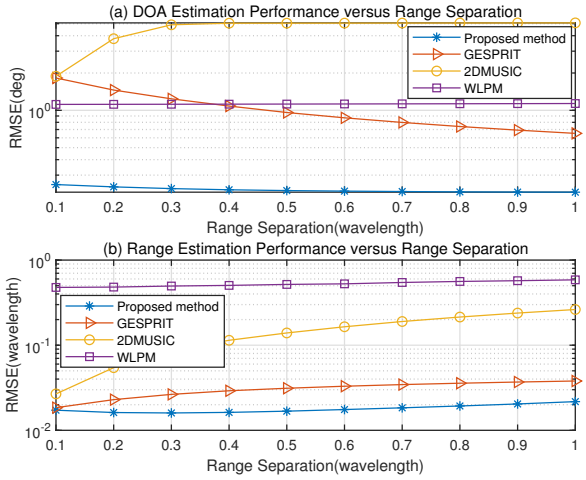

Fig. 6. RMSEs of the estimated (a) DOAs and (b) ranges versus the range separation

varied from -10dB to 20dB. Note that the proposed method outperforms other algorithms, especially at the lower SNR scenario. Furthermore, if the SNR level is used as the evaluation standard, under the same RMSEs, the SNR level required by the proposed method is at least lower than that of other algorithms by nearly 15dB and 5dB in the estimated DOAs and ranges, respectively.

Similarly, Fig. 4 displays the RMSEs of the estimated DOAs and ranges with respect to the number of snapshots. In this experiment, we attempt to estimate the location of two near-field sources at $(8.1^\circ, 2.19\lambda)$ and $(19.8^\circ, 2.66\lambda)$. The SNR is set as 10dB, while the number of snapshots varies from 10 to 1000. It can be seen from the figure that the proposed method performs better than other algorithms regardless of the number of snapshots. In particular, the proposed method does not saturate with the increase in the number of snapshots, which proves the positioning potential of the network. From the above two figures, we can conclude that, the proposed method can effectively solve the problems of model-driven algorithms, which have a poor performance in the environment of low SNRs and small snapshots.

Furthermore, we explore the effect of the angular separation on the estimation performance in Fig. 5, where the SNR is fixed at 10dB, and the number of snapshots is set at 200. In this experiment, two near-field sources located in $(6.4^\circ, 2.19\lambda)$ and $(6.4^\circ+\Delta\theta, 2.66\lambda)$ are considered, where $\Delta\theta$ is varied from 0° to 18° with $\Delta\theta = 1^\circ$. We can see that the robustness of different angular intervals for the proposed method is better than that of other algorithms. It is also worth noting that the proposed method performs better in the small angular intervals scenario compared with other algorithms.

Finally, we show the estimation performance under the different range separations in Fig. 6. The simulation conditions are similar to those in Fig. 5, except that two sources are located at $(6.5^\circ, 2.36\lambda)$ and $(13.7^\circ, 2.36\lambda+\Delta\lambda)$, where $\Delta\lambda$ is varied from 0.1λ to 1.0λ with $\Delta\lambda = 0.1\lambda$. It is observed that the proposed method also has the competitive performance in different range intervals compared with other algorithms.

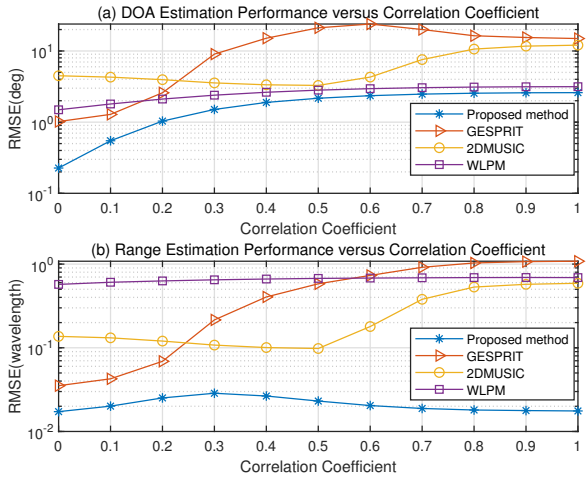

Fig. 7. RMSEs of the estimated (a) DOAs and (b) ranges versus the correlation coefficient

TABLE I
MODEL TRAINING AND INFERENCE TIME

	Proposed method	GESPRIT	2DMUSIC	WLPM
Train time	2h 4min 35s	\	\	\
Test time	0.29ms	2.03ms	227.11ms	0.43ms

Furthermore, like angular separation, range separation also has little effect on the estimation performance of our algorithm.

C. Impact of correlation coefficient

The impact of the source correlation coefficient on the estimation performance is presented in Fig. 7. In particular, we consider two correlated sources located in $(7.2°, 2.36\lambda)$ and $(13.7°, 2.86\lambda)$ with $R_S = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, where ρ is the correlation coefficient, and other simulation conditions are similar to those in Fig. 5. We can see that the proposed method is not only superior in the range estimation, but also is relatively robust for different degrees of correlation, though the DOA estimation performance of the proposed method and the WLPM algorithm is similar in the large correlation coefficient.

D. The averaged training and inference time

To evaluate the computational complexity of the proposed algorithm, we record the running time of various methods in Table I (averaged by 1000 tests). Note that the proposed algorithm has the smallest test time compared to other baselines methods. In addition, although the inference time of the WLPM is close to that of our proposed algorithm, the estimation performance of the WLPM is the worst among all methods. The above results demonstrate the excellent real-time localization capability of our algorithm.

V. CONCLUSION

In this paper, an end-to-end deep neural network based on the residual learning and regression structure is designed for near-field source location in an unknown spatially colored noises. The proposed algorithm makes full use of the multi-dimensional information of the array covariance, and can accurately realize near-field source localization in near real-time by rationally designing the network structure. Numerical results show the proposed approach can yield a substantial improvement in estimation accuracy and robustness in harsh environments over the model-driven methods.

REFERENCES

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, 1996.
[2] J. C. Chen, R. E. Hudson and K. Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1843–1854, 2002.
[3] Y. D. Huang and M. Barkat, "Near-field multiple source localization by passive sensor array," *IEEE Trans. Antennas Propag.*, vol. 39, no. 7, pp. 968–975, 1991.
[4] W. Zhi and M. Y. Chia, "Near-Field Source Localization via Symmetric Subarrays," *IEEE Signal Process. Lett.*, vol. 14, no. 6, pp. 409–412, 2007.
[5] A. Barthelme and W. Utschick, "Doa estimation using neural network-based covariance matrix reconstruction," *IEEE Signal Process. Lett.*, vol. 28, pp. 783–787, 2021.
[6] Y. Guo, Z. Zhang, Y. Huang and P. Zhang, "Doa estimation method based on cascaded neural network for two closely spaced sources," *IEEE Signal Process. Lett.*, vol. 27, pp. 570–574, 2020.
[7] Z. Liu, C. Zhang and P. S. Yu, "Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections," *IEEE Trans. Antennas Propag.*, vol. 66, no. 12, pp. 7315–7327, 2018.
[8] Q. Li, X. Zhang and H. Li, "Online direction of arrival estimation based on deep learning," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 2616–2620, 2018.
[9] G. K. Papageorgiou, M. Sellathurai and Y. C. Eldar, "Deep networks for direction-of-arrival estimation in low snr," *IEEE Trans. Signal Process.*, vol. 69, pp. 3714–3729, 2021.
[10] S. Chakrabarty and E. A. P. Habets, "Multi-speaker doa estimation using deep convolutional networks trained with noise signals," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 1, pp. 8–21, 2019.
[11] L. Wu, Z. Liu and Z. Huang, "Deep convolution network for direction of arrival estimation with sparse prior," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1688–1692, 2019.
[12] A. M. Elbir, "DeepMUSIC: Multiple signal classification via deep learning," *IEEE Sensors Lett.*, vol. 4, no. 4, pp. 1–4, 2020.
[13] O. Bialer, N. Garnett and T. Tirer, "Performance advantages of deep neural networks for angle of arrival estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 3907–3911, 2019.
[14] W. Liu, J. Xin, W. Zuo, J. Li, N. Zheng and A. Sano, "Deep learning based localization of near-field sources with exact spherical wavefront model," in *Proc. IEEE Eur. Signal Process. Conf. (EUSIPCO)*, pp. 1–5, 2019.
[15] Y. Cao, T. Lv, Z. Lin, P. Huang and F. Lin, "Complex resnet aided doa estimation for near-field mimo Systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11139–11151, 2020.
[16] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 770–778, 2016.
[17] A. L. Swindlehurst and T. Kailath, "Passive direction-of-arrival and range estimation for near-field sources," in *Proc. IEEE 4th ASSP Workshop Spec. Est. Mod.*, pp. 123–128, 1988.
[18] J. Tao, L. Liu and Z. Lin, "Joint doa, range, and polarization estimation in the fresnel region," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 4, pp. 2657–2672, 2011.
[19] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations.*, vol. 12, pp. 1–9, 2014.
[20] T. Li and A. Nehorai, "Maximum likelihood direction finding in spatially colored noise fields using sparse sensor arrays," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1048–1062, 2011.
[21] E. Grosicki, K. Abed-Meraim and Y. Hua, "A weighted linear prediction method for near-field source localization," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3651–3660, 2005.