MSEDOA: Enhancing DOA Estimation with Multiscale Squeeze-and-Excitation Networks for Automotive Millimeter-Wave Radar

Tingkai Hu[1], Shuang Sun[1], Zhenyu Wu[1], Chuandong Li[1], Choujun Zhan[2], Hailing Xiong[1*], Zhen Luo[1*]
[1] College of Electronic and Information Engineering, Southwest University, 400700, Chongqing, China,
[2] School of Computer, South China Normal University, 510631, Guangdong, China,
Email: {htkstudy@163.com, suns524@163.com, zheny_wu@163.com, cdli@swu.edu.cn, zchoujun2@gmail.com, xionghl@swu.edu.cn, zhenluo@swu.edu.cn}

2024 IEEE International Symposium on Product Compliance Engineering - Asia (ISPCE-ASIA) | 979-8-3315-2300-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISPCE-ASIA64773.2024.10756221

***Abstract*—Direction-of-Arrival (DoA) estimation is a critical task in automotive millimeter-wave radar systems, enabling the localization of targets and the implementation of essential safety functions. Traditional DoA methods struggle under low SNR and limited snapshot conditions. To address these challenges, we introduce MSEDOA, a novel deep learning method that processes raw IQ signals with a neural network featuring multiscale feature extraction, channel attention mechanisms, and Squeeze-and-Excitation Residual Networks. Experimental results demonstrate that the proposed model outperforms traditional DoA estimation methods, providing precise estimates in challenging environments.**

***Index Terms*—Direction-of-arrival (DOA) estimation, Convolutional Neural Network, Multi-label classification, Millimeter-Wave Radar**

I. Introduction

With the continuous advancements in advanced driver-assistance systems (ADAS) and autonomous driving technologies, automotive radar systems are becoming increasingly critical. Direction of Arrival (DOA) estimation is a fundamental aspect of radar target localization, directly influencing the safety of key functions such as collision avoidance, lane departure warning, and emergency braking [1]. Traditional DOA estimation methods rely on beamforming techniques that extend the Fourier transform to the spatial domain. However, these methods restricted by the Rayleigh limit, which hinders angular super-resolution [2]. To overcome these limitations, researchers developed various subspace-based DOA estimation techniques, with Multiple Signal Classification (MUSIC) [3] and Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [4] as the most prominent. While these methods surpass the Rayleigh limit, they depend heavily on pre-established models and struggle to provide stable and accurate estimates under low signal-to-noise ratio (SNR) and limited snapshot conditions. This hinders their practical application in automotive millimeter-wave radar systems.

In recent years, data-driven deep learning techniques have gained significant traction as a robust solution for DOA estimation [5]. These approaches utilize supervised learning on extensive datasets to autonomously discover and extract pertinent features, demonstrating enhanced robustness and adaptability across diverse application scenarios. Current deep learning methodologies for DOA estimation can be broadly classified into two categories: regression-based methods and multi-label classification strategies [6]. However, regression-based methods are inherently limited in practical applications due to their dependence on prior knowledge of the number of targets and their inability to adapt to dynamic multi-target scenarios. In contrast, multi-label classification approaches, which simultaneously predict the DOAs of multiple targets, offer superior flexibility and robustness, making them more suitable for complex environments, such as automotive radar systems. Despite the inherent flexibility of multi-label classification methods in DOA estimation, they exhibit poor convergence under conditions of limited snapshots, leading to instability in the estimation results. Additionally, deep learning techniques that rely on the covariance matrix of received signals as input impose a significant computational burden on automotive radar systems, a challenge that is particularly pronounced in real-time applications [7], [8].

To address the challenges in DOA estimation, we introduce a novel End-to-End method that directly inputs raw in-phase and quadrature (IQ) signals into a deep neural network. Our approach enhances the SE-ResNet architecture with multi-scale feature extraction and channel attention mechanisms, specifically designed to improve accuracy under low SNR and limited snapshot conditions. Despite the constraint of approximately 10 snapshots, our experimental results show that the proposed method significantly outperforms traditional techniques, delivering precise DOA estimates even in challenging environments. The code is publicly available at: https://github.com/Armorhtk/MSEDOA.

II. Problem Formulation

In the context of DoA estimation using a Uniform Linear Array (ULA), consider an array consisting of M elements. The array is tasked with detecting K narrowband sources, each emitting a signal from distinct directions $\Theta = [\theta_1, \theta_2, \ldots, \theta_K]$. The inter-element spacing d of the ULA is set to $d = \lambda/2$,

* Corresponding author: (Hailing Xiong, Zhen Luo)

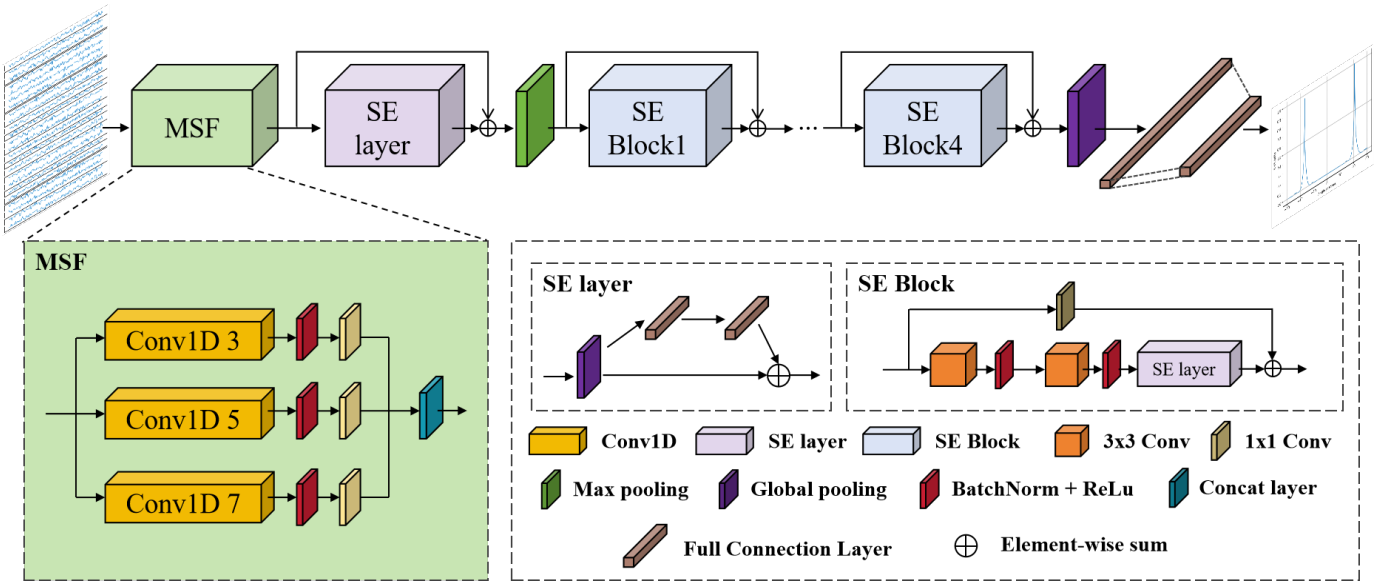

Fig. 1. The MSEDOA network architecture.

where λ denotes the wavelength of the incident signals. The received signal at the array across different time instants is

$$x(t) = \mathbf{A}(\Theta)s(t) + n(t), \quad t = 1, 2, \ldots, L, \tag{1}$$

where $x(t) \in \mathbb{C}^{M\times 1}$ is the vector of signals received by the array at time t, capturing the composite effect of all M array elements. The matrix $\mathbf{A}(\Theta) \in \mathbb{C}^{M\times K}$ represents the array manifold, encapsulating the spatial response of the array to the incoming signals from the K distinct directions. Here, $s(t) \in \mathbb{C}^{K\times 1}$ denotes the vector of source signals at time t, while $n(t) \in \mathbb{C}^{M\times 1}$ represents additive noise. The array response matrix $\mathbf{A}(\Theta)$ formulated as

$$\mathbf{A}(\Theta) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \ldots, \mathbf{a}(\theta_K)], \tag{2}$$

with each steering vector $\mathbf{a}(\theta_k)$ for $k = 1, 2, \ldots, K$ defined by

$$\mathbf{a}(\theta_k) = \left[1, e^{-j\frac{2\pi d}{\lambda}\sin(\theta_k)}, \ldots, e^{-j\frac{2\pi d}{\lambda}(M-1)\sin(\theta_k)}\right]^T. \tag{3}$$

Given that the array collects signals over discrete time instants, referred to as snapshots L, the resulting received signal matrix is $\mathbf{X} = [x(1), x(2), \ldots, x(L)] \in \mathbb{C}^{M\times L}$. The goal of DOA estimation is to infer the directions $\theta_1, \theta_2, \ldots, \theta_K$ from the received signal $\mathbf{X}$. We recasting DoA estimation as a multi-label classification task. Specifically, the Field of View (FOV) of the automotive radar, spanning the interval $[-g, g]$, is discretized into a series of grid points with a fixed resolution of δ degrees, thus forming a discrete set $\mathcal{G}$ is expressed as

$$\mathcal{G} = \{-g, -g+\delta, ..., -\delta, 0°, \delta, ..., g-\delta, g\}, \tag{4}$$

Here, the DoA estimation is treated as a classification problem where the number of classes equals the number of grid points, $G = 2g/\delta + 1$. For multiple signal sources, the task reduces to predicting the set of most probable angles, $\hat{\Theta} = \left\{\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_K\right\}$, by maximizing the posterior probability given by

$$\hat{\Theta} = \arg\max_{\Theta\in\mathcal{G}^K} P(\Theta|\mathbf{X}) \tag{5}$$

where $P(\Theta|\mathbf{X})$ denotes the posterior probability of angles Θ given the given the received signal $\mathbf{X}$.

III. Proposed Methodology

A. Designed Network Architecture

We propose a novel deep learning framework MSEDOA, as illustrated in Fig. 1, which integrates multiscale feature extraction, channel attention mechanisms, and an SE-ResNet to capture the complex nonlinear relationships between raw signals and their corresponding DOAs. Specifically, the complex-values received signal $\mathbf{X} \in \mathbb{C}^{M\times L}$ is decomposed into the real $\mathbf{X}_{Re}$ and imaginary $\mathbf{X}_{Im}$ components. These components are concatenated along the time axis L to form a new matrix $\mathbf{X}_{IQ} = \begin{pmatrix}\mathbf{X}_{Re}\\ \mathbf{X}_{Im}\end{pmatrix} \in \mathbb{R}^{2M\times L}$, which serves as the input to the neural network.

To address the limitations of traditional CNNs in capturing subtle differences in signals from different array elements [8], we designed a Multi-Scale Feature Extraction (MSF) module. This module employs three parallel convolutional paths to extract multi-scale features, and then concatenates their outputs to form a comprehensive feature representation:

$$F_{MSF} = \text{Cat}[C_3, C_5, C_7], \tag{6}$$

where Cat$[\cdot]$ denotes the concatenation operation, C_i is the feature map from the i-th convolutional path:

$$C_i = \text{Conv1D}_{1\times 1}(\text{ReLU}(\text{BN}(\text{Conv1D}_{2M\times i}(\mathbf{X}_{IQ})))) \tag{7}$$

where $\text{Conv1D}_{2M\times i}$ represents a convolutional layer with a kernel size of $(2M, i)$, $\text{Conv1D}_{1\times 1}$ denotes pointwise convolutional layer, BN denotes Batch Normalization, and ReLU is the activation function.

We introduce the channel attention mechanism of the SE layer [9] to enhance feature extraction. The SE layer performs global average pooling on F_{MSF} to generate a channel descriptor $\mathbf{z} \in \mathbb{R}^{3C}$, where C is the number of channels in the feature map, and learns the channel weights $\mathbf{w}$ through fully connected layers.

$$\mathbf{w} = \text{ReLU}(\mathbf{W}_2(\mathbf{W}_1(\mathbf{z}))), \tag{8}$$

where $\mathbf{W}_1 \in \mathbb{R}^{3C\times\frac{3C}{r}}$ and $\mathbf{W}_2 \in \mathbb{R}^{\frac{3C}{r}\times 3C}$ are the weight matrices of the fully connected layers, and r is the reduction ratio. Finally, the SE layer applies the learned weights $\mathbf{w}$ to the feature map through an element-wise multiplication, yielding the weighted feature map F_{SE}:

$$F_{SE} = \mathbf{w} \odot F_{MSF}, \tag{9}$$

where $\odot$ denotes the Hadamard product. The weighted feature map F_{SE} is subsequently processed by a customized SE-ResNet [9], adapted from its traditional image processing form to enhance signal processing for DOA estimation. Incorporating SE blocks, the SE-ResNet applies channel-wise attention to refine feature maps, boosting model sensitivity and accuracy. This enables it to generate a probability distribution over potential DOAs, resulting in the prediction:

$$\hat{\Theta} = \text{Sigmoid}(\text{SE-ResNet}(F_{SE})). \tag{10}$$

The binary cross-entropy (BCE) loss function is employed to minimize the error between the predicted and true labels. The BCE loss is defined as:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}\left[\theta_{ij}\log(\hat{\theta}_{ij}) + (1-\theta_{ij})\log(1-\hat{\theta}_{ij})\right], \tag{11}$$

where N denotes the number of training samples, K is the number of DOA labels, $\theta_{ij} \in \Theta$ is the ground truth label, and $\hat{\theta}_{ij} \in \hat{\Theta}$ is the predicted label.

B. Training and Infehrence

In automotive millimeter-wave radar scenarios, the antenna array is typically composed of 3 transmitting and 4 receiving antennas, forming a 12-element uniform linear array (ULA) with a field of view (FOV) from -60° to 60°. This angular range was discretized into 121 angle labels with 1° resolution. To strike a balance between real-time processing and computational efficiency, we limited the number of radar snapshots to 10. Our proposed network model was trained via supervised learning on a simulation dataset covering seven SNR levels, from 0 dB to 30 dB, with 200,000 samples per level. The total dataset of 1.4 million samples was split, with 90% for training and 10% for validation. The model was trained over 100 epochs using the AdamW optimizer, with a learning rate of 0.0001, adjusted via Cosine Annealing. During inference, test samples were processed to generate a pseudo-angle spectrum, with the top-k local maxima identifying the estimated DoAs. The implementation utilized PyTorch and was trained on an NVIDIA GeForce RTX 4090 GPU. To evaluate the accuracy of DoA estimation, the root mean square error (RMSE) was used as the performance metric:

$$RMSE(\theta, \hat{\theta}) = \min_{\hat{\theta}\in\mathcal{G}} \sqrt{\frac{1}{SK}\sum_{s=1}^{S}\sum_{k=1}^{K}(\theta_k - \hat{\theta}_k)^2} \tag{12}$$

where S is the number of samples, K is the number of sources, θ_k is the true angle of arrival, and $\hat{\theta}_k$ is the estimated angle of arrival.

IV. Performance Analysis

We conduct two experiments to evaluate the effectiveness and performance of the proposed MSEDOA model. In all of our experiments, we compare our method against several state-of-the-art traditional methods, including traditional algorithms like Conventional Beamforming (CBF), MUSIC, ESPRIT.

A. DoA Estimation Performance

In the first set of experiments, we fixed the angle between the two incoming sources at 4.7°, with the angle of the first incoming source varying from -60° to 55° in steps of 1°. Consequently, the angle of the second incoming source varies from -55.3° to 59.7° in steps of 1°, resulting in a total of 116 test samples. The DOA estimation results for these samples under SNR = 12 dB using CBF, MUSIC, ESPRIT, and MSEDOA are shown in Fig. 2. (a)-(d), with the corresponding angle errors shown in Fig. 2. (e)-(h). The experimental results demonstrate that MSEDOA outperforms traditional methods across all test samples, with an error range of [-1.2°, 1.0°], while CBF, MUSIC, and ESPRIT exhibit error ranges of [-20°, 26°], [-33°, 21°], and [-5.3°, 4.2°], respectively. Particularly in low-sample scenarios, CBF struggles to distinguish between targets, while MSEDOA demonstrates superior robustness and a smaller error range.

B. Robustness to Noise

In the second set of experiments, we assessed DOA estimation performance across varying SNR levels, calculating the RMSE over 100 simulations per SNR. The first source was sampled uniformly from $[-60°, 53°]$ for a realistic distribution, with a 6.1° separation. The second source was generated within $[-53.9°, 59.1°]$. Fig. 3 presents the relationship between RMSE and SNR on a logarithmic scale for different algorithms. As SNR increases, RMSE decreases across all methods. In the mid-to-low SNR range (-5 to 15 dB), MSEDOA shows a clear advantage. However, at higher SNRs (20-30 dB), MSEDOA's performance slightly lags behind MUSIC and ESPRIT, with RMSE values being about 0.16 and 0.34 higher, respectively. This minor difference may result from the 1° grid resolution, potentially causing grid mismatch [10]. Nonetheless, MSEDOA maintains RMSE around 0.5 at high SNR, excelling particularly in low-to-medium SNR conditions.

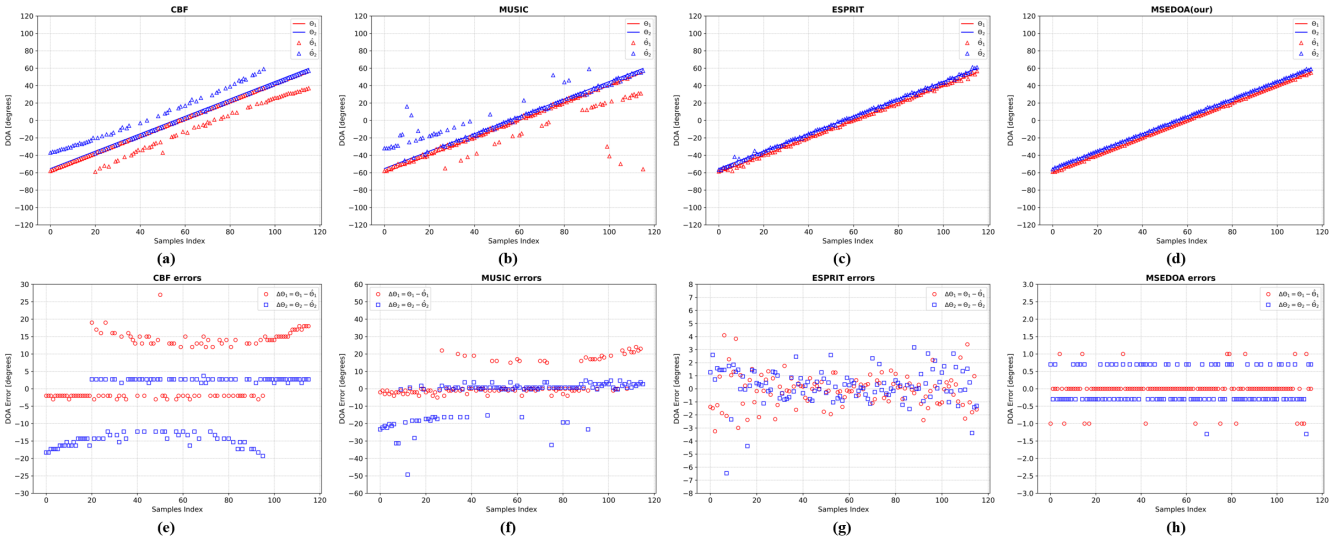

Fig. 2. The DOA estimation performance is evaluated for angles θ_1, θ_2 within the range of $[-60°, 60°]$ at 12 dB SNR with interval $\Delta\theta = 4.7°$. The results are presented for the following methods: (a) CBF, (b) MUSIC, (c) ESPRIT, (d) MSEDOA. Additionally, the DOA estimation errors for each method are shown in: (e) CBF, (f) MUSIC, (g) ESPRIT, (h) MSEDOA.

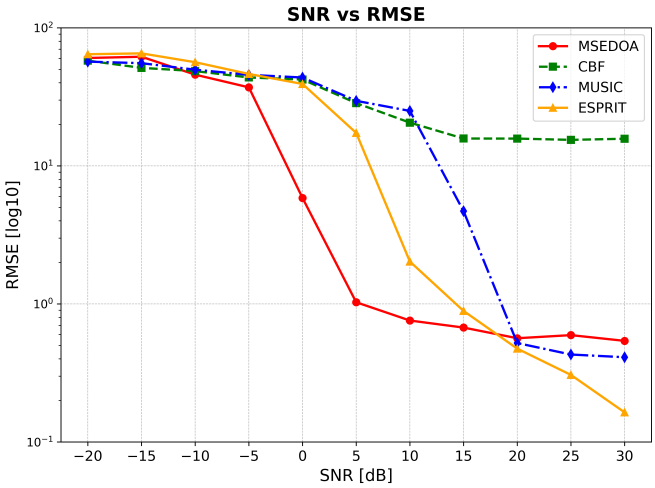

Fig. 3. The RMSE of DOA estimation for different algorithms under various SNR conditions.

V. CONCLUSION

We present MSEDOA, a deep learning-based method for DoA estimation designed for automotive millimeter-wave radar systems. MSEDOA leverages a Squeeze-and-Excitation Residual Network with multiscale feature extraction and channel attention to deliver robust and precise DoA estimates under challenging conditions. Experimental results demonstrate that MSEDOA outperforms conventional techniques, providing reliable estimates in complex environments. Future work will explore integrating classification and regression strategies for super-resolution DoA estimation with unknown source counts and assess its feasibility for real-time deployment in automotive radar.

ACKNOWLEDGMENT

This work was financially supported by the Fundamental Research Funds for the Central Universities of China (SWU2009107) and and Project of Chongqing Natural Science Foundation under Grant (CSTB2022NSCQ-MSX0990) and Project of Guangdong Provincial Education Department (2022ZDZX1040) and Educational Science Planning Topic of Guangdong Province (2022GXJK382).

REFERENCES

[1] Wan, Liangtian, et al. "Deep learning based autonomous vehicle super resolution DOA estimation for safety driving." IEEE Transactions on Intelligent Transportation Systems 22.7 (2020): 4301-4315.

[2] Merkofer, Julian P., et al. "DA-MUSIC: Data-driven DoA estimation via deep augmented MUSIC algorithm." IEEE Transactions on Vehicular Technology (2023).

[3] Schmidt R O. A signal subspace approach to multiple emitter location and spectral estimation[M]. Stanford University, 1982.

[4] Roy, Richard, A. Paulraj, and Thomas Kailath. "Estimation of signal parameters via rotational invariance techniques-ESPRIT." MILCOM 1986-IEEE Military Communications Conference: Communications-Computers: Teamed for the 90's. Vol. 3. IEEE, 1986.

[5] Liu, Zhang-Meng, Chenwei Zhang, and S. Yu Philip. "Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections." IEEE Transactions on Antennas and Propagation 66.12 (2018): 7315-7327.

[6] Zheng, Shilian, et al. "Deep learning-based DOA estimation." IEEE Transactions on Cognitive Communications and Networking (2024): 819-835.

[7] Papageorgiou, Georgios K., Mathini Sellathurai, and Yonina C. Eldar. "Deep networks for direction-of-arrival estimation in low SNR." IEEE Transactions on Signal Processing 69 (2021): 3714-3729.

[8] Barthelme, Andreas, and Wolfgang Utschick. "DoA estimation using neural network-based covariance matrix reconstruction." IEEE Signal Processing Letters 28 (2021): 783-787.

[9] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[10] Wu, Xiaohuan, et al. "A gridless DOA estimation method based on convolutional neural network with Toeplitz prior." IEEE Signal Processing Letters 29 (2022): 1247-1251.