International Journal of Antennas and Propagation

Efficient Processing and Learning with DNNS for Multidimensional Signal Processing

Lead Guest Editor: Xianpeng Wang
Guest Editors: Qi Liu and Bin Zhang

Efficient Processing and Learning with DNNS for Multidimensional Signal Processing

International Journal of Antennas and Propagation

Efficient Processing and Learning with DNNS for Multidimensional Signal Processing

Lead Guest Editor: Xianpeng Wang
Guest Editors: Qi Liu and Bin Zhang

Chief Editor

Associate Editors

Academic Editors

Contents

Hindawi
International Journal of Antennas and Propagation
Volume 2022, Article ID 5072666, 8 pages
https://doi.org/10.1155/2022/5072666

Research Article

Identification of Pneumonia in Chest X-Ray Image Based on Transformer

Yongjun Ma [1,2] **and Wei Lv** [2]

[1]*City University of Macau, Macau 999078, China*
[2]*Zhuhai College of Science and Technology, Zhuhai 519040, China*

Correspondence should be addressed to Wei Lv; luwei@zcst.edu.cn

Received 22 May 2022; Accepted 4 July 2022; Published 1 August 2022

Academic Editor: Xianpeng Wang

The research of application models based on traditional convolutional neural networks has gradually entered the bottleneck period of performance improvement, and the improvement of chest X-ray image models has gradually become a difficult problem in the study. In this paper, the Swin Transformer is introduced into the application model of pneumonia recognition in chest X-ray images, and it is optimized according to the characteristics of chest X-ray images. The experimental results based on the model in this paper are compared with those of the model built with the traditional convolutional neural network as the backbone network, and the accuracy of the model is proved to be greatly improved. After the comparison experiments on two different datasets, the experimental results show that the accuracy of the model in this paper improves from 76.3% to 87.3% and from 92.8% to 97.2%, respectively. The experiments show that the accuracy of image enhancement based on the features of chest X-ray images in this model will be higher than the accuracy without image enhancement. In the experiments of this paper, the identification decision factors in the chest X-ray images were extracted by grad-cam combined with a transformer to find the corresponding approximate lesion regions.

1. Introduction

Pneumonia is a common and dangerous disease that is mainly caused by viruses, bacteria, or fungi. If left untreated, its mortality rate is high. According to the literature, pneumonia is one of the ten deadliest diseases in the United States and has a higher mortality rate in developing countries [1]. Chest X-ray imaging (hereafter referred to as CXR) is widely used in general routine examinations because it is not only low cost, but also its radiation is less harmful than computed tomography. Relevant papers indicate that the mean effective radiation dose per exam of CXR is about 0.04 ± 0.19 msv, while the principle of computed tomography is that X-rays penetrate the human body for multiple times for tomography, so the mean effective radiation dose per exam can reach 1.09 ± 1.11 msv, about 25 times that of CXR [2]. Doctors often use CXR as an important aid in diagnosing pneumonia. In today's world, artificial intelligence is playing a huge role in the transformation of science, industry, and society, and its techniques are widely used in medical image processing. The application and improvement of artificial intelligence in CXR to identify pneumonia can assist doctors in making the correct diagnosis, help them speed up the diagnosis, reduce the proportion of missed and misdiagnosis, and be of great importance in saving lives.

Since the explosive development of deep learning in 2012, amazing achievements have been made in the research and application of artificial intelligence. Compared with other machine learning algorithms, deep learning algorithms can rely on their own learning methods for feature extraction. Deep learning has achieved great success in many fields such as computer vision, natural language processing, and big data analysis. In addition, it has become a mainstream approach to machine learning and has achieved record-breaking results in various competitions in artificial

intelligence. Deep learning can be traced back to AlexNet in 2012 [3]. The accuracy of this convolutional neural network algorithm, which won the championship in the famous international image classification competition ImageNet, has been improved by more than ten points compared with other algorithms in the past. It uses many methods for the first time, uses ReLu as a nonlinear activation function, uses dropout to prevent overfitting, uses data enhancement, and so on. After AlexNet, there have been many excellent convolutional neural networks. VGGNet is a convolutional neural network developed by the Visual Geometry Group of Oxford University on the basis of AlexNet [4]. The improvement of VGGNet is that it uses a smaller convolution kernel and a deeper network structure, which enhances the feature learning ability of the convolutional neural network, which also verifies the advantages of small convolution kernels and can improve network performance by deepening the network structure. In addition, VGGNet uses the multi-Scale method to train and predict, reducing the occurrence of model overfitting and improving the prediction accuracy. Inspired by the Network in Network theory, the concept of the Inception module emerged, that is, a convolutional layer contains multiple convolutional operations of different sizes. A typical convolutional neural network with Inception is GoogLeNet [5]. In addition, two auxiliary classifiers are added to the middle layer of GoogLeNet to strengthen supervision information and alleviate the problem of gradient disappearance. In simple theory, the deeper the network level, the more complex feature extraction can be carried out, so better results should be obtained. But in fact, it was found in the experiment that there was a problem of degradation after the network was deepened to a certain extent, that is, after a large increase in the network depth, the accuracy began to saturate and degrade. The main reason is that when the data are transmitted in a deep network, the gradient becomes smaller and gradually disappears, making it impossible to perform the backpropagation algorithm, so it is difficult for the network to train and find a good parameter after deepening the level to a certain extent. For this reason, He et al. proposed a residual unit with a "short-circuit connection" structure to solve this degradation problem, instead of directly connecting each layer. ResNet is modified on the basis of VGGNet, and it uses residual units [6]. Compared with VGGNet, it adds a "short-circuit connection" mechanism between every two layers, which gives an implementation idea for building a much deeper network. In addition to the ways of deepening the network such as ResNet and widening the network such as GoogLeNet to improve the effect, there are also multiplexing schemes, the typical representative is DenseNet, which can achieve better results while achieving fewer parameters [7]. Other scholars have proposed EfficientNet, which is based on an artificial neural network to obtain the optimal composite coefficient of network depth, network width, and image resolution [8].

With the research and development of the convolutional neural backbone network, it has also promoted the improvement of medical image processing model capabilities. As early as 2017, Wang's team built medical image processing models based on the classic convolutional neural network AlexNet [3], VGGNet [4], GoogLeNet [5], and ResNet [6] in deep learning, and tested and compared them on the public CXR dataset named Chest X-ray. Through their research and experiments, it was proved that resnet50 has the best effect of disease identification in CXR compared with other backbone networks [9]. Yao et al. optimized the convolutional neural network DenseNet [7], and the model they proposed was tested on the Chest X-ray dataset and achieved ideal results [10]. Later, Rajpurkar and other scholars built a 121-layer network based on the convolutional neural network DenseNet and used the weighted cross entropy as the loss function to propose the chexnet model for medical image classification. The model was tested with a higher accuracy score than four human medical imaging experts correctly judged [11]. Later, many scholars further improved the models based on the convolutional neural network according to the features of CXR [12–20]. But accuracy of models began to encounter bottlenecks, and there are still some unsolved or imperfect problems in the current models.

In this paper, a new model scheme based on the backbone network of the new transformer and optimized according to the features of CXR will be proposed, and it can greatly improve the accuracy of identification of pneumonia in CXR. The image enhancement and parameter optimization scheme are designed based on the features of CXR, and the lesion area is found to the greatest extent from the decision factors of transform. Experiments in this paper show that under the same circumstances, the model for identification of pneumonia in CXR based on the transformer backbone network has higher accuracy than that based on the traditional convolutional neural backbone network. The image enhancement scheme for CXR in this model will play a positive role in improving the accuracy rate of the model.

Through the research in this paper, the bottleneck problem of improving the accuracy of the model for identification of pneumonia in CXR based on the traditional convolutional neural network can be overcome, and better results can be achieved. To sum up, the research in this paper has its value both theoretically and practically in the identification of pneumonia and even more diseases in CXR.

2. Proposed Scheme

In order to better compare the difference between the model for identification of pneumonia in CXR based on transformer backbone network and the models based on traditional convolutional neural backbone network, the experiment in this paper was done on the Chest X-ray data set [9] and CXR images (pneumonia) data set [21], because a large number of scholars used these data sets when testing the models based on the traditional convolutional neural backbone network. It should be noted that the former data set comes from the National Institutes of health, and the latter data set comes from Guangzhou Women and Children's Medical Center, and these data sets are publicly available for free use in scientific research.

The chest X-ray data set is a data set of more than 100000 anonymous chest X-ray images released by the National Institutes of health to the scientific community. The copyright of this data set is announced on https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community, "The release will allow researchers across the country and around the world to freely access the datasets and increase their ability to teach computers how to detect and diagnose disease." The number of samples in the Chest X-ray data set [9] is shown in Table 1.

In the CXR images (pneumonia) data set [21], there are 5,856 anonymous chest X-ray images from Guangzhou Women and Children's Medical Center with "license CC BY 4.0". The text of the CC BY 4.0 was retrieved from https://creativecommons.org/licenses/by/4.0/, and for more information, view the full license text at https://creativecommons.org/licenses/by/4.0/legalcode. This data set is divided into two categories: pneumonia and normal. The number of samples in the CXR images (pneumonia) data set is shown in Table 2.

The model used in the experiment in this paper is based on the Swin Transformer backbone network [22] and optimizes the CXR accordingly. The basic steps are as follows (Figure 1): in addition to the obvious feature of a gray-scale image, CXR generally has its own characteristics such as low brightness, poor contrast, and high noise, so the first step is to improve the brightness, contrast, and suppress noise of the image according to the features of CXR. The second step is to obtain the best parameters of the model, the images are divided into a training set and validation set, normalize the images in the training set, and after random scaling, clipping, and flipping send them to the transformer network and fully connected network for training to obtain the best parameters of this model. The purpose of normalizing images is to facilitate the speedy contingency of the network. The purpose of random scaling, clipping, and flipping is to make the model not "see" the same image twice during training, so it has better generalization ability. A transformer network is used for feature extraction and a fully connected network is used for classification. In the third step, the images in the validation set are scaled and sent to the transformer network with trained parameters for feature extraction, and then send to the fully connected network with trained parameters for classification. The fourth step is to extract the decision factor from the Transformer network. The last step is to map the decision factor to the original image to output the lesion area.

The first was to do experiments with the model in this paper on the Chest X-ray data set, and then the experimental results are compared with the experimental results of models based on AlexNet [3], GoogLeNet [4], VGGNet16 [5], and ResNet50 [6] from the Wang's team on the same data set, the experimental results of model based on DenseNet [8] from Yao and other scholars, and the experimental results of model based on DenseNet121 [8] from Rajpurkar et al.

In order to verify the effectiveness of image enhancement according to the features of CXR, a comparison experiment between enhanced and nonenhanced images in the preprocessing with the model based on the Transformer backbone network was carried out, and the two experimental results were compared. As shown in Figure 2, it is a contrast map for CXR enhancement, in which the left side is before enhancement, and the right side is after enhancement.

Table 1: The number of samples in Chest X-ray data set.

Focus of infection	Samples
Atelectasis	5789
Cardiomegaly	1010
Effusion	6331
Infiltration	10317
Mass	6046
Nodule	1971
Pneumonia	1062
Pneumothorax	2793
Normal	84312

Data set source: https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community.

Table 2: The number of samples in CXR images (pneumonia) data set.

Focus of infection	Samples
Pneumonia	4273
Normal	1583

Data set source: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia.

In order to further verify the versatility of this model for the identification of pneumonia in CXR based on the transformer backbone network, a comparative experiment on the CXR Images (Pneumonia) dataset [21] was carried out and compared its result with the experimental results of other models on the same dataset.

Finally, in the experiment, the decision factors of the identification in the chest X-ray image from the Swin transformer were extracted, and with the Grad-CAM [23] they were superimposed on the original image to perform the discriminative output of the lesion area.

3. Experimental Result

The accuracy of the experiment results with the model in this paper on the Chest X-ray data set reached 87.3%. From the comparison in Table 3, it can be seen that the model based on the Swin Transformer backbone network and optimized for CXR is obviously better than other models based on traditional convolutional neural network.

In the experiment to verify the effectiveness of preprocessing of image enhancement according to the features of CXR, this paper collects the accuracy data of the model based on the Swin Transformer backbone network during the training process. As shown in Figure 3, in order to show the details more clearly, the figure draws a line graph of the accuracy from batches 32 to 128 on the first epoch without image enhancement and with image enhancement, in which the blue dotted line is no enhancement, and the orange

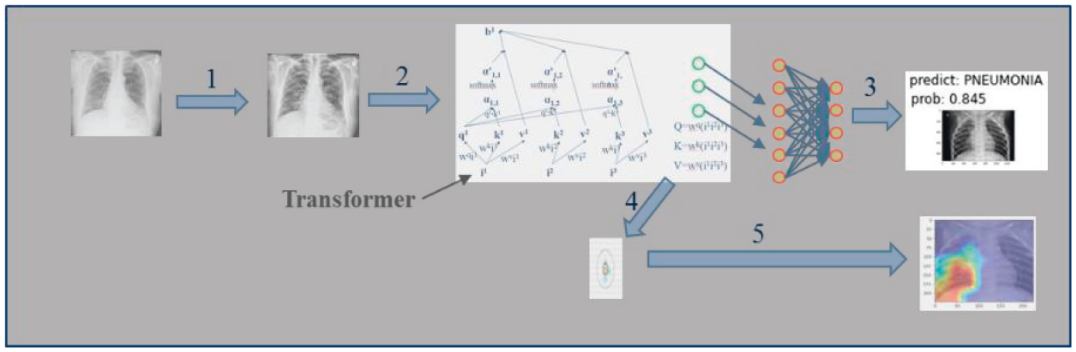
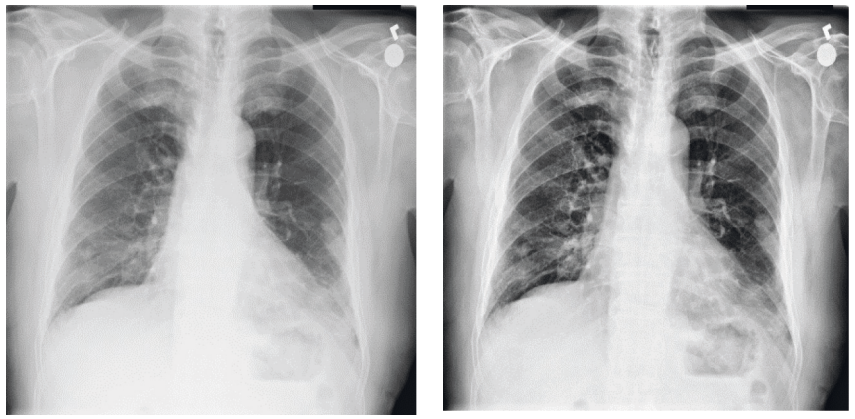


Figure 1: Schematic of a transformer network for CXR.



(a) (b)

Figure 2: Image enhancement (a) is before enhancement and (b) is after enhancement.

Table 3: Comparison of the models based on different backbone network on data set 1.

	Backbone network	Validate-accuracy (%)
Wang, et al. [9]	AlexNet	54.9
Wang et al. [9]	GoogLeNet	59.9
Wang et al. [9]	VGGNet-16	51.0
Wang et al. [9]	ResNet-50	63.3
Yao et al. [10]	DenseNet	71.3
Rajpurkar et al. [11]	DenseNet-121	76.3
This paper	SwinTransformer	87.3

dashed line is enhancement. As can be seen from Figure 3, the accuracy of image enhancement according to the features of CXR will be higher than that without image enhancement under the same circumstances.

In the comparative experiment on the CXR Images (Pneumonia) data set, the model based on the Swin Transformer backbone network and optimized for CXR in this paper achieved the best accuracy of 97.2% after only five epochs of training, which is much higher than the accuracy rate of 92.8% from the model based on the convolutional neural network proposed by Kermany's team [13]. It is also higher than the competition results in the Kaggle on the CXR Images (Pneumonia) data set (https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia/discussion/). The comparison data of accuracy from different models are shown in Table 4.

In Figures 4 and 5, the cross-entropy loss and accuracy during the training process using the model in this paper are shown. The top figure shows the change of the cross-entropy loss on epochs (the blue dotted line is from the data of the training set, and the orange dashed line is from the data of the validation set), and the bottom figure shows the change of the accuracy on epochs (the blue dotted line is from data of the training set, and the orange dashed line is from the data of the validation set).

What is the reason for the higher accuracy on the validation set than on the corresponding training set (Figure 5)? Because in order to enhance the generalization ability of the model, the data of the training set are randomly scaled, cropped, and flipped before entering the transformer network to extract features, while the data of the validation set has not undergone this transformation.

Before the transformation of the Softmax function and entering the fully connected classification network, the decision factors of the identification in chest X-ray image from the transformer are extracted. In our experiments, the decision factors are from the norm layer following the transformer backbone network, which can be obtained by back-propagating the result value of the latter classification network. The reverse derivation according to the Grad-CAM

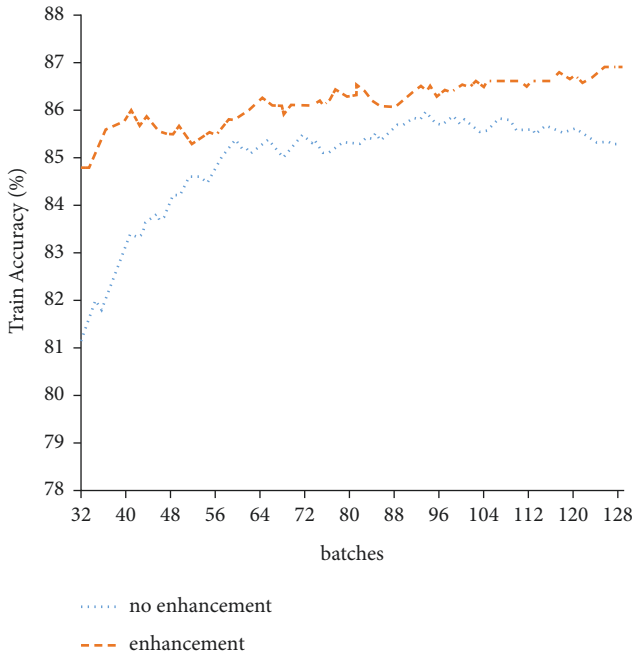


Figure 3: Comparison accuracy between no enhancement and enhancement on 1st epoch in the training process.

Table 4: Comparison of the models based on different backbone network on data set 2.

	Backbone network	Validate-accuracy (%)
Sharma et al. [13]	—	92.8
Grzegorz on Kaggle*	ResNet	94
This paper	SwinTransformer	97.2

*https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia/discussion/313883.

algorithm is superimposed with the original image to form a heat map as shown in Figure 6. The area with high color temperature is the area that plays an important role in the formation of network discrimination, so the corresponding lesion area can be obtained.

4. Discussion and Analysis

After several years of research, the research of the application models based on the traditional convolutional neural backbone networks such as AlexNet, VGGNet, GoogLeNet, ResNet, DenseNet, and EfficientNet has gradually entered the bottleneck period of network performance improvement, and the improvement effect in the application models research of CXR also gradually becomes less obvious. At this time, a new backbone network is urgently needed to solve this problem.

Vaswani et al. from the Google team proposed the transformer backbone network in 2017 [24]. Compared with the traditional Recurrent Neural Network [25], Transformer has many advantages such as infinite memory length in theory and parallel operation. The theory of the self-attention algorithm is the basis of the transformer (1).

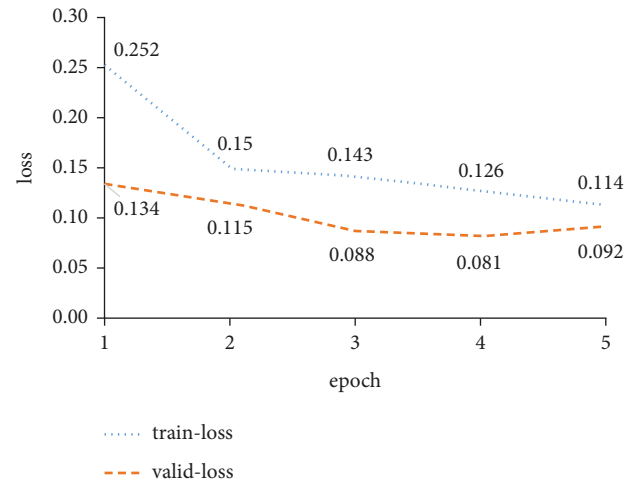


Figure 4: Cross-entropy loss in the training process and validation process.

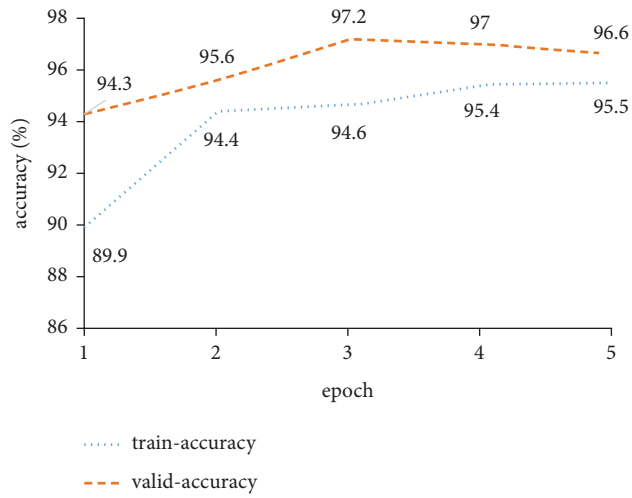


Figure 5: Accuracy in the training process and validation process.

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \tag{1}$$

In transformer multihead, self-attention extended from the self-attention algorithm is used, and it is split by linear mapping according to the number of headers and is usually divided equally (2).

$$\begin{aligned}\text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O,\\ \text{where head}_i &= \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right),\\ \text{Attention}(Q, K, V) &= \text{Softmax}\left(\frac{QK^T}{\sqrt{.d_k}}\right)V.\end{aligned} \tag{2}$$

The transformer was originally used for natural language processing. At the 2020 International Conference on Computer Vision and Pattern Recognition (CVPR), the

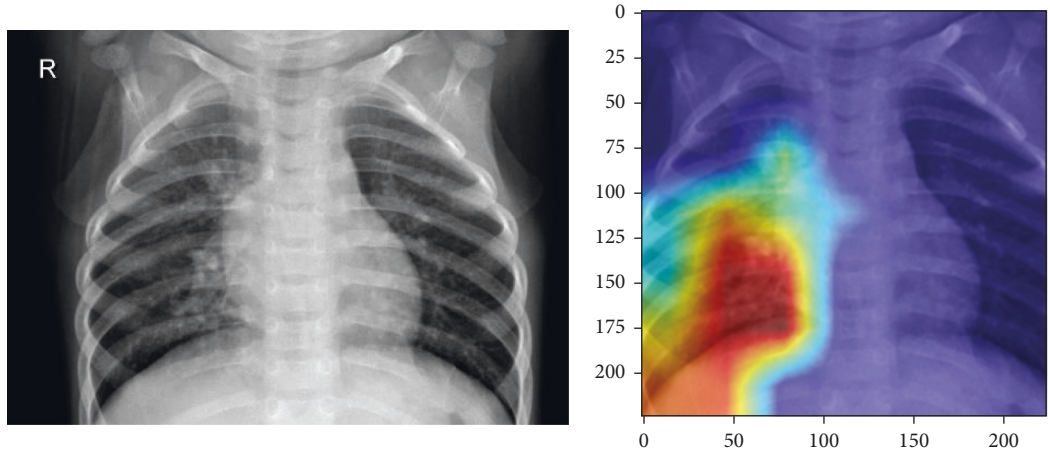


Figure 6: Heat map for lesion area.

Google team proposed a scheme to apply transformers to the field of computer vision and achieved good results [26]. The core of the transformer comes from the self-attention algorithm, and the self-attention algorithm and the convolution algorithm are very closely related, and the latter can be considered as a subset of the former [27]. As the scale of the data set increases, the performance of the transformer backbone network will exceed the traditional convolutional neural backbone networks, and large-scale training data can encourage the transformer to learn the more translation equivariance and locality than possessed by the convolutional neural networks. In 2021, the Swin Transformer proposed by Microsoft Research Asia has become a bright spot. It overcomes the bottleneck of the traditional convolutional neural backbone network to a certain extent and further improves the accuracy [22]. The experimental comparison results [22] of its effect on the ImageNet-1K data set are as follows shown in Table 5. The difference from convolutional neural networks such as ResNet is that the Swin Transformer no longer uses traditional convolution kernels in feature extraction, and the core at each level is window multihead self-attention and shifted window multihead self-attention. Window multihead self-attention is the multihead self-attention matrix operation performed inside the window. The advantage is to reduce the amount of computation, but the disadvantage is that information interaction between windows is not possible. The shifted window multihead self-attention is designed to overcome this shortcoming, and it can realize the information interaction between different windows by shifting the window position.

Since the Swin transformer backbone network has not been proposed for a long time, its application in various fields has not been sufficiently studied, and few studies have been conducted to optimize the model for the application of CXR images. Regarding the application model in the CXR field, the model proposed in this paper is no longer based on the traditional convolutional neural backbone network, but the Swin transformer backbone network is introduced to build the model. The experimental results on the two CXR data sets prove that the accuracy of the model based on the transformer backbone network is higher than that of the model based on the traditional convolutional neural backbone network model on the same data set, and the effect is improved obviously, which can overcome the existing bottleneck of improving the accuracy of model based on the convolutional neural backbone network.

Table 5: Comparison of different backbone network.

	Throughput (images/s)	Accuracy (%)
EfficeintNet-B3	732.1	81.6
EfficeintNet-B4	349.4	82.9
Swin-T	755.2	81.3
Swin-S	436.9	83.0
Swin-B	278.1	83.5

It is necessary to highlight image details and suppress noise because CXR is usually characterized by low brightness, low contrast, and large noise. In the histogram of CXR, the area with the highest pixel distribution is usually the background, which is a nonconcern area, so this part can be peaked. If the values in the histogram are evenly distributed, it means that the distribution on each gray level is balanced, and the contrast is the best at this time, that is, the image is generally clear. Therefore, a certain degree of equalization processing on each gray level in the histogram of CXR is helpful to the subsequent processing. It can be seen from the data chart of the experimental results (Figure 3) that this processing method is effective for the model based on the transformer backbone network.

The concept of class activation mapping (CAM) originated from the interpretability research of deep neural networks [28] and was later introduced into application research by some scholars. On this basis, Selvaraju et al. proposed gradient weighted activation mapping (Grad-CAM) [29].

In the experiment of CXR in this paper, gradient weighted activation mapping is combined into the transformer, so that by extracting the decision factors of the identification in CXR from the transformer, the heat map through reverse derivation is superimposed on the original

image, so the corresponding approximate lesions area can be found.

When the new type of pneumonia caused by the 2019 novel coronavirus appeared, the study of the corresponding model was carried out. Narin et al. used ResNet50, ResNet101, ResNet152, Inception V3, and Inception-ResNet-V2 five models for identification of the new type of pneumonia and compared them. Their experimental results also show that ResNet50 achieves better results [30]. In similar cases, it is estimated that the model with the Transformer backbone network will be better.

5. Conclusion

Pneumonia is a disease with a high mortality rate. Chest X-ray imaging is widely used in the routine examination of pneumonia. CXR as an important adjunct to the diagnosis of pneumonia can diagnose pneumonia quickly and accurately. Machine learning methods based on deep learning have been effective in chest X-ray imaging. In this paper, the Swin Transformer is applied to the application model of CXR image recognition and analysis, and the model is optimized accordingly according to the characteristics of CXR. The experimental results show that the model outperforms the model based on the traditional convolutional neural backbone network.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Guangdong Basic and Applied Basic Research Foundation (Grant no. 2021A1515310003).

References

[1] S. L. Murphy, J. Xu, K. D. Kochanek, S. C. Curtin, and E. Arias, "Deaths: final data for 2015. National vital statistics reports: from the centers for disease control and prevention, national center for health statistics," *National Vital Statistics System*, vol. 66, pp. 1–75, 2017.

[2] K. Y. Yoon, S. H. Joo, K. M. Joon, and L. M. Jung, "Comparison of radiation dose from X-ray, CT, and PET/CT in paediatric patients with neuroblastoma using a dose monitoring program," *Diagnostic and Interventional Radiology*, vol. 22, no. 4, pp. 390–394, 2016.

[3] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, 2012.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Science*, 2014.

[5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, and A. Rabinovich, "Going Deeper with Convolutions," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, Boston, MA, June 2015.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[7] G. Huang, Z. Liu, V. Laurens, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society. IEEE Computer Society, Honolulu, HI, USA, July 2017.

[8] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," 2019, https://arxiv.org/pdf/1905.11946.pdf.11946.

[9] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE. Proceedings of Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3462–3471, Honolulu, HI, USA, December 2017.

[10] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, "Learning to Diagnose from Scratch by Exploiting Dependencies Among Labels," 2017, https://arxiv.org/abs/1710.10501.

[11] P. Rajpurkar, J. Irvin, K. Zhu et al., "Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning," 2016, https://arxiv.org/abs/1711.05225.

[12] S. A. Khoiriyah, A. Basofi, and A. Fariza, "Convolutional neural network for automatic pneumonia detection in chest radiography," in *Proceedings of the 2020 International Electronics Symposium (IES)*, Surabaya, Indonesia, September 2020.

[13] H. Sharma, J. S. Jain, P. Bansal, and S. Gupta, "Feature extraction and classification of chest X-ray images using CNN to detect pneumonia," in *Proceedings of the 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence) IEEE.*, Noida, India, January 2020.

[14] I. Masad, A. Alqudah, A. M. Alqudah, and S. Almashaqbeh, "A hybrid deep learning approach towards building an intelligent system for pneumonia detection in chest X-ray images," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 6, Article ID 5530, 2021.

[15] Z. Parveen, S. Adinarayana, R. Aamani, S. Santoshi, and S. Tulasi, "Efficient pneumonia detection in chest xray images using convolution neural network," *International Journal of All Research Education and Scientific Methods*, vol. 9, no. 7, pp. 2900–2905, 2021.

[16] R. K. Gupta, Y. Sahu, N. Kunhare, A. Gupta, and D. Prakash, "Deep learning based mathematical model for feature extraction to detect corona virus disease using chest X-ray images," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 29, no. 06, pp. 921–947, 2021.

[17] R. Alsharif, Y. Alissa, A. M. Alqudah, I. A. Qasmieh, A. M. Wan, and H. Alquran, "PneumoniaNet: Automated Detection and Classification of Pediatric Pneumonia Using Chest X-ray Images and CNN Approach," *Electronics*, vol. 10, no. 23, pp. 1–13, 2021.

[18] W. Xue, "Classification of pulmonary lesions based on CNN and chest X-ray images," *Journal of Physics: Conference Series*, vol. 1952, no. 2, Article ID 022025, 2021.

[19] M. J. Alam, S. N. Ali, and M. Z. Hasan, "A Robust CNN Framework with Dual Feedback Feature Accumulation for Detecting Pneumonia Opacity from Chest X-ray Images," in *Proceedings of the 2020 11th International Conference on Electrical and Computer Engineering (ICECE)*, Dhaka, Bangladesh, December 2020.

[20] C. Han, T. Okamoto, K Takeuchi et al., "Tips and Tricks to Improve CNN-Based Chest X-ray Diagnosis: A Survey," 2021, https://arxiv.org/abs/2106.00997.

[21] D. S. Kermany, M. Goldbaum, W. Cai et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[22] Z. Liu, Y. Lin, Y. Cao et al., "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows," 2021, https://arxiv.org/abs/2103.14030, Article ID 14030.

[23] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-cam: why did you say that?," 2016, https://arxiv.org/abs/1611.07450.

[24] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention Is All You Need," 2017.

[25] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," *Computer Science*, 2015.

[26] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, and N. Houlsby, "An Image Is worth 16x16 Words: Transformers for Image Recognition at Scale," 2020, https://arxiv.org/abs/2010.11929.

[27] J. B. Cordonnier, A. Loukas, and M. Jaggi, "On the Relationship between Self-Attention and Convolutional Layers," 2020, https://arxiv.org/abs/1911.03584.

[28] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, June 2016.

[29] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, and D. Batra, "Grad-CAM: visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, 2020.

[30] A. Narin, C. Kaya, and Z. Pamuk, "Automatic Detection of Coronavirus Disease (Covid-19) Using X-ray Images and Deep Convolutional Neural Networks," *Pattern Anal Appl*, vol. 24, no. 3, 2020.

Hindawi
International Journal of Antennas and Propagation
Volume 2022, Article ID 2996862, 10 pages
https://doi.org/10.1155/2022/2996862

Research Article

HRNet Encoder and Dual-Branch Decoder Framework-Based Scene Text Recognition Model

Meiling Li, Xiumei Li, Junmei Sun, and Yujin Dong

Hangzhou Normal University, School of Information Science and Technology, Hangzhou 311121, China

Correspondence should be addressed to Xiumei Li; xiumei_li@hotmail.com

Received 11 February 2022; Accepted 28 May 2022; Published 15 June 2022

Academic Editor: Giuseppe Castaldi

Scene text recognition (STR) is designed to automatically recognize the text content in natural scenes. Different from regular document text, text in natural scenes has the characteristics of irregular shapes, complex background, and distorted and blurred contents, which makes STR challenging. To solve the problems of STR for distorted, blurred, and low-resolution texts in natural scenes, this paper proposes a HRNet encoder and dual-branch decoder framework-based STR model. The model mainly consists of an encoder module and a dual-branch decoder module composed of a super-resolution branch and a recognition branch in parallel. In the encoder module, the HRNet is adopted to realize the cross-parallel aggregation representation with multiple resolutions during feature extraction and then outputs four kinds of feature maps with different resolutions. Moreover, the supervised attention module is used to strengthen the learning of the important feature information. In the decoder module, the dual-branch structure is adopted, in which the super-resolution branch takes the feature maps with the highest resolution obtained in the encoder module as input and restores images by upsampling through transposed convolution. The four kinds of feature maps with different resolutions are fused through independent transposed convolution layers for multiscale fusion in the recognition branch and then inputted into the attention-based decoder for text recognition. To improve the accuracy of text recognition, the feature extraction effect of the encoder module is together supervised by the super-resolution branch loss and the recognition branch loss. In addition, the super-resolution branch is only used for training and is abandoned during testing to reduce the complexity of the model. The proposed model is trained on Synth90K and SynthText datasets and tested on seven natural scene datasets. Compared with classical models such as ASTER, TextSR, and SCGAN, the recognition accuracy of the proposed model is improved and better recognition results can be achieved on irregular and blurred datasets such as IC15, SVTP, and CUTE80.

1. Introduction

Natural scene text refers to the text content in natural situations, such as billboards and road signs. Due to the high diversity of text in orientation, shape, and blurring, scene text recognition (STR), which is designed to automatically recognize the text content in natural scene images, is challenging [1]. With the development of deep learning, the deep learning-based STR can obtain good text recognition results and has become a research highlight in the field of document analysis and recognition [2]. Moreover, the deep learning-based STR is an essential research technology, which can be employed in many computer vision applications, such as image retrieval, autonomous driving, and handwriting recognition [3–6].

Early STR models are usually based on temporal feature classification, such as the convolutional recurrent neural network (CRNN) [7]. CRNN uses convolutional neural networks to extract visual features and uses recurrent neural networks to learn the bidirectional dependence of feature sequences and predict the probability of character sequences. Then, the predicted probabilities of character sequences are transcribed into text character sequences according to the predefined transformation mode in the transcription layer. However, the setting of the transcription layer in the CRNN requires that the feature sequences of

image and text are aligned with each other, which is not beneficial to predict the text sequences with spatial dependence. The model based on the encoder-decoder framework [8] can avoid the alignment problem by training to predict the corresponding relationship between any two sequences. Generally, the visual features of an image are extracted using an encoder and then are converted into a fixed-length intermediate semantic feature sequence by means of a recurrent neural network. Then, the intermediate semantic feature sequence is decoded into a text character sequence through a decoder. The models based on the encoder-decoder framework have achieved higher performance than the earlier models based on temporal feature classification and provide an effective baseline model for further research [9].

However, scene text images are often disturbed by complex background and text distortion, which often cause the information loss of the visual features extracted by the encoder and then lead to the decoder's inaccurate recognition of the target sequences in the noisy decoding time steps. To alleviate the above problems, the ASTER model [10] based on the encoder-decoder framework is proposed and the thin-plate-spline (TPS) [11] is introduced to improve the text distortion, so that the encoder can extract more sufficient visual features from the rectified images. Based on the sequence model, the visual features are converted into the textual features and finally the attention mechanism is introduced to decode the textual features. When confronted with blurred images and low resolution, the models based on the encoder-decoder framework also suffer from low-recognition accuracy, which prompts researchers to introduce auxiliary networks to improve the resolution of scene text images and learn more accurate text information. Inspired by the success of multitask learning, the super-resolution network SRGAN [12] is used as a preprocessing method in text super-resolution (TextSR) [13] to restore the low-resolution image with the corresponding super-resolution image and then input it into the STR model to improve the recognition effect. The super-resolution network RCAN [14] is used as an auxiliary network in PlugNet [15] to update the parameters of the encoder, to achieve better recognition results. Similarly, in the text super-resolution network (TSRN) [16], a sequential-residual block is proposed to extract the sequential information of the scene text images and accomplish the super-resolution task. However, the super-resolution networks adopted by the above three models have complex structure and a large number of parameters, which increases the complexity of the model.

Recently, a parallel high-resolution network (HRNet) is proposed [17]. Instead of restoring high-resolution representations from low-resolution representations, HRNet maintains high-resolution representations at any time and performs multiscale fusion across parallel convolutions to enhance high-resolution representations, therefore greatly improving the detection and segmentation difficulties caused by image blurring and low resolution. Due to its advantages, HRNet is introduced into the STR task to effectively alleviate the text recognition difficulties caused by the scene text images with blurry and low resolution.

In this paper, a HRNet encoder and dual-branch decoder framework-based STR model is proposed to recognize the distorted and blurred text with low resolution. This model innovatively introduces a HRNet encoder to extract visual features and adopts dual-branch decoder structure composed of a super-resolution branch and a recognition branch following the encoder. The feature maps with highest resolution are inputted into the super-resolution branch for upsampling and image recovery. The feature maps with multiple resolutions are fused at multiscale in the recognition branch to accomplish the transformation of feature sequences and obtain the recognized text. The loss of the super-resolution branch and the loss of the recognition branch are together propagated back to enhance the feature extraction effect of the encoder module, therefore improving the performance of text recognition. The main contributions of this paper are as follows:

(1) The HRNet is innovatively used for feature extraction in STR and also performs as a super-resolution network. Moreover, the HRNet encoder provides effective feature maps for the super-resolution branch, therefore decreasing the model complexity caused by the introduction of an auxiliary super-resolution network, such as TextSR. Experiments on several natural scene datasets verify the effectiveness of the proposed model.

(2) In the encoder module, four kinds of feature maps with different resolutions are generated at the end of the HRNet. By using the supervised attention module (SAM) on the feature maps with the highest resolution, the important features are enhanced and the features with a small amount of information are suppressed. In the decoder module, the feature maps enhanced by SAM are upsampled through transposed convolution (Trans Conv2D) in the super-resolution branch to restore the super-resolution images. The other three feature maps with lower resolution are upsampled through independent transposed convolution layers (Independent Trans Conv2D Layers) in the recognition branch. The feature maps with the same size as the feature maps with the highest resolution are generated, and multiscale fusion is implemented to enhance the representation of the feature maps with multiple resolutions.

(3) The parallel dual-branch structure is adopted. In the training stage, the super-resolution branch and the recognition branch are adopted together to strengthen the feature extraction effect of the encoder module and constantly update the effective parameters in the model, so that the recognition branch can recognize the text on the more-effective feature maps. In the testing stage, the model is simplified, the super-resolution branch is abandoned, and only the recognition branch is used to

obtain the recognition results, which is helpful to reduce the model complexity.

2. Related Work

The ASTER model [10] based on the encoder-decoder framework introduces a rectification network TPS to alleviate the recognition difficulties caused by arbitrary arrangement and text distortion. The residual network (ResNet) is used in the encoder to encode the rectified images and obtain the visual feature sequences. Based on bidirectional long-short-term memory (Bi-LSTM), the visual feature sequences are converted into textual feature sequences and the text content is obtained by the decoder with attention mechanism. However, when the scene text images are severely distorted and the rectification is insufficient, the visual features extracted from the encoder are not sufficient. Based on the ASTER model, the ESIR model via iterative rectifications is proposed [18], which iteratively removes text distortion as driven by better recognition performance. However, for the images with few distortions, the rectified images obtained through the rectification network will be over rectified, the resolution of images will be reduced, and the rectified images will even lose some edge information. Meanwhile, to reduce the recognition errors caused by insufficient rectification, the SAR model is proposed [19], which abandons the rectification network and adopts a two-dimensional attention module to process the two-dimensional visual feature maps from the encoder. Then, the text characters are located and recognized by considering the regional information of each position in the feature maps.

In the encoder-decoder framework, the attention mechanisms are usually used in the decoder of the models. Most of the attention-based methods usually suffer from serious alignment problems due to its recurrent alignment operation, where the alignment relies on historical decoding results. Cheng et al. [20] put forward the concept of attention drift. The attention mechanism is easily affected by some problems, such as image blurring and complex background, and cannot get accurate alignment between feature maps and the targets of input images. The DAN model is proposed to alleviate the problem of attention drift [21]. The attention maps are obtained through the convolution alignment module based on visual features from the encoder. Moreover, a decoupled decoder is used to make the final prediction by jointly using the visual feature maps and attention maps. In addition, the RobustScanner model is proposed to alleviate the problem by introducing two branches [22]. Specifically, the position enhancement branch is specially designed to improve the ability of position encoding in the decoder. The hybrid branch is the traditional decoder with attention mechanism. The outputs of the two branches are combined through the dynamic fusion module and connected to an elementwise gate mechanism in the channel dimension. By the selection of features, the RobustScanner can adaptively adjust the importance of contextual information and positional information to obtain better performance of text recognition. However, the above models are complex in terms of the model structure.

Since the transformer [23] has achieved remarkable achievements in the tasks of natural language processing, researchers are beginning to explore its application in the field of STR. The 2DOCR model is proposed [24], which uses the transformer to decode twice at the end of the encoder. The second decoder is fine-tuned and optimized on the basis of the result of the first decoding, which can effectively improve the recognition performance. Moreover, the Bi-STET model is proposed [25] to solve the problem of information loss in the process of converting visual features to textual features. After extracting visual features from the ResNet, the encoder of the transformer is used to enhance the visual features, to better integrate visual information and text information. Besides, text recognition on the decoder of the transformer also has better recognition effect.

However, integrating the transformer into the STR models greatly increases the number of model parameters and training time. Therefore, researchers try to introduce auxiliary network modules with a relatively low number of parameters to improve the recognition accuracy when facing the problems of image blurring and low resolution. Wang proposes the TextSR model [13], which introduces a content-aware text super-resolution network SRGAN to restore low-resolution images with super-resolution images under the guidance of adversarial loss and then uses the ASTER model to identify the text content of super-resolution images. To solve the problems of low brightness in images and text occlusion, the SPIN model [26] proposes the structure preserving network (SPN) and the auxiliary inner-offset network (AIN), respectively. Specifically, SPN adjusts the intensity value between pixel points based on the structure-preserving transformation to alleviate the problem of low brightness in images. Based on the theory of offset from geometric transformation, the AIN introduces colour offsets to distinguish the colour intensity, to alleviate the problems of text occlusion and shadow. To solve the problem of complex background, the SCGAN model is put forward [27], which outputs binary images through the generator and inputs into the attention-based decoder to generate the attention feature maps. After the fusion of binary images and attention feature maps, the recognized texts are outputted to the discriminator and compared with the ground truth texts. The loss is propagated back to optimize the network parameters of the generator and to improve the recognition performance. The SEED model is proposed [28] to alleviate the problems of uneven illumination and incomplete characters. Based on the ASTER model, the SEED model innovatively introduces the pretrained language model FastText in the stage of visual features conversion to textual features. Moreover, the cosine embedding loss is calculated with semantic information and word embedding of target texts from the FastText, to supervise the effect of feature extraction in the encoder, to obtain more comprehensive text information and better recognition results.

3. HRNet Encoder and Dual-Branch Decoder Framework-Based Scene Text Recognition Model

This paper proposes a STR model based on HRNet encoder and dual-branch decoder framework, as shown in Figure 1. A single scene text image is taken as the input, and after the process of TPS network and Gaussian blur, the encoder module and dual-branch decoder module are adopted, in which the super-resolution image and recognized text are outputted by the super-resolution branch and recognition branch, respectively. Specifically, HRNet is adopted as the feature extraction network in the encoder module to output the feature maps with multiple resolutions. The SAM is acted on the feature maps with the highest resolution to strengthen the learning of important feature information. The input of the super-resolution branch is the feature maps with the highest resolution enhanced by the SAM, and the super-resolution image is generated by upsampling through Trans Conv2D. The input of the recognition branch is the feature maps with multiple resolutions. Through the Independent Trans Conv2D Layers, the lower resolution feature maps are expanded, so that the final multiscale feature maps can be fused in the channel dimension. The attention-based decoder is used to decode the fused feature maps to obtain text recognition results. In the parallel dual-branch decoder module, the super-resolution branch and the recognition branch together enhance the feature extraction effect of the encoder module and then improve the effect of STR. In the testing stage, the super-resolution branch is abandoned to simplify the model and reduce the complexity of the model.

3.1. Encoder Module. The encoder module of the model is shown in Figure 2, which innovatively adopts HRNet as the feature extraction network and maintains a high-resolution representation throughout the whole process. A high-resolution subnet is taken as the first stage, and multiresolution subnets from high to low are added one by one to form more stages. The multiresolution subnets are connected in parallel, and the information is repeatedly exchanged during the whole process to perform the multifeature fusion. At the end of the encoder module, the SAM is used to strengthen the learning of important feature information of the feature maps with the highest resolution outputted by the HRNet encoder. The feature with less information is suppressed by using the attention mask, so that the encoder module can transfer the most effective learned features to the super-resolution branch and the recognition branch. Finally, different connection operations are adopted according to the different purposes of the super-resolution branch and the recognition branch. The feature maps with the highest resolution enhanced by the SAM are inputted to the super-resolution branch, and four kinds of feature maps with different resolutions are inputted to the recognition branch.

The SAM is constituted by a series of convolution operation and sigmoid activation function, as shown in Figure 3. The feature maps with the highest resolution are added to the input image after the 1×1 convolution operation; that is, the feature maps are supervised by the input image. Then, the attention maps are obtained by the activation function and then are acted on the feature maps by weighted summation. In this way, important features can be enhanced and features with less information can be suppressed.

3.2. Super-Resolution Branch. The super-resolution branch of the model employs the Trans Conv2D for upsampling on feature maps with the highest resolution enhanced by the SAM, to restore the super-resolution images. No extra super-resolution network is introduced, the super-resolution branch is directly connected to the encoder module, and a simple upsampling recovery operation is adopted, so the super-resolution branch of the proposed model is more dependent on the feature maps outputted from the encoder module. The effect of feature extraction of the HRNet encoder is strengthened through the supervision of the super-resolution branch. Meanwhile, the super-resolution branch is only used in the training stage and is abandoned in the testing stage, which helps to reduce the model complexity. The Trans Conv2D is composed of 3×3 transposed convolution operation, BatchNorm layer, and ReLu layer. The average absolute error loss L_{sr} of the restored super-resolution image and the original image is calculated, as shown in

$$L_{\mathrm{sr}} = \frac{1}{W \times L} \sum_{i=1}^{W} \sum_{j=1}^{L} \left\| O^{i,j} - I^{i,j} \right\|, \tag{1}$$

where W and L represent the width and length of the image, respectively, O represents the super-resolution image restored by the super-resolution branch, and I represents the original scene text image.

3.3. Recognition Branch. The recognition branch of this proposed model consists of a multiscale fusion structure and an attention-based decoder structure. Specifically, in the multiscale fusion structure, in contrast to expanding the size of the feature maps by bilinear interpolation, the Independent Trans Conv2D Layer is used on all low-resolution feature maps, to obtain the feature maps with the same size as the feature maps with the highest resolution. The resolutions of feature maps decrease from the top to bottom, and the number of input channels and output channels of a single Independent Trans Conv2D Layer is determined according to the size of corresponding feature maps. Furthermore, the multiscale fusion is carried out in the channel dimension through the splicing operation, as shown in Figure 4. Then, by employing the channel attention mechanism [29], the weights on different channels of the multiscale feature maps are calculated and important channels of the feature maps are adaptively selected to help the network obtain more effective information.

After obtaining the structure of multiscale fusion, the attention-based decoder is connected to achieve complete text recognition. To realize effective sequence conversion from visual features to textual features, the multiscale feature maps are processed by a 3×3 basic convolution module in

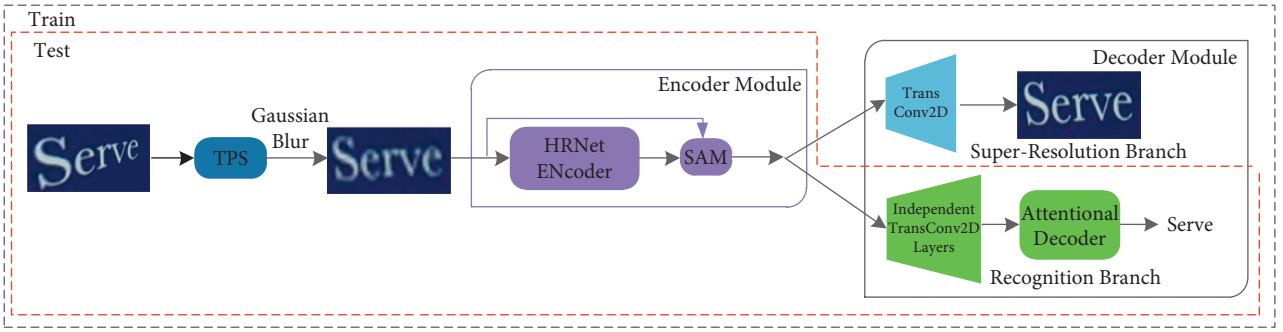


Figure 1: Structure of the proposed model.

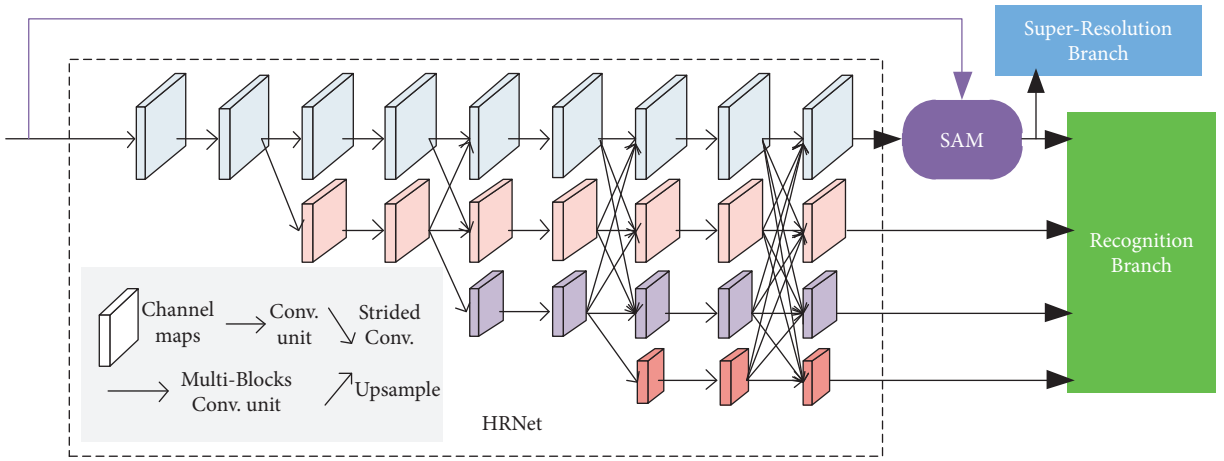


Figure 2: Structure of the encoder module.

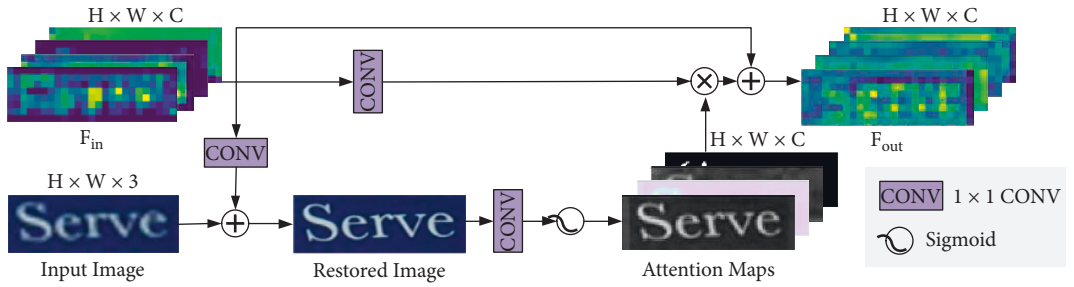


Figure 3: Structure of the SAM.

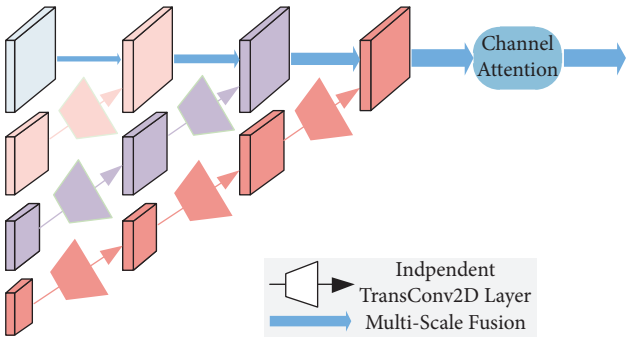


Figure 4: Structure of multiscale fusion in the recognition branch.

the HRNet, to adjust the channel numbers without changing the size of the feature maps and then rearrange the dimension of the feature maps. In other words, the channel dimension and the width dimension of feature maps are converted, to transform the two-dimensional visual feature maps into one-dimensional textual feature vectors. Then, the semantic information of one-dimensional feature vectors is strengthened through the Bi-LSTM network. Finally, the textual feature vectors are decoded by the GRU based on the attention mechanism to recognize the characters, as described in the ASTER [10]. The structure of the attention-based decoder is shown in Figure 5, and <EOS> represents the last character of the text sequence.

The sequence cross entropy loss L_{SCE} is calculated between the recognized text and the ground truth text, as shown in equation (2), to improve the decoding effect of the decoder module and the feature extraction effect of the encoder module and then improve the accuracy of STR.

$$L_{\mathrm{SCE}} = -\frac{1}{\mathrm{MN}} \sum_{i=1}^{M} \sum_{j=1}^{N} y_{i,j} \log\left(s_{i,j}\right), \tag{2}$$

where M represents the number of samples in a batch, N represents the number of text characters, y represents the ground truth text, and s represents the recognized text of the proposed model.

The loss function of the proposed model is shown in (3), where λ_1 is the corresponding weight parameter of the super-resolution branch loss and λ_2 is the corresponding weight parameter of the recognition branch loss.

$$L = \lambda_1 L_{\mathrm{sr}} + \lambda_2 L_{\mathrm{SCE}}. \tag{3}$$

4. Experiments and Results

The experimental environment of the proposed model is based on Pycharm integrated development environment, the PyTorch deep learning framework is adopted, and hardware is based on 1 NVIDIA GeForce GTX 2080Ti 11GB GPU. According to the unified experimental data and effective comparison models advocated by Baek et al. [9], the training data are the public synthetic datasets Synth90K [30] and SynthText [31] and the testing data are the testing set of seven natural scene datasets. The verification data are the training set of seven natural scene datasets. The seven natural scene datasets are as follows: IIIT5K-Words (IIIT5k) [32] refers to the regular scene text images such as billboards and posters in Google image search. Street View Text (SVT) [33] refers to the regular outdoor images in Google street view. ICDAR 2003 (IC03) [34] is a competition-based regular dataset published by the ICDAR conference, excluding scene text images of less than three characters or non-alphanumeric. ICDAR 2013 (IC13) [35] is a regular dataset, which is mostly taken from the IC03 dataset and expands some clear scene text images such as road signs and book covers. ICDAR 2015 (IC15) [36] is an irregular dataset, which mostly consists of some random images of blurred and occluded in streets or shopping malls. SVT-Perspective (SVTP) [37] refers to the irregular scene text images with perspective interference in Google street view. CUTE80 [38] mainly contains distorted and irregular scene text images.

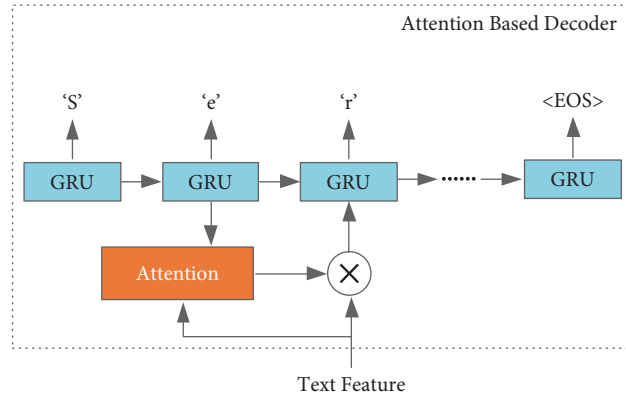


Figure 5: Structure of the attention-based decoder in the recognition branch.

The scene text images as input of the network are three-channel RGB images with a unified size of 64×256, and the size of the images is unified to 32×100 after TPS. The four kinds of feature maps with different resolutions outputted by the encoder module are 8×25, 4×13, 2×7, and 1×4, from the feature maps with the highest resolution to the feature maps with the lowest resolution, respectively. Due to the setting of super-resolution branch, pairs of low-resolution images and high-resolution images are required. Therefore, to simulate the recovery process of super-resolution network, the original image after random Gaussian blur is used as a low-resolution image and the original image is used as a high-resolution image. The Adadelta optimizer is used to update the network parameters, the weight attenuation factor is set as 0.1, the initial training learning rate is 1, and the fine-tuned training learning rate is 0.1. To ensure that the values of L_{sr} and L_{SCE} are in the same magnitude, λ_1 and λ_2 is set as 0.1 and 1 and the word accuracy is used as the evaluation metric.

4.1. Experiments of Model Comparison. To evaluate the effect of the proposed model, an experiment is performed to compare with other recent models, as shown in Table 1. For the fairness of comparison, the models using additional datasets for training are not compared. Synth90K and SynthText are used as training sets in all comparative experiments, and no lexicon is provided in the experiments. Word accuracy is taken as the evaluation metric. Meanwhile, the speed of the proposed model is 4.3 ms and 54 ms per image in the training stage and in the testing stage, respectively. Specifically, the proposed model innovatively introduces the HRNet, which combines with some methods such as the super-resolution branch, the SAM, and the Independent Trans Conv2D Layers. Compared with the ASTER and TextSR, the accuracy of the proposed model is improved in most datasets, especially in IC15, SVTP, and CUTE80, which are irregular and blurry, and the accuracy is improved by more than 3%. Compared with the Bi-STET, which uses the transformer to enhance and decode

Table 1: The accuracy comparison between the proposed model and recent models (%).

Model	Benchmark							Average	
	IIIT5k	SVT	IC03	IC13	IC15	SVTP	CUTE80	Regular	Irregular
ASTER	93.4	89.5	94.5	91.8	76.1	78.5	79.5	92.3	78.0
TextSR	92.5	87.2	93.2	91.3	75.6	77.4	78.9	91.0	77.3
ESIR	93.3	90.2	—	91.7	76.9	79.6	83.3	91.7	79.9
2DOCR	94	90.1	94.3	92.7	76.3	82.3	86.8	92.7	81.8
Bi-STET	94.7	89	**96**	93.4	75.7	80.6	82.5	**93.2**	79.6
SEED	93.8	89.6	—	92.8	80	81.4	83.6	92.0	81.6
DAN	94.3	89.2	95	93.9	74.5	80	84.4	93.1	79.6
SPIN	94.7	87.6	93.4	91.5	79.1	79.7	85.1	91.8	81.3
RobustScanner	**95.3**	88.1	—	**94.8**	77.1	79.5	**90.3**	92.7	82.3
SCGAN	94	90	95.6	93.3	81.6	**85.1**	78.1	**93.2**	81.6
Proposed model	93.7	**91.3**	93.3	94.3	**82.8**	83.1	83.0	93.1	**82.9**

Note: bold font is the optimal value in each column, and the underline font is the suboptimal value in each column.

Table 2: Comparison of accuracy of ablation models (%).

Model	IIIT5k	SVT	IC03	IC13	IC15	SVTP	CUTE80
Baseline (HRNet)	91.7	88.4	93.4	92.2	78.6	80.2	80.9
Baseline + SR (Bilinear Interpolation)	93.0	89.5	92.7	92.7	81.1	81.1	78.1
Baseline + SR (Bilinear Interpolation) + SAM	93.0	92.1	91.9	93.2	81.7	83.3	81.2
Baseline + SR (Trans Conv2D) + SAM	93.4	91.8	93.3	93.6	81.8	82.6	81.6
Proposed model	93.7	91.3	93.3	94.3	82.8	83.1	83.0

Figure 6: Qualitative comparison of recognition results.

information, the accuracy of the proposed model is also improved in three kinds of irregular datasets. Compared with the SCGAN, which introduces GAN to alleviate the background interference, the recognition accuracy of the proposed model is more balanced for different datasets. In addition, compared with other recent models, the proposed model can achieve better performance in average accuracy for irregular datasets as well as good performance for regular datasets.

4.2. Validity Experiments of the Proposed Methods. To verify the effectiveness of the proposed methods, such as the super-resolution branch, the SAM, and the Independent Trans Conv2D Layers, several comparative experiments are set up. The HRNet is used as the feature extraction network in the baseline model, and the proposed methods are gradually added to fine-tune in ablation models. The baseline model is trained for up to 3 epochs, and the fine-tuned models are trained for up to 4 epochs based on the baseline model. On the whole, the proposed model is trained for up to 13 epochs. The setting of hyperparameters is consistent all time.

Quantitative comparison is made based on the testing sets, and the results are shown in Table 2. Compared with the classical ASTER model in Table 1, which uses ResNet as the feature extraction network, the recognition accuracy of IC15, SVTP, and CUTE80 is improved by 2.5%, 1.7%, and 1.4%, respectively, by using the HRNet in the baseline model. The recognition accuracy is also improved in natural scenes by adding the super-resolution branch composed of bilinear interpolation to the baseline, which verifies that HRNet can be used as both a feature extraction network and a super-resolution network to provide effective high-resolution feature maps. In addition, the recognition accuracy can be further improved by the addition of the SAM, and instead of bilinear interpolation, we use Trans Conv2D as the upsampling method to recover super-resolution images in the super-resolution branch.

As shown in Figure 6, the qualitative comparison is given based on the irregular testing sets, such as IC15, SVTP, and

Table 3: PSNR results of restored images by Trans Conv2D and bilinear interpolation (dB).

	Bilinear interpolation	Trans Conv2D	Improved
IIIT5k	27.57	30.88	+3.31
SVT	31.93	36.05	+4.12
IC03	27.79	31.68	+3.89
IC13	27.82	32.06	+4.24
IC15	33.08	38.11	+5.03
SVTP	32.76	37.86	+5.10
CUTE80	24.98	28.39	+3.41

Figure 7: Performance of the adoption of Trans Conv2D in the super-resolution branch. In each image, from the top to bottom are the original image, blurred low-resolution image, and super-resolution image, respectively.

CUTE80. The text content below each picture is in lower case. The first line is the ground truth text, the recognition results of Baseline and Baseline + SR (Bilinear Interpolation), respectively. The second line is the recognition results of the Baseline + SR (Bilinear Interpolation) +SAM, Baseline + SR (Trans Conv2D) + SAM, and the proposed model (Baseline + SR (Trans Conv2D) + SAM + Independent Trans Conv2D Layers), respectively. It can be seen that the baseline model has some problems of misrecognition for individual characters. However, the proposed methods, such as the super-resolution branch, the SAM, and the Independent Trans Conv2D Layers, can be used gradually to effectively recognize the characters, which are relatively difficult to recognize, and then the proposed model can obtain better recognition results.

4.3. Validity Experiments of the Dual-Branch Structure. In the super-resolution branch of the proposed model, the methods of Trans Conv2D and bilinear interpolation are used to compare the effect of image recovery, respectively. The values of PSNR metric of the restored images are calculated, as shown in Table 3. Compared with the bilinear interpolation, the Trans Conv2D could increase the PSNR by more than 3 dB, which verifies the effectiveness of adopting Trans Conv2D in the super-resolution branch. Moreover, qualitative comparison is carried out with regard to the adoption of Trans Conv2D in the super-resolution branch, as shown in Figure 7. Experimental results on seven natural scene datasets verify that super-resolution branch can better accomplish the super-resolution task and assist the feature extraction network to effectively encode the scene text images; therefore, the accuracy of STR can be improved.

In the recognition branch of the proposed model, the Independent Trans Conv2D Layers are used for size expansion. The comparison between the feature maps generated by the Independent Trans Conv2D Layers and bilinear interpolation is shown in Figure 8, and the generated feature maps of five channels are randomly selected. The brighter regions in the feature maps represent the higher feature values of the regions and the more information contained. Four kinds of feature maps with different resolutions are outputted in the encoder module, with sizes of 8×25, 4×13, 2×7, and 1×4, respectively. The single Independent Trans Conv2D Layer is used to expand the size of feature maps with lower resolutions, so that the size of each resolution feature map is the same, that is, 8×25. From Figure 8, it can be seen that the feature maps generated by the Independent Trans Conv2D Layers contain more text information than the bilinear interpolation in the size of 4×13, which can reduce the loss of feature information in the process of size expansion. However, for the size of 2×7, the feature maps generated by the bilinear interpolation can only maintain some edge information, so very little information is transmitted to the recognition branch for text recognition. Meanwhile, the feature maps generated by the Independent Trans Conv2D Layers can retain some visual information even at the lowest resolution. Moreover, the multiscale fusion results transmitted to the attention-based decoder can contain more effective text information. In other words, the recognition effect of the model is significantly improved by several proposed methods on various testing sets, as shown in Table 2. As shown in Table 4, the three ablation models and the proposed model all use the super-resolution branch in the training stage and abandon it

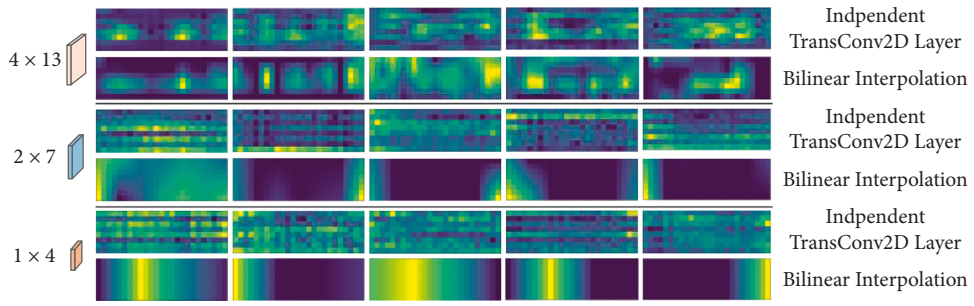


Figure 8: Comparison on the feature maps generated by the Independent Trans Conv2D layers and bilinear interpolation.

Table 4: Parameter comparison in ablation models during training and testing (M).

Model	Parameters	
	Training	Testing
Baseline (HRNet)	35.564	35.564
Baseline + SR (Bilinear Interpolation)	35.565	35.564
Baseline + SR (Bilinear Interpolation) + SAM	35.568	35.567
Baseline + SR (Trans Conv2D) + SAM	35.573	35.567
Proposed model (Baseline + SR (Trans Conv2D) + SAM + Independent Trans Conv2D Layers)	37.582	37.576

in the testing stage, which can reduce the model complexity. Moreover, the proposed model, which adds the effective methods, such as super-resolution branch, the SAM, and the Independent Trans Conv2D Layers, does not increase too many model parameters.

5. Conclusions

This paper proposes a HRNet encoder and dual-branch decoder framework-based STR model to recognize distortion, blurred, and low-resolution text in natural scenes. Based on the encoder-decoder framework, the model innovatively introduces the HRNet as feature extraction network and introduces the SAM to enhance the learning of important features. The feature maps with multiple resolutions extracted by the HRNet encoder are inputted to the dual-branch decoder module composed of the super-resolution branch and the recognition branch. Specifically, the feature maps with the highest resolution are inputted to the super-resolution branch to restore the super-resolution images and to strengthen the feature extraction effect of the encoder module. After multiscale fusion through the Independent Trans Conv2D Layers in the recognition branch, the four kinds of feature maps with different resolutions are decoded by the attention-based decoder and finally the recognized text is obtained. Through ablation experiments and comparative experiments, the effectiveness of the proposed methods such as the HRNet encoder, the super-resolution branch, and the Independent Trans Conv2D Layers is verified. Compared with the ASTER model and other recent models, the proposed model can better perform STR on multiple public natural scene datasets, especially for the text with distortion, blurring, and low resolution. In the future, STR for images with complex background and jitter imaging will be further studied.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant nos. 61801159 and 61571174 and Scientific and Technological Innovation Plan for Undergraduates in Zhejiang Province under Grant no. 2021R426066.

References

[1] S. Long, X. He, and C. Yao, "Scene text detection and recognition: the deep learning era," *International Journal of Computer Vision*, vol. 129, no. 1, pp. 161–184, 2021.

[2] C. Y. Liu, X. X. Chen, C. J. Luo, L. Jin, Y. Xue, and Y. Liu, "Deep learning methods for scene text detection and recognition," *Journal of Image and Graphics*, vol. 26, no. 06, pp. 1330–1367, 2021.

[3] J. X. Wang, Z. Y. Wang, and X. Tian, "Review of natural scene text detection and recognition based on deep learning," *Journal of Software*, vol. 31, no. 5, pp. 1465–1496, 2020.

[4] X. P. Wang, L. T. Yang, D. Meng, M. Dong, and H. Wang, "Multi-UAV cooperative localization for marine targets based on weighted subspace fitting in SAGIN environment," *IEEE Internet of Things Journal*, vol. 9, no. 8, 2021.

[5] H. Wang, L. Wan, M. Dong, K. Ota, and X. Wang, "Assistant vehicle localization based on three collaborative base stations via SBL-based robust DOA estimation," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5766–5777, 2019.

[6] X. Wang, L. Wan, M. Huang, C. Shen, and K. Zhang, "Polarization channel estimation for circular and non-circular signals in massive MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 5, pp. 1001–1016, 2019.

[7] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, pp. 2298–2304, 2016.

[8] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proceedings of the Neural Information Processing Systems*, pp. 3104–3112, Montreal, Canada, December 2014.

[9] J. Baek, G. Kim, J. Lee et al., "What is wrong with scene text recognition model comparisons? dataset and model analysis," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 4715–4723, Seoul, Korea (South), November 2019.

[10] B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao, and X. Bai, "ASTER: an attentional scene text recognizer with flexible rectification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 9, pp. 2035–2048, 2018.

[11] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[12] L. Christian, T. Lucas, H. Ferenc et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4681–4690, Honolulu, HI, USA, July 2017.

[13] W. Wang, E. Xie, P. Sun et al., "TextSR: Content-Aware Text Super-resolution Guided by Recognition," 2019, https://arxiv.org/abs/1909.07113.

[14] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 286–301, July 2018.

[15] Y. Mou, T. Lei, H. Yang et al., "PlugNet: degradation aware scene text recognition supervised by a pluggable super-resolution unit," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 158–174, November 2020.

[16] W. Wang, E. Xie, X. Liu et al., "Scene text image super-resolution in the wild," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 650–666, November 2020.

[17] J. Wang, K. Sun, T. Cheng et al., "Deep high-resolution representation learning for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3349–3364, 2020.

[18] F. Zhan and S. Lu, "ESIR: end-to-end scene text recognition via iterative image rectification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2059–2068, April 2019.

[19] H. Li, P. Wang, C. Shen, and G. Zhang, "Show, attend and read: a simple and strong baseline for irregular text recognition," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 8610–8617, 2019.

[20] Z. Cheng, F. Bai, Y. Xu, G. Zheng, S. Pu, and S. Zhou, "Focusing attention: towards accurate text recognition in natural images," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 5076–5084, September 2017.

[21] T. Wang, Y. Zhu, L. Jin et al., "Decoupled attention network for text recognition," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 12216–12224, 2020.

[22] X. Yue, Z. Kuang, C. Lin, H. Sun, and W. Zhang, "RobustScanner: dynamically enhancing positional clues for robust text recognition," *Computer Vision - ECCV 2020*, vol. 12364, pp. 135–151, 2020.

[23] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," *Advances in Neural Information Processing Systems (NIPS)*, pp. 5998–6008, 2017.

[24] P. Lyu, Z. Yang, X. Leng, X. Wu, R. Li, and X. Shen, "2D Attentional Irregular Scene Text Recognizer," 2019, https://arxiv.org/abs/1906.05708.

[25] M. Bleeker and M. de Rijke, "Bidirectional scene text recognition with a single decoder," in *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, December 2019.

[26] C. Zhang, Y. Xu, Z. Cheng et al., "SPIN: Structure-Preserving Inner Offset Network for Scene Text Recognition," 2020, https://arxiv.org/abs/2005.13117.

[27] C. Luo, Q. Lin, Y. Liu, L. Jin, and C. Shen, "Separating content from style using adversarial learning for recognizing text in the wild," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 960–976, 2021.

[28] Z. Qiao, Y. Zhou, D. Yang, Y. Zhou, and W. Wang, "SEED: semantics enhanced encoder-decoder framework for scene text recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13528–13537, Seattle, WA, USA, June 2020.

[29] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19, July 2018.

[30] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Synthetic data and artificial neural networks for natural scene text recognition," in *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, June 2014.

[31] A. Gupta, A. Vedaldi, and A. Zisserman, "Synthetic data for text localisation in natural images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2315–2324, June 2016.

[32] A. Mishra, K. Alahari, and C. V. Jawahar, "Top-down and bottom-up cues for scene text recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2687–2694, Providence, RI, USA, June 2012.

[33] K. Wang, B. Babenko, and S. Belongie, "End-to-end scene text recognition," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1457–1464, Barcelona, November 2011.

[34] S. M. Lucas, A. Panaretos, L. Sosa et al., "ICDAR 2003 robust reading competitions: entries, results, and future directions," *International Journal on Document Analysis and Recognition*, vol. 7, no. 2, pp. 105–122, 2005.

[35] D. Karatzas, F. Shafait, S. Uchida et al., "ICDAR 2013 robust reading competition," in *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1484–1493, Washington, DC, USA, August 2013.

[36] D. Karatzas, L. Gomez-Bigorda, A. Nicolaou et al., "ICDAR 2015 competition on robust reading," in *Proceedings of the International Conference on Document Analysis and Recognition*, pp. 1156–1160, ICDAR, Tunis, Tunisia, August 2015.

[37] T. Q. Phan, P. Shivakumara, S. Tian, and C. L. Tan, "Recognizing text with perspective distortion in natural scenes," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 569–576, Sydney, NSW, Australia, December 2013.

[38] A. Risnumawan, P. Shivakumara, C. S. Chan, and C. L. Tan, "A robust arbitrary text detection system for natural scene images," *Expert Systems with Applications*, vol. 41, no. 18, pp. 8027–8048, 2014.

Hindawi
International Journal of Antennas and Propagation
Volume 2022, Article ID 5417440, 11 pages
https://doi.org/10.1155/2022/5417440

Research Article

DANC-Net: Dual-Attention and Negative Constraint Network for Point Cloud Classification

Hang Sun,[1,2] **Yuanyue Zhang**,[1] **Jinmei Shi**,[3] **Shuifa Sun**,[1] **Guanqun Sheng**,[1] **and Yirong Wu**[1]

[1]*College of Computer and Information Technology, China Three Gorges University, Yichang 443002, China*
[2]*Hubei Engineering Technology Research Center for Farmland Environment Monitoring, China Three Gorges University, Yichang 443002, Hubei, China*
[3]*College of Information Engineering, Hainan Vocational University of Science and Technology, Haikou, Hainan 571158, China*

Correspondence should be addressed to Shuifa Sun; watersun@ctgu.edu.cn

Received 17 January 2022; Revised 30 March 2022; Accepted 25 April 2022; Published 28 May 2022

Academic Editor: Hervé Aubert

Convolutional neural networks, as a branch of deep neural networks, have been widely used in multidimensional signal processing, especially in point cloud signal processing. Nevertheless, in point cloud signal processing, most point cloud classification networks currently do not consider local feature correlation. In addition, they only adopt ground-truth as positive information to guide the training of networks while ignoring negative information. Therefore, this paper proposes a network model to classify point cloud signals based on feature correlation and negative constraint, DANC-Net (dual-attention and negative constraint on point cloud classification). In the DANC-Net, the dual-attention mechanism is utilized to strengthen the interaction between local features of point cloud signal from both channel and space, thereby improving the expression ability of extracted features. Moreover, during the training of the DANC-Net, the negative constraint loss function ensures that the features in the same categories are close and those in the different categories are far away from each other in the representation space, so as to improve the feature extraction capability of the network. Experiments demonstrate that the DANC-Net achieves better classification performance than the existing point cloud classification algorithms on synthetic datasets ModelNet10 and ModelNet40 and real-scene dataset ScanObjectNN. The code is released at https://github.com/sunhang1986/DANC-Net.

1. Introduction

Signal processing is usually understood as the processing of electronic signals [1–5]. Point cloud processing can be described as the processing of point cloud, a kind of multi-dimensional signal. However, the classification task of point cloud is still facing enormous challenges due to its unordered and sparse characteristics.

3D objects can be represented in two ways according to the spatial distribution of the 3D point cloud. (1) regular structure representation, which is represented by multi-view and voxel representation, and (2) irregular and unstructured representation, which is represented by point cloud and grid representation. Point cloud processing methods based on regular structured representations include 3D volumetric convolutional neural networks (CNNs) [6–8] and the multi-view CNN [9, 10]. These methods transform irregular/unstructured point clouds to regular/structured images (or volume grids), and use two-dimensional (2D) CNNs to extract local features and global features of the point cloud. Although these methods solve the unordered distribution issues of point clouds, they bring a lot of challenges in calculation and issues in memory consumption. Octree-based method [11] alleviates these problems to a certain extent and can apply 3D CNN to higher resolution grid. Le and Duan [12] and Hua et al. [13] studied different 3D convolution operators based on grid cells, which can better learn local features. On the contrary, methods based on irregular unstructured representation do not need to transform the representation of point cloud. They can learn point cloud features using special CNNs designed for raw point cloud data [14–16]. Because of low memory

consumption and simple structure of this type of representation, point cloud classification methods based on irregular unstructured representation have attracted more and more attention from researchers.

In the study of point cloud classification based on irregular and unstructured representations, Qi et al. [14] designed a PointNet network capable of point-by-point coding in order to use deep learning to process point cloud data. However, the details are lost because the whole PointNet network does not divide the point cloud regions and extract the region features. PointNet++ [15], which is based on PointNet, adopts a hierarchical structure that allows repeated capture of local information. Therefore, the overall accuracy (OA) of PointNet++ in ModelNet40 dataset is greatly improved compared with the OA of PointNet, which effectively demonstrates the importance of local information. However, because the processes of extracting local features are mutually independent, information is not exchanged between subclouds, resulting in a loss of structural information. Since then, in order to simplify the training process and save computing resources, a large number of researchers have proposed methods based on CNNs, such as PointCNN [16], tangent convolutions [17], and point cloud classification networks [18], which strengthen the geometric structure acquisition of point cloud data. However, these methods do not consider the effects of local structure relationship that are essential in 3D object recognition.

In summary, how to efficiently learn in-depth local features and their relationship from point cloud has become a pressing problem. In addition, most of the existing point cloud classification networks only use positive information to guide the training of network, lacking of effective use of negative information, which limits the network capability to extract more distinguishing features for point cloud classification.

In order to efficiently learn the correlation between local features of point cloud signals and utilize the negative information which is crucial to the classification results, we propose an effective point cloud classification network. Our point cloud classification network, based on a dual-attention mechanism and contrastive learning constraints, is named DANC-Net. The main components of the network are the channel attention and self-attention (CASA) module and the negative constraint loss function (NC-loss). The CASA module is used before the global features are aggregated. Channel attention and self-attention are used to capture the relationship between local features. In NC-loss, the output point cloud features with local feature relationships are divided into the output feature, positive sample features, and negative sample features. The output feature is constrained by negative information, in order to be approach-positive sample features and stay away from negative sample features. Positive information and negative information are used effectively at the same time, which improves the classification ability of our DANC-Net.

To sum up, our contributions are three-folds as follows:

(1) We propose a dual-attention module, CASA. It can strengthen the extraction of local feature correlation from channel and spatial, thereby helping the network to further develop the geometric structure between points.

(2) We propose a negative constraint loss function, NC-loss. Besides the positive information constraints, the effective constraint of negative information has also been strengthened; thus, the ability of the network to extract more distinctive features is improved.

(3) We propose a dual-attention negative constraint network, DANC-Net, which achieves superior performance compared with the recently proposed point cloud classification methods on open datasets ModelNet10 [8] and ModelNet40 [8] and the real-scene dataset ScanObjectNN [19].

2. Related Work

In recent years, deep learning continues to make breakthroughs in computer vision [20–23]. The early point cloud classification methods based on deep learning transform point cloud to regular volume grids and then extract features from the point cloud by using 3D CNNs [6, 8]. However, 3D CNN takes up more computing resources than 2D CNN. To make computation affordable, the volume grids are usually in low resolution, resulting in the loss of geometric information of 3D mesh shape, especially when dealing with large-scale point cloud. Therefore, the 3D point cloud is mapped to the 2D space, and then, the 2D image CNNs are used to classify [7, 10]. With well-engineered image CNNs, these methods have achieved the expected performance. Nevertheless, the selection of projection angle and projection plane has a significant impact on the classification accuracy, so the generalization ability of these models is poor.

PointNet [14], a kind of end-to-end network, is the first method to deal with point cloud directly based on deep learning. The method takes N points as input and uses a 3×3 affine transformation matrix (T-Net) to realize input alignment and feature alignment. The aligned point cloud learns global feature vectors through multiple three-layer perceptrons (MLPs) and max pooling, and finally realizes end-to-end point cloud classification. However, vital local information is ignored in the PointNet. PointNet++ [15] proposed by Qi et al. is a point cloud classification network based on PointNet. It refers to the feature extraction method of PointNet to process each group of point clouds independently. Then, the global features are aggregated using max pooling. The hierarchical structure of PointNet++ exploits local information to a certain extent. In PointNet++, multi-scale algorithm is used to group point clouds. In the process of grouping, it is inevitable that there will be repeated grouping points, which will result in local information redundancy and reduce the classification ability of the network. For the purpose of reducing the redundancy of local information, the authors of A-CNN [24] proposed the constraint-based k-nearest neighbor (k-NN) algorithm and annularly convolution on the basis of hierarchical structure. As shown in Figure 1, the input point cloud is sampled and the constraint-based k-NN algorithm is used to construct groups in each layer of the network. Then, the features within each group are extracted by combining annular convolution with max pooling. Compared with multi-scale

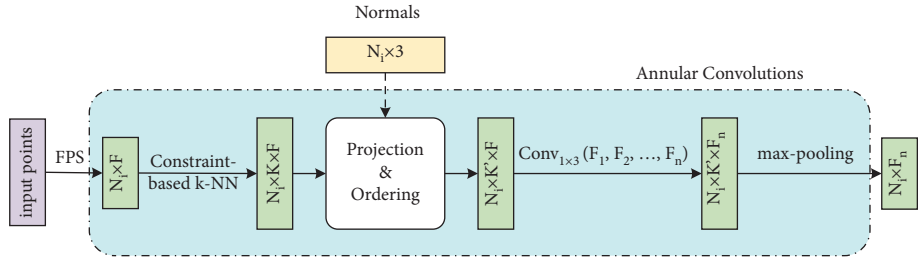


Figure 1: A-CNN abstract layer structure.

grouping, rings of annularly convolution do not contain duplicate points, which allows the network to learn more discriminant features. Therefore, A-CNN achieved a higher classification performance than PointNet++ on the ModelNet dataset. The progress of the above end-to-end point cloud classification methods is undeniable, However, their approach of extracting local features independently leads to inadequate identification of correlations between points or local neighborhoods.

Recently, attention mechanism [25] has achieved remarkable achievements in natural language processing, image recognition [26], and other fields. In point cloud classification, Bhattacharyya et al. [27] proposed an altitude attention model, which can achieve superior classification performance of airborne laser scanning (ALS) by considering the altitude information of points. Lee et al. [28] proposed a simple and efficient network based on self-attention, called set transformer, which can process set data, such as a point cloud. Shajaha et al. [29] proposed a multi-view CNN with self-attention. Multiple views of a roof point cloud were taken as the input, an adaptive weight learning algorithm was used to assign weights corresponding to each view, and the category of the roof was the output. However, the generalization ability of the model [29] is poor and is limited to special field. On the contrary, the DANC-Net we proposed can be applied to any point cloud classification tasks.

Currently, most of the point cloud classification networks only use ground-truth as positive information to guide the training of the network while negative information is ignored, which leads to the limitation of network discrimination capabilities. Therefore, in order to further explore the correlation of local features of point cloud and the constraints of negative information on features, this paper proposes the DANC-Net based on dual-attention mechanism and negative information constraints.

3. Method

This section details the proposed DANC-Net in this paper. First, the architecture of DANC-Net point cloud classification method is introduced in Section 3.1. Then, Section 3.2 performs detailed analysis of dual-attention CASA module for capturing correlations between local features. Next, Section 3.3 presents the loss function NC-loss under the negative information constraint. Finally, Section 3.4 summarizes the total loss function of the DANC-Net.

3.1. DANC-Net Architecture. For a clear understanding of our DANC-Net, we show the network architecture and the output feature map size of each layer in the network in Figure 2. Our DANC-Net consists of five layers.

(1) Input layer: for a given 3D shape point cloud, the coordinates and normals of N points are used as input.

Feature map size: each input consists of a 3D coordinates (x, y, z) and a normal, i.e., two $N \times 3$-dimensional tensors.

(2) A-CNN layer: local features are extracted from point cloud. This layer performs two feature extractions, and each feature extraction includes two operations, namely, the farthest point sampling (FPS) algorithm [30] and the A-CNN abstraction layer.

Feature map size: after two feature extractions, the previous-level output point cloud is divided into N_1 and N_2 local regions, and the number of channels for each local region feature is 128 and 256, respectively.

(3) CASA layer: a new feature map $\boldsymbol{f_r}$ with geometric relationship and positional relationship among local features of point cloud is obtained, and then, the global feature vector $\mathbf{f}_g$ is aggregated through the PointNet [14] layer.

Feature map size: the point cloud global feature vector $\mathbf{f}_g$ with correlation between local feature output by this layer has 1024 channels.

(4) Negative information constraint layer: the $\boldsymbol{f_r}$ is used to construct the NC-loss under the constraint of negative information, so as to restrain the mutual interference between similar categories (such as nightstands and dressing tables with similar spatial structure).

Feature map size: the feature map $\boldsymbol{f_r}$ that is fed into the loss function has 256 channels.

(5) Output layer: MLPs are used to obtain the probability score of the point cloud belonging to the c category.

Feature map size: the final output vector size of our DANC-Net is $1 \times c$.

3.2. Dual-Attention (CASA) Unit. Most point cloud classification networks only enhance the expression ability of the network from the perspective of enhancing local feature

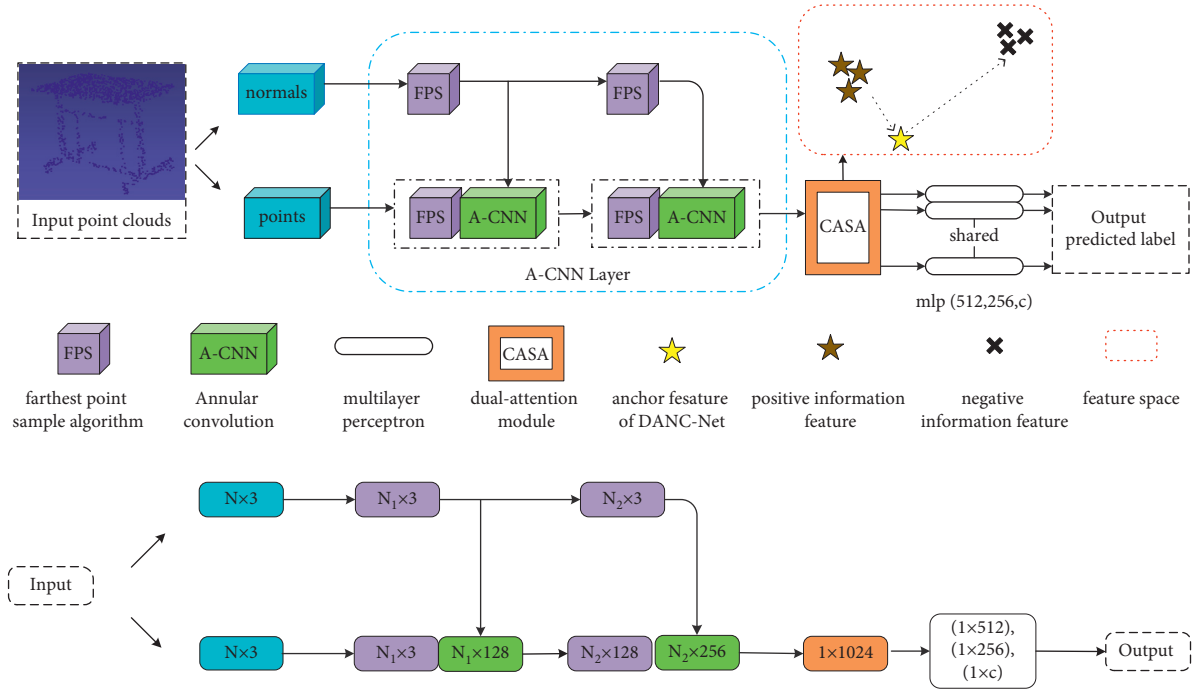


FIGURE 2: The architecture of DANC-Net (top) and the feature map size of each node of the model (bottom). The blue dashed box on the left of the network structure is the local feature extraction layer (A-CNN layer).

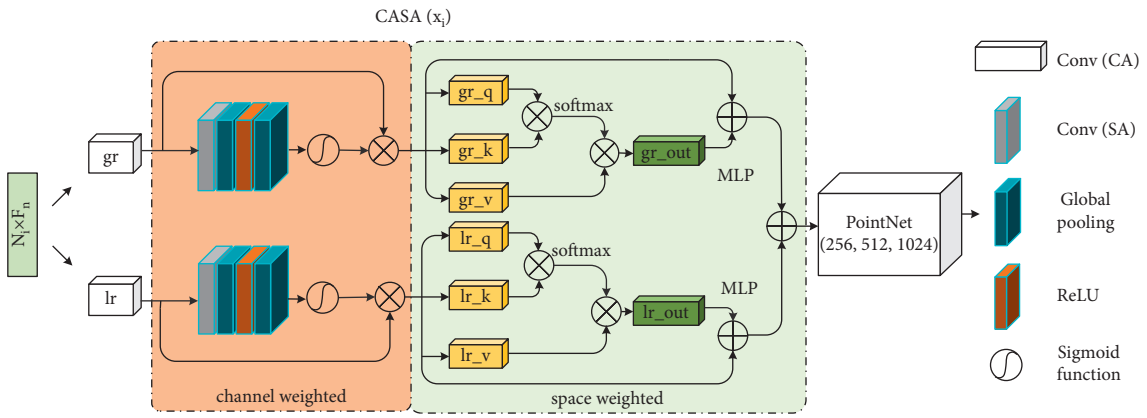


FIGURE 3: CASA module. ***gr*** and ***lr***, respectively, represent the geometric and positional features of the sub-cloud. $\boldsymbol{x_i}$ represents the *i*th input feature of our CASA module.

extraction, while ignoring the exchange of information between local features. CASA unit can adaptively learn feature weights and capture the correlation between local features. As shown in Figure 3, the CASA module consists of two branches, each consisting of channel attention [31] and self-attention [32]. Among them, the upper branch is used to extract the geometric feature relationship and the lower branch is used to extract the positional feature relationship. The CASA module considers that the features of different channels contain completely a different weighting information and point distribution is not uniform in different spatial positions. CASA treats different features and points unequally, providing extra flexibility in processing different types of information and expanding CNN's expressive ability.

For a set of points containing n points $P = \{P_i, \quad i = 1, \ldots, n\}$, A-CNN is used to extract the geometric feature vector $gr_i \in \mathbb{R}^d$ and location feature vector $lr_i \in \mathbb{R}^3$ of the local sub-cloud $\boldsymbol{G_i}$; the two feature vectors are then input into the CASA module. The output is a high-level global feature $\mathbf{f}_g$ that incorporates context information. The key to this process is how to generate different weights for each point feature. The detailed implementation process of the CASA module is analyzed as follows.

First, the input geometric information and location information are weighted by channel attention. In this process, we use global average pooling to transform the global information of the channel to the channel descriptors; that is, the channel dimension is kept unchanged, but the other dimensions of the feature map are reduced to 1. In

order to obtain the weights $\boldsymbol{CA}$ of the channels, channel descriptors sequentially pass through convolution layer, ReLU activation function, convolution layer, and sigmoid function.

$$CA = S(Conv(L(Conv(g_c(F_c))))), \tag{1}$$

where $\boldsymbol{F_c}$ is the feature graph of the input, g_c is the global pooling function, and S and L represent the sigmoid function and ReLU function, respectively. Finally, the input feature map $\boldsymbol{F_c}$ and channel weight $\boldsymbol{CA}$ are multiplied and the weighted feature map is obtained as shown in

$$F_c^* = CA \otimes F_c. \tag{2}$$

Second, the feature map is spatially weighted. Three 1×1 traditional convolutions of the input feature map are performed, and the three feature matrixes $\boldsymbol{q}$, $\boldsymbol{k}$, and $\mathbf{v}$ are obtained. The correlation matrix $\boldsymbol{M}$ is obtained by multiplying matrixes $\boldsymbol{q}$ and $\boldsymbol{k}$. The softmax normalization operation is performed on the correlation matrix $\boldsymbol{M}$ to obtain the attention weight in the range [0, 1]. The weight coefficient is applied to the feature matrix $\mathbf{v}$, and the residual connection is made with the input feature map F_c^*, so that each local feature is weighted by all local features. The weighted feature map obtained is shown in

$$F^* = \text{softmax}\left(q \otimes k^T\right) \otimes v + F_c^*. \tag{3}$$

Finally, the feature graph $\boldsymbol{F}^*$, which has been weighted by channel and space, is fused by the MLP, and local features $\boldsymbol{f_r}$ containing context correlation are obtained by matrix addition. All local regional features are then aggregated by PointNet to obtain the global feature $\mathbf{f}_g$. To demonstrate the correctness of the CASA module, we verify its point cloud classification effect in the ablation study (Section 4.3).

3.3. Negative Constraint Loss Function (NC-Loss) Unit. Inspired by [33, 34], in order to further improve the discrimination ability of the point cloud classification network, a loss function (NC-loss) with negative information constraints is proposed in this paper. The point cloud with the same label as the output feature is called positive information, while the point cloud with different labels from the output feature is called negative information. As shown in Figure 2, the red dotted box represents the feature space of the constructed NC-loss. In the feature space, NC-loss can not only close the features of positive information and the output features, but also push the features of negative information and the output features farther.

In the proposed NC-loss, an output feature is selected from the local features $\boldsymbol{f_r}$ containing context correlation. In this paper, the features of all input point cloud samples are traversed to ensure that the feature of each point cloud sample has the opportunity to be selected as output features. In the feature space, the distances between point clouds of the same class are minimized and the distances between point clouds of different classes are maximized. Therefore, the regularization loss function of contrastive learning is defined in

$$L_{nc} = \sum_{i=1}^{B} [D(F_i, F_P) - D(F_i, F_N)], \tag{4}$$

where $D(x, y)$ represents the L_1 distance between x and y. The number of input point cloud samples is B, and $i \in I \equiv \{1, \ldots, B\}$ represents the ith point cloud sample selected as the output feature. $P(i)$ is the set of positive point cloud samples, which contains all the point clouds in the B samples with the same label as the output feature. $N(i)$ is the set of negative point cloud sample, which contains all the point clouds in the B samples with the same label as the output feature. $F = \text{MLP}(f_r) \in \mathbb{R}^{D_M}$ means that the local features $\boldsymbol{f_r}$ containing context correlation become the feature matrixes $\boldsymbol{F}$ after MLP mapping; i.e., $\boldsymbol{F_i}$, $\boldsymbol{F_P}$, and $\boldsymbol{F_N}$ represent the output feature matrixes, the positive sample feature matrixes, and negative information sample feature matrixes, respectively, where D_M is a constant 256.

3.4. DANC-Net Totally Loss Function. For the task of point cloud classification, we use the cross-entropy loss function to measure the distance between the predicted values of 3D point cloud samples and the ground-truth. The calculation method of cross-entropy loss is as follows:

$$L(P) = -\sum_{k=1}^{c} \hat{y}_k \ln(y_k), \tag{5}$$

where $\hat{y}_k \in \{0, 1\}$ indicates the kth value in the label vector. $y_k \in [0, 1]$ indicates the probability that the prediction sample P belongs to the kth class.

Therefore, the finally loss function L of our DANC-Net consists of a classification loss function and a negative constraint loss function (NC-loss). The final loss function L can be expressed as follows:

$$L = L(p) + \lambda L_{nc} = L(p) + \lambda \sum_{i=1}^{B} D(F_i, F_P) - D(F_i, F_N). \tag{6}$$

In formula (6), λ is the penalty parameter used to balance the classification loss and NC-loss. The ablation experiment results show that the classification accuracy of the DANC-Net based on the dual-attention CASA module can be further improved by using NC-loss.

4. Experiments and Result Analysis

We used three benchmark datasets, ModelNet10, ModelNet40, and ScanObjectNN, to compare our DANC-Net with the state-of-the-art point cloud classification algorithms [6, 8, 14–16, 24, 35–40]. The synthetic datasets ModelNet10 and ModelNet40 are subsets of ModelNet (a large 3D CAD model dataset). Each point cloud sampled from the grid contains 10,000 points and normal vectors, and the coordinates are normalized to unit spheres. ScanObjectNN is a real-world dataset of point cloud objects, constructed from indoor scene scanning. More details about three benchmark datasets can be found in Table 1.

Table 1: Distribution of training and test sets.

Datasets	Class number	Training models	Testing models	Total models
ModelNet10	10	3991	908	4899
ModelNet40	40	9843	2468	12331
ScanObjectNN	15	80% of the total models were randomly selected	20% of the total models were randomly selected	2902

In robustness test, 80% and 20% of the total models were randomly selected as the training set and test set. In the experiment using ModelNet10 and ModelNet40, 1024 points with normals were sampled, and the normals were used in the ordering algorithm in A-CNN.

The hardware environment of the experiments included an RTX 2080 Ti graphics card, 12 GB video memory, Ubuntu 18.04 operating system, and CUDA 10.1 + cuDNN 7.6.5 + TensorFlow 1.3.0 + Python 3.6. In Tables 2–4, the highest, second highest, and third highest classification accuracies are indicated by bold, underline, and italic text, respectively.

4.1. Parameters. For experiments on three benchmark datasets, 1024 points from 3D meshes are sampled randomly as the input of the DANC-Net. The data augmentation method was the same as that of A-CNN. The loss included classification loss and comparison loss, as defined in (6), and the classification loss used the cross-entropy loss function. Using the Adam optimizer, the initial learning rate was set to 0.001 and attenuated at a decay rate of 0.7 per 200,000 steps. The classification model was trained for 250 epochs with a batch_size of 16. In the experiment, the penalty parameter was set to 1.0.

4.2. Comparison Experiments and Analysis. We demonstrate the effectiveness of our DANC-Net on ModelNet40 and ModelNet10 datasets though comparison experiments. As shown in Tables 2 and 3, using the datasets ModelNet10 and ModelNet40, our DANC-Net was compared with the state-of-the-art point cloud classification methods based on deep learning. The quantitative evaluation of the classification performance of models in experiments adopts the commonly used evaluation metrics for point cloud classification: mean per-class accuracy (mA) and overall accuracy (OA). mA and OA are defined by

$$mA = \sum_{i=1}^{c} \frac{\text{num}(TP)_i}{\text{num}_i} \times \frac{1}{c}. \tag{7}$$

$$OA = \frac{\sum_{i=1}^{c} \text{num}(TP)_i}{T}. \tag{8}$$

where $\text{num}(TP)_i$ represents the number of 3D meshed shapes correctly classified into category i; num_i represents the number of 3D meshed shapes that belong to category i; and Trepresents the total number of 3D meshed shapes to be predicted. In Tables 2 and 3, the top three mA and OA are highlighted by bold, underline, and italic text, respectively. Furthermore, the classification results of the methods used for comparison are obtained from corresponding papers. If the data and results are not given in the paper, we will download the codes of the models. The classification results are obtained by training and testing the models in the corresponding experimental environment. Moreover, "-" means that the dataset cannot be used by the methods or the codes are not provided in the paper.

4.2.1. Comparison Experiments on ModelNet10 Dataset. The DANC-Net achieves the best classification performance with 95.5 OA and 95.4 mA on ModelNet10 (Table 2).

(1) Comparison with the deep learning point cloud classification methods based on voxel grid.

In Table 2, we compare DANC-Net with 3DShapeNets and VoxNet methods based on voxel grid on ModelNet10. We observe that our DANC-Net achieves 12.0 OA and 3.5 OA more than 3DShapeNets and VoxNet. For 3DShapeNets and VoxNet, the classification performance is low because the invariance of point cloud cannot be maintained when point cloud is converted to 3D voxel grids. For our DANC-Net, conversion of point cloud to other forms is not required, and the invariance of point cloud is retained, achieving high classification accuracy.

(2) Comparison with the deep learning classification methods based on point cloud.

As shown in Table 2, we can observe that (1) compared with the classical PointNet and PointNet++, mA of our DANC-Net is increased by 1.2 and 0.7, and OA is increased by 1.1 and 0.6, respectively, on ModelNet10. (2) Compared with Kd-Net [35] and DGCNN [36], mA is increased by 1.9 and 0.6, and OA is improved by 1.5 and 0.6, respectively. (3) Compared with A-CNN, our DANC-Net has achieved an improvement of 1.0 mA and 0.9 OA on ModelNet10. Our DANC-Net achieves high performance because the CASA module can adjust both spatial weights of features and channel weights of features. Furthermore, our DANC-Net adds negative information constraint, so the classification accuracy is higher than that of other methods.

4.2.2. Contrast Experiments on ModelNet40 Dataset. The DANC-Net achieved the highest classification accuracy with 92.9 OA and 90.5 mA on ModelNet40 (Table 3).

(1) Comparison with the deep learning point cloud classification methods based on voxel grid.

Table 2: Classification performance on ModelNet10. (The top three accuracies are highlighted by bold, underline, and italic.)

Methods	Input	Points (k)	mA (%)	OA (%)
VoxNet (IROS 2015)	Points	1	—	92.0
3DShapeNets (CVPR 2016)	Points	1	—	83.5
PointNet (CVPR 2017)	Points	1	94.2	94.4
PointNet++ (CVPR 2017)	Points + normal	5	*94.7*	<u>94.9</u>
Kd-Net (ICCV 2017)	Points	32	93.5	94.0
DGCNN (TOG 2019)	Points	1	<u>94.8</u>	<u>94.9</u>
A-CNN (CVPR 2019)	Points + normal	1	94.4	*94.6*
Ours	Points + normal	1	**95.4**	**95.5**

Input and points represent the input data type and the number of sampling points, respectively.

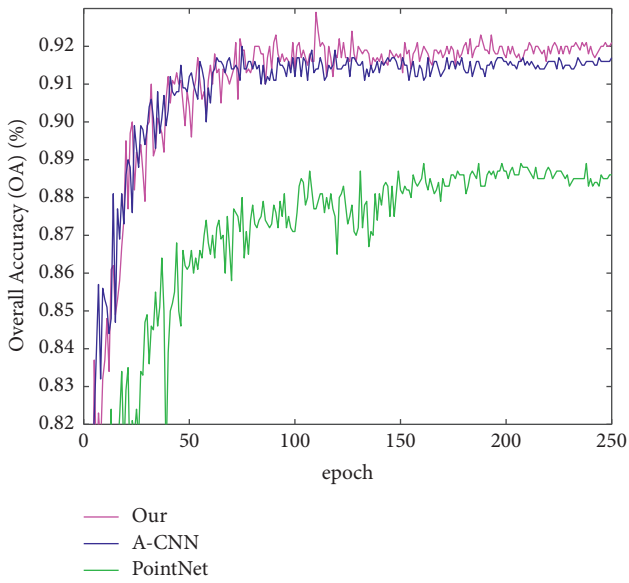


Figure 4: Variation of the overall accuracy (OA) in % on ModelNet40 with 250 epochs.

As shown in Table 3, our DANC-Net achieves 13.2 mA and 7.5 mA more than 3DShapeNets and VoxNet, respectively, on ModelNet40 dataset, and OA is increased by 8.2 and 7.0. Our DANC-Net has achieved the above improvement because it is more effective in learning the features of point cloud than 3DShapeNets and VoxNet methods based on voxel.

(2) Comparison with the deep learning classification methods based on point cloud.

As shown in Table 3, we can observe that (1) mA of our DANC-Net is increased by 4.3 and 2.3, and OA is improved by 3.7 and 2.3, respectively, compared with the classical PointNet and PointNet ++. (2) Compared to Kd-Net based on KD tree, mA of our DANC-Net is increased by 2.0, and OA is improved by 1.1. (3) Compared to DGCNN, PointCNN, and A-CNN based on convolution, mA of our DANC-Net is increased by 1.1, 2.4, and 0.6, respectively, and OA is increased by 1.1, 0.7, and 0.7, respectively. (4) Compared with the recent point cloud classification network PointHop [37] and MRFGAT [38], mA of our DANC-Net is increased by 6.1 and 0.4, and OA is increased by 3.8 and 0.4. Our DANC-Net demonstrates high classification performance because it extracts local features of point clouds to obtain more information. In addition, in order to obtain higher classification accuracy, CASA is added to our DANC-Net, which can take advantage of local feature correlations.

Meanwhile, mA of our DANC-Net is increased by 1.1, and OA is increased by 0.6, respectively, compared with DGANet [39]. There are two reasons for achieving high performance: (1) our DANC-Net can dynamically weight local features by CASA module. (2) Loss function with negative constraint is used to eliminate the interference between point cloud categories with similar structures. In contrast, DGANet introduces offset attention into graph-based methods. The accuracy of constructing local graph impacts the results of feature extraction. As a consequence, the classification accuracy of DGANet is inferior to our DANC-Net.

Besides, we compare our method with SRN-PointNet++ [40], which can extract geometrical relationship between points. OA of our DANC-Net is 1.4 higher than that of SRN-PointNet++. We think that (1) our DANC-Net uses CASA to assign different weights to features, improving the flexibility of network. SRN-PointNet++ uses MLPs to obtain geometrical relationship between points, and the weights are the same for each local feature. Therefore, the classification accuracy of SRN-PointNet++ is not as good as that of our DANC-Net. (2) ModelNet40 contains much more categories of 3D point cloud shape than ModelNet10. Therefore, the point clouds in ModelNet40 have higher shape similarity and smaller distance between categories than those in ModelNet10. Under the circumstances, the loss function with negative constraint plays a prominent role, which further improves the classification performance of DANC-Net. To sum up, our proposed DANC-Net obtains the better performance in the final results than other deep learning methods on the task of point cloud classification.

The OA changes of our DANC-Net, PointNet, and A-CNN with epoch times are depicted in Figure 4. It can be found that (1) the OA of our DANC-Net is consistently higher than that of A-CNN when the OA reaches a stationary stage. (2) The OA of our DANC-Net is always higher than that of PointNet throughout the testing process.

4.3. Ablation Study. In order to verify the effectiveness of the dual-attention CASA module and the loss function with negative constraint (NC-loss) in our DANC-Net network,

Table 3: Classification performance on ModelNet40. (The top three accuracies are highlighted by bold, underline, and italic.)

Methods	Input	Points (k)	mA (%)	OA (%)
VoxNet (IROS 2015)	Points	1	83.0	85.9
3DShapeNets (CVPR 2016)	Points	1	77.3	84.7
PointNet (CVPR 2017)	Points	1	86.2	89.2
PointNet++ (CVPR 2017)	Points + normal	5	87.9	91.9
Kd-Net (ICCV 2017)	Points	32	88.5	91.8
PointCNN (NeurIPS 2018)	Points + normal	1	88.1	92.2
DGCNN (TOG 2019)	Points	1	90.2	*92.3*
A-CNN (CVPR 2019)	Points + normal	1	89.9	92.2
SRN-PointNet++ (CVPR 2019)	Points	1	—	91.5
PointHop (IEEE *T* MULTIMEDIA 2020)	Points	1	84.4	89.1
DGANet (remote sensing 2021)	Points	1	89.4	*92.3*
MRFGAT (INT *J* ANTENN PROPAG 2021)	Points	1	90.1	92.5
Ours	Points + normal	1	**90.5**	**92.9**

Input and points represent the input data type and the number of sampling points, respectively.

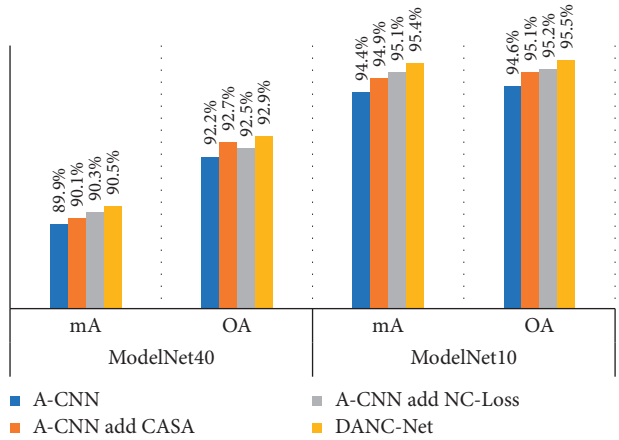


Figure 5: Classification effects of ablation experiments.

Table 4: Overall accuracy in % on ScanObjectNN. (The top three accuracies are highlighted by bold, underline, and italic.)

Methods	OBJ_BG	PB_T25	PB_T25_R	PB_T50_R	PB_T50_RS	Mean of OA
PointNet	79.0	74.5	73.2	69.3	67.8	72.7
PointNet++	83.5	85.4	82.8	80.9	78.7	82.2
DGCNN	85.3	**85.7**	83.8	*80.9*	*81.0*	83.3
A-CNN	85.1	83.2	83.5	81.8	81.4	*83.0*
Ours	**86.4**	84.3	**84.0**	**82.7**	**81.8**	**83.8**

this paper conducts ablation experiments on ModelNet10 and ModelNet40 datasets. Besides our DANC-Net, three additional models are designed in these experiments, including A-CNN, A-CNN + CASA module, and A-CNN + NC-loss.

In Figure 5, we observe that (1) CASA module and NC-loss contribute different degrees of improvement in classification performance from A-CNN to A-CNN + CASA and from A-CNN to A-CNN + NC-loss. (2) When both CASA module and NC-loss are added to the A-CNN, the classification accuracy will reach the maximum.

4.4. Robustness Test. We employ ScanObjectNN dataset to test the robustness of our DANC-Net. The objects in ScanObjectNN, which are selected from SceneNN [41] and ScanNet [42] scenes, are screened by the bounding boxes. In Table 4, we summarize the OAs of our DANC-Net and the state-of-the-art methods on ScanObjectNN dataset. In Table 4, OBJ_BG, PB_T25, PB_T25_R, PB_T50_R, and PB_T50_RS are five subsets of ScanObjectNN. OBJ_BG is point cloud with background, and PB denotes point cloud with random disturbance. T25 or T50 denotes point cloud after translation of 25% or 50% is performed. *R* and S denote rotation and scaling, respectively.

As presented in Table 4, we find that (1) our DANC-Net achieves the highest OA of 86.4 in OBJ_BG (without perturbation), compared with other methods. (2) Our DANC-Net also outperforms other methods on disturbed PB_T25_R, PB_T50_R, and PB_T50_RS datasets. (3) Overall, our DANC-Net has the highest average OA of 83.8

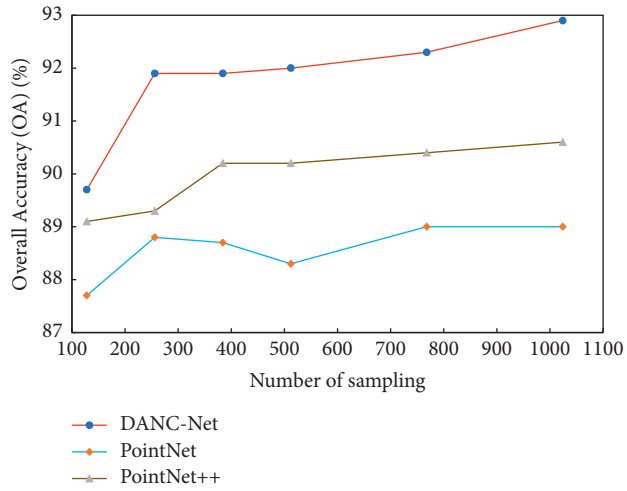


Figure 6: Robustness test of our DANC-Net to a sparse point cloud representation.

on ScanObjectNN. In summary, our DAC-Net performs better than other methods, which demonstrates its robustness on real-world datasets.

Figure 6 shows the robustness of our DANC-Net when ModelNet40 dataset is used for the test, in which 25%, 50%, 62.5%, 75%, and 87.5% of the input sampling points are randomly selected and discarded. The number of sampling points for training and testing is the same.

As shown in Figure 6, (1) our DANC-Net also achieves the highest classification accuracy, no matter whether the input point cloud is dense or sparse. (2) Compared with DANC-Net when 1024 sampling points are used, when 25% of sampling points are randomly dropped, the OA of the DANC-Net is only 0.6 lower. (3) Compared with PointNet when 1024 sampling points are used, when 87.5% of sampling points are randomly dropped, the OA of the DANC-Net is higher. It is shown that our DANC-Net is robust to point cloud sparsity.

5. Conclusions

At present, most point cloud classification models fail to explore the correlation between local regional features, and they use ground-truth as positive information to guide the network training, ignoring negative information. In view of this issue, we propose a new model of dual-attention and negative constraint network (DANC-Net). Our DANC-Net strengthens the interaction between local features of point cloud signals from both channel and space. At the same time, positive information and negative information are used effectively to improve the classification ability of our DANC-Net. Experimental results on synthetic datasets demonstrate that our DANC-Net successfully achieves high classification performance on point cloud classification tasks. Experimental results on real-world datasets confirm that our DANC-Net is robust. In the contrastive learning, the hard negative example is beneficial to enhance the ability of the network to distinguish between signal and noise. Therefore, in the future, we will explore new strategies to increase the number of the hard negative samples, so as to improve the classification ability of the DANC-Net.

Data Availability

The datasets used in this manuscript are available at https://github.com/artemkomarichev/a-cnn and https://hkust-vgd.github.io/scanobjectnn/.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Nos. 61871258 and U1703261), Natural Science Foundation of Hubei Province of China (No. 2021CFB004), and Open Fund of the Key Laboratory of Oil and Gas Resources and Exploration Technology of the Ministry of Education (K2021-01).

References

[1] J. Cong, X. Wang, C. Yan, L. T. Yang, M. Dong, and K. Ota, "CRB weighted source localization method based on deep neural networks in multi-UAV network," *IEEE Internet of Things Journal*, p. 1, 2022.

[2] X. Wang, L. T. Yang, D. Meng, M. Dong, K. Ota, and H. Wang, "Multi-UAV cooperative localization for marine targets based on weighted subspace fitting in SAGIN environment," *IEEE Internet of Things Journal*, vol. 9, 2021.

[3] F. Wen, J. Shi, and Z. Zhang, "Generalized spatial smoothing in bistatic EMVS-MIMO radar," *Signal Processing*, vol. 193, Article ID 108406, 2022.

[4] G. Zheng, Y. Song, and C. Chen, "Height measurement with meter wave polarimetric MIMO radar: signal model and MUSIC-like," *Algorithms*, vol. 190, Article ID 108344, 2022.

[5] J. Shi, Z. Yang, and Y. Liu, "On parameter identifiability of diversity-smoothing-based MIMO radar," *IEEE Transactions on Aerospace and Electronic Systems*, p. 1, 2021.

[6] D. Maturana and S. Scherer, "VoxNet: a 3D convolutional neural network for real-time object recognition," in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, Hamburg, Germany, October 2015.

[7] C. R. Qi, H. Su, M. Niebner, A. Dai, M. Yan, and L. J. Guibas, "Volumetric and multiview cnns for object classification on 3D data," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5648–5656, Las Vegas, NV, USA, June 2016.

[8] Z. Wu, S. Song, A. Khosla et al., "3D ShapeNets: a deep representation for volumetric shapes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1912–1920, Boston, MA, USA, June 2015.

[9] H. Huang, E. Kalogerakis, S. Chaudhuri, D. Ceylan, V. G. Kim, and E. Yumer, "Learning local shape descriptors from Part Correspondences with multiview convolutional networks," *ACM Transactions on Graphics*, vol. 37, no. 1, pp. 1–14, 2018.

[10] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3d shape recognition," in *Proceedings of the IEEE International Conference on*

Computer Vision, pp. 945–953, Santiago, Chile, December 2015.

[11] G. Riegler, A. Osman Ulusoy, and A. Geiger, "Octnet: learning deep 3d representations at high resolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3577–3586, Honolulu, HI, USA, July 2017.

[12] T. Le and Y. Duan, "PointGrid: a deep network for 3d shape understanding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9204–9214, Salt Lake City, UT, USA, June 2018.

[13] B.-S. Hua, M.-K. Tran, and S.-K. Yeung, "Pointwise convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 2018.

[14] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: deep learning on point sets for 3D classification and segmentation," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 77–85, Honolulu, HI, USA, July 2017.

[15] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: deep hierarchical feature learning on point sets in a metric space," in *Proceedings of the 31st Annual Conference on Neural Information Processing Systems*, pp. 5100–5109, Long Beach CA USA, December 2017.

[16] Y. Li, R. Bu, M. C. Sun, W. Wu, X. Di, and B. Chen, "PointCNN: convolution on x-transformed points," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 820–830, Montréal, Canada, December 2018.

[17] M. Tatarchenko, J. Park, V. Koltun, and Q. -Y. Zhou, "Tangent convolutions for dense prediction in 3d," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3887–3896, Salt Lake City, UT, USA, June 2018.

[18] S. Qiu, S. Anwar, and N. Barnes, "Geometric back-projection network for point cloud classification," *IEEE Transactions on Multimedia*, vol. 24, pp. 1943–1955, 2017.

[19] M. A. Uy, Q.-H. Pham, B.-S. Hua, T. Nguyen, and S. -K. Yeung, "Revisiting point cloud classification: a new benchmark dataset and classification model on real-world data," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1588–1597, Seoul, Korea (South), November 2019.

[20] J. Liu, Z. Chen, B. Du, and D. Tao, "ASTS: a unified framework for arbitrary shape text spotting," *IEEE Transactions on Image Processing*, vol. 29, pp. 5924–5936, 2020.

[21] X. Huang, B. Du, and W. Liu, "Multichannel color image denoising via weighted schatten p-norm minimization," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence IJCAI*, Yokohama, July 2020.

[22] H. Sun, Y. Zhang, P. Chen et al., "Scale-free heterogeneous cycleGAN for defogging from a single image for autonomous driving in fog," *Neural Computing & Applications*, pp. 1–15, 2021.

[23] H. Sun, J. Li, J. Chang, B. Du, and Z. Su, "Efficient compressive sensing tracking via mixed classifier decision," *Science China Information Sciences*, vol. 59, no. 7, pp. 072102–072115, 2016.

[24] A. Komarichev, Z. Zhong, and J. A. Hua, "CNN: annularly convolutional neural networks on point clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7421–7430, Long Beach, CA, USA, June 2019.

[25] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," in *Proceedings of the Advances in neural information processing systems*, Long Beach, CA, USA, December 2017.

[26] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, "Stand-alone self-attention in vision models," in *Proceedings of the Advances in Neural Information Processing System*, Vancouver, BC, Canada, December 2019.

[27] P. Bhattacharyya, C. Huang, and K. Czarnecki, "Sa-det3d: self-attention based context-aware 3d object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3022–3031, Montreal, BC, Canada, October 2021.

[28] J. Lee, Y. Lee, J. Kim, A. R. Kosiorek, S. hoi, and Y. W. Teh, "Set transformer: a framework for attention-based permutation-invariant neural networks," in *Proceedings of the International Conference on Machine Learning*, pp. 3744–3753, PMLR, Long Beach, CA, USA, June 2019.

[29] D. A. Shajahan, V. Nayel, and R. Muthuganapathy, "Roof classification from 3-D LiDAR point clouds using multiview CNN with self-attention," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 8, pp. 1465–1469, 2020.

[30] C. Moenning and N. A. Dodgson, *Fast Marching Farthest point Sampling*, University of Cambridge, Computer Laboratory, Cambridge, 2003.

[31] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: feature fusion attention network for single image dehazing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 11908–11915, New York, NY, USA, February 2020.

[32] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7794–7803, Salt Lake City, UT, USA, June 2018.

[33] P. Khosla, P. Teterwak, C. Wang et al., "Supervised contrastive learning," in *Proceedings of the Advances in Neural Information Processing Systems*, vol. 33, pp. 18661–18673, Curran Associates, Inc, December 2020.

[34] I. Misra and L. van der Maaten, "Self-supervised learning of pretextinvariant representations," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition CVPR*, Seattle, WA, USA, June 2020.

[35] R. Klokov and V. Lempitsky, "Escape from cells: deep kd-networks for the recognition of 3d point cloud models," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 863–872, Venice, Italy, October 2017.

[36] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," *ACM Transactions on Graphics*, vol. 38, no. 5, pp. 1–12, 2019.

[37] M. Zhang, H. You, P. Kadam, S. Liu, and C.-C. J. Kuo, "Pointhop: an explainable machine learning method for point cloud classification," *IEEE Transactions on Multimedia*, vol. 22, no. 7, pp. 1744–1755, 2020.

[38] X. A. Li, L. Y. Wang, and J. Lu, "Multiscale receptive fields graph attention network for point cloud classification," *Complexity*, vol. 2021, Article ID 8832081, 9 pages, 2021.

[39] J. Wan, Z. Xie, Y. Xu, Z. Zeng, D. Yuan, and Q. Qiu, "DGANet: a dilated graph attention-based network for local feature extraction on 3D point clouds," *Remote Sensing*, vol. 13, no. 17, p. 3484, 2021.

[40] Y. Duan, Y. Zheng, J. Lu, J. Zhou, and Q. Tian, "Structural relational reasoning of point clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 949–958, Long Beach, CA, USA, June 2019.

[41] B. S. Hua, Q. H. Pham, D. T. Nguyen, M. -K. Tran, L. -F. Yu, and S. -K. Yeung, "Scenenn: a scene meshes dataset with annotations," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. No.01, pp. 8778–8785, Honolulu, HI, USA, February 2019.

[42] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, "Scannet: richly-annotated 3d reconstructions of indoor scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5828–5839, Honolulu, HI, USA, July 2017.

Hindawi
International Journal of Antennas and Propagation
Volume 2022, Article ID 9604664, 14 pages
https://doi.org/10.1155/2022/9604664

Research Article

DOA Estimation for Non-Gaussian Signals: Three-Level Nested Array and a Successive SS-MUSIC Algorithm

Sha Shi,[1,2] **Haowei Zeng**,[1,2] **Heng Yue**,[1,2] **Changbo Ye**,[1,2] **and Jianfeng Li**[1,2]

[1]*College of Electronic Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China*
[2]*Key Laboratory of Dynamic Cognitive System of Electromagnetic Spectrum Space, Nanjing University of Aeronautics and Astronautics, Ministry of Industry and Information Technology, Nanjing 211106, China*

Correspondence should be addressed to Sha Shi; shisha@nuaa.edu.cn

Received 14 December 2021; Revised 28 February 2022; Accepted 11 March 2022; Published 7 April 2022

Academic Editor: Xianpeng Wang

Direction of arrival (DOA) estimation for non-Gaussian signals using three-level nested array (THL-NA) is investigated in this paper. Motivation from larger consecutive degree of freedom (DOF) and array aperture, the THL-NA is proposed, which can take full advantages of the consecutive coarrays of TL-NA and has the closed-form expression of DOF. Specifically, firstly, the array aperture is expanded by the second order sum coarray (2-SC) of the proposed array, secondly, the nested relationship between subarrays is employed to obtain the difference coarray of 2-SC (2-DCSC), finally, a consecutive virtual array with large array aperture is obtained. Besides, a successive SS-MUSIC algorithm is proposed, which employs the spatial smoothing estimating signal parameter via rotational invariance techniques (SS-ESPRIT) algorithm and partial spectrum searching multiple signal classification (PSS-MUSIC) to obtain initial estimations and fine estimations, respectively, resulting in a better balance between computational complexity and estimation accuracy.

1. Introduction

Direction of arrival (DOA) estimation, which is one of the fundamental issues in array signal processing, plays an important role in various fields, e.g., radar systems, acoustics, navigation, and wireless communications [1–3].

Compared with traditional uniform linear arrays (ULAs), sparse arrays can obtain less mutual coupling and higher degrees of freedom (DOF), which can obtain DOA estimation for signals with more sources. The nested array (NA) [4] is a sparse array composed of two ULAs with different spacings, which has strong expansion capability in DOA estimation algorithms based on FOC. The coprime array (CPA) [5, 6], consisting of two ULAs with coprime interelement spacings, has less mutual coupling compared with NA. These arrays can generate sets of uniformly distributed virtual second order difference coarray (2-DC) [4], in particular, the 2-DC of NA is free from holes. However, these sparse arrays have limitations due to their array geometries. The 2-DC of CPA generates a lot of holes that decrease consecutive DOF significantly, the dense part of NA leads to more serious mutual coupling than CPA.

Traditional DOA parameter estimation algorithms, such as multiple signal classification (MUSIC) algorithm [7, 8] and estimation of signal parameters via rotational invariance techniques (ESPRIT) algorithm [9, 10], mostly utilize the second order statistical characteristics of the signals. Furthermore, new algorithms have been proposed such as 2D-MUSIC and RD-MUSIC [1]. When the signals obey the Gaussian distribution, they can be described by the first order or second order statistics. However, the situation in practical applications is more complicated, and most sources to be processed are non-Gaussian signals, whose statistical characteristics can be described by fourth order cumulant (FOC) [11, 12], resulting in larger array aperture and better DOA estimation performance, as compared with second order cumulant (SOC). Considering the non-Gaussian signals, DOA estimation methods based on FOC, including MUSIC-LIKE [13, 14] and virtual-ESPRIT algorithm [15], are most exploited. The virtual coarrays utilized by

vectorized FOC methods can be obtained from fourth order difference coarray (4-DC) [16] or difference coarray of sum coarray (2-DCSC) [17] operation of physical sensors. The FOC can not only suppress Gaussian white noise or color noise, but effectively expand the length of array apertures. Therefore, the estimation performance of DOA based on FOC is greatly improved.

Combining the designing of sparse array and the exploiting of 4-DC or 2-DCSC, lots of array structures are proposed, such as fourth-level NA (FL-NA) [16], sparse array with fourth order difference coarray enhancement based on CPA (SAFE-CPA) [18]. FL-NA is investigated in detail as a special case of 2q-level NAs when q is equal 2 in reference [16] and SAFE-CPA constructs physical sensors structure by adding another subarray. Regarding the processing of the above sparse array structure, related scholars have proposed a series of FOC estimation algorithms based on spatial smoothing subspace methods [19, 20] such as SS-MUSIC algorithm, SS-ESPRIT algorithm, which provide a theoretical basis for the DOA estimation based on sparse arrays.

In this paper, a novel sparse array structure, named three-level nested array (THL-NA), is proposed for non-Gaussian incident signals, which can take full advantages of the consecutive coarrays of TL-NA and has the closed-form expression of DOF. Firstly, the 2-SC [21] of the proposed array is obtained to expand the array aperture, and then the consecutive 2-DCSC of the proposed array is obtained by employing the nested relationship between subarrays of 2-SC. To make a better balance between computational complexity and estimation accuracy, a successive SS-MUSIC algorithm is proposed, which employs the spatial smoothing ESPRIT (SS-ESPRIT) algorithm and partial spectrum searching MUSIC (PSS-MUSIC) to obtain initial estimates and fine estimates, respectively.

The three contributions of this paper are extracted as follows:

(1) From the viewpoint of constructing sum or difference coarray, the THL-NA based on the FOC is proposed to obtain large consecutive DOF and array aperture, which enhances the DOA estimation performance.

(2) From the perspective for expressing the length of the virtual array in the THL-NA, this paper derives the closed-form expressions with consecutive DOF and discusses the optimal array configuration to achieve the largest consecutive DOF.

(3) In terms of making a balance between computational complexity and estimation accuracy, a successive SS-MUSIC algorithm is proposed, which employs SS-ESPRIT algorithm and PSS-MUSIC to obtain initial angle estimations and fine angle estimations, respectively.

The chapter arrangement of this paper is as follows. In Section 2, we introduce 2-DC, 2-SC, 2-DCSC, 4-DC, and the properties of NA. We elaborate the array configuration of THL-NA and explain the closed-form expression in Section 3. The proposed algorithm called successive SS-MUSIC algorithm is presented in Section 4. Section 5 analyzes the performance of THL-NA and the computational complexity of the successive SS-MUSIC algorithm. Section 6 provides lots of simulations and the conclusions are drawn in Section 7.

Notations. Throughout the paper, matrices are expressed by upper-case bold characters and vectors are denoted by lower-case bold characters, respectively. $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^{-1}$, and $(\cdot)^*$ imply the transpose, the conjugate transpose operation, inverse and complex conjugation of a vector or matrix, respectively. $\otimes$, $\odot$, and $\oplus$ stand for the Kronecker product, Khatri-Rao product and Hadamard product, respectively. $\mathrm{vec}(\cdot)$ represents the vectorization operation and $\mathrm{cum}(\cdot)$ indicates the cumulant operator. $\mathrm{angle}(\cdot)$ signifies the phase operator and $\|\cdot\|_F$ shows the Frobenius norm. $\arcsin(\cdot)$ means the arcsine function.

2. Preliminaries

In this section, the 2-DC, 2-SC, 2-DCSC, 4-DC, and the properties of NA are introduced.

2.1. The Definitions of 2-DC, 2-SC, 4-DC, and 2-DCSC. Consider a P -sensors linear array, whose locations set can be indicated as [22]

$$S = \{v_1 \cdot d, v_2 \cdot d, \ldots, v_P \cdot d\}, \tag{1}$$

where $v_p \cdot d$ represents the position of the $p-th$ sensor, $p = 1, 2, \ldots, P$, and d is the unit spacing.

Definition 1. For the array with the set of sensor position S in equation (1), the 2-DC set is defined as [17]

$$C_{2-\mathrm{DC}} = \Phi_{2-\mathrm{DC}} d, \tag{2}$$

where the set of 2-DC lags

$$\Phi_{2-\mathrm{DC}} = \{v_{p1} - v_{p2}, 1 \le p1, p2 \le P\}. \tag{3}$$

Definition 2. The 2-SC set is defined as [21]

$$C_{2-\mathrm{SC}} = \Phi_{2-\mathrm{SC}} d, \tag{4}$$

with the set of 2-SC lags

$$\Phi_{2-\mathrm{SC}} = \{v_{p1} + v_{p2}, 1 \le p1, p2 \le P\}. \tag{5}$$

Definition 3. The 4-DC set is defined as [16]

$$C_{4-\mathrm{DC}} = \Phi_{4-\mathrm{DC}} d, \tag{6}$$

where the set of 4-DC lags

$$\Phi_{4-\mathrm{DC}} = \{(v_{p1} - v_{p2}) - (v_{p3} - v_{p4}), \quad 1 \le p1, p2, p3, p4 \le P\}. \tag{7}$$

Definition 4. The 2-DCSC set is defined as [17]

$$C_{2-\mathrm{DCSC}} = \Phi_{2-\mathrm{DCSC}} d, \tag{8}$$

where the set of 2-DCSC lags

$$\Phi_{2-\mathrm{DCSC}} = \left\{\left(v_{p1} + v_{p2}\right) - \left(v_{p3} + v_{p4}\right), \quad 1 \le p1, p2, p3, p4 \le P\right\}. \tag{9}$$

By permutation invariance, $\Phi_{2-\mathrm{DCSC}}$ can be rewritten as

$$\Phi_{2-\mathrm{DCSC}} = \left\{\left(v_{p1} - v_{p2}\right) - \left(v_{p3} - v_{p4}\right)\right\} = \left\{\left(v_{p1} + v_{p4}\right) - \left(v_{p2} + v_{p3}\right)\right\} = \Phi_{4-\mathrm{DC}}. \tag{10}$$

for $1 \le p1, p2, p3, p4 \le P$. That is, for a particular array, the 4-DC is equivalent to the 2-DCSC

$$C_{4-\mathrm{DC}} = C_{2-\mathrm{DCSC}}. \tag{11}$$

2.2. The Properties of TL-NA. The configuration of TL-NA [4] has been shown in Figure 1, which contains two sparse uniform subarrays. The first subarray has M_1 sensors with interelement spacing d, where $d = \lambda/2$ and λ is the wavelength, while the another subarray has M_2 sensors with interelement spacing $(M_1 + 1)d$. The position of TL-NA varies from 1 d to $M_2(M_1 + 1)d$ and it can be represented as

$$S = \{m_1 d, 1 \le m_1 \le M_1\} \cup \{m_2 (M_1 + 1)d, 1 \le m_2 \le M_2\}. \tag{12}$$

Refer to Definition 1, the 2-DC location set of TL-NA can be written as

$$C_{2-\mathrm{DC}} = \{-M_2 (M_1 + 1) + 1, \ldots, -1, 0, 1, \ldots, M_2 (M_1 + 1) - 1\} d. \tag{13}$$

Refer to Definitions 3 and4, the 4-DC (2-DCSC) location set of TL-NA can be expressed as [23]

$$C_{4-\mathrm{DC}} = C_{2-\mathrm{DCSC}} = \{-2M_2 (M_1 + 1) + 2, \ldots, -1, 0, 1, \ldots, 2M_2 (M_1 + 1) - 2\} d. \tag{14}$$

From equations (13) and (14), It can be concluded that the position of the 2-DC and 2-DCSC of TL-NA are consecutive.

2.3. Data Model Based on FOC. Assume K far-field narrowband uncorrelated sources impinging upon the array with locations set S from directions $\theta = [\theta_1, \ldots, \theta_K]$, the received data model of the array can be expressed as [24, 25]

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \tag{15}$$

where $\mathbf{A}(\theta) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \ldots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{P \times K}$ represents the steering matrix of array, and the steering vector at direction θ_k, which denotes the elevation angle of the $k - th$ target, is given by

$$\mathbf{a}(\theta_k) = \left[e^{-j2\pi v_1 d \sin\theta_k/\lambda}, e^{-j2\pi v_2 d \sin\theta_k/\lambda}, \ldots, e^{-j2\pi v_p d \sin\theta_k/\lambda}\right]^T \in \mathbb{C}^{P \times 1}. \tag{16}$$

and $\mathbf{s}(t) = [\mathbf{s}_1(t), \mathbf{s}_2(t), \ldots, \mathbf{s}_K(t)]^T \in \mathbb{C}^{K \times 1}$, $1 \le t \le L$ denotes non-Gaussian signal source matrix with zero mean, where L indicates the number of snapshots. And, $\mathbf{n}(t)$ is the received additive white Gaussian noise with mean zero and variance σ^2.

In this paper, for given $x_{k1}, x_{k2}, x_{k3}^*, x_{k4}^*$, the FOC definition can be written as [26]

Figure 1: Two-level nested array.

$$\begin{aligned}\mathbf{C}_4 &= \mathrm{Cum}\left(x_{k1}, x_{k2}, x_{k3}^*, x_{k4}^*\right)\\ &= E\left(x_{k1}x_{k2}x_{k3}^*x_{k4}^*\right) - E\left(x_{k1}x_{k2}\right)E\left(x_{k3}^*x_{k4}^*\right) - E\left(x_{k1}x_{k3}^*\right)E\left(x_{k2}x_{k4}^*\right) - E\left(x_{k1}x_{k4}^*\right)E\left(x_{k2}x_{k3}^*\right).\end{aligned} \tag{17}$$

For the array, the FOC of the received signal can be [26]

$$\begin{aligned}\mathbf{C}_{4,X} &= \mathrm{Cum}\left(\mathbf{X}, \mathbf{X}, \mathbf{X}^*, \mathbf{X}^*\right)\\ &= E\left[\left(\mathbf{X}\otimes\mathbf{X}^*\right)\left(\mathbf{X}\otimes\mathbf{X}^*\right)^H\right] - E\left(\mathbf{X}\otimes\mathbf{X}^*\right)E\left[\left(\mathbf{X}\otimes\mathbf{X}^*\right)^H\right] - E\left(\mathbf{X}\mathbf{X}^H\right)\otimes E\left[\left(\mathbf{X}\mathbf{X}^H\right)^*\right].\end{aligned} \tag{18}$$

The FOC matrix of the received signal is expressed by [26]

$$\mathbf{R}_4 = \mathbf{C}_{4,X} \tag{19}$$

where

$$\begin{aligned}\mathbf{B}(\theta) &= \left[\mathbf{b}(\theta_1), \mathbf{b}(\theta_2), \ldots, \mathbf{b}(\theta_K)\right]\\ &= \left[\mathbf{a}(\theta_1)\otimes\mathbf{a}^*(\theta_1), \mathbf{a}(\theta_2)\otimes\mathbf{a}^*(\theta_2), \ldots, \mathbf{a}(\theta_K)\otimes\mathbf{a}^*(\theta_K)\right]\\ &= \mathbf{A}(\theta)\odot\mathbf{A}^*(\theta),\\ \mathbf{C}_{\mathbf{S}} &= E\left[\left(\mathbf{S}\otimes\mathbf{S}\right)\left(\mathbf{S}\otimes\mathbf{S}\right)^H\right] - E\left(\mathbf{S}\otimes\mathbf{S}\right)E\left[\left(\mathbf{S}\otimes\mathbf{S}\right)^H\right] - E\left(\mathbf{S}\mathbf{S}^H\right)\otimes E\left[\left(\mathbf{S}\mathbf{S}^H\right)^*\right].\end{aligned} \tag{20}$$

$\mathbf{B}(\theta)$ is the steering matrix after array manifold ($\mathbf{A}(\theta)\Rightarrow\mathbf{B}(\theta)$) by using the FOC. To further improve the performance of DOA estimation, we use a vectorized method [23, 26, 27] for $\mathbf{R}_4$ to obtain a vector, which is equivalent to the received signal from the virtual ULA.

$$\mathbf{z} = \mathrm{vec}\left(\mathbf{R}_4\right) \tag{21}$$

where

$$\begin{aligned}\mathbf{\Lambda}(\theta) &= \left[\mathbf{b}^*(\theta_1)\otimes\mathbf{b}(\theta_1), \mathbf{b}^*(\theta_2)\otimes\mathbf{b}(\theta_2), \ldots, \mathbf{b}^*(\theta_K)\otimes\mathbf{b}(\theta_K)\right]\\ &= \mathbf{B}^*(\theta)\odot\mathbf{B}(\theta),\end{aligned} \tag{22}$$

and $\mathbf{p} = [c_{4,s_1}, c_{4,s_2}, \ldots, c_{4,s_K}]^T$ represents the FOC matrix of source vector, and c_{4,s_k} represents the FOC of the $k-th$ source vector.

2.4. Data Model with Mutual Coupling Based on FOC. Equation (16) assumes that the sensors in the array has not interference with each other. Actually, the output of each sensor is influenced by its adjacent elements owing to the presence of mutual coupling. Therefore, by introducing a mutual coupling matrix $\mathbf{C}$, the received data model can be rewritten as follows [28, 29].

$$\mathbf{x}(t) = \mathbf{CAs}(t) + \mathbf{n}(t), \tag{23}$$

where the mutual coupling $\mathbf{C}$ can be obtained from reference [28], whose define is determined by varieties of factors involving in the distance between sensors, the operating frequency, e.g., and it can be approximated by employing the B-banded model [30–32].

$$\mathbf{C}_{i,j} = \begin{cases} 0, & \left|v_i d - v_j d\right| > B,\\ c_{\left|v_i d - v_j d\right|}, & \left|v_i d - v_j d\right| \le B,\end{cases} \tag{24}$$

where $v_i d, v_j d \in S$ and $1 = c_0 > |c_1| > \cdots > |c_B| > |c_{B+1}| = 0$, $c_1 = 0.3e^{j\pi/3}$, $c_l = c_1 e^{-j(l-1)\pi/8}/l$, for $l \in [2, B]$, $B = 100$ represents the maximum spacing of sensor pairs with mutual coupling. Besides, for a given array, the total strength of the

mutual coupling effect can be measured by coupling leakage as [22, 28]

$$L(\mathbf{M}) = \frac{\|\mathbf{C} - \mathrm{diag}\{\mathbf{C}\}\|_F}{\|\mathbf{C}\|_F}. \tag{25}$$

Based on the mutual coupling model in equation (24), the received signal from the virtual array in equation (21) can be reconstructed as

$$\tilde{z} = \mathbf{C}_{\mathrm{vec}}\mathbf{\Lambda}(\theta)\mathbf{p}, \tag{26}$$

where $\mathbf{C}_{\mathrm{vec}} = (\mathbf{C} \otimes \mathbf{C}^*)^* \otimes (\mathbf{C} \otimes \mathbf{C}^*)$.

3. The Proposed Array

In this section, the proposed array is specifically introduced, including the position of physical sensors, the derivation of the consecutive elements part in the 2-SC and the 2-DCSC, and the properties of it. Next, we analyze the 2-SC and the 2-DCSC of the array for given parameters, and reveal the influence of different array configurations on the consecutive DOF of the array with the same total number of physical sensors.

3.1. Array Structure. Motivation from larger consecutive DOF and array aperture, the THL-NA is proposed, which has a concrete closed expression form with consecutive DOF. The 2-SC of the proposed array is employed to enlarge the array aperture, and then a consecutive virtual array with large array aperture is obtained by performing the difference operation on the 2-SC, whose consecutive character utilizes the nested relationship between subarrays in the 2-SC. Next, the position of physical sensors in the proposed array is given.

Proposition 1. *The THL-NA is composed of three ULAs with a dense ULA and two sparse ULAs, which is shown in Figure 2. The position of THL-NA sensors can be represented as*

$$\begin{aligned} S_1 &= \{n_1 d, 1 \le n_1 \le N_1\}, \\ S_2 &= \{n_2 (N_1 + 1)d, 1 \le n_2 \le N_2\}, \\ S_3 &= \{n_3 (N_2 + 1)(N_1 + 1)d, 1 \le n_3 \le N_3\}, \\ S_{\mathrm{THL-NA}} &= S_1 \cup S_2 \cup S_3. \end{aligned} \tag{27}$$

Figure 2 shows the distribution position of physical sensors in the proposed array. The array consists of three ULAs. The number of sensors is N_1 with interelement spacing d starting from 1 d in the first subarray, the number of sensors is N_2 with interelement spacing $(N_1 + 1)d$ starting from $(N_1 + 1)d$ in the second subarray and the number of sensors is N_3 with interelement spacing $(N_2 + 1)(N_1 + 1)d$ starting from $(N_2 + 1)(N_1 + 1)d$ in the third subarray. The total number of physical sensors is N and $N = N_1 + N_2 + N_3$.

3.2. The Deriving of Closed-Form Expression with Consecutive Elements. The 2-SC of the proposed array ranges from 2 d to $2N_3(N_2 + 1)(N_1 + 1)d$, which is not a virtual consecutive array. It is worth noting that there are N_3 TL-NAs and a THL-NA with the same array structure in the 2-SC. In fact, the 2-SC can be regarded as the panning of a certain position of the proposed array. For instance, the THL-NA at the last part of the 2-SC can be regarded as the number of elements position in the physical array plus the largest number of elements position, which can be understood as the panning of the initial physical array.

Proposition 2. *The 2-SC can be divided into $N_3 + 1$ parts. The 2-SC position of THL-NA can be represented as*

$$\begin{aligned} S_1 &= \{n_1 d, 2 \le n_1 \le (N_1 + 1)(N_2 + 1)\}, \\ S_2 &= (N_1 + 1)(N_2 + 1)d + \{n_1 d, 1 \le n_1 \le N_1\} \cup \{n_2 (N_1 + 1)d, 1 \le n_2 \le N_2 + 1\}, \\ S_3 &= 2(N_1 + 1)(N_2 + 1)d + \{n_1 d, 1 \le n_1 \le N_1\} \cup \{n_2 (N_1 + 1)d, 1 \le n_2 \le N_2 + 1\}, \\ &\vdots \\ S_{N_3+1} &= N_3 (N_1 + 1)(N_2 + 1)d + \{n_1 d, 1 \le n_1 \le N\}_1 \cup \{n_2 (N_1 + 1)d, 1 \le n_2 \le N_2\} \cup \{n_3 (N_2 + 1)(N_1 + 1)d, \quad 1 \le n_3 \le N_3\}, \\ S_{2-SC} &= S_1 \cup S_2 \cup \cdots \cup S_{N_3+1}. \end{aligned} \tag{28}$$

According to the consecutive character of the 2-DC of the TL-NA, it can be seen that a virtual consecutive 2-DCSC is obtained by performing another difference operation on the 2-SC.

Proposition 3. *The 2-DCSC position of THL-NA can be represented as*

$$S_{2-\mathrm{DCSC}} = \{-2N_3(N_2 + 1)(N_1 + 1) + 2, \ldots, -1, 0, 1, \ldots, 2N_3(N_2 + 1)(N_1 + 1) - 2\}. \tag{29}$$

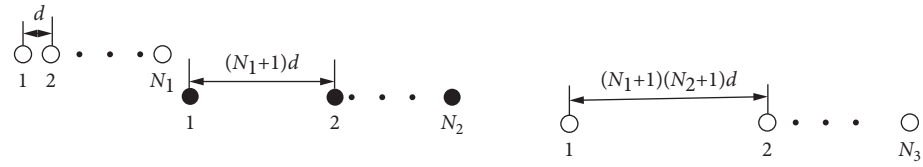


Figure 2: Three-level nested array structure.

Summarize the closed-form expression with the number of the consecutive lags of the array structure based on FOC as follows:

$$\begin{aligned} c\text{DOF} &= 2\left[2N_3\left(N_2+1\right)\left(N_1+1\right)-2\right]+1 \\ &= 4N_3\left(N_2+1\right)\left(N_1+1\right)-3. \end{aligned} \tag{30}$$

Next, a specific array configuration example is used to deepen understanding. Figure 3 shows when the array configuration satisfies $N_1 = 3, N_2 = 3, N_3 = 4$, the position of physical sensors in the proposed array, the position of array elements in the 2-SC and 2-DCSC. Next, the specific example is used to analyze and discuss the structure and properties of the proposed array.

The location of the physical sensors is distributed in $\{0.5, 1, 1.5, 2, 4, 6, 8, 16, 24, 32\}$, The 2-SC of the array ranges from $2\ d$ to $2N_3\left(N_2+1\right)\left(N_1+1\right)d$ namely 1 to 64 with $d = 0.5$. According to the above analysis, the 2-SC is divided into $N_3 + 1 = 5$ parts, and the position of coarray elements in each part can be expressed as

$$\begin{aligned} S_1 &= \left\{n_1 d, 2 \le n_1 \le \left(N_1+1\right)\left(N_2+1\right)\right\} \\ S_2 &= \left(N_1+1\right)\left(N_2+1\right)d + \left\{n_1 d, 1 \le n_1 \le N_1\right\} \cup \left\{n_2\left(N_1+1\right)d, 1 \le n_2 \le N_2+1\right\} \\ &= 8 + \{0.5, 1, 1.5, 2, 4, 6, 8\}, \\ S_3 &= 2\left(N_1+1\right)\left(N_2+1\right)d + \left\{n_1 d, 1 \le n_1 \le N_1\right\} \cup \left\{n_2\left(N_1+1\right)d, 1 \le n_2 \le N_2+1\right\} \\ &= 16 + \{0.5, 1, 1.5, 2, 4, 6, 8\}, \\ S_4 &= 3\left(N_1+1\right)\left(N_2+1\right)d + \left\{n_1 d, 1 \le n_1 \le N_1\right\} \cup \left\{n_2\left(N_1+1\right)d, 1 \le n_2 \le N_2+1\right\} \\ &= 24 + \{0.5, 1, 1.5, 2, 4, 6, 8\}, \\ S_5 &= 4\left(N_1+1\right)\left(N_2+1\right)d \\ &\quad + \left\{n_1 d, \quad 1 \le n_1 \le N_1\right\} \cup \left\{n_2\left(N_1+1\right)d, \quad 1 \le n_2 \le N_2\right\} \cup \left\{n_3\left(N_2+1\right)\left(N_1+1\right)d, \quad 1 \le n_3 \le N_3\right\} \\ &= 32 + \{0.5, 1, 1.5, 2, 4, 6, 8, 16, 24, 32\}, \\ S_{2-\text{SC}} &= S_1 \cup S_2 \cup \cdots \cup S_5. \end{aligned} \tag{31}$$

There is nested relationship among subarrays circled in the black box in Figure 3. The elements in the position set S_2, S_3, S_4 construct NA structure respectively, which can be regarded as TL-NA. The last part of the 2-SC is structurally equivalent to the proposed array, which can be regarded as a THL-NA. The consecutive 2-DCSC is obtained by performing a difference operation on the 2-SC. Since the 2-DC of the TL-NA is consecutive, the 2-DC of the sets S_2, S_3, S_4 is equivalent to the 2-DCSC of the proposed array, which is a consecutive virtual array. From Figure 3, it can be observed that the longest length of the consecutive holes is 8, and the longest length of the consecutive elements is 9. By performing the difference operation on set S_5, a consecutive array is obtained. As a result, a consecutive virtual array can be obtained through the above array processing. The length of the consecutive DOF is $c\text{DOF} = 4N_3\left(N_2+1\right)\left(N_1+1\right) - 3 = 253$.

Figure 4 shows when the array configuration satisfies $N_1 = 3, N_2 = 3, N_3 = 3$, the position of physical sensors in the proposed array, the position of array elements in the 2-SC and 2-DCSC. According to the above analysis, the consecutive DOF is $c\text{DOF} = 4N_3\left(N_2+1\right)\left(N_1+1\right) - 3 = 189$.

3.3. Optimal Array Configuration. For a fixed total number of physical sensors, there are multiple schemes about allocating sensors of each subarray in THL-NA. The problem of determining the optimal array configuration is addressed to maximize the consecutive DOF in this section. According to Proposition 3, the length of consecutive virtual elements in THL-NA is $4N_3\left(N_2+1\right)\left(N_1+1\right) - 3$. The consecutive DOF optimization problem can be formulated as

$$\begin{aligned} &\max, 4N_3N_2N_1 + 4N_3N_2 + 4N_3N_1 + 4N_3 - 3, \\ &\text{subject to}, N_3 + N_2 + N_1 = N. \end{aligned} \tag{32}$$

In equation (32), we focus on how to configure the number of sensors of each subarray (N_1, N_2, N_3), so that the possible largest length of consecutive lags is obtained from

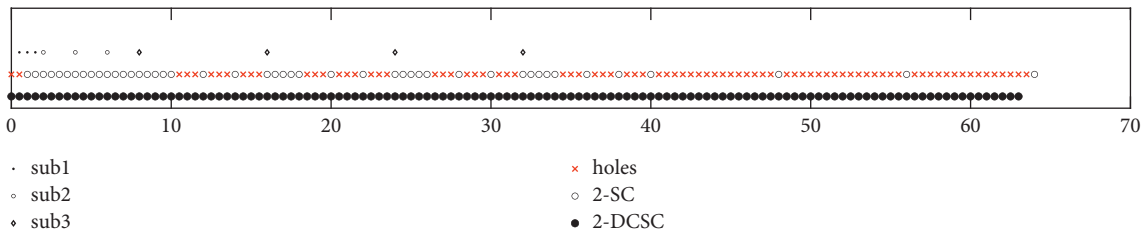


Figure 3: The location of physical sensors, 2-SC, 2-DCSC when $N_1 = 3, N_2 = 3, N_3 = 4$.

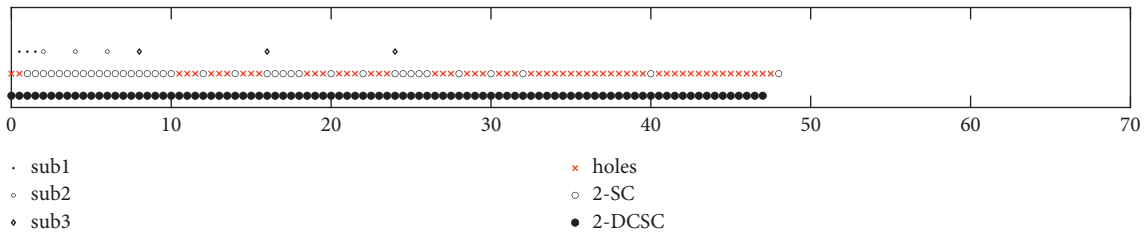


Figure 4: The location of physical sensors, 2-SC, 2-DCSC when $N_1 = 3, N_2 = 3, N_3 = 3$.

the virtual ULA when the total number of array sensors N is certain. We define integers m and n to be the remainder and quotient of N modulo 3, where $N = 3n + m$, $0 \le m \le 2$.

Proposition 4. *One solution to the optimization problem in equation (32) is can be presented as*

$$\begin{cases} N_3 = N_2 = N_1 = n, & \text{if } m = 0, \\ N_3 = n + 1, \quad N_2 = N_1 = n, & \text{if } m = 1, \\ N_3 = N_2 = n + 1, \quad N_1 = n, & \text{if } m = 2. \end{cases} \tag{33}$$

According to equation (33), the corresponding consecutive DOF can be evinced as

$$c - \text{DOF} = \begin{cases} 4n(n+1)(n+1) - 3, & \text{if } m = 0, \\ 4(n+1)(n+1)(n+1) - 3, & \text{if } m = 1, \\ 4(n+1)(n+2)(n+1) - 3, & \text{if } m = 2, \end{cases} \tag{34}$$

where $n = [N/3]$.

Table 1 shows the consecutive degree of freedom corresponding to different array configurations when the number of physical sensors is 9. Proposition 4 can be verified with specific examples in Table 1.

4. The Proposed Algorithm

The computational complexity of the FOC matrix and that of the conventional MUSIC method due to the global spectrum searching are expensive, the successive SS-MUSIC algorithm is presented, which employs partial spatial spectrum peak searching to drop the complexity. The SS-ESPRIT algorithm for the consecutive 2-DCSC of the THL-NA is utilized to obtain the initial estimates, which can be used to shrink the searching range of MUSIC algorithm to obtain the fine estimates.

4.1. Initial Estimates. From the previous discussion, there is a consecutive virtual array of the THL-NA with the range of elements $[-Wd, Wd]$, where $W = 2N_3(N_2 + 1)(N_1 + 1) - 2$, and the length of the virtual array is $T = 2W + 1$. Data model of the part refer to Section 2.3, the steering vector of the virtual array $\overset{*}{\mathbf{a}}(\theta_k) \in C^{T\times 1}$ is denoted as

$$\overset{*}{\mathbf{a}}(\theta_k) = \left[e^{-j2\pi(-W)\sin\theta_k/\lambda}, \ldots, 0, \ldots, e^{-j2\pi W \sin\theta_k/\lambda}\right]^T. \tag{35}$$

According to the position of the virtual array, remove the redundant part of $\mathbf{z}$ and sort $\mathbf{z}$, then construct $\mathbf{z}_1$.

$$\mathbf{z}_1 = \mathbf{\Lambda}_1 \mathbf{p}, \tag{36}$$

where $\Lambda_1 \in \mathbb{C}^{T\times K}$ is the direction matrix through sorting it after removing the redundant parts of $\Lambda(\theta)$, and it is mathematical form can be represented by

$$\mathbf{\Lambda}_1 = \begin{bmatrix} e^{-j(2\pi/\lambda)(-W)d\sin\theta_1} & e^{-j(2\pi/\lambda)(-W)d\sin\theta_2} & \cdots & e^{-j(2\pi/\lambda)(-W)d\sin\theta_K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j(2\pi/\lambda)Wd\sin\theta_1} & e^{-j(2\pi/\lambda)(-W)d\sin\theta_2} & \cdots & e^{-j(2\pi/\lambda)Wd\sin\theta_K} \end{bmatrix}. \tag{37}$$

Table 1: C-DOF of different array configurations.

(N_1, N_2, N_3)	3, 3, 3	2, 3, 4	3, 2, 4	2, 4, 3	4, 2, 3	4, 3, 2	3, 4, 2	2, 2, 5	5, 2, 2	2, 5, 2
c-DOF	189	189	189	177	177	157	157	177	141	141

Under the situation, the traditional DOA algorithms, which use the eigenvalue decomposition of covariance, is invalid because of the vectorized interrelated signal. The received signal matrix is updated by intercepting the consecutive part of $\mathbf{z}_1$. After obtaining the consecutive vector $\mathbf{z}_1$, the spatial smoothing algorithm [19, 20], which divides the ULA into several overlapping subarrays, is exploited to take the place of the invalid traditional DOA algorithms. After obtaining the sum of the covariance matrices of the divided subarrays, which have the same array structure, the spatial smoothing covariance matrix is obtained by taking the average of it. As shown in Figure 5, $W+1$ overlapping subarrays are obtained by dividing equally the virtual array, and each subarray incorporates $W+1$ elements. Where, the sensors position of the $i-th$ subarray is

$$\{(-i+1+n)d, n=0,1,\ldots,W\}. \tag{38}$$

The received signal matrix $\mathbf{z}_{1i}$ is from row $W+2-i$ to row $2W+2-i$ of $\mathbf{z}_1$.Construct the covariance matrix:

$$\mathbf{R}_i = \mathbf{z}_{1i}\mathbf{z}_{1i}^H. \tag{39}$$

To obtain the spatial smoothing covariance matrix $\mathbf{R}$, which has the same form as the signal covariance matrix based on classical subspace algorithms, we sum the covariance matrices of all $W+1$ subarrays and calculate the mean of it as shown in

$$\mathbf{R} = \frac{1}{W+1}\sum_{i=1}^{W+1}\mathbf{R}_i. \tag{40}$$

In the following, the results of DOA estimation are obtained by using the classic ESPRIT algorithm [33]. Perform eigenvalue decomposition on $\mathbf{R}$ to obtain the signal subspace $\mathbf{U}_N$ and the noise subspace $\mathbf{U}_S$.

$$\mathbf{R} = \mathbf{U}_S\sum_S\mathbf{U}_S^H + \mathbf{U}_N\sum_N\mathbf{U}_N^H. \tag{41}$$

The rotation invariance of array makes the matrix $\mathbf{U}_S$ decomposed into $\mathbf{U}_X \in \mathbb{C}^{(W-1)\times K}$, $\mathbf{U}_Y \in \mathbb{C}^{(W-1)\times K}$,corresponding to two subarrays, namely,

$$\mathbf{U}_S = \begin{bmatrix}\mathbf{U}_X\\ \mathbf{U}_Y\end{bmatrix} = \begin{bmatrix}\mathbf{U}_X\\ \mathbf{U}_X\boldsymbol{\Psi}\end{bmatrix},$$
$$\boldsymbol{\Psi} = \mathbf{T}^{-1}\boldsymbol{\Phi}\mathbf{T}. \tag{42}$$

The signal subspace $\mathbf{U}_X$ is similar to $\mathbf{U}_Y$, and the diagonal elements of Φ are equal to the eigenvalues of Ψ, so by calculating the eigenvalues λ_k of $\Psi = \mathbf{U}_X^+\mathbf{U}_Y$, where $(k=1,2,\ldots,K)$, the diagonal elements of Φ is estimated to obtain the angle estimation value:

$$\widehat{\theta}_k^{ini} = \arcsin\left(\text{angle}\frac{(\lambda_k)}{2\pi d}\right). \tag{43}$$

4.2. Fine Estimates. The initial estimated angles $\widehat{\theta}_k^{ini}$ are employed to shrink the range of spectrum searching and the MUSIC algorithm is utilized to perform interval processing on the spatial spectrum function, where the interval range is expressed by $[\widehat{\theta}_k^{ini}-\Delta, \widehat{\theta}_k^{ini}k+\Delta]$ (Δ is a small number), and then the more accurate angle estimation parameters are obtained.

The spatial spectrum function of MUSIC algorithm can be constructed as

$$P_{\text{MUSIC}} = \frac{1}{\mathbf{a}^H(\theta)\mathbf{U}_N\mathbf{U}_N^H\mathbf{a}(\theta)}, \tag{44}$$

where $\mathbf{a}(\theta)$ is the direction vector, and $\mathbf{a}(\theta) = [1, e^{-j2\pi\sin\theta/\lambda}, \ldots, e^{-j2\pi(W-1)\sin\theta/\lambda}]^T \in \mathbb{C}^{W\times 1}$. While changing the value of θ, and performing the spectrum searching in the space domain. When the denominator of the spectral function tends to 0, in where the spatial spectral function reaches a peak, the noise vector is orthogonal to the signal vector, and the signal arrival angle $\widehat{\theta}_k$ is equivalent as θ at this time.

4.3. The Advantages of the Proposed Algorithm. The main advantages of the proposed algorithm can be summarized as follows:

(1) The computational complexity of the successive SS-MUSIC algorithm is much lower than SS-MUSIC algorithm

(2) The successive SS-MUSIC algorithm owns approximately the same DOA estimation performance as the SS-MUSIC algorithm

(3) The successive SS-MUSIC algorithm makes a balance between computational complexity and estimation accuracy

5. Performance Analysis

Aiming to evaluate the performance of the proposed array geometry and the proposed algorithm, the array performance index, including the consecutive DOF, the array aperture, the SS-MUSIC spectrum, the coupling leakage e.g., and the computational complexity are compared in this section.

5.1. Analysis of the Proposed Array. Direction of arrival In this part, we compare the consecutive DOF of ACA, SAFE-CPA, TL-NA, FL-NA, and THL-NA, as shown in Figure 6. Figure 6 demonstrates that the THL-NA has a higher consecutive DOF than other arrays. It is worth noting that when the number of array elements exceeds 12, the consecutive DOF of the THL-NA begins to be lower than that of FL-NA. From the overall trend of the image, the consecutive

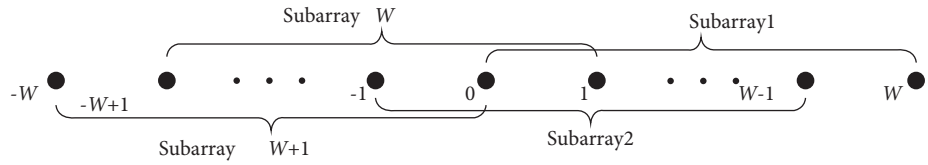


Figure 5: The spatial smoothing algorithm.

DOF of the THL-NA is close to but still greater than SAFE-CPA, which always surpasses that of FL-NA and TL-NA.

Table 2 lists the number of sensors, the consecutive DOF and the location of different arrays with 10 sensors. The length of array aperture in the proposed array is larger than that in TL-NA, ACA, which is smaller than that in FL-NA, SAFE-CPA.

$$M_0 = 4N_1N_2N_3 + 3N_1N_2 + 2N_1N_3 - N_2N_3 + N_1 - N_2 + N_3 - 1. \quad (45)$$

Table 3 shows the position, consecutive DOF, the weight distribution diagrams of the 2-SC, 2-DCSC, the SS-MUSIC spatial spectrum and mutual coupling coefficients of each array with 10 sensors in ACA, TL-NA, FL-NA, SAFE-CPA, THL-NA, and the consecutive DOF of the 2-DCSC with the THL-NA, SAFE-CPA, CPA, FL-NA, TL-NA can reach 253, 241, 85, 215, 117 respectively. The incident angles are evenly distributed between −50 and 50, SNR = 0dB, SNAPSHOTS as $L = 500$.

The comparison of the MUSIC spatial spectrum of the array structure proposed in this paper with SAFE-CPA, CPA, FL-NA, TL-NA at the same incident angles is shown as Table 3. It can be seen that the incident angle information of all signals can be effectively detected only in the spectral peak diagram of the THL-NA, while the DOA estimation based on the other four arrays fail to find target estimation and search false peaks, which fully demonstrates that the THL-NA is capable of achieving higher accuracy DOA estimation.

5.2. Analysis of the Proposed Algorithm. For evaluating the proposed algorithm, the computational complexity is introduced into the analysis in this part, which is compared among the SS-MUSIC algorithm, the SS-ESPRIT method and the proposed algorithm. Specifically, the total number of sensors is N, the number of signal source is K, L indicates the number of snapshots, and the total number of the virtual array is $T = 2W + 1$, where $W = 2N_3(N_2 + 1)(N_1 + 1) - 2$.

Based on the number of complex multiplications, the main complexities of the methods involve that, the calculation of FOC matrix needs $O(\mathrm{LN}^4)$. The virtual array is divided into $W + 1$ overlapping subarrays, and each subarray contains $W + 1$ elements. After the operation of spatial smoothing, the covariance matrix calculation of each virtual subarray needs $O\{(W+1)^3\}$ and eigenvalue decomposition of the covariance matrix requires $O\{(W+1)^3\}$, calculating Ψ needs $O\{2K^3 + 3K^2W\}$ and the eigenvalue decomposition

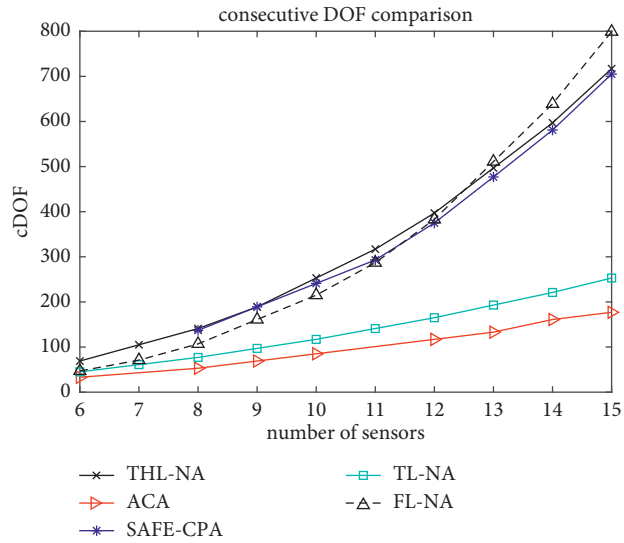


Figure 6: c-DOF comparison versus different arrays.

of Ψ requires $O\{K^3\}$. The complexity of spectral peak searching is $O\{n_1(W+1)(2(W-K)+3)\}$, where the number of searching $n_1 = 2K\Delta/0.01$, where Δ is a very small number and 0.01 is the searching accuracy.

Consequently, the total complexity of the proposed algorithm is given by $O\{\mathrm{LN}^4 + 2(W+1)^3 + 3K^2(W+K) + n_1(W+1)(2(W-K)+3)\}$, the total complexity of SS-ESPRIT algorithm is $O\{\mathrm{LN}^4 + 2(W+1)^3 + 3K^2(W+K)\}$, the total complexity of SS-MUSIC algorithm is $O\{\mathrm{LN}^4 + 2(W+1)^3 + n(W+1)(2(W-K)+3)\}$, where $n = 90/0.01$ stands for the peak search times over angle domain.

The computational complexity of algorithms is exhibited in Table 4. The histogram and the line graph are utilized to display the comparison results between the complexity of the three algorithms, as shown in Figures 7 and 8 respectively, where $K = 2$ and $L = 500$ in comparison versus different elements and $N_1 = 3, N_2 = 3, N_3 = 4$ in comparison versus different snapshots. It is observed that the complexity of the proposed algorithm is significantly lower than the SS-MUSIC method but is higher than the SS-ESPRIT algorithm.

6. RMSE Results

In this section, the root mean square error (RMSE) via 200 Monte-Carlo trials are utilized to validate the superior performance of the proposed array and the proposed algorithm by comparing with four arrays and two algorithms, respectively. The RMSE is defined by [34, 35]

Table 2: Comparison of the closed-form expression with consecutive DOF.

Arrays structure	Number of sensors ($N_i, i = 1, 2, \ldots, 4$)	Consecutive DOF ($N_i, i = 1, 2, \ldots, 4$)	Location ($T = 10$)
TL-NA	$N_1 + N_2$	$4N_2(N_1 + 1) - 3$	{0.5, 1, 1.5, 2, 2.5, 3, 6, 9, 12, 15}
FL-NA	$\sum_1^4 N_i - 3$	$2\prod_{i=1}^4 N_i - 1$	{0.5, 1, 1.5, 2, 4, 6, 12, 18, 36, 54}
ACA	$2N_1 + N_2 - 1$	$6N_1N_2 + 2N_1 - 2N_2 - 1$	{0, 1.5, 2.5, 3, 4.5, 5, 6, 7.5, 10, 12.5}
SAFE-CPA	$2N_1 + N_2 - 1 + N_3$	$2M_0 + 1$	{0, 1, 1.5, 2, 3, 4.5, 17.5, 30.5, 43.5, 56.5}
Proposed	$N_1 + N_2 + N_3$	$4N_3(N_2 + 1)(N_1 + 1) - 3$	{0.5, 1, 1.5, 2, 4, 6, 8, 16, 24, 32}

Table 3: The 2-SC, 2-DCSC, c-DOF, spectrum, and $L(M)$ comparison of different arrays.

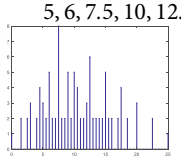
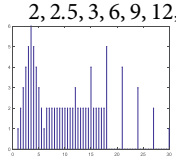
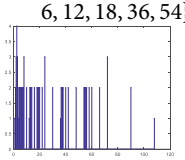
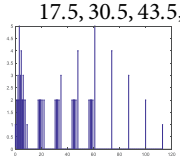
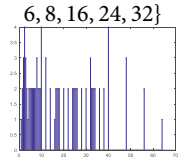
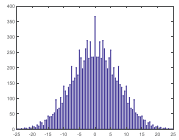
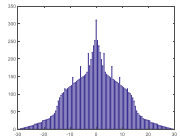
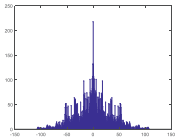
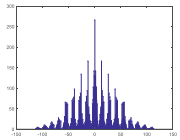
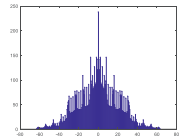
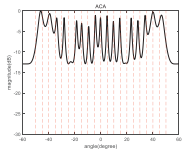
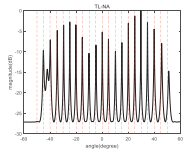
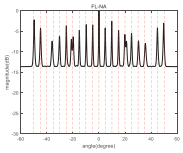
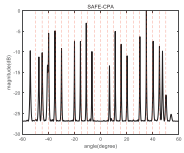
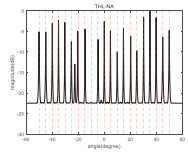
	ACA	TL-NA	FL-NA	SAFE-CPA	THL-NA
Antennas	{0, 1.5, 2.5, 3, 4.5, 5, 6, 7.5, 10, 12.5}	{0.5, 1, 1.5, 2, 2.5, 3, 6, 9, 12, 15}	{0.5, 1, 1.5, 2, 4, 6, 12, 18, 36, 54}	{0, 1, 1.5, 2, 3, 4.5, 17.5, 30.5, 43.5, 56.5}	(0.5, 1, 1.5, 2, 4, 6, 8, 16, 24, 32}
2-SC locations					
2-DCSC locations					
Spectrum					
C-DOF	85	117	215	241	253
$L(M)$	0.4582	0.6363	0.4686	0.4409	0.4728

Table 4: Complexity of different algorithms.

Algorithm	Computational complexity
Proposed	$O\{LN^4 + 2(W+1)^3 + 3K^2(W+K) + n_1(W+1)(2(W-K)+3)\}$
SS-ESPRIT	$O\{LN^4 + 2(W+1)^3 + 3K^2(W+K)\}$
SS-MUSIC	$O\{LN^4 + 2(W+1)^3 + n(W+1)(2(W-K)+3)\}$

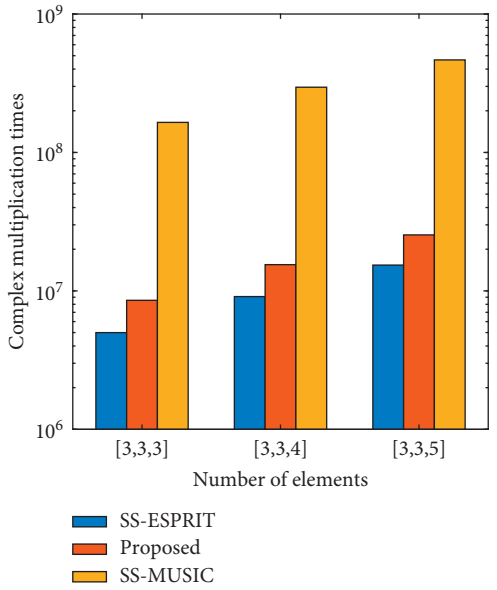


Figure 7: Complexities of different methods versus different elements.

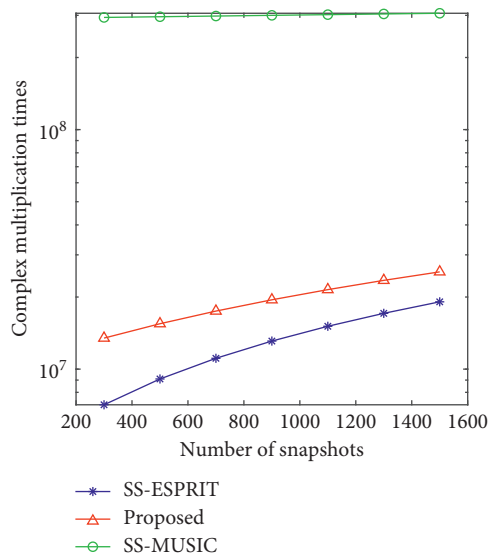


Figure 8: Complexities of different methods versus different snapshots.

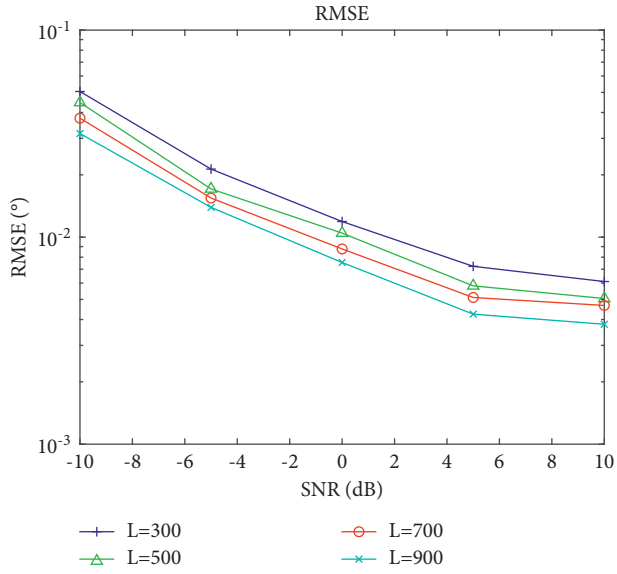


Figure 9: RMSE performance of different snapshots versus SNR.

$$\text{RMSE} = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{1}{200}\sum_{l=1}^{200}\left(\widehat{\theta}_{k,l}-\theta_k\right)^2}, \tag{46}$$

where θ_k denotes the true elevation of the $k-th$ target, $\widehat{\theta}_{k,l}$ is estimated value of θ_k in the $l-th\,(l=1,\ldots,200)$ Monte-Carlo simulation. The non-Gaussian sources with $\theta=[5^\circ,45^\circ]$ incident on the proposed array, SAFE-CPA, ACA, FL-NA, TL-NA, the total number of physical sensors is set as 10, and the consecutive DOF of the virtual coarray with the proposed array, SAFE-CPA, ACA, FL-NA, and TL-NA can reach 253, 241, 85, 215, and 117, respectively.

Sections 6.1–6.3 all employ the proposed successive SS-MUSIC algorithm to conduct the simulation experiments. Furthermore, Section 6.4 performs the comparison of different algorithms based on the proposed array geometry.

6.1. RMSE Comparison versus Snapshots. Figure 9 substantiates the effectiveness of the proposed array in estimated accuracy, and manifests that the increase of sampled data leads to the improvement of estimation performance, owing to the more accurate covariance estimation.

6.2. RMSE Comparison of Different Array Configurations. In this simulation, we compare the RMSE results of different array configurations, involving $N_1=3,N_2=3,N_3=3$, $N_1=3,N_2=3,N_3=4$ and $N_1=3,N_2=3,N_3=5$, using the successive SS-MUSIC algorithm, where SNAPSHOTS = 500. Meanwhile, the RMSE results of each algorithm versus the number of snapshots are exhibited in Figure 10. As depicted in Figure 10, with the number of elements increasing, the parameter estimation performance is enhanced because of diversity gain.

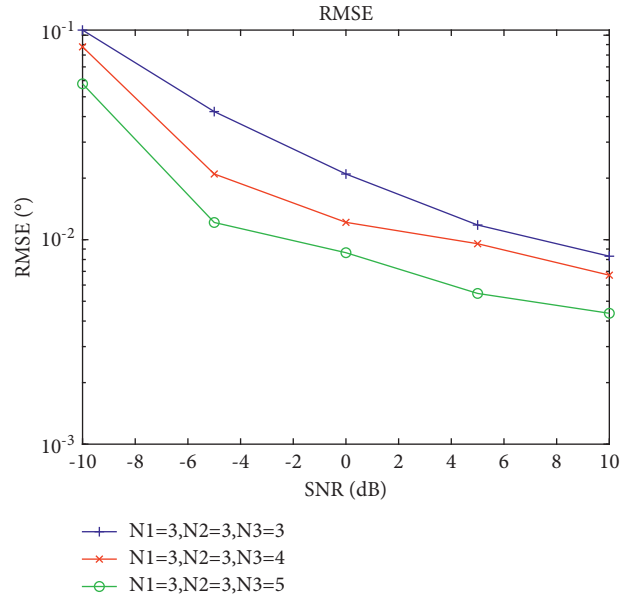


Figure 10: RMSE performance of different array configurations versus SNR.

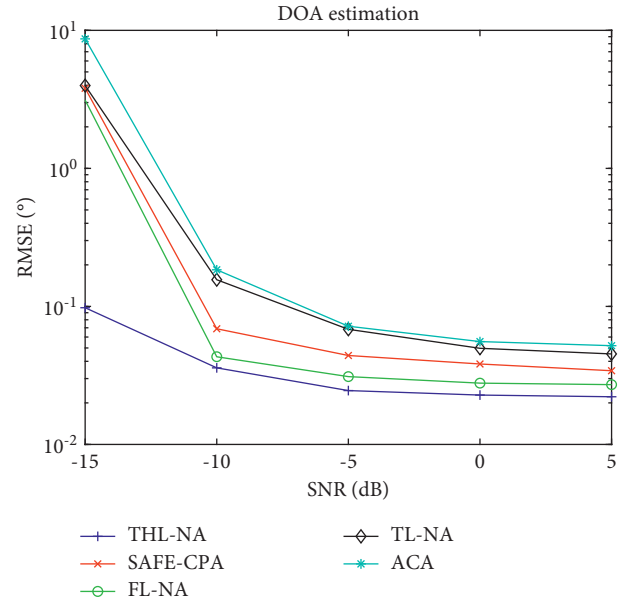


Figure 11: RMSE performance of different arrays versus SNR.

6.3. RMSE Performance Comparison of Different Arrays. To specifically examine the DOA estimation ability of the proposed array, the experiment uses 200 independent Monte Carlo experiments to statistically compares the RMSE of the array structure proposed in this paper with SAFE-CPA, ACA, FL-NA, TL-NA at different signal-to-noise ratios and snapshots, respectively. The number of physical sensors is 10. It is obviously seen that the performance of the proposed array is better than others.

Figure 11 presents the RMSE comparison of DOA estimation in different arrays versus SNR, where SNR is from −15 dB to 5 dB and SNAPSHOTS = 500. With the increase of

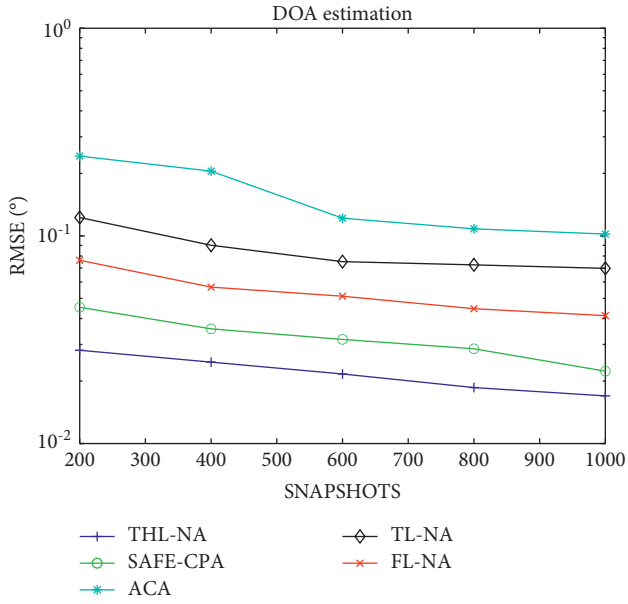


Figure 12: RMSE performance of different arrays versus SNAPSHOTS.

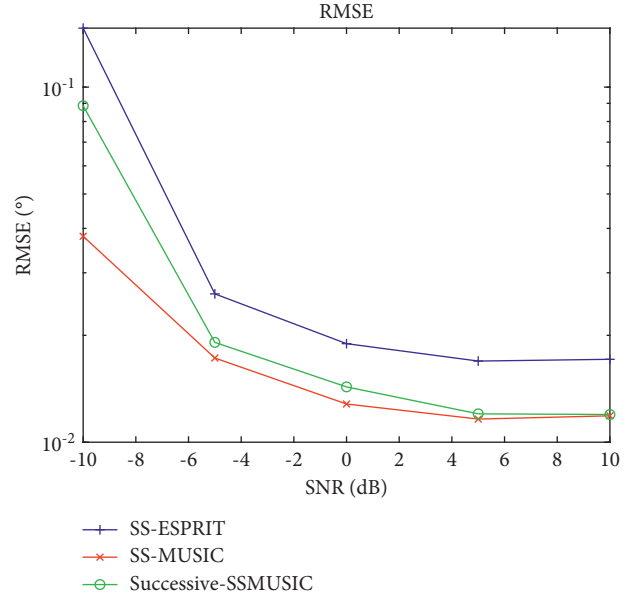


Figure 13: RMSE performance of different algorithms.

SNR, the THL-NA can reach the best DOA estimation performance compared to SAFE-CPA, ACA, FL-NA, and TL-NA.

Figure 12 presents the RMSE comparison of DOA estimation in different arrays versus SNAPSHOTS, where SNAPSHOTS is from 200 to 1000 and SNR = −5 dB. With the increase of the SNAPSHOTS, the THL-NA can reach the best DOA estimation performance compared to SAFE-CPA, ACA, FL-NA, and TL-NA due to the largest consecutive DOF.

6.4. RMSE Comparison of Different Algorithms. Figure 13 gives the comparison of estimation performance with

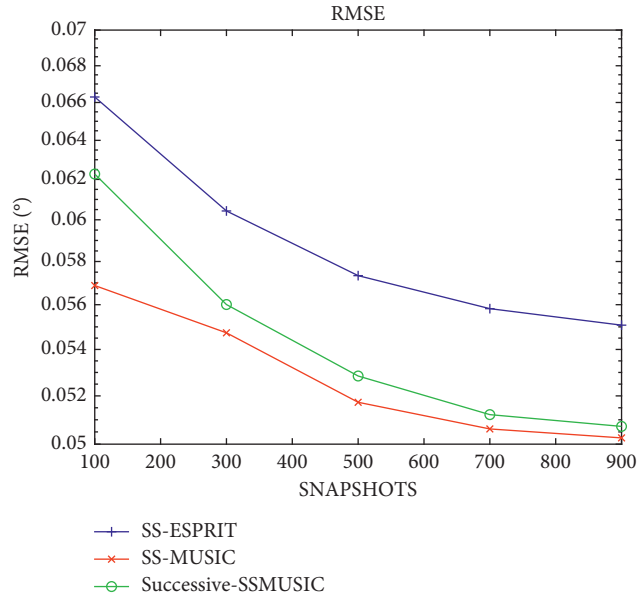


Figure 14: RMSE performance of different algorithms.

different algorithms versus SNR, where SNAPSHOTS = 500. In addition, the searching range of the successive SS-MUSIC algorithm and the SS-MUSIC method is $(\theta_k^{ini} - 1^\circ, \theta_k^{ini} + 1^\circ), k = 1, \ldots K$ and $(-60^\circ, 60^\circ)$, where θ_k^{ini} represents the initial DOA estimates according to Section 4.3. In terms of estimated performance according to following image, the successive SS-MUSIC algorithm is close to SS-MUSIC method and far better than SS-ESPRIT algorithm.

Figure 14 gives the comparison of estimation performance with different algorithms versus snapshots, where SNR = 0. It can be concluded from Figures 12 and 13 that the DOA estimation performance of the successive SS-MUSIC algorithm is more accurate than the SS-ESPRIT algorithm and slightly worse than the SS-MUSIC algorithm.

7. Conclusion

In this paper, the THL-NA is proposed for DOA estimation of non-Gaussian signals based on FOC, which can provide large consecutive DOF and a specific closed-form expression. The proposed array utilizes the characteristic of consecutive coarrays in TL-NA. Firstly, the array aperture is enlarged by obtaining the 2-SC of the proposed array, and then the nested relationship between subarrays in the 2-SC is employed to obtain 2-DCSC to construct a consecutive virtual array. The simulation results prove that the proposed array has higher performance than SAFE-CPA, CPA, FL-NA, TL-NA in DOA estimation. Besides, the successive SS-MUSIC method proposed in the paper has lower computational complexity than SS-MUSIC algorithm and more accurate estimate than SS- ESPRIT algorithm. In future research, sparse array design under the incidence of non-Gaussian sources based on MIMO system [36, 37] may be considered.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by China NSF (Grants 61971217, 61971218, and 61631020), Jiangsu NSF (Grant BK20200444), and National Key Research and Development Project (Grant 2020YFB1807602).

References

[1] X. Zhang, L. Xu, L. Xu, and D. Xu, "Direction of departure (DOD) and direction of arrival (DOA) estimation in MIMO radar with reduced-dimension MUSIC," *IEEE Communications Letters*, vol. 14, no. 12, pp. 1161–1163, 2010.

[2] J. Li, X. Zhang, and H. Chen, "Improved two-dimensional DOA estimation algorithm for two-parallel uniform linear arrays using propagator method," *Signal Processing*, vol. 92, no. 12, pp. 3032–3038, 2012.

[3] F. Wen, "Computationally efficient DOA estimation algorithm for MIMO radar with imperfect waveforms," *IEEE Communications Letters*, vol. 23, no. 6, pp. 1037–1040, 2019.

[4] P. Pal and P. P. Vaidyanathan, "Nested arrays: a novel approach to array processing with enhanced degrees of freedom," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167–4181, 2010.

[5] P. P. Vaidyanathan and P. Pal, "Sparse sensing with Co-prime samplers and arrays," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573–586, 2011.

[6] P. P. Vaidyanathan and P. Pal, "Theory of sparse coprime sensing in multiple dimensions," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3592–3608, 2011.

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

[8] M. Pesavento, A. B. Gershman, and M. Haardt, "Unitary root-MUSIC with a real-valued eigendecomposition: a theoretical and experimental performance study," *IEEE Transactions on Signal Processing*, vol. 48, no. 5, pp. 1306–1314, 2000.

[9] R. Roy and T. Kailath, "ESPRIT-Estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 37, no. 7, pp. 984–995, 1986.

[10] M. Haardt and J. A. Nossek, "Unitary ESPRIT: how to obtain increased estimation accuracy with a reduced computational burden," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1232–1242, 1995.

[11] J. K. Tugnait, "On time delay estimation with unknown spatially correlated Gaussian noise using fourth-order cumulants and cross cumulants," *IEEE Transactions on Signal Processing*, vol. 39, no. 6, pp. 1258–1267, 1991.

[12] M. C. Dogan and J. M. Mendel, "Applications of cumulants to array processing. II. Non-Gaussian noise suppression," *IEEE Transactions on Signal Processing*, vol. 43, no. 7, pp. 1663–1676, 1995.

[13] B. Porat and B. Friedlander, "Direction finding algorithms based on high-order statistics," *IEEE Transactions on Signal Processing*, vol. 39, no. 9, pp. 2016–2024, 1991.

[14] N. Yuen and B. Friedlander, "Asymptotic performance analysis of ESPRIT, higher order ESPRIT, and virtual ESPRIT algorithms," *IEEE Transactions on Signal Processing*, vol. 44, no. 10, pp. 2537–2550, 1996.

[15] M. C. Dogan and J. M. Mendel, "Joint array calibration and direction-finding with virtual-ESPRIT algorithm," in *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics*, pp. 146–150, South Lake Tahoe, CA, USA, June 1993.

[16] P. Pal and P. P. Vaidyanathan, "Multiple level nested array: an efficient geometry for $2q$th order cumulant based array processing," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1253–1269, 2012.

[17] E. BouDaher, F. Ahmad, and M. G. Amin, "Sparsity-based direction finding of coherent and uncorrelated targets using active nonuniform arrays," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1628–1632, 2015.

[18] Q. Shen, W. Liu, W. Cui, and S. Wu, "Extension of Co-prime arrays based on the fourth-order difference Co-array concept," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 615–619, 2016.

[19] R. T. Williams, S. Prasad, A. K. Mahalanabis, and L. H. Sibul, "An improved spatial smoothing technique for bearing estimation in a multipath environment," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 36, no. 4, pp. 425–432, 1988.

[20] G.-T. Pham, P. Loubaton, and P. Vallet, "Performance analysis of spatial smoothing schemes in the context of large arrays," *IEEE Transactions on Signal Processing*, vol. 64, no. 1, pp. 160–172, 2016.

[21] Z. Fu, P. Charge, and Y. Wang, "A virtual nested MIMO array exploiting fourth order difference coarray," *IEEE Signal Processing Letters*, vol. 27, pp. 1140–1144, 2020.

[22] C. L. Liu and P. P. Vaidyanathan, "Super nested arrays: linear sparse arrays with reduced mutual coupling—part I: fundamentals," *IEEE Transactions on Signal Processing*, vol. 64, no. 15, pp. 3997–4012, 2016.

[23] J. Liu, Y. Zhang, Y. Lu, S. Ren, and S. Cao, "Augmented nested arrays with enhanced DOF and reduced mutual coupling," *IEEE Transactions on Signal Processing*, vol. 65, no. 21, pp. 5549–5563, 2017.

[24] Q. Shen, W. Liu, W. Cui, and S. Wu, "Extension of nested arrays with the fourth-order difference co-array enhancement," in *Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2991–2995, Shanghai, China, March 2016.

[25] X. Wang, L. T. Yang, D. Meng, M. Dong, K. Ota, and H. Wang, "Multi-UAV cooperative localization for marine targets based on weighted subspace fitting in SAGIN environment," *IEEE Internet of Things Journal*, 2021.

[26] C. Ye, W. Chen, B. Zhu, and L. Tang, "Direction of arrival estimation of non-Gaussian signals for nested arrays: applying fourth-order difference co-array and the successive method," *ETRI Journal*, vol. 43, 2021.

[27] Y. Zhang, D. Wang, H. Xu, S. Yang, Z. Wang, and J. You, "High-accuracy DOA estimation based on vectorized fourth-order cumulant with coprime array," in *Proceedings of the 2019 IEEE 19th International Conference on Communication Technology (ICCT)*, pp. 210–215, Xi'an, China, October 2019.

[28] C. Liu and P. P. Vaidyanathan, "Super nested arrays: sparse arrays with less mutual coupling than nested arrays," in

Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2976–2980, Shanghai, China, March 2016.

[29] J. Shi, G. Hu, X. Zhang, and H. Zhou, "Generalized nested array: optimization for degrees of freedom and mutual coupling," *IEEE Communications Letters*, vol. 22, no. 6, pp. 1208–1211, 2018.

[30] B. Liao, Z. G. Zhang, and S. C. Chan, "DOA estimation and tracking of ULAs with mutual coupling," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 1, pp. 891–905, 2012.

[31] B. Friedlander and A. J. Weiss, "Direction finding in the presence of mutual coupling," *IEEE Transactions on Antennas and Propagation*, vol. 39, no. 3, pp. 273–284, 1991.

[32] Z. Ye, J. Dai, X. Xu, and X. Wu, "DOA estimation for uniform linear array with mutual coupling," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 1, pp. 280–288, 2009.

[33] C. Wu and C. Ye, "DOA estimation for unfolded coprime arrays: successive-MUSIC algorithm," *IOP Conference Series: Materials Science and Engineering*, vol. 719, 2020.

[34] H. Wang, L. Wan, M. Dong, K. Ota, and X. Wang, "Assistant vehicle localization based on three collaborative base stations via SBL-based robust DOA estimation," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5766–5777, 2019.

[35] J. Cong, X. Wang, C. Yan, L. T. Yang, M. Dong, and K. Ota, "CRB weighted source localization method based on deep neural networks in multi-UAV network," *IEEE Internet of Things Journal*, 2022.

[36] G. Zheng, Y. Song, and C. Chen, "Height measurement with meter wave polarimetric MIMO radar: signal model and MUSIC-like algorithm," *Signal Processing*, vol. 190, Article ID 108344, 2022.

[37] G. Zheng and Y. Song, "Signal model and method for joint angle and range estimation of low-elevation target in meter-wave FDA-MIMO radar," *IEEE Communications Letters*, vol. 26, no. 2, pp. 449–453, 2022.

Hindawi
International Journal of Antennas and Propagation
Volume 2022, Article ID 1462193, 8 pages
https://doi.org/10.1155/2022/1462193

Research Article

Design of a Novel Integral Sliding Mode-Based Composite Nonlinear Feedback Controller for Electrostatic MEMS Micromirror

Jun Wu,[1] **Wenbo Zhu**,[1] **Bingjie Guan**,[2] **and Hui Chen**[3]

[1]*School of Mechatronic Engineering and Automation, Foshan University, Foshan, China*
[2]*School of Artificial Intelligence, Henan University, Kaifeng, China*
[3]*College of Eletrical Engineering and Automation, Shandong University of Science and Technology, Qingdao, China*

Correspondence should be addressed to Hui Chen; chenhuijob@126.com

Received 15 December 2021; Accepted 4 March 2022; Published 31 March 2022

Academic Editor: Bin Zhang

In this study, the precise tracking problem for electrostatic micromirror systems with disturbances and input saturation is investigated. Inspired by the composite nonlinear feedback (CNF)'s improvement of the transient performance and the sliding mode control's enhancement of the robustness, a novel integral sliding mode with reaching law (ISMRL)-based composite nonlinear feedback (CNF) controller is proposed. Then, the stability of the closed-loop system is proved based on Lyapunov theorem. Finally, numerical simulations are investigated to evaluate the effectiveness of the proposed scheme. It is shown that the closed-loop system with the proposed scheme has precise positioning and improved transient performance in presence of time-varying disturbances.

1. Introduction

Micro-electro-mechanical system (MEMS) micromirror has experienced enormous commercial success in applications such as optical switches [1, 2], biomedical imaging [3], and high-resolution displays [4]. Compared with electrothermal, electromagnetic, and piezoelectric actuation, the advantages of electrostatic actuation are fast response, simple electronic driving, and low power consumption [5]. However, the electrostatic MEMS micromirror suffers unsatisfied transient performance and pull-in instability under open loop control. Lots of efforts and control strategies have been introduced to tackle the problem. Classic strategies, such as PID controller [6, 7], H-infinity robust controller [8], sliding mode control [9], and adaptive control [10] schemes have been reported to improve its tracking performance and eliminate the effect of external disturbances. The aforementioned control strategies have been verified benefits of improving the positioning performance and extending the stable operational range of micromirrors. However, the transient response, which is essential for micromirror's application, is not directly considered. For instant, fast setting time and low overshoot of micromirror-based optical switches are required in order to reduce the insertion loss.

For the past few years, composite nonlinear feedback (CNF) control scheme has attracted many attentions as its meaningful improvements in transient behaviors [11]. This scheme was first studied for a class of second-order linear system [12]. Then, it was extended to partially linear composite system [13]. CNF control methods are also investigated for master/slave synchronization of nonlinear system [14], nonlinear time-delay systems [15], strict-feedback nonlinear systems [16], and under actuated systems [17]. The output tracking problem of time-varying references in descriptor systems is investigated using CNF control technique in [18]. Besides developments in theory, the CNF controller is proposed for hard-disk-drive (HDD) servo system [19], spacecraft rendezvous systems [20], multiquadrotor systems [21], robot manipulators [22], and autonomous vehicles [23]. However, when considering the inevitable external

disturbances, the traditional CNF control reveals its lack of the ability to deal with it. Fortunately, in the field of control theory, sliding mode control is considered to be a solution for alleviating the effects of the parametric uncertainties and external disturbances [24–26]. As a result, the sliding mode control techniques combined with composite nonlinear feedback are proposed to improve system robustness in [27–38]. Recently, this scheme is also developed for more general class of linear and nonlinear systems with plant uncertainties [39].

Though sliding mode-based CNF controller has been fully studied and demonstrated the advantages of robustness, the chattering problem of sliding mode control is a serious situation which not only increases energy consumption but also leads to the instability. To avoid this, one solution is to introduce the second-order sliding mode control [40]. In [41], a super-twisting algorithm-based integral sliding mode control with composite nonlinear feedback control is proposed to eliminate the chattering effect for magnetic levitation system. Another effective solution to reduce the chattering is the sliding mode control based on the reaching law, which is first proposed in [42]. Recently, in [43–45], improved quick reaching law is proposed to speed up the response and reduce the chattering of a sliding mode control system simultaneously.

Motivated by the aforementioned problem through literature review, the main contribution of this research is that a novel integral sliding mode with reaching law-based composite nonlinear feedback (ISMRL-CNF) controller for angular control of an electrostatic MEMS micromiror is proposed. The CNF controller is designed to guarantee the system has fast dynamic performance and small overshoot. An integral sliding mode control with quick reaching law is designed to enhance the robustness, attenuate chattering, and achieve finite-time convergence of the sliding mode. Furthermore, the time-varying disturbances and input saturation are taken into account for controller design and stability analysis. Simulation study verifies that the closed-loop system with the proposed scheme has precise positioning and improved transient performance in presence of time-varying disturbances.

The rest of the paper is organized as follows. The dynamic model is described in Section 2. The design procedure of the proposed controller and stability analysis are developed in Section 3. Simulation study is given in Section 4. Finally, the conclusions are discussed.

2. The Simplified Dynamics of Electrostatic MEMS Micromirror

Figure 1 shows the schematic figure of a 2-degree-of-freedom electrostatic torsional MEMS micromirror. The studied micromirror consists of mirror plate, torsion bar, gimbal frame, and bottom and sidewall electrodes. The mirror plate is suspended by double frame structure and driven by electrostatic torque. When the driving voltage applied to the bottom and sidewall electrodes, the mirror plate is actuated about X-axis and Y-axis, respectively. The dynamic equations of the system are given as follows [5]:

$$\begin{aligned}(J_1 + J_2)\ddot{\alpha} + D_1\dot{\alpha} + K_1\alpha &= T_\alpha,\\ J_1\ddot{\beta} + D_2\dot{\beta} + K_2\beta &= T_\beta,\end{aligned} \tag{1}$$

where α and β represent the tilt angles of x-axis and y-axis, J_1 and J_2 denote the mass moment of inertias of the mirror plate and gimbal, respectively, D_1 and D_2 represent the damping coefficients, K_1, K_2 represent the stiffness coefficients, and T_α and T_β are electrostatic torque. Introducing the parameter $\tau = \sqrt{K_2/J_1}t$, let $x_1 = \alpha$, $x_2 = d\alpha/d\tau$, $x_3 = \beta$, and $x_4 = d\beta/d\tau$. As a result, system (1) can be described as [5]

$$\begin{aligned}\dot{x}_1 &= x_2,\\ \dot{x}_2 &= -R_1x_2 - \lambda_{\alpha\beta}x_1 + G_1T_\alpha,\\ \dot{x}_3 &= x_4,\\ \dot{x}_4 &= -R_2x_4 - x_3 + G_2T_\beta,\end{aligned} \tag{2}$$

where the parameters are $R_1 = 0.16$, $R_2 = 0.15$, $\lambda_{\alpha\beta} = 0.2251$, $G_1 = 3.0827 \times 10^6$, and $G_2 = 1.7894 \times 10^7$. Considering input saturation and external disturbances, the system is rewritten as

$$\begin{aligned}\dot{x} &= Ax + \mathrm{Bsat}(u) + B\,d,\\ y &= Cx,\end{aligned} \tag{3}$$

where $x \in R^n$ is the state of micromirror system, $u \in R$ is the control input, $y \in R$ is the measurement output, And $A, B,$ and C are constant appropriate dimensional matrices. The saturation function $\mathrm{sat}(\cdot)$: $R \longrightarrow R$ is defined as $\mathrm{sat}(u) = \mathrm{sgn}(u)\min\{u_{\max}, |u|\}$ with the maximum of control input $u_{\max}$. The system uncertainties and disturbances $d \in R$ are bounded, and $|d| \le d_{\max}$ with the maximum of disturbances $d_{\max}$. The following assumptions are satisfied for the investigated system (3) [19]: (1) (A, B) is stabilizable, (2) (A, B, C) is invertible, which means the system has no zeros at $s = 0$, and (3) (A, C) is detectable.

3. The Proposed Integral Sliding Mode-Based Composite Nonlinear Feedback Controller Design

In this section, a novel integral sliding mode-based composite nonlinear feedback (ISMRL-CNF) controller is designed for the micromirror system with input saturation and disturbances. The objective is to ensure that the controlled output can track step command input precisely with enhanced transient performance and robustness, in presence of external disturbances. The proposed controller consists of a CNF control law and an integral sliding mode control with reaching law (ISMRL):

$$u = u_N + u_{is}, \tag{4}$$

where the CNF control u_N is utilized to achieve good transient performance and the integral sliding mode control

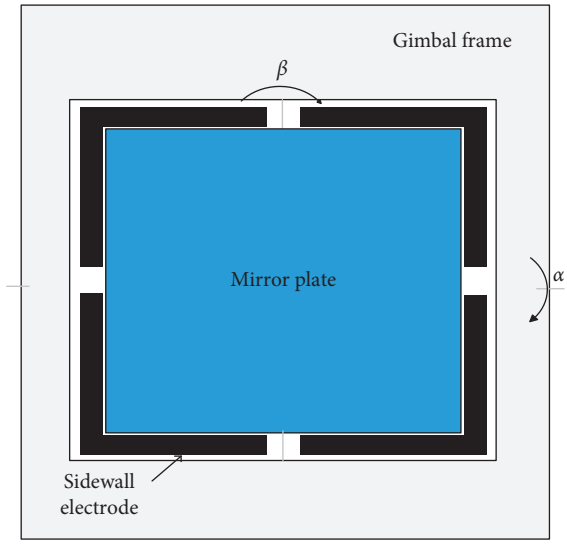


Figure 1: The electrostatic torsional MEMS micromirror.

with reaching law u_{is} is designed to guarantee the system robustness under disturbances and reduce chattering.

The CNF controller consists of a linear control law and a nonlinear control law. The linear control law is presented to achieve fast response by using small damping ration. The nonlinear control law is developed to change the damping ration in order to eliminate overshoot. The linear feedback control u_L is designed as [12]

$$u_L = Fx + Gr, \tag{5}$$

where F is chosen to guarantee that $(A + \mathrm{BF})$ an asymptotically stable matrix, r is a step command input, and G is scalar and calculated as

$$G = -\left[C(A + \mathrm{BF})^{-1}B\right]^{-1}, \tag{6}$$

where F and G are defined since (A, B, C) is assumed to have no invariant zeros at $s = 0$.

The nonlinear feedback law u_{NL} is designed as

$$u_{\mathrm{NL}} = \rho(y, r)B^T P(x - x_e), \tag{7}$$

where $\rho(y, r)$ represents any nonpositive nonlinear function. Different forms of $\rho(y, r)$ have been reported in previous works. A scaled nonlinear function is proposed to adapt the variation of tracking targets in [19] as

$$\rho(y, r) = -\overline{\beta}e^{-m(y-r)}, \tag{8}$$

where $\overline{\beta}$ and m are positive tunable parameters. Define P as a real positive definite symmetric matrix which can be solved from the Lyapunov equation:

$$(A + \mathrm{BF})^T P + P(A + \mathrm{BF}) = -W. \tag{9}$$

With a given positive definite symmetric matrix W, consider that P always exists since $(A + \mathrm{BF})$ is defined to be asymptotically stable. W can be chosen as

$$W = 10^{\theta} \cdot \widehat{E}, \tag{10}$$

where θ is a tunable parameter and $\widehat{E}$ is an identity matrix. Then, the new steady-state value x_e is computed as

$$x_e = -(A + \mathrm{BF})^{-1}\mathrm{BG}r. \tag{11}$$

Finally, a CNF controller is formed by combining the linear feedback law (5) and the nonlinear feedback law (7) as follows:

$$\begin{aligned} u_N &= u_L + u_{\mathrm{NL}} \\ &= Fx + Gr + \rho(y, r)B^T P(x - x_e), \end{aligned} \tag{12}$$

Theorem 1. *For any $\delta \in (0, 1)$, choosing $c_\delta > 0$ to be the largest positive scalar and satisfying the conditions [27],*

$$|Fx| \le (1 - \delta)u_{\max}, \forall x \in X_\delta := \left\{x | x^T P x \le c_\delta\right\}. \tag{13}$$

The initial state x_0 and r satisfy

$$x_0 - x_e \in X_\delta, |Hr| \le \delta_1 u_{\max}, \tag{14}$$

where $H = [I - F(A + BF)^{-1}B]G$, $0 \le \delta_1 < \delta$, and $|(\delta - \delta_1)u_{\max}| = d_{\max}$. Then, the control law (12) is capable to drive the system output y to track the commend input r asymptotically for any nonpositive function $\rho(y, r)$.

Inspired by the robustness enhancement of the integral sliding mode control with reaching law, such a controller is designed and added with the CNF controller to make overall system robust. Taking into account the input saturation, the integral sliding surface is designed as [38]

$$s = B^+\left(x(t) - x_0 - \int_o^t Ax(\tau) + B\left(\mathrm{sat}(u) - u_{is}\right)(\tau)d\tau\right), \tag{15}$$

where B^+ is the pseudoinverse of B and u_{is} is sliding mode control. The reaching law approach is first proposed in [42] which is utilized to force the system state quickly arrives at the sliding mode surface in the whole approaching process [43]. A novel improved quick reaching law is designed as [44]

$$\dot{s} = -k_1\left(b^{|s|} - 1\right)\mathrm{sgn}(s) - k_2|s|^a \mathrm{sgn}(s), \tag{16}$$

where $0 < a < 1$, $k_1 > 0$, $k_2 > 0$, and $b = 1 + k_2/k_1$. When the system state is far away from the sliding surface, the change rate of the first term in (16) is larger than that of the power function. It speeds up the reaching rate in the case $|s| > 1$. When the system state is near to the sliding surface, the second item in (16) can make the system approach the sliding surface with higher speed.

Finally, the proposed ISMRL-CNF scheme for system (3) is expressed as

$$\begin{aligned} u &= u_{\mathrm{CNF}} + u_{is} \\ &= Fx + Gr + \rho(y,r)B^T P(x - x_e) - k_1\left(b^{|s|} - 1\right)\mathrm{sgn}(s) - k_2|s|^a \mathrm{sgn}(s). \end{aligned} \tag{17}$$

Remark 1. In order to ensure the closed-loop system has a fast rise time, F is chosen such that the closed-loop poles of $C(sI - A + \mathrm{BF})^{-1}$ have a dominant pair with a small ration. F can be designed by using H_∞ optimization approach. In order to reduce the overshoot, $\rho(y,r)$ is selected to gradually change the damping ratio of the closed-loop system. To obtain the control parameters F, m, $\overline{\beta}$, and W properly, we have

$$\min_{F,m,\overline{\beta},W} = \int_0^\infty t|y - r|dt. \tag{18}$$

The integrated time and absolute error (ITAE) is utilized as the performance criteria, and the minimization problem (18) is solved by the particle swarm optimization (PSO) algorithm.

3.1. Demonstration of System Stability. Taking the derivative of the sliding surface (15) along the trajectories of system yields

$$\begin{aligned} \dot{s} &= B^+\left(\dot{x} - \left(Ax + B\left(\mathrm{sat}(u) - u_{is}\right)\right)\right) \\ &= B^+\left(Ax + \mathrm{Bsat}(u) + B\,d - \left(Ax + B\left(\mathrm{sat}(u) - u_{is}\right)\right)\right) \\ &= B^+\left(Bu_{is} + B\,d\right) \\ &= -k_1\left(b^{|s|} - 1\right)\mathrm{sgn}(s) - k_2|s|^a \mathrm{sgn}(s) + d. \end{aligned} \tag{19}$$

Defining a Lyapunov function $V_1 = 1/2s^2$ and taking the derivative of V_1 [44],

$$\begin{aligned} \dot{V}_1 &= s\dot{s} = s\left(-k_1\left(b^{|s|} - 1\right)\mathrm{sgn}(s) - k_2|s|^a \mathrm{sgn}(s) + d\right) \\ &= -k_1|s|\left(b^{|s|} - 1\right) - k_2|s|^{a+1} + s\,d \le -k_1|s|\left(b^{|s|} - 1\right) - k_2|s|^{a+1} + |s\|\,d| \le -k_2|s|^{a+1} - \left(k_1\left(b^{|s|} - 1\right) - d_{\max}\right)|s|, \end{aligned} \tag{20}$$

where $|d| \le d_{\max}$. It can be noted that when $k_1(b^{|s|} - 1) - d_{\max} \ge 0$ such that

$$|s| \ge \log_b\left(\frac{d_{\max} + k_1}{k_1}\right), \tag{21}$$

then $\dot{V}_1 \le -k_2|s|^{a+1}$. Thus, the sliding mode variable s can converge to the finite-time convergence region $|s| \le \log_b(d_{\max} + k_1/k_1)$ [44].

Take the control law into the system and let $\tilde{x} = x - x_e$, and we have

$$\dot{\tilde{x}} = (A + \mathrm{BF})\tilde{x} + B\psi, \tag{22}$$

where $\psi = \mathrm{sat}(F\tilde{x} + Hr + u_{\mathrm{NL}} + u_{\mathrm{is}}) - F\tilde{x} - Hr + d$.

Define a Lyapunov function:

$$V_2 = \tilde{x}^T P\tilde{x}. \tag{23}$$

Taking the derivative of V_2, we obtain

$$\begin{aligned} \dot{V}_2 &= \dot{\tilde{x}}^T P\tilde{x} + \tilde{x}^T P\dot{\tilde{x}} \\ &= \tilde{x}^T (A + \mathrm{BF})^T P\tilde{x} + \tilde{x}^T (A + \mathrm{BF})P\tilde{x} + 2\tilde{x}^T \mathrm{PB}\psi \\ &= -\tilde{x}^T W\tilde{x} + 2\tilde{x}^T \mathrm{PB}\psi. \end{aligned} \tag{24}$$

Then, we have

$$\left|F\tilde{x} + Hr + u_{is}\right| \le |F\tilde{x}| + |Hr| + \left|u_{is}\right| \le u_{\max}, \tag{25}$$

When $|F\tilde{x} + Hr + u_{NL} + u_{is}| \le u_{\max}$, then

$$\psi = F\tilde{x} + Hr + u_{\mathrm{NL}} + u_{\mathrm{is}} - F\tilde{x} - Hr + d, \tag{26}$$

where $\dot{s}$ would converge to zero in finite time; then, $u_{is} = -d$. Then, $\psi = u_{\mathrm{NL}} = \rho(y,r)B^T P\tilde{x}$. $\rho(y,r)$ is a nonpositive function. So, $\dot{V}_2 \le -\tilde{x}^T W\tilde{x}$.

When $(F\tilde{x} + Hr + u_{\mathrm{NL}} + u_{\mathrm{is}}) > u_{\max}$, then

$$\begin{aligned} u_{\mathrm{NL}} &< u_{\max} - \left(F\tilde{x} + Hr + u_{\mathrm{is}}\right) < 0, \\ \psi &= u_{\max} - (F\tilde{x} + Hr) + d \\ &= u_{\max} - \left(F\tilde{x} + Hr + u_{is}\right) \ge 0, \end{aligned} \tag{27}$$

where $\rho(y,r)$ is a nonpositive function. So, it implies that $\tilde{x}^T \mathrm{PB} \le 0$. Then, $\dot{V}_2 \le -\tilde{x}^T W\tilde{x}$.

When $(F\tilde{x} + Hr + u_{\mathrm{NL}} + u_{\mathrm{is}}) < -u_{\max}$, then

$$\begin{aligned} u_{\mathrm{NL}} &\le -u_{\max} - F\tilde{x} - Hr - u_{\mathrm{is}} \le 0, \\ \psi &= -u_{\max} - (F\tilde{x} + Hr) + d \\ &= -u_{\max} - \left(F\tilde{x} + Hr + u_{\mathrm{is}}\right) \le 0, \end{aligned} \tag{28}$$

where $\rho(y,r)$ is a nonpositive function. So, it implies that $\tilde{x}^T \mathrm{PB} \le 0$. Thus, $\dot{V}_2 \le -\tilde{x}^T W\tilde{x}$. Therefore, we can summarize that $\dot{V}_2 \le -\tilde{x}^T W\tilde{x} < 0$.

4. Simulation Results

In this section, various simulations are conducted to test the performance of the proposed strategy. For simplicity, system (2) is consider as linear systems with disturbances:

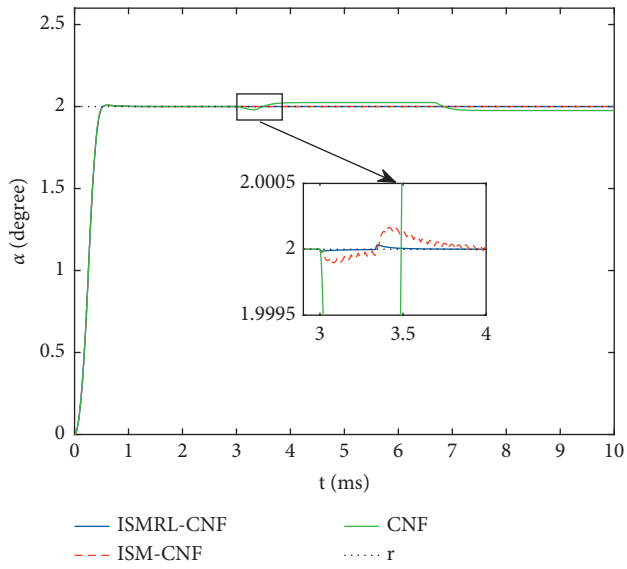


Figure 2: Comparison of angular α responses.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -0.2251 & -0.16 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 3.0827 \end{bmatrix} (\text{sat}(u_\alpha) + d_\alpha),$$
$$y_1 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \tag{29}$$

$$\begin{bmatrix} \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & -0.15 \end{bmatrix} \begin{bmatrix} x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 \\ 1.7894 \end{bmatrix} (\text{sat}(u_\beta) + d_\beta),$$
$$y_2 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_3 \\ x_4 \end{bmatrix}, \tag{30}$$

where $u_\alpha = 10^6 T_\alpha$ and $u_\beta = 10^7 T_\beta$. The maximum inputs are $u_{\alpha\max} = u_{\beta\max} = 10$. The control objective is to force the scan angle α or β and follow the reference trajectory r precisely with fast response in presence of disturbances.

The parameters of CNF for system (29) are tuned by PSO as $F_\alpha = [-40.3457 \ \ -5.6301]$, $m_\alpha = 62.9236$, $\overline{\beta}_\alpha = 49.4610$, and $\theta_\alpha = -0.1296$. Then, $G_\alpha = 40.4187$, $W_\alpha = \begin{bmatrix} 0.7420 & 0 \\ 0 & 0.7420 \end{bmatrix}$, and $P_\alpha = \begin{bmatrix} 2.7124 & 0.0030 \\ 0.0030 & 0.0214 \end{bmatrix}$. The parameters of ISMRL are chosen as $k_{1\alpha} = 1$, $k_{2\alpha} = 1.6$, and $a_\alpha = 0.05$. The parameters of CNF for system (30) are tuned by PSO as $F_\beta = [-40.5216 \ \ -7.1381]$, $m_\beta = 4.5160$, $\overline{\beta}_\beta = 35.0126$, and $\theta_\beta = 0.4681$. Then, $G_\alpha = 41.0804 W_\beta = \begin{bmatrix} 2.9383 & 0 \\ 0 & 2.9383 \end{bmatrix}$ and $P_\beta = \begin{bmatrix} 8.7290 & 0.0200 \\ 0.0200 & 0.1152 \end{bmatrix}$. The parameters of ISMRL are chosen as $k_{1\beta} = 1.5$, $k_{2\beta} = 2.2$, and $a_\beta = 0.04$.

Figure 2 shows the MEMS micromirror along the x-axes tracking trajectories using the CNF, ISM-CNF, and proposed ISMRL-CNF controller, respectively. The target references for α are set as $r = 2$; the disturbance $d_\alpha = -\text{sgn}(\sin(0.3\pi t))$ is introduced when $t \geq 3\,ms$. It can be noted that the proposed controller has better performance in comparison with the two other controller such as CNF and ISM-CNF. The actuation motion trajectory of the MEMS micromirror under the proposed ISMRL-CNF controller is consistent with the desired trajectory in presence of the disturbance. The closed-loop system has good transient performance such as very small overshoot and fast response. The control inputs of CNF, ISM-CNF, and ISMRL-CNF are shown in Figure 3; compared with ISM-CNF, the chattering problem is eliminated by using the proposed ISMRL-CNF.

Figure 4 shows the MEMS micromirror along the y-axes tracking trajectories using the CNF, ISM-CNF, and proposed ISMRL-CNF controller, respectively. The target references for β are set as $r = 2.2$ and the disturbance $d_\beta = -1.5\text{sgn}(\sin(\pi t))$ is introduced when $t \geq 3.5\,ms$. The results demonstrate that the CNF controller exhibits worst performance due to the disturbances. Compare with ISM-CNF, the proposed ISMRL-CNF controller can obtain a faster and more efficient performance in presence of the disturbances. The control inputs of CNF, ISM-CNF, and proposed controller

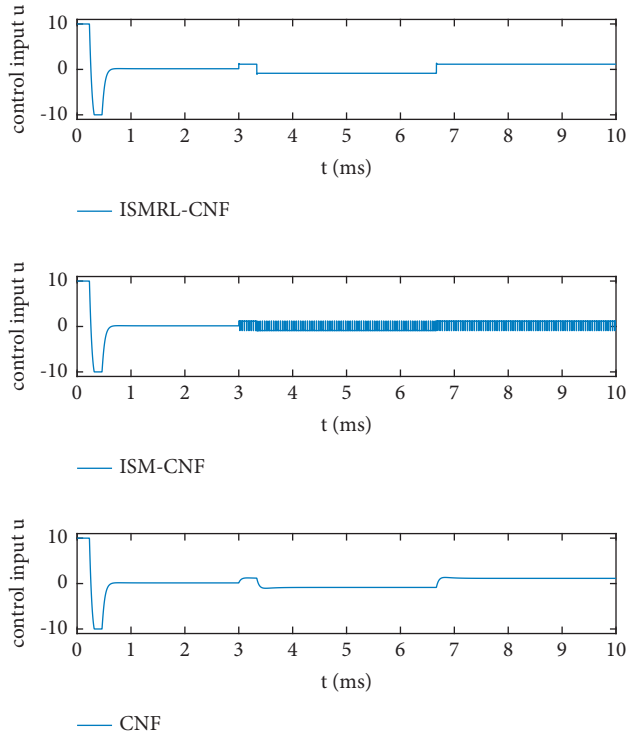


Figure 3: Control input u_α.

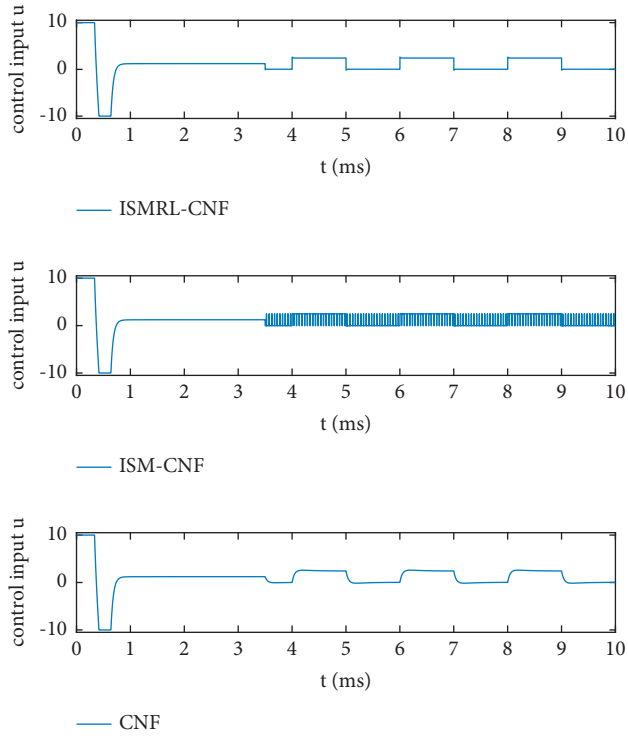


Figure 5: Control input u_β.

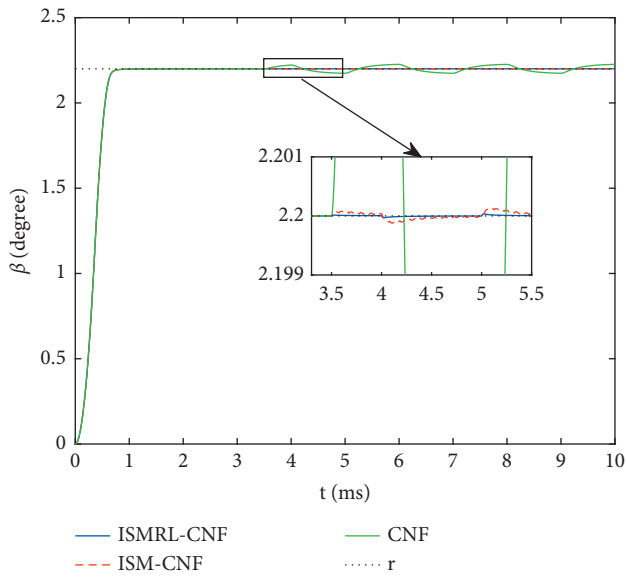


Figure 4: Comparison of angular β responses.

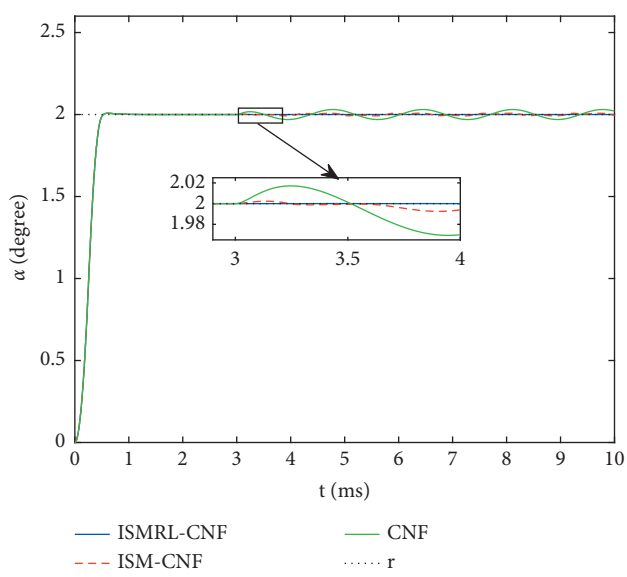


Figure 6: Comparison of angular α responses under disturbance $d = -1.5\sin(1.2\pi t)$.

are shown in Figure 5. It can be noted that the control input of ISMRL-CNF is more smoother than ISM-CNF.

To further verify the antidisturbances of the proposed controller, the time-varying disturbance is introduced.

The comparison of angular α responses using CNF, ISM-CNF, and proposed ISMRL-CNF controller under disturbance $d = -1.5\sin(1.2\pi t)$ is shown in Figure 6. It can be seen that the traditional CNF controller is not able to suppress the time-varying disturbance. Compared with ISM-CNF, the proposed ISMRL-CNF controller ensures the system has better response in presence of disturbance. The comparison of angular β responses using CNF, ISM-CNF, and proposed ISMRL-CNF controller under disturbance $d = -1.4\cos(1.3\pi t)$ is shown in Figure 7. The results demonstrate that the proposed controller ensures the system has better performance in presence of time-varying disturbance.

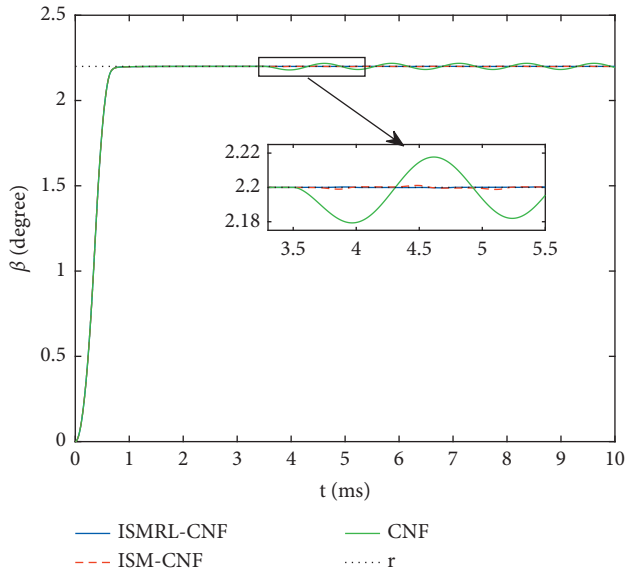


FIGURE 7: Comparison of angular β responses under disturbance $d = -1.4\cos(1.3\pi t)$.

5. Conclusions

In this study, the precise tracking problem for electrostatic micromirror systems with disturbances and input saturation is investigated. Inspired by the composite nonlinear feedback (CNF)'s improvement of the transient performance and the sliding mode control's enhancement of the robustness, a novel integral sliding mode with reaching law (ISMRL)-based composite nonlinear feedback (CNF) controller is proposed. Then, the stability of the closed-loop system is guaranteed based on Lyapunov theorem. Numerical simulations verify the effectiveness of the proposed scheme. It is shown that the closed-loop system with the proposed scheme has precise positioning and improved transient performance even in presence of time-varying disturbances. It should be noted that the proposed controller needs the accurate model knowledge; as a result, more inclusive methods about the model uncertainty and model inaccuracy, combined with faster convergence rate and smaller chattering, will be inquired in future work.

Data Availability

All data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by National Science Foundation of China, under Grants 61603093 and 61703142.

References

[1] D. Wang, C. Watkins, and H. Xie, "MEMS mirrors for LiDAR: a review," *Micromachines*, vol. 11, no. 5, pp. 456–480, 2020.

[2] M. Saif, B. Ebrahimi, and M. Vali, "A second order sliding mode strategy for fault detection and fault-tolerant-control of a MEMS optical switch," *Mechatronics*, vol. 22, no. 6, pp. 696–705, 2012.

[3] Y. Lin, L. Keeler, and G. Ethan, "Progress of mems scanning micromirrors for optical bioimaging," *Micromachines*, vol. 6, 2015.

[4] A. D. Yalcinkaya, H. Urey, D. Brown, T. Montague, and R. Sprague, "Two-Axis electromagnetic microscanner for high resolution displays," *Journal of Microelectromechanical Systems*, vol. 15, no. 4, pp. 786–794, 2006.

[5] Y. Bai, J. T. W. Yeow, and B. C. Wilson, "A characteristic study of micromirror with sidewall electrodes," *International Journal of Optomechatronics*, vol. 1, no. 3, pp. 231–258, 2007.

[6] P. Zuo, G. Li, W. Xie, and J. T. W. Yeow, "Angle tracking of MEMS hard-magnetic micromirror by PID control," in *Proceedings of the 2015 34th Chinese Control Conference (CCC)*, pp. 593–597, Hangzhou, China, July 2015.

[7] C. G. Agudelo, M. Packirisamy, G. Guchuan Zhu, and L. Saydy, "Nonlinear control of an electrostatic micromirror beyond pull-in with experimental validation," *Journal of Microelectromechanical Systems*, vol. 18, no. 4, pp. 914–923, 2009.

[8] H. Chen, M. Li, Y. Zhang et al., "H_∞ robust control of a large-piston MEMS micromirror for compact fourier transform spectrometer systems," *IEEE Sensors Journal*, vol. 18, no. 4, 2018.

[9] H. Chen, Z. Sun, W. Sun, and J. T. W. Yeow, "Twisting sliding mode control of an electrostatic MEMS micromirror for a laser scanning system," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 4, pp. 1060–1067, 2019.

[10] K. M Liao, Y. C. Wang, C. H. Yeh, and R. Chen, "Closed-loop adaptive control for electrostatically driven torsional micromirrors," *The Journal of Microlithography, Microfabrication, and Microsystems*, vol. 4, no. 4, 2005.

[11] Z. Lin, M. Pachter, and S. Banda, "Toward improvement of tracking performance nonlinear feedback for linear systems," *International Journal of Control*, vol. 70, no. 1, pp. 1–11, 1998.

[12] B. M. Chen, T. H. Lee, K. Kemao Peng, and V. Venkataramanan, "Composite nonlinear feedback control for linear systems with input saturation: theory and an application," *IEEE Transactions on Automatic Control*, vol. 48, no. 3, pp. 427–439, 2003.

[13] W. Lan, B. M. Chen, and Y. He, "On improvement of transient performance in tracking control for a class of nonlinear systems with input saturation," *Systems & Control Letters*, vol. 55, no. 2, pp. 132–138, 2006.

[14] S. Mobayen and F. Tchier, "Composite nonlinear feedback control technique for master/slave synchronization of nonlinear systems," *Nonlinear Dynamics*, vol. 87, no. 3, pp. 1731–1747, 2016.

[15] S. Singh, S. Purwar, and A. Kulkarni, "Two term composite nonlinear feedback controller design for nonlinear time-delay systems," *Transactions of the Institute of Measurement and Control*, vol. 40, no. 12, pp. 3424–3432, 2017.

[16] T. Lu and W. Lan, "Composite nonlinear feedback control for strict-feedback nonlinear systems with input saturation," *International Journal of Control*, vol. 92, no. 9, pp. 2170–2177, 2019.

[17] T. Yang, N. Sun, and Y. Fang, "Neuroadaptive control for complicated underactuated systems with simultaneous output and velocity constraints exerted on both actuated and unactuated states," *IEEE Transactions on Neural Networks and Learning Systems*, 2021, in press.

[18] E. Jafari and T. Binazadeh, "Observer-based improved composite nonlinear feedback control for output tracking of time-varying references in descriptor systems with actuator saturation," *ISA Transactions*, vol. 91, pp. 1–10, 2019.

[19] W. Weiyao Lan, C. K. Chin Kwan Thum, and B. M. Chen, "A hard-disk-drive servo system design using composite nonlinear-feedback control with optimal nonlinear gain tuning methods," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 5, pp. 1735–1745, 2010.

[20] H. Namdari, F. Allahverdizadeh, and A. Sharifi, "Robust composite nonlinear feedback control for spacecraft rendezvous systems under parameter uncertainty, external disturbance, and input saturation," *Proceedings of the Institution of Mechanical Engineers - Part G: Journal of Aerospace Engineering*, vol. 234, no. 2, pp. 143–155, 2018.

[21] Z. Hou and I. Fantoni, "Interactive leader-follower consensus of multiple quadrotors based on composite nonlinear feedback control," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 5, pp. 1732–1743, 2018.

[22] Y. Jiang, K. Lu, C. Gong, and H. Liang, "Robust composite nonlinear feedback control for uncertain robot manipulators," *International Journal of Advanced Robotic Systems*, vol. 17, no. 2, pp. 1–9, 2020.

[23] Y. Chen, C. Hu, and J. Wang, "Motion planning with velocity prediction and composite nonlinear feedback tracking control for lane-change strategy of autonomous vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 1, pp. 63–74, 2020.

[24] F. Castanos and L. Fridman, "Analysis and design of integral sliding manifolds for systems with unmatched perturbations," *IEEE Transactions on Automatic Control*, vol. 51, no. 5, pp. 853–858, 2006.

[25] G. Sun, Z. Ma, and J. Yu, "Discrete-time fractional order terminal sliding mode tracking control for linear motor," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 4, pp. 3386–3394, 2018.

[26] H. Wang, L. Hua, Y. Guo, and C. Lu, "Control of Z-axis MEMS gyroscope using adaptive fractional order dynamic sliding mode approach," *IEEE Access*, vol. 7, pp. 133008–133016, 2019.

[27] B. Bandyopadhyay, D. Fulwani, and Y. J. Park, "A robust algorithm against actuator saturation using integral sliding mode and composite nonlinear feedback," in *Proceedings of the 17th IFAC World Congress*, pp. 14174–14179, Seoul, Korea, July 2008.

[28] J. Tan, W. Sun, and J. T. W. Yeow, "An enhanced robust control algorithm based on CNF and ISM for the MEMS micromirror against input saturation and disturbance," *Micromachines*, vol. 8, no. 8, 2017.

[29] S. Mobayen, V. J. Majd, and M. Sojoodi, "An LMI-based composite nonlinear feedback terminal sliding-mode controller design for disturbed MIMO systems," *Mathematics and Computers in Simulation*, vol. 85, pp. 1–10, 2012.

[30] S. Mobayen, M. H. Asemani, and V. J. Majd, "Transient performance improvement using composite nonlinear feedback and integral sliding surface for matched and unmatched uncertain MIMO linear systems," in *Proceedings of the 3rd ICCIA*, pp. 83–88, Tehran, Iran, December 2013.

[31] S. Mobayen, "Design of CNF-based nonlinear integral sliding surface for matched uncertain linear systems with multiple state-delays," *Nonlinear Dynamics*, vol. 77, no. 3, pp. 1047–1054, 2014.

[32] V. J. Majd and S. Mobayen, "An ISM-based CNF tracking controller design for uncertain MIMO linear systems with multiple time-delays and external disturbances," *Nonlinear Dynamics*, vol. 80, no. 1-2, pp. 591–613, 2015.

[33] C. Hu, R. Wang, and F. Yan, "Integral sliding mode-based composite nonlinear feedback control for path following of four-wheel independently actuated autonomous vehicles," *IEEE Transactions on Transportation Electrification*, vol. 2, no. 2, pp. 221–230, 2016.

[34] C. Hu, R. Wang, F. Yan, M. Chadli, Y. Huang, and H. Wang, "Robust path-following control for a fully actuated marine surface vessel with composite nonlinear feedback," *Transactions of the Institute of Measurement and Control*, vol. 40, no. 12, pp. 3477–3488, 2017.

[35] S. Mobayen and F. Tchier, "Composite nonlinear feedback integral sliding mode tracker design for uncertain switched systems with input saturation," *Communications in Nonlinear Science and Numerical Simulation*, vol. 65, pp. 173–184, 2018.

[36] S. Mobayen, "Chaos synchronization of uncertain chaotic systems using composite nonlinear feedback based integral sliding mode control," *ISA Transactions*, vol. 77, pp. 100–111, 2018.

[37] P. S. Babu and B. Bandyopadhyay, "Integral sliding mode based composite nonlinear feedback control for descriptor systems," *IFAC-Papers On Line*, vol. 51, no. 1, pp. 598–603, 2018.

[38] P. S. Babu, B. Bandyopadhyay, and M. Thomas, "Robust composite non-linear feedback control for descriptor systems with general reference tracking," in *Proceedings of the IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, pp. 2434–2439, Washington, DC, USA, October 2018.

[39] T. Yang, N. Sun, and Y. Fang, "Adaptive fuzzy control for a class of MIMO underactuated systems with plant uncertainties and actuator deadzones: design and experiments," *IEEE Transactions on Cybernetics*, vol. 99, pp. 1–14, 2021.

[40] H. Chen, B. Guan, Y. Zhou, W. Sun, and J. T. W. Yeow, "Optimal second order integral sliding mode based composite nonlinear feedback approach for an electrostatic micromirror," *IEEE Access*, vol. 8, pp. 145960–145967, 2020.

[41] A. Pati and R. Negi, "Super-twisting algorithm-based integral sliding mode control with composite nonlinear feedback control for magnetic levitation system," *International Journal of Automation and Control*, vol. 13, no. 6, pp. 717–734, 2019.

[42] W. Weibing Gao and J. C. Hung, "Variable structure control of nonlinear systems: a new approach," *IEEE Transactions on Industrial Electronics*, vol. 40, no. 1, pp. 45–55, 1993.

[43] H. Ma, J. Wu, and Z. Xiong, "A novel exponential reaching law of discrete-time sliding-mode control," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 3840–3850, 2017.

[44] C. B. Xiu and P. H. Gao, "Global terminal sliding mode control with the quick reaching law and its application," *IEEE Access*, vol. 6, pp. 1–8, 2018.

[45] H. Pan, G. Zhang, H. Ouyang, and L. Mei, "A novel global fast terminal sliding mode control scheme for second-order systems," *IEEE Access*, vol. 8, pp. 22758–22769, 2020.

Hindawi
International Journal of Antennas and Propagation
Volume 2022, Article ID 7984443, 11 pages
https://doi.org/10.1155/2022/7984443

Research Article

Parameter Estimation of Non-Gaussian Signals for Polarization-Sensitive Augmented Coprime Array: Fourth-Order Cumulant Reduced-Dimensional Capon Algorithm

Meng Yang,[1,2] Haowei Zeng,[1,3] and Xiaofei Zhang[1,3]

[1]*College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China*
[2]*Jiangsu Automation Research Institute, Lianyungang 222000, China*
[3]*Key Laboratory of Dynamic Cognitive System of Electromagnetic Spectrum Space (Nanjing University of Aeronautics and Astronautics), Ministry of Industry and Information Technology, Nanjing 211106, China*

Correspondence should be addressed to Haowei Zeng; zenghaowei@nuaa.edu.cn

Received 19 January 2022; Accepted 10 March 2022; Published 27 March 2022

Academic Editor: Xianpeng Wang

In this paper, we investigate the problem of direction of arrival (DOA) and polarization estimation for non-Gaussian signal in polarization-sensitive augmented coprime array. Instead of the second-order cumulant statistics, the fourth-order cumulant statistics of the received signals are used for parameter estimation because they can detect more information of the non-Gaussian signal. First, polarization-sensitive augmented coprime array is designed, where each sensor element is equipped with a pair of orthogonal electric dipoles. Furthermore, a low-complexity reduced-dimensional Capon algorithm which uses the fourth-order cumulant of the received array signal is proposed for DOA and polarization estimation. Only one-dimensional peak search is required by reconstructing peak search function. Theoretical analysis has proven the effectiveness of the algorithm, and simulation results demonstrate that the proposed fourth-order cumulant reduced-dimensional Capon algorithm outperforms the other algorithms.

1. Introduction

The research on directional of arrival (DOA), as a fundamental project in array signal processing, has attracted great attention from relevant scholars all around the world [1, 2]. In recent years, it is proved to be feasible in many fields, such as sonar, radar, and communication system. With the proposal and improvement of related algorithms for different scenarios, the estimate accuracy of DOA is gradually improved, the complexity is also gradually reduced, and more sources could be detected [3–7]. However, existing DOA estimation techniques are usually based on scalar sensor array structures with equal element spacing not greater than half-of-wavelength, which is also called uniform linear array [8]. This array structure can obtain unambiguous DOA estimates while suffering from heavy mutual coupling effects and low system resolution [9]. Simultaneously, only relying on DOA information received by scalar sensor array, it is hard to distinguish different signals impinging on the array from similar directions of arrival. Meanwhile, polarization-sensitive array (PSA) [10] is proposed to obtain another two-dimensional information combined with DOA information, which can distinguish different signals in higher dimensions, especially signals with similar incident angles. Many kinds of polarization-sensitive sensors have been designed for receiving different electromagnetic vector signals, like triad [11], cocentered orthogonal loop and dipole (COLD) [12], and crossed dipoles [13]. Relative algorithms are also proposed for measuring joint DOA and polarization estimates [14–16].

In addition, coprime array (CA) [17, 18] is proposed by extending the array aperture and reducing mutual coupling

effect sparse array, in order to improve the array performance. It solves the problem of compact array structure in uniform linear array, whose phase ambiguity can be eliminated by vectorizing the covariance matrix of the received signal to construct a virtual array. Based on CA, some high-performance DOA estimation algorithms have been proposed in the field of PSA [19, 20]. In [19], the CA is split up into two uniform linear arrays, and a minimum distance parallel factor algorithm is proposed to obtain an accurate estimate, though the advantage of high degrees of freedom (DOF) in CA is not exploited. In order to take full advantage of the large virtual aperture of the CA, a coarray interpolation method is proposed in [20]. However, it suffers from high computational complexity when solving the large stacked recovered covariance matrix.

Augmented CA (ACA) [21], as an improved array, is equipped with more virtual elements than other CAs after vectorization and a closed-form formula in its virtual array's continuous part can be derived, which is more suitable for covariance matrix vectorization. Furthermore, compared with the covariance matrix which only uses second-order statistics of the received signals, fourth-order cumulants can demonstrate more information of non-Gaussian signals [22]. Besides, by constructing the fourth-order cumulants of the received signals, the array aperture of the equivalent virtual array is significantly extended [23], while the noise is suppressed in the fourth-order cumulant due to its Gaussian property [24]. As a result, fourth-order cumulant and its corresponding algorithms have shown great advantages in parameter estimation. However, the CA research on PSA and the corresponding algorithms has just begun, and more and more updated array structure designs and high-precision estimation algorithms are required in this field.

In this paper, we introduce the augmented CA to polarization-sensitive arrays in electromagnetic environment and design a kind of new sparse array called polarization-sensitive ACA (PSACA). Each sensor element is equipped with a pair of orthogonal dipoles, which are orthogonal to each other while the array element location is the same as ACA. On the other hand, a fourth-order cumulant reduced-dimensional Capon (FOC-RD-Capon) algorithm is proposed for non-Gaussian signals. First, the fourth-order cumulant is constructed by received signals, replacing the covariance matrix in traditional Capon algorithm, which can obtain a large amount of consecutive virtual elements. Benefited from the fourth-order cumulant, DOF are increased and virtual array aperture is extended in contrast to the virtual array by vectorization of covariance matrix. Second, considering the high calculation burden of three-dimensional spectral peak search for DOA and polarization estimates, a reduced-dimensional Capon algorithm is proposed, which only requires one-dimensional spectral peak search. The accuracy of the algorithm can be improved by shortening the search interval. The proposed algorithm is proved to have much lower complexity than traditional Capon algorithm. Numerous simulations illustrate the effectiveness of the ACA in polarization-sensitive scenarios and the proposed FOC-RD-Capon algorithm.

To summarize, the contributions of this paper are as follows:

(1) We design the ACA in polarization-sensitive scenario, where each sensor element is a pair of orthogonal dipoles to receive electric field strength vector. Compared with traditional CA, ACA can achieve higher DOF.

(2) We use the fourth-order cumulant signal instead of second-order cumulant signal for non-Gaussian signal, which can demonstrate more information to obtain high estimation performance. Moreover, the equivalent virtual array for the fourth-order cumulant signal has more consecutive elements.

(3) We propose a low-complexity reduced-dimensional algorithm for polarization-sensitive augmented CA, avoiding three-dimensional peak search. Meanwhile, the proposed algorithm enjoys high estimation accuracy with multi-parameter autopairing.

The rest of the paper is arranged as follows.

Section 2 demonstrates the data model of the signal impinging on the PSA. Section 3 introduces the array structure of PSACA and its virtual array element model after employing the fourth-order cumulant. Section 4 elaborates the proposed reduced-dimensional Capon algorithm. Section 5 depicts the simulation results, and Section 6 concludes the paper.

Notations. We use lower-case (upper-case) bold character to imply vector (matrix). $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ denote the conjugate, transpose, and the conjugate transpose of a matrix or vector, respectively. $\otimes$ denotes the Kronecker product, and $\odot$ represents Khatri–Rao product. $(\cdot)^{-1}$ represents matrix inverse. angle$(\cdot)$ means phase operator. diag$\{\cdot\}$ means diagonalization operator. $\langle a, b\rangle$ means the integer set from a to b, and $E[\cdot]$ means the mathematical expectation of a vector or matrix.

2. Data Model

Consider that K far-field non-Gaussian narrowband signals impinge on the linear array with M sensors where each sensor element consists of a pair of orthogonal electric dipoles to ensure vector receiving capability. The two dipoles with different polarization modes in an element are supposed to be distributed in the direction along with x-axis and y-axis, respectively, which are able to receive the electric component in the corresponding direction. The elevation angles of the signals are $\theta_k \in [-\pi/2, \pi/2]$, $k \in \langle 1, K\rangle$. The corresponding polarization information in the signal can be modeled as two polarization parameters: polarization auxiliary angle $\gamma_k \in [0, \pi/2]$, $k \in \langle 1, K\rangle$ and polarization phase difference $\eta_k \in [-\pi, \pi)$, $k \in \langle 1, K\rangle$. As a result, the single snapshot data model of the received signal on the array is written as [16]

$$\mathbf{y}(t) = [\mathbf{a}_1 \otimes \mathbf{s}_1, \mathbf{a}_2 \otimes \mathbf{s}_2, \ldots, \mathbf{a}_K \otimes \mathbf{s}_K]\mathbf{b}(t) + \mathbf{n}(t), \quad (1)$$

where $\mathbf{a}_k = [e^{j2\pi d_1 \sin\theta_k/\lambda}, e^{j2\pi d_2 \sin\theta_k/\lambda}, \ldots, e^{j2\pi d_M \sin\theta_k/\lambda}]$ denotes the directional vector and $d_m, m \in \langle 1, M\rangle$ is the location of $m-$th sensor element. $\mathbf{b}(t) \in \mathbb{C}^{K\times 1}$ represents the non-Gaussian signal vector while $\mathbf{n}(t) \in \mathbb{C}^{M\times 1}$ is the white Gaussian noise vector. $\mathbf{s}_k \in \mathbb{C}^{2\times 1}$ symbolizes the polarization-space steering vector of $k-$th signal for orthogonal electric dipoles, which can be represented as [13]

$$\begin{aligned}\mathbf{s}_k &= \mathbf{\Phi}(\theta_k)\boldsymbol{\omega}(\gamma_k, \eta_k)\\ &= \begin{bmatrix} 0 & -1 \\ \cos\theta_k & 0 \end{bmatrix}\begin{bmatrix} \sin\gamma_k e^{j\eta_k} \\ \cos\gamma_k \end{bmatrix},\end{aligned} \tag{2}$$

where $\Phi(\theta_k)$ denotes the coordinate matrix only containing DOA information, which is determined by the polarization mode of the electric dipoles, and $\omega(\gamma_k, \eta_k)$ is the polarization vector. Assume that J snapshots are received during a period of time for parameter estimation, and the overall received signal is modeled as [25]

$$\mathbf{Y} = [\mathbf{A} \odot \mathbf{S}]\mathbf{B} + \mathbf{N}, \tag{3}$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_K]$, $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K]$ are the directional matrix and electric matrix, respectively. $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_J] \in \mathbb{C}^{K\times J}$ denotes the non-Gaussian signal vector while $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_J] \in \mathbb{C}^{2M\times J}$ symbolizes the additive Gaussian noise matrix.

Note that the number of the signals to be estimated in this paper is known. If the number is uncertain, methods such as matrix decomposition [26], information theory [27], or Geist's circle [28] can be used for estimation.

3. ACA and Virtual Elements

Based on the array model in [21], Figure 1 depicts the array structure of the ACA, which is composed of two subarrays. The two subarrays are both uniform linear arrays. Subarray 1, which is marked with white rectangles, is equipped with N array elements with adjacent interval $M\lambda/2$. Meanwhile, subarray 2 is marked with black circles, whose total array elements are $2M$ with adjacent interval $N\lambda/2$. The total element number of the ACA is $2M + N - 1$ because the first element of subarray 1 and subarray 2 is located at the same place. ACA is considered as the improvement of traditional CA because it keeps subarray 1 unchangeable while extending the array aperture of subarray 2 twice as that of CA by equipping $2M$ sensor elements.

It can be obviously concluded from Figure 1 that the sensor element location of each subarray is that

$$\begin{cases} \mathbb{S}_1 = \langle 0, N-1\rangle M\lambda/2, \\ \mathbb{S}_2 = \langle 0, 2M-1\rangle N\lambda/2, \end{cases} \tag{4}$$

where $\langle a, b\rangle$ denotes the integers from a to b. Therefore, the overall position set of the sensor elements is described as $\mathbb{S}_{ACA} = \mathbb{S}_1 \cup \mathbb{S}_2$.

According to the data model for PSA presented in Section 2, the fourth-order cumulant matrix of received signal $\mathbf{Y}$ with J snapshots can be expressed as [24]

$$\begin{aligned}\mathbf{C}_4 = {} & E\left[(\mathbf{Y}\otimes\mathbf{Y}^*)(\mathbf{Y}\otimes\mathbf{Y}^*)^H\right] - E[(\mathbf{Y}\otimes\mathbf{Y}^*)]E\left[(\mathbf{Y}\otimes\mathbf{Y}^*)^H\right] \\ & - E\left[\mathbf{Y}\mathbf{Y}^H\right]\otimes E\left[\left(\mathbf{Y}\mathbf{Y}^H\right)^*\right].\end{aligned} \tag{5}$$

Assume that the signals are independent with each other, and fourth-order cumulant matrix can be considered as the covariance matrix of virtual received signal corresponding to an equivalent virtual array whose element location set is the difference co-array of the physical sensor elements, which can be defined as

$$\mathbb{S}_{ACA\_4} = \left\{d_i - d_j | d_i, d_j \in \mathbb{S}_{ACA}\right\}. \tag{6}$$

Figure 2 demonstrates the virtual array structure after fourth-order cumulant extension for ACA and CA with the same physical sensor elements. There are both 12 physical sensor elements with $M = 4$, $N = 5$ in ACA and $M = 7$, $N = 6$ in CA. As depicted in the figure, CA has 25 consecutive virtual elements while ACA has 47 consecutive virtual elements, which is obviously more than that of CA. Benefited from the merit, ACA can achieve high DOF to detect more signals [29].

4. FOC-RD-Capon Algorithm

4.1. Algorithm Process Introduction. Capon algorithm can obtain multi-parameter estimates simultaneously by a multi-dimensional peak search, which has been widely used in DOA estimation. However, the direct introduction of Capon algorithm from scalar sensor array to polarization-sensitive array will result in the dramatic increase in search dimension, which causes the inevitable high-dimensional search process. To reduce the calculation burden of the Capon algorithm in polarization-sensitive arrays while improving the estimation performance for non-Gaussian signals, a FOC-RD-Capon algorithm is proposed in this section. The introduction will be presented as follows.

Traditional Capon search function for DOA and polarization estimation is expressed as [30]

$$F_{\text{Capon}}\left(a_1, a_2, \cdots, a_Q\right) = \frac{1}{\mathbf{M}^H\mathbf{R}^{-1}\mathbf{M}}, \tag{7}$$

where $\mathbf{R}$ is the covariance matrix of the received signal and $\mathbf{M}$ denotes the array manifold for Q parameters. In PSA scenarios, the array manifold includes three parameters, DOA θ, polarization auxiliary angle γ, and polarization phase difference η, which means three-dimensional peak search is required, but it is unrealistic in engineering application. This part introduces a kind of parameter elimination method for polarization dimensional reduction.

Replacing the covariance matrix by the fourth-order cumulant matrix $\mathbf{C}_4$, define $\mathbf{M} = \mathbf{a}_c(\theta)\otimes\mathbf{s}(\theta, \gamma, \eta)$, where $\mathbf{a}_c(\theta)$ is the directional vector for virtual array structure which is the difference coarray of physical array. Then, the peak search function for PSA fourth-order cumulant scenario is given by

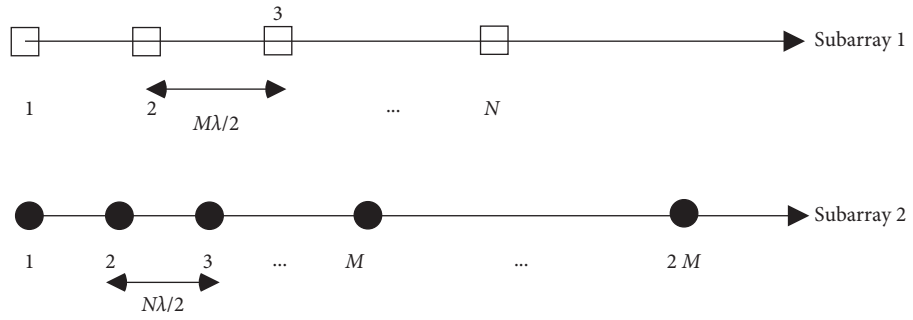


Figure 1: ACA structure.

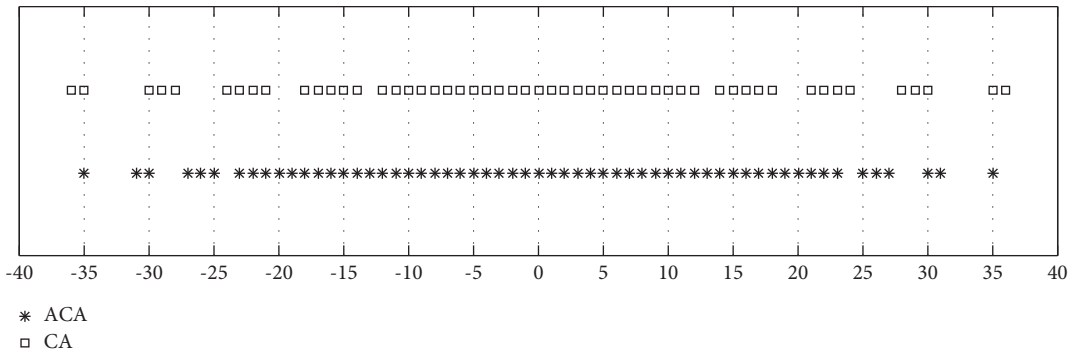


Figure 2: The virtual array structure after fourth-order cumulant extension for ACA and CA.

$$F_{\text{Capon}}(\theta,\gamma,\eta)=\frac{1}{\mathbf{M}(\theta,\gamma,\eta)^H\mathbf{C}_4^{-1}\mathbf{M}(\theta,\gamma,\eta)}. \tag{8}$$

Note that $\mathbf{s}=\Phi(\theta)\omega(\gamma,\eta)$ and a characteristic of Kronecker product is defined as $(\mathbf{AB})\otimes(\mathbf{C}\ \mathbf{D})=(\mathbf{A}\otimes\mathbf{C})(\mathbf{B}\otimes\mathbf{D})$ [25]. The array manifold can be reconstructed as $\mathbf{M}=[\mathbf{a}(\theta)\otimes\Phi(\theta)]\omega(\gamma,\eta)$; therefore, (8) is written as

$$F_{\text{Capon}}(\theta,\gamma,\eta)=\frac{1}{\boldsymbol{\omega}(\gamma,\eta)^H[\mathbf{a}_c(\theta)\otimes\boldsymbol{\Phi}(\theta)]^H\mathbf{C}_4^{-1}[\mathbf{a}_c(\theta)\otimes\boldsymbol{\Phi}(\theta)]\boldsymbol{\omega}(\gamma,\eta)}. \tag{9}$$

Classifying the terms in the function according to the parameters, the terms only about DOA is expressed as $\mathbf{U}(\theta)=[\mathbf{a}_c(\theta)\otimes\Phi(\theta)]^H\mathbf{C}_4^{-1}[\mathbf{a}_c(\theta)\otimes\Phi(\theta)]$. Besides, there is an identity $\omega(\gamma,\eta)^H\omega(\gamma,\eta)=1$ that holds. Then, the peak search function has been transformed into the minimum finding under constraints that

$$F_m(\sigma,\boldsymbol{\omega})=\boldsymbol{\omega}^H\mathbf{U}\boldsymbol{\omega}+\sigma\left(1-\boldsymbol{\omega}^H\boldsymbol{\omega}\right). \tag{10}$$

In order to find the minimum value of (10), we calculate the partial derivative of $F_m(\sigma,\omega)$ about ω^H that

$$\frac{\partial F_m(\sigma,\boldsymbol{\omega})}{\partial\boldsymbol{\omega}^H}=\mathbf{U}\boldsymbol{\omega}-\sigma\boldsymbol{\omega}. \tag{11}$$

Let (11) equal 0 to calculate the final estimates that $\mathbf{U}\omega=\sigma\omega$. Instead of solving the equations, we assume it as the form of eigenvalue and eigenvector pairs, where σ is the eigenvalue and ω is the corresponding eigenvector. Based on the analysis above, the peak search function (9) is expressed as

$$\begin{aligned}F_{\text{Capon}}(\sigma,\boldsymbol{\omega})&=\frac{1}{\boldsymbol{\omega}^H\mathbf{U}\boldsymbol{\omega}}\\&=\frac{1}{\boldsymbol{\omega}^H\sigma\boldsymbol{\omega}}.\end{aligned} \tag{12}$$

Because σ is a constant, (12) is actually the search about the eigenvalue [32].

$$\begin{aligned}F_{\text{Capon}}(\sigma)&=\frac{1}{\sigma\boldsymbol{\omega}^H\boldsymbol{\omega}}\\&=\frac{1}{\sigma_{\mathbf{U},\min}}.\end{aligned} \tag{13}$$

Every peak calculation can be considered as finding the minimum eigenvalue about $\mathbf{U}$ which is only constructed by DOA information. Polarization information is eliminated in Capon search function to tremendously reduce the computational complexity.

After finding K peaks about the DOA estimates $\widehat{\theta}_k, k \in \langle 1, K\rangle$, the corresponding $\mathbf{U}$ is obtained as

$$\mathbf{U}\left(\widehat{\theta}_k\right) = \left[\mathbf{a}_c\left(\widehat{\theta}_k\right) \otimes \mathbf{\Phi}\left(\widehat{\theta}_k\right)\right]^H \mathbf{C}_4^{-1} \left[\mathbf{a}_c\left(\widehat{\theta}_k\right) \otimes \mathbf{\Phi}\left(\widehat{\theta}_k\right)\right]. \quad (14)$$

Polarization matrix $\widehat{\omega}_k$ is measured by the eigen decomposition of $\mathbf{U}(\widehat{\theta}_k)$. Note that $\mathbf{U}(\widehat{\theta}_k) \in \mathbb{C}^{2\times 2}$ has 2 eigenvalue and eigenvector pairs, and the eigenvector whose eigenvalue is the smallest is the estimate $\widehat{\omega}_k$. Ultimately, polarization estimates are obtained as

$$\begin{cases} \widehat{\gamma}_k = \arctan\left(abs\left(\boldsymbol{\omega}_k^{[1]}/\boldsymbol{\omega}_k^{[2]}\right)\right), \\ \widehat{\eta}_k = \text{angle}\left(\boldsymbol{\omega}_k^{[1]}/\boldsymbol{\omega}_k^{[2]}\right), \end{cases} \quad (15)$$

where $k \in \langle 1, K\rangle$ and $\omega^{[i]}$ denotes the $i-$th element in vector ω.

The main steps of the proposed FOC-RD-Capon algorithm are summarized as follows:

Step 1. Compute the fourth-order cumulant matrix of received signal $\mathbf{C}_4$ according to (5).

Step 2. Obtain the peak search function with (8) and reconstruct it according to the characteristic of Kronecker product (9).

Step 3. Establish the search dictionary $\theta_i \in [-\pi/2, \pi/2]$ to calculate $\mathbf{U}(\theta_i)$ with each dictionary element θ_i.

Step 4. Perform eigen decomposition of $\mathbf{U}(\theta_i)$ and find its smallest eigenvalue $\sigma_{\mathbf{U},\min}$.

Step 5. Find K peaks in (13) to estimate $\widehat{\theta}_k$, $k \in \langle 1, K\rangle$.

Step 6. Calculate the eigenvector $\widehat{\omega}_k$ corresponding to the smallest eigenvalue in $\mathbf{U}(\theta_i)$, $k \in \langle 1, K\rangle$.

4.2. Discussion. The calculation burden of the FOC-RD-Capon algorithm mainly results from the following steps. Assume that there are P sensor elements in the ACA. Computing the fourth-order cumulant of the received signal requires $O[J(2P)^4]$. The inverse of fourth-order cumulant matrix needs the complexity of $O[(2P)^6]$. Every peak search consists of $O(64P^2L)$ for L-times peak search. Therefore, it can be summed up that the proposed FOC-RD-Capon algorithm is composed of the complexity of $O(64P^2L + 16JP^4 + 64P^6)$.

Meanwhile, the traditional Capon algorithm without dimensional reduction requires three-dimensional peak search. Computing the fourth-order cumulant and inverse of it requires totally $O[J(2P)^4 + (2P)^6]$, which is the same as the proposed algorithm. However, the three-dimensional peak search needs the complexity of $O(64P^2L^3)$ for L-times search in each dimension. To sum up, the approximate calculation burden of the traditional Capon algorithm is $O(64P^2L^3 + 16JP^4 + 64P^6)$, which is obviously higher than that of the proposed algorithm. Figure 3 shows the comparison of the complexity of the two algorithms with the number of sensor elements.

In addition, ACA is employed for the proposed algorithm due to the following merits:

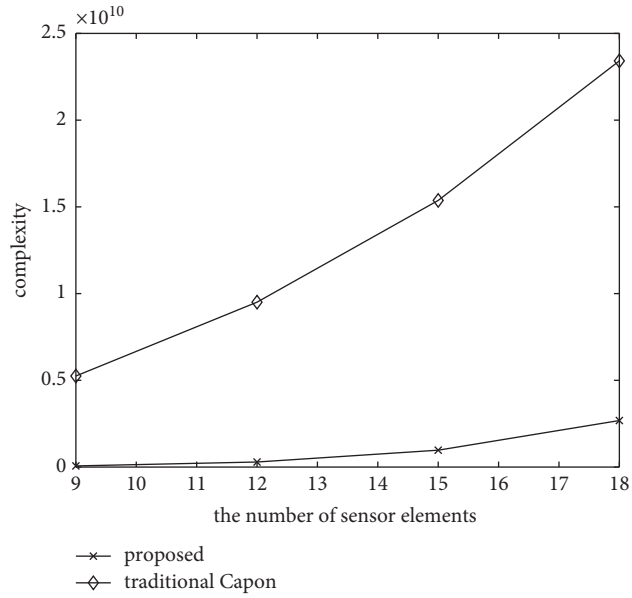


Figure 3: The complexity comparison between the proposed algorithm and the traditional Capon algorithm.

(1) ACA is a kind of sparse array, where array aperture is extended and mutual coupling effect is eliminated. In practice, the coprime numbers M and N can be chosen according to the requirement to satisfy different scenarios. Benefited from the sparse structure, the DOA and polarization estimation results are tremendously improved.

(2) The proposed algorithm uses the fourth-order cumulant statistic instead of the second-order cumulant, where the received signal can be regarded as the signal received by the equivalent virtual array. The virtual array is the difference coarray of physical array. Based on the conclusion and the analysis in Section 3, ACA can achieve more consecutive elements than CA, which means being able to detect more signals.

Cramér–Rao bound (CRB) is often used as the standard error which can be calculated. We derive the CRB formula to evaluate the RMSE performance of the algorithms on the designed ACA. Define $\mathbf{A}_s = \mathbf{A} \odot \mathbf{S}$, $\Pi_{\mathbf{A}_s}^{\perp} = \mathbf{I}_{2P} - \mathbf{A}_s(\mathbf{A}_s^H\mathbf{A}_s)^{-1}\mathbf{A}_s^H$, and $\mathbf{P} = \sum_{j=1}^{J} \mathbf{b}_j\mathbf{b}_j^H/J$, and the formula of CRB for polarization-sensitive ACA is demonstrated as [32]

$$\text{CRB} = \frac{\kappa^2}{2J}\left\{\text{Re}\left[\left(\mathbf{D}^H\Pi_{\mathbf{A}_s}^{\perp}\mathbf{D}\right) \oplus \mathbf{P}^T\right]\right\}^{-1}, \quad (16)$$

where $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_K, \mathbf{e}_1, \ldots, \mathbf{e}_K, \mathbf{f}_1, \ldots, \mathbf{f}_K]$, $\mathbf{d}_k = \partial(\mathbf{a}_k \otimes \mathbf{s}_k)/\partial\theta_k$, $\mathbf{e}_k = \partial(\mathbf{a}_k \otimes \mathbf{s}_k)/\partial\gamma_k = \mathbf{a}_k \otimes \Gamma_k\mathbf{s}_k$, $\Gamma_k = \text{diag}\{\sin\gamma_k, \cos\gamma_k\}$, $\mathbf{f}_k = \partial(\mathbf{a}_k \otimes \mathbf{s}_k)/\partial\eta_k = \mathbf{a}_k \otimes \Xi_k\mathbf{s}_k$, $\Xi_k = \text{diag}\{0, j\eta_k\}$. κ^2 is the power of noise.

5. Simulation Results

Numerous simulations are performed to verify the effectiveness of the designed ACA and proposed FOC-RD-Capon

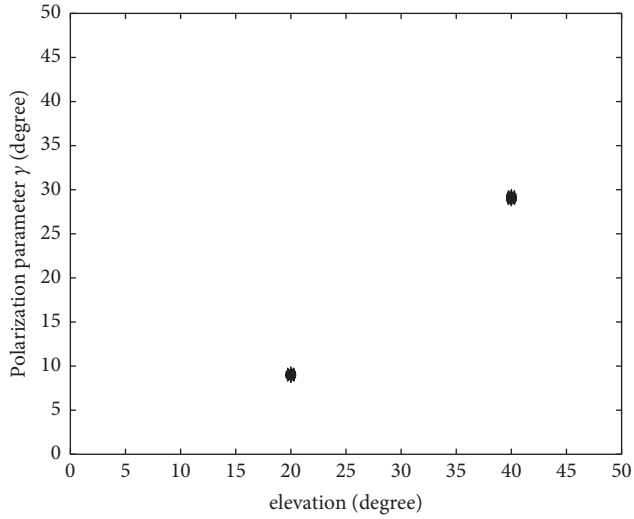


Figure 4: Scatter plot of DOA and polarization parameter γ.

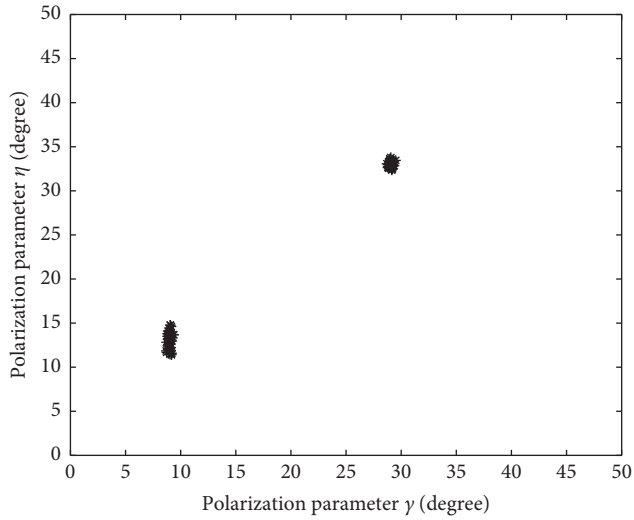


Figure 5: Scatter plot of polarization parameters γ and η.

algorithm, where root mean square error (RMSE) is employed as an evaluation standard to judge the performance. RMSE is defined as

$$\mathrm{RMSE}_a = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{1}{L}\sum_{l=1}^{L}\left[\left(\widehat{a}_{k,l} - a_k\right)^2\right]}, \tag{17}$$

which is the RMSE of parameter a. L denotes the times of independent Monte Carlo simulations. a_k is the actual parameter, and $\widehat{a}_{k,l}$ is its estimate in $l-$th simulation. The two sinusoidal waveform signals are impinging on the ACA, and the DOA and polarization parameters are set as $(\theta_1, \gamma_1, \eta_1) = (20^\circ, 9^\circ, 13^\circ)$, $(\theta_2, \gamma_2, \eta_2) = (40^\circ, 29^\circ, 33^\circ)$. The array structure of ACA is $M = 4, N = 5$, and the total number of physical array elements is 12, which means subarray 1 is equipped with 5 sensor elements and subarray 2 has 8 sensor elements.

5.1. Scatter Plot of the Proposed Algorithm. Figures 4 and 5 depict the scatter plots of DOA and polarization estimates, where 100 independent Monte Carlo simulations are performed. Simulation environments are set as follows: the signal-to-noise ratio (SNR) is 15 dB and the number of snapshots J is 300. As is exhibited in the figures, DOA estimates and two polarization estimates are accurately measured and different parameters are correctly paired. In addition, it can be noticed that the variances of polarization estimates are higher than DOA estimates.

5.2. RMSE Performance of Different Algorithms versus SNR. Figures 6–8 demonstrate the RMSE performance of the proposed FOC-RD-Capon algorithm versus SNR, where CRB curve is given as the standard. Snapshots are fixed as 200 while SNR varies from −10 dB to 20 dB. Compared with the propagator method (PM) [33], estimating signal parameter via rotational invariance techniques (ESPRIT) [34],

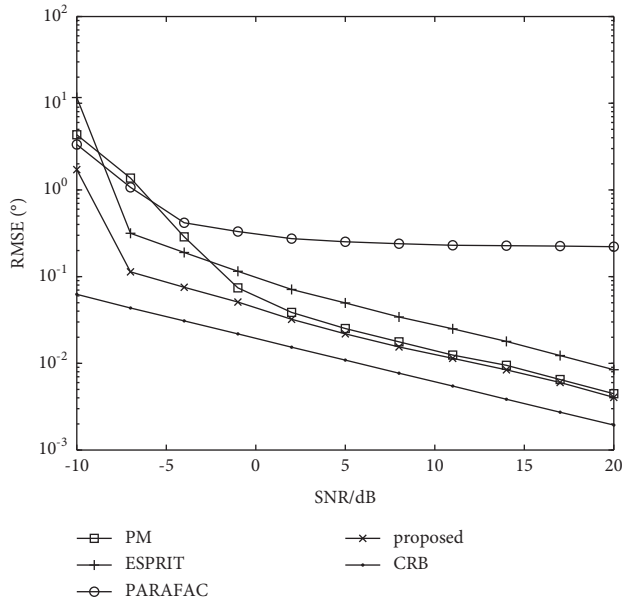


Figure 6: DOA estimation of different algorithms versus SNR.

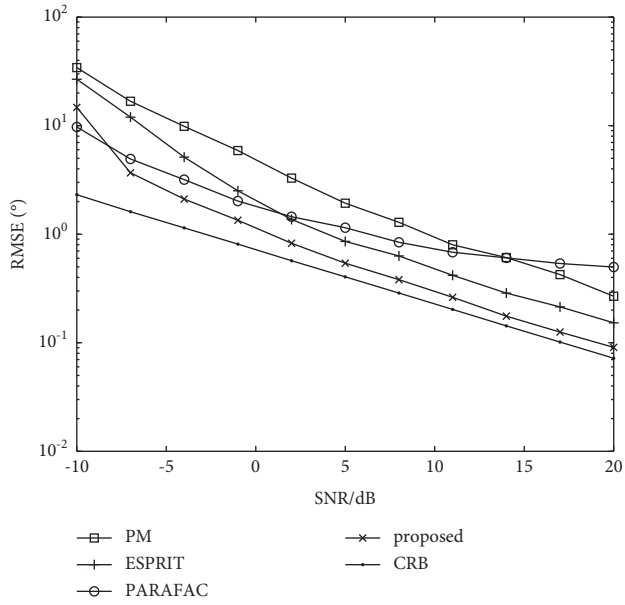


Figure 7: γ estimation of different algorithms versus SNR.

and parallel factor (PARAFAC) [35] algorithms, the proposed algorithm has the lowest RMSE, and its curve is closest to the CRB. Meanwhile, the RMSE of all the four algorithms decreases as the SNR improves, which indicates its influence to the estimates.

5.3. RMSE Performance of Different Algorithms versus Snapshots. Figures 9–11 show the RMSE performance versus snapshots. Similar to Section 5.2, PM, ESPRIT, and PARAFAC algorithms are used for comparison and CRB curve is depicted as the standard. With the increase of

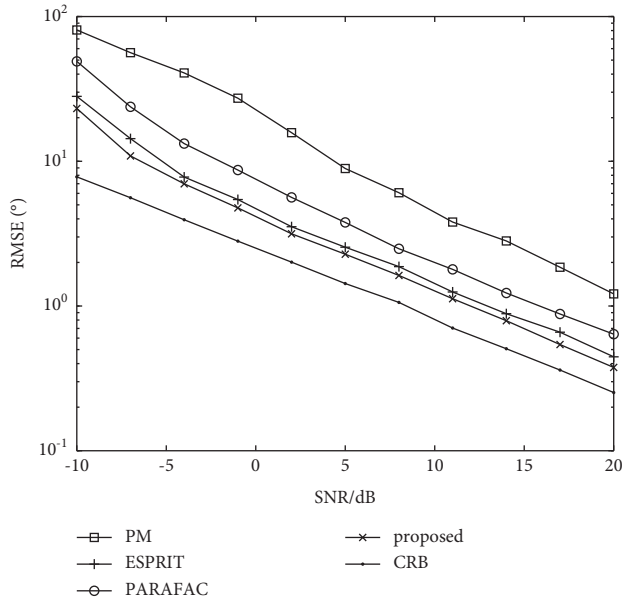


Figure 8: η estimation of different algorithms versus SNR.

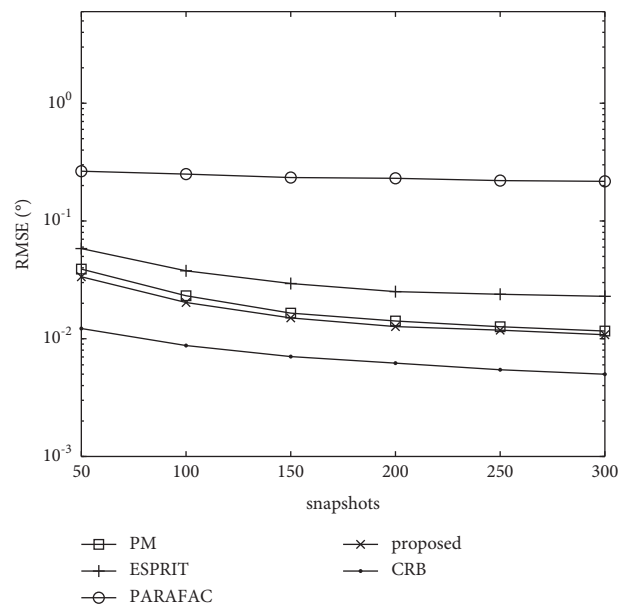


Figure 9: DOA estimation of different algorithms versus snapshot.

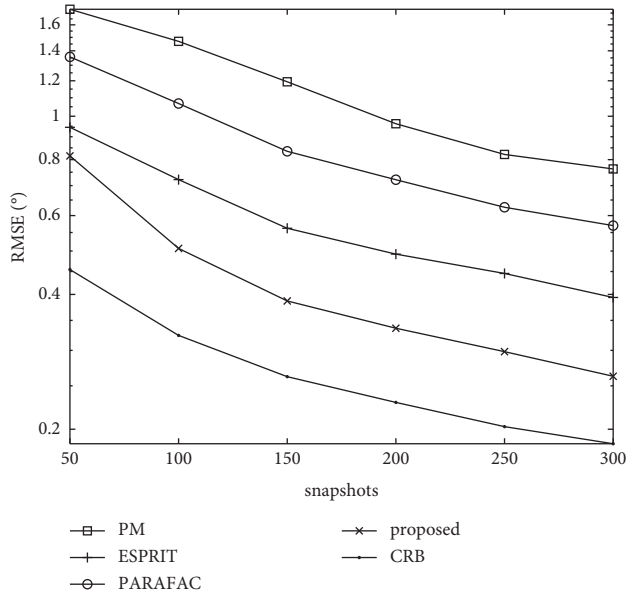


Figure 10: γ estimation of different algorithms versus snapshot.

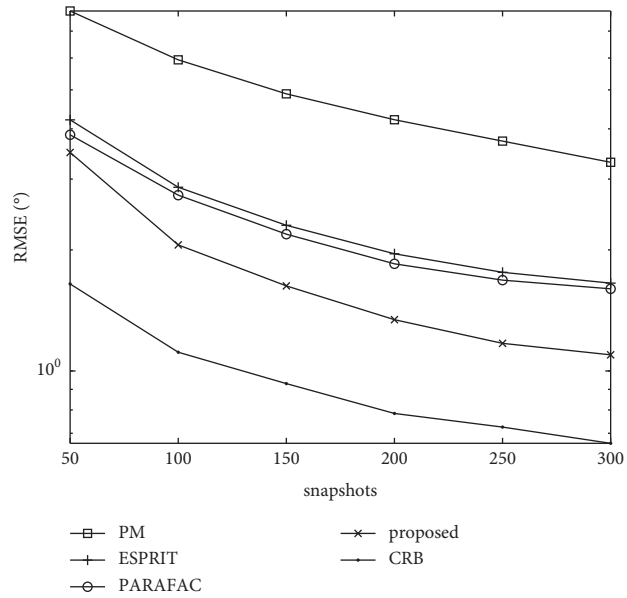


Figure 11: η estimation of different algorithms versus snapshot.

snapshots, the estimation accuracy of all the algorithms is improved. Meanwhile, the proposed FOC-RD-Capon algorithm has the least RMSE whatever snapshots are, which shows its great performance.

6. Conclusion

In this paper, we make use of the inherent information in fourth-order cumulant of non-Gaussian signals, proposing a FOC-RD-Capon algorithm for PSACA. Compared with the scalar sensor arrays, each sensor element of PSASA is equipped with a pair of orthogonal electric dipoles to receive vector signals, which is able to obtain joint DOA and polarization estimates. The proposed algorithm reduces the search dimension of Capon method from three dimensions to one dimension, tremendously reducing the calculation burden. According to the analysis and the numerous simulations, the effectiveness of the proposed FOC-RD-Capon algorithm is confirmed.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by China NSF (61971217, 61971218, and 61631020), Jiangsu NSF (BK20200444), and National Key Research and Development Project (2020YFB1807602).

References

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[2] F. Wen, "Computationally efficient DOA estimation algorithm for MIMO radar with imperfect waveforms," *IEEE Communications Letters*, vol. 23, no. 6, pp. 1037–1040, 2019.

[3] X. Zhang, L. Xu, L. Xu, and D. Xu, "Direction of departure (DOD) and direction of arrival (DOA) estimation in MIMO radar with reduced-dimension MUSIC," *IEEE Communications Letters*, vol. 14, no. 12, pp. 1161–1163, 2010.

[4] Z. Yang, L. Xie, and C. Zhang, "Off-Grid direction of arrival estimation using sparse bayesian inference," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 38–43, 2013.

[5] X. Wu, W.-P. Zhu, and J. Yan, "A toeplitz covariance matrix reconstruction approach for direction-of-arrival estimation," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 8223–8237, 2017.

[6] J. Shi, Z. Yang, and Y. Liu, "On parameter identifiability of diversity-smoothing-based MIMO radar," *IEEE Transactions on Aerospace and Electronic Systems*, to be published, 2021.

[7] J. Cong, X. Wang, C. Yan, L. T. Yang, M. Dong, and K. Ota, "CRB Weighted Source Localization Method Based on Deep Neural Networks in Multi-UAV Network," *IEEE Internet of Things Journal*, 2022.

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

[9] B. D. Steinberg, *Principles of Aperture and Array System Design: Including Random and Adaptive arrays*, Wiley, Hoboken, NJ, USA, 1976.

[10] J. W. P. Ng and A. Manikas, "Polarisation-sensitive array in blind MIMO CDMA system," *Electronics Letters*, vol. 41, no. 17, pp. 970–972, 2005.

[11] K. T. Wong, "Direction finding/polarization estimation-dipole and/or loop triad(s)," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 2, pp. 679–684, 2001.

[12] H. Chen, W. Wang, and W. Liu, "Joint DOA, range, and polarization estimation for rectilinear sources with a COLD array," *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1398–1401, 2019.

[13] X. Zhang, C. Chen, J. Li, and D. Xu, "Blind DOA and polarization estimation for polarization-sensitive array using dimension reduction MUSIC," *Multidimensional Systems and Signal Processing*, vol. 25, no. 1, pp. 67–82, 2014.

[14] B. Hochwald and A. Nehorai, "Polarimetric modeling and parameter estimation with applications to remote sensing," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 1923–1935, 1995.

[15] M. Hurtado and A. Nehorai, "Polarimetric detection of targets in heavy inhomogeneous clutter," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1349–1361, 2008.

[16] B. Zhong and N. Wang, "DOA Estimation Algorithm for Wideband Polarization Sensitive arrays," in *Proceedings of the 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pp. 2516–2520, IEEE, Xi'an, China, May 2018.

[17] C. Zhou, Y. Gu, S. He, and Z. Shi, "A robust and efficient algorithm for coprime array adaptive beamforming," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1099–1112, 2017.

[18] P. P. Vaidyanathan and P. Pal, "Sparse sensing with Co-prime samplers and arrays," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573–586, 2011.

[19] A. Tanveer and X. Zhang, "DOA estimation for coprime EMVS arrays via minimum distance criterion based on PARAFAC analysis," *IET Radar, Sonar & Navigation*, vol. 13, no. 1, pp. 65–73, 2019.

[20] M. Fu, Z. Zheng, W.-Q. Wang, and H. C. So, "Coarray interpolation for DOA estimation using coprime EMVS array," *IEEE Signal Processing Letters*, vol. 28, pp. 548–552, 2021.

[21] K. Zhang, C. Shen, H. Li et al., "Direction of arrival estimation and robust adaptive beamforming with unfolded augmented coprime array," *IEEE Access*, vol. 8, pp. 22314–22323, 2020.

[22] B. Porat and B. Friedlander, "Direction Finding Algorithms Based on High-Order statistics," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pp. 2675–2678, IEEE, Albuquerque, NM, USA, April 1990.

[23] F.-G. Yan, X.-W. Yan, J. Shi et al., "MUSIC-like direction of arrival estimation based on virtual array transformation," *Signal Processing*, vol. 139, pp. 156–164, 2017.

[24] M. C. Dogan and J. M. Mendel, "Applications of cumulants to array processing. II. Non-Gaussian noise suppression," *IEEE Transactions on Signal Processing*, vol. 43, no. 7, pp. 1663–1676, 1995.

[25] M. D. Zoltowski and K. T. Wong, "Closed-form eigenstructure-based direction finding using arbitrary but identical subarrays on a sparse uniform Cartesian array grid," *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 2205–2210, 2000.

[26] A. Angzhao Di, "Multiple source location--A matrix decomposition approach," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 33, no. 5, pp. 1086–1091, 1985.

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

[28] H. T. Hsien-Tsai Wu, J. F. Jar-Ferr Yang, and F. K. Fwu-Kuen Chen, "Source number estimators using transformed Gerschgorin radii," *IEEE Transactions on Signal Processing*, vol. 43, no. 6, pp. 1325–1333, 1995.

[29] P. Pal and P. P. Vaidyanathan, "Coprime sampling and the MUSIC algorithm," in *Proceedings of the Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*, pp. 289–294, Sedona, AZ, USA, January 2011.

[30] P. Stoica, P. Handel, and T. Soderstrom, "Study of Capon method for array signal processing," *Circuits, Systems, and Signal Processing*, vol. 14, no. 6, pp. 749–770, 1995.

[31] L. Wang, L. Yang, and G. Wang, "Uni-vector-sensor dimensionality reduction MUSIC algorithm for DOA and polarization estimation," *Mathematical Problems in Engineering*, vol. 2014, no. 21, pp. 682472.1–682472.9, 2014.

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

[33] S. Marcos, A. Marsal, and M. Benidir, "The propagator method for source bearing estimation," *Signal Processing*, vol. 42, no. 2, pp. 121–138, 1995.

[34] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT--A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 34, no. 5, pp. 1340–1342, 1986.

[35] T. D. Pham and J. Möcks, "Beyond principal component analysis: a trilinear decomposition model and least squares estimation," *Psychometrika*, vol. 57, no. 2, pp. 203–215, 1992.

Hindawi
International Journal of Antennas and Propagation
Volume 2022, Article ID 4038830, 7 pages
https://doi.org/10.1155/2022/4038830

Research Article

Mobile Performance Intelligent Evaluation of IoT Networks Based on DNN

Zhen Tang, Xiaobin Fu, and Pingping Xiao

College of Physical Science and Engineering, Yichun University, Yichun 336000, Jiangxi, China

Correspondence should be addressed to Xiaobin Fu; fxb8508457@163.com

Received 15 December 2021; Revised 14 January 2022; Accepted 21 February 2022; Published 16 March 2022

Academic Editor: Qi Liu

The rapid development of the sensor equipment has promoted the rapid growth of the Internet of Things (IoT). The IoT has been widely employed in the multidimensional signal processing and gradually formed the IoT networks. Mobile communication promotes the wide application of the IoT networks. In this study, the transmit antenna selection (TAS) scheme is employed to investigate the average symbol error probability (ASEP) performance of mobile IoT networks over the 2-Rayleigh channels. We first employ moment-generating function (MGF) approach to derive the exact ASEP expressions. We also investigate the outage probability (OP) performance and derive OP expressions. Employing the deep neural network (DNN), an OP intelligent prediction algorithm is proposed. Then, the numerical simulations are conducted to confirm the ASEP and OP performance analysis. The effect of different channel parameters is also analyzed. Compared with Nakagami and Rayleigh channel models, the 2-Rayleigh model has 83.6% and 59.1% increase in ASEP values, respectively. Compared with ELM and RBF models, the DNN model has 31.7% and 22.5% increase in OP prediction accuracy, respectively.

1. Introduction

In recent years, the information society is facing the transformation from information age to intelligent age. Sensor industry is developing towards the direction of intelligence, low power consumption, and high precision. The intelligent sensor industry plays an important supporting role in the development of the Internet of Things (IoT) and gradually forms the IoT networks [1–3]. With the extensive application of mobile communication, the IoT networks are widely used in transportation, agriculture, and manufacturing [4, 5]. The perspectives and challenges of physical Internet employed in the IoT networks were investigated in [6]. Due to IoT device mobility, the IoT networks are facing many challenges. A new secure user authenticated scheme was employed in the IoT environment to establish a secure transmission process [7]. In the transmission of the biomedical information, smartphone was employed to power the electrochemical biosensing dongle in the IoT networks [8].

Multiple-input multiple-output (MIMO) is an important method to obtain the high data-rate in mobile IoT networks [9]. To improve the system throughput, a Kalman filtering combining scheme was employed in the MIMO system [10]. To improve the power efficiency, the robust beamforming was designed in magnetic MIMO systems [11]. In [12], a deep neural network-based linear precoding method was proposed for multiuser MIMO systems, which can achieve a higher downlink rate. With sparse channels, a low complexity parameter estimation method was proposed for MIMO big data communication system [13]. In [14], the authors considered covert MIMO communications and investigated the covert capacity with variational distance constraint. The sparse Bayesian learning method was used to realize the channel estimation in industrial IoT networks [15].

Transmit antenna selection (TAS) has been widely used in the MIMO system to reduce complexity [16, 17]. With deep neural network, the data-driven prediction method was employed to achieve the TAS in the MIMO system [18]. In

[19], the exact average symbol error probability (ASEP) and outage probability (OP) expressions for underwater MIMO system with TAS were derived. The TAS was employed to improve the ASEP performance of underlay spectrum-sharing MIMO system in [20]. In [21], using the moment-generating function (MGF) method, the TAS scheme was employed to derive the error rate expressions for space-time line code system.

However, the TAS performance of the IoT networks over the 2-Rayleigh channels is very difficult. Motivated by the above discussion, employing TAS and SC schemes, the ASEP and OP performance of the mobile IoT networks is investigated. The main contributions are as follows:

(1) The MGF approach was employed to derive the exact ASEP expressions with TAS, which are used to investigate the ASEP performance of different modulation methods.

(2) Employing the CDF-based approach, the OP expressions are also derived. The derived OP results of the 2-Rayleigh model are more complicated than those of Nakagami and Rayleigh channel models. However, they have a high computational complexity.

(3) Employing the deep neural network (DNN), an OP intelligent prediction algorithm is proposed in this study. The prediction algorithm can achieve rapid analysis of OP performance.

(4) Through different conditions, the effect of different channel parameters is analyzed. Compared with Nakagami and Rayleigh channel models, the 2-Rayleigh model has 83.6% and 59.1% increase in ASEP values. Compared with ELM and RBF models, the DNN model has 31.7% and 22.5% increase in OP prediction accuracy, respectively.

Table 1 shows the notations.

Table 1: Notations.

Notations	Designation
PDF	Probability density function
MSE	Mean square error
SC	Selection combining
CDF	Cumulative density functions
SNR	Signal-to-noise ratio

2. The Mobile IoT Networks

In the mobile IoT networks, it has a mobile source (S) and mobile destination (D). There are K antennas at S and L antennas at D. E_s is the transmit power. Figure 1 shows the system model.

Firstly, Si transmits the signal. The received SNR of Dj is given as

$$\gamma_{ij} = \left|h_{ij}\right|^2 \frac{E_s}{N_0}, \quad i = 1, \ldots, K, j = 1, \ldots, L, \tag{1}$$

where h_{ij} is the channel gain, which follows the 2-Rayleigh model and N_0 is the noise power.

The average SNR is given as

$$\begin{aligned} \overline{\gamma_{ij}} &= E\left(\left|h_{ij}\right|^2\right)\frac{E_s}{N_0} \\ &= E\left(\left|h_{i1}\right|^2\right)e^{-(j-1)W}\frac{E_s}{N_0}, \end{aligned} \tag{2}$$

where W is the power attenuation factor.

Karagiannidis et al. [22] give the CDF and PDF of γ_{ij} as

$$\begin{aligned} F_{\gamma_{ij}}(r) &= G_{1,3}^{2,1}\left[\frac{r}{\overline{\gamma_{ij}}}\middle|_{1,1,0}^{1}\right], \\ f_{\gamma_{ij}}(r) &= \frac{1}{r}G_{0,2}^{2,0}\left[\frac{r}{\overline{\gamma_{ij}}}\middle|_{1,1}^{-}\right]. \end{aligned} \tag{3}$$

D employs the SC combiner, which can select the best γ_{ij}. The output SNR γ_i is

$$\gamma_i = \max\left(\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{iL}\right). \tag{4}$$

To derive the ASEP and OP results, it needs to obtain the CDF and PDF of γ_i. The CDF of $\gamma_i = \max(\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{iL})$ is

$$\begin{aligned} F_{\mathrm{SC}}(r) &= \Pr\left[\gamma_i \le r\right] = \Pr\left[\max\left(\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{iL}\right) \le r\right] \\ &= \prod_{j=1}^{L} \Pr\left[\gamma_{ij} \le r\right] \\ &= \prod_{j=1}^{L} F_{\gamma_{ij}}(r) \\ &= \prod_{j=1}^{L} G_{1,3}^{2,1}\left[\frac{r}{\overline{\gamma_{ij}}}\middle|_{1,1,0}^{1}\right]. \end{aligned} \tag{5}$$

With the derivative of (5), it can obtain the PDF of γ_i

$$f_{\mathrm{SC}}(r) = \sum_{j=1}^{L}\frac{1}{r}G_{0,2}^{2,0}\left[\frac{r}{\overline{\gamma_{ij}}}\middle|_{1,1}^{-}\right] \times \prod_{\substack{p=1 \\ p\neq j}}^{L} G_{1,3}^{2,1}\left[\frac{r}{\overline{\gamma_{ip}}}\middle|_{1,1,0}^{1}\right]. \tag{6}$$

S employs the TAS scheme. For K antennas at S, the g is selected as

$$g = \arg\max\left(\gamma_i\right), \quad 1 \le i \le K. \tag{7}$$

Redha [23] gives the PDF of γ_g

$$f_{\gamma_g}(r) = K\left[F_{\mathrm{SC}}(r)\right]^{K-1} f_{\mathrm{SC}}(r). \tag{8}$$

So, it can obtain the MGF of γ_g [21]:

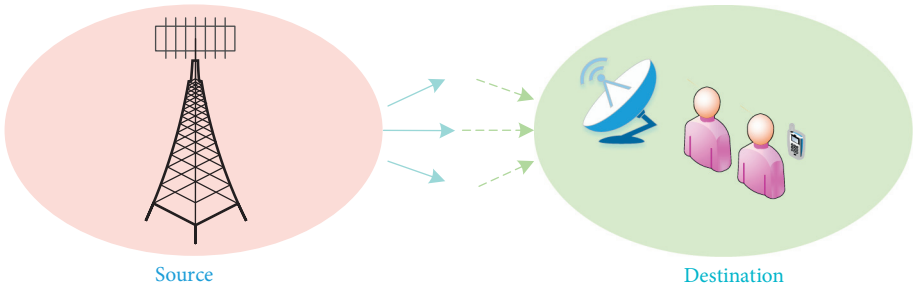


Figure 1: The TAS/SC model.

$$M_{\gamma_g}(s) = \int_0^{\infty} e^{-sr} f_{\gamma_g}(r)\mathrm{d}r$$
$$= K \int_0^{\infty} e^{-sr} \times \left[\prod_{j=1}^{L} G_{1,3}^{2,1}\left[\frac{r}{\gamma_{ij}}\Big|_{1,1,0}^{1}\right]\right]^{K-1} \times \left(\sum_{j=1}^{L} \frac{1}{r} G_{0,2}^{2,0}\left[\frac{r}{\gamma_{ij}}\Big|_{1,1}^{-}\right] \times \prod_{\substack{p=1 \\ p\neq j}}^{L} G_{1,3}^{2,1}\left[\frac{r}{\gamma_p}\Big|_{1,1,0}^{1}\right]\right)\mathrm{d}r. \tag{9}$$

3. The ASEP and OP Performance

Yilmaz and Kucur [24] give the ASEP of q-ary PSK:

$$P_{\mathrm{ASEP}} = \frac{1}{\pi} \int_0^{(q-1)\pi/q} M_{\gamma_g}\left(\frac{\sin^2(\pi/q)}{\sin^2\theta}\right)\mathrm{d}\theta. \tag{10}$$

Yilmaz and Kucur [24] give the ASEP of q-ary QAM:

$$PP_{\mathrm{ASEP}} = \frac{4(\sqrt{q}-1)}{\pi\sqrt{q}} \int_0^{\pi/2} M_{\gamma_g}\left(\frac{3}{2(q-1)\sin^2\theta}\right)\mathrm{d}\theta$$
$$- \frac{4(\sqrt{q}-1)^2}{\pi q} \int_0^{\pi/4} M_{\gamma_g}\left(\frac{3}{2(q-1)\sin^2\theta}\right)\mathrm{d}\theta. \tag{11}$$

The OP is

$$P_{\mathrm{out}} = \Pr\left[\gamma_g \le r_{\mathrm{th}}\right]$$
$$= \Pr\left[\arg\max_{1\le i\le K, 1\le j\le L}\left(\gamma_{ij}\right) \le r_{\mathrm{th}}\right]$$
$$= \prod_{i=1}^{K}\prod_{j=1}^{L} \Pr\left[\gamma_{ij} \le r_{\mathrm{th}}\right] \tag{12}$$
$$= \prod_{i=1}^{K}\prod_{j=1}^{L} \int_0^{r_{th}} f_{\gamma_{ij}}(r)\mathrm{d}r,$$

where r_{th} is a given threshold.

Employing (3) and (4), it can obtain the OP as

$$P_{\mathrm{out}} = \prod_{i=1}^{K}\prod_{j=1}^{L} G_{1,3}^{2,1}\left[\frac{r_{\mathrm{th}}}{\gamma_{ij}}\Big|_{1,1,0}^{1}\right]. \tag{13}$$

4. OP Prediction Algorithm

However, (14) has a high computational complexity. In order to realize real-time analysis of OP performance, we propose an OP prediction method based on DNN.

From (13), K, L, r_{th}, and W constitute the DNN input X. The DNN output y is P_{out}.

The DNN structure is shown in Figure 2. It has six layers, which are input layer, two hidden layers, ReLU layer, sigmoid layer, and output layer, respectively.

For the input layer, it has four characteristics. Then, it uses full connection to connect with two hidden layers. The hidden layers have 512 and 1024 neurons, respectively. Meanwhile, we use ReLU and Sigmoid functions after hidden layer 1 and hidden layer 2, respectively. To make regression prediction, Sigmoid function is employed to finally predict the output.

MSE is widely used as an evaluation criterion in deep learning. So, we also employ the MSE to evaluate the OP prediction effect. MSE is given as [16, 17]

$$\mathrm{MSE} = \frac{\sum_{i=1}^{S}(t_i - y_i)^2}{S}, \tag{14}$$

where t_i is the predicted output and S is the number of the testing date.

5. Simulation Analysis

Figures 3 and 4 present the impact of W on the ASEP and OP performance, respectively. The parameters are in Table 2. We can obtain that the ASEP and OP performance is degraded as W increases. For example, when SNR = 16 dB, the W values are 0.5, 2, and 4, respectively, the ASEP values are 9×10^{-4},

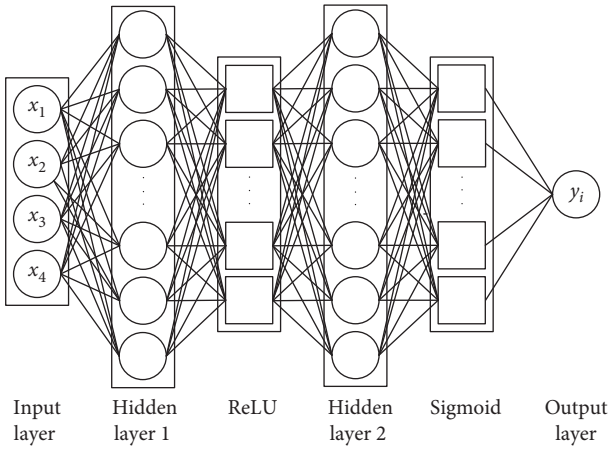


Figure 2: The DNN structure.

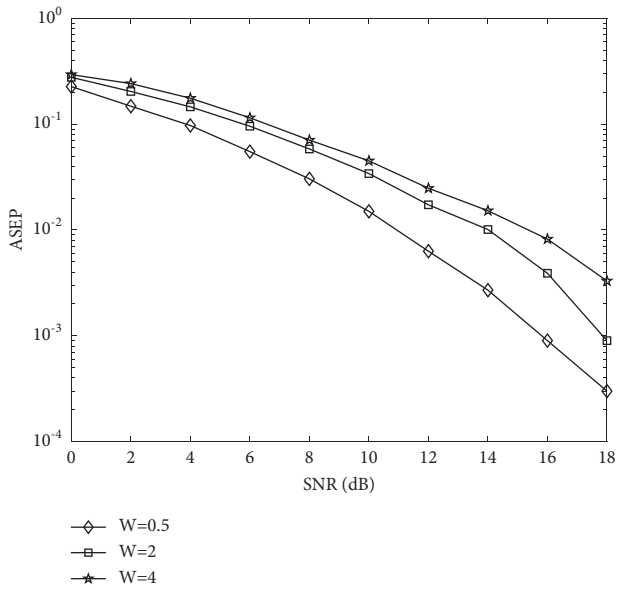


Figure 3: The ASEP performance with different (W).

3.9×10^{-3}, and 8.2×10^{-3}, respectively. With W increases, the fading severity is more serious.

Figure 5 presents the ASEP performance comparison. We can obtain that the ASEP performance of the 2-Rayleigh model is worse than that of Rayleigh and Nakagami models. When SNR = 10 dB, the ASEP values are 0.0450, 0.0184, and 0.0074, respectively. Compared with Rayleigh and Nakagami channel models, the 2-Rayleigh model has a 59.1% and 83.6% increase in ASEP values, respectively. This shows that the communications environment of the 2-Rayleigh model is more complex than that of Rayleigh and Nakagami models.

Figure 6 presents the OP performance comparison. When SNR = 6 dB, the OP values are 0.3214, 0.1902, and 0.0968, respectively. Compared with Nakagami and Rayleigh channel models, the 2-Rayleigh model has a 69.9% and 49.1% increase in OP values, respectively.

For the DNN, ELM, and RBF algorithms, Figures 7–9 show the prediction results, and Table 3 shows the

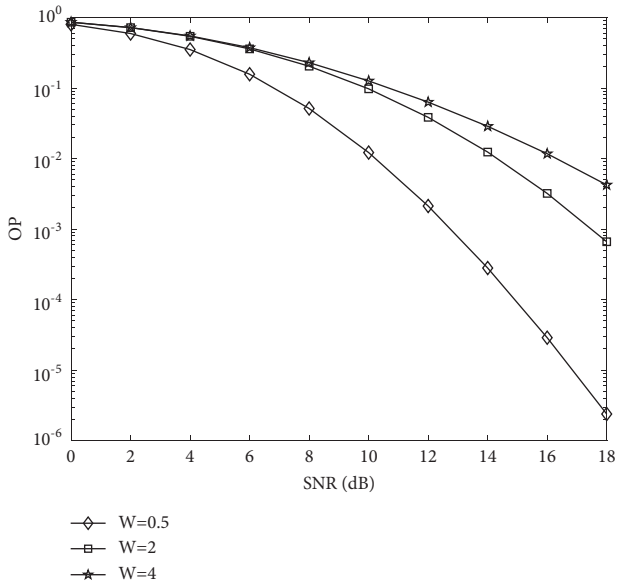


Figure 4: The OP performance with different (W).

Table 2: Simulation parameters for Figures 3–6.

Parameters	Value
E	1
W	0.5, 2, 4
K	2
L	2
γ_{th}	5 dB

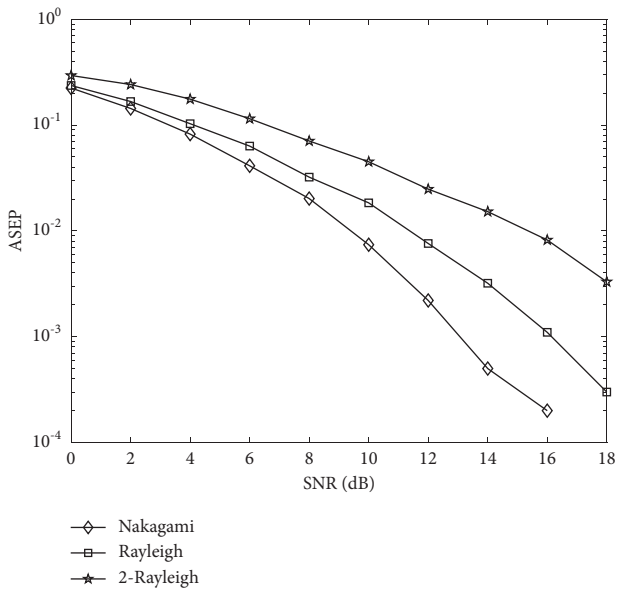


Figure 5: The ASEP performance comparison with different channel models.

simulation parameters. The prediction results of the DNN algorithm are better than ELM and RBF methods. Table 4 shows the MSE comparison. The MSE with the DNN

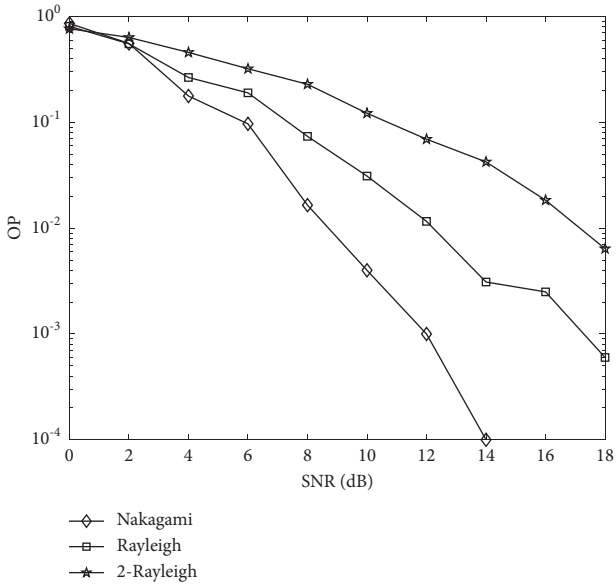


Figure 6: The OP performance comparison with different channel models.

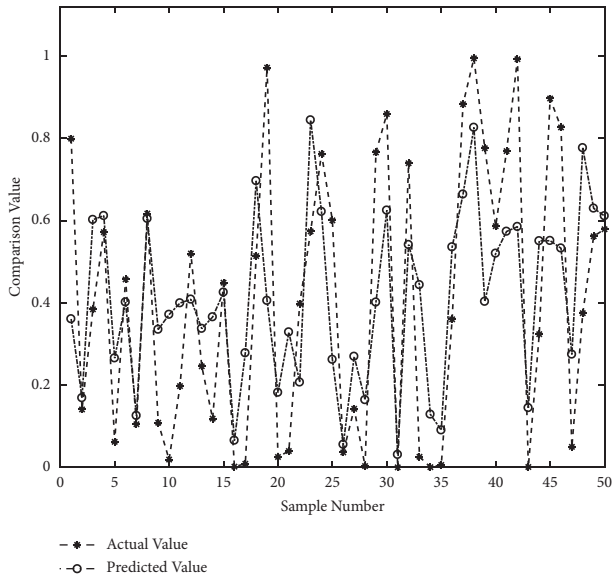


Figure 7: DNN prediction.

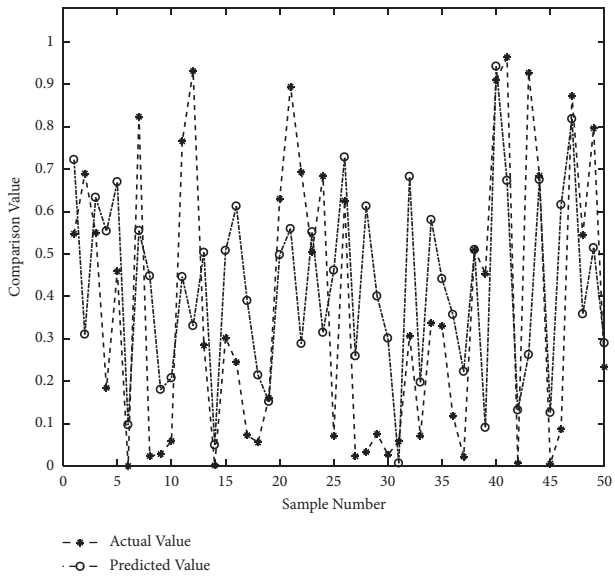


Figure 8: ELM prediction.

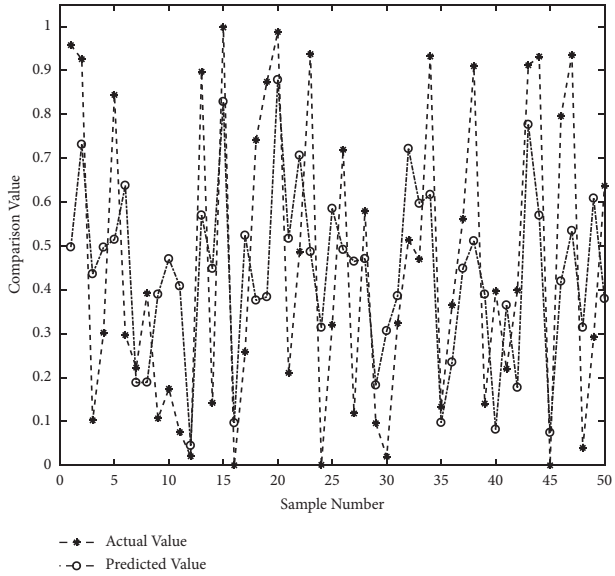


Figure 9: RBF prediction.

Table 3: Simulation parameters for Figures 7–9.

Parameters	DNN	ELM	RBF
	Training set: 1120; testing set: 50		
	X:4	X:4	X:4
	y:1	y:1	y:1
	q: [512, 1024]	q:[100]	q:[120]
	lr: 0.00024		

Table 4: MSE comparison for the three algorithms.

Algorithms	MSE
DNN	0.0557
ELM	0.0815
RBF	0.0719

algorithm is only 0.0557, which is lower than ELM and RBF methods. Compared with ELM and RBF models, the DNN model has 31.7% and 22.5% increase in OP prediction accuracy, respectively.

6. Conclusions

We investigated the ASEP and OP performance of the 2-Rayleigh model, respectively. Through different conditions, the effect of K, L, and W on the ASEP and OP performance was analyzed. K or L values increased, and the system performance was improved. The ASEP and OP performance degraded as W increased. Compared with Nakagami and Rayleigh channel models, the 2-Rayleigh model had 83.6% and 59.1% increase in ASEP values, respectively. The DNN model had a better OP prediction accuracy than the ELM and RBF algorithms.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request and with permission of funders.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This project was supported by the National Natural Science Foundation of China (no. 11664043).

References

[1] L. Xu, X. Zhou, Y. Tao, L. Liu, X. Yu, and N. Kumar, "Intelligent security performance prediction for IoT-enabled healthcare networks using an improved CNN," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 2063–2074, 2022.

[2] H. Wang, P. Xiao, and X. Li, "channel parameter estimation of mmWave MIMO system in urban traffic scene: a training channel-based method," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–9, 2022.

[3] L. Xu, H. Wang, and T. A. Gulliver, "Outage probability performance analysis and prediction for mobile IoV networks based on ICS-BP neural network," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3524–3533, 2021.

[4] X. P. Wang, L. T. Yang, D. D. Meng, M. Dong, K. Ota, and H. Wang, "Multi-UAV Cooperative Localization for Marine Targets Based on Weighted Subspace Fitting in SAGIN Environment," *IEEE Internet of Things Journal*, 2022.

[5] H. Wang, W. C. Du, X. P. Wang, and L. W. Xu, "Channel estimation performance analysis of FBMC/OQAM systems with bayesian approach for 5G-enabled IoT applications," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 2389673, 9 pages, 2020.

[6] H. Tran-Dang, N. Krommenacker, P. Charpentier, and D.-S. Kim, "Toward the Internet of Things for physical Internet: perspectives and challenges," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 4711–4736, 2020.

[7] A. Jabbari and J. B. Mohasefi, "A secure and LoRaWAN compatible user authentication protocol for critical applications in the IoT environment," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 1, pp. 56–65, 2022.

[8] J. Guo, "Smartphone-powered electrochemical biosensing dongle for emerging medical IoTs application," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 6, pp. 2592–2597, 2018.

[9] H. Wang, L. Wan, M. Dong, K. Ota, and X. Wang, "Assistant vehicle localization based on three collaborative base stations via SBL-based robust DOA estimation," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5766–5777, 2019.

[10] S. Park, "Kalman filtering based combining for MIMO systems with hybrid ARQ," *IEEE Transactions on Signal Processing*, vol. 69, no. 9, pp. 5250–5258, 2021.

[11] Y. Zhang, D. Chen, and T. Jiang, "Robust beamforming design for magnetic MIMO wireless power transfer systems," *IEEE Transactions on Signal Processing*, vol. 69, no. 9, pp. 5066–5077, 2021.

[12] K. Kong, W. J. Song, and M. Min, "Knowledge distillation-aided end-to-end learning for linear precoding in multiuser MIMO downlink systems with finite-rate feedback," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, Article ID 11095, 2021.

[13] H. Wang, L. Xu, Z. Yan, and T. A. Gulliver, "Low-complexity MIMO-FBMC sparse channel parameter estimation for industrial big data communications," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3422–3430, 2021.

[14] S.-Y. Wang and M. R. Bloch, "Covert MIMO communications under variational distance constraint," *IEEE Transactions ;on Information Forensics and Security*, vol. 16, no. 9, pp. 4605–4620, 2021.

[15] H. Wang, X. Li, R. H. Jhaveri et al., "Sparse bayesian learning based channel estimation in FBMC/OQAM industrial IoT networks," *Computer Communications*, vol. 176, pp. 40–45, 2021.

[16] L. Xu, X. Yu, and T. A. Gulliver, "Intelligent outage probability prediction for mobile IoT networks based on an IGWO-elman neural network," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 2, pp. 1365–1375, 2021.

[17] L. W. Xu, X. P. Zhou, X. W. Li, R. H Jhaveri, T. R. Gadekallu, and Y. Ding, "Mobile collaborative secrecy performance prediction for artificial IoT networks," *IEEE Transactions on Industrial Informatics*, 2021.

[18] H. Liu, Y. Xiao, P. Yang, J. Fu, S. Li, and W. Xiang, "Transmit antenna selection for full-duplex spatial modulation based on machine learning," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 10695–10708, 2021.

[19] I. S. Ansari, L. Jan, Y. T. Tang, L. Yang, and M. H. Zafar, "Outage and error analysis of dual-hop TAS/MRC MIMO RF-UOWC systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 10695–10708, 2021.

[20] R. Sarvendranath and N. B. Mehta, "Statistical CSI driven transmit antenna selection and power adaptation in underlay spectrum sharing systems," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 2923–2934, 2021.

[21] S. C. Lim and J. Joung, "Transmit antenna selection for space-time line code systems," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 786–798, 2021.

[22] G. K. Karagiannidis, N. C. Sagias, P. T. Mathiopoulos, and N. Nakagami, "A novel stochastic model for cascaded fading channels," *IEEE Transactions on Communications*, vol. 55, no. 8, pp. 1453–1458, 2007.

[23] M. R. Redha, "Impact of delayed arbitrary transmit antenna selection on the performance of rectangular QAM with receive MRC in fading channels," *IEEE Communications Letters*, vol. 13, no. 6, pp. 390–392, 2009.

[24] F. Yilmaz and O. Kucur, "Exact performance of wireless multi-hop transmission for M-ary coherent modulations over generalized gamma fading channels," in *Proceedings of the 2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–5, Cannes, France, September 2008.

Hindawi
International Journal of Antennas and Propagation
Volume 2022, Article ID 3949988, 9 pages
https://doi.org/10.1155/2022/3949988

Research Article

DOA Estimation Method of Weak Signal under the Compound Background of Strong Interference and Colored Noise

Bin Lin,[1,2] **Guoping Hu**,[1] **Hao Zhou**,[1] **Guimei Zheng**,[1] **and Yuwei Song**[1]

[1]*Air and Missile Defense College, Air Force Engineering University, Xi'an, Shaanxi 710051, China*
[2]*Graduate College, Air Force Engineering University, Xi'an, Shaanxi 710051, China*

Correspondence should be addressed to Guoping Hu; hgp6068@163.com

Received 3 December 2021; Accepted 8 January 2022; Published 4 February 2022

Academic Editor: Qi Liu

The traditional algorithm performing direction of arrival (DOA) estimation under the background of strong interference and colored noise has the problems of low estimation accuracy and small measurement targets. Based on the construction of a fourth-order cumulant (FOC) matrix to suppress colored noise, this paper adopts the extended noise subspace (ENS) algorithm and the fixed projection blocking (FPB) algorithm to estimate the DOA of weak targets. Firstly, a FOC matrix of the received signal vector is established to curb the noise component, and the eigenvalue decomposition is performed. Then, two approaches of weak signal DOA estimation are proposed. One approach is to merge the space where the strong interference steering vector lies into the noise subspace to construct an extended noise subspace, and then, the multisignal classification (MUSIC) algorithm is used to obtain the DOA estimation of the weak signal on the basis of the extended noise subspace. Another approach is to build the orthogonal projection matrix of the interference subspace as the interference blocking matrix, and the receiving array signal is preprocessed, and on the basis of it, the eigen decomposition is performed again to obtain the DOA information of the weak signal. Both algorithms make breakthroughs in the aperture limitation of the traditional algorithm, effectively expand the aperture, and promote the accuracy of estimation. The simulation tests the effectiveness of the proposed method.

1. Introduction

With the increasing intricacy of the electromagnetic environment, radar detection is interference by more and more electromagnetic, making it increasingly difficult for radar to detect targets [1, 2]. Strong electromagnetic interference will cover up the target's echo signal or overload the signal processor and receiver [3, 4]. When the power of the interference signal is greater than that of the echo signal, the received data is directly used in the estimation of the direction of arrival, and the false peak resulting from strong interference will be regarded as the peak of the source, which will give birth to the misinterpretation of the correct angle [5, 6]. Firstly, at present, one method to address this problem is signal separation, mainly including the Relaxation (Relax) algorithm [7] and the CLEAN algorithm [8]. This type of algorithm separates the output part of the array of all signals into multiple data blocks and then determines and removes the data blocks containing the interference signals according to the characteristics of the interference signal so as to achieve target recognition. However, this type of algorithm is extremely complex. Secondly, the jamming jam method (JJM) was proposed by Fang et al. and its extended application [9, 10]. The core concept is to use the angle information of the signal receiving matrix to establish an interference blocking matrix to eliminate the interference signal part of the received signal covariance matrix. However, the algorithm requires a precise prediction of the direction of interference. In addition, there is an Extended Noise Subspace algorithm [11]. The algorithm first constructs an extended noise subspace for strong interferences and noises, and then performs conventional DOA estimation. Dong et al. [12] deduced an interference blocking method based on modified projection. The interference direction is not a necessity in this approach, and it is of more practical use.

However, the abovementioned strong interference rejection algorithms are characterized by low estimation accuracy and small measurement targets in the direction finding under the background of nonideal colored noise. According to related literature, the calculation of high-order cumulants is blind Gaussian which can achieve aperture expansion and provides a powerful tool in response to Gaussian colored noise. Literature [13, 14] accordingly proposed a variety of FOC-based algorithms, which have better angle estimation performance under Gaussian colored noise conditions. Nonetheless, there are still some problems in terms of angular resolution and complexity. Tufail and Ahmed [15] used the FOC and ESPRIT algorithm to propose DOA estimation based on the genetic algorithm (GA) and obtained the multiple invariant cumulant ESPRIT algorithm, which has a better angular resolution, but the problem of excessive complexity remains unsolved. Literature [16] used the real-valued sparse Bayesian learning method to transform the FOC matrix into a real-valued matrix and simplify the algorithm through unitary transformation. The abovementioned FOC algorithm has made certain progress in the field of array signal processing, but the research on the background of strong interference is scarce.

To tackle the above problems, this paper puts forward the fourth-order cumulant-expanded noise subspace (FOC-ENS) algorithm and the fourth-order cumulant-fixed projection blocking (FOC-FPB) algorithm. One is to use the FOC matrix to substitute for the traditional covariance matrix. After eigenvalue decomposition, the space of the strong interference steering vector is incorporated into the noise subspace, and thus, the extended noise subspace is constructed. And the conventional MUSIC algorithm is used to obtain the DOA estimation of the weak signal on the basis of the extended noise subspace. The other method is to construct the orthogonal projection matrix of the strong interference signal subspace as the interference blocking matrix, and the received array signal is preprocessed. And the eigenvalue decomposition is performed again to obtain the noise subspace. Eventually, the MUSIC algorithm is used to search for spectral peaks to obtain the azimuth angle of the weak signal. Compared with traditional algorithms, this algorithm enjoys higher estimation accuracy under a variety of conditions.

2. Signal Model and FOC Vector Formulation

2.1. Signal Model. Assuming that K far-field narrowband signals and J strong interference signals $S_i(t)\,(i = 1, 2, \ldots, J, \ldots, K+J)$ are incident on a uniform linear array, the incident angle is θ_i, and if the first J is a strong interference signal, their power satisfies $\sigma_1^2 > \sigma_2^2 > \cdots > \sigma_J^2 \gg \sigma_{J+1}^2 > \cdots > \sigma_{K+J}^2$. The distance d of the array elements is half of the signal wavelength. The array is composed of M array elements, each of which is omnidirectional. The received signal of the m-th array element of the array at time t can be expressed as

$$x_m(t) = \sum_{i=1}^{K+J} s_i(t)\exp\{j(m-1)\pi\sin(\theta_i)\} + n_m(t), \tag{1}$$

where $n_m(t)$ is the Gaussian colored noise of the m-th element at time t and $s_i(t)$ is the complex envelope of the i-th source.

Assuming that the plane wave propagates along a straight line, the array response vector corresponding to the direction can be expressed as

$$\mathbf{a}(\theta_i) = \left[1,\ \exp(-j\pi\sin\theta_i), \ldots,\ \exp(-j\pi(M-1)\sin\theta_i)\right]^{\mathrm{T}}, \tag{2}$$

among which, $(\cdot)^{\mathrm{T}}$ represents the transpose of the matrix.

The direction matrix is defined as

$$\mathbf{A}(\theta) = \left[\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \ldots, \mathbf{a}(\theta_J), \ldots, \mathbf{a}(\theta_{\mathrm{K+J}})\right]. \tag{3}$$

This matrix is a $M \times (K+J)$ dimensional direction matrix.

The vector output of the array element is expressed as

$$\mathbf{X}(t) = \mathbf{A}(\theta)\mathbf{S}(t) + \mathbf{N}(t). \tag{4}$$

In the above formula, $\mathbf{S}(t)$ is the $(K+J) \times 1$ dimensional signal vector and $\mathbf{N}(t)$ is the $\mathrm{M} \times 1$ dimensional Gaussian colored noise vector.

2.2. Observation Model Based on FOC Vector. Compared with the second-order cumulant, the high-order cumulant can better characterize the signal characteristics and has the blind Gaussian property. Therefore, in the actual array signal processing, the fourth-order cumulant is frequently used for processing.

Regarding a uniform linear array, according to the symmetric definition of the FOC of the zero-mean stable random process, the FOC of the array received data is

$$\begin{aligned}&\mathrm{cum}\left(x_{k_1}, x_{k_2}, x_{k_3}^*, x_{k_4}^*\right)\\ &= E\left(x_{k_1}x_{k_2}x_{k_3}^*x_{k_4}^*\right) - E\left(x_{k_1}x_{k_3}^*\right)E\left(x_{k_2}x_{k_4}^*\right)\\ &\quad - E\left(x_{k_1}x_{k_4}^*\right)E\left(x_{k_2}x_{k_3}^*\right) - E\left(x_{k_1}x_{k_2}\right)E\left(x_{k_3}^*x_{k_4}^*\right).\end{aligned} \tag{5}$$

In the above formula, $(\cdot)^*$ represents conjugate and *cum*() represents cumulant.

According to literature [13], it is easy to get

$$\begin{aligned}\mathbf{C}_x = {} & E\left\{\left(\mathbf{X}\otimes\mathbf{X}^*\right)\left(\mathbf{X}\otimes\mathbf{X}^*\right)^{\mathrm{H}}\right\} -\\ & E\{\left(\mathbf{X}\otimes\mathbf{X}^*\right)\}E\left\{\left(\mathbf{X}\otimes\mathbf{X}^*\right)^{\mathrm{H}}\right\} - .\\ & E\left\{\left(\mathbf{XX}\right)^{\mathrm{H}}\right\}\otimes E\left\{\left(\mathbf{XX}^{\mathrm{H}}\right)^*\right\}\end{aligned} \tag{6}$$

In the above formula, $\otimes$ represents the Kronecker product.

Substituting the signal vector into the above formula can obtain the FOC matrix of the signal as follows:

$$\begin{aligned}\mathbf{C}_{\mathrm{s}} = {} & E\left\{\left(\mathbf{S}\otimes\mathbf{S}^*\right)\left(\mathbf{S}\otimes\mathbf{S}^*\right)^{\mathrm{H}}\right\}\\ & - E\{\left(\mathbf{S}\otimes\mathbf{S}^*\right)\}E\left\{\left(\mathbf{S}\otimes\mathbf{S}^*\right)^{\mathrm{H}}\right\}\\ & - E\left\{\left(\mathbf{SS}\right)^{\mathrm{H}}\right\}\otimes E\left\{\left(\mathbf{SS}^{\mathrm{H}}\right)^*\right\},\end{aligned} \tag{7}$$

where the signal FOC matrix $\mathbf{C}_s$ is a $K^2 \times K^2$ complex matrix, and its $(i+1)K + j$ row and $(k-1)K + l$ column elements can be expressed as

$$\mathrm{cum}\left(s_i, s_j, s_k^*, s_l^*\right), \forall i, j, k, l \in \{1, 2, \ldots, K\}. \tag{8}$$

With respect to an independent signal source, according to the nature of the higher-order cumulant, the elements in the FOC $\mathbf{C}_s$ of the signal are not zero if and only if $i = j = k = l$, and all other elements are zero, namely,

$$\mathbf{C}_s = \mathrm{cum}\left(s_i, s_j, s_k^*, s_l^*\right)\begin{cases} \neq 0, i = j = k = l, \\ = 0, \text{ otherwise}. \end{cases} \tag{9}$$

Hence, there are only K nonzero elements in $\mathbf{C}_s$, and they are situated at position $(k-1)K + k (k = 1, 2, \ldots, K)$ on the diagonal of $\mathbf{C}_s$. Now, the rows and columns are deleted with all zero elements in $\mathbf{C}_s$; $\mathbf{C}_s$ is reduced from the $K^2 \times K^2$ dimensional matrix to the $K \times K$ dimensional diagonal matrix. Then, the simplified signal FOC matrix can be expressed as

$$\mathbf{C}_s = \mathrm{diag}(\gamma_1, \gamma_2, \ldots, \gamma_k). \tag{10}$$

In the above formula, $\gamma_k = \mathrm{cum}(s_k, s_k, s_k^*, s_k^*)$.

It can also be obtained that the FOC matrix of noise is

$$\begin{aligned} \mathbf{C}_n = {} & E\left\{(\mathbf{N} \otimes \mathbf{N}^*)(\mathbf{N} \otimes \mathbf{N}^*)^H\right\} \\ & - E\{\mathbf{N} \otimes \mathbf{N}^*\} \cdot E\left\{(\mathbf{N} \otimes \mathbf{N}^*)^H\right\} \\ & - E\left\{\mathbf{N} \cdot \mathbf{N}^H\right\} \otimes E\left\{\left(\mathbf{N} \cdot \mathbf{N}^H\right)^*\right\}. \end{aligned} \tag{11}$$

In addition, the array steering vector after array expansion according to the FOC defined above is

$$\mathbf{b}(\theta) = \mathbf{a}(\theta) \otimes \mathbf{a}^*(\theta). \tag{12}$$

The expanded direction matrix is

$$\begin{aligned} \mathbf{B}(\theta) &= \left[\mathbf{b}(\theta_1) \ \mathbf{b}(\theta_2) \ \cdots \ \mathbf{b}(\theta_k)\right] \\ &= \left[\mathbf{a}(\theta_1) \otimes \mathbf{a}^*(\theta_1)\mathbf{a}(\theta_2) \otimes \mathbf{a}^*(\theta_2) \cdots \mathbf{a}(\theta_K) \otimes \mathbf{a}^*(\theta_K)\right]. \end{aligned} \tag{13}$$

If each signal source is completely independent, the following formula holds [17]:

$$\mathbf{C}x = \mathbf{B}(\theta)\mathbf{C}_s\mathbf{B}^{\mathrm{H}}(\theta) + \mathbf{C}n. \tag{14}$$

Assuming that the noise is Gaussian noise, whether it is white noise or colored noise with unknown spectral characteristics, the theoretical definition by the above formula $\mathbf{C}_n$ should always be a zero matrix. Due to noise deviation from Gaussian or finite data length calculation accuracy, $\mathbf{C}_n$ is a matrix approaching $\boldsymbol{0}$.

3. DOA Estimation Based on FOC Vector

Eigenvalue decomposition is performed on $\mathbf{C}x$, and its eigenvalues are arranged from large to small as $\lambda 1, \lambda 2, \ldots, \lambda_{M^2}$, and the corresponding eigenvector is $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{M^2}$, where the eigenvectors corresponding to the J large eigenvalues of the matrix $\mathbf{C}x$ are transformed into a fourth-order strong interference signal subspace [18–20]:

$$\mathbf{E}_J = \left[\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_J\right]. \tag{15}$$

The eigenvectors corresponding to the K larger eigenvalues are transformed into a fourth-order signal subspace:

$$\mathbf{E}_s = \left[\mathbf{e}_{J+1}, \mathbf{e}_{J+2}, \ldots, \mathbf{e}_{J+K}\right]. \tag{16}$$

The eigenvectors corresponding to the other $M^2 - K - J$ small eigenvalues are transformed into a fourth-order noise subspace:

$$\mathbf{E}_N = \left[\mathbf{e}_{J+K+1}, \mathbf{e}_{J+K+2}, \ldots, \mathbf{e}_{M^2}\right]. \tag{17}$$

3.1. Expanded Noise Subspace DOA Estimation. Taking the MUISC algorithm as an example, the spatial spectrum value of a certain angle is the reciprocal of the projection modulus of the steering vector in direction to the projection space, that is, the noise subspace [21–23]. The peak value represents that the steering vector at this angle projected in the noise subspace is smaller than its surrounding angle. The direction steering vector is merged into the noise subspace to form an expanded noise subspace [24]. When the expanded noise subspace is used for spectrum estimation, there must be no peak in this direction, and the steering vector of the weak signal will be in the expanded noise subspace. The projection of the steering vector of the weak signal on the extended noise subspace is a small nonzero value, and its reciprocal will generate a larger peak so that strong interference can be curbed, and the DOA of a weak signal can be estimated [11].

Accordingly, the feature vector of the interference signal can be incorporated into the noise subspace to form an extended noise subspace, namely,

$$\mathbf{E}_{JN} = \left[\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_J, \mathbf{e}_{J+\mathrm{K}+1}, \ldots, \mathbf{e}_{M^2}\right]. \tag{18}$$

Based on the interference-noise subspace, since the weak signal steering vector and the interference signal steering vector are often not orthogonal, in order to ensure the correct formation of the weak signal peak, the steering vector is transformed as follows:

$$\mathbf{c}(\theta) = \left(\mathbf{I} - \mathbf{E}_J\mathbf{E}_J^{\mathrm{H}}\right)\mathbf{b}(\theta), \tag{19}$$

where $\mathbf{b}(\theta)$ is the steering vector after the expansion of the FOC.

The conventional MUSIC method is adopted to search for spectral peaks and estimate the DOA of the weak signal:

$$P_1(\theta) = \frac{\mathbf{c}^{\mathrm{H}}(\theta)\mathbf{c}(\theta)}{\mathbf{c}^{\mathrm{H}}(\theta)\mathbf{E}_{JN}\mathbf{E}_{JN}^{\mathrm{H}}\mathbf{c}(\theta)}. \tag{20}$$

In the above formula, since the total signal subspace is orthogonal to the noise subspace, there is $\mathbf{E}_J^{\mathrm{H}}\mathbf{E}_N = 0$. If θ_i is the incident angle of the weak signal, then

$$\mathbf{E}_N\left(\mathbf{I} - \mathbf{E}_J\mathbf{E}_J^{\mathrm{H}}\right)\mathbf{a}(\theta_i) = \mathbf{E}_N\mathbf{a}(\theta_i) - 0 = 0. \tag{21}$$

So the weak signal steering vector is orthogonal to the interference-noise subspace after transformation by equation (20).

Finally, the angle corresponding to the maximum point obtained through the spatial spectrum is the incident direction of the weak signal.

According to the above analysis, the FOC-ENS algorithm is summarized, as shown in Table 1.

3.2. Modified Projection Blocking Method DOA Estimation. After the division of the strong interference signal subspace, signal subspace, and noise subspace according (after formulas (15)–(17)) to the subspace theory, a modified projection blocking matrix $\boldsymbol{G}$ orthogonal to the strong interference signal subspace is established [12] as follows:

$$\mathbf{G} = \mathbf{I} - \mathbf{E}_J\left(\mathbf{E}_J^{\mathrm{H}}\mathbf{E}_J\right)^{-1}\mathbf{E}_J. \tag{22}$$

The global subspace is performed to get

$$\mathbf{D} = \mathbf{GE} = \mathbf{G}\left[\mathbf{E}_J, \mathbf{E}_S, \mathbf{E}_N\right] = \left[\mathbf{0}_J, \mathbf{D}_{K+1}, \mathbf{D}_{K+2}, \ldots \mathbf{D}_{M^2}\right]. \tag{23}$$

That is, the matrix $\boldsymbol{D}$ after the modified projection blocking transformation is merely related to the weak signal and noise and has nothing to do with the interference, achieving the suppression of the strong interference. Therefore, in order to avoid interference, the transformed data vector should be

$$\mathbf{Y} = \mathbf{GC}x. \tag{24}$$

The data covariance matrix after modified projection transformation is

$$\mathbf{R}_{\mathrm{Y}} = \mathbf{GER}_{\mathrm{s}}\mathbf{E}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}} + \sigma^2\mathbf{GG}^{\mathrm{H}}. \tag{25}$$

After the modified projection transformation, the strong interference is blocked, thereby eliminating the influence of the strong interference on the weak signal. Then, eigenvalue decomposition is performed on the receiving matrix again to obtain a new signal subspace $\mathbf{E}_S'$ and noise subspace $\mathbf{E}_N'$. At this time, $\mathbf{E}_S'$ is only correlated with the weak signal, $\mathbf{E}_N'$ orthogonal to the steering vector in the direction of the weak signal, and the spatial spectrum function is used as follows [25–27]:

$$P_2(\theta) = \frac{1}{\mathbf{b}^{\mathrm{H}}(\theta)\mathbf{E}_N'\mathbf{E}_N'\mathrm{H}\mathbf{b}(\theta)}. \tag{26}$$

Finally, the azimuth angle of the weak signal can be estimated with the help of the spectral peak search.

According to the above analysis, the algorithm of this paper is summarized, as depicted in Table 2.

4. Simulation Results and Analysis

Suppose the signal-to-noise ratio of the signal is defined as $10\log_{10}(\sigma_k^2/\sigma_n^2)$, the interference-to-signal ratio is defined as $10\log_{10}(\sigma_{\mathrm{j}}^2/\sigma_{\mathrm{k}}^2)$, and the interference-to-noise ratio is defined as $10\log_{10}(\sigma_{\mathrm{j}}^2/\sigma_{\mathrm{n}}^2)$. Among them, σ_{k}^2 is the power of the k-th signal, σ_{j}^2 is the power of the strong interference signal, and σ_{n}^2 is the noise power.

The root mean square error (RMSE) is

$$RMSE = \sqrt{\frac{1}{KN}\sum_{i=1}^{N}\sum_{k=1}^{K}\left(\left|\theta_i - \theta_{ik}\right|^2\right)}. \tag{27}$$

Among them, N is the number of Monte-Carlo experiments, K is the number of weak signals, and θ_i and θ_{ik} are the true value and estimated value of the azimuth angle of the i-th signal in the k-th experiment.

4.1. Spatial Spectrum Estimation. Experiment 1 sets the incident angles of 3 target signals at −30°, 0°, and 30°, respectively; the incident angles of 2 strong interference signals are −60° and 60°; the number of array elements is 10; the signal-to-noise ratio (SNR) is 10 dB; the signal-to-interference ratio (SIR) is 30 dB; the number of snapshots is 200. The advantages of the proposed algorithm in DOA estimation accuracy are analyzed. Figure 1 illuminates the spatial spectrum curves of the conventional FOC-MUSIC algorithm, JJM algorithm, FOC-ENS algorithm, and FOC-FPB algorithm.

Figure 1 illustrates that the conventional FOC-MUSIC algorithm is inclined to regard the false peak formed by the strong interference signal as the peak of the real target, thus failing to estimate the arrival angle of the target. Although the JJM algorithm can suppress strong interference signals, it has a mediocre effect on the estimation of the target's arrival angle, and it requires precise prediction of the direction of the strong interference signal. The FOC-ENS algorithm can better estimate the direction of arrival of the target. However, the direction of strong interference will exist during the estimation of small spikes, which will affect DOA estimation under certain conditions. The FOC-FPB algorithm can determine the target wave arrival angle best.

4.2. Comparative Analysis of Errors of Different Algorithms. Experiment 2 sets the number of array elements at 10, the SIR is 30 dB, and the number of Monte Carlo experiments is 500. The errors of different algorithms in the Monte Carlo simulation experiment are also analyzed. Figure 2(a) illuminates that under the condition of SNR = 10 dB, the number of snapshots varies from 50 to 500 in step of 50 and the variation curves of the RMSE of the JJM algorithm, ENS algorithm, FPB algorithm, FOC-ENS algorithm, and FOC-FPB algorithm with the number of snapshots. Figure 2(b) shows the curve of the RMSE of the five algorithms changing from −10 dB to 10 dB in step of 2 dB under the condition of 200 snapshots.

It can be inferred from Figure 2 that the RMSE of these methods dwindles as the SNR and the number of snapshots grow. In a composite background, when the number of snapshots and the SNR climb to a certain extent, the RMSE of several algorithms tends to be stable. Due to the suppression of spatial colored noise, the FOC-ENS algorithm and the FOC-FPB algorithm are better than other algorithms in low SNR.

Table 1: FOC-ENS algorithm basic steps.

FOC-ENS algorithm
Step 1: estimate the data covariance matrix R of the array from the output vector $\mathbf{X}(t)$ of the array element
Step 2: perform eigen decomposition on the covariance matrix $\mathbf{C}x$ and arrange the eigenvalues in descending order so as to obtain interference subspace $\mathbf{E}_J$, signal subspace $\mathbf{E}_s$, and noise subspace $\mathbf{E}_N$
Step 3: take a set of orthogonal bases $\mathbf{E}_J$ determined by interference in the total signal subspace and merge it into the noise subspace $\mathbf{E}_N$ to form the interference-noise subspace $\mathbf{E}_{JN}$
Step 4: based on the interference-noise subspace $\mathbf{E}_{JN}$, perform a spectral peak search on $P_1(\theta)$ to estimate the DOA of the weak signal

Table 2: FOC-FPB algorithm basic steps.

FOC-FPB algorithm
Step 1 and Step 2 are the same as the FOC-ENS algorithm
Step 3: construct a modified projection matrix $\mathbf{G}$ orthogonal to the interference steering vector
Step 4: preprocess the received data vector to get $\boldsymbol{Y}$
Step 5: similarly duplicate Step 2 to do the second eigenvalue decomposition to get the noise subspace $\mathbf{E}_N^{'}$
Step 6: perform a spectral peak search on $P_2(\theta)$ and estimate the DOA of the weak signal

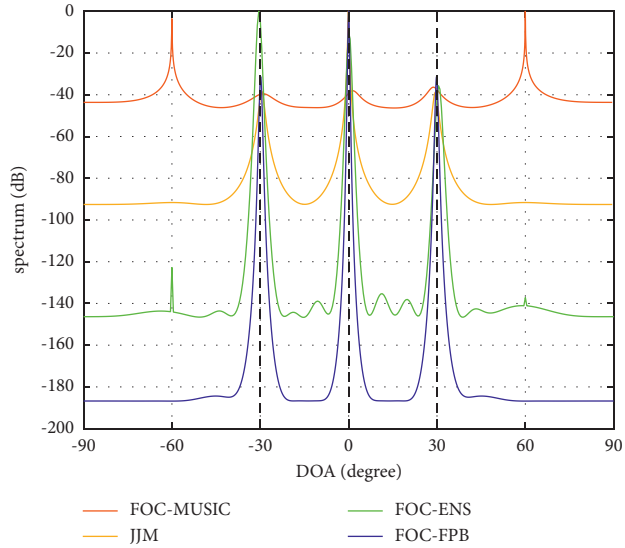


Figure 1: Comparison of spatial spectrum curves.

4.3. Comparative Analysis of Errors under Different Numbers of Interference Sources. Experiment 3 sets the number of array elements at 10, the SIR is 30 dB, the number of snapshots is 200, and the number of Monte Carlo experiments is 500. The RMSE of the proposed algorithm under different numbers of interference sources is analyzed. Figure 3(a) shows the SNR of the FOC-ENS algorithm from −10 dB to 10 dB in step of 2 dB and the curves of the RMSE with the SNR under the different numbers of interference sources. Figure 3(b) depicts the FOC-FPB algorithm SNR changes from −10 dB to 10 dB in step of 2 dB, reflecting the variation curves of the RMSE with the SNR under the different numbers of interference sources.

As presented in Figure 3, the two algorithms can suppress interference and accurately estimate weak signals in a composite background. It effectively addresses the defect that the traditional algorithm cannot accurately estimate the weak signal under the background of strong interference and colored noise. As the number of interference sources increases, the direction finding performance of the proposed algorithm is slightly worse. The reason for the poor performance may be that the increase in the number of interferences gives rise to the leakage of the interference subspace, which affects the noise subspace.

4.4. Comparative Analysis of Errors under Different Noise Backgrounds. Experiment 4 sets the number of array elements at 10, the SIR is 30 dB, the number of snapshots is 200, and the number of Monte Carlo experiments is 500. The RMSE under different noise backgrounds is analyzed. Figure 4(a) shows the SNR of the FOC-ENS algorithm from −10 dB to 10 dB in step of 2 dB and the curves of the RMSE versus SNR under the two noise backgrounds. Figure 4(b) illuminates the FOC-FPB algorithm SNR changes from −10 dB to 10 dB in step of 2 dB, providing the curves of the RMSE with the SNR under different backgrounds.

Due to the unique blind Gaussian type of the FOC, the performance of this method under colored noise is similar to that under white noise.

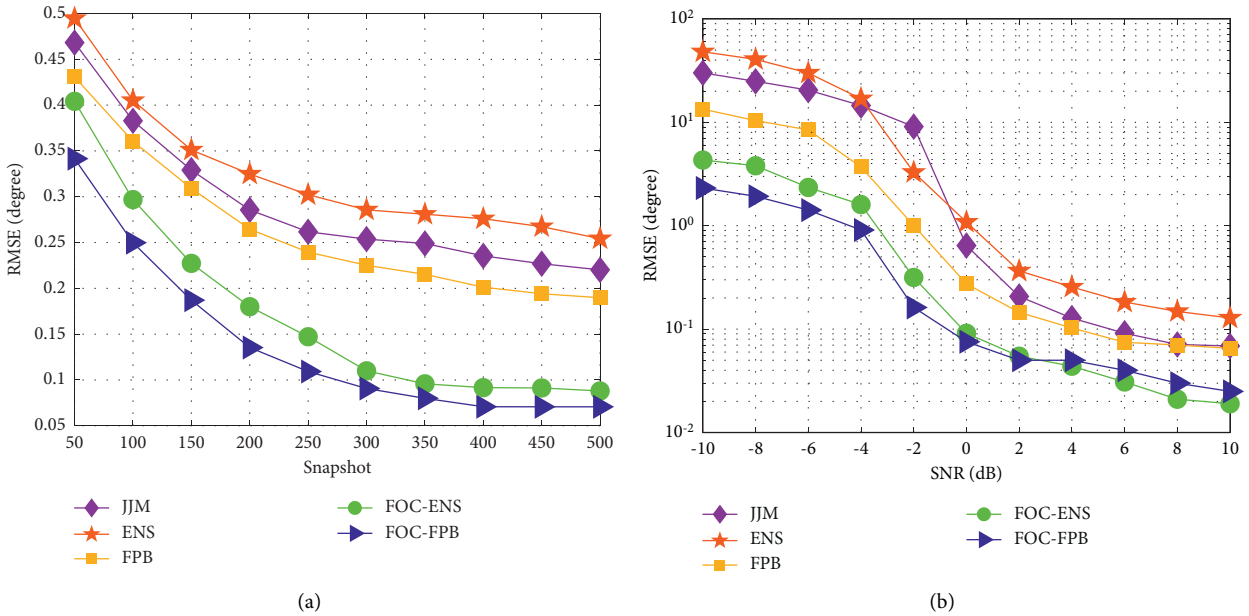


Figure 2: RMSE error analyses of different algorithms. (a) Relationship between RMSE and the number of snapshots. (b) Relationship between RMSE and SNR.

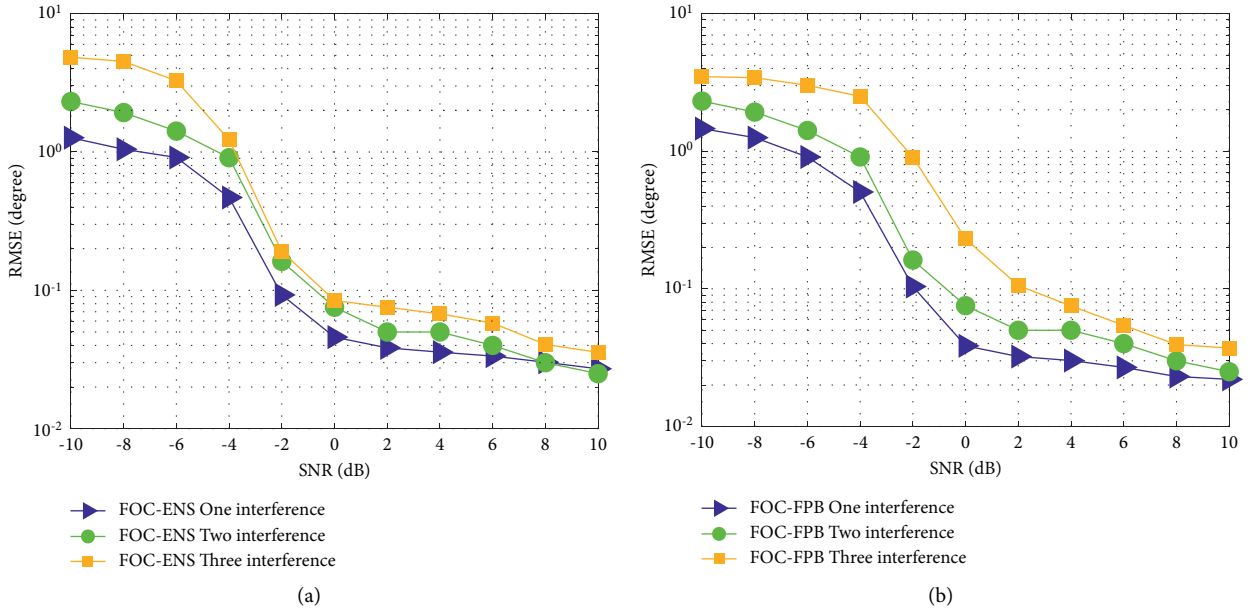


Figure 3: The statistical performance of the algorithm varying with SNR under different numbers of interference sources. (a) FOC-ENS algorithm. (b) FOC-FPB algorithm.

4.5. Algorithm Aperture Expansion Effect. Experiment 5 analyses the multitarget direction finding performance of the proposed algorithm. Figure 5(a) shows the simulation of the FOC-ENS algorithm. It sets the incidence angle of 12 target signals to be uniformly distributed from −50° to 60°, and the incidence angle of 2 strong interference signals to be 70° and 80°, and other settings are the same as in Section 4.1. Figure 5(b) is the FOC-FPB simulation. It sets the incident angle of 11 target signals to be uniformly distributed from −40° to 60°, and the incident angle of 2 strong interference signals to be −60° and −50°.

Owing to the aperture expansion characteristics of the FOC algorithm, the combined ENS algorithm and the FPB algorithm do not experience aperture loss, and the two improved algorithms perform well in multiobjective situations.

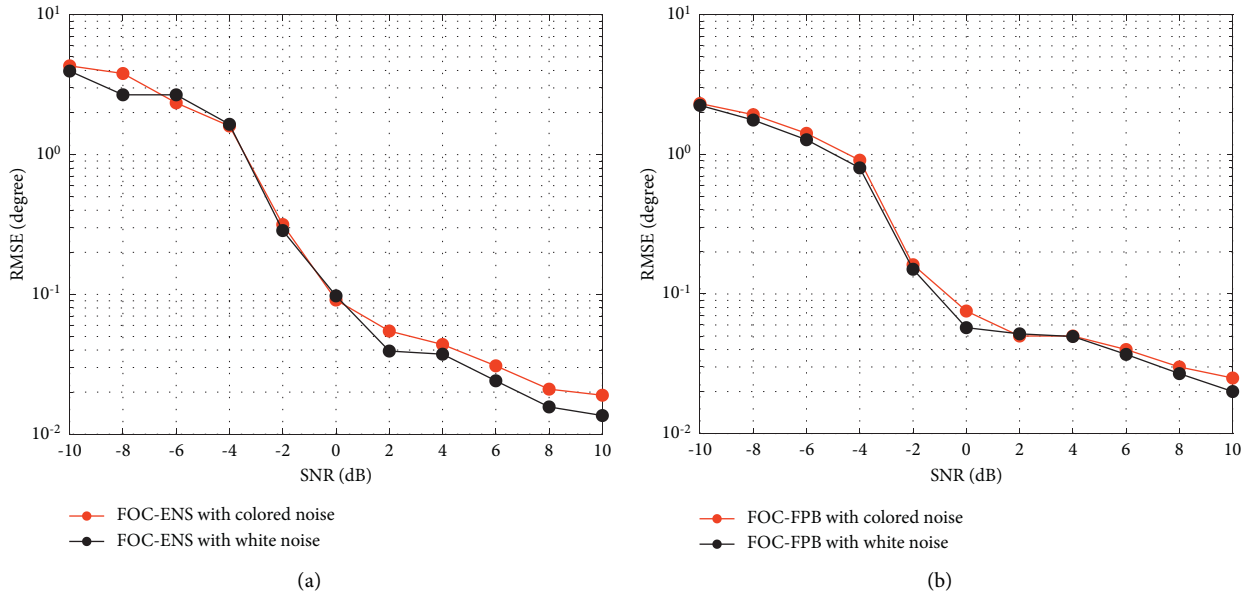


FIGURE 4: The statistical performance of different noise background algorithms varying with SNR. (a) FOC-ENS algorithm. (b) FOC-FPB algorithm.

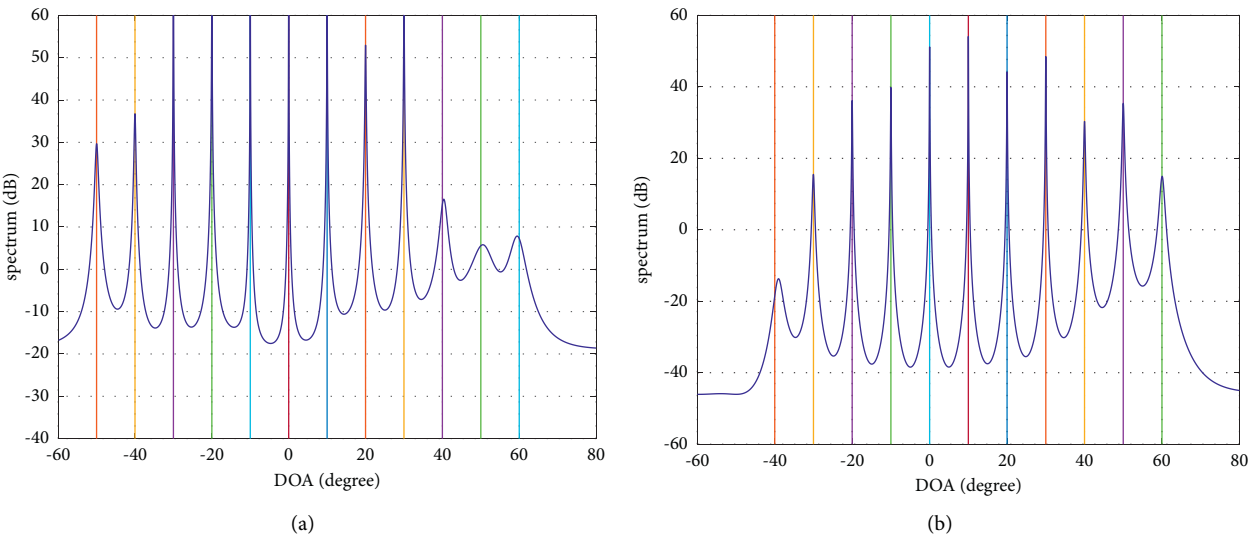


FIGURE 5: Multitarget direction [1] finding effect. (a) FOC-ENS algorithm. (b) FOC-FPB algorithm.

4.6. Time Complexity Analysis. Supposing that the number of sampling points is N, and the number of angle searches is N_θ. The calculation amount of FOC-ENS and FOC-FPB algorithms mentioned in this paper is mainly the construction of the cumulant matrix, eigenvalue decomposition, and spectral peak search. The calculation amount is $O(9NM^4 + M^6 + N_\theta M^4)$ and $O(9\mathrm{N}M^4 + 2M^6 + N_\theta M^4)$, respectively. Compared with the JJM and ENS algorithms with complexity $O(NM^4 + M^6 + N_\theta M^4)$ and the FPB algorithm with complexity $O(\mathrm{N}M^4 + 2M^6 + N_\theta M^4)$, the complexity is slightly higher.

Firstly, the calculation time of the two improved algorithms is simulated, and the number of target signals and the number of interference sources are set. K represents the number of target, and J represents the number of interference sources. Figures 6(a) and 6(b) show the simulation time comparison of FOC-ENS and FOC-FPB algorithms at different target numbers and interference sources, respectively. Then, $K = 8$ and $J = 2$ are set, and they are compared with traditional algorithms. Figure 6(c) shows the simulation time comparison of different algorithms.

It can be seen from Figure 6 that the FOC-ENS and FOC-FPB algorithms construct FOC matrix, so the calculation time is slightly longer than that of the JJM, ENS, and FPB algorithms. FOC-ENS algorithm simulation time is shorter than the FOC-FPB algorithm in that the FOC-FPB algorithm has one more feature decomposition than the FOC-ENS algorithm, which enhances the complexity to a certain extent.

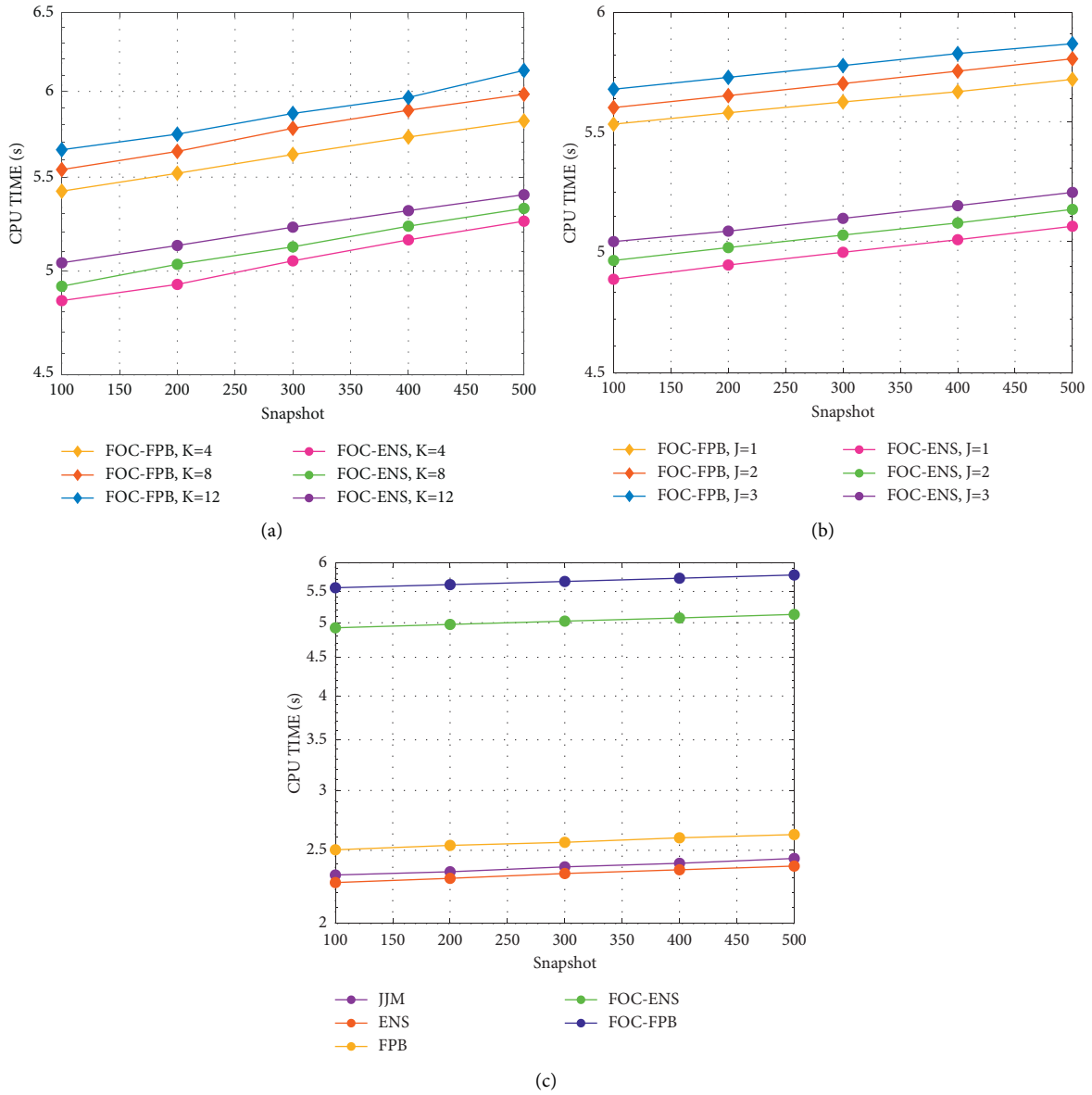


Figure 6: Comparison of simulation time. (a) Number of different targets. (b) Number of different interference sources. (c) Different algorithms.

5. Conclusion

To address the direction finding problem under the compound background of strong interference and colored noise, this paper proposes two algorithms based on FOC. Both algorithms need to construct a FOC matrix by receiving signal vectors to suppress noise components and then perform eigenvalue decomposition on the matrix and divide the subspace in accordance with the eigenvalues. The FOC-ENS algorithm merges the space where the strong interference steering vector is located into the noise subspace to construct an extended noise subspace for spectral peak search. The FOC-FPB algorithm constructs the orthogonal projection matrix of the strong interference signal subspace as the interference blocking matrix, uses the matrix to pre-transform the global subspace, and then eigen decomposes again to obtain the target DOA. Compared with the traditional algorithm, the algorithm proposed in this paper has a smaller RMSE, and there is no array aperture loss. It functions in a complex and volatile electromagnetic environment.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

This work was supported by the Natural Science Foundation of China (Grant no. 62071476).

References

[1] Z. Wang, R. Schaeger, M. Skoglund, M. Xiao, and H. Poor, "Strong secrecy for interference channels based on channel resolvability," *IEEE Transactions on Information Theory*, vol. 64, p. 1, 2018.

[2] Q. Liu, Y. Gu, and H. C. So, "DOA estimation in impulsive noise via low-rank matrix approximation and weakly convex optimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 6, pp. 3603–3616, 2019.

[3] W. Fu, Z. Hu, and D. Li, "A sorting algorithm for multiple frequency-hopping signals in complex electromagnetic environments," *Circuits, Systems, and Signal Processing*, vol. 39, no. 4, pp. 1–23, 2020.

[4] S. Salari, F. Chan, T. Chan, and R. Guay, "DOA estimation using compressive sampling-based sensors in the presence of interference," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 99, p. 1, 2020.

[5] Q. Liu, J. Xu, Z. Ding, and H. C. So, "Target localization with jammer removal using frequency diverse array," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 11685–11696, 2020.

[6] X. Wang, M. Amin, F. Ahmad, and E. Aboutanios, "Interference DOA estimation and suppression for GNSS receivers using fully augmentable arrays," *IET Radar, Sonar & Navigation*, vol. 11, no. 3, pp. 474–480, 2017.

[7] H. R. Park and J. Li, "Efficient sparse parameter estimation based methods for two-dimensional DOA estimation of coherent signals," *IET Signal Processing*, vol. 14, no. 7, 2020.

[8] F. Izedi, M. Karimi, and M. Derakhtian, "Joint DOA estimation and source number detection for arrays with arbitrary geometry," *Signal Processing*, vol. 140, pp. 149–160, 2017.

[9] Q. Fang, M. Jin, W. Liu, and Y. Han, "DOA estimation for sources with large power differences," *International Journal of Antennas and Propagation*, vol. 2021, Article ID 8862789, 12 pages, 2021.

[10] Y. Yang, Y. Zhang, and L. Yang, "Wideband sparse spatial spectrum estimation using matrix filter with nulling in a strong interference environment," *Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3891–3898, 2018.

[11] J. Zhang, G. Liao, and J. Zhang, "Weak signal DOA estimation method based on noise subspace expansion under strong signal background," *Systems Engineering and Electronics*, vol. 31, no. 06, pp. 1279–1283, 2009.

[12] H. Dong, T. Xu, and C. Wang, "Modified projection blocking method for two-dimensional weak signal DOA estimation under strong interference background," *Signal Processing*, vol. 29, no. 02, pp. 221–227, 2013.

[13] D. Xu, M. Li, W. Wang, and X. Wang, "Joint DOD and DOA angle estimation of coherent targets for bistatic MIMO radar," in *Proceedings of the Third International Conference on Instrumentation, Measurement, Computer, Communication and Control*, pp. 1025–1028, Washington, DC, USA, September, 2013.

[14] H. Li, W. Cui, B. Ba, H. Xu, and Y. Zhang, "Central DOA estimation method for exponential-type coherent distributed source based on fourth-order cumulant," *International Journal of Antennas and Propagation*, vol. 2020, no. 2, 11 pages, Article ID 8968136, 2020.

[15] M. Tufail and A. Ahmed, "Genetic algorithm-based improved DOA estimation using fourth-order cumulants," *International Journal of Electronics Theoretical & Experimental*, vol. 104, 2017.

[16] M. Pan, G. Zhang, Z. Hu, and Q. Zheng, "Real-valued off-grid DOA estimation based on fourth-order cumulants using sparse Bayesian learning in spatial coloured noise," *IET Communications*, vol. 13, no. 19, pp. 3344–3349, 2019.

[17] H. Shi, N. Ma, and Z. Guan, "A fourth-order cumulants orthonormal propagator rooting method based on toeplitz approximation," *EURASIP Journal on Wireless Communications and Networking*, vol. 13, no. 4, pp. 409–414, 2020.

[18] Q. Liu, H. C. So, and Y. Gu, "Off-grid DOA estimation with nonconvex regularization via joint sparse representation," *Signal Processing*, vol. 140, pp. 171–176, 2017.

[19] G. Zheng, Y. Song, and C. Chen, "Height measurement with meter wave polarimetric MIMO radar: signal model and MUSIC-like algorithm," *Signal processing*, vol. 190, no. 6, pp. 1872–7557, 2021.

[20] L. Qiu, T. Lan, and Y. Wang, "A sparse perspective for direction-of-arrival estimation under strong near-field interference environment," *Sensors*, vol. 20, no. 1, p. 163, 2019.

[21] Y. Gao, X. U. Jia, and T. Long, "Performance analysis of anti-jamming method via block matrix," *Journal of Signal Processing*, vol. 31, 2015.

[22] J. Gong, S. Lou, and Y. Guo, "DOA estimation method of weak sources for an array antenna under strong interference conditions," *International Journal of Electronics*, vol. 105, no. 9, pp. 1522–1531, 2018.

[23] H. Chen and H. Su, "A new method of DOA estimation in the background of strong interference signal," *Chinese Journal of Electronics*, no. 03, pp. 530–534, 2006.

[24] H. Chen, C. Hou, Q. Wang, and L. Huang, "Cumulants-based toeplitz matrices reconstruction method for 2-D coherent DOA estimation," *IEEE Sensors Journal*, vol. 14, no. 8, pp. 2823–2832, 2014.

[25] Y. Wang, M. Trinkle, and B. W. H. Ng, "Two-stage DOA estimation of independent and coherent signals in spatially coloured noise," *Signal Processing*, vol. 128, no. Nov, pp. 350–359, 2016.

[26] H. Yang, L. Yi, and W. Xi, "DOA estimation of coherent signals on coprime arrays exploiting fourth-order cumulants," *Sensors*, vol. 17, no. 4, 2016.

[27] Y. Wang, L. Wang, and X. Yang, "Efficient cumulant-based methods for joint angle and frequency estimation using spatial-temporal smoothing," *Electronics*, vol. 8, no. 1, 2019.