Recognition of acoustic vortex fields based on a convolutional attention neural network

Haicai Xiao,[1,†] Xinwen Fan,[2,†] Yang Kang,[1] Xiaolong Huang,[1] Can Li,[1] Ning Li,[1] Chunsheng Weng,[1] and Xudong Fan[1,*]

[1] *National Key Laboratory of Transient Physics, Nanjing University of Science and Technology, Nanjing 210094, China*

[2] *College of Electronic Science and Engineering, Jilin University, Changchun 130012, China*

(Received 11 April 2024; revised 20 June 2024; accepted 28 June 2024; published 19 July 2024)

In this work, we propose a deep convolutional attention neural network to realize the recognition of acoustic vortex fields. Convolutional attention mechanisms are introduced into the network together with the residual idea. The deep neural network is trained and then evaluated, and the performance is compared with those of several typical convolutional neural networks, including AlexNet, GoogLeNet, and ResNet. The results show that the improved neural network model based on the convolutional attention mechanism proposed here has better classification performance and stronger stability. The classification accuracy of the improved model on the whole test set reaches more than 95%, indicating that the model has a stable classification ability. Our work helps further study the detection and recognition of acoustic vortex fields, which will find many applications in scientific research and industry.

DOI: 10.1103/PhysRevApplied.22.014051

I. INTRODUCTION

Acoustic waves are the only fluctuation form to propagate over long distances in the ocean. With the increase of human underwater activities, underwater acoustic communication is crucial for underwater applications such as deep-sea scientific exploration, offshore industrial control, and long-range monitoring of the marine environment. Compared to acoustic waves, electromagnetic waves are more strongly absorbed in underwater environments, which limits the potential applications for such waves. Therefore, using acoustic waves to transmit information is currently the dominant technology for underwater applications.

The recent emergence of underwater communications based on acoustic vortex waves provides an alternative method for data transmission. Acoustic vortex waves possess a spiral phase distribution, characterized by an integer number, named topological charge. The generation of acoustic vortex waves has been investigated for a long time and can be realized either by active methods, which utilize multiple transducers to control the initial phase [1–5], or by passive acoustic metamaterials, which adjust the phase and amplitude distribution from artificial structures [6–17].

However, the detection and recognition of acoustic vortex waves has only recently started. In 2014, Gao *et al.* [18] investigated the linear phase distributions of phase-encoded acoustic vortices, and designed an eight-acoustic-source sensor system to verify the feasibility of precise phase control for acoustic vortices. In 2017, Shi *et al.* [19] designed a system consisting of 64 acoustic sources, which realized the phase encoding of acoustic vortices from -4 to $+4$, and can detect and separate multiplexed orbital angular momenta (OAM) based on the inner-product algorithm. Recently, passive acoustic metamaterials [20,21] and a circular receiver array [22] were used to detect and separate different acoustic OAM states. In 2020, Tong *et al.* [23] used an acoustic metaskin insulator as a waveguide to effectively transmit data based on OAM multiplexing. Although great success has been achieved in OAM detection, the recognition of acoustic vortex waves remains to be explored.

In recent years, the emergence and rapid development of machine learning and deep learning methods have provided a new idea to solve this problem. For example, Stankevich [24] proposed an approach to solve multiplexed OAM sound beams using a two-layer convolutional neural network in 2019. In 2023, Cao *et al.* [25] reported a strategy of phase-dislocation-mediated high-dimensional fractional acoustic vortex communication based on machine learning, which demonstrates the potential to infinitely increase the channel capacity in acoustic-vortex communications. Although OAM detection based on the machine learning method has exhibited fascinating results, it is still in its infancy. Further

*Contact author: fanxudong@njust.edu.cn

†H.X. and X.F. contributed equally to this work.

in-depth exploration is still highly required for the study of machine-learning-based acoustic-vortex-wave recognition and decoding.

In this paper, a deep learning approach is introduced into conventional acoustics for the recognition of acoustic vortex waves. A deep learning network model based on the convolutional attention mechanism is constructed together with the residual idea inspired by ResNet [26]. Although only partial information of the acoustic vortex field is used during the identification, the detection and demodulation of acoustic vortex waves with topological charges from -8 to $+8$ are successfully realized via the deep learning model proposed here. The breakthrough of the convolutional attention module here greatly improves the classification accuracy, stability, and robustness of the neural network. The deep-learning-based acoustic-vortex-wave recognition method proposed in this paper paves a new way for the detection and demodulation of vortex acoustic waves, which has potential applications in both scientific research and engineering fields.

II. EXPERIMENTAL METHOD

A. Deep learning model

The specific structure of the whole deep learning network is shown schematically in Fig. 1(a). The network is stacked by a series of basic building blocks, each of which consists of several operation layers, including convolutional (conv) layers, pooling layers, batch normalization (BN) [27], linear rectification functions (ReLUs) [28], and convolutional attention modules (CAMs). To avoid gradient disappearance and accuracy saturation [29,30], two types of residual modules are used within the neural network, i.e., conv block and identity block, as shown in Figs. 1(b) and 1(c), respectively. The input and output dimensions of the conv block are not necessarily the same, so it cannot be connected in series continuously. The role of the conv block is to change the representation dimension of input features. On the other hand, the identity block has the same input-output dimension and can be connected in series to deepen the network. In this way, the entire residual block can output high-dimensional features and provide a better representation based on the input features.

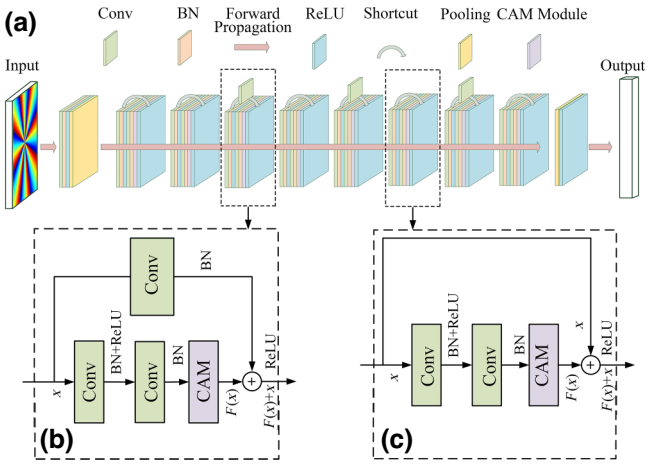

FIG. 1. Illustration of acoustic-vortex-wave recognition based on deep learning: (a) structure of the deep learning model; (b) the convolutional (conv) block; and (c) the identity block.

The BN is used to normalize the output features from the convolutional layers to permit the selection of a larger initial learning rate and efficient tuning of hyperparameters. The formulas for BN are given below [27]:

$$\begin{aligned}\mu_B &= \frac{1}{m}\sum_{i=1}^{m} x_i,\\ \sigma_B^2 &= \frac{1}{m}\sum_{i=1}^{m}(x_i-\mu_B)^2,\\ \widetilde{x_i} &= \frac{x_i-\mu_B}{\sqrt{\sigma_B^2+\epsilon}},\\ y_i &= \gamma\widetilde{x_i}+\beta = \mathrm{BN}_{\gamma,\beta}(x_i),\end{aligned} \tag{1}$$

Here μ_B and σ_B are the expectation and standard deviation of a batch of input data, respectively; and γ and β are the parameters to be learned, involving the propagation of the entire neural network.

Besides, the rectified linear unit (ReLU) [28], a commonly used activation function in deep learning, is defined as follows:

$$f(x) = \max(0, x), \tag{2}$$

where x is the input value. The output of the ReLU function is equal to the input if the input is positive; otherwise, the output is zero. Compared with some complex activation functions, ReLU is very simple to compute, which helps improve the training and inference speed of the model.

The detailed parameters of the model are as follows. The first layer is the input layer, where the input image size is $224 \times 224 \times 3$. A convolutional layer is then used with the kernel size of 7×7, a stride of 2, and a padding of 3. The number of output channels is 64. Batch normalization and ReLU activation function are performed, followed by a maximum pooling layer with the kernel size of 3×3, a stride of 2, and a padding of 1. Eight residual blocks consisting of a channel attention module and a spatial attention module follow. Each residual block consists of two convolutional layers with the kernel size of 3×3, a stride of 1, and a padding of 1. The extracted features go through a global average pooling layer and are finally mapped to the 17 labels of the classification through a fully connected layer.

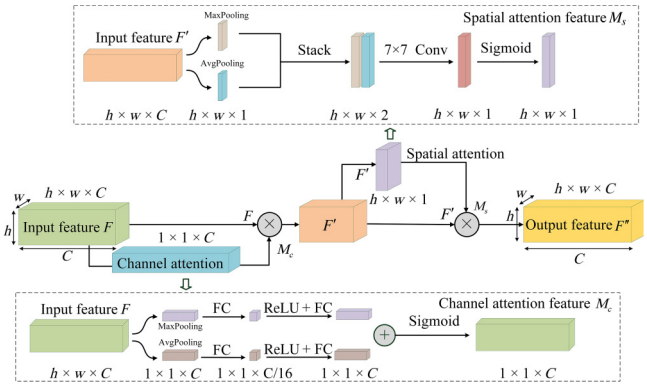

FIG. 2. Illustration of the convolutional attention module.

B. Convolutional attention mechanism module

Specifically, the convolutional attention modules are introduced and inserted after the second convolutional layer within the residual blocks [see Figs. 1(b) and 1(c)]. The attention mechanism in deep learning is a method of modeling the human visual and cognitive system that allows neural networks to focus on relevant parts of the input data as they are processed. The benefit of introducing the attention mechanism is that the neural network can automatically learn and selectively focus on the important information in the input data, which could then improve the performance and generalization of our model.

In this work, the convolutional attention module is composed of the channel attention mechanism and the spatial attention mechanism, as shown in Fig. 2. The traditional attention mechanism based on convolutional neural networks mainly focuses on the analysis of the channel domain, which considers only the interaction between the channels of the feature map. However, CAM focuses on both the channel domain and the space domain by introducing two analysis dimensions of spatial and channel attention. Specifically, spatial attention makes the neural network pay more attention to the pixel regions in the image that play a decisive role in classification while ignoring irrelevant regions, while channel attention is used to deal with the allocation relationship of the feature map channels. CAM allocates attention to both dimensions at the same time, which enhances the effect of the attention mechanism on the performance of the model.

The overall flowchart of CAM is shown in Fig. 2. The input feature map F first passes through the channel attention mechanism. The output feature M_c is multiplied by the input feature maps F and then sent to the spatial attention mechanism, where the normalized spatial weights are multiplied by the input feature maps of the spatial attention mechanism to obtain the final weighted feature maps. The main formulas are as follows:

$$\begin{aligned} F' &= M_c(F) \otimes F, \\ F'' &= M_s(F') \otimes F'. \end{aligned} \tag{3}$$

Here $F \in R^{C\times H\times W}$ is the input feature, where $\otimes$ denotes element-by-element multiplication, C is the channel weight, H and W are the height and width of the input feature, respectively; and M_c and M_s represent the operations of attention extraction on the channel dimension and the spatial dimension, respectively. The expression for M_c is as follows:

$$\begin{aligned} M_c &= \sigma[\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MLP}(\mathrm{MaxPool}(F))] \\ &= \sigma\left[w_1\left(w_0(F^c_{\mathrm{avg}})\right) + w_1\left(w_0(F^c_{\mathrm{max}})\right)\right] \end{aligned} \tag{4}$$

where σ represents the sigmoid function, and the weights w_0 and w_1 of the multilayer perceptron (MLP) are shared. The main process of the channel attention module is to compress the spatial dimension of the input feature mapping first to get a one-dimensional vector to operate later. The spatial dimension compression of the input features is performed using average pooling and maximum pooling, respectively, to obtain a total of two one-dimensional vectors, which has the advantage that the average pooling efficiently learns the target object, while the maximum pooling gathers another important cue about the features of the unique object to infer the attention in terms of the channel. After obtaining the two one-dimensional vectors, they are placed into a shared network, which is composed of a hidden layer and the MLP, and after applying the shared network to the vectors, the output feature vectors are merged using element-by-element summation.

In addition, M_s is expressed as

$$\begin{aligned} M_s &= \sigma(f^{7\times7}[\mathrm{AvgPool}(F); \mathrm{MaxPool}(F)]) \\ &= \sigma(f^{7\times7}[F^s_{\mathrm{avg}}; F^s_{\mathrm{max}}]), \end{aligned} \tag{5}$$

where f represents the convolution operation with the kernel size of 7×7. The main process of the spatial attention module is to first apply the average pooling and maximum pooling operations along the channel axis, which compresses the input feature on the channel level, and then the average pooling and maximum pooling operations are done for the input features on the channel dimension. Finally, two one-dimensional features are obtained, which are spliced together by channel dimension to obtain a feature with channel number 2, after which a hidden layer containing a single convolutional kernel is used to perform convolutional operations on it to ensure the final feature obtained is consistent with the input feature in terms of spatial dimension.

TABLE I. Experimental parameters.

Experimental parameters	Values
Image size	$224 \times 224 \times 3$
Epoch number	100
Batch size	32, 64, 128
Learning rate	0.00003
Weight decay	0.00005

The CAM module here can help enhance the feature interaction between different channels and improve the feature representation. By introducing the CAM module, the model can automatically learn the important information in the input, and then target attention to specific features, including both global and local information that contribute to the task. The CAM module can improve the performance of the network with a negligible increase in the computational cost.

C. Data preparation and training parameters

Acoustic vortex fields with topological charges from -8 to $+8$ for training and testing are artificially synthesized. Acoustic vortex phase fields are randomly cropped with the resolution of 224×224. A total of 1000 random samples are generated for each topological charge (i.e., OAM mode), giving a total of 17 000. The obtained acoustic field images are randomly disrupted and divided into the training set, validation set, and test set in the ratio of 8:1:1. The training set is used to train the deep neural network model; the validation set is used to tune the network hyperparameters such as learning rate, weight decay, and batch size [31]; and the test set is used to evaluate the model performance.

The dataset is normalized for each channel of the image, which helps the training of the model and improves the robustness of the model regarding the input data. The loss function adopts CrossEntropyLoss, which is a function used to measure the gap between the model prediction and the ground truth. The Adam optimizer is used to adjust network parameters according to the current parameters and gradient during the training process so that the loss function can approach its minimum value. The details of experimental parameters are shown in Table I.

The experimental platform is based on the Pytorch framework using the Python language with the details as follows: Windows 10 operating system; Intel Core i5-13600K processor; 16 GB memory; NVIDIA GeForce RTX 4060Ti graphics card; CUDA 11.7 platform; Python version 3.9; and Pytorch version 1.13.1.

III. RESULTS

We examine the loss function and the accuracy of the model as a function of the epochs for three datasets

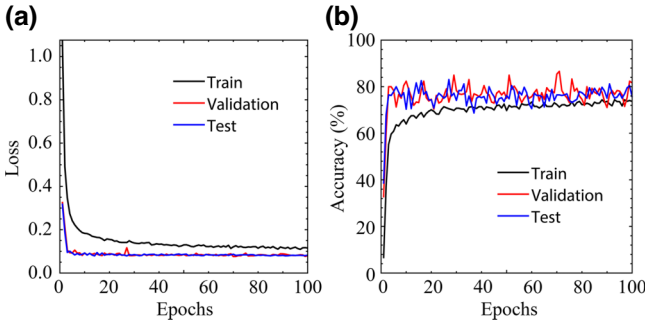

FIG. 3. Variation of (a) loss function and (b) accuracy during the training process for three datasets: black, training set; red, validation set; and blue, testing set.

(Fig. 3). The batch size is 64 and the convolutional attention neural network (CANN) is trained for 100 epochs. From the results, all three loss curves in Fig. 3(a) decrease with the epoch number, indicating that the model is learning and the fitting ability is improving. The loss functions all converge to below 0.2 during the training process, although the loss function is slightly higher on the training set and slightly lower on the validation and test sets. All three losses decrease to around 0.08 at the 100th round, which indicates that the model has a good fit at the end of the training, and has a better generalizability on the validation and test sets.

The three curves in Fig. 3(b) increase with the number of training rounds, showing that the classification performance and generalizability of the model are improving once again. In both Figs. 3(a) and 3(b), the test set and validation set work better than the training set since data augmentation is used during the training process to increase the diversity and difficulty of the training set to improve the generalizability of the model. In that case, the accuracy on the training set may be lower than the accuracy on the test set because the data on the test set is not data-enhanced and is relatively easier to recognize by the model. Another reason is that we use regularization methods (weight decay and batch normalization) during the training process to prevent model overfitting and thus improve the generalizability of the model.

When the network reaches the optimal parameters, we test the classification accuracy on several randomly selected vortex fields, as shown in Fig. 4. From the results, the deep learning model successfully recognizes the vortex fields with relatively high confidence larger than 90%. We further examine the classification accuracy for all topological charges from -8 to $+8$, as shown in Fig. 5(a). The classification accuracy of all labels is about 95%, except for labels 2 and -2, whose accuracy is about 90%. All other labels have a classification accuracy of more than 94%, there is no significant difference in the classification accuracy between different labels, and label 0 has the best classification accuracy of 100%.

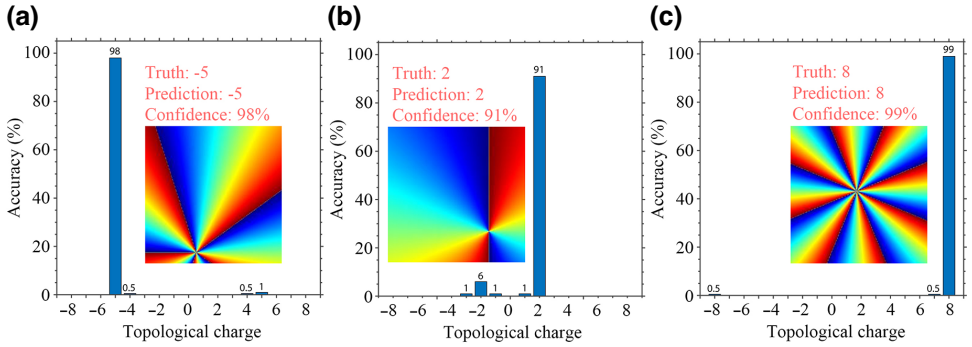

FIG. 4. Prediction on several typical vortex fields based on the convolutional attention neural network.

The confusion matrix for the test set is shown in Fig. 5(b). We can find that each categorical label is most likely to be confused with its opposite number of labels, i.e., topological charges with opposite positive and negative signs are difficult to distinguish. This confusion is most serious for topological charges of 2 and −2, which have 91% and 89% correct classification rates, respectively. In addition, we have also visualized the extracted feature from the model with the attention mechanism [see Fig. 5(c)]. It can be seen that the attention mechanism can focus more clearly on the key areas of the image, thus better capturing important

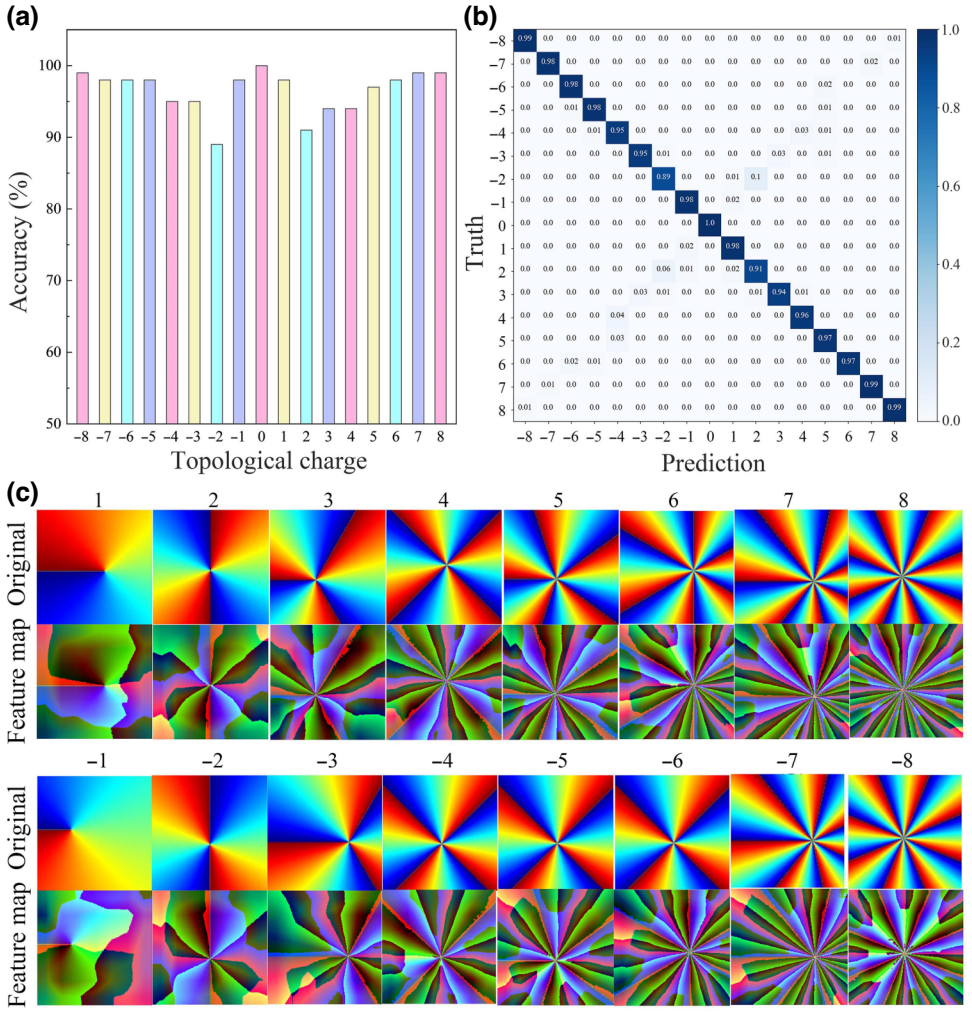

FIG. 5. (a) Classification accuracy for all topological charges. (b) Confusion matrix. (c) Visualization of feature maps extracted from the model.

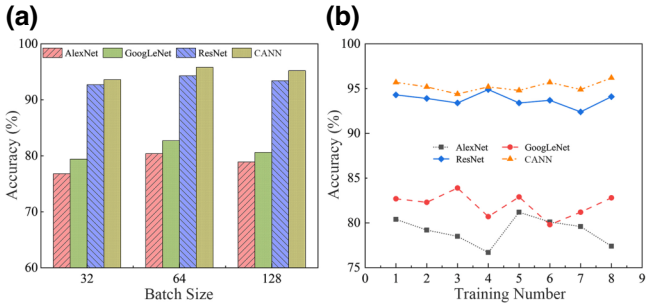

FIG. 6. Comparison of the performance for different networks.

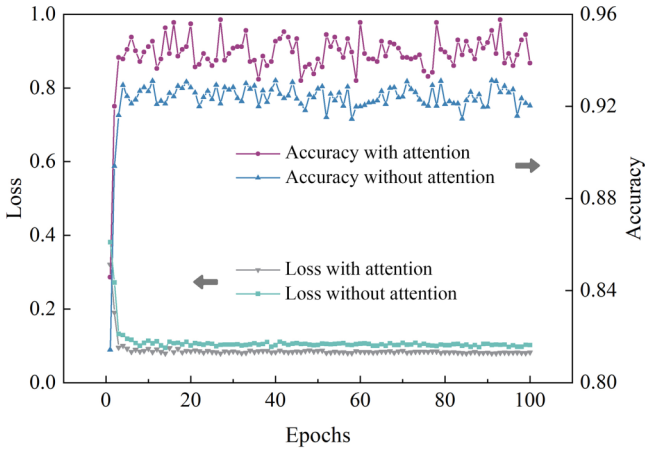

FIG. 7. Comparison of the models without the attention mechanism.

structural information and details of acoustic vortex fields.

We further compare the performance of our model with that of other commonly used networks for image classification, including AlexNet [32], GoogLeNet [33], and ResNet18 [26]. Three different batch sizes of 32, 64, and 128 are also considered to test the performance of the four models. Other setups have all the same experimental parameters as shown in Table I. Figure 6 shows the comparison results of the four network models during the dataset training process. Each model takes the optimal weights during the 100 training epochs. From Fig. 6(a), it can be seen that the classification accuracy of each model achieves good results under different batch sizes, but when the batch size is 64, the classification accuracy of the four models is 1%–5% higher than for other batch sizes. In addition, when the batch size is kept the same, the classification accuracy of CANN proposed in this work is significantly better than that of the other three models.

The four models above are run eight times with a batch size of 64 and a maximum epoch number of 100. Best weights are saved for each run, which in turn leads to the accuracy of eight different sets as shown in Fig. 6(b). From the results, it can be seen that the deep learning model proposed here is more stable than the other three models. A 1%–2% fluctuation indicates the stability and the robustness of the network.

We have further compared the results of two deep learning models without the attention mechanism (see Fig. 7). The results show that the model with the attention mechanism significantly improves the classification accuracy and stability. We compared the accuracy and loss function of the two models trained for 100 epochs. The loss function of the model with the attention mechanism is finally stable at around 0.08 during the training process, which is 0.02 lower than that of the model without the attention mechanism, which means the model with the attention mechanism converges faster. Moreover, the accuracy of the model with the attention mechanism during the training process is consistently higher than that of the model without the attention mechanism, which again proves the advantages and improvements of using the attention mechanism in this work.

IV. DISCUSSION

In this study, we proposed a deep learning neural network based on convolutional attention mechanisms and residue ideas, aiming to realize the recognition of nonaxial acoustic vortex fields. The experimental results show that the model achieves a total classification accuracy of more than 95% on the whole dataset, and the model has no obvious overfitting or degradation phenomenon during the training process, which exhibits good recognition performance and fitting ability. In addition, the classification accuracy for different topological charges was also compared, exhibiting strong classification ability and stability.

The deep learning approach proposed here provides an alternative method for the recognition and detection of acoustic vortex waves and also helps acoustic communications based on orbital angular momenta. The results here could be potentially used in other vortex fields, such as electromagnetic waves with the aid of the domain adaption approach. Although these strategies such as normalization and data augmentation have been used in the current model, it may not achieve high accuracy if directly applied to experimental results. Therefore, to achieve better performance, experimental data containing both intensity and phase information could be included to further train and fine-tune the model. In addition, the typical characteristics of acoustic vortex fields include helical phase wavefronts and donut-shaped sound intensity distributions.

In experiments, noise, turbulence, and disturbance from many different sources will impact the quality of the received signals. In addition, measured data with different signal-to-noise ratios should be considered to estimate the effect caused by noise. Hence, intensity field

information can be included in the future together with the phase information to further improve the performance and the robustness of the model. This combined approach is expected to improve the accuracy and reliability of vortex field recognition. Furthermore, in OAM multiplexed communication, hybrid vortex beams with different topological modes can significantly enhance communication capacity. However, this study has not yet explored this aspect. We recognize that our current model may not achieve high recognition accuracy when directly applied to the classification of multiplexed OAM modes. To improve performance, a dataset containing multiplexed OAM modes could be used to further train and fine-tune the model. Additionally, employing a more complex network structure and advanced training strategies will be beneficial in achieving this goal.

ACKNOWLEDGMENTS

The authors acknowledge the support of the National Natural Science Foundation of China (Grant No. 12204241), the Natural Science Foundation of Jiangsu Province (Grant No. BK20220924), and the Fundamental Research Funds for the Central Universities (Grants No. 30923011019 and No. 020414380195).

[1] B. T. Hefner and P. L. Marston, An acoustical helicoidal wave transducer with applications for the alignment of ultrasonic and underwater systems, J. Acoust. Soc. Am. **106**, 3313 (1999).

[2] Régis Marchiano and Jean-Louis Thomas, Synthesis and analysis of linear and nonlinear acoustical vortices, Phys. Rev. E **71**, 066616 (2005).

[3] Timothy M. Marston and Philip L. Marston, Acoustic beams with modulated helicity and a simple helicity-selective acoustic receiver, J. Acoust. Soc. Am. **126**, 2187 (2009).

[4] Timothy M. Marston and Philip L. Marston, Modulated helicity for acoustic communications and helicity-selective acoustic receivers, J. Acoust. Soc. Am. **127**, 1856 (2010).

[5] Ling Yang, Qingyu Ma, Juan Tu, and Dong Zhang, Phase-coded approach for controllable generation of acoustical vortices, J. Appl. Phys. **113**, 154904 (2013).

[6] Régis Wunenburger, Juan Israel, Vazquez Lozano, and Etienne Brasselet, Acoustic orbital angular momentum transfer to matter by chiral scattering, New J. Phys. **17**, 103022 (2015).

[7] Stefan Gspan, Alex Meyer, Stefan Bernet, and Monika Ritsch-Marte, Optoacoustic generation of a helicoidal ultrasonic beam, J. Acoust. Soc. Am. **115**, 1142 (2004).

[8] Xue Jiang, Jiajun Zhao, Shi-lei Liu, Bin Liang, Xin-ye Zou, Jing Yang, Cheng-Wei Qiu, and Jian-chun Cheng, Broadband and stable acoustic vortex emitter with multi-arm coiling slits, Appl. Phys. Lett. **108**, 203501 (2016).

[9] Tian Wang, Manzhu Ke, Weiping Li, Qian Yang, Chunyin Qiu, and Zhengyou Liu, Particle manipulation with acoustic vortex beam induced by a brass plate with spiral shape structure, Appl. Phys. Lett. **109**, 123506 (2016).

[10] Noé Jiménez, Rubén Picó, Vıctor Sánchez-Morcillo, Vicent Romero-García, Lluís M. García-Raffi, and Kestutis Staliunas, Formation of high-order acoustic Bessel beams by spiral diffraction gratings, Phys. Rev. E **94**, 053004 (2016).

[11] Noé Jiménez, Vicent Romero-García, Luis M. García-Raffi, Francisco Camarena, and Kestutis Staliunas, Sharp acoustic vortex focusing by Fresnel-spiral zone plates, Appl. Phys. Lett. **112**, 204101 (2018).

[12] Hongping Zhou, Jingjing Li, Kai Guo, and Zhongyi Guo, Generation of acoustic vortex beams with designed Fermat's spiral diffraction grating, J. Acoust. Soc. Am. **146**, 4237 (2019).

[13] Xue Jiang, Yong Li, Bin Liang, Jian-chun Cheng, and Likun Zhang, Convert Acoustic Resonances to Orbital Angular Momentum, Phys. Rev. Lett. **117**, 034301 (2016).

[14] Chen Liu, Houyou Long, Chengrong Ma, Yurou Jia, Chen Shao, Ying Cheng, and Xiaojun Liu, Broadband acoustic vortex beam generator based on coupled resonances, Appl. Phys. Lett. **118**, 143503 (2021).

[15] Tuo Liu, Shuowei An, Zhongming Gu, Shanjun Liang, He Gao, Guancong Ma, and Jie Zhu, Chirality-switchable acoustic vortex emission via non-Hermitian selective excitation at an exceptional point, Sci. Bull. **67**, 1131 (2022).

[16] Xudong Fan, Yifan Zhu, Zihao Su, Ning Li, Xiaolong Huang, Yang Kang, Can Li, Chunsheng Weng, Hui Zhang, Weiwei Kan *et al.*, Transverse particle trapping using finite Bessel beams based on acoustic metamaterials, Phys. Rev. Appl. **19**, 034032 (2023).

[17] Xudong Fan, Yifan Zhu, Zihao Su, Ning Li, Xiaolong Huang, Yang Kang, Can Li, Chunsheng Weng, Hui Zhang, Bin Liang *et al.*, Ultrabroadband and reconfigurable transmissive acoustic metascreen, Adv. Funct. Mater. **33**, 2300752 (2023).

[18] Lu Gao, Haixiang Zheng, Qingyu Ma, Juan Tu, and Dong Zhang, Linear phase distribution of acoustical vortices, J. Appl. Phys. **116**, 024905 (2014).

[19] Chengzhi Shi, Marc Dubois, Yuan Wang, and Xiang Zhang, High-speed acoustic communication by multiplexing orbital angular momentum, Proc. Nat. Acad. Sci. U.S.A **114**, 7250 (2017).

[20] Xue Jiang, Bin Liang, Jian-Chun Cheng, and Cheng-Wei Qiu, Twisted acoustics: Metasurface-enabled multiplexing and demultiplexing, Adv. Mater. **30**, 1800257 (2018).

[21] Xue Jiang, Chengzhi Shi, Yuan Wang, Joseph Smalley, Jianchun Cheng, and Xiang Zhang, Nonresonant metasurface for fast decoding in acoustic communications, Phys. Rev. Appl. **13**, 014014 (2020).

[22] Han Zhang and Jun Yang, Transmission of video image in underwater acoustic communication, arXiv:1902.10196.

[23] Lei Tong, Zhu Xiong, Ya-Xi Shen, Yu-Gui Peng, Xin-Yu Huang, Lei Ye, Ming Tang, Fei-Yan Cai, Hai-Rong Zheng, Jian-Bin Xu *et al.*, An acoustic meta-skin insulator, Adv. Mater. **32**, 2002251 (2020).

[24] D Stankevich, Orbital angular momentum acoustic modes demultiplexing by machine learning methods, Inform Tech Nanotech **2416**, 300 (2019).

[25] Ruijie Cao, Gepu Guo, Wei Yue, Yang Huang, Xinpeng Li, Chengzhi Kai, Yuzhi Li, Juan Tu, Dong Zhang, Peng

Xi *et al.*, Phase-dislocation-mediated high-dimensional fractional acoustic-vortex communication, Research **6**, 0280 (2023).

[26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, Deep residual learning for image recognition, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. **6**, 770 (2016).

[27] Sergey Ioffe and Christian Szegedy, in *International Conference on Machine Learning* (PMLR, 2015), Vol. 37, p. 448.

[28] Abien Fred Agarap, Deep learning using rectified linear units (ReLU), arXiv:1803.08375.

[29] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng, in *Proceedings of the 26th Annual International Conference on Machine Learning* (2009), Vol. 1, p. 609.

[30] Senem Tanberk, Volkan Dağlı, and Mustafa Kağan Gürkan, in *2021 6th International Conference on Computer Science and Engineering (UBMK)* (IEEE, 2021), p. 506.

[31] Zhao Chen, Yin Jiang, Xiaoyu Zhang, Rui Zheng, Ruijin Qiu, Yang Sun, Chen Zhao, and Hongcai Shang, ResNet18DNN: Prediction approach of drug-induced liver injury by deep neural network with ResNet18, Brief. Bioinformatics **23**, bbab503 (2022).

[32] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, ImageNet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. **25**, 1097 (2012).

[33] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), Vol. 4, p. 1.