APPLIED GEOPHYSICS, Vol.21, No.1 (March 2024), P. 1–13, 9 Figures.
DOI:10.1007/s11770-024-1058-y

Microseismic Event Recognition and Transfer Learning Based on Convolutional Neural Network and Attention Mechanisms

Jin Shu[1,2], Zhang Shichao[3,4], Gao Ya[1,2], Yu Benli[1,2], Zhen Shenglai[1,2*]

Abstract: Microseismic monitoring technology is widely used in tunnel and coal mine safety production. For signals generated by ultra-weak microseismic events, traditional sensors encounter limitations in terms of detection sensitivity. Given the complex engineering environment, automatic multi-classification of microseismic data is highly required. In this study, we use acceleration sensors to collect signals and combine the improved Visual Geometry Group with a convolutional block attention module to obtain a new network structure, termed CNN_BAM, for automatic classification and identification of microseismic events. We use the dataset collected from the Hanjiang-to-Weihe River Diversion Project to train and validate the network model. Results show that the CNN_BAM model exhibits good feature extraction ability, achieving a recognition accuracy of 99.29%, surpassing all its counterparts. The stability and accuracy of the classification algorithm improve remarkably. In addition, through fine-tuning and migration to the Pan II Mine Project, the network demonstrates reliable generalization performance. This outcome reflects its adaptability across different projects and promising application prospects.
Keywords: Microseismic; Convolutional Neural Networks; Multi-classification; Attentional mechanism; Transfer learning

Introduction

In today's exploration of the Earth's depths, humans increasingly realize the vast resources and space in the subsurface [1,2]. Microseismic analysis is rapidly developing as a crucial technology in disaster monitoring and early warning; it is widely used in the capitalization of subsurface space [3,4,5]. With advancements in sensing and computer technologies, researchers can collect more signals from highly sensitive sensors and then manually classify and label the data to localize events using mathematical algorithms [6,7,8,9]. Microseismic data processing is divided into three main steps: waveform identification, microseismic arrival detection, and event location.

Precise categorization of microseismic data can filter out valid events requiring localization. Microseismic

Manuscript received by the Editor August 13, 2023; revised manuscript received November 21, 2023
1.Information Materials and Intelligent Sensing Laboratory of Anhui Province, Anhui University, Hefei, 230601, China
2.Key Laboratory of Opto-Electronic Information Acquisition and Manipulation, Ministry of Education, Anhui University, Hefei, 230601, China
3.School of Public Safety and Emergency Management, Anhui University of Science and Technology, Huainan, 232000, China
4.IDETECK CO., LTD., Chuangxin Avenue, Hefei 230601, Anhui, China
*Corresponding author: Zhen Sheng-lai(Email: slzhen@ahu.edu.cn).

monitoring in geotechnical engineering entails a long construction period and accumulates a large amount of data, necessitating substantial manual effort and time for engineers to calibrate. Therefore, many algorithms have been developed for automatic classification [10,11,12,13]. In practical engineering, construction conditions are complex and multichannel- microseismic signals (MSs) are usually mixed with various background noises [14], prompting the development of denoising algorithms [15,16,17]. Traditional algorithms may not fully utilize the raw data; therefore, existing automatic classification methods continue to need engineering assistance. More powerful and advanced techniques must be urgently developed. Some researchers have tackled this challenge by incorporating neural networks [18, 19]. They have selected certain parameters to replace the waveform as the input for neural network classification. By feeding 11 extracted features into CNN, Peng et al. [20] achieved an astounding 98.2% accuracy in categorizing five signals. As deep learning techniques become more widely used and computer technology develops, researchers are increasingly using the original waveform data as input to prevent potential information loss. For instance, Lin et al. [21] classified multichannel microseismic waveforms with an accuracy of 91.13% using a deep convolutional neural network (CNN) with spatial pyramid pooling (DCNN-SPP). Additionally, Zhang et al. [22] investigated the use of CNN-MDN on raw microseismic data. They demonstrated the low sensitivity of CNN-MDN to noise of various intensities by examining semi-synthetic data. These results demonstrate that deep learning is a promising approach to microseismic data processing. However, most datasets used for academic research are selected only for MSs with high signal-to-noise ratios (SNRs) and typical noise signals (NSs), which can be easily categorized.

In practical engineering, especially in the initial stages of construction or when data availability is limited, manual calibration of data and collection of high-quality data require considerable time and physical labor. In addition, the reusability of the model for other projects remains problematic because of the large differences between different datasets. In recent years, researchers have proposed a migration learning approach that can be used to address these challenges. Tang et al. [23] introduced a unified transfer learning strategy aimed at transferring knowledge for microseismic identification across various items with different levels of difficulty. Similarly, Zhang et al. [24] proposed a fine-tuning approach for the full convolution U-Net architecture model after pre-training. The experimental results highlight the effectiveness of knowledge transfer for downhole microseismic data obtained from different sources. In other words, pre-trained models can be fine-tuned to perform well on new datasets, thus demonstrating the potential and versatility of transfer learning. However, the networks constructed by most researchers are overly complex and rely on hyperparameter tuning.

In this study, we collected a dataset containing typical microseismic waveforms and highly deceptive noise waveforms from the Hanjiang-to-Weihe River Diversion Project (HW). Subsequently, based on the improved VGG13 network, we proposed a basic CNN that may be best suited to identify microseismic events. Additionally, we employed the attention module in place of traditional denoising algorithms, which enables the network to handle multchannel signal processing effectively and greatly enhances network performance. The model was successfully adjusted and used to assess the network's suitability and dependability in real-world engineering applications, such as the Pan II Mine Project. The network's resilience, capacity for generalization, and potential for practical use were all validated by the experiments.

Preliminaries

1 Project Introduction and Data Description

The HW, also known as the South-North Water Transfer Project in Shaanxi, straddles the Yellow River and Yangtze River basins. It channels water from a designated source into the Han River, then through the Wei River water transfer tunnel, and eventually to the Huangchi Gorge water transfer project supplying water to the Guanzhong region. To prevent the risk of rock explosion, a microseismic monitoring system is used to monitor microseismic activities in the tunneling process during construction. For the monitoring configuration, A30 acceleration sensors are used to monitor cracks within the rock mass. Fig. 1(a) shows the layout of the sensor array with every two sensors 50 m apart, three on the right wall and the others on the left wall. The field experiment is shown in Fig. 1(b), and Fig. 1(c) shows the sensor collecting data by coupling the grout to the rock mass. The captured MSs can be efficiently transmitted to the transceiver using fiber optic cables. These signals are then converted to digital waveforms. Fig.2 illustrates the waveforms for six channels, with Fig. 2(a) and 2(f) representing the first and sixth channel waveforms, respectively.

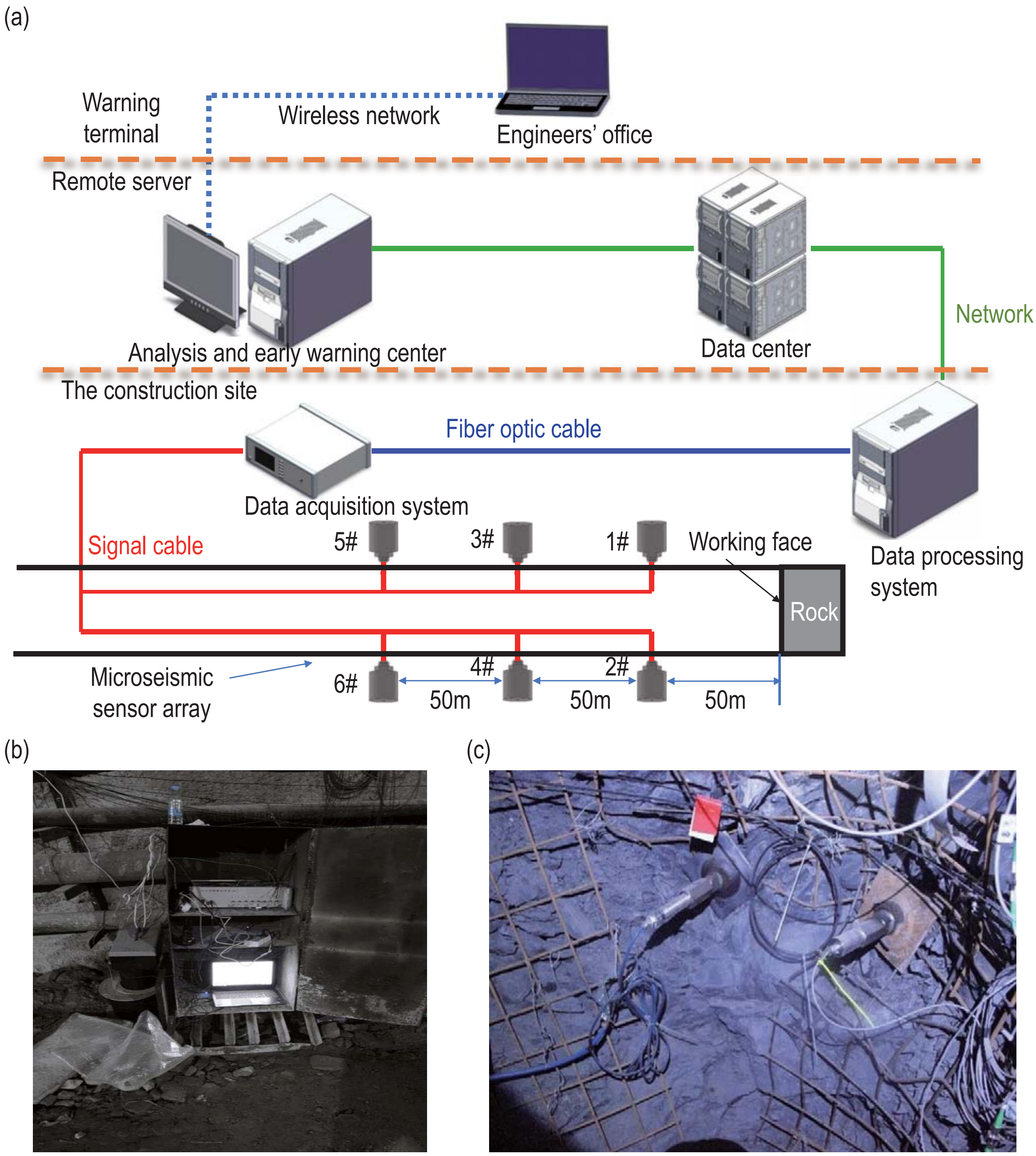

Fig. 1. (a) Topology diagram of the microseismic monitoring system, (b) field experiment diagram, (c) sensor installation diagram.

When observing Figs. 2(d), 2(e), and 2(f), distinguishing whether the signals are generated by rock fracture is challenging. Some examples of typical microseismic and atypical waveforms are given in Fig. 3. Fig. 3(a) shows a typical microseismic waveform. Figs. 3(b) and 3(c) show segments of low SNR signals that merge with the microseismic waveform and interfere with machine recognition. Fig. 3(d) is a highly deceptive noise, which can seriously affect the accuracy of signal identification in this case.

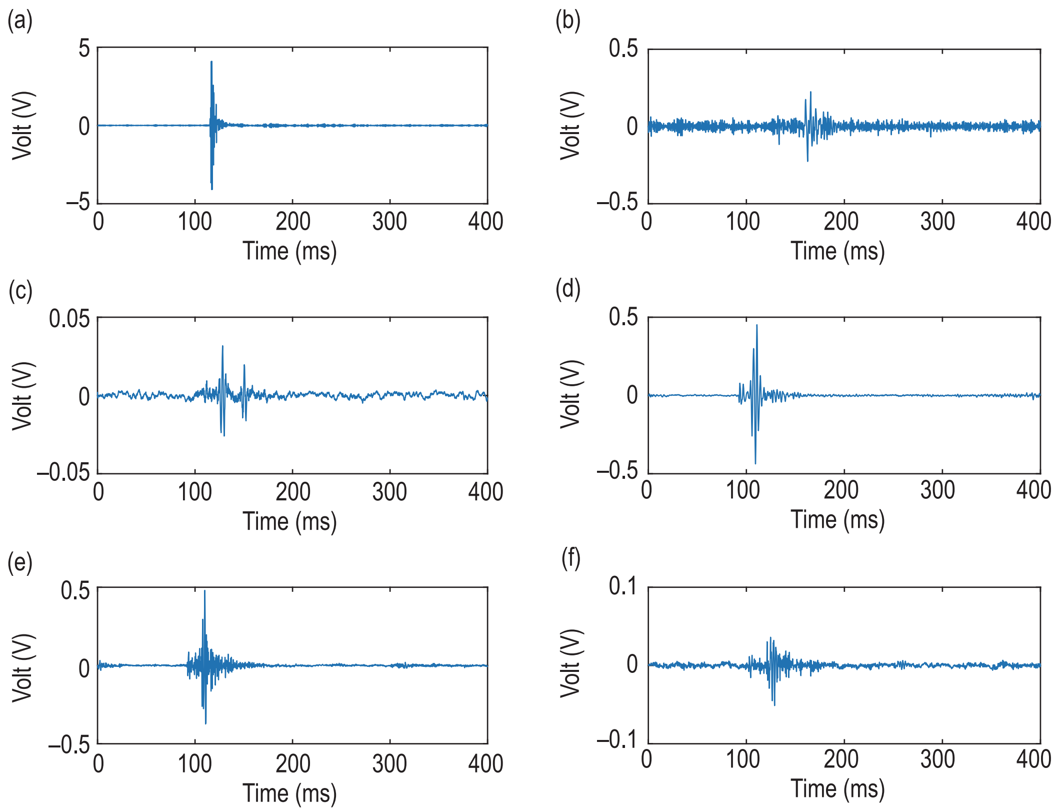

Fig. 2. Examples of multichannel microseismic data. (a), (b), (c), (d), (e), and (f) represent the waveforms acquired by six channels.

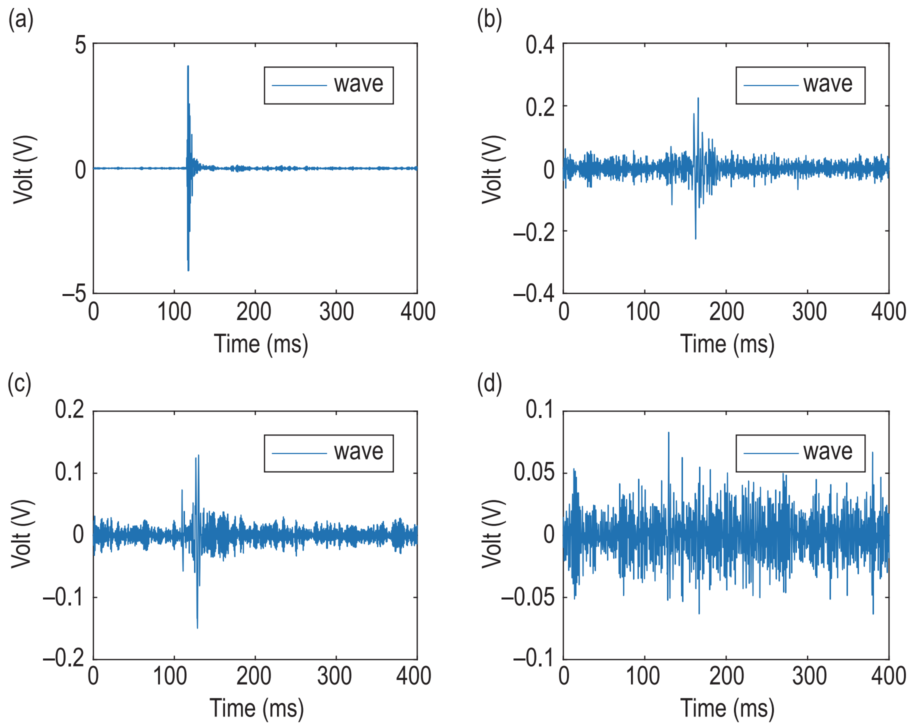

Fig. 3. Examples of microseismic waveform. (a) Typical waveform; (b), (c), and (d) atypical waveforms.

In practical engineering, many interference signals exist besides MSs. For example, Figs. 4 (a), 4(c), 4(e), 4(g), 4(i), and 4(k) depict MSs, blasting signals, rockburst signals, drilling signals, mechanical vibration

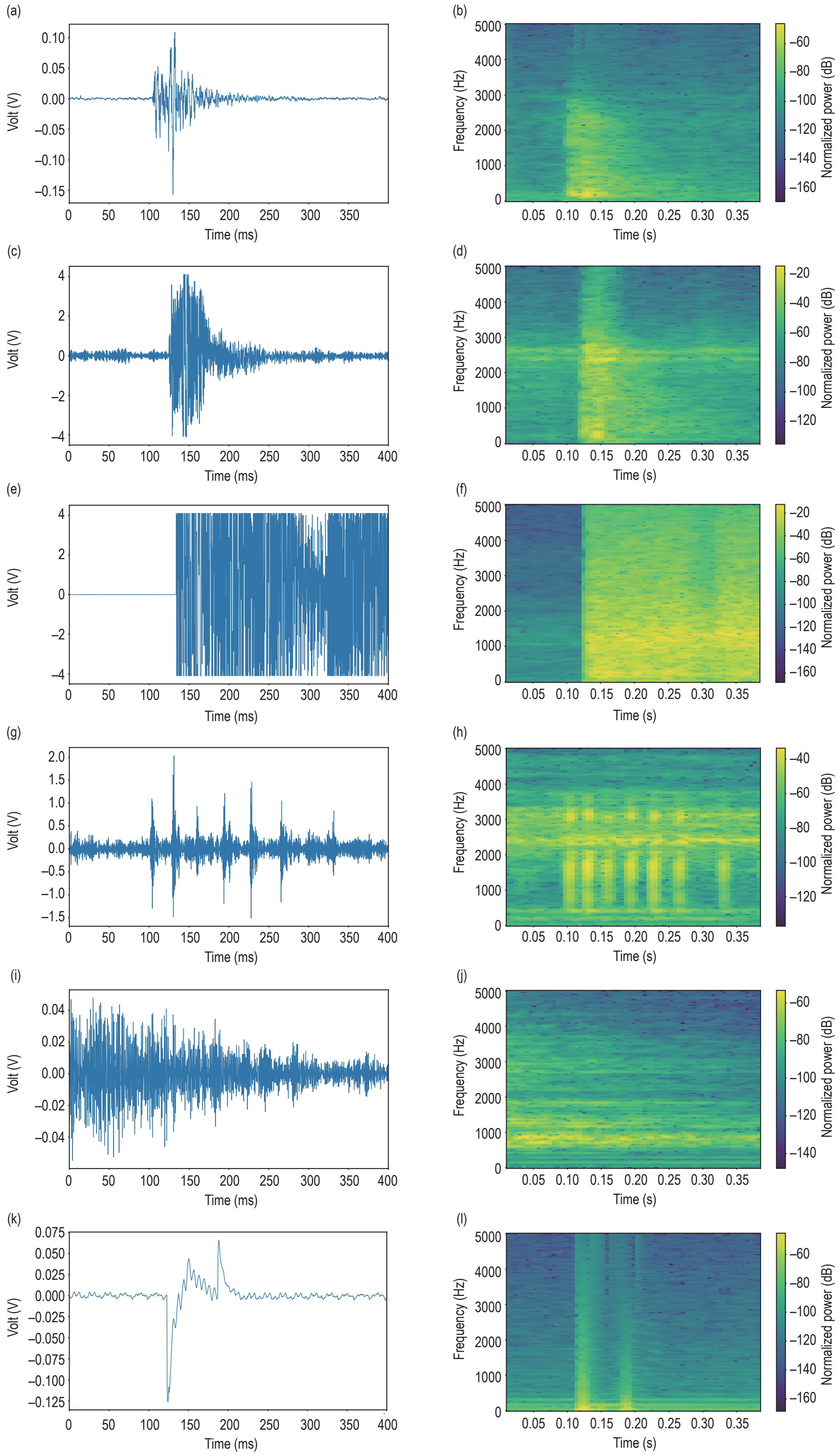

Fig. 4. Various types of signal waveforms. (a), (c), (e), (g), (i), and (k) are MSs, blasting signals, rockburst signals, borehole signals, mechanical vibration signals and electromagnetic interference signals, respectively. (b), (d), (f), (h), (j), and (l) are the corresponding time frequency domain features.

signals and electromagnetic interference signals, respectively, Figs. 4 (b), 4(d), 4(f), 4(h), 4(j), and 4(l) represent the corresponding time frequency domain characteristics. MS and drilling signals are similar to the waveform, with their peaks suddenly appearing and attenuating, and their peaks are relatively close. Rockburst and explosion signals have higher peaks and decay more slowly. Mechanical vibration signals present a disorganized waveform image, electromagnetic interference signals and micro-vibration signal waveforms are alike. These specific noises tend to interfere with machine recognition, so collecting a diverse range of waveforms in the dataset enables the model to learn from various scenarios.

1DCNN

As MS is typically 1D the networks must be converted to 1D versions so that they can be easily combined with the network modules proposed in this study. Therefore, the second dimension of the 1DCNN (vertical or horizontal dimension in the visualization) is set to 1. The CNN is mainly composed of the following structures:

(1) Convolution layer

The convolutional layer is performs feature extraction on the input data. Each neuron is connected to multiple neurons in a similarly located region in the previous layer, and the weight of that part of the line is constant. Moreover, each convolutional layer takes the input from the previous layer. By using different convolutional kernels after performing operations before inputting to the next layer, this process can be expressed by the following equation:

$$y_i^l = f\left(\sum_{i=0}^{m-1} w_i y_{p+i}^{l-1}\right), \qquad (1)$$

where wi denotes the weight of the position in the filter, y^{l-1} represents the output of the previous layer, $p+i$ denotes the position of the convolution kernel applied in the convolution operation, and f denotes the activation function, usually the tanh function.

(2) Pooling layers

The pooling layer is sandwiched between successive convolutional layers and it involves feature selection and information filtering. Although some information is lost in the process, the parameters and computation are reduced, finding a balance between model effectiveness and computational performance.

(3) Activation function

Each neuron node in a neural network accepts the output value of the neuron in the previous layer as the input value of that neuron and passes the input value to the next layer. A functional relationship exists between the output of the previous layer and the input of the nodes in the next layer, and this function is called the activation function. The tanh activation function used in the model, which converges faster, can be expressed by the following equation:

$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}, \qquad (2)$$

(4) Dropout

Dropout is a regularization method to prevent model overfitting, and it is widely used in deep learning. The layer temporarily removes some neurons randomly during the training to reduce the complexity and parameters of the neural network, thus effectively avoiding the overfitting problem.

3 Attention Mechanism

Diverse different signals have different durations, so we introduce an attention mechanism to empower the network with adaptive learning capabilities.

The CBAM is a lightweight general-purpose module that can be seamlessly integrated into any CNN architecture using two main independent attention mechanisms, namely channel attention and spatial attention. The basic principles of the attention mechanism draw their inspiration from the human tendency to show a high degree of attention when processing information. Through its ability to learn on its own, the attention mechanism recognizes and assigns different levels of importance to various pieces of information [25]. In this study, we mainly use the channel attention mechanism to perform feature reconstruction on the input features. We then use the spatial attention mechanism to reconstruct the reconstructed features again to obtain the final features. Fig. 5 shows the structure of CBAM.

The channel attention mechanism plays a vital role in recognizing important features. It first performs average and maximum pooling on the input elements. Two features are combined through a fully connected network. A sigmoid activation function is then applied, ensuring that the weights assigned to each feature component sum to 1. Subsequently, the resulting weight matrix is multiplied by the feature matrix, effectively assigning weights to different parts of the feature matrix. When analyzing multichannel waveform data, one can learn to selectively emphasize the more meaningful features.

Spatial attention focuses on recognizing the most informative region or portion of the input, thus complementing the channel attention mechanism. The signals collected in the field inevitably carry highly deceptive noise and contain small informative parts. The attention mechanism can effectively suppress the effect of noise, and to a certain extent can replace the traditional denoising algorithm while avoiding information loss from noise removal.

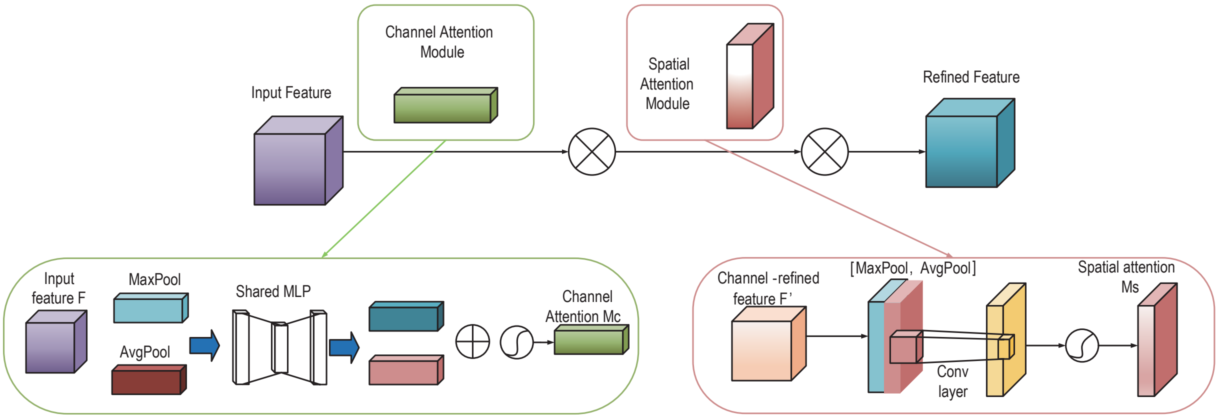

Fig. 5. Structure of the attention module.

Network Construction and Model Training

The VGG network is a deep CNN proposed by the Visual Geometry Group at the University of Oxford [26]. It stands out because of its well-structured network, making it highly compatible with hardware acceleration techniques.

1 Model Overview

The network model in this study is an improved network structure on the standard VGG13. Fig.6 shows the network structure: the convolutional layers are eight, and data features are extracted by convolutional computation. Setting the convolution kernel to 3 requires fewer parameters and speeds up training when the sensory field is the same. The use of the tanh activation function and dropout (0.25 in this study) after every two convolutional and pooling layers helps mitigate overfitting and enhance the generalization ability of the model. The Adam optimizer is selected, and the learning rate is set to 0.001. Softmax activation function is used to ensure that the outputs of each category form a probability distribution. Unlike VGG13, this study uses a global mean pool instead of a fully connected layer. It reduces the complexity of the model and improves the

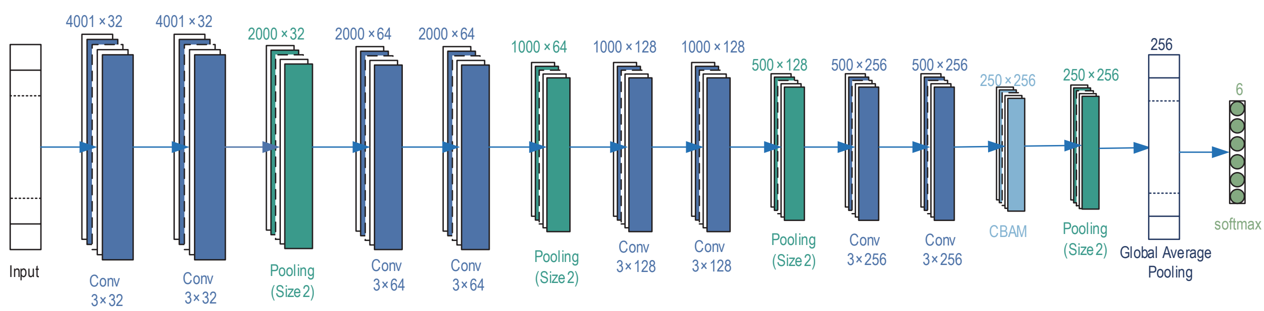

Fig. 6. Network structure.

overall performance. Table 1 provides the parameters for each layer, which include the different layers and their outputs, the convolutional kernel size, the activation function and the parameters.

Table 1. Parameters of each layer

Layer	Output	Kernel	Activation function	Parameters
Conv1	(None,4001,32)	3	tanh	608
Conv2	(None,4001,32)	3	tanh	3104
Maxp1	(None,2000,32)	2		0
Conv3	(None,2000,64)	3	tanh	6028
Conv4	(None,2000,64)	3	tanh	12352
Maxp2	(None,1000,64)	2		0
Conv5	(None,1000,128)	3	tanh	24704
Conv6	(None,1000,128)	3	tanh	24704
Maxp3	(None,500,128)	2		0
Conv7	(None,500,256)	3	tanh	98560
Conv8	(None,500,256)	3	tanh	196864
CBAM	(None,250,256)			16398
Global Avepl	(None, 256)			0
Dense	(None, 6)		softmax	1542

2 Experimental Study and Analysis

To verify the feasibility and effectiveness of the network, we fed the datasets collected in the field into the network for multiple classification identification. In this study, six types of datasets consisting of 1272 MSs, 486 blasting signals, 349 rock blasting signals, 566 drilling planning noises, 1042 mechanical vibration noises and 505 electromagnetic disturbances were selected. We divided them into two parts according to the ratio between the training set and the validation set, which is 8:2. All networks in this study were implemented using the TensorFlow framework. The PC (with i5-11400 CPU, Intel(R) UHD Graphics 730 and 8 GB RAM) fully distributes the computer's resources so that the model is capable of full learning. In Fig. 7, the accuracy and loss of CNN_BAM iterations on the training and validation datasets are shown. Accuracy reflects the probability of correctly detecting a waveform, whereas loss indicates the effectiveness of model learning. Lower loss values indicate better training results.

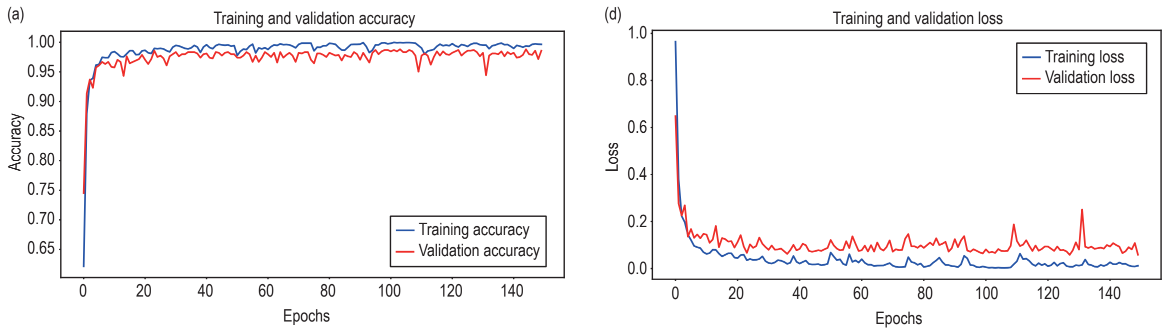

Fig. 7. Model training process and results. (a) Variation of accuracy with the epochs; (b) variation of loss with the epochs. The blue and red lines represent the accuracy and loss of the training and validation datasets, respectively.

The results show that as the number of iterations increases, the model converges to a steady state at roughly 120 iterations, which indicates that the model is finally close to fitting. At this stage, the model achieves an accuracy of roughly 99.8% with a loss of approximately 0.01 in the training set and an accuracy of roughly 99.29% with a loss of approximately 0.09 in the validation set. The model clearly demonstrates excellent performance in the training and validation phases.

3 Comparative Analysis

An effective model should exhibit strong generalization ability. In addition, standardized metrics are crucial for assessing the generalization ability of various classification models. Different tasks require diverse performance metrics, and commonly used metrics such as precision, recall, and F1-score are used to evaluate how well a model generalizes to a classification task. The multi-classification is made, in which the category is a positive category, and all other categories form an inverse category. The predicted results can be classified into four categories: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). Precision is the proportion of correct predictions, and recall is the proportion of correct predictions in the actual positive sample (both TP and FN). F1-score is used to assess the overall performance of the model. The higher the F1-score, the better the performance of the model. MSs and NSs are positive and negative examples in signal classification, corresponding to TP and TN, respectively. They can be correctly identified by the learner and expressed by the following equation:

$$P = \frac{TP}{TP + FP}, \tag{3}$$

$$R = \frac{TP}{TP + FN}, \tag{4}$$

$$F1\_\text{score} = \frac{2 \times P \times R}{P + R}, \tag{5}$$

To verify the stability and generalization ability of the network, we compare it with different networks. Among them, ResNet [27], as a deep residual network, has the ability to learn more complex features. AlexNet [28] is a classical deep CNN, which achieved a notable victory in the ImageNet image classification competition in 2012. SVM [29] and RF [30] are classical machine learning models. The former performs classification by finding an optimal hyperplane in the feature space, and the latter constructs multiple decision trees to handle a large number of features and samples. The results show that our proposed model achieves an accuracy of 99.29% on the validation set. Compared with other models, CNN_BAM outperforms other networks in all evaluation metrics of MSs, as shown in Table 2.

Table 2. Comparison between different networks

Methods	Validation Accuracy	Precision	Recall	F1-Score
CNNBAM	99.29	0.99	1.00	0.99
CNN	97.99	0.96	1.00	0.98
ResNet	93.48	0.98	0.93	0.96
AlexNet	92.00	0.98	0.91	0.94
SVM	52.13	0.45	0.52	0.44
RF	91.94	0.93	0.80	0.86

CNN_BAM and CNN are selected for comparison of different signals. As shown in Table 3, the evaluation metrics of CNN_BAM multi-classification outperform those of the other networks.

Table 3. Comparison of the two methods on a validation dataset

CNN_BAM				
Classes	**Precision**	**Recall**	**F1_score**	**Reported Runtime**
MS	0.99	1.00	0.99	4.4h
Blast	0.97	0.99	0.98	
Rock blast	0.97	0.94	0.95	
Drilling and planning	0.95	1.00	0.98	
Mechanical vibration	1.00	0.98	0.99	
Electromagnetic interference	0.95	0.99	0.97	

CNN				
Classes	**Precision**	**Recall**	**F1_score**	**Reported Runtime**
MS	0.95	0.98	0.97	3.6h
Blast	1.00	0.99	0.99	
Rock blast	0.98	0.97	0.98	
Drilling and planning	0.97	0.99	0.98	
Mechanical vibration	0.97	0.99	0.98	
Electromagnetic interference	1.00	0.86	0.92	

Migration Learning and Reusability

Differences in geological conditions, construction methods, and monitoring equipment lead to variability between diverse rock projects. To save considerable labor and time, we would like the network structure to be trained without having to start from scratch, using only a limited number of labeled data samples from the new project. This technique is known as transfer learning. It starts from trained models and trains methods for new related domains. One of the common solutions is the fine-tuning technique, which learns from the original task and adapts to the specific requirements of the new task [31]. This approach has proven its effectiveness in practical applications.

1 Model Overview and Analysis of Results

Owing to the small size of the new two datasets, to avoid overfitting, we train only the classifier of the network, keeping all other layers unchanged. As shown in Fig. 8, we add a fully connected layer and change the last layer of the network to binary classification. All the other layers are frozen, keeping the pre-trained parameters.

Two sets of data are input: the first set of data HW project of different periods and the second set of data from the PJ-2. A total of 250 sets of MSs and 250 sets of NSs are selected for the two sets of data. PJ-2 is located in the northern part of the Huaihe River and in the eastern part of the Huainan Mining Area. To effectively monitor the impact of the bottom slab mining damage zone on the gray rock aquifer and the change of stratum stress during the working face back mining, a microseismic monitoring system is used to collect the MSs caused by the rock layer rupture of the working face bottom slab during the mining process. Through precise calculation, the damaged location of the

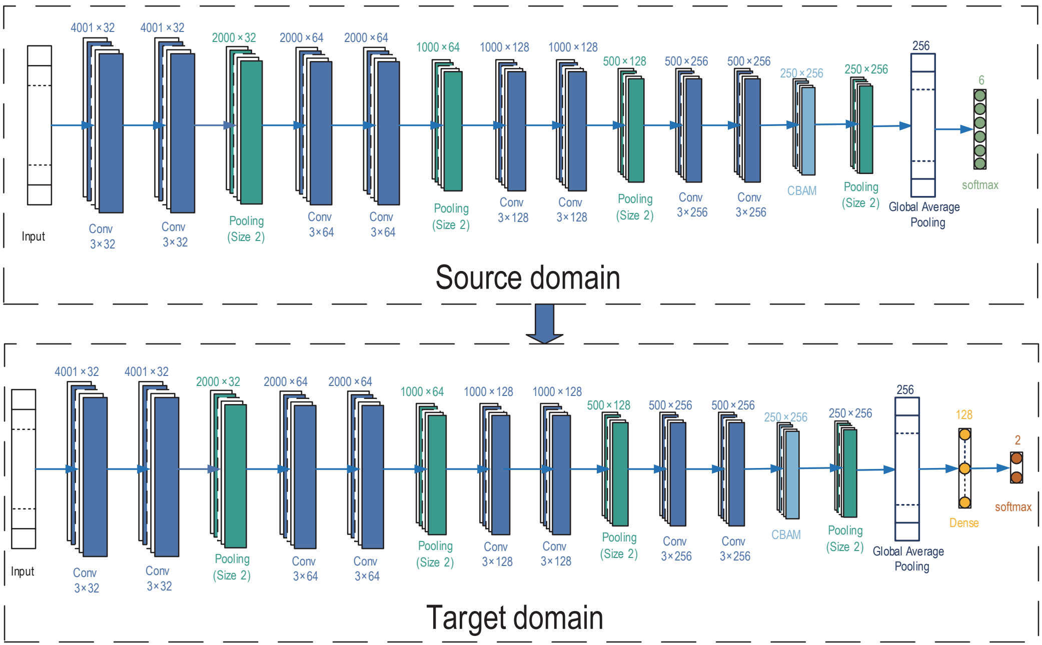

Fig. 8. Schematic of the structure of the fine-tuning network.

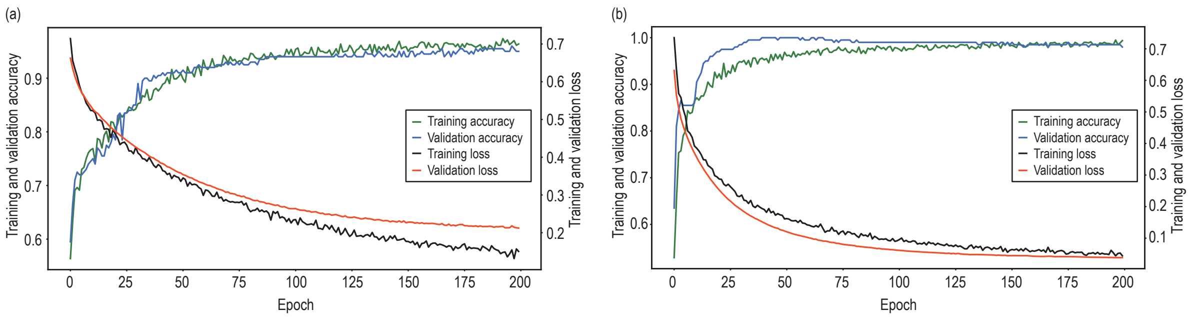

Fig. 9. Fine-tuning model training process and results.

microfracture and the area that may be affected can be determined to improve the safety and efficiency of the mine production. The classification results are shown in Fig. 9. At a learning rate of 0.0005, Fig. 9(a) shows the superior performance of the first dataset (96% accuracy). Data for PJ-2 are collected in coal mines owing to differences in geological conditions, construction methods and monitoring configurations. The PJ-2 dataset is less similar to the main dataset, and Fig. 9(b) shows an accuracy rate of approximately 92%, as shown in Table 4.

Table 4. Fine-tuned results of the network generalization ability validation in datasets

	HW			
	Precision	**Recall**	**F1-Score**	**Accuracy**
MS	0.94	0.96	0.95	0.96
NS	0.96	0.94	0.95	

	PJ-2			
	Precision	**Recall**	**F1-Score**	**Accuracy**
MS	1.00	0.84	0.91	0.92
NS	0.86	1.00	0.93	

2 Discussion

Through model training and validation, CNN_BAM shows better performance in microseismic multi-classification than other networks. When variability between two datasets is inevitable, and the data amount is minimal, the optimized CNN_BAM network model demonstrates superiority in information transfer across various engineering contexts. It still has the following drawbacks. (1) Training samples are crucial for neural networks to function properly. The model is prone to overfitting and inadequate training if the samples available are insufficient for the network to completely learn the features of the categorized subjects. The environment of geotechnical engineering is complicated, so gathering datasets with a variety of waveform patterns is difficult. (2) For the fine-tuning model proposed in this study, future work should increase the signal samples of various waveform signals from different regions and improve the network structure through further training. Doing so can enhance the intelligence of automatic classification and subsequent analysis of microseismic data, saving labor and time costs.

Acknowledgments

This work was supported by the Key Research and Development Plan of Anhui Province (202104a05020059) and the Excellent Scientific Research and Innovation Team of Anhui Province (2022AH010003). Partial financial support from Hefei Comprehensive National Science Center is highly appreciated.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Conclusion

On the basis of CNN and attention mechanism in deep learning, this study proposes a new lightweight network structure for efficient recognition of microseismic multichannel events. The classification performance is evaluated by combining CNN_BAM with several state-of-the-art networks. CNN_BAM outperforms other networks in terms of precision, recall, and F1-score, demonstrating its effectiveness and robustness. By fine-tuning the network structure, knowledge migration under different projects is achieved, which successfully solves the problem that the deep learning model cannot be repurposed among different microseismic monitoring projects in diverse geological settings. This capability greatly saves time and financial resources and improves the efficiency of researchers in different projects. The network can be utilized for intelligent monitoring in different applications related to rock engineering because of its generality and versatility. As neural network algorithms continue to progress, the network is expected to handle different kinds of data and carry out a variety of tasks. This progression is poised to contribute to the field of intelligent monitoring in rock engineering.

References

Wang, Q.H., Zhang, Y.T., Xie, Z., Zhao, Y.W., Zhang, C., Sun, C., Wu, G.H., 2022, The Advancement and Challenges of Seismic Techniques for Ultra-Deep Carbonate Reservoir Exploitation in the Tarim Basin of Northwestern China: Energies, **15**, 7653.

Nan, L. I., En-Yuan, W., & Mao-Chen, G. E., 2017, Microseismic Monitoring Technique and Its Applications at Coal Mines: Present Status and Future prospects: Journal of China Coal Society, **42**, 83–96.

Li, J., Yu, B. S., Tian, Y. K., Kang, H. X., Wang, Y. F., and Zhou, H., 2020, Effect analysis of borehole microseismic monitoring technology on shale gas fracturing in western hubei: Applied Geophysics, **17**, 764–775.

Tang, Shibin, J. Wang, and C. Tang., 2021., Identification of microseismic events in rock engineering by a convolutional neural network combined with an attention mechanism: Rock Mechanics and Rock Engineering, **54**(1), 1–23.

Zhang, C., Jin, G., Liu, C., Li, S., Xue, J., Cheng, R., Wnag, X., Zheng, X., 2021, Prediction of rockbursts in a typical island working face of a coal mine through microseismic monitoring technology: Tunnelling and Underground Space Technology, **113**, 103972.

Liu, L., Song, W., Zeng, C., and Yang, X., 2021, Microseismic event detection and classification based on convolutional neural network: Journal of Applied Geophysics, **192**(6), 104380.

Xu, S., Zhang, C., Chen, Z., Li, Y., and Liu, J., 2021, Accurate identification of microseismic waveforms based on an improved neural network model: Journal of Applied Geophysics, **190**, 104343.

Zhu, M., Wang, L., Liu, X., Zhao, J., and Peng, P., 2018, Accurate identification of microseismic p- and s-phase arrivals using the multi-step AIC algorithm: Journal of Applied Geophysics, **150**, 284–293.

Angus, D. A., Aljaafari, A., Usher, P., and Verdon, J. P., 2014, Seismic waveforms and velocity model heterogeneity: towards a full-waveform microseismic location algorithm: Journal of Applied Geophysics, **111**, 228–233.

A, P. B. Q., A, P. G., A, Y. C., and B, M. U., 2015, Detection and classification of seismic events with progressive multi-channel correlation and hidden markov models: Computers & Geosciences, **83**, 110–119.

Dong, L., Wesseloo, J., Potvin, Y., and Li, X., 2016, Discrimination of mine seismic events and blasts using the fisher classifier, naive bayesian classifier and logistic regression: Rock Mechanics and Rock Engineering, **49**(1), 183–211.

Kuyuk, H. S., Yildirim, E., Dogan, E., and Horasan, G., 2011, An unsupervised learning algorithm: application to the discrimination of seismic events and quarry blasts in the vicinity of Istanbul: Natural hazards and earth system sciences, **11**(1), 93–100.

Dong, Long-Jun, Wesseloo, Johan, Potvin, and Yves, et al., 2016, Discriminant models of blasts and seismic events in mine seismology: International Journal of Rock Mechanics & Mining Sciences, **86**, 282–291.

Zhou, Y., & Wu, G., 2020, Unsupervised machine learning for waveform extraction in microseismic denoising: Journal of Applied Geophysics, **173**, 103879.

Iqbal N, Liu E, McClellan J, Al-Shuhail A, Kaka S, Zerguine A., 2018, Detection and denoising of microseismic events using time–frequency representation and tensor decomposition: IEEE Access, **6**, 22993–23006.

Chen., Y., 2016, Dip-separated structural filtering using seislet transform and adaptive empirical mode decomposition based dip filter: Geophysical Journal International, **206**(1), 457–469.

Xing Li, Z., Lian Yue, C., Yan, C. et al., 2022, Microseismic signal denoising by combining variational mode decomposition with permutation entropy: Applied Geophysics, **19**, 65–80.

Zhang, XL., Jia, RS., Lu, XM. et al., 2018, Identification of blasting vibration and coal-rock fracturing microseismic signals: Applied Geophysics, **15**, 280–289.

Lin, B., Wei, X., and Junjie, Z., 2019, Automatic recognition and classification of multi-channel microseismic waveform based on DCNN and SVM: Computers & geosciences, **123**, 111–120.

Peng, P., He, Z., Wang, L., and Jiang, Y., 2020, Automatic classification of microseismic records in underground mining: a deep learning approach: IEEE Access, **8**, 17863–17876.

Lin, B. I., Wei, X. I. E., Junjie, Z. H. A. O., and Hui, Z. H. A. O., 2018, Automatic classification of multi-channel microseismic waveform based on DCNN-SPP: Journal of Applied Geophysics, **159**, 446–452.

Girshick, R., Donahue, J., Darrell, T., and Malik, J., 2014, Rich feature hierarchies for accurate object detection and semantic segmentation: IEEE Computer Society, 580–587.

Wang, J., Tang, S., 2022, Novel Transfer Learning Framework for Microseismic Event Recognition Between Multiple Monitoring Projects: Rock Mechanics and Rock Engineering, **55**, 3563–3582.

Yilun, Z., Jiaxuan, L., Yihan, D., Zhichao, Y., Tianyue, H., and Chuan, H., 2022, Phase arrival picking for bridging multi-source downhole microseismic data using deep transfer learning: Journal of Geophysics and Engineering, **19**(2), 178–191.

Jaderberg, M., Simonyan, K., Zisserman, A., and Kavukcuoglu, K., 2015, Spatial transformer networks: MIT Press.

Simonyan, K., & Zisserman, A., 2014, Very deep convolutional networks for large-scale image recognition: Computer Science.

He, K., Zhang, X., Ren, S., & Sun, J., 2016, Deep Residual Learning for Image Recognition: IEEE Conference on Computer Vision and Pattern Recognition IEEE.

Krizhevsky, A., Sutskever, I., & Hinton, G., 2012, Imagenet classification with deep convolutional neural networks: Advances in neural information processing systems, **25**(2).

Cortes, C., & Vapnik, V., 2009, Support-vector networks: Chemical Biology & Drug Design, **297**(3), 273–297.

Statistics, L. B., & Breiman, L., 2001, Random forests: Machine Learning, 5–32.

Kading C, Rodner E, Freytag A, Denzler J., 2017, Fine-tuning deep neural networks in continuous learning scenarios: Lecture Notes in Artificial Intelligence, 10118, 588–605.

Shu Jin is currently working toward her MS degree in optical engineering in the School of Physics and optoelectronic engineering, Anhui University, Hefei, China. She is a master's candidate directed by Prof. Shenglai Zhen.

Shenglai Zhen is a professor at the School of Physics and Optoelectronic Engineering at Anhui University, Hefei, China. He received an MS degree in optics in 2003 and a PhD degree in electromagnetic field and microwave technology in 2008, both from the Anhui University. He is mainly engaged in the research of optoelectronic sensors and laser sensors, especially in the field of interferometric sensors.