Computers and Geosciences 123 (2019) 111–120

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

Automatic recognition and classification of multi-channel microseismic waveform based on DCNN and SVM

Bi Lin*, Xie Wei, Zhao Junjie

School of Resources and Safety Engineering, Central South University, China

ARTICLE INFO

Keywords:
Multi-channel waveform
Recognition
Classification
DCNN
SVM
Feature

ABSTRACT

The recognition and classification of the multi-channel microseismic waveform are important for mine hazard prediction. It is widely used to design the corresponding waveform feature for recognition and classification of the microseismic waveform by hand. The process of designing features manually is arduous and the results of recognition and classification are not ideal. In this paper, we propose a method combining Deep Convolutional Neural Networks (DCNN) with Support Vector Machine (SVM) for identifying the microseismic waveform automatically. We constructed a DCNN structure to train the optimal weight model named the DCNN-Model. The DCNN-Model is used as a tool for extracting features from multi-channel waveforms. After combining the extracted features, we used SVM to classify multi-channel waveforms. We compared the outputs of other classifiers, such as Random Forest and k-Nearest Neighbor (KNN). To extend the dataset of DCNN training and extract the essential characteristics of waveform images more accurately, we pre-process the raw data by means of filtering and de-nosing. The experiment shows that the recognition and classification method is of practical value, and the accuracy rate can reach as high as 98.18%.

1. Introduction

In deep mining, the problem of safety in mine production is becoming more and more serious. To ensure safe production, the microseismic monitoring system was used to monitor the seismicity dynamically in the mine, and real-time monitoring of rock mass stability is realized through analysis of the microseismic waveform during production activities. Due to the complexity of the mine geological environment, monitored microseismic signals are often subject to interference from background noise. It is therefore difficult to recognize the seismic waveform effectively.

Owing to the complexity of the application of mine field, it is still difficult to classify microseismic signals, blasting signal and noise automatically. At home and abroad, researchers have performed many practical studies for recognition and classification in the field. Quan-jie et al. (2012) used an algorithm involving the range and scale-free fractal box dimensions that addressed the fractal characteristics of microseismic signals, and identify mechanical vibration waveforms, blasting waveforms and rock burst waveforms based on SVM. Tan et al. (2010) used the microseismic data from Cold Lake, Alberta provided by Imperial Oil Ltd. to research and develop new methods such as frequency filtering, event length detection and statistical analysis, and they were used to accurately and automatically classify microseismic event signals, respectively. The test proved that the statistical analysis algorithm and the principal component analysis method are combined and the classification is optimal. Dong et al. (2016) used logarithmic logic and generalized logic distribution to establish the origin time difference (ODT) probability density model of the adjacent explosion zone, and employed 7 parameters as discriminant indicators and logistic regression to construct a discriminant model, which was applied to the classification of two mines, and the accuracy rates were 96% and 95%, respectively. Bui Quang et al. (2015) first used the progressive multi-channel correlation (PMCC) detector to detect the coherent wavefront of the sensor array, and converted the detected signal into a feature vector sequence to train and test the hidden Markov model (HMM), which was used to classfy the seismic events.The antecedent researchers have performed much prospective work for extraction and classification of microseismic waveform features. However, these methods designed associated features from the relevant waveform by hand, and classify the designed features by a certain method. Typically, manual design of features is troublesome and time consuming, and professional knowledge is necessary to determine the proper method. In this paper, Deep Convolutional Neural Networks (DCNN) was used for training the optimal weight model named the DCNN-Model which extracted features of multi-channel waveforms automatically; then, all channel features of the multi-channel waveform were combined as the

* Corresponding author. School of Resources and Safety Engineering, Central South University, Changsha, 410083, China.
E-mail address: mr.bilin@163.com (B. Lin).

https://doi.org/10.1016/j.cageo.2018.10.008
Received 2 October 2016; Received in revised form 3 August 2018; Accepted 22 October 2018
Available online 26 October 2018

input data of Support Vector Machine (SVM) for microseismic waveform recognition and classification.

The outline of the paper is as follows. In Section 2, according to the characteristics of the microseismic signals, we prepared two data sets for the microseismic recognition and classification. The first data set was used to train and test the optimal weight model, the DCNN-Model. The other data set was used to train and test the SVM classifier. In Section 3, the DCNN structure was constructed to train the DCNN-Model with massive image data, and automatic classification of the multi-channel waveform was designed by SVM. Meanwhile, we compared SVM with other classifiers such as Random forest and k-Nearest Neighbor (KNN). Our experiment, combining DCNN with SVM, is presented in Section 4 and trained and tested on real data. Based on the experimental results, we analysed the DCNN model and compared the classification results of classifiers. Finally, conclusions are drawn in Section 5.

2. Data preparation

In this paper, the microseismic monitoring data are from the Dongguashan Copper Mine in Tongling City, Anhui Province, China, which uses the Institute of Mine Seismology (IMS) microseismic monitoring system from South Africa. The Dongguashan Copper Mine is a typical deep buried ore body with high in situ stress, so the possibility of pressure damage and bump burst hazard is very high.

When a microseism event occurs in a mine, it usually triggers an array of sensors, and the microseismic monitoring system will collect and store the corresponding multi-channel waveform. Fig. 1 shows 8 waveforms of one event; the distance between the sensors and the source is different or the signal-to-noise ratio is variable, and thus the waveforms received by the sensor can change dramatically.

We prepared two data sets for recognition and classification microseismic events. The first data set consists of the images of all channels and was used to train and test the DCNN model. According to the characteristics of the waveforms, all images were classified into three types: microseismic image, blasting image and noise image. To ensure the amount of data required for DCNN model training and to ensure its extraction to the essential characteristics of the waveform image, the first data sets are processed by denoising and filtering, and the processed data are added to the first data set to extend the number of the first data set(Chen and Lin, 2014). The second data set is composed of a single event multichannel waveform image. Each event contains the waveform images of the number of triggered sensors to train and test the SVM classifier, including blasting events, microseismic events and noise events. Two data sets were divided into a training set and test set according to a certain proportion for training and testing DCNN and SVM, respectively. The data processing procedure is shown in Fig. 2.

2.1. Artificial recognition method

At present, according to the characteristics of the waveform (Abdelwahed, 2012) detected by the microseismic monitoring system, the technicians of the Dongguashan Copper Mine analyze the morphological characteristics of the waveforms, and manually classify and identify its waveform types. Fig. 3 shows that the morphological characteristics of artificial recognition waveforms, including wave amplitude, rise time, duration, interval time, and trigger threshold.

In addition, artificial recognition distinguishes signal types through mining vibration mechanisms: (1) Typically, the generation of the waveform is related to blasting methods, such as deep hole blasting, medium deep hole blasting and drilling blasting. There are many peaks in the waveform detected by millisecond blasting, and the energy of blasting increases continuously, and the amplitude of the latter is larger than that of the preceding one. (2) The waveform types are identified and classified by the characteristics of wave energy released by the damage mechanism of the rock mass. There are two types of mechanisms for rock mass damage, which refers to shear failure and tensile failure. Shear failure releases pressure waves (P-wave) and shear waves (S-wave), and the propagation velocity of the P-wave is faster than that of S-wave, while tensile failure generates only the P-wave. Therefore, the waveform type can be distinguished by the principle whereby there exists an S-wave or the P-wave arrival is faster than the S-wave.

2.2. Image de-noising and filtering

The microseismic signals will be seriously disturbed during the transmission process due to the complicated underground environment conditions. The interference signals include motor vehicle noise,

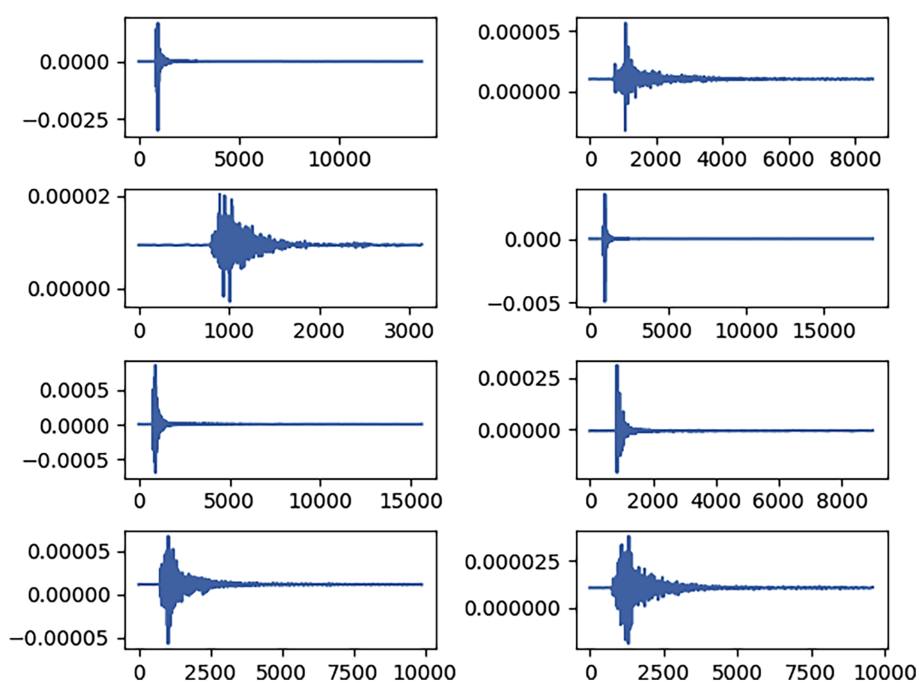


Fig. 1. The multi-channel waveform display.

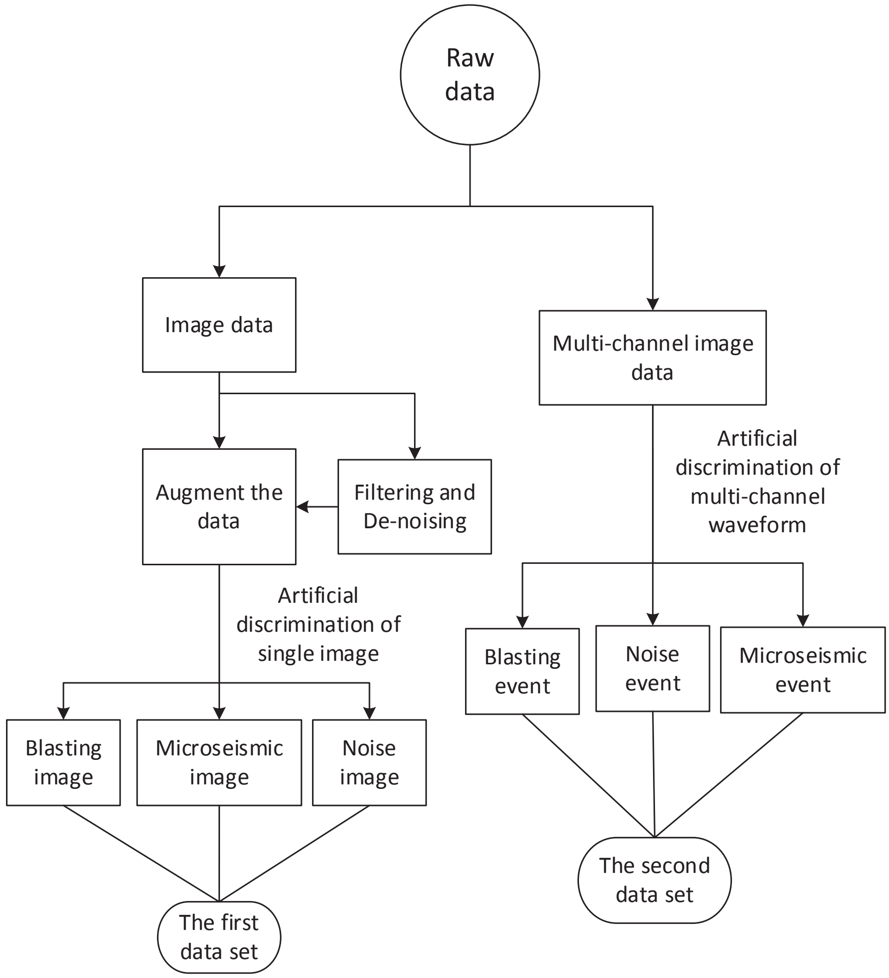


Fig. 2. The process of data preparation.

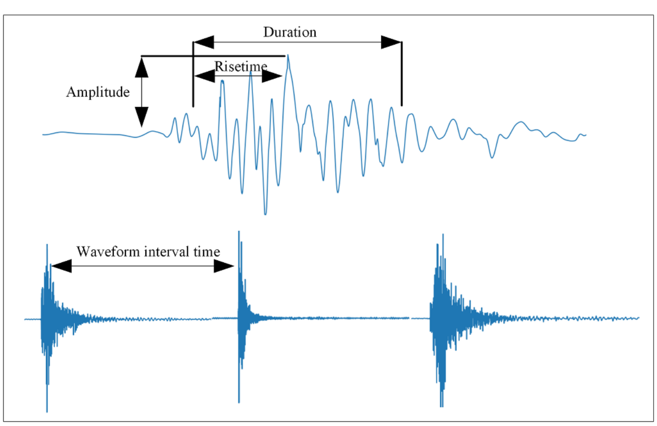


Fig. 3. Characteristic of artificial recognition signal.

machinery, equipment noise, and electromagnetic interference. Most monitoring data are composed of a variety of effective signals and interfering signals. Therefore, to expand the data set of DCNN model training and to extract the essential features of the waveform image more accurately, this paper uses the wavelet threshold denoising method to denoise and filter the original waveform image data.

Wavelet threshold de-noising is a method to address nonlinear signals. In the process of denoising and filtering, the amplitude of the effective signal does not vary with the increase in the scale of wavelet analysis, but the interference signal can be attenuated rapidly to eliminate the excessive interference signal in the effective signal (Xuelong et al., 2015). Formula (1) represents a one-dimensional model containing noise signals.

$$s(k) = f(k) + \varepsilon e(k)\,(k=0,1,\cdots,n-1) \tag{1}$$

where s(k) , f(k) , ε , e(k) are, respectively, the signal with containing noise, a stationary signal of low frequency, the amplification factor of noise and white Gaussian noise. The process of wavelet de-noising is shown in Fig. 4.

In this paper, Sym8 is selected as the wavelet basis function of wavelet threshold denoising (Chavan et al., 2011), the decomposition layer number is selected as 5 layer decomposition, and the threshold function selects the heuristic SURE fixed threshold for signal denoising. The contrast between the de-noising waveform and the original waveform is shown in Fig. 5.

Raw data are discrete points, and they can be described in a two-dimensional axis, where the horizontal axis stands for time and the vertical axis represents the amplitude. Because the depth of the sensor is different, the distance from the signal source is different, and the propagation process is disturbed by noise and obstacles, it usually

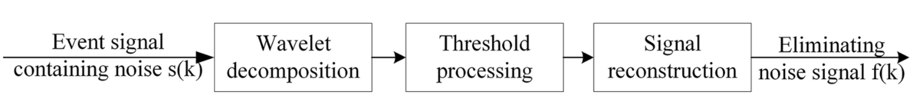


Fig. 4. Wavelet de-noising process.

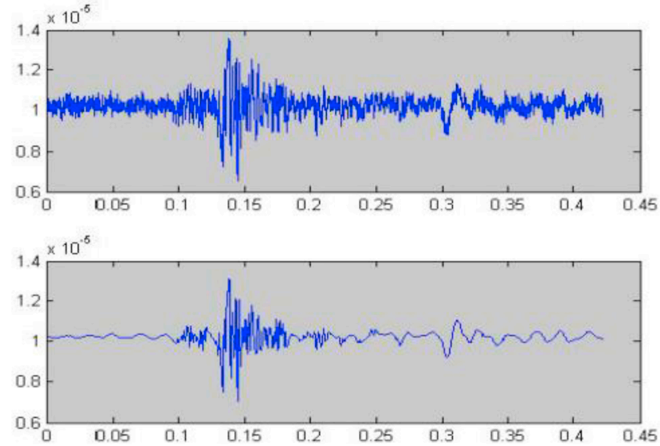


Fig. 5. Comparison of the original waveform and the de-noising waveform.

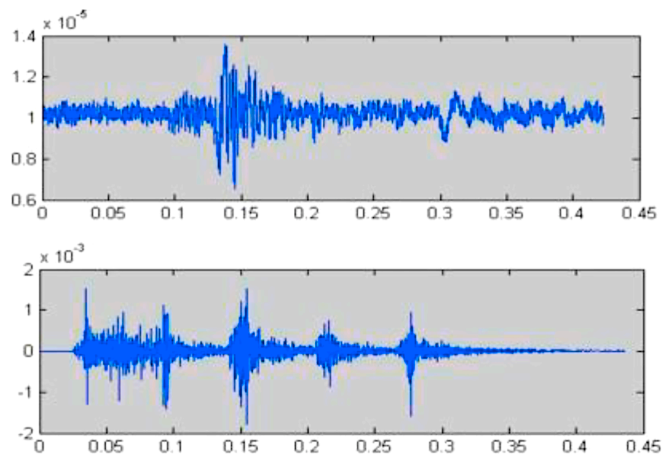


Fig. 6. Comparison of the amplitude for different waveforms.

triggers a number of sensors, and the amplitude of the received signal is not consistent. Therefore, the amplitude of different channels is different, and the difference is greater than 10^2. Due to the large difference in image size, it is necessary to zoom in to a certain scale (see Fig. 6) to prevent the image size from affecting the accuracy of feature extraction.

In this paper, the wavelet threshold de-noising method is used to denoise the original waveforms, and thus the total amount of the first data set is nearly doubled to improve the generalization ability of the DCNN weight model and prevent overfitting of the network. In addition, wavelet threshold denoising can also enhance the effect of the effective signal and suppress the influence of noise interference to extract the essential feature of the waveform image data more accurately.

3. Multi-channel recognition and classification based on DCNN and SVM

3.1. Building the DCNN structure to extract features

In recent years, the development of deep learning(LeCun et al., 2015) has been extremely rapid, DCNN has been successfully applied in some fields, such as behavior recognition and handwritten recognition (Ciresan et al., 2010; Lawrence et al., 1997; LeCun and Bengio, 1995; Wang et al., 2012). In 2016, Hinton et al. (Krizhevsky et al., 2012; Srivastava et al., 2014) applied it in the NIPS2012 to the largest database ImageNet in the field of image recognition, achieving good results. Ji Shuiwang(Ji et al., 2013) and other researchers extended the DCNN from the original 2D CNN to 3D CNN for human action recognition and video surveillance of the airport; the recognition results show its superior performance.

Traditional image recognition methods require features (f1,f2, …fn) to be designed by hand and use a certain classifier to classify the designed features, as shown in Fig. 7(a); the manual design of waveform characteristics requires a good professional knowledge background and is difficult and time-consuming. Therefore, in this paper, the deep convolutional neural network (DCNN) is used to extract the waveform feature automatically, and identify and classify single event multi-channel microseismic events.

DCNN is a type of neural network with supervised learning. In the hidden layer, the convolution and pooling layers are the core module of the DCNN. To accelerate convergence or prevent over fitting, the hidden layer also uses a number of other network layers, such as the normalized layer or the dropout layer. Forward propagation is used to drive network abstraction and abstraction features, and the back propagation algorithm is used to optimize network weights (Jin et al., 2000). Due to the avoidance of manual design features, as shown in Fig. 7(b), the waveform image data can be directly input and calculated and output calculation results, that is, it is an end-to-end learning process and has a good recognition effect.

To extract the image feature effectively, the 3 × 3 and 5 × 5 types of convolution kernels of DCNN model were used as the basic components of convolution operations. The whole network structure of DCNN is 13 layers deep, including an input layer, 7 convolution layers, 3 max

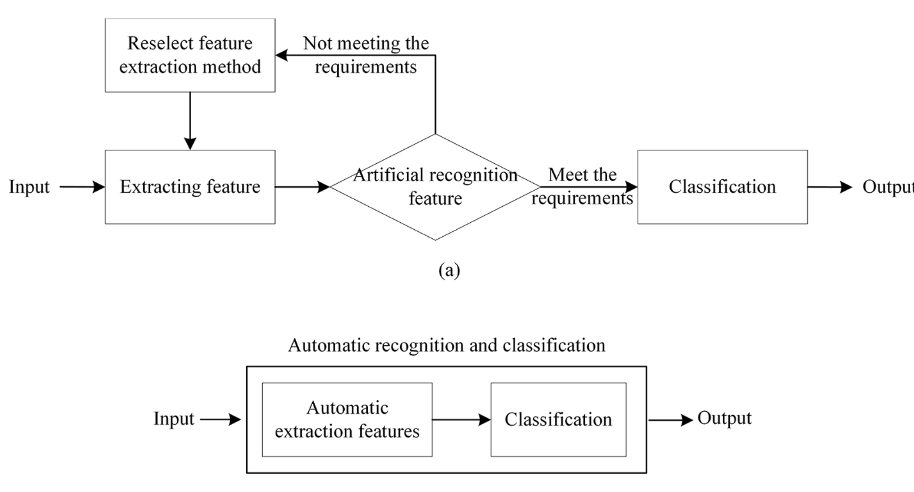
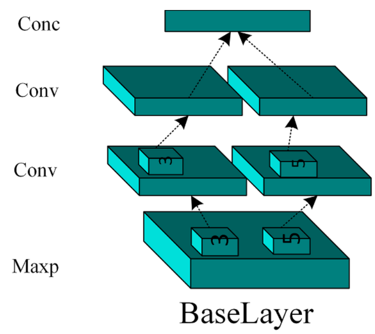
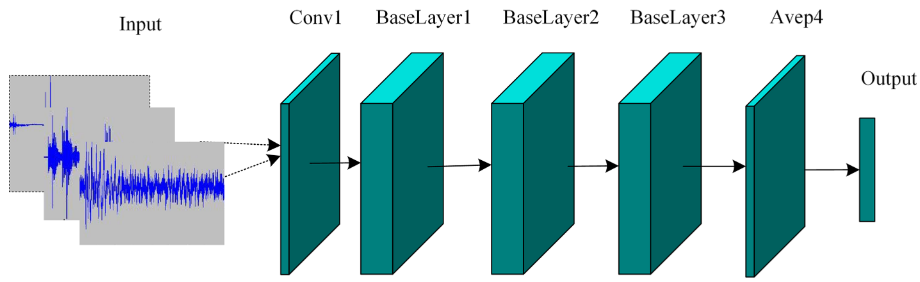


Fig. 7. Comparison of traditional and DCNN image pattern recognition (a) Traditional image pattern recognition; (b) DCNN image pattern recognition.



(a)



(b)

Fig. 8. (a) The BaseLayer of DCNN; (b)The overall structure of DCNN. Conv represents the convolution layer; Conc represents the concatenate layer; Maxp is the max pooling layer; Avep4 is the average pooling layer; 3, 5 represent 3 x 3, 5 x 5 size of convolution kernel.

pooling layers, 1 average pool layer and 1 output layer. The overall structure of DCNN is shown in Fig. 8.

The convolutional layer is to extract features of input feature maps (or images) with the operation of convolution, which uses a certain learnable convolution kernel which stands for extracting a certain feature of the input maps. The convolution kernels are randomly initialized and optimized by the backpropagation algorithm. The output of the convolutional layer is defined as:

$$y_j^l = Relu\left(\sum_{i \in M_j} x_{ij}^l * k_{ij}^l + b_j^l\right) \tag{2}$$

where *Relu* represents the activation function, x_{ij}^l denotes the i-th input of the j-th neuron of the layer l, k_{ij}^l represents the size of the convolution kernel between neuron j of the layer l and neuron i of layer $l-1$, * denotes a convolution operation, M_j represents a selection of input maps, and b_j^l is an additive bias of neuron j of layer l.

The first convolution layer of the DCNN network structure is obtained with a 7 × 7 convolution kernel, followed by the max-pooling layer, followed by two convolution layers obtained with the 3 × 3 and 5 × 5 mixed convolution kernel, followed by the max-pooling layer, where a total of 6 stacked convolution layers and 3 max-pooling layers are used. The mixed convolution kernel (3 × 3 and 5 × 5) can extract features of different sizes in the convolution layer and reduce the connection parameters between neurons. As shown in Fig. 9(a) and Fig. 9(b), all feature maps of the third convolutional layer and the second max-pooling layer are visualized. Fig. 9 (a) is a feature map using a convolution operation with a 3 × 3 convolution kernel, and Fig. 9 (c) indicates a feature map that uses a 5 × 5 convolution kernel to carry out the convolution operation. It can be seen that the output size of the feature map depends on the stride of the move and the size of the convolution kernel. In this paper, in addition to the stride of the first convolution layer and the first max-pooling layer being set to 2, the stride for the other convolution layers and pooling layer of the DCNN structure is 1. The step size of the first convolution layer and max-pooling layer of the DCNN structure used in this paper is 2, while the other convolution layer and the pool level step size is 1. Therefore, to maintain consistency of the output feature maps in the convolution layer, all of the convolution layers with a 5 × 5 kernel need to be filled with the 2 × 2 size of padding.

The parameters of each layer are shown in Table 1, including weight and bias parameters. As seen from Table 1, in this paper, all convolution layers follow *Relu*, which improve nonlinearity and introduce sparseness for the network. In addition, the effect of the pooling operation is to reduce the influence of the resolution of the feature map and the precise location of the feature map to prevent the network from overfitting and influencing the effect of the network recognition. The pooling effect is shown in Fig. 9(b). The output of the pooling layer is formally defined as manifest in formula (3):

$$y_j^l = Relu(\beta_j^l pooling(x_j^{l-1}) + b_j^l) \tag{3}$$

where *pooling*() represents a max or average pooling function. β is a multiplicative bias (Lin et al., 2013).

Since the full connection layer occupies 80% of the overall network layer parameters and the number of neuron connections, the global average pooling layer can significantly reduce network parameters to prevent network overfitting and improve network performance. Therefore, the full connection layer of the DCNN model adopted in this paper is replaced by the global average pooling, and the multi-dimensional features become a one-dimensional vector through the global average pooling layer. Finally, Softmax regression(Hinton and Salakhutdinov, 2009) is employed as the output layer of the DCNN for classification,and its function is to output features that are highly abstracted through the hidden layer as probability values that describe the classification to which the original input image belongs.

Because the image of the multi-channel waveform is simpler than the natural image, the deeper level of the neural network will increase the burden of computation and have little impact on the accuracy of the model. Meanwhile, a shallow network would not contribute to extracting features of the image effectively, which makes the recognition accuracy rate decrease, thereby reducing the generalization ability of the network applications. Therefore, the deep convolutional neural network structure shown in Fig. 7 is designed to train a large amount of

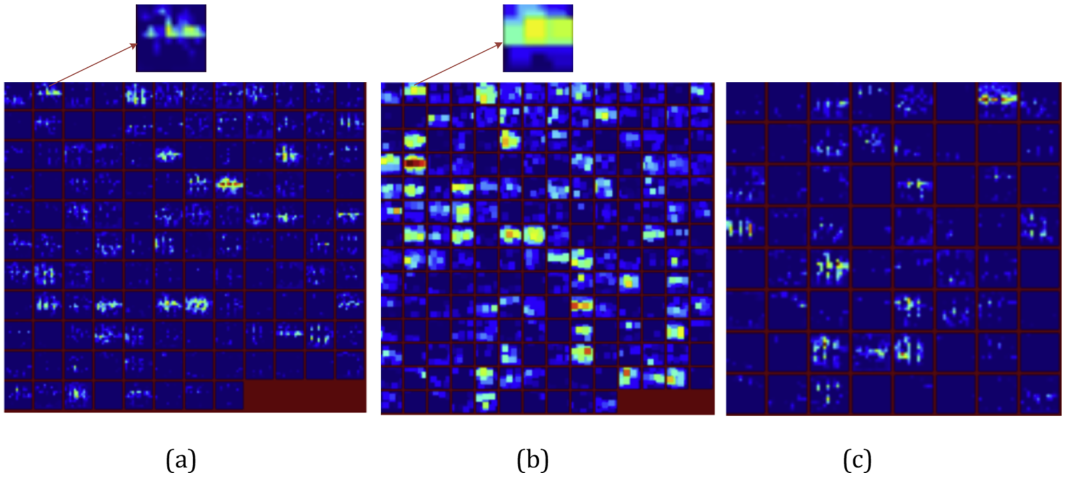


Fig. 9. Visualization of features of convolutional and pooling layers. (a) The convolutional layer of the 3 × 3 kernel; (b) The pooling layer; (c) The convolutional layer of the 5 × 5 kernel.

Table 1
Parameter of each layer. Conv2_1, Conv2_2 denote the second convolution layer of the 3 × 3 kernel and 5 × 5 kernel, respectively.

Layer	Output	Kernel/stride	Pad	Parameters
Conv1	115 × 115 × 96	7 × 7/2	0	14.2k
Maxp	57 × 57 × 96	3 × 3/2	0	
Conv2_1	55 × 55 × 128	3 × 3/1	0	110.7k
Conv2_2	55 × 55 × 64	5 × 5/1	1	153.7k
Conv3_1	53 × 53 × 192	3 × 3/1	0	147.6k
Conv3_2	53 × 53 × 96	5 × 5/1	1	153.8k
Maxp	26 × 26 × 288	3 × 3/2	0	
Conv4_1	24 × 24 × 256	3 × 3/1	0	663.8k
Conv4_2	24 × 24 × 128	5 × 5/1	1	921.7k
Conv5_1	22 × 22 × 256	3 × 3/1	1	590k
Conv5_2	22 × 22 × 128	5 × 5/1	2	409.7k
Maxp	11 × 11 × 384	3 × 3/2	0	
Conv6_1	9 × 9 × 128	3 × 3/1	0	442.5k
Conv6_2	9 × 9 × 64	5 × 5/1	1	614.5k
Conv7_1	7 × 7 × 64	3 × 3/1	0	73.8k
Conv7_2	7 × 7 × 32	5 × 5/1	1	51.2k
Avrp	1 × 1 × 96	7 × 7/1	0	
Linear	1 × 1 × 3			0.288k

waveform image data, and the optimal weight model named DCNN-Model is selected to extract features of the multi-channel waveform.

3.2. Recognition and classification of the multi-channel waveform

Microseismic signal detection instruments often bury multiple sensors underground, and an event often triggers multiple sensors. Because of the location of each sensor, the depth of the embedding is different, the number of triggered sensors is different, and the signal is often disturbed by all kinds of noise in the process of transmission, which leads to the same signal source as well as different signals received by each sensor. Therefore, judging an event requires a comprehensive analysis of each channel waveform (sensor receiving waveform). However, the DCNN network model can only classify single waveform images and can not satisfy the classification and recognition of single event multi channel waveform images.

To solve these problems, this paper proposes using DCNN and SVM to classify and identify microseismic events. First, the optimal weight model (DCNN-Model) is selected as a tool for extracting multi-channel waveform characteristics through the DCNN model structure. Second, the feature of the last global pooling layer of each channel waveform is extracted by DCNN-Model for the single event multi-channel waveform image. Then, all the extracted waveform features are sorted according to the order of each channel waveform input to the DCNN model in the event. Finally, the combined multi-channel waveform image features are used as the input data of SVM, and then the final feature classification is carried out by SVM. The whole process of joint classification and recognition of microseismic events by DCNN and SVM is shown in Fig. 10.

SVM is a type of machine learning proposed by Vapnik (Cortes and Vapnik, 1995) in the 1990s. The core idea of SVM is to map the sample space in the input to a high-dimensional space and obtain the optimal classification hyper plane in the high dimension space. The goal of SVM is to find a hyper plane that minimizes the average loss in the training data, so the following optimization problems can be derived:

$$\varphi(w, \xi) = \frac{1}{2} w^{T} \cdot \mathrm{w} + \mathrm{C} \sum_{i=1}^{N} \xi_i \tag{4}$$

where ξ, w, c respectively represent the relaxation variable, weight vector and punishment coefficient, which denotes the punishment degree of SVM. Applying the Lagrange multiplier method to solve the hyper plane of the optimal classification, it can be translated into the following constrained optimization problem:

$$Q(a) = \sum_{i=1}^{N} \mathrm{a} - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j y_i y_j K(x_i, x_j) \tag{5}$$

Constrained optimization is as follows:

$$\sum_{i=1}^{N} a_i y_i = 0,\ a_i \geq 0,\ i = 1,2,\ldots,N \tag{6}$$

where $\{a_i\}_{i=1}^{N}$ represent the Lagrange multiplier, the value of most a_i was 0, and the sample corresponding to whose value is not 0 what is called support vector. $K(x_i, x_j)$ denotes a kernel function which meets the Mercer 'Theorem. There are commonly 4 types of kernel function, including linear kernel (Linear), polynomial kernel (Polynomial), Sigmoid kernel and Gauss radial basis kernel (RBF). In this paper, we selected RBF, and the formula is as follows:

$$K(x_i, x_j) = \exp(-\frac{x_i - x_j^2}{2\sigma^2}) \tag{7}$$

where σ is a free parameter, and is determined as follows:

$$gamma = \frac{1}{2\sigma^2} \tag{8}$$

We select the best parameter c and by experiment, which can optimize the SVM model.

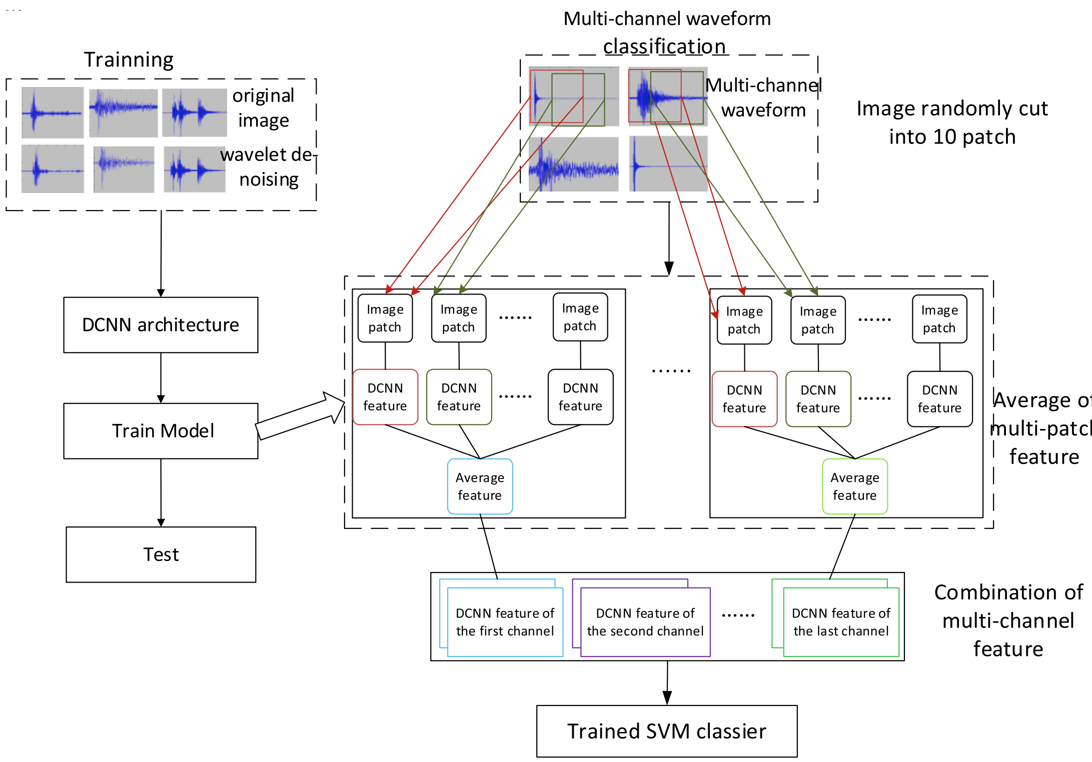


Fig. 10. Multi-channel waveform classification design.

Fig. 11. Visualization of a sample of the classifier.

It is concluded that the number of sensors triggered by an event is 5–11 (11 is the total number of sensors) by analyzing the monitoring data of the Dongguashan copper mine. Since the number of channels of a single event is different, the final feature dimension of a single event multi-channel is $1 \times 1 \times 96xN$ (N is the number of channels), resulting in inconsistent total combined feature dimensions of micro-seismic events of different channel numbers; thus, such features cannot be used as input data for the SVM model. Therefore, multiple different SVM models need to be trained for different sensor channel numbers. In this paper, we took the 7 channel waveform as an example. The forward calculation of the DCNN model is performed for each waveform to extract the last global average pooling layer features of the waveform image, and then the features extracted by each channel waveform are arranged and combined, which is used as the input data for training and testing the SVM classifier. The dimensions of the 7-channel waveform feature vectors extracted by the DCNN model are $1 \times 1 \times 96 \times 7$; the feature dimension is then converted to 7×96. Finally, the features of each multi-channel waveform are transformed into 1×672 dimensions. To see the representation of the feature maps clearly, we visualized the feature maps for the 7 channel waveform, as shown in Fig. 11. Each row matrix represents a feature map corresponding to a channel.

4. Experiment and analysis

4.1. Experimental platform and data processing

The hardware platform of the experiment is GTX TITAN X GPU GeForce. In addition, the software platform is Caffe(Jia et al., 2014), which is a deep learning framework. To construct a multi-channel microseismic event recognition and classification model, this paper analyzes the data collected by the IMS microseismic monitoring system of the Dongguashan Copper Mine to construct the DCNN model and SVM training and test dataset. The data set is divided into two classes: the first data set consists of a single waveform image, including micro-seismic images, blast images and noise images, each of which has 10,000 samples, for a total of 30,000 samples used to train and test the DCNN model; the second data set consists of single event 7-channel waveform images, each of which contains 7 waveform images, including blast events, microseismic events, and noise events. Each event has 1320 samples for a total of 3960 event samples used to train and test SVM classifiers. Further, the waveform image of the first data set is marked; 0 represents a microseismic image, 1 represents a blast image, and 2 represents a noise image. Finally, the data in the first data set and the second data set are put into the training set and the test set with a ratio of 3:1.

The first data set contains wavelet threshold denoising and filtered images. Wavelet threshold denoising not only can double the number of original images, to ensure that the DCNN model has enough data for training and testing, but can also obtain high-quality waveform images, which can greatly suppress noise interference. The impact makes the effective information input into the DCNN model larger than the useless information, and the learned features are more convenient to identify and classify. At the same time, the original image without wavelet threshold denoising and filtering in the first data set can prevent the effective information loss of the filtered image to solve the incomplete problem of the DCNN model learning feature.

To increase the randomness of the sample, improve the generalization ability of the DCNN network model, and make the feature extraction more accurate, the denoised processing and the labeled waveform image are randomly divided into 10 graphs. As shown in Fig. 10, we do not perform forward calculation directly on the image, but perform forward calculation on the 10 cropped images formed from the original image to extract features, and obtain the feature mean of 10 cropped images extracted by the DCNN model. To avoid the adverse effects caused by the migration of images, In this paper, the image size of $256 \times 256 \times 3$, is randomly cut into 10 cropped images of size $235 \times 235 \times 3$. Among them, 256×256 represents the length and width of the image, and 3 represents the number of channels of the image.

In this paper, the maximum number of iterations of the DCNN network is set to 9000 times, the learning rate in the training progress is set to 0.0001, and the momentum factor u is set to 0.9 to make weight updates smooth, stable, and fast. In addition, all training sample values are subtracted from their mean values in order to improve and speed up convergence.

4.2. Experimental results and comparative analysis of DCNN network structure construction

To prove the good recognition and classification ability of the DCNN network structure, we have performed a series of comparative experiments.

1) A deep convolutional neural network is constructed with a 3×3 convolution kernel, That is, except for the 1st layer using the 7×7 convolution kernel, the remaining convolutional layers all use the 3×3 convolution kernel, which is recorded as the DCNN-3 model. Parameters such as the number and size of the image features of each layer of the DCNN-3 model are the same as the DCNN model; see Table 1.
2) All 3×3 convolution kernels of the DCNN-3 model are replaced with 5×5 convolution kernels, which are recorded as DCNN-5 models. To ensure the same size as the DCNN output image features, all 5×5 convolution kernels are filled with the 2×2 size. The other parameters of the DCNN-5 model are the same as the DCNN model parameters; see Table 1.
3) The seventh convolutional layer in the hidden layer of the DCNN model is removed, and the kernel size of the global average pooling is changed to 9×9, which is recorded as the DCNN-reduce model. The DCNN-reduce model is equivalent to reducing the convolutional layer in the DCNN model, and the other parameters are the same as the DCNN model; see Table 1.
4) After the seventh convolutional layer in the hidden layer of the DCNN model, the 1 maxpool layer and the 2 convolutional layers are sequentially added, and the global pooling layer is removed, which is recorded as the DCNN-increase model. The 3×3 convolution kernel image feature output size of the DCNN-increase model is 64, and the 5×5 convolution kernel feature map output size is 32, so the dimension of the last convolutional layer has become $1 \times 1 \times 96$. The other parameters of the DCNN-increase model are the same as the DCNN model; see Table 1.

The DCNN model, DCNN-increase model, DCNN-reduce model, DCNN-3 model and DCNN-5 model were tested by experiments. The relationship between the accuracy of the test and the number of iterations is shown in Fig. 12. The relationship between the loss rate and the number of iterations of the training is shown in Fig. 13. The relationship between the loss rate and the number of iterations is shown in Fig. 14.

It can be seen from Fig. 12 that with respect to the classification of a single waveform image, the DCNN model has a maximum accuracy of 94.13% compared with other models, followed by the DCNN-increase model and the DCNN-5 model. It can be seen that the DCNN model has better expression ability, which combines the characteristics of 3×3 and 5×5 convolution kernels. First, the larger receptive field can make more use of the regional information of the image context, and has stronger expression ability; second,the use of a smaller convolution kernel can reduce the number of weight connections to reduce computation and improve the efficiency of the model.

Fig. 13 shows the fitting effect of the network on the training data, and Fig. 14 shows the fitting effect of the network on the test data to show the generalization ability of the model. By analyzing Figs. 12 and 13, the training loss of the DCNN-increase model is the lowest, and the model fitting training data are the best, but the loss rate and accuracy of the test are higher than the DCNN model, meaning that the deeper the network level, the better the fitting of the training data, but the ability of the model to migrate to other datasets is lacking and there is a risk of overfitting. By analyzing Figs. 13 and 14, compared with the DCNN-3

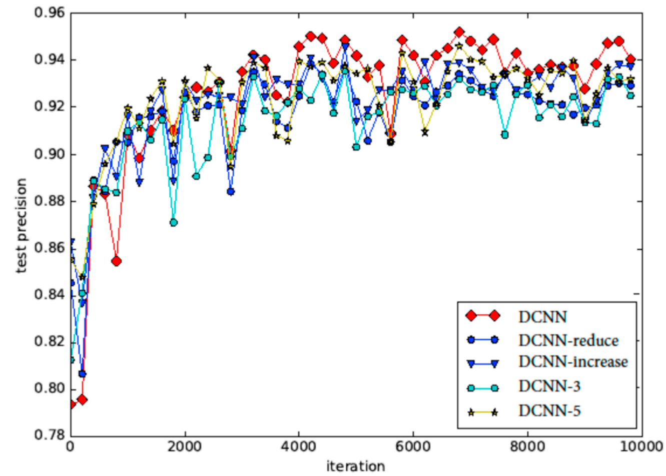


Fig. 12. Relationship between test accuracy and number of iterations for each model.

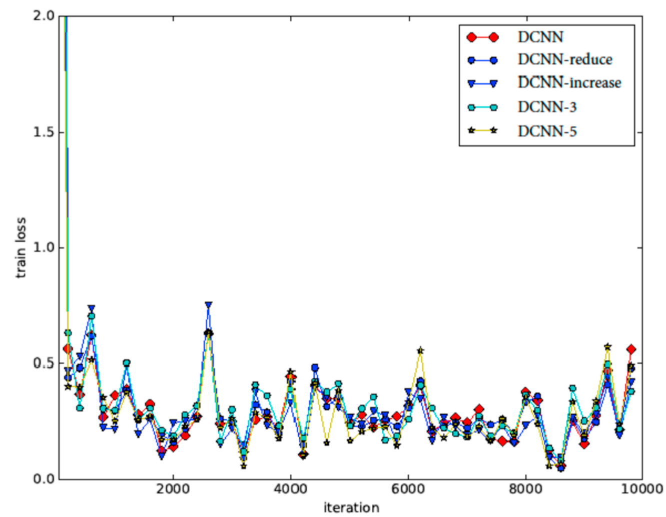


Fig. 13. Relationship between training loss rate and the number of iterations of each model.

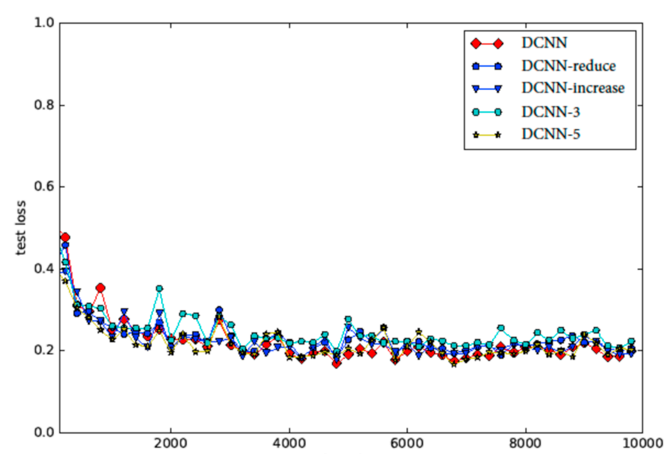


Fig. 14. Relationship between test loss rate and number of iterations for each model.

model, the DCNN-5 model can obtain enough receptive fields to maintain good network recognition.

In summary, the following conclusions can be drawn: (1)the deeper the network level, the better its expression ability, but the network that is too deep is prone to over-fitting, and the deep network connection parameters are relatively large; (2) the 5 × 5 convolution kernel has a slightly stronger expression than the 3 × 3 convolution kernel, but the large convolution kernel has a larger weight parameter. Therefore, this paper adopts the DCNN model constructed by 3 × 3 and 5 × 5 convolution kernels and 13-layer network, which not only enhances the expressive power of the convolutional neural network and reduces the parameters of the network but also ensures accurate extraction of the features of the microseismic waveforms and prevents the risk of over-fitting on the network.

To better demonstrate the relationship between the test accuracy and loss rate in the DCNN model test, both are displayed on Figs. 12 and 13. It can be seen from Fig. 12 that the overall trend of the test accuracy of the DCNN network model is first increased rapidly, then the speed is slowed down, and finally, reaches a steady state. The corresponding loss rate is such that the overall trend first drops rapidly, then slowly declines, and finally, it reaches a steady state. Therefore, its classification accuracy, stability and convergence speed can better meet the needs of the joint recognition and classification of DCNN and SVM.

4.3. Automatic recognition and classification results and analysis of microseismic waveforms

4.3.1. Automatic recognition and classification of single event multi-channel waveform images

Since the number of sensors triggered by each event collection and data collection is inconsistent, it is difficult to use the DCNN model for end-to-end training. Therefore, this paper proposes using the classifier to classify the high-level features extracted by DCNN. In the classification experiment of the multi-channel waveform, we will randomly crop each image (256 × 256 x 3) into 10 patches, and the size of each patch is 235 × 235 x 3. The features of each image patch are extracted by the DCNN model, and then the feature values of the 10 image patches are averaged. According to the above method, the average value of all waveform image features in 7 channels of a single event is calculated as the input data of the SVM classifier. To improve the classification effect of SVM classifier, this paper searches the optimal parameters (the penalty factor C of SVM and the radius G of the kernel function) through the grid parameter optimization method (grid search method) (Huang et al., 2007), where C and G are meshed within a certain range, and all points in the grid are traversed, and the optimal values are obtained by cross-validating the determined C and G.

The SVM classification accuracy and the experimental results of the grid search method for selecting the optimal parameters are shown in Fig. 15. It can be seen from Fig. 15 that when the radius of the SVM kernel function and the penalty factor C are constantly changing, the classification accuracy is changed by approximately 80% first, and then the accuracy is slowly rising, and the final accuracy reaches a stable state of approximately 98%. The experimental results show that the accuracy of the event classification test of the two dataset reaches 98.18%, and the values of the optimal parameters C and G obtained by the grid search method are 77 and 195, respectively.

4.3.2. Performance comparison of each classifier

The k-Nearest Neighbor (KNN) classification algorithm is a non-parametric method for classification and is also a type of instance-based learning or lazy learning. KNN has the following characteristics (Altman, 1992): (1) the input of the KNN is the feature vector of the instance, that is, the point on the corresponding feature space, and the output is the category of the instance; (2) KNN does not have an explicit learning process, which determines the type of classification by a voting mechanism based on k nearest neighbors; (3) the three basic elements of KNN include the selection of K values, distance metrics, and classification decision rules.

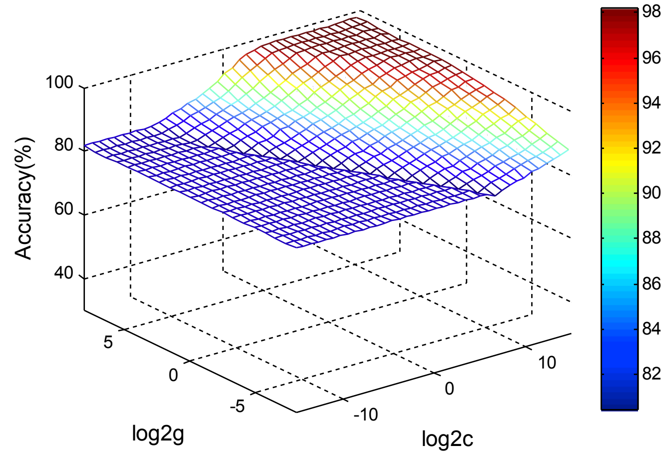


Fig. 15. SVM classification accuracy and parameter. Choice parameters of GridSearch (3D View); Best c = 77 g = 195.

Random Forests (RF) (Ho, 1995) is a statistically based combination classification algorithm that combines resampling and decision tree methods, its essence is composed of multiple Classification And Regression Tree (CART). RF uses reciprocal sampling for each tree in the training set; that is, some samples in the total training set may appear in the training set of a tree multiple times, or may never appear in the training set of a tree (Gao et al., 2009). In addition, because the algorithm integrates multiple single classifiers, it effectively improves the classification performance of the classifier and is widely used in various types of classification screening and prediction (Min et al., 2015).

To further verify the accuracy and efficiency of joint recognition and classification of single event multi-channel waveform images using DCNN model and SVM, this paper combines the DCNN network model with KNN, random forest classifier and SVM to identify the same set of waveform image data. This includes comparing the accuracy of the classification of each classifier with the time required to classify the sample. In this paper, the DCNN model is respectively combined with KNN, random forest classifier and SVM to identify and classify the same dataset of the waveform image, and the accuracy and the time required of each classifier classification are compared. As a result of several experiments, the most ideal parameter factor k of KNN is 3. The random forest can effectively improve the classification performance by increasing the number of trees, but it will greatly increase the calculation time of the classifier. The number of trees in the random forest experiment is 1000, 2000 and 3000, respectively. The experimental results are shown in Table 2.

The experiment demonstrates that the accuracy of the SVM is the best when the parameter of C is 77 and the parameter of G is 195, the accuracy rate of which reaches 98.18%. Followed by random forests, the accuracy rate can reach approximately 94%, but its calculation time is too long; it is not suitable for the practical application of mine microseismic waveform classification and recognition. While KNN has a running time similar to SVM, its accuracy is only 92.8%. According to the comprehensive consideration above, we conclude that the performance of SVM is perfect.

Table 2
Comparison with three algorithms for accuracy rate and taking times.

Classifier	Classifier parameters	Accuracy rate	times
KNN	K = 3	92.80%	3.64s
SVM	C = 77; G = 195	98.18%	3.63s
Random forest	Tree = 1000	94.17%	87.09s
	Tree = 2000	94.62%	175.80s
	Tree = 3000	94.55%	265.49s

5. Conclusion

In this paper, we propose a new method for mine micro-seismic pattern recognition and classification. This method combines DCNN with SVM for classification of the multi-channel waveform. DCNN-Model learns and recognizes the feature of multi-channel waveforms automatically, and automatically classifies them using SVM by combining the feature of all channels waveforms, and the accuracy of classification reaches 98.18%. The method does not require designing features by hand, which guarantees accuracy, realtime application and intelligence in recognition and classification. The method is of practical value in the industry.

Author contributions

BI Lin and XIE Wei conceived of the study and wrote the paper; ZHAO Junjie analysed the data and verified the algorithm; all authors discussed the results and revised the paper.

Acknowledgement

Thanks to the following organizations who provided provide basis data and technique support for this research: Tongling Nonferrous Metals Group Holding Co., Ltd (Dongguashan Copper Mine), Digital Mine Co., Ltd, Berkeley Vision and Learning Center (Deep Learning Framework). The authors also gratefully acknowledge the financial support from the National Natural Science Foundation of China (41572317) and the National Key R&D Program of China (2017YFC0602905).

References

Abdelwahed, M.F., 2012. SGRAPH (SeismoGRAPHer): seismic waveform analysis and integrated tools in seismology. Comput. Geosci. 40, 153–165.

Altman, N.S., 1992. An introduction to kernel and nearest-neighbor nonparametric regression. Am. Statistician 46, 175–185.

Bui Quang, P., Gaillard, P., Cano, Y., Ulzibat, M., 2015. Detection and classification of seismic events with progressive multi-channel correlation and hidden Markov models. Comput. Geosci. 83, 110–119.

Chavan, M.S., Mastorakis, N., Chavan, M.N., Gaikwad, M., 2011. Implementation of SYMLET wavelets to removal of Gaussian additive noise from speech signal. In: Proceedings of the 10th International Conference on Recent Researches in Communications, Automation, Signal Processing, Nanotechnology, Astronomy and Nuclear Physics, pp. 37–41.

Chen, X.-W., Lin, X., 2014. Big data deep learning: challenges and perspectives. IEEE Access 2, 514–525.

Ciresan, D.C., Meier, U., Gambardella, L.M., Schmidhuber, J.r., 2010. Deep, big, simple neural nets for handwritten digit recognition. Neural Comput. 22, 3207–3220.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Dong, L.-J., Wesseloo, J., Potvin, Y., Li, X.-B., 2016. Discriminant models of blasts and seismic events in mine seismology. Int. J. Rock Mech. Min. Sci. 86, 282–291.

Gao, D., Zhang, Y.-X., Zhao, Y.-H., 2009. Random forest algorithm for classification of multiwavelength data. Res. Astron. Astrophys. 9, 220.

Hinton, G.E., Salakhutdinov, R.R., 2009. Replicated softmax: an undirected topic model. Adv. Neural Inf. Process. Syst. 1607–1614.

Ho, T.K., 1995. Random decision forests, document analysis and recognition, 1995. In: Proceedings of the Third International Conference on. IEEE, pp. 278–282.

Huang, C.-L., Chen, M.-C., Wang, C.-J., 2007. Credit scoring with a data mining approach based on support vector machines. Expert Syst. Appl. 33, 847–856.

Ji, S., Xu, W., Yang, M., Yu, K., 2013. 3D convolutional neural networks for human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. 35, 221–231.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia. ACM, pp. 675–678.

Jin, W., Li, Z.J., Wei, L.S., Zhen, H., 2000. The improvements of BP neural network learning algorithm. In: Signal Processing Proceedings, 2000. WCCC-ICSP 2000. 5th International Conference on. IEEE, pp. 1647–1649.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 1097–1105.

Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D., 1997. Face recognition: a convolutional neural-network approach. IEEE Trans. Neural Network. 8, 98–113.

LeCun, Y., Bengio, Y., 1995. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks 3361, 1995.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Lin, M., Chen, Q., Yan, S., 2013. Network in Network. arXiv preprint arXiv:1312.4400.

Min, L., Rongling, L., Yongbin, C., 2015. Number of trees in random forest. Computers Engineering and Applications 51, 126–131.

Quan-jie, Z., Fu-xing, J., Yong-ming, Y., Zheng-xing, Y., Jing-lin, W., 2012. Classification of mine microseism events based on wavelet-fractal method and pattern recognition. Chin. J. Geotech. Eng. 34, 2036–2042.

Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Tan, J.F., Bland, H.C., Stewart, R.R., 2007. Classification of Microseismic Events from Bitumen Production at Cold Lake, Alberta. CREWES Research Report, 19. pp. 1–24.

Wang, T., Wu, D.J., Coates, A., Ng, A.Y., 2012. End-to-end text recognition with convolutional neural networks. In: Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, pp. 3304–3308.

Xuelong, L., Zhonghui, L., Enyuan, W., Linbo, J., Junjun, F., 2015. Study of mine microcosmic signals interference characteristic and its de-noising method. J. China Inst. Min. Technol. 44, 788–792.