IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Recognition of Microseismic and Blasting Signals in Mines Based on Convolutional Neural Network and Stockwell Transform

Guangdong Song[1,2], Jiulong Cheng[2] and K T V Grattan[1,3]

[1]Laser Institute, Qilu University of Technology (Shandong Academy of Sciences), Ji'nan, Shandong 250014, China
[2]State Key Laboratory of Coal Resources and Safe Mining, China University of Mining & Technology (Beijing), Beijing 100083, China
[3]School of Mathematics, Computer Science and Engineering, City, University of London, London, EC1V 0HB, United Kingdom

Corresponding author: Jiulong Cheng (e-mail: JLCheng@cumtb.edu.cn).

This work was supported by Primary Research & Development Plan of Shandong province (Grant no. 2018GSF120008), and the National Key Research and Development Program of China (Grant no. 2018YFC0807804). Grattan acknowledges support from the Royal Academy of Engineering in the United Kingdom.

ABSTRACT The microseismic monitoring signals which need to be determined in mines include those caused by both rock bursts and by blasting. The blasting signals must be separated from the microseismic signals in order to extract the information needed for the correct location of the source and for determining the blast mechanism. The use of a convolutional neural network (CNN) is a viable approach to extract these blast characteristic parameters automatically and to achieve the accuracy needed in the signal recognition. The Stockwell Transform (or S-Transform) has excellent two-dimensional time-frequency characteristics and thus to obtain the microseismic signal and blasting vibration signal separately, the microseismic signal has been converted in this work into a two-dimensional image format by use of the S-Transform, following which it is recognized by using the CNN. The sample data given in this paper are used for model training, where the training sample is an image containing three RGB color channels. The training time can be decreased by means of reducing the picture size and thus reducing the number of training steps used. The optimal combination of parameters can then be obtained after continuously updating the training parameters. When the image size is 180 × 140 pixels, it has been shown that the test accuracy can reach 96.15% and that it is feasible to classify separately the blasting signal and the microseismic signal based on using the S-Transform and the CNN model architecture, where the training parameters were designed by synthesizing LeNet-5 and AlexNet.

INDEX TERMS Convolutional Neural Network (CNN), Stockwell Transform (S-Transform), Microseismic and Blasting Signal, Signal Recognition

I. INTRODUCTION

The technology of Microseismic monitoring in mines is very important, as an effective means for forecasting and creating the early warning of a rock burst disaster is needed. To create an effective early warning of such a rock burst, it is important to classify accurately both the microseismic and the blasting signals obtained from mines accurately [1-3]. Since the 1960s, seismic researchers have carried out automatic identification of seismic signals and blasting signals, using the same approach as for nuclear explosion signals [4-6]. In the 1990s, Musil and Pleginger utilized a multi-layer perceptron model of an artificial neural networks to identify microseismic signals and stope blasting signals. Using a multi-layer perceptron approach, 20% of the previous misclassifications could then be classified correctly by using self-organizing feature maps [7]. On the assumption that the microseismic and blasting signals are linearly separable, Dong et al. adopted the Fisher discriminant method to realize the signal classification needed, where the accuracy achievable from 50 groups of test samples was shown to reach 94% [8]. In his work, Li [9] used an approach in which he decomposed the signal adaptively, based on local mean decomposition, and obtained the feature vector in the pattern recognition. The classification accuracy rate of the Support

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2978392, IEEE Access

Vector Machine (SVM), based on the knowledge of the spectral coefficients of the local mean decomposition principal components, reached 93%. Zhao et al. [10] studied the probability density distribution of the signal characteristic parameters, where the highest accuracy of 97.1% could be achieved by artificially setting the signal characteristic parameters. In the above recognition methods, it is necessary to prepare the feature parameters in advance, but these are only a small part of all the features of the original signals seen and do not make full use of all the features available from the original signal. Further, Vallejos et al. have proposed a method for identifying microseismic and blasting signals based on logistic regression and a neural network, achieving an accuracy rate of more than 95%. However, the drawback in that approach is that it is an arduous task to apply the model to different mines [11].

With the development both of the internet and computing power, the acquisition and analysis of massive data sets is now feasible, which further encourages the rapid use of neural network technology in such applications. A traditional neural network contains a simplistic input layer, a hidden layer, and an output layer. Manual feature extraction is needed, following which weight learning is carried out to obtain the results needed and to predict the outcomes. Feature extraction is an extremely complex task, and it is sometimes impossible to extract the features required manually and using a multitude of complex objects represents an arduous approach to extracting features directly [12-15].

To solve the above problems, Hinton et al. have proposed the concept of deep learning, based on multiple hidden layers [16]. Deep learning originates from neural network research, where it is a neural network with a higher level of nonlinear operation in the function derived from the neural network learning and contains the structure of multiple hidden layers [17]. A large number of training data sets are utilized in deep learning to achieve feature learning. There is a connection between the adjacent layers, while there is no connection between the nodes of the cross-layer and between the same layers. A neural network with multiple hidden layers has excellent feature learning abilities [18]. The convolutional neural network (CNN), a quintessential algorithm in deep learning, is an improved algorithm approach on the traditional neural network, and which has achieved significant success in the field of image processing. It is a feed-forward neural network derived from the concept of the receptive field, proposed by Hubel and Wiesel in the 1960s. CNN uses spatial structure relationships to reduce parameters, allowing the efficiency of parameter training to be improved. It is not required manually to extract features and it achieves automatic abstraction and extraction of the features in training. It reduces the difficulty seen in image recognition and improves the recognition accuracy. The structure of the CNN is similar to the spatial structure of images, where they are all two-dimensional plane structures [19].

A large number of methods or models based on CNN and associated images for various purposes have been proposed in recent years. Shao et al. proposed a remote sensing image fusion method to generate remote sensing images at both high spatial and spectral resolution based on the CNN. The spatial and spectral features were respectively extracted from the multispectral and panchromatic images by convolutional layers with different depths. Then the extracted features were utilized to yield fused images. By evaluating the performance on the QuickBird and Gaofen-1 images, the method provided better results compared with other classical methods [20]. A saliency-aware CNN model for real-time detection of inshore ships was proposed by Shao et al. This model used CNN to predict the category and the position of ships, and used the global contrast based salient region detection to correct the location. The experimental results showed that the model outperforms representative counterparts in terms of accuracy and speed [21].

This paper focuses on the existing problems of the mixture of mine microseismic signal and blasting signal, analyses the characteristics of the time-domain waveform of the monitoring signals, studies the CNN algorithm and the concept and algorithm implementation of the Stockwell Transform (S-transform), and puts forward a recognition method for mine microseismic and blasting signals. The original blasting and microseismic time-domain signals are changed to the time-frequency domain by use of the S-Transform, and a two-dimensional time-frequency image is obtained and trained by using CNN technology. The model architecture for the classification of the microseismic signals and the blasting signals has been designed. The training parameters of the model, such as the training steps and image size are changed and used with different combinations for training. Finally, a special combination with the highest prediction accuracy is selected from the model training parameters.

II. RECOGNITION ALGORITHM FOR MICROSEISMIC AND BLASTING SIGNALS

A. WAVEFORM OF THE MICROSEISMIC SIGNAL AND THE BLASTING SIGNAL IN THE TIME-DOMAIN

The energy at the explosion center in a mine is released rapidly and propagates mainly in the form of a shock pressure wave. The surrounding rock mass is affected by the compression elastic wave, which is the source of the expansion wave and propagates mainly in the form of a longitudinal wave. Therefore, in the time-domain waveform diagram (Figure 1), the time associated with the amplitude peak of the blasting signal is very close to the first arrival time, and it can be seen that the energy decays rapidly. The diagram shows that the maximum energy of the signal only lasted for 1 cycle and from the time-domain waveform (Figure 2), it can be seen that the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2978392, IEEE Access

amplitude of the signal increases gradually after the first arrival time and takes a long time to reach its maximum value.

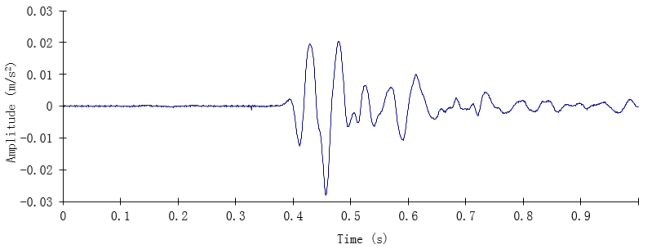

FIGURE 1. Time-domain view of the waveform of the blasting signal

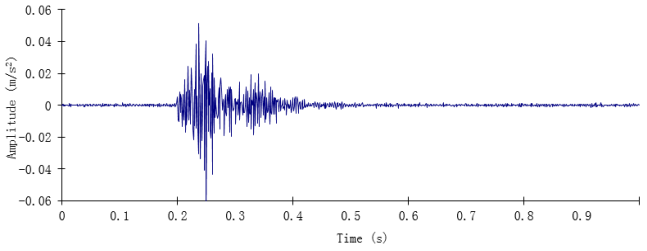

FIGURE 2. Time-domain view of the waveform of the microseismic signal

From these two figures it can be seen that the two monitored signals are obviously different near the first arrival time, and therefore the time-domain waveform may be utilized to classify the signals directly. Nevertheless, the frequency-domain characteristics of the signals will be neglected, which will result in unrecognizable signals or large recognition errors, if the time-domain signals are used directly for signal recognition. The time-domain waveform signal cannot express the frequency-domain characteristics needed and the time-domain signal can be displayed synchronously after using the S-Transform approach. Therefore, it is necessary to display both the time-domain and the frequency-domain information on the same picture, and then use the signal recognition algorithm to understand the nature of the signals. Experiments are carried out to verify whether the time-domain signal or the time-frequency synchronous signal can be used to allow the recognition that is needed. The CNN method is an efficient image classification approach without manual feature parameter extraction when compared to the classification algorithm using manual feature parameter extraction. The CNN approach makes use of translation invariance: that is it still can produce the same features as before after a small translation of the image and can demonstrate the visual content concisely and effectively. In this paper, CNN is utilized to recognize both the microseismic signals of rock mass fracture and the blasting signals. The original time-domain signal and the time-frequency signal obtained after the S-Transformation is applied are respectively compared and verified to select the data types that can meet the requirements of high-precision recognition. The CNN and S-Transform algorithms used are discussed respectively below.

B. ALGORITHM – CNN

The error back-propagation (BP) algorithm is used to update the weights used with the CNN approach. In a way that is similar to the use of the back-propagation algorithm, the CNN approach uses forward propagation to calculate the output value, and the BP to adjust the weight and bias. Unlike all neurons that are completely connected in the BP algorithm, neurons between the adjacent layers of CNN are partly connected [22-23]. A simple CNN architecture can be described as including: the convolution layer, the nonlinear transformation layer, the pooling layer and the fully connected layer. A well-designed architecture highlights the crucial feature: information – and ignores the noise.

The convolution layer is the core layer of the CNN. Generally, the input nodes are multiplied in a 3 × 3 or a 5 × 5 way at the convolution kernel, and the bias term is added to obtain a nodal value of the next layer. The local eigenvalue of the next layer is more abstract. At the same time, the depth of the node matrix is increased and a brand new 2D image can then be obtained. The convolution kernel, also known as a filter, can transform the node matrix of the upper layer into the unit node matrix of the next layer, in which length and width are unity and the depth is unlimited.

Unlike the full connection approach, the neuron node of a feature map in the convolution layer only connects to one node of the corresponding feature map in the pooling layer, which uses a 1-to-1 non-overlapping sampling. Equation 1 indicates that to obtain updates for each neuron weight, the residual δ^l should be required. To calculate the residual of the convolution layer, it is essential to calculate the corresponding residual of the pooling layer, so that the residual map and the feature map of the convolution layer are of the same size. The partial derivative of the activation value of the feature map of the l layer is multiplied by the residual map of the pooling layer element-by-element [24] and so:

$$\delta_j^l = \beta_j^{l+1}(f'(u_j^l)^{\circ} up(\delta_j^{l+1})) \tag{1}$$

where δ_j^l is the residual or sensitivity value of the jth feature map in the l layer, 'up' means to extend the residual tensor of layer l+1 to the same size as layer l, ° indicates dot multiplication and β_j^{l+1} is a multiplicative bias of layer l+1. Thus:

$$\frac{\partial E}{\partial b_j} = \sum_{u,v}(\delta_j^l)_{uv} \tag{2}$$

$$\frac{\partial E}{\partial k_{ij}^l} = \sum_{u,v}(\delta_j^l)_{uv}(p_i^{l-1})_{uv} \tag{3}$$

where k_{ij}^l is the convolution kernel of the jth feature map of layer l and it is connected to the ith map of layer l-1. u and v are the position coordinates of the output convolution feature

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2978392, IEEE Access

map. $\left(p_i^{l-1}\right)_{uv}$ is the result of the convolution of x_i^{l-1} and k_{ij}^l.

The output obtained after filtering by use of the convolution kernel needs to be processed by a nonlinear activation function. The commonly used activation function is the rectified linear unit (ReLU) [19].

The pooling layer reduces the size of the node matrix (and the parameters of the whole neural network) by reducing the resolution. The forward propagation algorithm [19] of the pooling layer is given by:

$$x_j^l = f(\beta_j^l down(x_j^{l-1}) + b_j^l) \tag{4}$$

where $down(x_j^{l-1})$ illustrates the fact that the pooling layer resamples the convolution input of the previous layer. The maximum pooling sampling is used to find the maximum among all the pixels of the n × n block of the input feature map; consequently, the output feature map of the pooling layer is reduced n times in each of two-dimensions. β_j^l operates on a multiplicative bias, b_j^l on an additive bias. Reducing the 3 × 3 images to 1 × 1, maximum pooling is generally utilized, which can improve the model distortion tolerance property.

The use of the convolution layer and the pooling layer allow the automatic extraction of image features. Finally, one or two fully connected layers are needed to create the final image classification.

C. THE CONCEPT AND ALGORITHM IMPLEMENTATION OF S-TRANSFORM

Because the time-domain signals obtained by using the microseismic monitoring system do not contain frequency-domain information, if the original time-domain signals are directly identified and classified by use of the CNN, the frequency-domain characteristics of the signals will be ignored, which will inevitably create a negative impact on the classification accuracy.

A Fourier Transform approach can only map the signal from the time-domain to a one-dimensional frequency domain, and cannot analyze the change of the signal frequency in the time-domain, nor locate the time and frequency simultaneously [25]. A short-time Fourier transform (STFT) can only analyze the time-frequency with one dimensional resolution, while the wavelet transform cannot directly correspond to frequency. Stockwell et al. put forward the S-Transformation method [26] in 1996. The S-Transform is an extension and phase correction of the continuous wavelet transform. It combines the advantages of the short-time Fourier transform and the wavelet transform and avoids their shortcomings. The Gauss window function is utilized in the S-Transform, the drawback of a fixed window width is improved upon and the window function is not needed. It has outstanding time-frequency characteristics [27] which suit this application well.

In view of the superior time-frequency characteristics of the S-Transform, the original microseismic data are transformed first by the algorithm to obtain a two-dimensional time-frequency image, followed by classifying them by use of the CNN algorithm to achieve a high accuracy in the microseismic signal and the blasting signal classification. The two-dimensional time-frequency image of the microseismic signal after the S-Transformation is applied is shown in Figure 3.

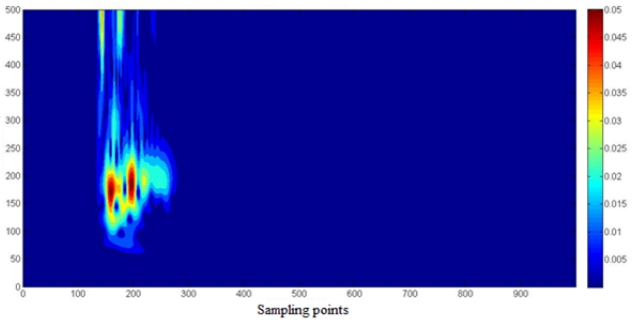

FIGURE 3. Two-dimensional time-frequency image of microseismic signal showing the sampling points along the x axis and frequency value along the y axis.

The standard definition of the S-Transformation is:

$$S(\tau, f) = \int_{-\infty}^{\infty} h(t) \frac{|f|}{\sqrt{2\pi}} e^{-\frac{(\tau-t)^2 f^2}{2}} e^{-i2\pi f t} dt \tag{5}$$

where t is time, i is an imaginary unit, $S(\tau,f)$ is the S-Transform of the time function of $h(t)$, f is the frequency, d is the width of the window and its value is the reciprocal of f. The value of f affects the resolution of the time-frequency spectrum of the S-Transform. τ is a translation factor determining the position of W on the time axis.

The inverse transformation of the S-Transform is given by:

$$h(t) = \int_{-\infty}^{+\infty} [\int_{-\infty}^{+\infty} S(\tau, f) d\tau] e^{i2\pi f t} df \tag{6}$$

where $h[kT]$ is the discrete series of $h(t)$, T is the sampling interval，k=0,1,2,…,N-1，the Fourier transform of $h[kT]$ is:

$$H\left[\frac{n}{NT}\right] = \frac{1}{N} \sum_{k=0}^{N-1} h[kT] e^{-\frac{i2\pi nk}{N}} \tag{7}$$

and n is the sampling number, n=0,1,2,…, N-1.

The reciprocal correlation between the local Gauss window width and the frequency represents a great improvement when compared with the use of the STFT with fixed-width windows. The phase of the S-Transform refers to the the starting time and provides useful and supplementary information about the spectrum, which is not available in the local reference phase information of the continuous wavelet transform (CWT) [26].

III. MODEL ARCHITECTURE DESIGN OF THE CNN AND SAMPLE DATA ACQUISITION

A. MODEL ARCHITECTURE DESIGN OF THE CNN

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2978392, IEEE Access

LeCun et al. [28] put forward the LeNet-5 model, which was successfully applied to image recognition in 1998. It has seven layers and can achieve a 99.2% accuracy rate when applied to digital recognition. The model architecture is as follows: an input layer - convolution layer 1 - pooling layer 1 - convolution layer 2 - pooling layer 2 - fully connected layer 1 - fully connected layer 2 - output layer. Nonetheless, LeNet-5 cannot solve the relatively large image dataset very well.

Krizhevsky et al. [29] then put forward the AlexNet model in 2012, this being a deeper and wider version of LeNet-5. The error rate has been greatly reduced, and its dominance in machine vision has been established. The model architecture consists of five convolution layers, three pooling layers, and three fully connected layers. The Local response normalization (LRN) is added to the pooling layer 1 and 2, and the ReLU activation function is added to the convolution layer and the fully connected layer. The dropout function is added into the fully connected layer to realize a nonlinear function transformation of the input data, where quite a few neurons are randomly ignored to avoid over-fitting. The model architecture of the AlexNet is: input layer - convolution layer 1 (ReLU) - pooling layer 1 (LRN) - convolution layer 2 (ReLU) - pooling layer 2 (LRN) - convolution layer 3 (ReLU) - convolution layer 4 (ReLU) - convolution layer 5 (ReLU) - pooling layer 3 (Dropout) - fully connected layer 2 (Dropout) - output layer.

The deeper neural network can be designed by repeatedly stacking the convolution layer and the pooling layer. However, its corresponding features are increasing, the parameters are increasing significantly and the search space is increased, so that the computational complexity is improved.

The model architecture for the classification of the microseismic signal and the blasting signal is designed based on the combined advantages of LeNet-5 and AlexNet. The model architecture of the two-dimensional time-frequency image by use of the S-Transform is: input layer - convolution layer 1 (ReLU, batch standardization) - pooling layer 1 (LRN) - convolution layer 2 (ReLU, batch standardization) - pooling layer 2 (LRN) - fully connected layer 1 (ReLU, Dropout) - fully connected layer 2 (ReLU, Dropout) - SOFTMAX - Classification result.

B. PROCEDURE OF THE RECOGNITION METHOD

The procedure used for the recognition method for the mine microseismic and the blasting signals is shown in Figure 4.

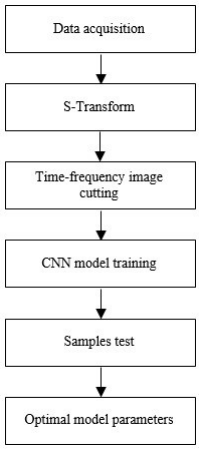

FIGURE 4. Procedure figure of the recognition method

The procedure can be described as follows: 1) the original blasting and microseismic time-domain signals are obtained in mines; 2) then the original samples are changed to the time-frequency domain images by use of the S-Transform; 3) the white edge of the new images are cut; 4) the two-dimensional time-frequency images are obtained and trained by using CNN technology. The training parameters of the model, such as the training steps and image size are changed and used with different combinations for the training; 5) the test results with different parameters are obtained; and 6) finally, an optimal combination with the highest prediction accuracy is selected from the model training parameters.

C. SAMPLE DATA ACQUISITION

The data obtained are divided into two types: the blasting signal and the microseismic signal. The sample data come from the microseismic monitoring data obtained in the coal mine. In order to monitor the situation dynamically, thus allowing for monitoring of a potential disaster that could occur in the coal mine in real-time, eight microseismic sensors were placed and arranged in the working coalface, with the sensor arrangement as shown in Figure 5. Sensors labelled s6, s7, and s8 are arranged in the upper roadway, with the other five sensors s1 to s5 are arranged in the lower roadway: respectively in the roof, coal seam, and floor. A total of 116 blasting samples were obtained in one month, 90 of which were taken as blasting training samples and 26 as blasting test samples. 116 microseismic samples were also obtained, 90 of which were taken as microseismic training samples and 26 as microseismic test samples.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2978392, IEEE Access

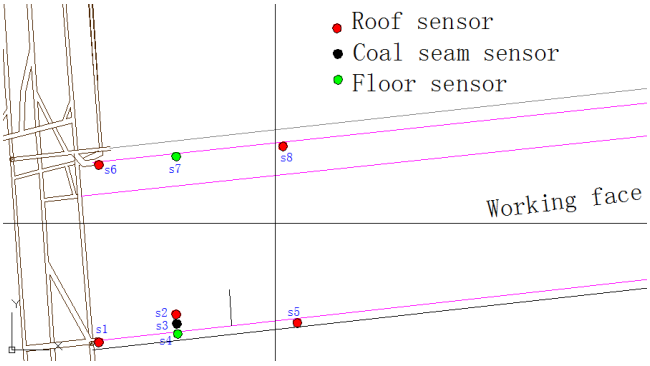

FIGURE 5. Schematic measuring points of the eight sensors, s1 to s8, arranged for optimal microseismic monitoring

IV. TRAINING RESULT ANALYSIS

The workstation processor used for model training was an Intel 2.40GHz dual-core CPU with 32GB of memory and a 64-bit operating system. A two-dimensional time-frequency diagram could be obtained by using the S-Transformation of the original vibration acceleration signal. The original size of the time-frequency image is 1200 × 900 pixels, and the size after removing the white edge is 930 × 732 pixels. When the number of training steps is set to 2000, it will take 568 minutes to train the model with 930 × 732 pixels samples. However, when the sample pixels are reduced to 450 × 350 or 180 × 140, the training time will be reduced to 127 minutes or 20 minutes respectively. Therefore, it is necessary to accept a reduced image resolution, to achieve a consequently reduced training time. As a result, images of 450 × 350 pixels and 180 × 140 pixels were respectively used for model training. The test results with different parameters when the size of sample image was 450 × 350 pixels are shown in TABLE I.

TABLE I
TEST RESULTS WITH DIFFERENT PARAMETERS WHEN THE SIZE OF SAMPLE IMAGE IS 450 × 350 PIXELS

Conv1weight	Conv2weight	Test accuracy/%
[3,3,3,32]	[3,3,32,64]	80.77
[3,3,3,16]	[3,3,16,32]	80.77
[5,5,3,32]	[5,5,32,64]	86.54
[3,3,3,16]	[3,3,16,16]	90.38

When the size of sample image is 450 × 350 pixels, the following were the data used: the training batch size was 16; the 'keep_prob' value to avoid over-fitting function dropout was 0.45; the ksize of the maximum pooling function was [1,3,3,1]; the step parameter strides were [1,2,2,1]; the shape of weight tensor of convolution layer 1 (Conv1weight) was [3,3,3,32]; and the shape of the weight tensor of the convolution layer 2 (Conv2weight) was [3,3,32,64]. Several key points should be noted: 1) After 1000 steps of training, 52 test samples were tested with the model obtained, 42 samples were predicted correctly and the accuracy obtained is 80.77%. 2) When the weight tensor shape of the convolution layer 1 was changed to [3,3,3,16] and the weight tensor shape of convolution layer 2 was changed to [3,3,16,32], the test accuracy obtained was still 80.77%. 3) When the shape of the weight tensor of convolution layer 1 was changed to [5,5,3,32] and the shape of weight tensor of convolution layer 2 was changed to [5,5,32,64], the test accuracy then obtained was 86.54%. 4) If the weight tensor shape of convolution layer 1 was changed to [3,3,3,16] and the weight tensor shape of convolution layer 2 was changed to [3,3,16,16], the test accuracy resulting increased to 90.38%. This latter accuracy is the highest, compared with the test results obtained before with the three previous sets of parameters. The test results with different parameters when the size of sample image is 180 × 140 pixels are shown in TABLE II.

TABLE II
TEST RESULTS WITH DIFFERENT PARAMETERS WHEN THE SIZE OF SAMPLE IMAGE IS 180 × 140 PIXELS

Batch size	Keep_ prob	St ep	Ksize	Strides	Conv1 weight	Conv2 weight	Test accura cy/%
16 or 32	0.45	20 00	[1,3, 3,1]	[1,2,2, 1]	[3,3,3,1 6]	[3,3,16 ,16]	90.38
16	0.4 or 0.5	20 00	[1,3, 3,1]	[1,2,2, 1]	[3,3,3,1 6]	[3,3,16 ,16]	90.38
16	0.45	10 00	[1,3, 3,1]	[1,2,2, 1]	[3,3,3,1 6]	[3,3,16 ,16]	90.38
16	0.45	10 00	[1,2, 2,1]	[1,2,2, 1]	[3,3,3,1 6]	[3,3,16 ,16]	86.54
16	0.45	10 00	[1,3, 3,1]	[1,3,3, 1]	[3,3,3,1 6]	[3,3,16 ,16]	90.38
16	0.45	10 00	[1,3, 3,1]	[1,2,2, 1]	[5,5,3,1 6]	[5,5,16 ,16]	90.38
16	0.45	10 00	[1,3, 3,1]	[1,2,2, 1]	[5,5,3,3 2]	[5,5,32 ,64]	92.31
16	0.45	10 00	[1,3, 3,1]	[1,2,2, 1]	[3,3,3,3 2]	[3,3,32 ,64]	96.15

The next step was to continue to reduce the sample image size to 180 × 140 pixels and use the convolution weight tensor shape from example 4) above. Thus when the size of the sample image was 180 × 140 pixels, the following were the data used: the training batch size was 16; the 'keep_prob' value of avoiding over-fitting function dropout was 0.45; the ksize of maximum pooling function was [1,3,3,1]; the step parameter strides were [1,2,2,1]; the shape of weight tensor of convolution layer 1 (Conv1weight) was [3,3,3,32]; the shape of weight tensor of convolution layer 2 (Conv2weight) was [3,3,32,64]; and the number of training steps used was also 2000. Again, several points should be noted. 1) 52 test samples were tested with the model obtained after the training process, 47 samples are predicted correctly, and the accuracy obtained was 90.38%. 2) When the training batch size was changed to 32, the test accuracy was still 90.38%. Simply increasing the training batch size could not change the test result and therefore the training batch size was kept as 16. 3) So as to test the influence of different values of 'keep_prob' of the dropout on the training results, the value of 'keep_prob' was changed from 0.45 to 0.4 and then 0.5 respectively and as a result, the test accuracy could not be improved. 4) To test the effect of the training steps on the training results, the number of training steps was changed from 2000 to 1000 and the results show that the accuracy of the test was not decreased. As a result, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2978392, IEEE Access

number of training steps was changed to 1000, to take advantage of the fact that this would further reduce the training time. 5) When the ksize value of the maximum pooling function was changed to [1,2,2,1], after 1000 steps of training, the test accuracy was reduced to 86.54%; then the ksize value of the maximum pooling function was changed back to a value of [1,3,3,1], the strides of the step parameter was changed to [1,3,3,1] and the test accuracy was still 90.38%. Therefore, it was concluded that the initial maximum pooling parameter should still be used. 6) When the weight tensor shape of the convolution layer 1 was changed to [5,5,3,16], the weight tensor shape of convolution layer 2 was changed to [5,5,16,16], it was shown that the test accuracy was still not improved. Then the weight tensor shape of convolution layer 1 was changed to [5,5,3,32], the weight tensor shape of convolution layer 2 was changed to [5,5,32,64] and so the test accuracy was increased to 92.31%. Finally, the weight tensor shape of convolution layer 1 was changed to [3,3,3,32] and the shape of the weight tensor of convolution layer 2 was changed to [3,3,32,64] and as a result, the test accuracy further improved to achieve a result of 96.15%.

From the above test data, it can be concluded that when the shape of the weight tensor of convolution layer 1 was [3,3,3,32], the shape of the weight tensor of convolution layer 2 was [3,3,32,64] and the size of the sample image was 180 × 140 pixels, the highest test accuracy of 96.15% was obtained. However, when the sample image size was 450 × 350 pixels, the accuracy of the same training parameters was less, at only 80.77%, with the same training parameters used. The analysis carried out has shown that the optimal training model parameters are different, with different sample image sizes being used.

The following parameters were the same data used in the model architecture of AlexNet for comparison: the number of training steps was 2000; the training batch size was 32; the size of sample image was 180 × 140 pixels. The accuracy obtained was only 50%, which was far below the above maximum discussed above of 96.15%. Therefore, it is feasible to classify successfully the blasting signals and the microseismic signals based on using the model architecture and parameters proposed in this paper.

V. CONCLUSIONS

Taking advantage of the use of CNN in image recognition, here CNN is utilized in the classification and recognition of both blasting signals and microseismic signals obtained from actual coal mines. Extracting the characteristic parameters manually is tedious and the use of the automated technique described has shown that it can be used to train the original image directly, in that way to avoid the tedious work caused by data preprocessing. The neurons between the adjacent layers are partly connected and so the number of training parameters and the number of connections can be greatly reduced.

A new model training architecture of the two-dimensional time-frequency image by use of the S-Transform has thus been designed. It uses the approach: the input layer - convolution layer 1 (ReLU, batch standardization) - pooling layer 1 (LRN) - convolution layer 2 (ReLU, batch standardization) - pooling layer 2 (LRN) - fully connected layer 1 (ReLU, Dropout) - fully connected layer 2 (ReLU, Dropout) - SOFTMAX - Classification result. The high accuracy recognition achieved was successfully realized by combining the training image with the S-Transform and the corresponding model training parameters.

This work done has shown that it is feasible to classify blasting signals and microseismic signals based on the CNN and the S-transform. When the size of the sample image is too large and the model needs to be trained for a long time, the training time can be reduced by reducing the size of the image and the number of training steps. When the sample image size is 180 × 140 pixels and after testing different parameters, the test accuracy can reach as high as 96.15%. It should be noted that for any one image size, the optimal training parameters are not necessarily the optimal parameters for any other sizes of image, and so the network needs to be retrained to optimize the training parameters for that specific case.

Thus, in summary, a new approach has been put forward in this paper, which provides a new method that can be used for the recognition and classification of microseismic signals and blasting signals in actual coal mines. The time-frequency image by S-Transform was utilized for the training, testing and predicting of samples with the CNN and the prediction accuracy rate achieved can achieve a high level as a result. By optimizing the CNN architecture and the model parameters, the accuracy rate of the sample prediction can be further improved. With a view to continuously enhancing the approach, in future studies work will be done to optimize the CNN model architecture, design a deeper and wider network and optimize the model parameters, such as changing the convolution stride and convolution weight tensor to further improve the accuracy of the sample prediction. Furthermore, the test accuracy will be improved by adding more samples and the optimized approach.

REFERENCES

[1] J. A. Vallejos, S. D. Mckinnon, "Logistic regression and neural network classification of seismic records," International Journal of Rock Mechanics & Mining Sciences, 2013, 62(9):86-95.
[2] D. Malovichko, "Discrimination of blasts in mine seismology," Deep and High Stress Mining, 2012, 72.
[3] X. Y. Shang, X. B. Li, A. M. Esteban et al., "Improving microseismic event and quarry blast classification using Artificial Neural Networks based on Principal Component Analysis," Soil Dynamics and Earthquake Engineering, 2017, 99:142-149.
[4] A. Booker, W. Mitronovas, "An application of statistical discrimination to classify seismic events," Bulletin of the Seismological Society America, 1964, 54:961 -971.
[5] B. A. Bolt, "Nuclear explosions and earthquakes-the parted veil," W. H. Freeman and Company, San Francisco USA, 1976.

[6] A. Douglas, "Forensic seismology revisited," Survey in Geophysics, 2007, 28(6):1-31.
[7] M. Musil, A. Pleginger, "Discrimination between Local Microearthquakes and Quarry Blasts by Multi-Layer Perceptrons and Kohonen Maps," Bulletin of the Seismological Society of America, 1996, 86(4):1077-1090.
[8] L. J. Dong, D. Y. Sun, X. B. Li et al., "A statistical method to identify blasts and microseismic events and its engineering application," Chinese Journal of Rock Mechanics and Engineering, 2016, 35(7):1423-1433.
[9] W. Li, "Feature extraction and classification method of mine microseismic signals based on LMD and pattern recognition," Journal of China Coal Society, 2017, 42(5):1156-1164.
[10] G. Y. Zhao, J. Ma, L. J. Dong et al., "Classification of mine blasts and microseismic events using starting-up features in seismograms," Transactions of Nonferrous Metals Society of China, 2015, 25(10):3410-3420.
[11] J. A. Vallejos, S. D. Mckinnon, "Logistic regression and neural network classification of seismic records," International Journal of Rock Mechanics & Mining Sciences, 2013, 62(9):86-95.
[12] W. McCulloch, W. Pitts, "A logical calculus of the Ideas immanent in nervous activity," Bulletin of Mathematical Biophysics, 1943, 5.
[13] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," Psychological review, 1958, 65(6):386.
[14] D. E. Rumelhart, G. E. Hinton, R. J. Williams, "Learning representations by back-propagating errors," Nature, 1986, 323(6088):399-421.
[15] C. Cortes, V. Vapnik, "Support-Vector Networks," Machine Learning, 1995, 20:273-297.
[16] G. E. Hinton, R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, 2006, 313(5786):504-507.
[17] J. W. Liu, Y. Liu, X. L. Luo, "Research and development on deep learning," Application Research of Computers, 2014, 31(7):1921-1930.
[18] Z. Y. Zheng, S. Y. Gu, "TensorFlow：Google deep learning framework," Publishing House of Electronics Industry, Beijing, 2017, 4-5.
[19] W. J. Huang, Y. Tang, "TensorFlow in Actual combat," Publishing House of Electronics Industry, Beijing, 2017, 76-141.
[20] Z. F. Shao, J. J. Cai, "Remote sensing image fusion with deep convolutional neural network," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018:1-14.
[21] Z. F. Shao, L. G. Wang, Z. Y. Wang, et al., "Saliency-Aware Convolution Neural Network for Ship Detection in Surveillance Video," IEEE Transactions on Circuits and Systems for Video Technology, 2019:1-1.
[22] B. Yang, J. Y. Zhong, "Review of Convolution Neural Network," Journal of University of South China (Science and Technology), 2016, 30(3):66-72.
[23] M. Yu, Z. B. Chen, L. F. Wang, "Research on BP-GA mixture algorithm in micro-seismic source inversion," Computer Technology and Its Application, 2013, 39(5):135-137.
[24] J. Bouvrie, "Notes on Convolutional Neural Networks," Neural Nets, 2006.
[25] X. H. Chen, Z. H. He, D. J. Huang, "Generalized S Transform and Its Time-Frequency Filtering," SIGNAL PROCESSING, 2008, 24(1):28-31.
[26] R. G. Stockwell, L. Mansinha, R. P. Lowe, "Localization of the complex spectrum: The S-transform," IEEE Transactions on Signal Processing, 1996, 44:998-1001.
[27] C. R. Pinnegar, L. Mansinha, "The S-Transform with windows of arbitrary and varying shape," Geophysics, 2003, 68(10):381.
[28] Y. Lecun Y, L. Bottou, Y. Bengio et al., "Gradient-based learning applied to document recognition," Proceedings of the IEEE, 1998.
[29] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet classification with deep convolutional neural networks," International Conference on Neural Information Processing Systems, pp. 1097-1105, Lake Tahoe, USA, December, 2012.