Accurate Identification Strategy of Coal and Gangue Using Infrared Imaging Technology Combined With Convolutional Neural Network

FENG HU[ID][1,2] AND KAI BIAN[ID][1]

[1]School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan 232001, China
[2]State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mines, Anhui University of Science and Technology, Huainan 232001, China

Corresponding author: Feng Hu (hufeng0106@163.com)

ABSTRACT To effectively separate coal and gangue, accurate classification is an important prerequisite. Here, a new recognition solution for coal and gangue is proposed, in which the convolutional neural network (CNN) is trained to achieve the automatically identifying coal and gangue based on the infrared images without considering the selection of feature extraction and classifier. Firstly, the specific architecture and detailed parameters of the model are optimized and the CNN model based on only one Inception Block contains three different convolution kernels are considered to be the most appropriate model. Next, performance of the proposed identification model is analyzed and evaluated by the infrared image dataset, and we discovered that the CNN model is capable of correctly identifying 192 training samples and 48 test samples. Finally, compared with the traditional recognition model and other CNN recognition model, it is proved that the proposed CNN model has superior recognition performance. The results state clearly that the combination of infrared image and CNN can quickly and accurately identify coal and gangue without complex image processing steps. At the same time, the model has a certain anti-interference ability for different noises. And it has a certain reference value for the research and development of intelligent coal preparation equipment.

INDEX TERMS Coal-gangue identification, convolutional neural network, infrared imaging technology, image identification.

I. INTRODUCTION

China is a large coal country with a very rich stock of coal [1]. The basic characteristics of energy resources of "coal rich, poor oil, and less gas" determine the important position of coal in primary energy [2]. As a companion of coal mining [3], [4], gangue is harder than coal and its carbon content is low. Gangue is mainly composed of SiO_2 and Al_2O_3 [5], [6], which contains not only high levels of sulphur but also many heavy metals (such as arsenic, chromium and mercury). When coal is used as the fossil fuel, if it is blended with the gangue, it will reduce the utilization rate of coal,

and worse, it will cause serious environmental pollution [7]. With these in account, it is very important to separate gangue precisely from raw coal before the comprehensive utilization of coal. At the same time, precisely classify coal and gangue is the key and the prerequisite for achieving precise separation [8].

In coal (gangue) separation technology [9], [10], besides manually selecting the gangue and coal, the automatic separation method can be split into dry selection and wet selection depending on the use of water resources. Wet selection of gangue mainly includes moving sieve jigging and heavy media separation of the gangue and so on. For wet separation of gangue, it has a large footprint, high investment cost, and requires a lot of water resources. Besides, it will

produce a significant amount of coal slime pollution, which is complicated to deal with. Over the past few years, there has been rapid development in the research and application of the dry separation technique, mainly using different sensor technology to identify coal and gangue, including dual-energy gamma rays [11], X-ray [12], laser [13], image [14], and other means. It is easy to integrate for the application of ray detection technology to realize coal (gangue) separation, but the bad thing is that there is radiation during use, so radiation insulation is necessary. In the mean time, with the vigorous development and wide application of image processing and pattern recognition technology [15], the use of image recognition technology for gangue separation is considered to have a broad application prospect [16]. However, this method presents certain problems, environmental factors (such as light, dust, etc.) will affect the results of identification and separation for coal and gangue.

Infrared imaging technology [17]–[19], as a common technology for image acquisition, is less affected by light and environment, and has many applications in the fields of security monitoring, environmental monitoring, and medical detection [20]. Considering the bad separation environment of coal and gangue, infrared imaging technology is proposed to identify coal and gangue. Recently, deep learning [21], [22], especially convolution neural network (CNN) [23], [24], has become a hot topic in current research and has been widely used in face recognition [25], license plate recognition [26], spectral recognition [27] and other fields. In particular, CNN also has many applications in the processing and recognition of infrared images. Kuang *et al.* [28] offered a method of deep learning for the removal of optical noise from a single infrared image, and the optical noise can be eliminated by using the full CNN. Introducing a new method based entirely on the extreme learning machine (ELM) to learn the useful features of CNN, Khellal *et al.* [29] realized the rapid and precise classification of images, which is appropriate for the infrared recognition system and is verified on the VAIS dataset.

Taking these factors into account, a new recognition strategy for coal and gangue is proposed, in which the CNN is trained to achieve the automatically identifying coal and gangue on the basis of their infrared images without considering the selection of feature extraction and classifier. Firstly, we introduce the experimental instrumentations and materials, and then briefly explain the structure and training of CNN used in this paper. Next, we focus on the analysis of the experimental results. On the basis of comparing the performance of the CNN models of the two basic structures, the structure with better performance is selected, its parameters are optimized, and the CNN model most suitable for identifying coal and gangue is obtained. In particular, we also show the structure and training process of the CNN model. For the purpose of verifying the reliability of the CNN model, we also compare the CNN model with the traditional strategy of image feature extraction [30] combined with classifier, and analyze the robustness of CNN model to different noises.

Lastly, we present the summary and the perspectives of this study.

II. MATERIALS AND METHODS
A. INSTRUMENTATION

For the purpose of collecting infrared images of coal and gangue, we first built an infrared imaging system. Figure 1 provides the experimental configuration diagram for an infrared imaging system, it is comprised principally of an infrared light source, an infrared camera, and a computer system. The light source used in the experiment is an LED array (Guangzhou Tianjian Electronics Co., Ltd., Guangzhou, China) consisting of six infrared diodes with a peak wavelength of 940 nm. The infrared images of coal and gangue are captured via an infrared camera (S908; Shenzhen Linbaishi Technology Co., Ltd., Shenzhen, China) equipped with an OV2710, which is a true full HD (1080p) CMOS image sensor designed specifically for digital video camcorders, PC webcam other applications. More specifically, we installed a 940 nm narrowband filter in front of the CMOS to capture only infrared images near the 940 nm band. In this way, we can effectively avoid the influence of other light on the images of coal and gangue.
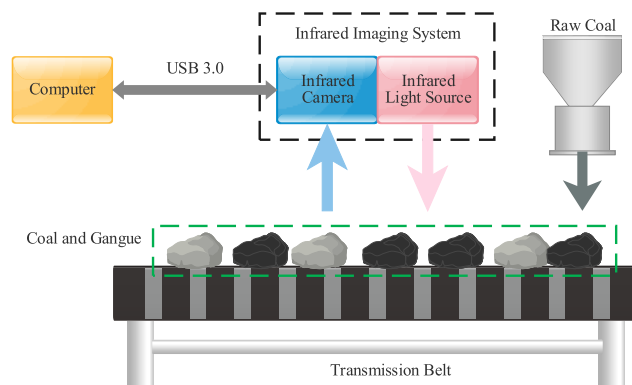

FIGURE 1. Schematic diagram of the infrared imaging system.

B. MATERIALS AND SAMPLES

Because Huainan is one of the main coal-producing regions in China, we choose coal and gangue in Huainan as the subject of research. Experimental materials, coal and gangue used to collect infrared images, were gathered from the Huainan mining region of Anhui Province on January 8, 2019. 120 pieces of coal and gangue with similar size and shape were selected for infrared image acquisition. All samples were detected and analyzed under the same conditions to ensure more realistic and reliable experimental data were available. In particular, for the infrared images of coal and gangue, we randomly select 96 samples from 120 samples as training samples and the rest as test samples. In order to ensure the effectiveness of the recognition model of coal and gangue, we adopt 5-fold cross validation in the process of model training, that is, the data set is divided into 5 parts, 4 of which are trained and 1 is verified in turn, and the mean value of the results of 5 times is used as the estimation of the accuracy of the recognition model.
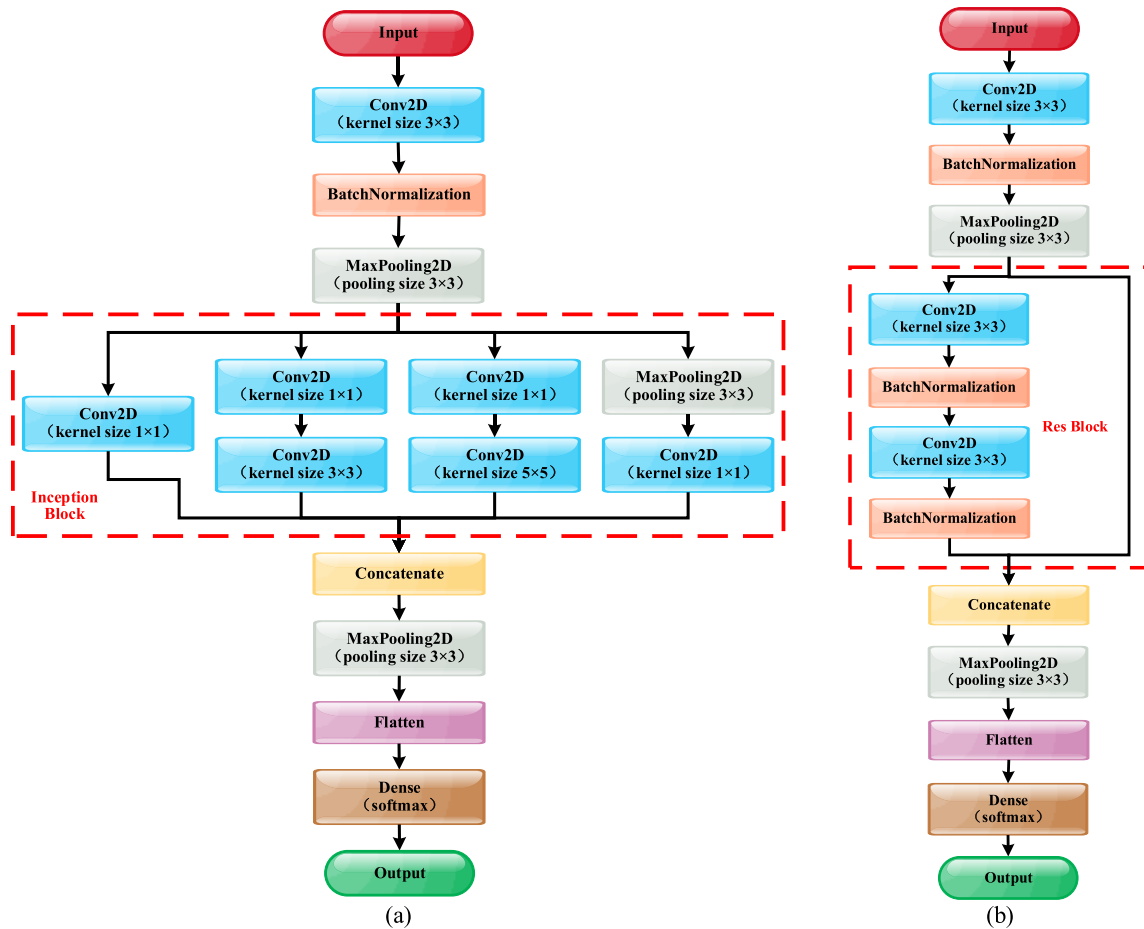
FIGURE 2. Two typical CNN structures. (a) Structure A. (b) Structure B.

C. CNN FOR INFRARED IMAGE CLASSIFICATION

CNN, as an important element of deep learning and a focal point of research, usually includes the convolutional layer, activation layer, and pooling layer. CNN such as LeNet [31] and AlexNet [32] has a good performance in image recognition. For the purpose of improving the performance of the neural network, two new ideas emerged: one is to introduce the inception unit [33], and the other is to introduce the residual unit [34]. As a result, two typical CNN structures were designed as shown in Figure 2. In structure A, the adaptability of the network is enhanced mainly by using multiple convolution kernels of different sizes (using Inception Blocks). The problem of gradient vanishing and performance degradation with the increase of network depth is solved by calculating the residual (using Res Blocks) in structure B. Whether in structure A or structure B, a deeper CNN model can be easily established by increasing the number of Inception Blocks or Res Blocks. Moreover, to determine the best structure, the performance of the different CNN models is compared in Section III.B.

The main components of CNN are detailed below:

The original size of infrared images of coal and gangue is 1920×1080. Considering the real-time performance of the recognition model of coal and gangue, we scale the

original infrared image information to 10% of the original size, that is, the infrared image information with the size of 192×108 is used for the recognition of coal and gangue. So the input information regarding the classification model of coal and gangue is the infrared image, which has a size of $192 \times 108 \times 3$.

In the Inception Block or the Res Block, features are extracted by convolution layer and each layer applies various numbers (initial default is 16) of convolution kernels of the same size (usually 3×3).

Batch normalization (BN) [35], as a means to improve the reliability and stability of the model, which may inhibit gradient vanishing and over-fitting.

Max pooling, as one of the most widely used strategies for pooling, can decrease the number of dimensions in feature maps and network settings. More specifically, the two frameworks designed in this paper use max pooling with a size of 3×3.

The upper layers of the CNN are fully connected through a dense layer that has the same number of outlets as the class of the samples. For the two structures shown in Fig. 2, only the Dense layer uses softmax as the activation function, and the remaining activation functions are Rectified linear units (ReLU).
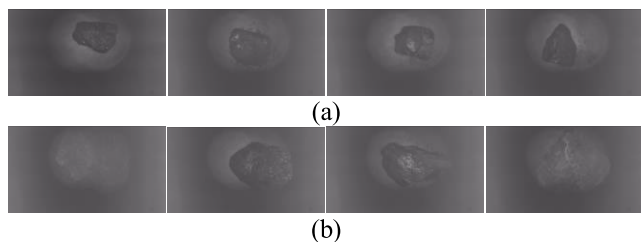
FIGURE 3. Infrared images of some coal and gangue samples. (a) Coal. (b) Gangue.

The softmax activation produces values between [0,1] as confidence scores in the classification. Loss of classification is calculated by comparing the confidence ratings with the actual labels of the samples. The softmax function and loss function are expressed as below,

$$p_{mn} = \frac{e^{z_{mn}}}{\sum_{l=1}^{K} e^{z_{ml}}} \quad n = 1, 2, \cdots, K \tag{1}$$

$$Loss = -\sum_{m} \sum_{n} \text{label}_{mn} \log(p_{mn}) \tag{2}$$

where z indicates the input of the softmax layer, m indicates a sample, n indicates a class and K indicates the total number of classifications.

D. TRAINING OF CNN MODEL

Two typical CNN models were first constructed on Keras (v 2.2.4), with a TensorFlow background (v 1.10.0). The templates were developed on a computer that has a Core I7-9700K processor and a GeForce RTX 2070 graphic processor. The CNN training was carried out using the SGD algorithm and the Adam algorithm. Firstly, the performance of two CNNs with different optimizers is compared by using the default parameters of Keras. Subsequently, the parameters of the preferred optimizer are further adjusted. For the optimization, the batch size was set to 128 and the training epoch was set to 1000.

III. RESULTS AND DISCUSSION

A. INFRARED IMAGE OF SAMPLES

The infrared images of coal and gangue are collected through the infrared imaging system shown in Figure 1, with the original size of 1920×1080, and several infrared images of samples are displayed in Figure 3. Some coal and gangue can be easily distinguished by infrared images, but some are not easily distinguishable. Consequently, there is a need to classify coal and gangue using image processing and pattern recognition methods.

B. COMPARISON BETWEEN THE TWO CNN STRUCTURES

For testing the two structures mentioned in Section II.C, the identification model with a single Inception Block or a single Res Block is built on the basis of Fig. 2, and the performance and representation of different CNN identification models of the infrared images of coal and gangue

TABLE 1. Results of the two CNN structures under different optimizers.

	Structure A		Structure B	
	Accuracy (%)	Loss	Accuracy (%)	Loss
SGD	97.22± 1.21	0.0517± 0.0197	96.53± 2.41	0.0793± 0.0577
Adam	**100.00± 0.00**	**0.0018± 0.0021**	98.61± 2.41	0.0383± 0.0446
Adamax	97.92± 2.09	0.0603± 0.0484	98.61± 2.41	0.0677± 0.0539
Nadam	98.61± 1.20	0.1188± 0.1884	**99.31± 1.20**	**0.0319± 0.0349**

was compared. By changing the optimizers (SGD, Adam, Adamax, and Nadam), the evaluation indexes (accuracy and loss) of the test set under two structures were recorded under 3 tests, respectively. The mean accuracy and loss of different CNN models were computed and displayed, as indicated in Table 1. We note that the CNN model, whether structural A or structural B, has a good effect on identifying coal and gangue. When using the A structure to build a CNN model, the optimizer selects Adam to achieve the highest recognition rate, that is, 100.00%. And when using structure B to build a CNN model, the optimizer selects Nadam to achieve the highest recognition rate, that is, 99.31%. This indicates that Structure A is better suited to identifying coal and gangue through infrared images. Based on Table 1, we determine that the CNN of structure A is utilized in the establishment of the CNN recognition model for coal and gangue by using the infrared images, and Adam is selected as the optimizer.

C. OPTIMIZATION OF THE CNN ARCHITECTURE

With the results of in Section III.B, we know that a CNN model with only one Inception Block can achieve an identification rate of 100.00%, so there is no need for increasing the depth of the network. To simplify the structure of the neural network, we gradually decrease the number of convolution kernels in the convolutional layer, and the accuracy of the test set is shown in Figure 4. We can see that the best recognition effect can be attained when the number of kernels is 10, 12, 14 and 16.

Next, for obtaining the model with the best performance and the simplest structure, the parameters, such as the loss, training time and trainable parameters, were taken into account in the choice of the most suitable identification model. Table 2 illustrates the average results from the three trials. To begin with, we made a comparison of the loss and noted that the loss of the CNN model with 14 convolution kernels can reach a minimum of 0.0007. Then, by comparing the training time, we observed that the training time that uses 10 convolution kernels is shorter and steadier. Then we found that the number of the trainable parameters that apply to the identification model using 10 convolution kernels is almost two-thirds that of the CNN model with 14 kernels. In addition, as the number of convolution kernels goes fewer, the trainable
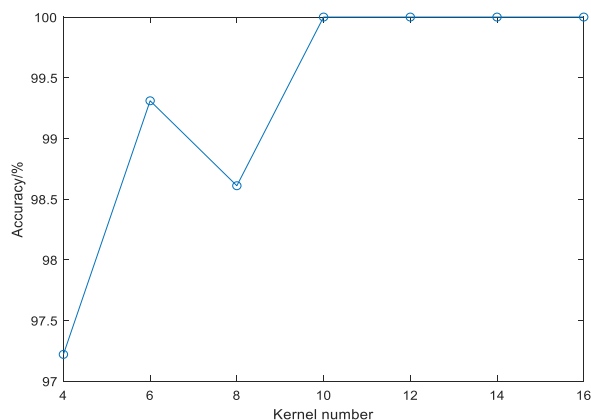
FIGURE 4. Accuracy of the test set under different kernel numbers in structure A.

parameters and the training time are gradually decreasing. Considering these indicators, the CNN identification model with 10 kernels in all convolutional layers is more appropriate for classifying coal and gangue by using infrared images.

D. PERFORMANCE OF THE DEVELOPED CNN IDENTIFICATION MODEL

According to the above results, we get the CNN structure for establishing the identification model for coal and gangue by using infrared images. Further details about the CNN model are given in Figure 5.

The perfect architecture is a one Inception Block CNN model: (1) The Inception Block contains three different convolution kernels with the size of 1×1, 3×3, and 5×5; (2) All convolutional layers have 10 convolution kernels. In order to more intuitively observe the types and parameters of each layer, we have established Table 3 to show all layers and all parameters used in the CNN model. In Table 3, we can clearly see the type of layers used in each layer of the CNN model, the connection relationship between layers, and the output size of each layer. In addition, we can also see the specific parameters of each layer, such as kernel sizes, kernel numbers, activation function and pool size.

The accuracy and loss in a variety of epochs of the proposed CNN model are set out in Figure 6. The results indicate that accuracy increases and the loss decrease with the increase of epochs. Following 200 iterative learning epochs, the accuracy and loss gradually become stable when applied to identifying coal and gangue. In this way, the established model achieves an accuracy of 100.00% in both training set and test set, which means that the CNN recognition model is capable of correctly identifying 192 training samples and 48 test samples. That is to say, the proposed CNN recognition model is achievable and efficient for identifying coal and gangue through infrared images.

E. COMPARED WITH THE TRADITIONAL CLASSIFICATION MODEL

As is well known, traditional image recognition usually includes two parts: feature extraction and classifier.

TABLE 2. Comparison of corresponding results of different convolution kernel numbers.

Kernel numbers	Trainable parameters	Time (ms/epoch)	Loss
16	28738	125.48±1.22	0.0018±0.0021
14	24082	124.96±0.53	**0.0007±0.0007**
12	19730	121.81±0.75	0.0047±0.0045
10	**15682**	**120.19±0.45**	0.0013±0.0020

For feature extraction, the three most commonly used features are local binary pattern (LBP) features, histogram of oriented gradient (HOG), and Haar features. These three features describe three types of local information and have different performances in different fields. As a traditional classification model, support vector machine (SVM) has an enormous application range and a good performance in the field of image recognition. Grid search (GS), genetic algorithm (GA), and particle swarm optimization (PSO) are the common parameter optimization methods of SVM. Hence, the performance of various feature extraction methods (LBP, HOG, and Haar) and SVM classifier (GS-SVM, GA-SVM, and PSO-SVM) combinations are mainly compared here. It is worth noting that all the image features here are processed in the following two steps before being fed into the SVM: normalized processing and principal component analysis (PCA) dimension reduction (with the cumulative contribution rate of 95%).

For the purpose of ensuring the reliability of the experimental results, we conducted three trials and calculated the average value, and the results are shown in Fig. 7. The figure displays the average accuracy of the test set under various recognition models. When the infrared images of coal and gangue are classified by using the LBP feature, combined with GS-SVM, the average recognition ratio for the test samples can reach a maximum of 88.19%. What stands out in the figure is the classification of coal and gangue by using Haar features, and the three classifiers have the same accuracy, which can reach 97.92%. The infrared images of coal and gangue are classified by the CNN model, and the accuracy of the test set reaches an identification rate of 100.00%. In comparison to SVM, the proposed CNN model has superior performance in identifying coal and gangue from infrared images. At the same time, it can also be seen that feature extraction methods of infrared images have a certain influence on classification performance. In other words, the traditional image recognition method needs to choose the extracted features reasonably in order to achieve a good classification result. All in all, the CNN model has better performance and it can achieve the automatically identifying coal and gangue on the basis of their infrared images without considering the selection of feature extraction and classifier.

To more intuitively analyse the performance of different methods, we use the confusion matrix to count the recognition results of coal and gangue of different methods under three different experiments, as shown in Table 4. It should be noted that the statistics in the table are the summary of the results of three different experiments using 5-fold cross validation.
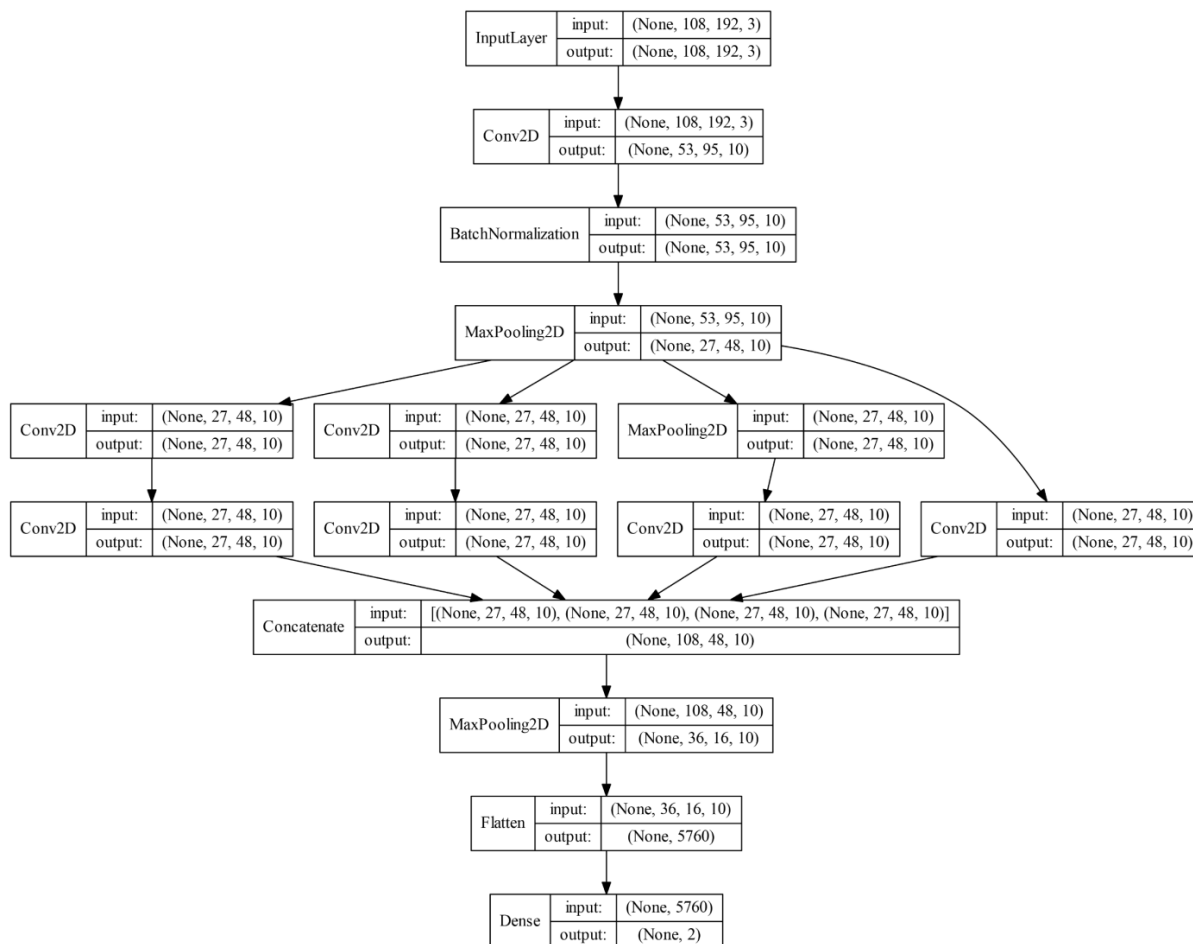
FIGURE 5. Details of the developed CNN identification model.

By observing this confusion matrix, we can more clearly understand the identification differences of coal and gangue by different methods. For example, when GS-SVM is used as the classifier, 22 coal samples are mistakenly identified as gangue, whereas 43 gangue samples are mistakenly identified as coal. When GA-SVM is used as the classifier, 25 coal samples are misidentified as gangue, while 30 gangue samples are misidentified as coal. On balance, the number of coal samples wrongly identified as gangue is lower than the number of gangue samples wrongly identified as coal samples.

F. COMPARED WITH THE STATE-OF-THE-ART CNN CLASSIFICATION MODEL

In order to further verify the effectiveness of the recognition model of coal and gangue proposed in this paper, we compare it with LeNet5, AlexNet, VGG, ResNet, DenseNet and other mainstream CNN model. Considering that the input size of other CNN models is inconsistent with the infrared image size collected in this paper, we modify the input of other CNN models to $108 \times 192 \times 3$. During the experiment, we counted the accuracy, training time and model parameters of different CNN models, as shown in Table 5.
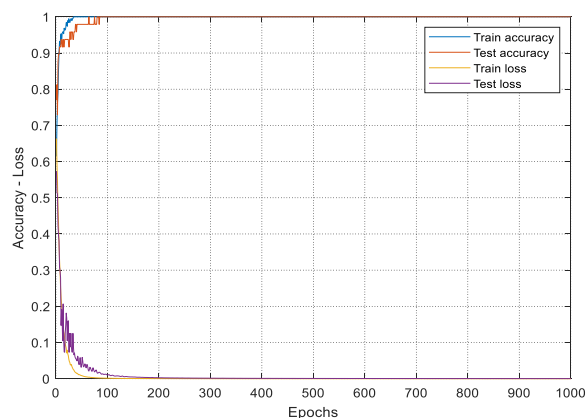
FIGURE 6. Variation trend of accuracy and loss during CNN training.

Firstly, we pay attention to the accuracy of the model. We can find that the overall performance of using CNN for the recognition of coal and gangue with infrared images is good. Only VGG-16, GoogleNet V3 and ResNet 152 have poor recognition accuracy, and the recognition rate of other models can be maintained at more than 95.00%. In particular, our proposed model can achieve 100.00% recognition accuracy. Secondly, when we observe the training time, we find that the training time of different models is very different.

TABLE 3. Layers and all parameters used in the CNN model.

Layer	Type	Note	Output size	Parameters	Connected to
input_1	InputLayer	/	(108, 192, 3)	0	/
conv2d_1	Conv2D	Kernel_size=3×3 Kernel_number=10 Activation: ReLU	(53, 95, 10)	280	input_1
batch_normalization_1	BatchNor	/	(53, 95, 10)	40	conv2d_1
max_pooling2d_1	MaxPooling2D	Pool_size=3×3	(27, 48, 10)	0	batch_normalization_1
conv2d_2	Conv2D	Kernel_size=1×1 Kernel_number=10 Activation: ReLU	(27, 48, 10)	110	max_pooling2d_1
conv2d_3	Conv2D	Kernel_size=1×1 Kernel_number=10 Activation: ReLU	(27, 48, 10)	110	max_pooling2d_1
max_pooling2d_2	MaxPooling2D	Pool_size=3×3	(27, 48, 10)	0	max_pooling2d_1
conv2d_4	Conv2D	Kernel_size=1×1 Kernel_number=10 Activation: ReLU	(27, 48, 10)	110	max_pooling2d_1
conv2d_5	Conv2D	Kernel_size=3×3 Kernel_number=10 Activation: ReLU	(27, 48, 10)	910	conv2d_2
conv2d_6	Conv2D	Kernel_size=5×5 Kernel_number=10 Activation: ReLU	(27, 48, 10)	2510	conv2d_3
conv2d_7	Conv2D	Kernel_size=1×1 Kernel_number=10 Activation: ReLU	(27, 48, 10)	110	max_pooling2d_2
concatenate_1	Concatenate	/	(108, 48, 10)	0	conv2d_4 conv2d_5 conv2d_6 conv2d_7
max_pooling2d_3	MaxPooling2D	Pool_size=3×3	(36, 16, 10)	0	concatenate_1
flatten_1	Flatten		(5760)	0	max_pooling2d_3
dense_1	Dense	Activation: Softmax	(2)	11522	flatten_1

TABLE 4. Confusion matrix of recognition results under different methods.

		Predicted class							
		Coal				Gangue			
		GS-SVM	GA-SVM	PSO-SVM	Our CNN	GS-SVM	GA-SVM	PSO-SVM	Our CNN
Actual class — Coal		LBP 314	LBP 305	LBP 312		LBP 46	LBP 55	LBP 48	
		HOG 338	HOG 335	HOG 340	360	HOG 22	HOG 25	HOG 20	0
		Haar 354	Haar 354	Haar 354		Haar 6	Haar 6	Haar 6	
Actual class — Gangue		LBP 49	LBP 55	LBP 47		LBP 311	LBP 305	LBP 313	
		HOG 43	HOG 30	HOG 24	0	HOG 317	HOG 330	HOG 335	360
		Haar 9	Haar 9	Haar 9		Haar 351	Haar 351	Haar 351	

ResNet 152, we found that with the deepening of the model depth, the training time increased, but the accuracy showed a downward trend. It is worth noting that we found an unbelievable phenomenon that the performance of GoogleNet-V3 and ResNet 152 is very poor. This may be because the model is too complex, and our data set can not fully meet the training of the model.

G. ROBUSTNESS TEST OF THE IDENTIFICATION MODEL

The selection environment of coal and gangue is usually quite harsh, and there is dust around it, which will cause certain interference to the infrared image data acquisition process. There are inevitably some noise signals in the infrared image data collection process of coal and gangue. In order to verify the robustness of the identification model proposed in this paper, the environmental disturbance in the coal preparation process is approximated, and the robustness of the identification model proposed for coal and gangue is tested. The following two sets of anti-interference experiments are carried out: One is to add different image noise signals to the training set and test set samples at the same time; the other is to only add different image noise signals to the test set samples. Gaussian noise, Poisson noise and Salt and pepper

The training time of our CNN model is the least (only 87.63 s), while the training time of ResNet50 model is 920.84s, and the training time of DenseNet model increases very much, reaching 5751.50s. In general, our proposed model has certain advantages in recognition accuracy and training time. In addition, we also found an interesting phenomenon. When we compared ResNet 50, ResNet 101 and
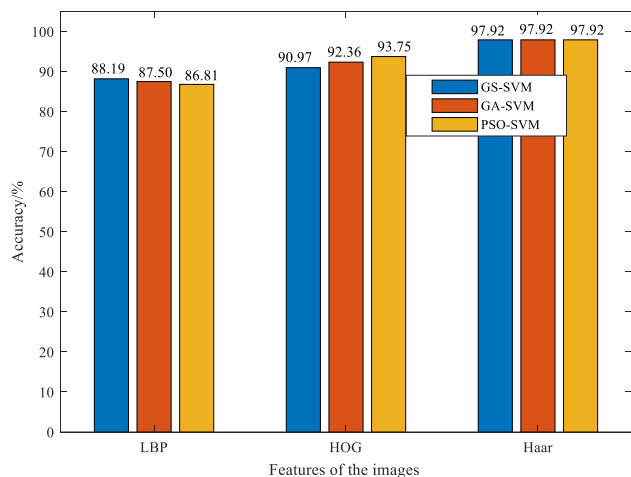
FIGURE 7. Accuracy of the test set under various feature extraction and classifiers.

TABLE 5. Comparison with other CNN models.

	Accuracy (%)	Training time(s)	Parameters
Our CNN	100.00	87.63	15702
LeNet5	97.92	98.03	37321610
AlexNet	95.83	120.30	28929410
VGG-16	69.53	749.81	69232450
GoogleNet V3	52.08	1125.43	46240709
ResNet 50	97.92	920.84	38136706
ResNet 101	95.83	2040.63	75050370
ResNet 152	52.08	2975.13	105458562
Xception	97.92	1179.74	20892842
DenseNet	95.83	5751.50	501980

TABLE 6. Results of adding different image noise.

Accuracy of the test set	Original	Gaussian	Poisson	Salt and pepper
Anti-interference test 1	100.00%	91.67%	97.92%	89.58%
Anti-interference test 2	100.00%	85.42%	97.92%	83.33%

noise were selected as image noise signals. The results of the anti-interference experiments are presented in Table 6.

For the first set of anti-interference experiments (simulating the same environmental disturbance), it is easy to observed that after adding different noise signals to the original infrared image of all samples, the classification accuracy of the test set decreases to different degrees, and the classification accuracy of the CNN model can be maintained above 89.00%. For the second group of anti-interference experiments (simulating the emergence of new environmental disturbances), it can be found that when only adding different image noise signals to the test set samples, the recognition rate decreases significantly. At this time, the recognition rate of the CNN recognition model can maintain

above 83.00%. Interestingly, when adding Poisson noise, the CNN identification model maintained a very good recognition effect (97.92%) under both anti-interference tests, that is, the recognition model has a very good anti-interference ability for Poisson noise. In summary, using infrared imaging technology combined with the CNN model for identifying coal and gangue, the identification model has a certain anti-interference ability when the same or different environmental disturbances occur.

IV. CONCLUSION

Here we have proposed a CNN model for classifying coal and gangue on the basis of infrared images which not only performs exceptionally well but also evades the challenge of choosing the appropriate image feature. At the beginning, we design two CNN models with typical structures. On the basis of determining the better structure, we optimize the parameters of the model, and finally get the CNN model which is most suitable for the recognition of coal and gangue with the infrared image. In addition, for the purpose of verifying the performance of the proposed CNN model, we also compare it with the traditional classification model and other CNN classification models.

The results show that CNN model based on only one Inception Block contains three different convolution kernels (with the size of 1×1, 3×3, and 5×5) are considered to be the most appropriate model. The CNN model proposed in this paper has an excellent performance in solving the identification problem of coal and gangue with infrared imaging. Compared with the traditional classification model, we found that the CNN model is more accurate, and utilizing CNN to analyze the infrared images does not need to consider the selection of extracting features (such as LBP, HOG, and Haar). In comparison with other CNN models, our proposed model offers some advantages in terms of recognition accuracy and training time. Furthermore, the proposed CNN model has some anti-interference capability for various noises.

This is significant because accurate identification of coal and gangue is an important prerequisite for intelligent separation of coal gangue, which is helpful to promote the intelligent process of coal industry. The new idea for identifying coal and gangue developed in this paper may have a certain reference value for the research and development of intelligent coal preparation equipment.

This study focuses on identifying coal and gangue from infrared images by using the CNN with only one Inception Block contains three convolution kernels of different sizes. It is worth extending this investigation to the identification with the help of a more straightforward structure (fewer types and numbers of convolution kernels) while planning to design a more general CNN identification model that can be used to identify more different substances.

CONFLICT OF INTEREST

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

REFERENCES

[1] D. Creedy, W. Lijie, Z. Xinquan, L. Haibin, and G. Campbell, "Transforming China's coal mines: A case history of the Shuangliu mine," *Natural Resour. Forum*, vol. 30, no. 1, pp. 15–26, Feb. 2006.

[2] J. Teng, Y. Qiao, and P. Song, "Analysis of exploration, potential reserves and high efficient utilization of coal in China," *Chin. J. Geophys.*, vol. 59, no. 12, pp. 4633–4653, 2016.

[3] C. Zhou, G. Liu, S. Wu, and P. K. S. Lam, "The environmental characteristics of usage of coal gangue in bricking-making: A case study at Huainan, China," *Chemosphere*, vol. 95, pp. 274–280, Jan. 2014.

[4] M. Yuan, Q. Zhang, and Y. Huo, "The effect of coal gangue pile on soil enzymatic activities," *China Mining Mag.*, vol. 27, no. 6, pp. 85–89, 2018.

[5] M. Yang, Z. Guo, Y. Deng, X. Xing, K. Qiu, J. Long, and J. Li, "Preparation of CaO–Al$_2$O$_3$–SiO$_2$ glass ceramics from coal gangue," *Int. J. Mineral Process.*, vols. 102–103, pp. 112–115, Jan. 2012.

[6] J. Galjak, J. Dokić, G. Milentijević, I. Dervišević, and S. Jović, "Characterization of the tailing waste deposit 'Gornje Polj,'" *Optik*, vol. 215, Aug. 2020, Art. no. 164684.

[7] S. Wang, K. Luo, X. Wang, and Y. Sun, "Estimate of sulfur, arsenic, mercury, fluorine emissions due to spontaneous combustion of coal gangue: An important part of Chinese emission inventories," *Environ. Pollut.*, vol. 209, pp. 107–113, Feb. 2016.

[8] D.-L. Yang, J.-P. Li, C.-L. Du, K.-H. Zheng, and S.-Y. Liu, "Particle size distribution of coal and gangue after impact-crush separation," *J. Central South Univ.*, vol. 24, no. 6, pp. 1252–1262, Jun. 2017.

[9] J. Tan, L. Liang, Y. Peng, and G. Xie, "The concentrate ash content analysis of coal flotation based on froth images," *Minerals Eng.*, vol. 92, pp. 9–20, Jun. 2016.

[10] F. Hu, M. Zhou, P. Yan, K. Bian, and R. Dai, "Multispectral imaging: A new solution for identification of coal and gangue," *IEEE Access*, vol. 7, pp. 169697–169704, 2019.

[11] M. Yazdi and S. A. Esmaeilnia, "Dual-energy gamma-ray technique for quantitative measurement of coal ash in the Shahroud mine, Iran," *Int. J. Coal Geol.*, vol. 55, nos. 2–4, pp. 151–156, Aug. 2003.

[12] Z. Dong, J. Xia, X. Duan, and J. Cao, "Based on curing age of calcined coal gangue fine aggregate mortar of X-ray diffraction and scanning electron microscopy analysis," *Spectrosc. Spectral Anal.*, vol. 36, no. 3, pp. 842–847, 2016.

[13] W. Wang and C. Zhang, "Separating coal and gangue using three-dimensional laser scanning," *Int. J. Mineral Process.*, vol. 169, pp. 79–84, Dec. 2017.

[14] K. Zheng, C. Du, J. Li, B. Qiu, and D. Yang, "Underground pneumatic separation of coal and gangue with large size ($\geq$50 mm) in green mining based on the machine vision system," *Powder Technol.*, vol. 278, pp. 223–233, Jul. 2015.

[15] K.-S. Fu, "Pattern recognition and image processing," *IEEE Trans. Comput.*, vol. C-25, no. 12, pp. 1336–1346, Dec. 1976.

[16] V. Maksimovic, M. Petrovic, D. Savic, B. Jaksic, and P. Spalevic, "New approach of estimating edge detection threshold and application of adaptive detector depending on image complexity," *Optik*, vol. 238, Jul. 2021, Art. no. 166476.

[17] C. Gao, C. Song, Y. Zhang, D. Qi, and Y. Yu, "Improving the performance of infrared and visible image fusion based on latent low-rank representation nested with rolling guided image filtering," *IEEE Access*, vol. 9, pp. 91462–91475, 2021.

[18] X. Wang, J.-F. Witz, A. El Bartali, and C. Jiang, "Infrared thermography coupled with digital image correlation in studying plastic deformation on the mesoscale level," *Opt. Lasers Eng.*, vol. 86, pp. 264–274, Nov. 2016.

[19] C. Ibarra-Castanedo, D. González, M. Klein, M. Pilla, S. Vallerand, and X. Maldague, "Infrared image processing and data analysis," *Infr. Phys. Technol.*, vol. 46, nos. 1–2, pp. 75–83, Dec. 2004.

[20] Q. Li, J. Nie, and S. Qu, "A small target detection algorithm in infrared image by combining multi-response fusion and local contrast enhancement," *Optik*, vol. 241, Sep. 2021, Art. no. 166919.

[21] Z.-M. Li, W.-H. Gui, and J.-Y. Zhu, "Fault detection in flotation processes based on deep learning and support vector machine," *J. Central South Univ.*, vol. 26, no. 9, pp. 2504–2515, Sep. 2019.

[22] X. Chen, L. Xie, Y. He, T. Guan, X. Zhou, B. Wang, G. Feng, H. Yu, and Y. Ji, "Fast and accurate decoding of Raman spectra-encoded suspension arrays using deep learning," *Analyst*, vol. 144, no. 14, pp. 4312–4319, 2019.

[23] S. Saponara, A. Elhanashi, and Q. Zheng, "Recreating fingerprint images by convolutional neural network autoencoder architecture," *IEEE Access*, vol. 9, pp. 147888–147899, 2021.

[24] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput.*, vol. 29, no. 9, pp. 2352–2449, Sep. 2017.

[25] J.-J. Lv, X.-H. Shao, J.-S. Huang, X.-D. Zhou, and X. Zhou, "Data augmentation for face recognition," *Neurocomputing*, vol. 230, pp. 184–196, Mar. 2017.

[26] L. Xie, T. Ahmad, L. Jin, Y. Liu, and S. Zhang, "A new CNN-based method for multi-directional car license plate detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 507–517, Feb. 2018.

[27] F. Hu, M. Zhou, P. Yan, D. Li, W. Lai, K. Bian, and R. Dai, "Identification of mine water inrush using laser-induced fluorescence spectroscopy combined with one-dimensional convolutional neural network," *RSC Adv.*, vol. 9, no. 14, pp. 7673–7679, 2019.

[28] X. Kuang, X. Sui, Y. Liu, Q. Chen, and G. Gu, "Single infrared image optical noise removal using a deep convolutional neural network," *IEEE Photon. J.*, vol. 10, no. 2, pp. 1–15, Apr. 2018.

[29] A. Khellal, H. Ma, and Q. Fei, "Convolutional neural network based on extreme learning machine for maritime ships recognition in infrared images," *Sensors*, vol. 18, no. 5, p. 1490, May 2018.

[30] Q. Yang, L. Tang, and R. Yu, "Efficient pattern unmixing of multiplex proteins based on variable weighting of texture descriptors," *Anal. Methods*, vol. 8, no. 46, pp. 8188–8195, 2016.

[31] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[33] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[35] X. Han and Q. Dai, "Batch-normalized Mlpconv-wise supervised pre-training network in network," *Appl. Intell.*, vol. 48, no. 1, pp. 142–155, Jan. 2018.

FENG HU received the Ph.D. degree in mine electrical and mechanical engineering from the Anhui University of Science and Technology, in 2020. He is currently a Lecturer with the Anhui University of Science and Technology. His research interests include machine learning, photoelectric information processing, and coal mine safety monitoring.

KAI BIAN is currently pursuing the Ph.D. degree with the Anhui University of Science and Technology. His research interests include machine learning and image processing.

• • •