A Real-Time Multilevel Fusion Recognition System for Coal and Gangue Based on Near-Infrared Sensing

ZIHAO DING1, GUODONG CHEN1, ZHENG WANG1,2, WENZHEN CHI1, ZHENHUA WANG1, AND YULIN FAN1

1Jiangsu Provincial Key Laboratory of Advanced Robotics, School of Mechanical and Electrical Engineering, Soochow University, Suzhou 215123, China
2Shanghai Jiao Tong University Affiliated Sixth People’s Hospital, Shanghai Jiao Tong University, Shanghai 200233, China

Corresponding authors: Guodong Chen (guodongxyz@163.com) and Wenzheng Chi (chiwenzheng@suda.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2019YBF1310201, and in part by the Key Project of Universities of Jiangsu Province in Natural Science Research under Grant 17KJA460007.

ABSTRACT Coal is an indispensable energy source for humans. As an important part of the mining industry, intelligent separation of coal and gangue will promote development. The traditional methods of recognition do not consider the interference created by a dynamic environment. There are many problems such as noise, complex backgrounds and occlusion, which lead to low accuracy and cannot satisfy real-time requirements in mining. Aiming at dynamic environments, a real-time multilevel fusion recognition system was built in this paper. First, we introduced a near-infrared camera into the field of separation, which was used to form a binocular system with a visible light camera. The SVM classifier was obtained by feature selection and fusion training of the binocular system, which overcomes the interference of environmental factors. Then, we proposed a new deep learning training method of two-sample fusion to improve the recognition network performance by expanding the number of samples and features. Finally, the SVM and deep learning algorithms were combined to establish a fast detection strategy. In addition, the length suppression algorithm was added to solve the occlusion problem. The accuracy of the fusion algorithm was 0.923 and the detection speed was increased to 26 fps. The experimental results indicated that the sorting system satisfied the requirements of real-time and robust of the coal industry.

INDEX TERMS Coal, near-infrared, sensing and fusion, separation, deep learning.

I. INTRODUCTION Clean processing and utilization of coal is the top priority for achieving green development in the coal industry. In many countries, coal-fired power plants provide most of the electricity [1]. The key to coal-fired power plants is how to reduce pollutant emissions while increasing heat release [2]. However, there is considerable gangue in the mining process, which has high density and low calorific value. The quality and efficiency of combustion will be seriously affected and the environment will be polluted if the gangue is mixed in with the coal for industrial production. Therefore, the separation of coal and gangue is an indispensable process in coal production [3]. Traditional artificial methods have high labor intensity. In addition, large quantities of coal ash have harmed people’s health. Many scholars have performed considerable studies on the separation of coal and gangue, which are mainly divided into wet separation processes and dry separation processes. See Table 1 for the characteristics of different methods.

The most widely used wet separation methods are dense-medium separators [4], jigging [5], and the flotation process [6]. The wet separation method is simple. However, it is difficult to dehydrate coal and requires considerable energy and investment. It also contains substances in the separation medium water, which causes environmental pollution. Currently, the wet separation method is being gradually eliminated by the market. Compared with wet separation methods, dry separation methods are low cost and use simple equipment and have become the focus of coal gangue
TABLE 1. Performance comparison of different separation methods.

| Category         | Methods      | Feature   | Advantage          | Disadvantage          |
|------------------|--------------|-----------|--------------------|-----------------------|
| Wet separation   | Dense-medium | Density   | High precision     | Complex process       |
| process          | Jigging      | Vibratio n | Simple process     | High investment       |
|                  | Flotation    | Wettabil ity | Effective for     | Only for less         |
|                  | Ray          | γ-rays    | Anti-interference  | Poor sorting ability  |
|                  |             | X-rays    | Long life          | Bad for health        |
| Dry separation   | ML           | Gray      | Easy to extract    | Sensitive             |
| process          | ML           | Texture   | Rich features      | Artificial selection  |
|                  | DL           | Neural network | High precision | Low efficiency        |

separation research. Dry separation mainly includes ray separation [7], [8], machine vision separation [9], and deep learning separation.

An intelligent dry sorting system, has been successfully applied in the market by using γ-rays as the detection sensor and a high-pressure air gun as the actuator [10]. However, the largest disadvantage of the ray method is that it will harm the human body [11]. With the rapid development of image processing technology, many scholars have proposed visual methods for separating coal and gangue. Among the most used features are density [12], grayscale, and texture features. Wang and Zhang [13] proposed separating gangue from coal on the basis of density, calculated from volume using three-dimensional (3D) laser scanning technology. Gao et al. [14] obtained the grayscale distributions of coal and gangue and employed the Bayesian Discriminant algorithm to differentiate them. Dou et al. [15] expanded the features, including the color and texture features and employed the relief-support vector machine (SVM) to identify optimal features and construct optimal classifiers. Sun et al. [16] introduced complementary textures into traditional textures. Perez et al. [17] improved the features selection method. He proposed a new method based on mutual information for feature selection and a voting process that considers boundary information to improve the classification. In recent years, neural networks have performed better in classification. Liang et al. [18] proved that it is possible to develop a high-precision coal gangue recognition system using a SVM and neural network. Li et al. [19] designed a four-layer Levenberg Marquart back-propagation neural network to classify coal gangue images. Based on a powerful trained image recognition model, VGG16, Pu et al. [20] introduced the idea of transfer learning to build a custom CNN model, which solves the problems of massive trainable parameters and limited computing power linked to the building of a brand-new model from scratch. Lai et al. [21] and Hu et al. [22] applied multispectral technology. Their research showed that the recognition accuracy was different for different wavelength spectral images.

Through extracting features, the SVM can identify the differences in features values and has good classification performance. However, the above studies were conducted under ideal laboratory conditions, in which the surface of gangue was clean and ashless. Under the actual conditions, the background and light change frequently and the scene is dusty. The target is in motion, resulting in occlusion between the targets. The ordinary visible light camera is greatly affected by the environment, which makes it difficult to extract obvious feature differences and affects the effect of SVM classification. In addition, the traditional machine vision method requires artificially designing the coal gangue image features. The recognition result depends on the designed algorithm and experience.

Therefore, this paper proposed a real-time multilevel fusion recognition system of coal and gangue based on near-infrared sensing. It adopts multilevel fusion: different sensor fusion in the hardware, sample fusion, feature fusion, and algorithms fusion in the software. Finally, the accurate identification of coal and gangue is realized.

II. REAL-TIME MULTILEVEL FUSION RECOGNITION SYSTEM

FIGURE 1 is the flow diagram of the coal and gangue separation system in this paper. Raw coal is mined and crushed and then enters the roller screen. The clearance between the roller screen gears is strictly set. Smaller coal (diameter < 50 mm) is directly discharged from the gear clearance. The rear clearance is slightly larger so that the mixture of coal and gangue (50 mm ∼ 150 mm) falls on the belt. Larger gangue (diameter > 150 mm) is discharged along the pipe because it cannot be dropped. The mixture of coal and gangue (50 mm ∼ 150 mm) is the target for the recognition system to separate. After the array unit on the belt, the recognition system distinguishes the coal and gangue and sends the information to the robot. At the end of the belt, the robot sorts the coal and gangue into different areas to complete the separation of the coal and gangue.

FIGURE 1. Schematic diagram of a coal sorting system.

Because coal and gangue are in motion, it creates many recognition challenges. Background light changes frequently in a dynamic environment. As shown in FIGURE 2, the brightness of the belt at the same position changes at
different times. A large amount of coal ash will remain on the belt during the movement, as shown in Fig. 2 (b). These noises will affect the background and the distribution of gangue characteristics, increase the difficulty of target segmentation from the background, and create interference in the recognition algorithm. In addition, the actual detection process is characterized by fast belt movement and easy occlusion between samples, as shown in (a) in FIGURE 2. When the sample volume is small, the number of samples in a picture is large, which affects the speed of image processing, as shown in FIGURE 2 (c).

FIGURE 3 is the algorithm flow used in the recognition system for solving dynamic problems. First, a sensor fusion device was built by combining a NIR (near-infrared) camera and a VIS (visible light) camera. Then, the SVM classifier was trained by using feature selection and feature fusion by using the characteristics of the NIR camera, which overcame the shortcoming that ordinary cameras cannot obtain effective features. A deep learning network of two-sample fusion was also built to improve sorting accuracy. Finally, the above two different algorithms and length suppression algorithms were integrated to establish a set of fast detection mechanisms.

A. FAST RECOGNITION ALGORITHM BASED ON NEAR-INFRARED FEATURES FUSION

1) FEATURE EXTRACTION AND ANALYSIS

To solve the problem of light mutation and background noise in the process of dynamic detection, a NIR camera was introduced to the field of coal mining. By using the characteristics of insensitivity to the change in light and high resolution of the NIR camera, differences between features of coal and gangue were extracted. In addition, in image processing, the matching relationship between the NIR camera and the sample was combined to select the features with better performance. The segmented images are shown in FIGURE 4.

Generally, the surface of coal is bright and black, while gangue is dim and gray. A gray histogram of coal and gangue can directly reflect their grayscale range and frequency distribution. Texture is an attribute that reflects the spatial distribution of the gray level of the pixels in a region of images. Using a gray-level cooccurrence matrix to describe the texture, the characteristic parameters could be extracted to quantitatively describe target characteristics. In this paper, gray mean, energy, contrast, correlation, and homogeneity are selected to analyze the difference between coal and gangue. The calculation formulas are shown in (1) to (10).

Gray mean:

\[ G = \frac{1}{N} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} n_{ij} \]  

(1)

Energy:

\[ f_1 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta) \cdot (i - j)^2 \]  

(2)

Contrast:

\[ f_2 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta) \cdot (i - \mu_1)(j - \mu_2) \]  

\[ \sigma_1 \sigma_2 \]  

(3)

Correlation:

\[ f_3 = \frac{1}{N} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta) \]  

\[ (i - \mu_1)(j - \mu_2) \]  

\[ \sigma_1 \sigma_2 \]  

(4)

Homogeneity:

\[ f_4 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta) \cdot \frac{1}{1 + ||i - j||^2} \]  

(5)
heat of an object’s surface, thus showing a better energy feature than the VIS camera.

2) SVM ALGORITHM BASED ON FEATURES FUSION
After analyzing the distribution characteristics of coal and gangue in different cameras and different features, the classification effect of each feature was tested in the test set, then selected the better features as the SVM feature vector. The SVM classifier has the advantage of high accuracy, which provides a good theoretical guarantee for avoiding overfitting. Manuel [23] demonstrated that SVM performed well on two-class data sets, which is very suitable for our task. Moreover, Hu et al. [22] achieved good results in the recognition of coal and gangue under the multispectral camera by using the SVM algorithm. Therefore, this paper chose the SVM algorithm as the feature fusion classifier. Finally, the classification results of the selected features are normalized to obtain the weight of the features and realize feature fusion. The accuracy $P(c_i)$ of each feature $c_i$ under the training sample is obtained by SVM. Then, it is normalized with (11) to obtain the weight $w_i$ of each feature.

$$w_i = \frac{P(c_i)}{\sum_{i=1}^{K} P(c_i)}$$

Machine learning algorithm is currently the most widely used algorithm for sorting coal and gangue. But with the development of the neural networks, the deep learning shows better performance in stability.

B. TWO-SAMPLE FUSION DETECTION ALGORITHM BASED ON TRANSFER LEARNING
Although the machine learning algorithm can extract and recognize the features of coal and gangue, the shallow features are easily disturbed by the background environment. If coal ash has adhered to the gangue surface, the traditional machine learning method may not be able to accurately identify the gangue. In addition, when the target is in a state of motion, the clarity of the captured image decreases, resulting in a change in features, and the robustness of the algorithm is seriously affected. Deep learning is composed of a multilayer neural network, which can obtain more basic feature expressions and extract more stable deep features between objects. In the actual mine environment, coal and gangue need to be separated on a fast-moving belt. Therefore, Yolov3 [24], a real-time deep learning network, was selected as the basic network in our algorithm.

There is a risk of over fitting due to the small data set. Therefore, this paper proposed a two-sample training method based on transfer learning and the device shown in FIGURE 7 was built to collect samples, which included both the pictures collected by the VIS camera and the NIR camera. The training method was modified and was different from the existing transfer learning based on different types of samples. The only difference was that the sensors were different. First,
the images collected by the VIS camera were input into the deep learning network for training. In addition, the middle model with certain recognition ability was trained from an initial meaningless parameter. Then, the 249 layers at the beginning of the network model were froze and the NIR images were input for training to obtain the final network model. Because the two training samples were all targets, both local features and global features were relatively similar. Most of the parameters in the intermediate model after the initial training could be applied to the identification of the final target. On the basis of transfer learning, the CNN network was equivalent to a feature extractor. The two-sample fusion method not only increased the training samples but also enriched features and improved the robustness of the network. The training process is shown in FIGURE 8.

To further improve network performance, the YOLOv3 deep learning network was modified, shown in FIGURE 9. The SVM algorithm has the advantage of fast recognition speed. However, it is easy to make errors when gangue has adhered with coal ash. The deep learning method can eliminate the interference and accurately identify but takes a long time. To improve the detection rate of coal and gangue and ensure accuracy, this paper proposed an effective fusion strategy based on the threshold of SVM detection confidence.

In addition, under actual working conditions, the coal and gangue samples on the belt are mixed, overlapped and unevenly distributed, which leads to the formation of narrow connecting areas between the extracted binary image target samples. The original multiple targets are regarded as one, which affects the correct extraction of the edge of the sample target and leads to incorrect positioning. Neither the SVM algorithm nor the deep learning algorithm can effectively solve the occlusion problem. To separate them to obtain independent sample targets, a length suppression algorithm was added to the fusion algorithm.

C. ALGORITHMS FUSION BASED ON DETECTION CONFIDENCE AND LENGTH SUPPRESSION

The SVM algorithm has the advantage of fast recognition speed. However, it is easy to make errors when gangue has adhered with coal ash. The deep learning method can eliminate the interference and accurately identify but takes a long time. To improve the detection rate of coal and gangue and ensure accuracy, this paper proposed an effective fusion strategy based on the threshold of SVM detection confidence.

In addition, under actual working conditions, the coal and gangue samples on the belt are mixed, overlapped and unevenly distributed, which leads to the formation of narrow connecting areas between the extracted binary image target samples. The original multiple targets are regarded as one, which affects the correct extraction of the edge of the sample target and leads to incorrect positioning. Neither the SVM algorithm nor the deep learning algorithm can effectively solve the occlusion problem. To separate them to obtain independent sample targets, a length suppression algorithm was added to the fusion algorithm. After sieving through a roller screen, the diameter of the coal and gangue to be tested is in the range of approximately 50-150 mm. Therefore, in the detection results, the size of the bounding box should also meet this range. When the two targets are connected or overlapped, the bounding box will increase significantly. If the size of the bounding box exceeds the expected range, the original detection area is divided into two, and the...
The detection algorithm is performed again to obtain the correct result. The specific algorithm flow is shown in FIGURE 10.

![Detection process of fusion algorithm.](image)

The SVM runs so fast that the algorithm starts the SVM algorithm for detection. Only when the recognition result is lower than a certain threshold due to a complex background or uneven target distribution, the deep learning algorithm is called to make a more accurate judgment on the target. Assuming the current SVM detection result is coal, (12) was used to calculate whether the confidence level for coal satisfies the condition.

$$\sum \omega_i (\alpha_i - C_i)^2 < \varepsilon$$ (12)

where, $\alpha_i$ is the value of detected features, $C_i$ and are the average values of the coal and gangue, respectively, in the previous feature analysis based on all samples, $\omega_i$ represents the weights calculated in (11), and $\varepsilon$ denotes the set thresholds.

When the formula is true, the deep learning algorithm is called to detect. Otherwise, the SVM algorithm continues to be applied. Then, (13) was used to verify the results. When the length of the bounding box is not satisfied, the detection frame is divided into two parts by (14). In addition, the targets are detected again to effectively eliminate the problem of occlusion.

$$W(H) > 150 \times \lambda$$ (13)

$$W'(H') = W(H)/2$$ (14)

where, $\lambda$ is the reliability coefficient, $W$ is the length of the bounding box and $H$ is the width of the bounding box.

### III. EXPERIMENT

The experiment was performed in the device shown in FIGURE 11. The NIR camera used in the detection system was a BV-C2901, and the VIS camera was an acA2440-20 gc from Basler. An FST-CLL370 was used as the NIR light source, and an ordinary bar light source was used as VIS light source. The PC was configured with a 2.7 GHz frequency and 16 GB of memory. It was equipped with an NVIDIA GeForce 2080Ti with a computing capability of 7.5.

In the experiment, coal samples were selected as positive samples and gangue samples as negative samples. The accuracy, precision, and recall rate were selected as evaluation criteria. The accuracy represents the proportion of samples predicted correctly in the total sample number. It was the main indicator for judging the performance of the classifier. The precision (also called positive predictive value) is the fraction of coals among the positive samples. In addition, recall (also known as sensitivity) is the fraction of the total amount of coals that are actually retrieved. The higher the precision, the more pure the coal quality. While a low recall rate indicates a considerable waste of coal resources. The calculation of each indicator is shown in the (15) to (17).

$$\text{Accuracy} = \frac{TP + TN}{N(+) + N(-)} \quad (15)$$

$$\text{Precision} = \frac{TP}{TP + N(-) - TN} \quad (16)$$

$$\text{Recall} = \frac{TP}{N(+)} \quad (17)$$

where $N(\cdot)$ represents the total number of actual positive samples, $N(-)$ represents the total number of actual negative samples, TP is the number of correctly classified positive examples, and TN is the number of correctly classified negative examples.

### A. FEATURE ANALYSIS AND CONTRAST EXPERIMENT UNDER DIFFERENT CAMERAS

In this experiment, the differences between the features of the samples collected by the VIS camera and the NIR camera...
TABLE 2. Comparison of characteristics of samples collected by the two cameras.

| Feature  | VIS simples | NIR simples |
|----------|-------------|-------------|
|          | mean difference | deviation | mean difference | deviation |
|          | gangue | coal | gangue | coal |
| Correlation | 0.016 | 0.035 | 0.027 | -0.007 | 0.018 | 0.017 |
| Homogeneity | 0.040 | 0.023 | 0.035 | 0.044 | 0.019 | 0.019 |
| Contrast   | -0.238 | 0.045 | 0.224 | -0.470 | 0.212 | 0.318 |
| Energy     | 0.018 | 0.030 | 0.036 | 0.038 | 0.023 | 0.018 |
| Gray mean  | 1.120 | 1.100 | 1.042 | 6.244 | 5.370 | 2.963 |

TABLE 3. Test results of visible light camera sample feature classifier.

| Features | Coal(+) | Gangue(-) | Total | Accuracy | Precision | Recall |
|----------|---------|-----------|-------|----------|-----------|--------|
|          | N(+) | TP | N(-) | TN |       |         |        |
| Correlation | 500 | 351 | 500 | 272 | 1000 | 0.623 | 0.606 | 0.702 |
| Homogeneity | 500 | 343 | 500 | 385 | 1000 | 0.728 | 0.749 | 0.686 |
| Contrast   | 500 | 364 | 500 | 436 | 1000 | 0.800 | 0.851 | 0.728 |
| Energy     | 500 | 290 | 500 | 264 | 1000 | 0.554 | 0.551 | 0.58  |
| Gray mean  | 500 | 386 | 500 | 337 | 1000 | 0.723 | 0.703 | 0.772 |

TABLE 4. Test results of near-infrared camera sample feature classifier.

| Features | Coal(+) | Gangue(-) | Total | Accuracy | Precision | Recall |
|----------|---------|-----------|-------|----------|-----------|--------|
|          | N(+) | TP | N(-) | TN |       |         |        |
| Correlation | 500 | 281 | 500 | 238 | 1000 | 0.519 | 0.517 | 0.562 |
| Homogeneity | 500 | 289 | 500 | 457 | 1000 | 0.746 | 0.870 | 0.578 |
| Contrast   | 500 | 408 | 500 | 436 | 1000 | 0.844 | 0.864 | 0.816 |
| Energy     | 500 | 433 | 500 | 419 | 1000 | 0.852 | 0.842 | 0.866 |
| Gray mean  | 500 | 350 | 500 | 331 | 1000 | 0.681 | 0.674 | 0.700 |
| Fusion     | 500 | 439 | 500 | 431 | 1000 | 0.870 | 0.864 | 0.878 |

were analyzed quantitatively. In addition, the classification performance of each feature was detected by the designed SVM classifier. The mean value and deviation of coal and gangue were calculated under the same cameras and the mean value difference was obtained by the difference in the mean value. The mean difference reflects the distribution difference between coal and gangue while the deviation measures the degree of deviation of the characteristic value from the arithmetic mean value. When the mean difference is greater and the deviation is smaller, the effect of distinguishing coal from gangue is better. The experimental results are shown in Tables 2 to 4.

First, according to the data analysis in Table 2, except for the correlation feature, the differences in mean value of the other features under the NIR camera were significantly greater than those under the VIS camera, which meant that the features of coal and gangue were distributed in different regions under the NIR camera. Therefore, the classifier of the NIR samples could find the threshold with segmentation effect and obtain better classification performance. Under the NIR camera, the deviations of the energy and homogeneity features were in a small range, which led to better stability of the classifier. Comparing the results of Table 3 and Table 4 and combining with the characteristic analysis in FIGURE 5 and FIGURE 6, it could be seen that under the NIR camera, the feature distribution of coal and gangue had more significant differences. In addition, the classifier effect obtained by the NIR camera was better. The accuracy of the classifier obtained by the energy and contrast feature reached 0.852 and 0.844, respectively, which were higher than the feature classifier under the VIS camera. And the accuracy and recall rate were also high. In particular, the homogeneity classifier was not as good as the energy in accuracy and recall rate. It had excellent performance in precision and overall comprehensive function. While in the VIS camera, the comprehensive performance of each classifier was poor. Although the accuracy rate of the homogeneity classifier reached 0.728, the recall rate was only 0.686. The recall rate of the gray classifier was 0.772, but the accuracy rate was only 0.723 and the precision was worse than that of the NIR camera. In summary, the separation method based on the NIR camera proposed in this paper was better than the traditional method based on the VIS camera. The NIR camera could obtain more effective features and train a classifier with higher accuracy.

In Table 4, a comprehensive comparison of the test results of each feature classifier shows that the energy and contrast feature classifiers had better performance in accuracy, precision, and recall rate. Although the recall rate of the homogeneity feature classifier was lower, the precision was better.
TABLE 5. Units for magnetic properties.

| Algorithms                        | Noise samples | Occlusion samples | Overall samples |
|-----------------------------------|---------------|-------------------|-----------------|
|                                   | Precision     | Recall            | Precision       | Recall       | P     | FPS  |
| SVM                               | 0.600         | 0.522             | 0.446           | 0.391        | 0.870 | 52   |
| Yolov3                            | 0.714         | 0.696             | 0.630           | 0.591        | 0.830 | 17   |
| Two-sample deep learning          | 0.867         | 0.678             | 0.636           | 0.670        | 0.901 | 15   |
| Fusion algorithms                 | 0.823         | 0.687             | 0.807           | 0.765        | 0.923 | 26   |

FIGURE 12. Detection effect of different algorithms; (a) SVM; (b) Yolov3; (c) two-sample deep learning; (d) fusion algorithms; the real label of the top-down target is: gangue, gangue, gangue, coal, and coal.

B. REAL-TIME COAL AND GANGUE DETECTION EXPERIMENT

The real-time detection experiment was divided into three groups. The first group selected the three features under the NIR camera of energy, contrast, and homogeneity for feature fusion to obtain the SVM fusion classifier. The second group selected the NIR samples to train the deep learning network. In the third group, the two-sample deep learning network was selected. First, the visible light sample was used to pretrain the network. Then, migration learning was performed on the VIS deep learning model and the NIR images were input for further training. The deep learning networks all used the modified YOLOv3 network. The fourth group selected the fusion algorithm for the experiments. $\varepsilon$ of experimental threshold was set to 0.01 and $\lambda$ was set to 1.5. Each group of experiments separately obtained the recognition results in the overall sample, noise sample, and occlusion sample. The overall samples were used to evaluate the overall performance of the algorithm. The noise samples and occlusion samples tested the algorithm’s ability to deal with the challenges in dynamic scenes. Four groups of experimental results were compared in Table 5 and FIGURE 12.

First, according to all the data in Table 5, the performance of each algorithm in noise samples and occlusion samples was lower than that of the synthesis samples, which showed that noise and occlusion were important factors affecting the classifier. Except that the recall rate of the occlusion sample in the two-sample deep learning was slightly higher than that of the noise sample, the accuracy and recall rate of the occlusion sample were lower. The complete features of the target could not be obtained under occlusion, which created a considerable challenge for the classifier. Then, by comparing the data of each group, it was found that the SVM algorithm could not effectively deal with noise and occlusion. The accuracy of the SVM algorithm in Table 4 was 0.864, while the accuracy of the noise samples in Table 5 was reduced to 0.600 and the accuracy of occlusion samples, 0.446, was worse. As can also be seen in FIGURE 12(a), SVM divided two close targets into one object. In addition, it is easy to divide the background area into objects because the SVM algorithm only relies on gray and texture features of samples. When occlusion occurs, the camera could not acquire effective features.

Compared with SVM, deep learning extracted deeper target features, which effectively distinguished the background and different objects. In the second group of experiments, the precision and recall rate of the noise samples and occlusion samples significantly improved. In FIGURE 12(b), the deep learning method eliminated most of the background area. However, at the same time, the detection target still contained other targets and the accuracy of the overall sample was 0.830, which was slightly lower than the SVM because insufficient training samples led to overfitting of the network.

The two-sample deep learning further expanded the number of samples and features. In training mode, VIS samples were trained first, and then NIR samples were trained.
Therefore, the performance of the network was mainly determined by the near-infrared samples. Moreover, the VIS samples were similar to the near-infrared samples. Through pretraining, the neutral performance became better and the network parameters were close to the optimal value in advance. The experimental results also verified the feasibility of two-sample transfer. The accuracy and precision of the third group of experiments improved compared with the second group. FIGURE 12 (c) shows the third experimental detection results. The results of target detection were basically the same as the actual object. The accuracy rate of the two-sample deep learning method reached 0.901. Except that the recall rate of noise samples was lower than that of traditional deep learning, other performances were better. However, it was found from Table 5 that the detection speed of deep learning was slow with 15 fps frame rate. The coal and gangue were in motion during the experiment. Because the speed of deep learning detection was slow, some targets were missed, which is the main reason for limiting the accuracy of detection. Moreover, the above algorithms could not effectively deal with the problem of occlusion and attachment.

Therefore, in the fourth set of experiments, the SVM algorithm and the deep-learning algorithm were integrated and the length suppression algorithm was introduced. It can be seen from the experimental results that the accuracy and recall rate of the occlusion samples were greatly improved at the expense of little loss of accuracy on the noise samples. In addition, the detection rate of the algorithm after fusion was accelerated. Because the fusion algorithm takes the SVM algorithm as the main body and suppresses the deep learning algorithm to a certain extent, the accuracy in the noise sample was reduced compared with the two-sample method. However, the length suppression algorithm in the fusion algorithm effectively overcame the problem of occlusion, increasing the accuracy of the occlusion samples to 0.807, the recall rate to 0.765, and the overall sample accuracy rate of 0.923, which were the highest among all the algorithms. In addition, the detection speed of the fusion algorithm was increased to 26 fps, which was much higher than the deep learning algorithm. In conclusion, the fusion algorithm combined the advantages of a fast speed of the SVM and the detection accuracy of deep learning.

To further verify the accuracy and robustness of the algorithm proposed in this paper, the algorithm was tested on the actual moving belt. There were 500 test targets, including 250 gangue and 250 coal. All targets were randomly poured into the roller screen. The accuracy of total detection was calculated by 10 targets at every interval and a total of 50 sets of data were recorded. The variation in accuracy with the sample is shown in FIGURE 13.

The experimental environment was the same as a real coal mine environment. The belt speed was 2.5 m/s. Coal and gangue fell on the moving belt after passing through the roller screen. The position, posture, and clearance were all random. It can be seen from the data in FIGURE 13 that the detection accuracy basically remained within the interval of 0.90 to 0.93. Experiments showed that the algorithm model could satisfy the complex environment of the actual coal mine. It detected targets in real time and had good robustness.

In addition, three different experiments were added to test the stability of the system in different environments. The first experiment used a black belt. The angle of the light was 45 degrees. In the second set of experiments, the belt was replaced with a green belt. In the third set of experiments, image enhancement was performed on the training samples. In the fourth and fifth sets of experiments, the angle of the light was adjusted. Each group of experimental test sample was the same with a total of 500. The accuracy rates in different environments were shown in Table 6.

| No. | Belt color | Whether to enhance | Lighting angle/degree | Accuracy |
|-----|------------|--------------------|-----------------------|----------|
| 1   | black      | N                  | 45                    | 0.923    |
| 2   | green      | N                  | 45                    | 0.918    |
| 3   | black      | Y                  | 45                    | 0.935    |
| 4   | black      | N                  | 15                    | 0.857    |
| 5   | black      | N                  | 75                    | 0.884    |

By comparing the first group and the second group, the background had no effect on the algorithm. The NIR camera was not sensitive to the color and material of the belt. Comparing the first group with the fourth group and the fifth group, the change of the light would have a greater impact on the experimental results. The effect is better only at 45 degrees. When the light on the surface of the object was insufficient, the target surface features were not obvious, which reduced the detection accuracy. On the contrary, comparing the first group with the third group, the characteristics of the target were enhanced through image enhancement. This made it easier for the classifier to find the difference between coal and gangue.

C. COMPARISON WITH OTHER ALGORITHMS

For verifying the effectiveness of the algorithm in this paper, different recognition algorithms of coal and gangue were
To ensure the fairness of the experiment, the training samples were reduced to 300 representatives.

### TABLE 7: Units for magnetic properties.

| Author | Feature                        | Algorithm | Samples | Accuracy |
|--------|--------------------------------|-----------|---------|----------|
| Dou    | color and textural              | SVM       | 560     | 0.926    |
| Wang   | bioriental wavelet              | SVM       | 150     | 0.951    |
| Pu     | transfer learning              | Deep learning | 240  | 0.825    |
| Ours   | grayscale, and transfer learning| Deep learning | 300  | 0.901    |

Dou et al. [15] and Wang et al. [25] both used the SVM algorithm. Dou extracted 19 features of coal and gangue pictures, including color and textural features. He employed the relief-SVM method to identify optimal features and construct optimal classifiers. This method had a complicated extraction process for many features, and did not consider the relationship between the features. Wang used bioriental wavelet to transform the coal and gangue pictures, and used K-fold cross-validation to optimize the penalty factor and kernel function coefficient of the SVM mathematical model.

In his paper, the training accuracy rate was 100% and the number of test sets was only forty. Although the test accuracy rate of this algorithm was as high as 0.951. It may fail in other complex environments. Pu et al. [20] introduced the idea of transfer learning to build a custom CNN model, which solved the problem of massive trainable parameters and limited computing power linked to the building of a brand-new model from scratch. The deep-learning model converged in advance during the training process, so it did not fully release its potential, so the accuracy rate was only 0.825. If a larger database is used, the accuracy rate should improve. Although all of the above methods showed good recognition ability, they did not consider the challenges created by dynamic sorting in the actual coal industry. This paper combined gray and texture in features and integrated SVM and deep learning in algorithms.

IV. CONCLUSION

As an important part of the coal mining industry, sorting coal and gangue has gradually developed from traditional manual to intelligent sorting, which has greatly improved the efficiency. Aiming at the problems of low recognition accuracy and the current complex sorting method in dynamic environments, the real-time multilevel fusion recognition system proposed in this paper effectively overcame the interference of the dynamic environment and improved the recognition performance. First, the experiment in this paper proved that the NIR camera had a better ability to distinguish the features of coal and gangue. In detail, a binocular fusion system was established and features under different sensors were analyzed to select effective features of SVM. Then, the VIS samples and the NIR samples were fused in the coal mine field for the first time to increase the robustness of the network to various complex environments. A new deep learning training method of double sample fusion was proposed, which not only increased the number of samples but also enriched the diversity of features to improve accuracy. Finally, two different algorithms were fused to establish a set of fast detection strategies, which satisfied the real-time requirement of the actual industry. In addition, a length suppression algorithm was introduced to solve the most common problems of occlusion in the detection field. Overall, the sorting system in this paper satisfied the real-time, robust, and economical requirements of the coal industry through innovations in sensors and methods.

However, there are still some defects that need to be improved. First, the environment of the coal mine is unique. It was not analyzed whether environmental changes such as humidity and temperature in the mine would affect the characteristics of coal gangue. In addition, the actual coal composition on the coal mine site is more complex. And there is “medium coal”: the mixture of coal and gangue. The surface of “medium coal” contains the characteristics of coal as well as gangue. It is difficult for this kind of coal to be sorted successfully. Second, in terms of algorithm fusion, the time complexity of the algorithm was not calculated in detail because we focused on solving engineering problems. Only FPS was used to judge the runtime. And due to the “black box” nature of deep learning, we did not make an in-depth analysis of the deep learning network structure. If the network structure could be modified according to the characteristics of coal and gangue, the network sorting performance can be further improved.

REFERENCES

[1] S. Dai, X. Zheng, X. Wang, R. B. Finkelman, Y. Jiang, D. Ren, X. Yan, and Y. Zhou. “Stone coal in China: A review,” *Int. Geol. Rev.*, vol. 60, nos. 5–6, pp. 736–753, Apr. 2018.
[2] H.-D. Lv, J.-S. Zhou, L. Yang, Y.-M. Li, L. Liu, “An accounting of the external environmental costs of coal in inner Mongolia using the pollution damage method,” Environ. Develop., Sustainability, vol. 22, pp. 1299–1321, Sep. 2018.

[3] M. Ahmaruzzaman, “A review on the utilization of fly ash,” Prog. Energy Combustion Sci., vol. 36, no. 3, pp. 327–363, 2010.

[4] E. J. Meyer and I. K. Craig, “Dynamic model for a dense medium drum separator in coal beneficiation,” Minerals Eng., vol. 77, pp. 78–85, Jun. 2015.

[5] H. A. M. Myburgh, “The influence of control and mechanical conditions of certain parameters on jiggings,” J. Southern Afr. Inst. Mining Metall., vol. 110, no. 11, pp. 655–661, 2010.

[6] Y. Zhou, B. Albijanic, B. Tadesse, Y. Wang, and J. Yang, “Investigation of bubble–particle attachment interaction during flotation,” Minerals Eng., vol. 133, pp. 91–94, Mar. 2019.

[7] C. Robben, J. de Korte, H. Wotрубa, and M. Robben, “Experiences in dry coarse coal separation using X-ray-transmission-based sorting,” Int. J. Coal Preparation Utilization, vol. 34, nos. 3–4, pp. 210–219, Jul. 2014.

[8] N. Zhang and C. Liu, “Radiation characteristics of natural gamma-ray from coal and gangue for recognition in top coal caving,” Sci. Rep., vol. 8, no. 1, p. 190, Dec. 2018.

[9] X. Zhu, J. Xiong, and Q. Liang, “Fault diagnosis of rotation machinery based on support vector machine optimized by quantum genetic algorithm,” IEEE Access, vol. 6, pp. 33583–33588, 2018.

[10] D. Dou, D. Zhou, J. Yang, and Y. Zhang, “Coal and gangue recognition under four operating conditions by using image analysis and relief-SVM,” Int. J. Coal Preparation Utilization, vol. 40, no. 7, pp. 1–10, 2018.

[11] J. D. Boice, K. D. Held, and R. E. Shore, “Radiation epidemiology and health effects following low-level radiation exposure,” J. Radiol. Protection, vol. 39, no. 4, pp. S14–S27, Dec. 2019.

[12] Z. Zhang and J. Yang, “The density fraction estimation of coarse coal by use of the kernel method and machine vision,” Energy Source, Part A, Recovery, Utilization, Environ. Effects, vol. 37, no. 2, pp. 181–191, Jan. 2015.

[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. Gao, C. Du, H. Wang, and S. Zhang, “An efficient of coal and gangue recognition algorithm,” Int. J. Signal Process., Image Process. Pattern Recognit., vol. 6, no. 4, pp. 345–354, 2013.

[15] D. Dou, W. Wu, J. Yang, and Y. Zhang, “Classification of coal and gangue under multiple surface conditions via machine vision and relief-SVM,” Powder Technol., vol. 356, pp. 1024–1028, Nov. 2019.

[16] Z. Sun, W. Lu, P. Xuan, H. Li, S. Zhang, S. Niu, and R. Jia, “Separation of gangue from coal based on supplementary vector by morphology,” Int. J. Coal Preparation Utilization, pp. 1–17, Mar. 2019.

[17] C. A. Perez, P. A. Estévez, P. A. Vera, L. E. Castillo, C. M. Aravena, D. A. Schulz, and L. E. Medina, “Ore grade estimation by feature selection and voting using boundary detection in digital image analysis,” Int. J. Mineral Process., vol. 101, nos. 1–4, pp. 28–36, Nov. 2011.

[18] H. Liang, H. Cheng, T. Ma, Z. Pang, and Y. Zhong, “Identification of coal and gangue by self-organizing competitive neural network and SVM,” in Proc. 2nd Int. Conf. Intell. Hum.-Mach. Syst. Cybern., vol. 2, Aug. 2010, pp. 41–45.

[19] W. Li, Y. Wang, B. Fu, and Y. Lin, “Coal and coal gangue separation based on computer vision,” in Proc. 5th Int. Conf. Frontier Comput. Sci. Technol., Aug. 2010, pp. 467–472.

[20] Y. Pu, D. B. Apel, A. Szmigiel, and J. Chen, “Image recognition of coal and coal gangue using a convolutional neural network and transfer learning,” Energies, vol. 12, no. 9, p. 1735, May 2019.

[21] W. Lai, M. Zhou, F. Hu, K. Bian, and H. Song, “A study of multispectral technology and two-dimension autoencoder for coal and gangue recognition,” IEEE Access, vol. 8, pp. 61834–61843, 2020.

[22] 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.

[23] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, “Do we need hundreds of classifiers to solve real world classification problems?” J. Mach. Learn. Res., vol. 15, no. 1, pp. 3133–3181, 2014.

[24] A. Farhadi, “YOLOv3: An incremental improvement.”

[25] W. Wang, Z. Lv, and H. Lu, “Research on methods to differentiate coal and gangue using image processing and a support vector machine,” Int. J. Coal Preparation Utilization pp. 1–14, Jun. 2018.