IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 5, 2019, accepted December 12, 2019, date of publication December 19, 2019, date of current version December 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2961075

An Image-Based Hierarchical Deep Learning Framework for Coal and Gangue Detection

DONGJUN LI[1], ZHENXIN ZHANG[2], ZHIHUA XU[3,4], LILI XU[5], GUOYING MENG[1], ZHEN LI[2], AND SIYUN CHEN[2]

[1]School of Mechanical, Electronic, and Information Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China
[2]Beijing Advanced Innovation Center for Imaging Theory and Technology, Capital Normal University, Beijing 100048, China
[3]State Key Laboratory of Coal Resources and Safe Mining, China University of Mining and Technology (Beijing), Beijing 100083, China
[4]College of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China
[5]Community Service and Management, University of Science and Technology Beijing, Beijing 100083, China

Corresponding author: Zhenxin Zhang (zhangzhx@cnu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFC0600907, in part by the National Natural Science Foundation of China under Grant 41701533 and Grant 41701534, in part by the Open Fund for Guangdong Key Laboratory of Ocean Remote Sensing (South China Sea Institute of Oceanology Chinese Academy of Sciences) under Grant 2017B030301005-LORS1804, in part by the Open Fund of the State Key Laboratory of Remote Sensing Science under Grant OFSLRSS201818, in part by the Open Fund Key Laboratory for National Geography State Monitoring (National Administration of Surveying, Mapping and Geoinformation) under Grant 2017NGCM06, and in part by the Open Fund of State Key Laboratory of Coal Resources and Safe Mining under Grant SKLCRSM19KFA01.

ABSTRACT The efficient separation of coal and gangue in the mining process is of great significance for improving coal mining efficiency and reducing environmental pollution. Automatic detection of coal and gangue is the key and foundation for the separation of coal and gangue. In this paper, we proposed a hierarchical framework for coal and gangue detection based on deep learning models. In this framework, the Gaussian pyramid principle is first used to construct multi-level training data, leading to the sets of coal and gangue image features with multiple scales. Then, the coal and gangue regional proposal networks (CG-RPN) are designed to determine the candidate regions of the target objects in the image. Next, convolution neural networks (CNNs) are constructed to recognize coal and gangue objects on the basis of extracted candidate regions. We performed our method on three different datasets. Experimental results showed that the proposed method improves the detection accuracy of coal and gangue objects by 0.8% compared with the previous methods, reaching up to 98.33%. In addition, our proposed method enables the detection of multiple coal and gangue objects in an individual image and solves the problem of queuing requirements in traditional methods.

INDEX TERMS Coal and gangue, detection, deep learning, CNNs, multi-level training data.

I. INTRODUCTION

The automatic and efficient separation of coal and others in the mining process is of great significance for improving coal mining efficiency [1], [2]. Gangue is the main solid waste generated by coal mining. The accumulation of gangue leads to the waste of land resources and environmental damage [3], [4]. Therefore, the automatic detection of coal and gangue plays an important role in gangue-caused environment pollution [5], mineral processing automation [6], digital mining [7], and other applications [8].

The detection of coal and gangue objects from optical images is a cost-effective way for mineral processing [9], [10]. The image of coal and gangue has abundant and important characteristic information, such as color, grayscale, texture, etc. However, due to the illumination change and background color difference, the shape scales of coal and gangue objects are uneven, posing challenges in extracting stable features for recognizing coal and gangue objects from optical images. In addition, the arrangement of coal and gangue on the conveyor belt is complex and irregular. The method which combines traditional image features [11] or measures density [12] is only suitable to neatly arranged coal and gangue detection, which leads to certain limitations. How to simultaneously and accurately detect coal and gangue is still a problem.

In the field of object feature extraction, the traditional machine learning methods have achieved certain advanced

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

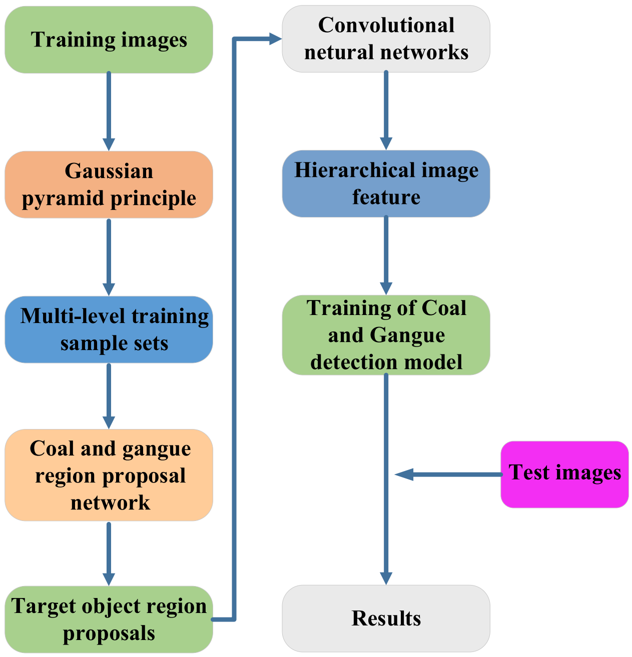


FIGURE 1. Overview of our method.

results in unsupervised classification task by using the priori knowledge and the intrinsic multiple relationships of images [13], [14]. With the development of computing devices and computer vision technology, the advancement of deep learning theory has been demonstrated to be more effective for object detection [15]. In particular, the combination of shallow convolutional and deep convolutional layers enables to extract stable image features, thus, to improve the recognition accuracy [16]. In addition, for a large number of image data, the way of combining the weakly supervised learning (WSL) and high-level feature learning can substantially improve the detection efficiency [17]. For the problem of object rotation variations, the performance of object detection can be improved by introducing and learning a new rotation-invariant layer based on the existing CNN architectures [18].

This paper aims to propose a hierarchical deep learning framework to detect coal and gangue objects from optical images (Fig.1). The proposed framework consists of two main phases. In training phase, we first constructed the multi-level training samples using the Gaussian pyramid principle. Then, we built a coal and gangue suggestion regional network (CG-RPN) to identify the regions of candidate objects. This configuration improved the time efficiency and simultaneously displayed the position information of the plurality of objects to be detected. Finally, we used the convolutional neural networks (CNNs) and the pre-extracted multi-level features to create multi-level coal and gangue detection models. In test phase, the learned CNNs were used to detect coal and gangue objects. Similar to Faster R-CNN [19], our proposed object detection method of coal and gangue adopts the strategy of generating candidate regions recommendation to improve the detection efficiency. Different from the ignoring hierarchical space information in Faster R-CNN which only utilized the single scale of training image, our proposed method is to construct the network by combining the spatial hierarchy structure and deep learning model of multi-level training data set. In addition, we used the coal and gangue image datasets to train the network and optimize network parameters, and improved the detection network's ability to detect two specific targets of coal and gangue.

The main contributions of this paper are summarized as follows:

(1) We proposed a multi-level learning framework for coal and gangue detection based on CNN model. It enables to extract the image features at multiple scales and to train deep learning model for obtaining high detection accuracy.

(2) The CG-RPN is constructed to identify the candidate regions of the target objects, which improves the time efficiency and detection accuracy. Moreover, the target objects of a plurality of random positions in an individual image can be identified at the same time, which solves the problem that the coal to be identified and the gangue are queued in the front section in the traditional detection.

II. RELATED WORK

A. COAL AND GANGUE SEPARATION

In coal preparation technology, in addition to the manual operation of separation between coal and gangue materials, automated sorting techniques can be summarized into two categories based on whether water is used or not, i.e., wet selection and dry selection. Representative techniques for wet selection include dynamic jump screening [20] and heavy media selection [21]. This practice is efficient and mature, but it requires a large amount of land resources, water resources, and generates a large amount of slime, which damages to the environment. In contrast, dry selection techniques require no water, and have the advantages of simple structure, cost-effective, and no harmful waste, which have recently became a hot research field [22]. The representative techniques for dry selection are ray-based, i.e. dual-energy gamma ray [23], X-ray [24], density-based [12], and image-based [25]. Among them, the ray-based methods can achieve high accuracy, but the radiation sources should be carefully managed to avoid the damage to human body. Density-based method requires dynamic weighing and measurement of the target objects. Before detection, coal and gangue should be neatly arranged, which leads to a complex system and unpredictable measurement accuracy. Image-based method only needs to obtain the image of object, the structure is simple, and it will not cause harm to the human body and the environment, and it does not need to arrange coal and gangue neatly before detection. Therefore, selection based on image detection is a promising trend for coal and gangue sorting [4].

B. IMAGE-BASED DETECTION OF COAL AND GANGUE

In general, the detection of coal and gangue objects from optical images involves two main challenges, i.e. feature extraction and model construction. One practice is to use grayscale features to detect coal and gangue objects from optical images [25]. This practice, however, results in low recognition accuracy. In addition to grayscale features, texture features, i.e. co-occurrence matrix [26]–[28] and fractal dimension [29], [30] are alternative features for coal and gangue identification. Nevertheless, the coal and gangue images acquired under poor working conditions have similar appearance and are greatly affected by ambient illumination changes, and it is difficult to extract features of target objects in the image, so the recognition rate is limited. For model construction, the models, i.e. support vector machine (SVM) and Bayesian models, are usually used. These methods improve the detection accuracy through different feature combinations and parameter optimization methods, but only pay attention to the low-level semantic features of the image, and the simplicity of the model structure makes it difficult to extract more comprehensive image features.

C. DEEP LEARNING FOR OBJECT DETECTION

Deep learning has been widely studied and used in the field of image classification due to its excellent feature representation ability. Compared with traditional image object detection and recognition techniques, deep learning methods do not require the extraction of hand-crafted image features. Simultaneously, deep learning methods can abstract deep features of image information, which has better ability for object detection [31]. The earlier convolutional neural network model is LeNet [32] proposed by LeCun, which was designed to solve the problem of handwritten digital recognition. The deep learning theory was first proposed by Geoffrey Hinton in 2006 [15], giving a solution to the gradient disappearance problem in deep neural network training. In 2012, the model of AlexNet [33] used ReLU as the activation function of CNN, and adopted dropout strategy to randomly ignore a part of neurons during network training to avoid over-fitting of the model. VGG [34] model can express more powerful features of input data through stacking small convolution kernel and pooling layer. GoogleNet [35] increases the depth and breadth of the network to improve the performance of deep convolutional networks, by using the Inception module. ResNet [36] solves the problem that the network cannot effectively converge as the network deepens by adding directly-connected channels to the network. These frameworks achieve higher recognition accuracy in the field of image recognition by increasing the depth of the network and using smaller convolution kernels. The above methods have also been applied to the identification of coal and gangue. Su *et al.* [37] proposed a simple LeNet-5 deep learning model for coal and gangue image classification. Compared with the traditional method, this model enables to extract image features without artificial selection process, and the recognition accuracy is improved. Hong *et al.* [38] designed a convolutional neural network based on AlexNet and proposed a method of image recognition and region clipping. The above methods are only suitable to identify a single target object in a single picture.

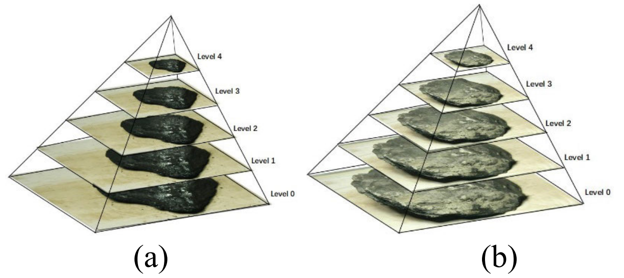


FIGURE 2. Multi-level training data: (a) coal and (b) gangue.

For multiple coals and gangues in an individual image, not only each target needs to be identified but also the positional information in the image needs to be detected.

For the field of object detection, deep learning in object detection methods based on candidate regions combined with CNN have been becoming more and more popular in recent years, such as R-CNN [39], Fast R-CNN [40] and Faster R-CNN [19]. R-CNN selects many candidate boxes by using selective search method, then, performs convolutional operations for each candidate region to extract features. Finally, the extracted convolutional features are input in the SVM classifier and regressed the bounding-box to determine the target category and location. Fast R-CNN uses a multi-task loss function, which makes the training and testing of the network convenient and improves the detection accuracy. However, the extraction of region proposal uses selective search, so this network does not implement end-to-end training and testing. In order to solve this problem, Faster R-CNN convolves the entire input image and the feature map is directly input to the RPN network. This approach enables end-to-end training and testing. Compared to the previous method, the efficiency of detection is improved.

III. THE PROPOSED METHOD

A. CONSTRUCTION OF MULTI-LEVEL TRAINING DATA

As the coal and gangue objects in the acquired image differ in size and shape, and the adjacent objects may be partially blocked each other, it is difficult to construct an efficient image features for the detection process. Therefore, it is necessary to obtain more stable image characteristics of coal and gangue to improve the generalization ability of the model, to meet the detection requirements in images with different sizes and shapes of coal and gangue.

For solving the above problems, one practice is to construct image pyramid to transfer the image into multiple resolution structure [31]. Thus, an original image can be generated into N different resolutions of images, and the image with the highest resolution (original image) is placed at the bottom (see Fig. 2). Then the other images are arranged on basis of the original image to form a pyramid according to resolutions from highest to lowest, and the upward is a series of pixels or images whose dimensions are gradually reduced. In this paper, we constructed a multi-level training samples by subsampling the original coal and gangue image, including

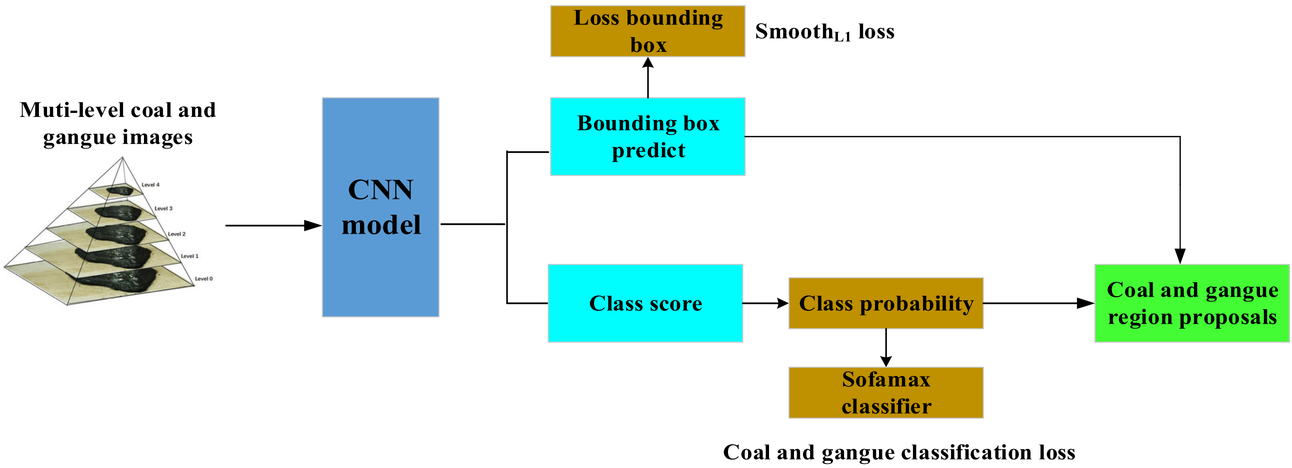


FIGURE 3. Process of constructing target object candidate regions.

the following steps: First, the sampled picture of the coal and gangue is smoothed by low-pass filter, and the automatic specification is 1024 × 1024 pixels. These images are then down sampled by Gaussian kernel image convolution, and multi-level training samples are generated step by step.

We resampled the image at (l+1) level (l = 0, 1, 2 ...N − 1, N represents the number of layers in the hierarchical training dataset, as shown in Figure 2, N = 5) as an example. Describe the process of Gaussian kernel convolution operations. After the low-pass filtering smoothing of the l^{th} layer image, the image is sampled as follows:

$$K_{l+1}(i,j) = \sum_{m=-2}^{2}\sum_{n=-2}^{2} p(m,n)K_l(2i-m, 2j-n) \tag{1}$$

where $K_l(\cdot)$ represents the l^{th} level image of the training dataset and $K_{l+1}(\cdot)$ is resampled from $K_l(\cdot)$. The $p(m, n) = k(m)k(n)$ is a 5 × 5 pixels window function with low-pass filtering characteristic as a Gaussian convolutional kernel, and the function $k(\cdot)$ is Gaussian density distribution function, which is

$$p(m,n) = \frac{e^{-(m^2+n^2)/2\sigma^2}}{2\pi\sigma^2}. \tag{2}$$

According to the above method, multilevel training images $K_0, K_1, \ldots, K_{N-1}$ can be created to form a multi-level structure of training data.

B. NETWORK STRUCTURE

1) CONSTRUCTION OF CG-RPN

In order to effectively detect the coal and gangue, we design coal and gangue region proposal networks to generate the target area of candidate coal and gangue. The following is a detailed description of the composition of CG-RPN.

The traditional regional proposal networks (RPNs) takes the original image as inputs and RPN outputs a set of regional proposals. At the same time, each region proposal is accompanied by an object score. This process is implemented by a fully convolutional network (FCN). The selection of candidate regions is usually carried out by sliding window, and the feature maps generated by the last shared convolutional layer are mapped to low-dimensional feature maps. Additionally, these feature maps are taken as the input for two fully connected layers: (1) A box-regression layer (reg) is used to predict the coordinate (x, y) corresponding to the central anchor point of the object detection box and the width and height of the box is (w, h); (2) A box-classification layer (cls) is used to determine whether the proposal is a target object or a background. This method needs the sliding window to traverse the whole image, which requires high computing time cost and has an adverse impact on the detection efficiency. This method requires manual operation to set the size and proportion of sliding window, so it increases the complexity of coal and gangue detection in the image.

In CG-RPN, the CNNs model is used to extract the candidate target area, which can effectively generate a small amount of high-quality coal and gangue candidate target area [19], and reduce the calculation time [16]. Firstly, the convolutional features of multi-level training dataset are extracted by a set of shared convolution layers, such as AlexNet, ZF, VGG-16, and Inception-v3. The input coal and gangue images can be extracted into a 512-dimensional convolutional feature (shared feature map) by the last layer of the CNN model. Then, the extracted shared feature maps are used in the reconstruction of region proposal networks, in which there are two parts (Fig. 3). One part is the bounding box which predicts the regression value of the location of the target object area box and obtains the location parameter information of the predicted target area recommendations; The other part is class score which is used to predict the probability of detecting an object as a target by calculating the intersection over union (IoU) between the initial target region and the marked region in samples.

The size and shape of the coal and gangue in the image are not uniform, so we specifically constructed the multi-scale object detection box for the coal and gangue as 9 rectangular

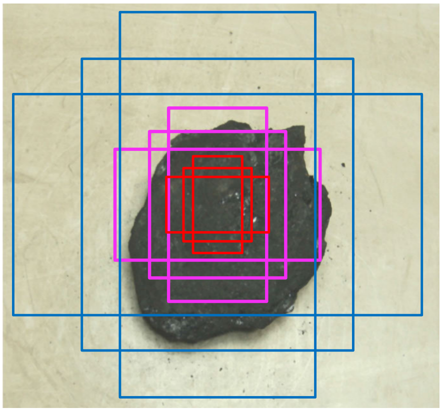


FIGURE 4. CNN feature maps of different sizes of initial sliding windows, with three different aspect ratios and sizes. The three colors of red, purple, and blue represent three scales or sizes: 128^2, 256^2 and 512^2. The area frame of each color has three types of height-to-width ratios with 1:1, 1:2 and 2:1.

sliding windows with three sizes of pixel (128, 256 and 512). Each rectangular box further includes three aspect ratios of 1:1, 1:2 and 2:1, as shown in Fig. 4.

We used the above nine rectangular sliding windows as the initial detection box [34]. Each size of window indicates a candidate target object area. If an object detection box tightly covers a target object, the box should be a positive example. Conversely, it should be a negative example. However, marking object detection boxes that partially overlap with objects is not easy. IoU measurement is introduced to solve this problem [40]. Selecting the appropriate IoU threshold can help to select a less number of candidate target object regions with high quality and basically cover all detected objects. Then, according to the regression mapping relation, the regression value of the target object detection box is calculated, and the position and size of each target object detection box are modified. The mapping relationship is defined by the position translation and scaling dimensions of the target object detection box.

The mapping parameters are shown as follows:

$$\begin{aligned} t_x &= (x - x_a)/w_a, \quad t_y = (y - y_a)/h_a \\ t_w &= \log(w/w_a), \quad t_h = \log(h/h_a) \\ t_x^* &= (x^* - x_a)/w_a, \quad t_y^* = (y^* - y_a)/y_a \\ t_w^* &= \log(w^*/w_a), \quad t_h^* = \log(h^*/h_a). \end{aligned} \tag{3}$$

where the parameters x, y, w, and h represent the centroid position, width, and height of the object detection bounding box respectively. The variables x, x_a, and x^* are denoted the horizontal coordinate values of center point in the predicted object detection box, initial object detection box and ground-truth box respectively. The y, y_a, and y^* are vertical coordinates of the above boxes respectively. The parameters w, w_a, and w^* represent the widths of the predicted object detection box, initial object detection box and ground-truth box respectively. The values t_x, t_y, t_w, and t_h represents four parameterized coordinates of the predicted object detection box. The values t_x^*, t_y^*, t_w^*, and t_h^* represent the coordinate position of ground truth bounding box corresponding to positive anchor.

In the detection process of multiple coal and gangue target objects in a single image, CG-RPN can be used to generate multi-scale target detection boxes for input images. Selecting the appropriate IoU threshold can help to select a small number of high-quality candidate target object regions and basically cover all detected objects. According to our experience, we set the IoU value as 0.7 [41] to determine the desired target detection area proposals (as shown in Fig. 5).

The loss function training process of CG-RPN is divided into two parts, one is classified loss, the other is the loss of bounding-box regression, which is defined in Eq. 4.

$$L(p_i, t_i) = \sum\nolimits_i L_{cls}(p_i, p_i^*) + \sum\nolimits_i p_i^* L_{loc}(t_i, t_i^*) \tag{4}$$

where i is the serial number of an object detection box in a batch and p_i is the predicted probability of object detection box i as an object. The ground-truth label p_i^* is equal to 1 if the object detection box is positive, and equal to 0 if the object detection box is negative. Variable t_i is a vector representing the parameters of the predicted bounding box, and the ground-truth box of a positive anchor box is represented by t_i^*. The classification loss function L_{cls} is a log loss over two categories (target object vs. no target object). The definition of L_{cls} is showed as follows:

$$L_{cls}\left(p_i, p_i^*\right) = \begin{cases} -\log p_i & if \ \ p_i^* = 1 \\ -\log\left(1 - \mathrm{p_i}\right) & if \ \ p_i^* = 0\,. \end{cases} \tag{5}$$

The regression loss (L_{loc}) is calculated by a classical loss function ($Smooth_{L1}$), which is defined as follows:

$$L_{loc}\left(t_i, t_i^*\right) = Smooth_{L1}(t_i - t_i^*). \tag{6}$$

And the $Smooth_{L1}$ function is a nonlinear regression function defined as:

$$Smooth_{L1}\left(a\right) = \begin{cases} 0.5a^2 & if \ \ |a| < 1 \\ |a| - 0.5 & otherwise, \end{cases} \tag{7}$$

where the parameter α represents the regression argument of $(t_i - t_i^*)$. In the optimization process, the loss function gradually approaches its minimum value and the parameters of the training network are obtained.

2) FEATURE EXTRACTION

Different from manually selected features (such as color, intensity, and edge direction) that are based on limited human knowledge, deep learning model can automatically extract more powerful different levels of features by using multi-level convolutional layers. For example, a shallow convolution layer can extract information about the edge contour, and a deep convolution layer can extract more abstract features of coal and gangue. Finally, it combines low-level features with more abstract high-level features to find estimable relationships in a given data set.

The process of feature extraction is as follows. The shared feature map extracted from the multilevel training dataset

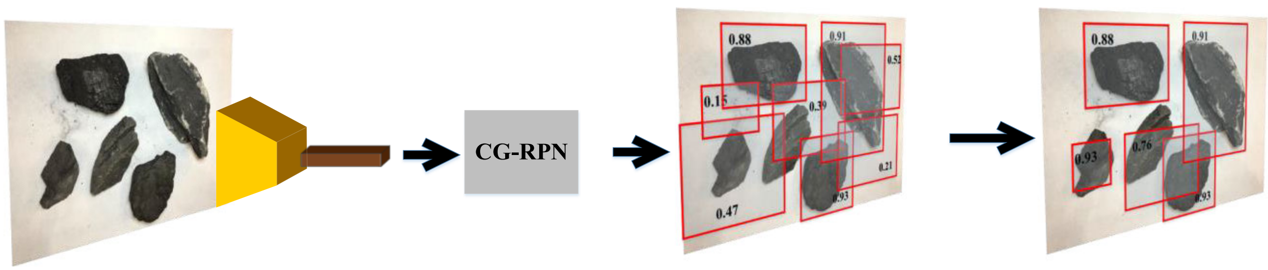


FIGURE 5. Schematic representation of the candidate regions of coal and gangue.

Algorithm 1 Process of Feature Extraction

Input: The training images consisting of hierarchical coal and gangues images.

1 The shared convolutional layers extract the features of the training images to obtain shared feature maps.

2 The feature extraction layers extract deeper feature from shared feature maps.

3 The ROI pooling layer converts the extracted features into a list of feature vectors.

Set $i = 1$;

while $i \leq$ max iteration **do**

  Perform feedforward pass process;

  Update weights of model through Adam optimizer [42] and backpropagate error;

  $i = i+1$;

  if training accuracy converges, then

    break;

end

Output: the extracted features.

by the shared convolutional layer is passed to the feature extraction layer and transmitted to the CG-RPN model. The region of interest pooling layer is set before the last fully connected layer [34]. The region of interest pooling layer is used to receive the region of interest pooling list generated by the CG-RPN, and convert the different sizes of feature map extracted by the CNN into fixed-length feature vectors. So the vector input to the fully connected layer is a feature vector of a fixed dimension (512 dimension) [42].

3) TRAINING PROCESS

Inspired by the transfer learning techniques, the pre-training model can improve the training efficiency in the image detection of coal and gangue. By keeping the structure of original model, the weight and bias can be corrected continuously in many iterations to obtain better training effects. Here we used the pre-training model as the feature extractor, and removed the output layer, and then employed the rest of the network as a fixed feature extractor to apply to our datasets.

In the process of model training, the weights and biases are updated by the learning strategies of back-propagation and stochastic gradient descent (SGD). According to the Gaussian pyramid principle, we constructed a multi-level training dataset and used the multi-level training dataset with labels to learn the deep model. The training process is as follows (**Algorithm 1**): we firstly used the ImageNet-based pre-training VGG-16 model to initialize the CG-RPN model, in order to prevent network over-fitting, and then the trained CG-RPN model generated the target object area proposals. We exploited the same pre-training model to initialize the detection network and trained the detection network based on the object area proposal obtained from the CG-RPN model. The parameters of the shared convolution layer are obtained by optimizing the classifier and detection box parameters [41].

IV. COAL AND GANGUE DETECTION

In this section, we used the trained deep learning model to detect the coal and gangue in image datasets. The process is shown in Fig. 6. First of all, the shared feature maps are extracted from test images by CNN model. Simultaneously, the candidate target object detection proposal is generated from images by using CG-RPN model, and these proposals are shared with the above shared feature maps. Then, features of each target object detection region proposal are extracted by convolutional layer and are mapped into ROI pooling layer, so as to obtain fixed size of feature vectors and bounding box list of target object detection region. Coal and gangue are marked by feature vectors in fully connected layers and softmax classification layer. Finally, the accurate detection results of coal and gangue in the images can be obtained by correcting the location parameters of the predicted target object detection box in the list of the bounding box with the target object bound box regression layer [39].

V. EXPERIMENT AND RESULTS

We conducted a comprehensive experiment for evaluating the effectiveness of our method. We tested the sensitivities of the model parameters and compared our method with other methods, to verify the efficiency and stability of the proposed model. The computer configuration environment was as follows: Intel(R) Xeon(R) E5-2620 v4 CPU, total eight of Nvidia TITAN XP GPUs with each have 12-GB RAM. The training process was performed by TensorFlow [43] on Ubuntu14.04 operating system.

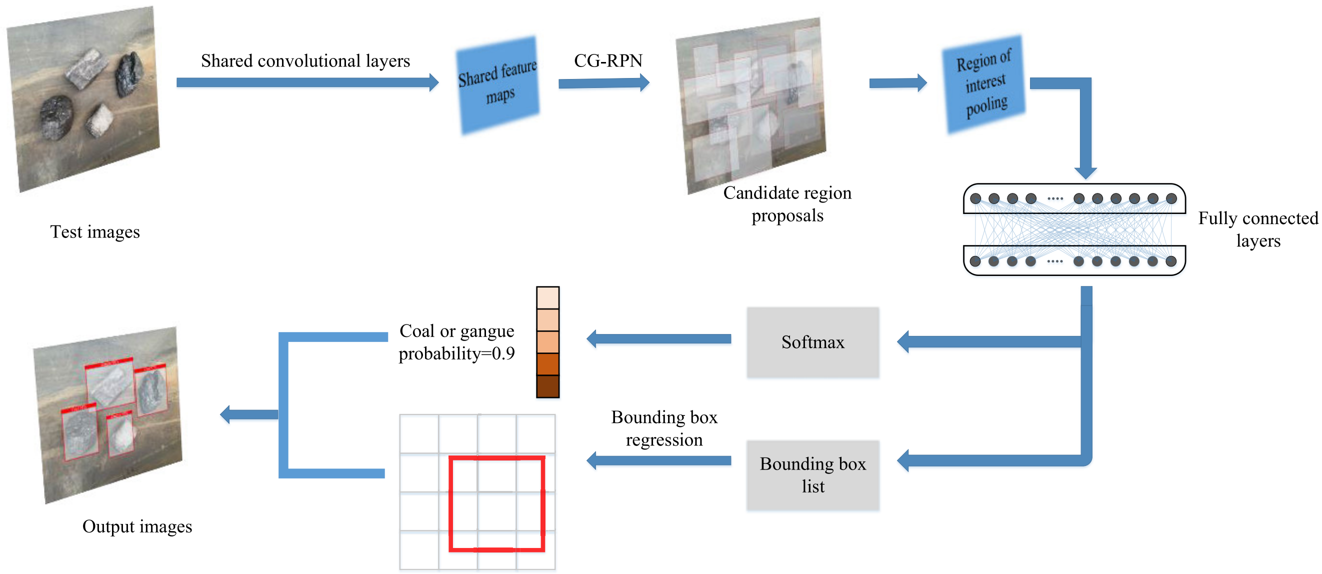


FIGURE 6. Process of coal and gangue detection.

A. DATASETS

For qualitative analysis, we used three datasets named Dataset I, Dataset II, and Dataset III, to test the performance of the proposed method. Dataset I was collected from Wukuang mine in the Pingdingshan Coal Group of China, and had 363 images only including individual coal, 355 images only including individual gangue, and 124 images including multiple coal and gangue respectively. Dataset II was collected from Zhaogu mine in Henan Energy Group of China. For Dataset II, we respectively captured 324, 310, and 95 images for individual coal, individual gangue, and multiple objects. Dataset III was collected from Wangzhuang mine in Luan Coal Group of China, including 305, 331, and 109 images for cover individual coal, individual gangue, and multiple coal and gangue respectively. The size of all images is 1024×1024, which is collected under the condition of natural light.

From the above datasets, we randomly respectively collected 300, 100, and 20 images for training, validating, and testing. We manually annotated labels and corresponding locations for each coal and gangue in optical images. Examples of coal and gangue were displayed in Fig. 7.

B. VALUATION METRICS

We used six indexes to evaluate the performance of the proposed method. Intersection over union (IoU) was employed to evaluate the performance of target detection, referring to the overlap ratio between the target window generated by the model and the original marked window (see Figure 8).

Besides, we evaluated the accuracies with the metric of mean intersection over union (MIoU), which is illustrated as follows:

$$\mathrm{MIoU} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ij} - p_{ii}} \tag{8}$$

where $k + 1$ represents the total number of categories, p_{ii} is the number of pixels of class i predicted to class i, that is true positive (TP). By contrast, the p_{jj} represents the true negatives (TN). The p_{ji} and p_{ij} represent the numbers of false positive (FP) and false negative (FN) pixels respectively.

The other evaluation parameters in this paper are accuracy, precision, and recall. The calculation formula is as follows:

$$Precison = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{11}$$

where TP is correctly classified into the number of positive examples, that is, the number of instances (samples) that are actually positive examples are classified into positive examples by the classifier; FP is incorrectly classified into the number of positive examples, that is, the number of instances are actually negative, but are classified into positive examples by the classifier; FN is the number of negative cases that are incorrectly classified into positive examples, that is, they are actually positive; TN is the number of negative cases that are correctly classified. Moreover, the number of instances that are actually negative are classified as negative by the classifier.

C. EFFECTS OF DIFFERENT TRAINING ITERATION NUMBER

In order to analyze the effects of training iteration number [43], we set the learning rate as 0.001 and adopt the Adam learning strategy [42] to optimize our model. The parameters of Adam learning strategy are as follows: exponential decay rates $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and minor amount $\varepsilon = 1e{-}08$. We used training accuracy and test accuracy to evaluate the detection results, and exploited the MIoU value to assess the

(a)

(b)

(c)

FIGURE 7. Datasets. (a) The images with an individual coal; (b) The images with an individual gangue; (c)The images with multiple coal and gangue.

TABLE 1. The experimental results with different number of iteration on Dataset I.

Iteration Number	200	400	600	800	1000
Training accuracy (%)	83.4	99.8	99.8	99.8	**99.8**
Test accuracy (%)	95.6	97.7	98.0	**98.3**	97.1
MIoU (%)	89.76	91.69	92.51	92.51	**92.51**

TABLE 2. The experimental results with different number of iteration on Dataset II.

Iteration Numbers	200	400	600	800	1000
Training accuracy (%)	84.9	99.3	99.3	99.3	**99.3**
Test accuracy (%)	95.44	98.39	**98.96**	96.15	96.27
MIoU (%)	88.01	88.74	88.74	88.74	**88.74**

TABLE 3. The experimental results with different number of iteration on Dataset III.

Iteration Numbers	200	400	600	800	1000
Training accuracy (%)	88.3	99.3	99.5	99.5	**99.5**
Test accuracy (%)	95.17	96.47	**98.91**	98.91	98.90
MIoU (%)	88.92	90.08	90.33	90.33	**90.33**

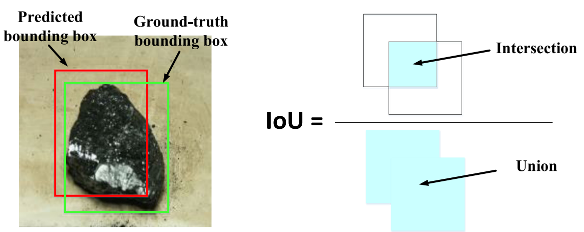


FIGURE 8. Schematic diagram of IoU.

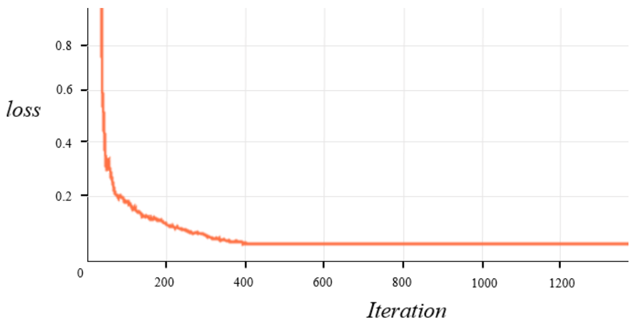


FIGURE 9. Convergence performance of our proposed algorithm of the coal and gangue image.

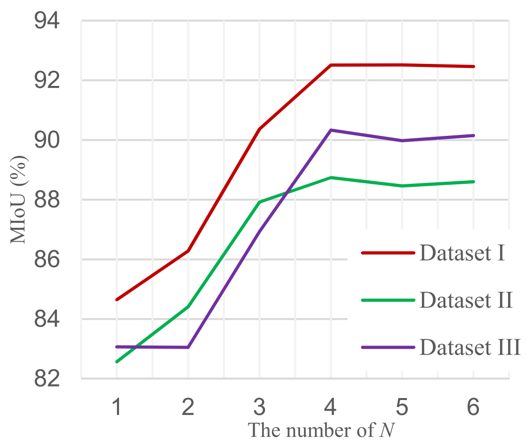


FIGURE 10. Curves of MIoU values under different numbers of layer.

accuracy of object detection box position. Here, VGG-16 model [34] is integrated into our method to extract features under different numbers of training iteration.

Firstly, the convergence performance of our proposed method is illustrated in figure 9. In this figure, the transverse axis indicates the number of iterations from 1 to 1200 while the longitudinal axis shows the value of the loss function. It is clear that our proposed algorithm converges in about 400 iterations, which shows that our proposed detection model is very effective in implementation.

Tables 1-3 show the obtained accuracy values and the MIoU values of training and test images in Datasets I – III. It can be seen from Tables 1-3 that in case of a small number of iteration, network learning is not sufficient, and the learned model is not ideal, so the classification effects of training data is poor. As the number of iterations increases, the network parameters are continuously optimized, and the classification accuracy rate also increases. It can be seen from Tables 1-3 that when the number of training iteration changes from 200 to 1000, MIoU values of the test images also increases synchronously, which indicates that the accuracy of the object detection and positioning are improved with the increasing of the training iteration number. However, the test accuracy of object detection decreases when the number of iteration is more than 600 in Tables 2 and 3, 800 in Table 1, which indicates that the overfitting may be occurred. It can be observed that our model has the best detection performance when the training iteration number ranges from 600 to 800. Moreover, it shows that our network has strong robustness by comparing the experimental results under different datasets, and its final test accuracy is about 98.33%, 98.96% and 99.16%,

TABLE 4. Test results of different model on Dataset I.

Model	Alexnet	ZF	VGG-16	Inception-v3
MIoU (%)	81.26	80.79	92.51	**93.67**
Coal precision	0.922	0.931	0.983	**0.992**
Coal recall	0.916	0.925	0.988	**0.994**
Gangue precison	0.915	0.925	0.989	**0.995**
Gangue recall	0.921	0.930	0.983	**0.992**

TABLE 5. Test results of different model on Dataset II.

Model	Alexnet	ZF	VGG-16	Inception-v3
MIoU (%)	81.59	82.03	**88.74**	87.57
Coal precision	0.918	0.928	**0.989**	0.994
Coal recall	0.935	0.919	0.985	**0.987**
Gangue precison	0.937	0.919	0.985	**0.987**
Gangue recall	0.919	0.927	0.989	**0.993**

TABLE 6. Test results of different model on Dataset III.

Model	Alexnet	ZF	VGG-16	Inception-v3
MIoU (%)	81.97	82.45	90.33	**90.81**
Coal precision	0.919	0.929	**0.989**	0.981
Coal recall	0.924	0.915	**0.982**	0.972
Gangue precison	0.924	0.914	**0.982**	0.972
Gangue recall	0.919	0.927	**0.989**	0.981

and MIoU is respectively 92.51%, 88.74% and 90.33% on Datasets I, II and III.

D. EFFECTS OF DIFFERENT DEEP NEURAL NETWORK MODELS

Large scale visual recognition challenge (ILSVRC) competition is as an object detection and identification event, and has globally significant influence. It has been held annually since 2010. Many excellent models have achieved excellent results in the competition. We used several representative deep learning models such as AlexNet [33], ZF [44], VGG-16 [34] and Inception-v3 [45] to integrate into our method, to test the effects of different deep neural network models. The common parameters for each model are set as 800 iterations, and 7 layers in the multi-level training datasets.

We used MIoU, precision and recall of coal and gangue detection to evaluate the test results of different methods on Datasets I-III, and the test results are shown in Tables 4-6. By observing the test statistics, we can see that the four methods have achieved good performance in our model, which indicates that our hierarchical target detection model has advantages on coal and gangue detection.

E. IMPACTS OF LAYER NUMBER N

In order to verify the effects of level number (N) in the multi-level training samples on the detection results, we set the number of layers N from 1 to 6, and integrated the Inception-v3 model into the training network of the object detection. The number of iterations is set as 700. The trained models and parameters are saved and the images in the Datasets I–III

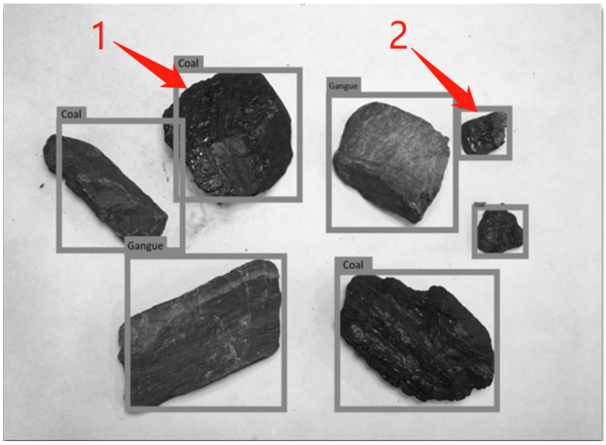


FIGURE 11. Different sizes of coal and gangue in an image.

are detected. MIoU is used to evaluate the results of test data, as shown in Fig. 10. Obviously, MIoU value rises with increasing of N, and can reach a stable situation when N is equal to 4. We can find that multi-level training structure can improve the accuracy of coal and gangue detection.

For a plurality of coal and gangue in an individual image, the size difference may sometimes be large (as shown in Fig. 11) since the shape and size are not uniform. In Fig. 11, the size of gangue with number 1 is about 6 times larger than the coal with number 2. If the ordinary training strategy is adopted, the target objects of different sizes cannot be well detected simultaneously. Multi-level training structure can help extract more adequate features for different sizes of coal and gangue. Compared with the method of judging by fractal

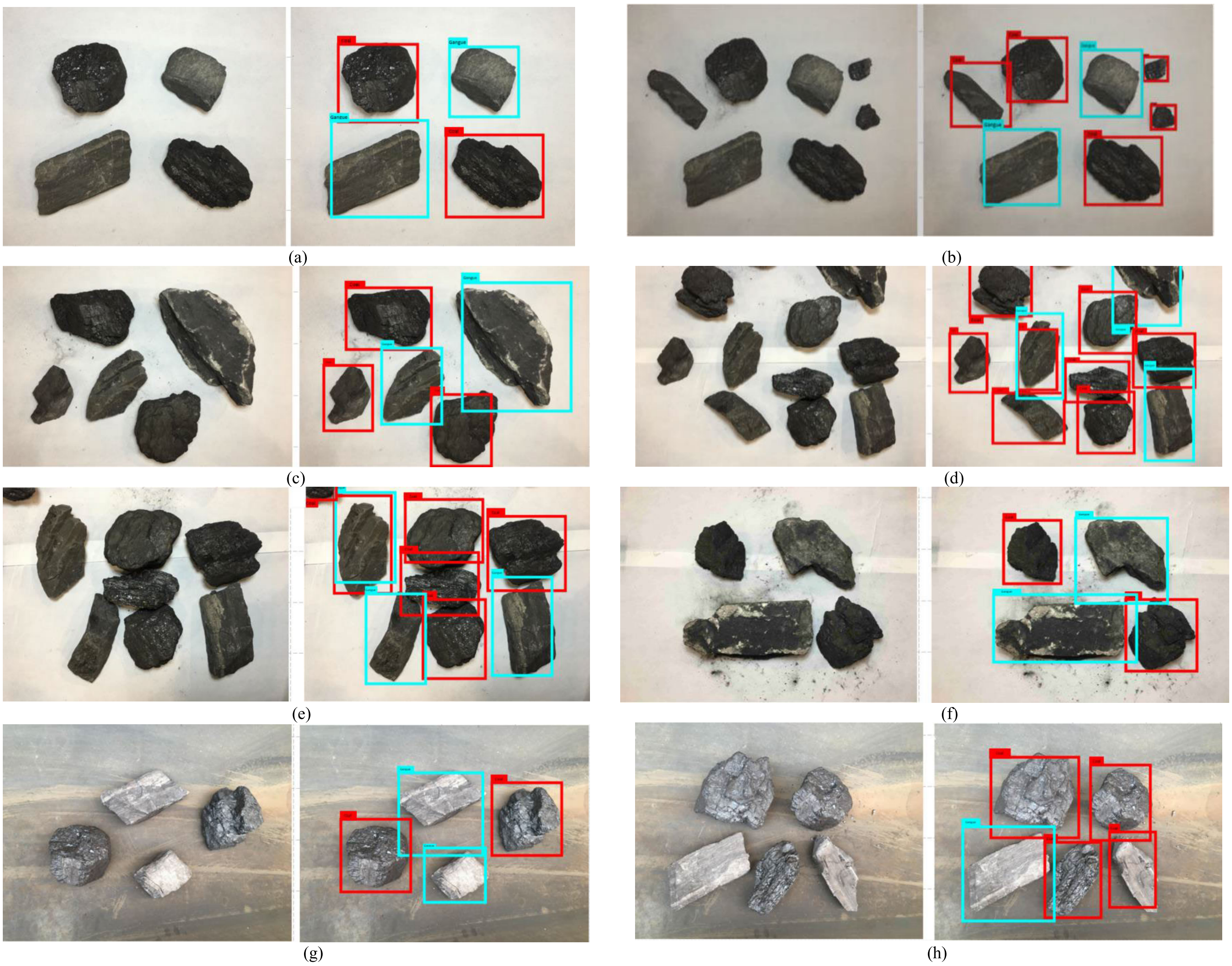


FIGURE 12. The results of coal and gangue detection. In sub-figure (a-h), the left side is original picture and the right side is detection results. Coal is marked with red detection box and gangue is marked with blue detection box.

dimension [30], our method can discriminate among multiple targets of different sizes at the same time.

F. TEST OF THE MULTIPLE OBJECT DETECTION

In order to verify the detection effects of our model on multiple coal and gangue in the image, we set the number of layers N as 4, and the other network parameter settings are same as those in the previous section. The visual detection results of partial image are shown in Fig. 11. In the sub-figures (a-h), the left side is original picture and the right side is detection results. The coal is marked with red detection box and gangue is marked with blue detection box. It can be seen that all the objects in the eight test images are successfully detected, which illustrates the capacity on simultaneous detection of multiple coal and gangue in an image.

G. COMPARISONS WITH OTHER METHODS

We compared our method with the other six methods for detection of coal and gangue, such as grayscale based detection methods(the mean and variance of grayscale is as the recognition feature), texture feature based detection methods (contrast, correlation, energy, and inverse different moment are used as identification features), joint grayscale and texture feature based detection methods, fractal dimension features based method [30] (the modified local singular spectrum widths of the outline curve series are extracted as the characteristic variables of the coal and gangue for pattern recognition), the LeNet-improved detection method [37] (the learning rate is set as 0.0001 with 10000 iterations), AlexNet detection method [38] (the exponential decay rate $\alpha = 0.9$ and 2000 iterations). Table 7 shows the differences between our method and the other six methods. The detection accuracy refers to the percentage of correctly identified coal and gangue samples in all test samples. Compared with these six methods, our method has three aspects of characteristic: (1) It can automatically select the image features of coal and gangue in image, avoiding manual process of selecting and combining features;

TABLE 7. Comparison with other methods.

Methods	Extract features manually	Extract features automatically	Deep learning model	CG-RPN model	Multilevel structure	Recognition accuracy
Grayscale features[37]	✓	✗	✗	✗	✗	68.46%
Texture features[37]	✓	✗	✗	✗	✗	73.86%
Grayscale& texture[37]	✓	✗	✗	✗	✗	79.72%
Fractal dimension[30]	✓	✗	✗	✗	✗	97.5%
Improved LeNet[37]	✗	✓	✓	✗	✗	95.88%
AlexNet[38]	✗	✓	✓	✗	✗	96%
Our method	✗	✓	✓	✓	✓	**98.33%**

(2) Automatically positioning multiple detection targets, which solves the complicated front-end queuing problem that can be only recognized by a single object in the process of coal and gangue detection; (3) Multi-level training strategies are used, which can accurately detect different sizes of coal and gangue simultaneously; (4) Using the deep learning model can extract the discriminative feature, which is verified by experiments. We can see that the recognition accuracy of our method to coal and gangue is higher than the other 6 methods.

VI. CONCLUSION

In this paper, we presented a method for detecting coal and gangue based on the principle of deep learning from optical images. We first used the Gaussian pyramid principle to construct multi-level training data for the original images of coal and gangue in the training dataset, and extracted image features of different scales and resolutions by constructing deep learning model. In order to improve the detection efficiency, we designed CG-RPN to generate the target area of candidate objects. Different from the traditional method, we can detect multiple coals and gangue in an irregular arrangement. The convolutional neural networks are integrated into our model as an efficient feature extractor. The proposed method was tested on three datasets and obtained the best parameter settings (the number of iterations was 700, and the number of layers for multi-layer training was set to 4). We also compared our approach with six other methods, including traditional artifact-based image features and simple deep learning models. Experiments showed that the proposed method has higher detection accuracy and solves the problem that coal and gangue needs to be regularly arranged when separation is performed. This also showed that deep learning-based approach has advantages in coal and gangue feature extraction.

In future work, we will focus on simplifying the model and improving multithreading capabilities to support online detection. By integrating the model into control system, a design of complete coal and gangue sorting system can be explored in future.

ACKNOWLEDGMENT

The authors would like to thank the editor and the anonymous reviewers for their constructive comments and suggestions, which improve the quality of this paper.

REFERENCES

[1] Y. Zhao, X. Liu, S. Wang, and Y. Ge, "Energy relations between China and the countries along the Belt and Road: An analysis of the distribution of energy resources and interdependence relationships," *Renew. Sustain. Energy Rev.*, vol. 107, pp. 133–144. Jun. 2019.

[2] 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, pp. 736–753, Sep. 2017.

[3] G. B. Stracher and T. P. Taylor, "Coal fires burning out of control around the world: Thermodynamic recipe for environmental catastrophe," *Int. J. Coal Geol.*, vol. 59, nos. 1–2, pp. 7–17, Jul. 2004.

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

[5] C. Hua, G. Zhou, X. Yin, C. Wang, B. Chi, and Y. Cao, "Assessment of heavy metal in coal gangue: Distribution, leaching characteristic and potential ecological risk," *Environ. Sci. Pollut. Res.*, vol. 25, no. 32, pp. 32321–32331, 2018.

[6] J. Ding, C. Yang, and T. Chai, "Recent progress on data-based optimization for mineral processing plants," *Engineering*, vol. 3, no. 2, pp. 183–187, Apr. 2017.

[7] J.-G. Li and K. Zhan, "Intelligent mining technology for an underground metal mine based on unmanned equipment," *Engineering*, vol. 4, pp. 381–391, Apr. 2018.

[8] X. Miao and M. Qian, "Current status and prospects of green mining research in China's coal resources," *J. Mining Saf. Eng.*, vol. 26, no. 1, pp. 1–14, Mar. 2009.

[9] D. Dou, J. Yang, J. Liu, and H. Zhang, "A novel distribution rate predicting method of dense medium cyclone in the Taixi coal preparation plant," *Int. J. Mineral Process.*, vol. 142, pp. 51–55, Sep. 2015.

[10] D. Dou, J. Yang, J. Liu, Z. Zhang, and H. Zhang, "Soft-sensor modeling for separation performance of dense-medium cyclone by field data," *Int. J. Coal Preparation Utilization*, vol. 35, no. 3, pp. 155–164, Mar. 2015.

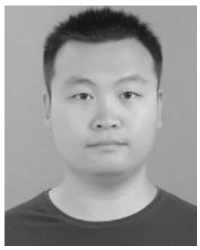
[11] X. Ma, "Research on Online identification and automatic sorting system of coal gangue," *J. Xi'an Univ. Sci. Technol.*, vol. 1, pp. 66–78, Mar. 2003.

[12] 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.

[13] L. Zhang, L. Zhang, B. Du, J. You, and D. Tao, "Hyperspectral image unsupervised classification by robust manifold matrix factorization," *Inf. Sci.*, vol. 485, pp. 154–169, Jun. 2019.

[14] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature learning using spatial-spectral Hypergraph discriminant analysis for hyperspectral image," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2406–2419, Jul. 2019.

[15] G. E. Hinton and R. R. Salakhutdinov, ''Reducing the dimensionality of data with neural networks,'' *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[16] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, ''ImageNet: A large-scale hierarchical image database,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[17] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, ''Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3325–3337, Jun. 2015.
[18] G. Cheng, P. Zhou, and J. Han, ''Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.
[19] S. Ren, K. He, R. Girshick, and J. Sun, ''Faster R-CNN: Towards real-time object detection with region proposal networks,'' in *Proc. 28th Int. Conf. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
[20] K. Castil, H. Yang, and H. Liu, ''Review of centrifugal re-election equipment,'' *Foreign Metal Ore Dressing*, vol. 40, no. 11, pp. 4–16, 2003.
[21] R. Zhang, X. Fu, and L. Wei, ''Fluid dynamics study of loose and layered jigging machine layers,'' *J. China Coal Socety*, vol. 28, no. 2, pp. 193–198, 2003.
[22] T. D. Nguyen, A. V. Nguyen, L. Chen-Luh, and J. D. Miller, ''Application of high-resolution X-ray microcomputed tomography for coal washability analysis,'' *Minerals Eng.*, vol. 124, pp. 137–148, Aug. 2018.
[23] N. Zhang, C. Liu, X. Chen, and B. Chen, ''Measurement analysis on the fluctuation characteristics of low level natural radiation from gangue,'' *J. China Coal Soc.*, vol. 40, no. 5, pp. 988–993, May 2015.
[24] L. K. Oikonomopoulos, M. Perraki, and N. Tougiannidis, ''A comparative study on structural differences of xylite and matrix lignite lithotypes by means of FT-IR, XRD, SEM and TGA analyses: An example from the Neogene Greek lignite deposits,'' *Int. J. Coal Geol.*, vol. 115, pp. 1–12, Aug. 2013.
[25] F. Liu, J. Qian, X. Wang, and J. Song, ''Automatic separation of coal mines based on image processing and recognition technology,'' *J. China Coal Soc.*, vol. 25, no. 5, pp. 534–537, 2000.
[26] C. Tan and J. Yang, ''Research on extraction of image gray information and texture features of coal and gangue image,'' *Ind. Mine Automat.*, vol. 43, no. 4, pp. 27–31, 2017.
[27] W. Hou, ''Identification of coal and gangue by feed-forward neural network based on data analysis,'' *Int. J. Coal Preparation Utilization*, vol. 39, no. 1, pp. 33–43, Mar. 2017.
[28] 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, Aug. 2013.
[29] X. Chen, X. Zhang, L. Xu, Y. Liu, and Y. Du, ''Study on fractal dimension differences of coal and rock,'' *Coal Sci. Technol.*, vol. 45, no. 7, pp. 196–199, 2017.
[30] K. Liu, X. Zhang, and Y. Chen, ''Extraction of coal and gangue geometric features with multifractal detrending fluctuation analysis,'' *Appl. Sci.*, vol. 8, no. 3, pp. 463–478, 2018.
[31] M. Suárez, V. M. Brea, J. Fernández-Berni, R. Carmona-Galán, D. Cabello, and Á. Rodríguez-Vázquez ''Low-power CMOS vision sensor for Gaussian pyramid extraction,'' *IEEE J. Solid-State Circuits*, vol. 52, no. 2, pp. 483–495, Feb. 2016.
[32] 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.
[33] A. Krizhevsky and I. H. G. Sutskever, ''Imagenet classification with deep convolutional neural networks,'' in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.
[34] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large-scale image recognition,'' *Comput. Sci.*, pp. 1904–1916, Apr. 2014.
[35] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, and D. Erhan, ''Going deeper with convolutions,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[36] 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.
[37] L. Su, X. Cao, H. Ma, and Y. Li, ''Research on coal gangue identification by using convolutional neural network,'' in *Proc. 2nd IEEE Adv. Inf. Manage., Commun., Electron. Automat. Control Conf. (IMCEC)*, May 2018, pp. 810–814.
[38] H. Hong, L. Zheng, J. Zhu, S. Pan, and K. Zhou, ''Automatic recognition of coal and gangue based on convolution neural network,'' Dec. 2017, *arXiv:1712.00720*. [Online]. Available: https://arxiv.org/abs/1712.00720
[39] R. Girshick, J. Donahue, T. Darrell, and J. Malik, ''Rich feature hierarchies for accurate object detection and semantic segmentation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 580–587.
[40] R. Girshick, J. Donahue, and T. Darrelland, ''Fast R-CNN,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2015, pp. 1440–1448.
[41] Y. Liu, Z. Zhang, R. Zhong, D. Chen, Y. Ke, J. Peethambaran, C. Chen, and L. Sun, ''Multilevel building detection framework in remote sensing images based on convolutional neural networks,'' *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 10, pp. 3688–3700, Oct. 2018.
[42] D. Kingma and J. Ba, ''Adam: A method for stochastic optimization,'' in *Proc. Conf. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 611–623.
[43] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, and J. Dean, ''TensorFlow: Large-scale machine learning on heterogeneous distributed systems,'' Mar. 2016, *arXiv:1603.04467*. [Online]. Available: https://arxiv.org/abs/1603.04467
[44] M. Zeiler and R. Fergus, ''Visualizing and understanding convolutional neural networks,'' in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2013, pp. 818–833.
[45] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, ''Rethinking the inception architecture for computer vision,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.



DONGJUN LI received the B.S. degree in electrical engineering and its automation from Hennan Polytechnic University, Jiaozuo, China, in 2015. He is currently pursuing the Ph.D. degree with the School of Mechanical, Electronic, and Information Engineering, China University of Mining and Technology, Beijing, China.

His current research interests include deep learning and non-destructive testing.

ZHENXIN ZHANG received the Ph.D. degree in geoinformatics from the School of Geography, Beijing Normal University, Beijing, China, in 2016.

He is currently an Associate Professor with the Beijing Advanced Innovation Center for Imaging Theory and Technology and the Key Laboratory of 3D Information Acquisition and Application, College of Resource Environment and Tourism, Capital Normal University, Beijing, China. His research interests include light detection and ranging data processing, quality analysis of geographic information systems, remote sensing image processing, and algorithm development.

ZHIHUA XU received the Ph.D. degree in cartography and geographic information system from the Key Laboratory of Environment Change and Natural Disaster, Ministry of Education, Beijing Normal University, Beijing, China, in 2016.

He is currently an Associate Professor with the College of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing, China. His research interests include UAV photogrammetry, light detection and ranging (LiDAR) data processing, and open-pit mining.

LILI XU was born in 1993. She received the M.E. degree in measuring and testing technologies and instruments from the School of Mechanical, Electronic, and Information Engineering, China University of Mining and Technology (Beijing), Beijing, China, in 2019. Her research interests include image processing and computer vision.

GUOYING MENG received the M.S. and Ph.D. degrees in fluid power driven and control, and mining machinery engineering from the China University of Mining and Technology (Beijing), Beijing, China, in 1993 and 1995, respectively. He is currently a Professor with the School of Mechanical, Electronic, and Information Engineering, China University of Mining and Technology (Beijing). His research interests include fault diagnostics, mining equipment, and coal mine robot.

ZHEN LI received the bachelor's degree in surveying and mapping engineering from the Shandong University of Science and Technology, Qingdao, China, in 2017. He is currently pursuing the master's degree in cartography and geographical information engineering with the Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing, China.

His research interests include deep learning and remote sensing images semantic segmentation.

SIYUN CHEN received the bachelor's degree in geographic information science from the Kunming University of Science and Technology, Kunming, China, in 2017. She is currently pursuing the master's degree in cartography and geographical information engineering with the Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing, China.

Her research interests include deep learning and vehicle laser scanning point clouds semantic segmentation.