Coal and Gangue Recognition Method Based on Local Texture Classification Network*

Yuting Xie, Xiaowei Chi, Haiyuan Li, *Member, IEEE*, Fuwen Wang

***Abstract*— Coal gangue is a kind of industrial waste in coal mines preparation process. Compared to conventional manual or machine-based separation technology, vision-based methods and robotic grasping are more superior in respect of cost and maintenance. However, the existing methods may have a poor recognition accuracy problem in diverse environments since a lack of learnable features in complete full-scale images of coals and gangues. In this paper, we propose a vision-based coal and gangue recognition model LCT-Net for separation systems. In order to solve the problem, we divide the preprocessed full-scale images into *n**n* local texture images since coals and gangues differ more in a smaller scale, enabling the model to recognize texture patterns in addition to limited features like absolute gray value. A VGG16-based model is trained to classify the local texture images, with results counted and prediction given by a threshold *t*. Experiments based on multi-environment datasets show a higher accuracy and stability of our method compared to existing methods. The effect of *n* and *t* is also discussed.**

Keywords—Coal and Gangue Classification, Deep Learning

I. INTRODUCTION

Coal gangue, or gangue, is a kind of black or gray solid waste discharged in the process of coal mining and coal washing. Compared to the widely-used fossil fuel, gangue provides much fewer energy supply and therefore contain less value, due to lower carbon content. It can also be a pollution source in the absence of proper treatments. Therefore, the separation of gangue and coal in raw coal production workshops is critical to improve the quality of coal, minimize the storage and transportation cost and protect the environment [1].

Conventional methods for coal and gangue separation are mainly manual or machine-based. A typical underground coal-mining production line transports mixed coals and gangues onto ground-level workshops with conveyor belts, and feeds them into separation process. Manual method requires workers on both sides of the belt, removing gangues manually. Since the dust and high temperature could render a harsh working environment, coal yards today incline to machine-based methods. Coal-washing machine, which separates coals and gangues automatically is the most widely used one to free workers from intensive, repetitive and arduous separation movements. It is essentially a centrifugal machine utilizing the different density of coal and gangue. X-rays or gamma-rays transmission sensors are also adopted by workshops among other machine-based methods [1]. However, the high mechanical quality requirement in order to ensure the productivity leads to high purchase and maintenance costs of these kinds of machines. Moreover, soaring energy consumption and potential air pollution can also be a problem.

With the rapid development of computer vision technology and growth of computing capability, low-cost vision-based coal and gangue automated separation systems have been emerging. Fig. 1 depicts the general framework of an automated vision-based coal and gangue separation system. It can be divided into three core components as following:

- Vision unit: Normally an industrial camera with illumination devices to capture digital images at a fixed frequency in correspondence with the constant-speed conveyor belt.
- Control unit: Normally a micro-computer with a coal and gangue recognition model (e.g. classification algorithm) with an image from vision unit as input and a recognition result as out-put.
- Separation unit: A physical separation device (e.g. air pressure gun or robotic hand) operating based on the output of control unit to remove recognized gangues.

Benefiting from the distributed structure and relatively lower-cost devices, vision-based systems have an advantage in cost and maintenance compared to machine-based methods. However, it is also worth noting that the reliability and effectiveness of vision-based systems in realistic industry depend significantly on the accuracy of the recognition model.

Promising vision-based methods with regard with coal and gangue image recognition model arose over recent years, and could be mainly classified into two types: two-step methods and one-step methods. The former is represented by Liang's work [2] which extracts eight characteristic parameters from the Gray Level Co-occurrence Matrix (GLCM) of the original full-scale images and trained an SVM model or a BP neural network as a prediction boundary.

Liu [3] notices the different patterns in the shape of coal and gangue edges could be an indicator of category. They applied Multifractal Detrending Fluctuation Analysis (MDFA) on the

*This work was supported by the National Key Research and Development Program of China (Grant No. 2019YFB1309802), and by National Natural Science Foundation of China (Grant No. 62003048).
Yuting Xie is with School of Software, Beijing University of Posts and Telecommunications, No. 10 Xitucheng Road, Beijing, China.
Xiaowei Chi is with International School, Beijing University of Posts and Telecommunications, No. 10 Xitucheng Road, Beijing, China.
Haiyuan Li and Fuwen Wang are with School of Automation, Beijing University of Posts and Telecommunications, No. 10 Xitucheng Road, Beijing, China (Corresponding author: lihaiyuan@bupt.edu.cn).

images and trained classifier for recognition. In Dou's paper [4], the case of coal and gangue covered by dust stains on the surface is discussed. A total of twelve colors and texture features are used for a support vector machine (SVM) classifier. The above-mentioned works have a common in the two-step structure which first extracts some specific feature values from original images and performs classification with trained models like SVM or neural networks on these values. One-step methods aims at classification directly on raw or preprocessed images. Pu [5] sets up a coal and gangue dataset with 100 images respectively, and trained a convolutional neural network VGG16 for direct classification. Transfer learning method is also applied to solve the problem of data volume, and to obtain an accuracy rate of 82.5% on their dataset. Hong [6] implements an improved convolutional neural network based on AlexNet [10] to solve the discrimination problem, and also discusses the problem of object detection and area clipping of target objects against the back-ground in images. Image enhancement and transfer learning is also introduced to solve the problem of insufficient data in the early stage. By studying the previous works, we notice a negligence of full discussion of diverse datasets in different environments and from different devices, which would result in a failure of these methods when applied to various datasets (different working environments). Therefore, we consider improving the solutions by proposing a method Local Texture Classification Network(LCT-Net). The rest of this paper is organized as following:

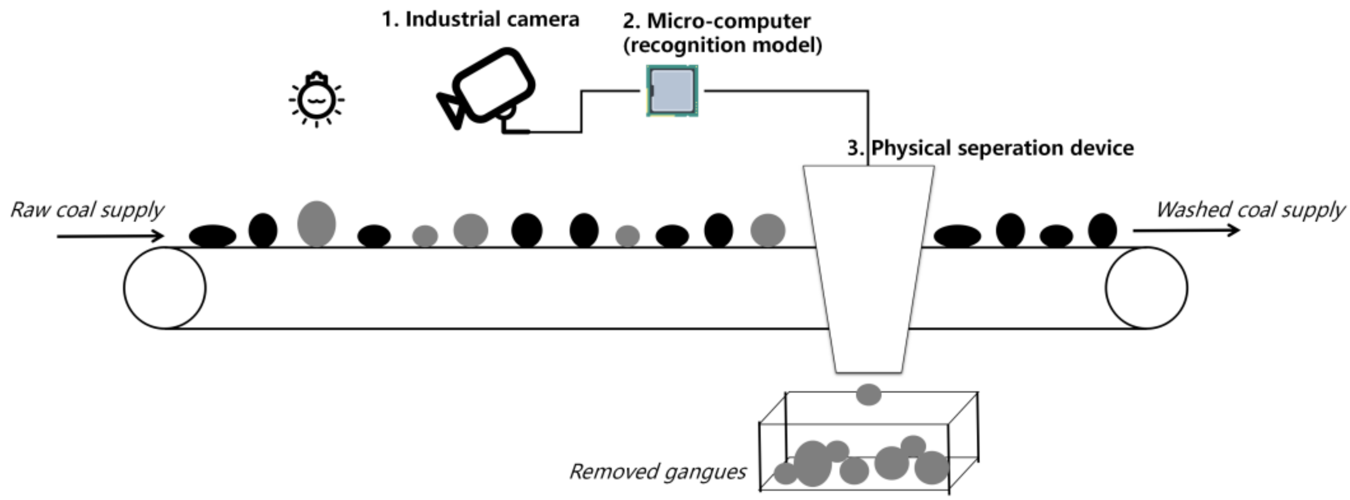

Fig. 1. General framework of an automated vision-based coal and gangue separation system.

In Section II, the limitation of existing methods under certain circumstances is discussed and a coal and gangue recognition method based on LCT-Net is proposed. In Section III, a group of experiments on various datasets with baseline methods and our method, and the result and the effect of parameters is discussed. At the end, we draw conclusion and discuss possible directions of future work.

II. Problem Analysis

To compare algorithms' performance, the homologous dataset is created: images in multiple datasets taken by identical device under same illumination condition from fixed view angle and distance. Generally speaking, coal and gangue images in homologous datasets are to some extant respectively identical in average gray level, average image pixel (resolution), average hue and other attributes. By contrast, images taken by different cameras from different distance and angle under different illumination conditions are dimmed as non-homologous dataset which might come from changing workcell environments in plant.

The current methods are applied on both homologous and non-homologous datasets (see section 4.2). The test results indicates two main problems, which we are prone to judge as a limitation of the learning capability of existing methods:

- For two-step methods, such as Gray Value+SVM, GLCM+SVM and MDFA, though decent accuracy can be reached on training dataset and its homologous datasets, a sharp drop occurs in both the accuracy and recall rate when tested on non-homologous datasets. Even a total failure in recognition could occur. We believe that the inherent differences between non-homologous datasets reflecting on significantly different features or characteristic numbers extracted in the first step, lead to a less effective recognition model trained in the second step.
- For convolutional neural network methods like LeNet [12], an upturn in accuracy could be expected when tested on non-homologous datasets. However, the result would also deteriorate with resolution degradation and can be adjusted and improved.

In short, the existing methods lack generalization ability among non-homologous datasets. However, vision-based coal and gangue separation systems are expected to be a prompt in efficiency and reduction in cost for coal yards. According to field investigation, the lighting conditions, system structures and camera installations differ among coal yards, failing to guarantee identical operating environments and homologous images between yards. Under this circumstance, a non-generalization method could be limited from wide application. We work on a new solution that secure an acceptable recognition accuracy on homologous datasets.

III. Method

Through further study of non-homologous datasets, we put forward a reasonable assumption that unlike normal image classification problems (i.g. MNIST [17], Cats vs. Dogs), the large inner-class difference within coal images and gangue images can be a possible explanation accounting for the non-generalization problem, since both coals and gangues are short of fixed shape or texture patterns in full-scale, least of the significant difference within class between datasets, as shown in Fig. 4.

Therefore, inspired by Nasiri's work [13], we consider a smaller scale and discover a less inner-class difference and larger difference between local textures, as shown in Fig. 6. Accordingly, we propose a novel method LCT-Net (Local Texture Classification Network), which first cuts and divides the full-scale preprocessed coal and gangue images into local texture images, and classifies them with a VGG16-based [14] network. The results would be counted with a threshold used to make the final judgement.

Fig. 2 shows the framework of the method, the details of which will be explained in the following sections.

A. Data setup

In order to test and verify the validation and usability of the LCT-net method with regard to the above-mentioned other ones, we set up one training dataset and four testing datasets by images captured with different devices, in different illustration conditions and from different views of angle. Besides, the image proportion of objects also differs. An uniform image background is the only same feature. Fig. 4 shows sample images from different datasets from the captured images. Among these datasets, Dataset1 is divided into two homologous datasets, a large-sized Dataset1_train and a small-sized Dataset1_test. Dataset2 is the same as Dataset1 except for imposed random lights which render the images random gray value features both inner-class and inter-class. Dataset3 and Dataset4 are taken by the same less pixel camera, however, the lights and distances to the objects differ, and this leads to a resolution difference after preprocessing (see section III. *B.*).

In short, all datasets are non-homologous except for Dataset1_train and Dataset1_test. Detailed information and comparison of the datasets are shown in Table I and Fig. 3.

Among these datasets, Dataset1_train is to be the training set of our method and baseline methods, with Dataset1_test, Dataset2, Dataset3 and Dataset4 being the testing sets.

B. Data preprocessing

As shown in Fig. 2, we apply preprocessing procedures on images in datasets as the first step of our method. An example of detailed processing procedures and intermediate results is given by Fig. 5.

For each original image (a), we transform it into gray image (b), followed by multiple Gaussian blur procedures to reduce the noisy and lower the level of details to (c). An image binarization with OTSU method [16] is then applied as shown by (d), with which each pixel is filtered by a threshold and fixed to an upper-bound gray value or a lower-bound one. To reduce noise interference in the maximum extent, we process the image with erosion and dilation operations as shown in (e). By doing the above-mentioned procedures, we can now perform the segmentation on (e) and acquire a contour line which outlines the desired object in the image. We cut the origin image (a) into an minimum rectangle that surrounds the contour line as shown in (f), and at the same time set the gray value to 0 for all back-ground pixels outside the contour line, in order to prevent a disturb in the following model training. (g) shows a final result after preprocessing.

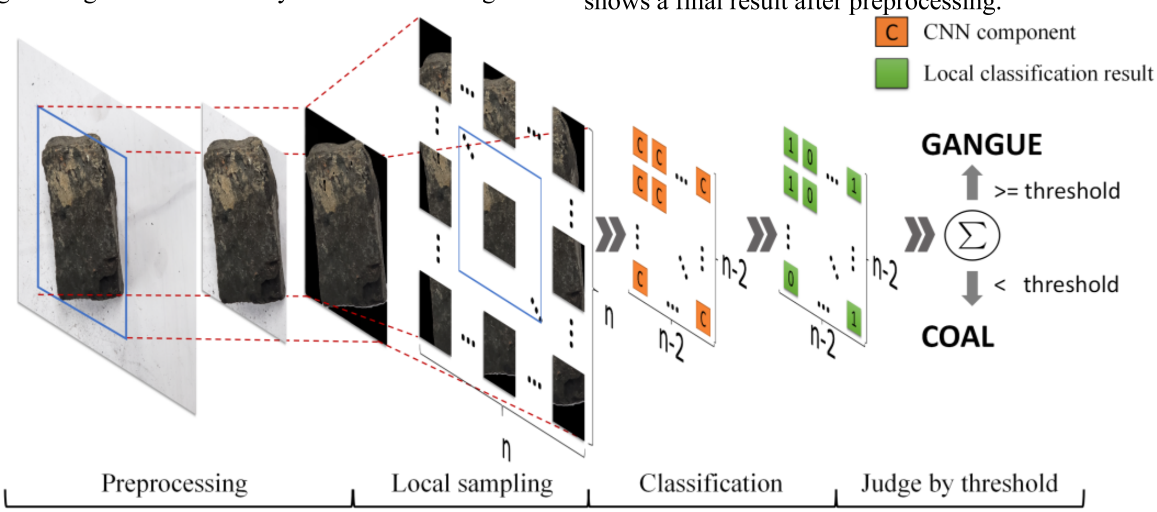

Fig. 2. LCT-net method.

TABLE I. META DATA OF DATASETS

Name	Attributes					
	Description	***Size of set (coal, gangue)***	***Avg. coal gray value***	***Avg. gangue gray value***	***Resolution (Million Pixels)***	***Avg. resolution after preprocessing***
Dataset1_train	High resolution, normal lights.	235,245	33	41	14.0	≈8.0
Dataset1_test		49,49				
Dataset2	High resolution, random lights.	49,49	37	43		
Dataset3	Low resolution, dark lights.	49,49	28	36	5.0	≈1.0
Dataset4	Very low resolution, bright lights.	49,49	56	77	5.0	≈0.06

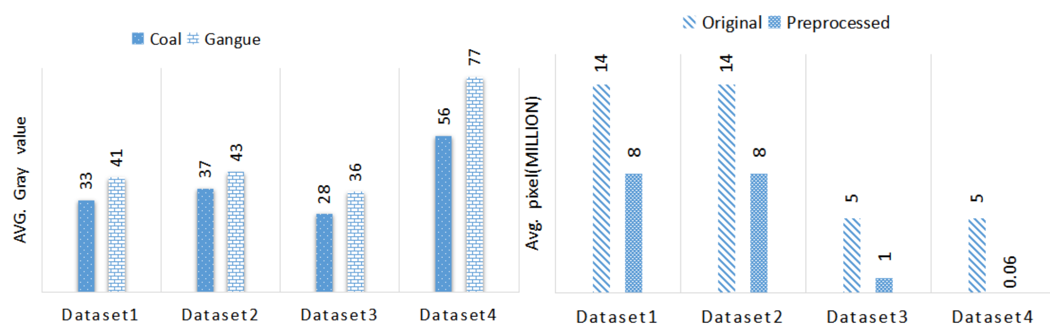

Fig. 3. The gray value of images and pixel numbers varies among datasets.

Fig. 4. Sample full-scale images in different datasets.

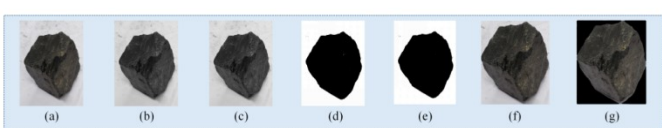

Fig. 5. Intermediate results in preprocessing procedures.

The preprocessing procedures also cause the above-described datasets with great differences in average pixel, as shown in Table I.

C. Local sampling

By preprocessing procedures we now have the segmented and background-reduced images for further treatments. As illustrated in the "Local sampling" part in Fig. 2 and Algorithm 1, we now aim at acquiring local texture images through cutting the original full-scale images. A n-by-n cutting is applied as n being a hyper-parameter, the effect of which is discussed in section IV. *C*. For cutting images, the outer-ring ones will be abandoned, while the inner (n-2)-by-(n-2) ones being used. This can be intuitively comprehended as a mechanism to minimize the number of local images that have a high proportion of background in datasets, and this could also be controlled by the value of n.

The number of cut images provided for following model training can be calculated by (1) where Sc and Sg are respectively the number of coals and gangues in datasets.

$$(Sc + Sg) * (n - 2) * (n - 2) \tag{1}$$

This process is depicted by Algorithm 1. With n = 5 as an example, the 235 coal images and 245 gangue images in

Dataset1_train will be able to yield 4320 training images for the following CNN component. Samples of cut local texture images are shown in Fig. 6.

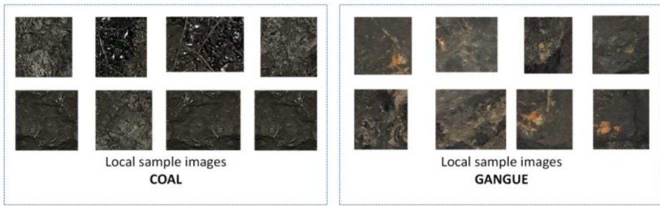

Fig. 6. Local texture images are more distinct and separable compared to full-scale images.

Algorithm 1 Local sampling of coal or gangue images.
Input: a preprocessed coal of gangue image $I(0,0,w,h)$ with width w and height h, hyper-parameter n.

```
for i = 1 to n-1 do
    for j = 1 to n-1 do
        yield Prepro(I(i*w/n, j*h/n, w/n, h/n))
```

D. Local texture classification with CNN component

This section introduces the local texture classification shown in Fig. 2. We design and implement a VGG16-based [14] CNN model, and use it as the local texture classifier, which is summarized by the left part in Fig. 7.

The network is trained with preprocessed cutting images from Dataset1_train, and tested on those from Dataset1_test. Before fed into the model, images are resized to (256, 256, 3). The right part of Fig. 7 depicts the fitting process of the model on training set, and shows a steadily rising trend of accuracy and dropping trend of loss both on training set and validation set. Details and test results of the model are shown in Table II.

TABLE II. DETAILED INFORMATION OF CNN COMPONENT

Name	Description
Input	(256,256,3)
Output	0 : coal texture 1 : gangue texture
Activation function	Output layer: SIGMOID Other: RELU
Loss function	Binary cross entropy
Validation	5-fold validation
Random dropout threshold	0.5
Epoch(s)	20
Batch size	60
Initial parameters	ImageNet
Other	Random rotation, zooming, cut, brightness and contrast change are imposed before training.
Test accuracy	82.33%
Test recall rate of coal textures	83.76%
Test recall rate of gangue textures	81.49%

It is worth noting that this model is set for the classification of the local texture images instead of the full-scale complete images. Though the local texture classification accuracy to a certain extant determines the overall success of LCT-Net method, it is not the only direct indication of the final recognition results, since the original intention of LCT-Net method is to reduce the instability in classification and improve the robustness.

E. Full-scale recognition

The test results shown in Table II from the above section testified the validation of the CNN component in terms of classification of the local texture images from coals and gangues with a test accuracy above 80%. We then apply it in the recognition process described in Fig. 2. As shown in Fig. 2, for each full-scale image, $(n-2)*(n-2)$ classification results will be given by the "Classification" procedure. The next procedure "Count and Judge" sum the results and compare it to a threshold value t. If the threshold is reached, a final recognition result of gangue is given, otherwise, coal. This process is depicted by Algorithm 2, and the value of t is discussed in subsection *F*.

Algorithm 2 Full-scale recognition.
Input: $(n-2)*(n-2)$ local texture images I sampled from one original full-scale image (the output of Algorithm. 1), a trained local texture classification CNN model m, a pre-calculated threshold t.

```
c = 0
for i = 1 to (n-2)*(n-2) do
    resize (I[i], (256,256,3))
    if 0 == m.predict(I[i]) do
        c = c + 1
if c > t do
    output GANGUE
else
    output COAL
```

F. Threshold computation

The above-mentioned threshold t is determined based on the counting result given on training set (Dataset1_train). Taking $n = 5$ as an example, Fig. 7 shows the counting results of each full-scale image in Dataset1_train. It is also shown that given $n = 5$, a maximum counting result 9 can be provided per full-scale image.

It also shows the distribution of coal and gangue counting results as general low values for coals and relatively high values for gangues. This corresponds to the intuition that a large proportion of gangue images actually "seems like" coal with a large range of darkness and several small areas in full-scale gangue images present a unique pattern or texture feature. It can be dimmed as a possible explanation that the lack of separable texture features or patterns in full-scale images for classification model to learn can cause the problem of existing methods, and also a possible reason for the robustness lift of LCT-Net method.

Given the counting results, we can calculate the threshold through by a one-dimensional classifier (e.g. SVM), regarding the counting results for coals and gangues as two classes. The threshold line for $n = 5$ is shown in Fig. 8.

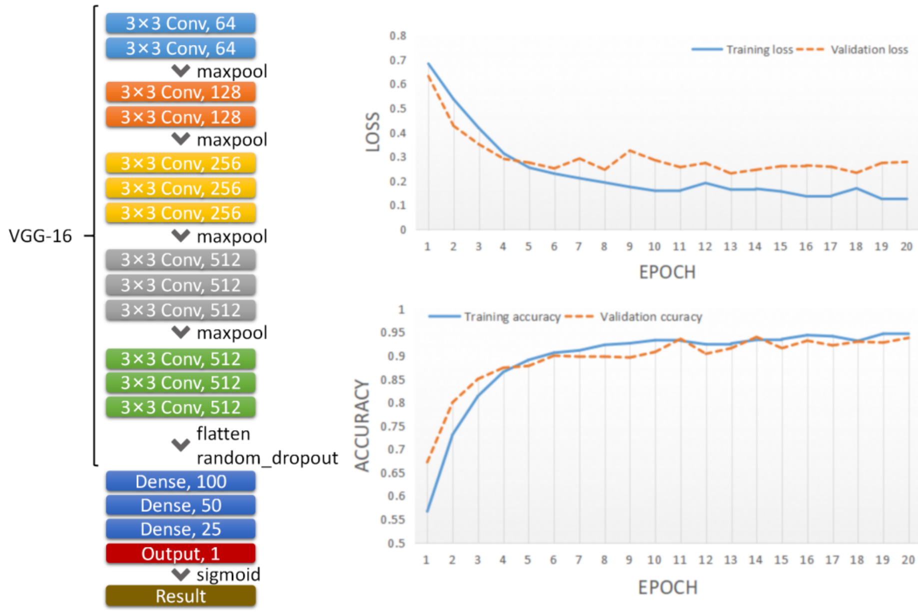

Fig. 7. Network structure and the fitting process of the model.

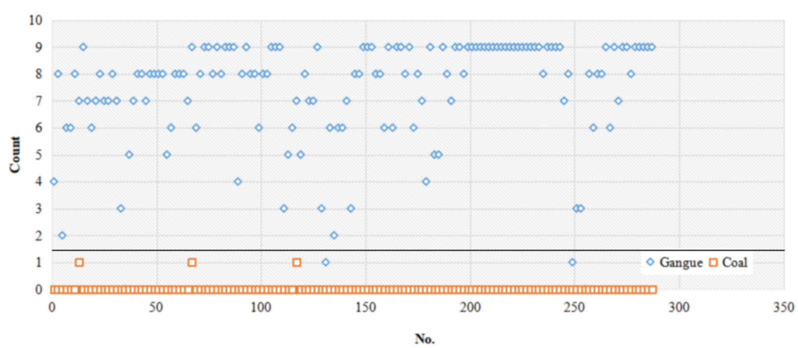

Fig. 8. The value of *t* is given by a one-dimensional linear classification model *(SVM)*.

G. Implementation

The method introduced above is implemented in Python3.6 with Numpy, Keras, TensorFlow, OpenCV and other necessary libraries.

IV. Experiment

In this chapter, we conduct experiments with LCT-Net method and baseline methods on Dataset1_test, Dataset2, Dataset3 and Dataset4 respectively. Overall accuracy, recall rates of coal and gangue are recorded correspondingly, with comparison discussed.

A. Baseline methods

1) Gray Value+SVM: A classification method is to perform direct average gray value calculation on specific images in Dataset1_train, and use a SVM model to acquire the category boundary to classify the images in testing datasets.

2) GLCM+SVM: To Four key parameters in GLCM, xxx, tat, yyy and eff respectively, are computed for each of Dataset1_train images, and an SVM model is trained to perform classification based on the four parameters.

3) CNN(VGG16): A VGG16 model is trained with preprocessed full-scale images in Dataset1_train, which takes an image(256, 256, 3) as input and present a zero or one output, as a category prediction at a full scale and at once. The detailed parameters are identical with the CNN component in Section III. *D*, shown in Table II.

B. Results and discussion

Similar to other studies, classification accuracy is an overall measure of LCT-Net method and other baselines. However, the recall rate of coal and the recall rate of gangue should also be a key indicator of the effectiveness and robustness. On one hand, the numbers of coals and gangues are basically balanced in our testing sets, therefore the offline method could by chance has an accuracy, greater than 50%, on the other hand, the ratio of coals to gangues can be very unbalanced in real-industrial workshops, and hence the methods should be evaluated based on all of the above indexes. A high recall rate of coal in mines with a low percentage of gangue leads to failure of the separation system, on the contrary coal mines with a high percentage of gangue results in a very waste of useful coals.

TABLE III. ACCURACY

Method	Dataset			
	Dataset1_test	*Dataset2*	*Dataset3*	*Dataset4*
Gray Value+SVM	0.800	0.710	0.772	0.485
GLCM+SVM	0.645	0.520	0.752	**0.879**
CNN(VGG16)	0.914	0.898	0.655	0.521
LCT-Net	**0.978**	**0.959**	**0.841**	0.788

TABLE IV. RECALL RATE OF COAL

Method	Dataset			
	Dataset1_test	*Dataset2*	*Dataset3*	*Dataset4*
Gray Value+SVM	0.880	0.680	1.000	0.000
GLCM+SVM	0.684	**0.980**	1.000	**0.833**
CNN(VGG16)	0.801	0.960	0.597	0.111
LCT-Net	**0.960**	0.979	**0.836**	0.778

TABLE V. RECALL RATE OF GANGUE

Method	Dataset			
	Dataset1_test	*Dataset2*	*Dataset3*	*Dataset4*
Gray Value+SVM	0.700	0.740	0.577	**1.000**
GLCM+SVM	0.583	0.060	0.538	0.933
CNN(VGG16)	1.000	0.833	0.706	**1.000**
LCT-Net	**1.000**	**0.939**	**0.846**	0.802

TABLE VI. AVG. OF FOUR DATASETS

Method	Index		
	ACC	*R (C)*	*R (G)*
Gray Value+SVM	0.692	0.640	0.754
GLCM+SVM	0.699	0.874	0.529
CNN(VGG16)	0.747	0.617	0.885
LCT-Net	**0.892**	**0.888**	**0.897**

As shown in Table III, IV, V and VI, LCT-Net method is testified stable and accurate on the four testing datasets. On Dataset1_test, the homologous dataset of the training set Dataset1_train, LCT-Net secure the 97.8% accuracy while baseline methods also have fairly good performances. However, when tested on non-homologous datasets, baseline methods show a decreasing accuracy while LCT-Net method can still have decent successful classification rates for both coals and gangues, even on Dataset4, the resolution of which is much lower.

As mentioned above, the accuracy should be interpreted along with the two recall rates together. For example, the CNN (VGG16) model on full-scale images seems to have a flawless performance on the Dataset4 in terms of the recognizing gangues, however it can barely recognize 11% of the coal images, resulting in a total accuracy of only 52.1% in Table IV. By contrast, LCT-Net could stably give reasonable results as the illustration condition and resolution of images change in datasets. This could be shown by Table VI.

In Section II, we discuss the problem that a lack of learnable features and patterns for models and methods to learn or extract may be the obstacle for other methods when tested on homologous datasets, since the inherent similarity of coals and gangues. Here, we consider it a possible explanation for the good performance of LCT-Net method. By dividing the full-scale images into local texture ones, we mitigate the affection of the inherent similarity problem, since the classification result of LCT-Net method is not given directly from a full-scale image or some parameters extracted from a full-scale image. In this way:

- The local texture images are more separable and have marked distinctions (larger inter-class distance), which leads to a higher recognition accuracy than those trained on full-images. After all, the non-homologous images could be considerably different in full-scale but more similar at a local texture scale.
- The final result is given by a counted value of local texture recognition middle results and a comparison to threshold, rather than directly at once. This improves the robustness, which could be intuitively interpreted as that LCT-Net is to some extant designed to work in a human-expert-like way. It is capable of capturing the distinctive textures in an image to decide its category. In general, the fewer remarkable patterns and signs of gangue an image contains, the more probable it would be recognized as a coal.

C. *Parameters study*

As introduced in section III. *C*, the scale of the local texture images is controlled by the value of the hyper-parameter n.

By introducing n, LCT-Net aims at zooming into a smaller scale and recognizing the local texture images, and therefore, when conducting the local sampling process in Fig. 2, the value of n actually represents the level of “local”. A too large n cuts the images into fragments, causing again the similarity problem, while smaller n eventually degenerates into the CNN(VGG16) method.

With a reasonable assumption that at a certain scale, the local texture images could be classified most effectively, we conduct controlled trials using Dataset2. As shown in Fig. 8, the peak of performance, considering the accuracy and recall rates of both coals and gangues combined, occurs when n equals 5. A gradual decreasing accuracy is also shown towards a total failure of recognition when n reaches 11.

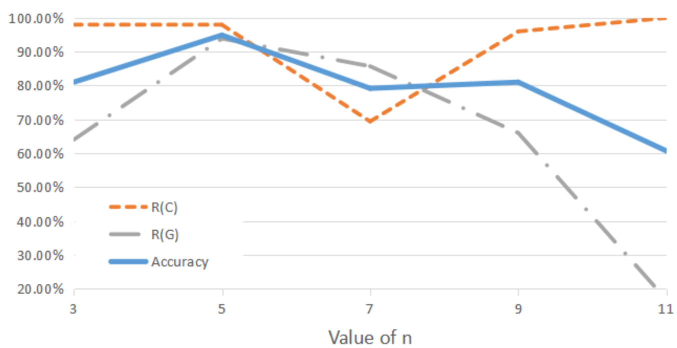

Fig. 9. The accuracy, recall rate of coal[R(C)] and recall rate of gangue[R(G)] tested on Dataset2 for different values of n

V. CONCLUSION

In this paper, a novel method LCT-Net for coal and gangue images recognition is proposed. Instead of recognizing full-scale images, a hyper-parameter n is introduced to acquire the local texture images of coals and gangues, refining the recognition input from complete coal and gangue images into local texture images, which display a higher repetitiveness and therefore have less inner-class difference. The final result is given by the sum of local recognition results and the comparison with a threshold t. Test results show a higher performance in terms of classification accuracy and stability on various datasets compared to other existing methods.

In the feature work, focus will be put on the preprocessing of local texture images, which is expected to amplify the inter-class difference and boost the accuracy and robustness. In addition, experiments will be performed in industry field.

REFERENCES

[1] X. Qi and Z. Yongchuan, “A novel automated separator based on dual energy gamma-rays transmission,” in Meas. Sci. Technol, 2000, pp. 1383.

[2] L. Haonan, S. Baojin, H. Yaqun, H. Jingfeng and H. Qiongqiong, "Research on identification of coal and waste rock based on GLCM and BP neural network," 2010 2nd International Conference on Signal Processing Systems, Dalian, pp. V2-275-V2-278, 2010, doi: 10.1109/ICSPS.2010.5555496.

[3] L. Kai, Z. Xi, and C. Yangquan, “Extraction of coal and gangue geometric features with multifractal detrending fluctuation analysis,″ Applied Sciences, vol. 8, no. 3, pp. 463, Mar 2018.

[4] D. Dongyang, W. Wenze, Y. Jianguo, and Z. Yong, “Classification of coal and gangue under multiple surface conditions via machine vision and relief-SVM,” in Powder Technology, pp. 1024-1028, 2019.

[5] P. Yuanyuan, A. B. Derek, S. Alicja, and C. Jie, “Image recognition of coal and coal gangue using a convolutional neural network and transfer learning,” in Energies, pp. 1735, December 2019.

[6] H. Huichao, Z. Lixin, Z. Jianqing, P. Shuwan, and Z. Kaiting, “Automatic recognition of coal and gangue based on convolution neural network,” in arXiv:1712.00720, December 2017.

[7] Y. Le, Z. Lixin, D. Yongzhao and H. Xuan, “Image recognition method of coal and coal gangue based on partial grayscale compression extended coexistence matrix,” in journal of huaqiao university (natural science), vol. 39, pp. 906-912, November 2018.

[8] L. Man and S. Kaikai, "An image recognition approach for coal and gangue used in pick-up robot," 2018 IEEE International Conference on Real-time Computing and Robotics (RCAR), Kandima, Maldives, 2018, pp. 501-507, doi: 10.1109/RCAR.2018.8621725.

[9] Z. Yuwei, Z. Peng, Y. Chun, and W. Zhi, "Texture and shape biased two-stream networks for clothing classification and attribute recognition," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 13535-13544, doi: 10.1109/CVPR42600.2020.01355.

[10] A. Krizhevsky, I. Sutskever and G. Hinton, “ImageNet classification with deep convolutional neural networks” in Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS 2012), vol. 1, pp. 1097-1105, December 2012.

[11] Z. Jingwang, and L. Lianghui, “Coal-gangue mixture degree recognition using different illuminant method in underground coal cining,” in proceedings of Frontiers in Optics, pp. JW4A.129, January 2019.

[12] L. Su, X. Cao, H. Ma and Y. Li, "research on coal gangue identification by using convolutional neural network," 2018 2nd IEEE Advanced Information Management,Communicates,Electronic and Automation Control Conference (IMCEC), Xi'an, pp. 810-814, 2018, doi: 10.1109/IMCEC.2018.8469674.

[13] N. Amin, T. G. Amin, and Z. Yudong, “Image-based deep learning automated sorting of date fruit,” in Postharvest Biology and Technology, pp. 133-141, 2019.

[14] K. Simonyan, and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Conference Track Proceedings of 3rd International Conference on Learning Representations (ICLR 2015), 2015.

[15] Y. A. Ohali, “Computer vision based date fruit grading system: Design and implementation,” in Journal of King Saud University Computer & Information Sciences, vol. 23 pp. 29-36, 2011.

[16] N. Otsu, "A threshold selection method from gray-level histograms," in IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.

[17] L. Yann, C. Corinna, and B. J. C. Christopher, “The mnist database of handwritten digits,” in yann.lecun.com/exdb/mnist