Coal and Gangue Recognition Method Based on Local Texture Classification Network for Robot Picking

Yuting Xie 1,†, Xiaowei Chi 2,†, Haiyuan Li 3,*; Fuwen Wang 3; Lutao Yan 3; Bin Zhang 4; and Qinjian Zhang 5

1 School of Software, Beijing University of Posts and Telecommunications, Beijing 100876, China; xyt@bupt.cn
2 International School, Beijing University of Posts and Telecommunications, Beijing 100876, China; litwellChi@gmail.com
3 School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China; wfsw990706@163.com (F.W.); yanlt@bupt.edu.cn (L.Y.)
4 School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China; zb362301@126.com
5 School of Mechanical & Electrical Engineering, Beijing Information Science & Technology University, Beijing 100192, China; 13671243929@126.com
* Correspondence: lihaiyuan@bupt.edu.cn
† Yuting Xie and Xiaowei Chi contribute equally in this work.

Abstract: Coal gangue is a kind of industrial waste in the coal mine preparation process. Compared to conventional manual or machine-based separation technology, vision-based methods and robotic grasping are superior in cost and maintenance. However, the existing methods may have a poor recognition accuracy problem in diverse environments since coals and gangues’ apparent features can be unreliable. This paper analyzes the current methods and proposes a vision-based coal and gangue recognition model LTC-Net for separation systems. The preprocessed full-scale images are divided into \( n \times n \) local texture images since coals and gangues differ more on a smaller scale, enabling the model to overcome the influence of characteristics that tend to change with the environment. A VGG16-based model is trained to classify the local texture images through a voting classifier. Prediction is given by a threshold. Experiments based on multi-environment datasets show higher accuracy and stability of our method compared to existing methods. The effect of \( n \) and \( t \) is also discussed.

Keywords: coal and gangue recognition; visual classification for robot picking; deep learning

1. Introduction

Coal gangue, or gangue, is a kind of black or gray solid waste discharged in coal mining and coal washing. Compared to the widely-used fossil fuel, gangue provides much less energy supply and contains less value due to lower carbon content. It can also be a pollution source in the absence of proper treatments [1]. Thus, the separation of gangue and coal in raw coal production workshops is critical to improving coal quality, minimizing the storage and transportation cost, and protecting the environment [2]. Higher utilization rates of coal could also help slow down global warming by reducing total coal burning [3].

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 the separation process. The 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. A 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 densities of coal and gangue. X-rays or gamma-ray transmission sensors are also adopted by workshops, among other machine-based
methods [2]. Besides, there are also other separating methods based on coal’s and gangue’s different physical characteristics such as vibration tests [4]. However, the high mechanical quality requirement to ensure productivity leads to high purchase and maintenance costs of these machines [5]; moreover, soaring energy consumption and potential air pollution can also be a problem.

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

- **Vision unit**: Normally, an industrial camera with illumination devices captures 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 control command based on the recognition result as output.
- **Separation unit**: A physical separation device (e.g., a robotic system with grippers or robotic hand [6,7]) operating based on the output command of the control unit to remove recognized gangues. Sun’s work put forward a coal and gangue separating robot system based on computer vision [8].

![Figure 1. General framework of automated vision-based coal and gangue separation system. An example of a physical separation device is a robot with picking end-effector, e.g., grippers.](image)

Figure 1 shows a simplified working flow: images are captured by the vision unit and fed into the control unit, which performs recognition on each image and gives a control command regarding the enabling status of the separation unit based on the recognition result. If an image is recognized as a gangue, the separation unit will execute the predefined mechanical operation once to remove the recognized gangue from the conveyor.

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

Promising vision-based methods about coal and gangue image recognition models arose over recent years. There are two types of methods: two-step and one-step. For two-step methods, predefined hyper characteristic parameters are first manually extracted from original images, and then classifiers are trained based on these parameters. While for one-step or direct methods, classifiers are trained directly to recognize images instead of manually-extracted parameters. In one-step method, features are usually learned automatically from the raw input compared to two-step method.
Liang’s work [9] was a representation of two-step methods. It extracted eight characteristic parameters from the Gray Level Co-occurrence Matrix (GLCM) of the original full-scale images and trained an SVM model or a B.P. neural network as a prediction boundary. Zhao [10] also analyzed the gray information feature parameters like gray mean value, gray histogram, and slope value. Several solutions are proposed to strengthen the performance of GLCM, like partial grayscale compression extended coexistence matrix [11]. Gray information was extracted to recognize coal and gangue images using partial grayscale compression with extended coexistence. However, their research strongly relies on coal’s and gangue’s distinct light reflection feature, which can be easily affected by the environment.

Liu [12] noticed that the different patterns in the shape of coal and gangue edges could be an indicator of category. The study applied Multifractal Detrending Fluctuation Analysis (MDFA [13]) on the images and trained the classifier for recognition. In Dou’s paper [14], 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. Tripathy [15] extracted color and texture features, respectively, for the latter classification. These works are in the two-step structure, which performs classification with trained models like SVM or neural networks on features extracted from original images.

One-step methods are direct. Pu trained a VGG16 network directly on a one-hundred-sized dataset for classification [16]. The transfer learning method was also applied to fit the model with a limited data volume and obtain an accuracy rate of 82.5%. Hong [17] implemented an improved convolutional neural network based on AlexNet [18] to solve the discrimination problem and discussed the detection and location of target objects against the background. Alfarzaeai [19] carried out a CNN model to classify thermal images of coal and gangue.

By studying the previous works and examining diverse datasets from different environments and devices, we noticed these methods would be prone to fail when applied to datasets generated in different conditions. Inspired by the texture recognition on clothes [20], we assumed that texture is an important factor in classification. Therefore, this paper proposes a Local Texture Classification Network (LTC-Net) method to improve the classification performance in different working conditions.

The rest of this paper is as follows. In Section 2, the recognition problem of coal and gangue classification is presented. In Section 3, the limitation of existing methods is discussed under certain circumstances and proposes a coal and gangue recognition method based on LTC-Net. In Section 4, a group of experiments on various datasets with baseline methods and our method as well as the results and parameters’ effect are discussed. Finally, we draw the conclusion of the paper.

2. Problem Analysis

As mentioned above, previous works mainly focus on the classification methods or algorithm frameworks, with less concentration on the experimental datasets. The results are also given on testing datasets originating the same as the training datasets. Therefore, we consider the lack of generalization ability as a potential problem that a good-performing pre-trained model may fail when adopted to another working environment, which is a common scenario given the variety of coal plants.

2.1. Non-Homogenous Datasets

Datasets are non-homogenous if they are collected under different external conditions and (or) with different devices. The external condition variables include but are not limited to illustration intensity, color temperature and background purity, and the devices could be internally different in resolution, capturing angle and light sensitivity. Figure 2 shows a sample of four non-homogenous datasets, which are detailed and utilized in the following sections. In contrast, datasets are homogenous if collected in an identical external environment with identical devices.
Datasets are non-homogenous if they are collected under different external conditions and (or) with different devices. The external condition variables include but are not limited to illustration intensity, color temperature and background purity, and the devices used in the different datasets. The following sections illustrate the differences between non-homogenous and homogenous datasets.

### 2.1. Non-Homogenous Datasets

As shown in Figure 2, the classification of coal and gangue may be straightforward within homogenous datasets since the physical and optical property inherently differs between these two kinds of rocks. A model that give results merely based on raw numerical information (e.g., contrast, entropy, energy, and inverse different moment) of images may perform well on homogenous datasets, even if no feature of patterns is learned. However, these numerical properties could be vastly different between non-homogenous datasets, causing the performance to intensively deteriorate when a model is experimented on non-homogenous datasets to which it is trained. This conclusion is drawn based on the analysis in Section 4. We consider this as a kind of significant trait trap, because of which a model may lack the ability to generalize to various working conditions, limiting its application.

For two-step indirect methods, such as Gray Value + SVM, GLCM + SVM, and MDFA, although decent accuracy can be reached on the training dataset and its homogenous datasets, a sharp drop occurs in both the accuracy and recall rate when tested on non-homogenous datasets. Even a total failure in recognition could occur. It will be proved in the following part that the inherent differences between non-homogenous 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 one-step direct methods like LeNet [21], an upturn in accuracy could be expected when tested on non-homogenous datasets. However, the result would also deteriorate with resolution degradation and can be adjusted and improved. Moreover, deep learning models trained under the homogenous datasets will also fall into a significant feature trap.

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

### 3. LTC-Net Method

As shown in Figure 3, it is noticed that the yellow spots on gangue and crystal reflection on coal are a neglectful feature on full-scale gray images but significant in the local RGB image. Therefore, inspired by Nasiri’s work [22], a sampling method is used to avoid large inner-class differences and focus more on the larger difference between local textures.

Accordingly, this paper proposes a novel method LTC-Net (Local Texture Classification Network) as shown in Figure 4. As an integer constant $n$ defined, we cut and divide the full-scale preprocessed coal and gangue images into $n \times n$ local texture images, then generate $(n - 1) \times (n - 1)$ classification results by a VGG-based [23] deep learning core. A voting machine produces the final classification result of this stone.
Accordingly, this paper proposes a novel method LTC-Net (Local Texture Classification Network) as shown in Figure 4. As an integer constant is defined, we cut and divide local texture images, then generate (see Section 3.2). Among these datasets, Dataset1 is divided into two homogenous datasets, a large-sized Dataset1_train, and a small-sized Dataset1_test. Dataset2 is the data degradation version of Dataset1 with random light, noises, blur, and resize. Dataset3 and Dataset4 are taken by the same camera with lower resolution which is different from the one Dataset1 used. Besides, Dataset4 is collected under yellow light. The lights and distances to the objects differ, and this leads to a resolution difference after preprocessing (see Section 3.2). 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. Detailed information and comparison of the datasets are shown in Table 1 and Figure 5. The importance of non-homogenous testing datasets is to be further discussed in Section 4.

3.1. Data Setup

One training dataset and four testing datasets are set up by images captured with different devices, resolutions, lights, and angles. We organized our datasets by following important concepts.

Among these datasets shown in Table 1, Dataset1 is divided into two homogenous datasets, a large-sized Dataset1_train, and a small-sized Dataset1_test. Dataset2 is the data degradation version of Dataset1 with random light, noises, blur, and resize. Dataset3 and Dataset4 are taken by the same camera with lower resolution which is different from the one Dataset1 used. Besides, Dataset4 is collected under yellow light. The lights and distances to the objects differ, and this leads to a resolution difference after preprocessing (see Section 3.2). 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. Detailed information and comparison of the datasets are shown in Table 1 and Figure 5. The importance of non-homogenous testing datasets is to be further discussed in Section 4.

3.2. Data Preprocess

In order to minimize the effect from external environment conditions, the original raw images are first processed with a series of blur, binarization, erosion, and dilation operations, as shown in Figure 6.
Table 1. Metadata of dataset.

| Name          | 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                     | 5.0                         | ≈1.0                              |
| Dataset3      | Low resolution, dark lights.       | 49, 49                     | 28                   | 36                     | 5.0                         | ≈0.06                             |
| Dataset4      | Very low resolution, bright lights.| 49, 49                     | 56                   | 77                     | 5.0                         | ≈0.06                             |

Figure 5. Feature parameters vary among datasets: (a) The gray value of images; (b) Pixel numbers.

Figure 6. Intermediate results in preprocessing procedures. (a) A sample original image; (b) Blur operation; (c) Binarization operation; (d) Erosion and dilation operation; (e) Edge detected and bounding-box acquired from binary image and utilized to segment object from original image. (f) Background noise reduced with black color filling.

The preprocessing procedures intends to conduct segmentation and box-bound the object within the image, so that the differences and noises of the background are greatly reduced.

3.3. Local Sampling

By preprocessing procedures, the segmented and background-reduced images are generated for further treatments. As illustrated in the “Local sampling” part in Figure 4 and Algorithm 1, local texture images are acquired by cutting the original full-scale images. By setting a hyper-parameter \( n \), the raw image will be divided equally into \( n \times n \) cut pieces. Parameter \( n \) controls the size of cut pieces and the effect of which is discussed in Section 4.3. For cut images, the outer-ring ones will be abandoned, while the inner \( (n-2) \times (n-2) \) ones are being used. This can be intuitively comprehended as a mechanism to minimize the number of local images with a high proportion of background in datasets, and this is also affected by the value of \( n \). The number of cut images provided for the following model training can be calculated by (1), where \( S_c \) and \( S_g \) are, respectively, the numbers...
of coals and gangues in datasets. The \( n \) is the parameter that controls the size of local sampling pieces.

\[
(Sc + Sg) \times (n - 2) \times (n - 2) \tag{1}
\]

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. Sample local texture images are shown above in Figure 3.

**Algorithm 1.** Local sampling of a coal or gangue image

**Input:** a preprocessed coal or gangue image \( I \) with size \( w \times h \), and a hyper-parameter \( n \).

```plaintext
for i = 1 to \( n-2 \) do
  for j = 1 to \( n-2 \) do
    yield \( I[i \times w/n: (i + 1) \times w/n][j \times h/n: (j + 1) \times h/n] \)
  end
end
```

3.4. **Data Degradation**

In order to provide stable results under the distraction of light, angles, viewpoints, and sizes, it is necessary to perform data augmentation with such methods as random rotate, random scale operations and random noises before training. A randomly selected area with a random light change is also used to simulate the unbalanced light environment. These processes can significantly reduce potential overfitting and improve the generalization ability.

3.5. **Local Texture Classification with CNN Components**

The classification process is local sampling, predicting, and voting. In predicting part, a VGG16-based [23] CNN model is designed and implemented as the local texture classifier. This network is trained with preprocessed cutting images from Dataset1_train and tested on those from Dataset1_test. Details and test results of the model are shown in Table 2.

**Table 2.** 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 extent determines the overall success of the LTC-Net method, it is not the only direct indication of the final recognition results since the original intention of the LTC-Net method is to reduce the instability in classification and improve the robustness.

The test results shown in Table 2 from the above section testified the CNN component’s validation in terms of classification of the local texture images from coals and gangues with a test accuracy above 80%. It is then applied in the recognition process described in Figure 4. For each full-scale image, \((n - 2) \times (n - 2)\) classification results will be given by the “Classification” procedure. In the following process, “Count and Judge” sum the results and compare them 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 Section 3.6.

Algorithm 2. Full-scale recognition.

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

\[
c = 0
\]

for \(i = 1\) to \((n - 2) \times (n - 2)\) 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

3.6. Threshold Computation

The threshold \(t\) is determined based on the counting result given on the training set (Dataset1_train). Taking \(n = 5\) as an example, Figure 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 of 9 can be provided per full-scale image.

![Figure 7](image-url)

**Figure 7.** When \(n = 5\), the value of \(t\) is given by a one-dimensional linear classification model (SVM).

The computation 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 the classification model to learn can cause the problem of existing methods, and also a possible reason for the robustness lift of the LTC-Net method.

Given the counting results, the threshold is calculated through 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 Figure 7.

### 3.7. Implementation

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

### 4. Experiment

#### 4.1. Basic Methods

**Gray Value + SVM:** A classification method performs direct average gray value calculation on specific images in Dataset1_train and uses an SVM model to acquire the category boundary to classify the images in testing datasets.

**GLCM + SVM:** To four key parameters in GLCM: entropy, energy, inverse different moment, and contrast, respectively, are computed for each of Dataset1_train images, and an SVM model is trained to perform classification based on the four parameters.

**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 presents a zero or one output as a category prediction at a full scale and at once. The detailed parameters are identical to the CNN component in Section 3, shown in Table 2.

#### 4.2. Results and Discussion

Similar to other studies, classification accuracy is an overall measure of the LTC-Net method and other baselines. However, the recall rate of coal and the recall rate of gangue should also be a key indicator of effectiveness and robustness. On the one hand, the numbers of coals and gangues are balanced in our testing sets. Therefore, the offline method could, by chance, have 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 the failure of the separation system. On the contrary, coal mines with a high percentage of gangue result in a very high waste of valuable coal.

As shown in Tables 3–6, the LTC-Net method is testified stable and accurate on the four testing datasets. On Dataset1_test, the homogenous dataset of the training set Dataset1_train, LTC-Net secure 97.8% accuracy while baseline methods also have fairly good performances. However, when tested on non-homogenous datasets, baseline methods show a decreasing accuracy while the LTC-Net method can still have decent successful classification rates for both coals and gangues, even on Dataset4, the resolution of which is much lower.
Table 3. Accuracy.

| Method          | 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    |
| LTC-Net         | 0.978         | 0.959    | 0.841    | 0.788    |

Table 4. Recall of coal.

| Method          | 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    |
| LTC-Net         | 0.960         | 0.979    | 0.836    | 0.778    |

Table 5. Recall of gangue.

| Method          | 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    |
| LTC-Net         | 1.000         | 0.939    | 0.846    | 0.802    |

Table 6. Average accuracy of four datasets.

| Method          | 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    |
| LTC-Net         | 1.000         | 0.939    | 0.846    | 0.802    |

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 to recognize gangues. However, it can barely recognize 11% of the coal images, resulting in a total accuracy of only 52.1%. By contrast, LTC-Net could stably give reasonable results, as the illustration condition and resolution of images change in datasets.

In Section 2, the problem of a lack of learnable features and patterns for models and methods was discussed to learn or extract and may be the obstacle for other methods when tested on homogenous datasets, since there was an inherent similarity of coals and gangues. Here, it is a possible explanation for the good performance of the LTC-Net.
method. By dividing the full-scale images into local texture ones, the affection of the inherent similarity problem is mitigated since the classification result of the LTC-Net method is not given directly from a full-scale image or some parameters extracted from a full-scale image. The benefit of this method is concluded.

- 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-homogenous 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 the threshold, rather than directly at once. With random data degradation, the model can capture the distinctive textures (like the significant yellow spots on some gangue) 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 coal.

Moreover, the choice of the threshold can also be a simple parameter to control coal and gangue recall. Gangue is an impurity and pollution in the coal industry. The threshold is adjusted to manipulate whether to remove all gangue with the cost of wasting coal or retain all coal and some contaminants.

### 4.3. Study of Parameters

As introduced in Section 3, the scale of the local texture images is controlled by the value of the hyper-parameter \( n \).

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

With a reasonable assumption that the local texture images could be classified most effectively at a certain scale, controlled trials are conducted using Dataset2. As shown in Figure 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.

![Figure 8](image-url)

**Figure 8.** The accuracy, recall rate of coal (R(C)) and recall rate of gangue (R(G)) tested on Dataset2 for different values of \( n \).

### 5. Conclusions

In this paper, a novel method LTC-Net, for coal and gangue image 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 a 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 classification accuracy and stability on various datasets than other baseline methods.

Recognition and classification are important for coal and gangue processing. In the next work, the robotic manipulator with picking end-effectors will be integrated, and a hand-eye system can be used to separate coal and gangue.

**Author Contributions:** Conceptualization, H.L. and Q.Z.; investigation and methodology, Y.X., X.C., H.L., L.Y. and B.Z.; software, Y.X. and X.C.; validation, Y.X., X.C. and F.W.; writing—original draft preparation, Y.X., X.C. and H.L.; writing—review and editing, Y.X., X.C. and H.L.; supervision, H.L.; project administration, H.L.; funding acquisition, H.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Key Research and Development Program of China, grant number 2019YFB1309800, National Natural Science Foundation of China, grant number 62003048, and Research Innovation Fund for College Students of Beijing University of Posts and Telecommunications, grant number 201904037.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

1. Zhang, J.; Li, L. Coal-Gangue Mixture Degree Recognition Using Different Illuminant Method in Underground Coal Mining. In *Frontiers in Optics + Laser Science APS/DLS*; Optical Society of America: Washington, DC, USA, 2019; p. JW4A-129.
2. Xu, Q.; Zhang, Y. A novel automated separator based on dual energy gamma-rays transmission. *Meas. Sci. Technol.* **2000**, *11*, 1383–1388.
3. Dunmade, I.; Madushele, N.; Adedeji, P.A.; Akinlabi, E.T. A Streamlined Life Cycle Assessment of a Coal-Fired Power Plant: The South African Case Study. *Environ. Sci. Pollut. Res.* **2019**, *26*, 18484–18492. [CrossRef] [PubMed]
4. Yang, Y.; Zeng, Q.; Yin, G.; Yan, L. Vibration Test of Single Coal Gangue Particle Directly Impacting the Metal Plate and the Study of Coal Gangue Recognition Based on Vibration Signal and Stacking Integration. *IEEE Access* **2019**, *7*, 106784–106805. [CrossRef]
5. Baic, I.; Blaschke, W.; Goralczyk, S.; Szafarczyk, J.; Buchalik, G. A New Method for Removing Organic Contaminants of Gangue from the Coal Output. *Rocz. Ochr. Środowiska* **2015**, *17*, 1274–1285.
6. Sun, Z.; Li, D.; Huang, L.; Liu, B.; Jia, R. Construction of Intelligent Visual Coal and Gangue Separation System based on CoppeliaSim. In Proceedings of the International Conference on Automation, Control and Robotics Engineering (CACRE), Dalian, China, 18–20 September 2020; pp. 560–564.
7. Li, M.; Sun, K. An Image Recognition Approach for Coal and Gangue Used in Pick-Up Robot. In Proceedings of the 2018 IEEE International Conference on Real-time Computing and Robotics (RCAR), Kandima, Maldives, 1–5 August 2018; Volume 29, pp. 501–507.
8. Sun, Z.; Huang, L.; Jia, R. Coal and Gangue Separating Robot System Based on Computer Vision. *Sensors* **2021**, *21*, 1349. [CrossRef] [PubMed]
9. Liang, H.; Su, B.; He, Y.; He, J.; He, Q. Research on Identification of Coal and Waste Rock Based on GLCM and B.P. Neural Network. In Proceedings of the 2nd International Conference on Signal Processing Systems, Dalian, China, 5–7 October 2010; pp. 275–278.
10. Zhao, M.; Ma, S.; Zhao, D. Image Processing Based on Gray Information in Sorting System of Coal Gangue. In Proceedings of the 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China, 25–26 August 2018; pp. 81–83.
11. Le, Y.; Zheng, L.; Du, Y.; Huan, X. Image Recognition Method of Coal and Coal Gangue Based on Partial Grayscale Compression Extended Coexistence Matrix. *J. Huazhong Univ.* **2018**, *39*, 906–912.
12. Kai, L.; Xi, Z.; Yangquan, C. Extraction of Coal and Gangue Geometric Features with Multifractal Detrending Fluctuation Analysis. *Appl. Sci.* **2018**, *8*, 463. [CrossRef]
13. Tripathy, D.P.; Reddy, K.G.R. Multispectral and Joint Colour-Texture Feature Extraction for Ore-Gangue Separation. *Pattern Recognit. Image Anal.* **2017**, *27*, 338–348. [CrossRef]
14. Dou, D.; Wu, W.; Yang, J.; Zhang, Y. Classification of Coal and Gangue under Multiple Surface Conditions via Machine Vision and Relief-SVM. *Powder Technol.* **2019**, *356*, 1024–1028. [CrossRef]
15. Ihlen, E.A.F. Introduction to Multifractal Detrended Fluctuation Analysis in Matlab. *Front. Physiol.* **2012**, *3*, 141. [CrossRef] [PubMed]

16. Pu, Y.; Apel, D.B.; Szmigiel, A.; Chen, J. Image Recognition of Coal and Coal Gangue Using a Convolutional Neural Network and Transfer Learning. *Energies* **2019**, *12*, 1735. [CrossRef]

17. Hong, H.; Zheng, L.; Zhu, J.; Pan, S.; Zhou, K. Automatic Recognition of Coal and Gangue Based on Convolution Neural Network. *arXiv 2017*, arXiv:1712.00720.

18. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

19. Alfarzaeai, M.S.A.; Qiang, N.; Zhao, J.; Eshaq, R.; Hu, E. Coal/Gangue Recognition Using Convolutional Neural Networks and Thermal Images. *IEEE Access* **2020**, *8*, 76780–76789. [CrossRef]

20. Zhang, Y.; Zhang, P.; Yuan, C.; Wang, Z. Texture and shape biased two-stream networks for clothing classification and attribute recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 13–19 June 2020; pp. 13535–13544.

21. Su, L.; Cao, X.; Ma, H.; Li, Y. Research on Coal Gangue Identification by Using Convolutional Neural Network. In *Proceedings of the 2nd IEEE Advanced Information Management Communicates, Electronic and Automation Control Conference (IMCEC)*, Xi’an, China, 25–27 May 2018; pp. 810–814.

22. Nasiri, A.; Taheri-Garavand, A.; Zhang, Y. Image-Based Deep Learning Automated Sorting of Date Fruit. *Postharvest Biol. Technol.* **2019**, *153*, 133–141. [CrossRef]

23. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv 2015*, arXiv:1409.1556.