energies

MDPI

Article

Automatic Coal and Gangue Segmentation Using U-Net Based Fully Convolutional Networks

Rong Gao, Zhaoyun Sun, Wei Li *, Lili Pei, Yuanjiao Hu and Liyang Xiao

School of Information Engineering, Chang'an University, Xi'an 710064, China; gr@chd.edu.cn (R.G.); zhaoyunsun@126.com (Z.S.); peilili@chd.edu.cn (L.P.); sandra@chd.edu.cn (Y.H.); liyang_xiao@163.com (L.X.)
* Correspondence: grandy@chd.edu.cn

Received: 23 November 2019; Accepted: 11 February 2020; Published: 14 February 2020



Abstract: Sorting gangue from raw coal is an essential concern in coal mining engineering. Prior to separation, the location and shape of the gangue should be extracted from the raw coal image. Several approaches regarding automatic detection of gangue have been proposed to date; however, none of them is satisfying. Therefore, this paper aims to conduct gangue segmentation using a U-shape fully convolutional neural network (U-Net). The proposed network is trained to segment gangue from raw coal images collected under complex environmental conditions. The probability map outputted by the network was used to obtain the location and shape information of gangue. The proposed solution was trained on a dataset consisting of 54 shortwave infrared (SWIR) raw coal images collected from Datong Coalfield. The performance of the network was tested with six never seen images, achieving an average area under the receiver operating characteristics (AUROC) value of 0.96. The resulting intersection over union (IoU) was on average equal to 0.86. The results show the potential of using deep learning methods to perform gangue segmentation under various conditions.

Keywords: coal and gangue; U-Net; feature extraction; segmentation

1. Introduction

Gangue is a solid waste with low carbon content, which is usually mixed into the raw coal during production. Due to the inefficiency of gangue separation, much gangue flows out of the mining area, which increased the transporting costs and caused severe pollution to the environment [1]. Therefore, effectively sorting gangue from the raw coal is essential in improving the quality of the coal and reducing the costs of transport. With the increase of labor costs and the need to avoid hazards to worker's health, automatically separating gangue from raw coal has become a critical issue in recent years. As a contactless inspection technology, computer vision has been widely applied in driverless cars [2], medical diagnostics, remote sensing, mineral processing, and many other fields. From an engineering perspective, it seeks to automate many tasks the human vision can perform. Human beings can distinguish gangue and coal by the differences in brightness, color, morphology, texture, and other features. So far in the literature, a number of studies have been conducted on the gangue image features extracting algorithms to separate gangue from coal.

Tripathy [3] suggested a vision-based gangue sorting model based on the analysis of color texture and a multilayer perceptron (MLP) neural network. Color texture features were extracted from hue saturation value (HSV) and luminance chrominance (YCbCr) color spaces, respectively, which were used as inputs to the MLP neural network to sort gangue. Hong [4] built a deep learning model using a convolutional neural network (CNN) and transfer learning to distinguish coal and gangue images. The typical workflow for CNN image recognition was presented and the model was tested with photos from a washing plant. Su [5] improved the LeNet-5 coal gangue identification model to achieve a recognition rate of 95.88%. Many other similar kinds of literature are not listed because of the limited

length. These studies have primarily focused on classifying the image category, (i.e., detecting whether gangue or coal existed in the image). The results of these studies would not provide the position and shape information of gangue in the images to guide the sorting manipulators.

Image segmentation is an important task in computer vision facilitated areas. The results of image segmentation can divide the image into object regions that can provide the position of the target in the image. Gao [6] proposed an algorithm based on grayscale feature decision theory to recognize gangue over a moving belt conveyor. The threshold of the grayscale distribution histogram was calculated by the Bayesian decision theory. Wei [7] proposed a method of combining image feature extraction and artificial neural network to identify gangue, and the performance of the algorithm was tested on a dataset consisted of 20 images. Liu [8] proposed a computer vision approach for feature extraction of coal and gangue. Grayscale and texture features were extracted using multifractal detrending fluctuation analysis (MFDFA). Sun [9] proposed a morphology-based method to separate gangue from coal, and the supplementary texture is extracted based on morphology. Pre-treatments of coal and gangue images have been introduced. Li [10] proposed an image-based hierarchical deep learning framework for coal and gangue detection. Gaussian pyramid principle was applied to construct the training dataset, and convolution neural networks were used as a feature extractor to generate the classification and detection box.

Though previous research has reported acceptable detection or segmentation accuracy for dealing with gangue images, the existing approaches still need to be improved because the gangue and coal in their images datasets were sparse, the entire image even contained individual categories. However, images collected on the conveyor belt in a coalfield are often complicated, with coal and gangue randomly heaped with ash. Thus, comprehensive studies on the gangue segmentation methods with regard to complex environmental conditions call for much attention.

Deep learning, as one of the most interested scientific research trends [11,12], has brought revolutionary advances in computer vision and machine learning [13]. Unlike traditional machine learning techniques, it can automatically learn representations from images without introducing hand-coded rules or human domain knowledge [14]. Convolutional neural networks (CNNs), one of the most popular methodologies in deep learning, have been widely applied as an efficient architecture for feature extraction [15]. A fully convolutional network (FCN) is a particular type of CNN, which replaces the fully connected layer with convolution operation, and has been proven to be an ideal way to accomplish the image segmentation tasks [16]. The U-Net has further improved this method, a network structure first created by Ronneberger [17] and applied to biomedical image segmentation. The outstanding performance of U-Net in biomedical image segmentation without the need for a massive number of training images inspired the authors to examine the performance of U-Net in gangue image segmentation. To the best of our knowledge, this is the first research work considering U-Net in gangue sorting.

In this article, a fully convolutional neural network model called U-Net was applied to segment gangue from raw coal images collected under complex conditions. The main approach was to build an end-to-end deep learning model, which could be generalized to the complex conditions in applying to a coalfield. The U-Net model was trained using 54 manually labeled images. The trained U-Net model was able to provide probability maps from six unseen testing images. The model was successfully deployed to guide two manipulators sorting gangue from a moving conveyor belt. The results of the present research indicated that the proposed solution is able to segment gangue from raw coal effectively. The rest of this paper is organized as follows. First, in Section 2, the collection and preparation of the raw coal images used in our study are described. This is followed by a detailed description of the proposed approach. The obtained results are presented and discussed in Section 4. Finally, Section 5 concludes the research.

2. Input Data

This research investigated the coal and gangue segmentation based on U-Net using raw coal images as the input. Figure 1 shows the schematic of the proposed method. During the training stage, manually labeled training data were used to fine-tune the model parameters until the difference between the predicted results and the ground-truth remained stable. In the testing stage, the trained U-Net model produced a pixel-level probability map instead of classifying an input image.

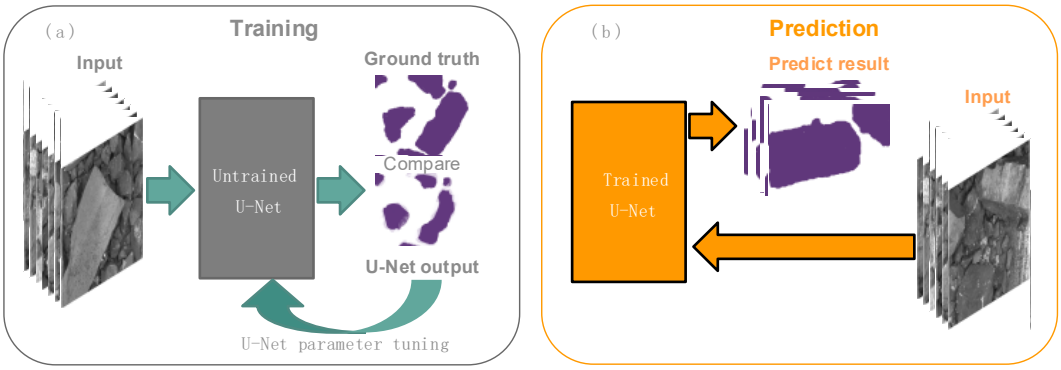


Figure 1. An overview of our deep learning-based coal and gangue segmentation technique. (**a**) Training stage; (**b**) Testing stage.

2.1. Raw Data Collection

The sample images were collected from the raw coal produced in Datong Coalfield. The Datong Coalfield is located in northern Shanxi Province. The data collection and gangue grabbing manipulator are shown in Figure 2. Sample images were captured by a BlueVision BV-C2901-GE air cooling SWIR camera illuminated by 8×500 W iodine tungsten lamps. The camera employs an InGaAs sensor with a pixel size of 20 µm. The camera and light source were mounted above a conveyor belt with a width of 800 mm. Images captured by the camera were transferred to a computer for display and storage by the Gigabit Ethernet (GigE) protocol. The encoder on the conveyor provided the frame trigger signal for the camera. To ensure the diversity of sample images, sixty 8-bit grayscale raw coal images with a spatial resolution of 512×512 pixels were collected from six different production batches.

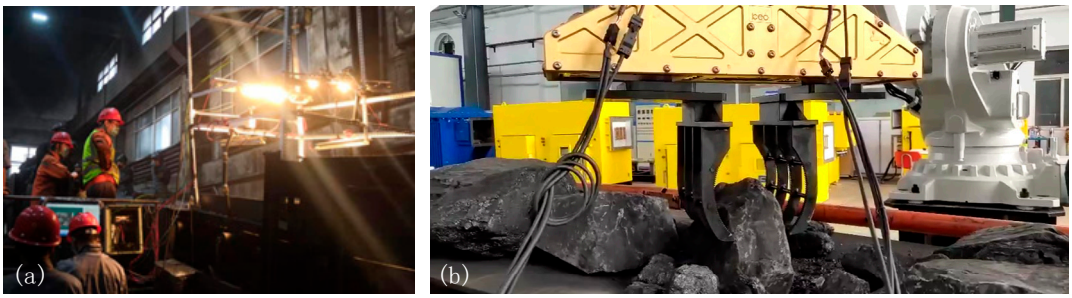


Figure 2. Data collection and gangue separating systems: (**a**) Data collection and; (**b**) gangue grabbing manipulator.

2.2. Data Preparation

Sixty gray-scale raw coal images were collected for model training, each image had a dimension of 512×512 pixels for the experiment. Human vision could easily distinguish the differences between gangue and the background (coal and belt) based on the surface color, texture, and edge features. The surface of coal was darker than that of gangue, but with reflective spots randomly distributed. On the other hand, the intensity of the gangue surface was distributed uniformly with only a few scratches that appeared as brighter lines or points. To imitate the powerful analytical and recognition capabilities

of human vision, and teach the U-Net to distinguish gangue from the background, we labeled each pixel in the images to mark it as gangue or the background. The Colabeler AI tool was used to carry out the labeling task by providing ground truth for gangue images. The ground truth was not perfect because of the time-consuming nature of labeling, especially since small gangue was ignored intentionally. Sixty pairs of images and ground truth were used to train and test the proposed U-Net.

2.3. Data Augmentation

Data augmentation is a strategy that enables practitioners to significantly increase the diversity of data and reduce over-fitting in model training without collecting new data. For each image on the available training dataset, we generated a new image by using a combination of traditional transformations that were rotated, flipped, shifted, or zoomed in and out [18]. Generated images were fed into the neural network with the original images. This allows the network to learn invariance to such deformations, without the need to see these transformations in the annotated images. Figure 3 shows the data augmentation results under the rotation transformation. The original image and the corresponding ground-truth were transformed by the same scale.

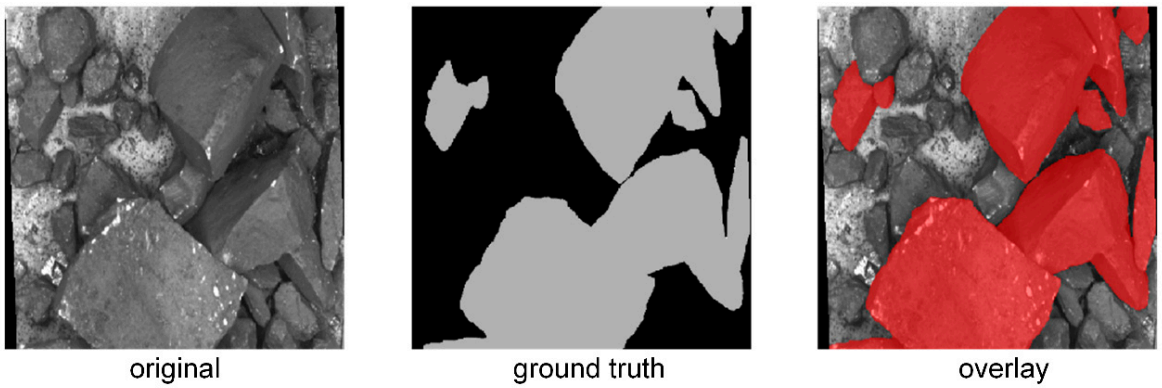


Figure 3. Data augmentation under rotation transformation.

3. The Proposed Approach

3.1. Network Architecture

The architecture of the CNN used for gangue image segmentation was presented in Figure 4. It was derived from the U-Net network proposed in [17]. The U-Net consists of an encoder to condense the spatial information and a decoder that enables precise localization. As a result, it outputs a pixel-level probability map instead of detecting an input image as a whole [19]. Compared to the original architecture, each convolution operation reduces the size of the feature map, which leads to an output image size smaller than the input image. To preserve the boundary of feature maps, we applied a 3×3 convolution with all-zero padding in both the encoder and decoder to extract features. Convolutional neural networks employ a shared-weights architecture and translation invariance strategy that leads to a shift-invariant or space invariant characteristic [20]. By utilizing the weights sharing strategy, neurons were able to perform convolutions on the input images with the convolution filters formed by the weights. The extracted feature maps from the proposed convolutional procedure were as follows:

$$\mathbf{X}^{l}_{i,j} = \sum_{m=0}^{k_1} \sum_{n=0}^{k_2} \mathbf{W}^{l}_{m,n} \cdot \mathbf{O}^{l-1}_{i+m,j+n} + b^{l} \tag{1}$$

where l is the l^{th} layer; $l = 1$ is the first layer; and $l = L$ is the last layer. Input $\mathbf{X}$ is of dimension 512×512 and has i by j as the iterators. The kernel $\mathbf{W}$ is of dimension 3×3 and has m by n as the iterators. $\mathbf{W}^{l}_{m,n}$ is the weight matrix connecting neurons of layer l with neurons of the layer $l - 1$. b^{l}

is the bias unit at layer l. $\mathbf{X}^l_{i,j}$ is the convolved input vector at layer l plus the bias. $\mathbf{O}^l_{i,j}$ is the output vector at layer l given by $\mathbf{O}^l_{i,j} = f(\mathbf{X}^l_{i,j})$. $f(\cdot)$ is the activation function.

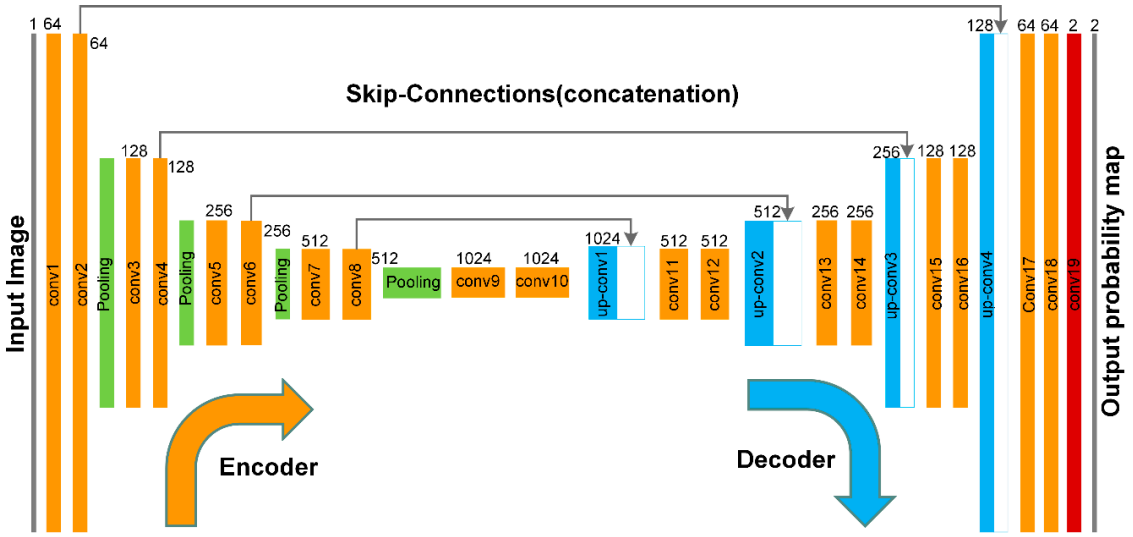


Figure 4. Structure of the proposed U-Net model.

Each convolution operation was followed with the rectified linear unit (ReLU) activation function, which allows the neural network to fit more nonlinear functions. ReLU [21] is a widely used activation function in neural networks and is defined as follows:

$$f(\cdot) = \text{ReLU}(x) = \max(0, x) \tag{2}$$

The feature maps extracted by the convolution layer were equal in size to the input image, and processing this number of feature maps will meet the computational challenges and increase the number of parameters in the model. In the proposed network, max-pooling operations with kernel size $= 2 \times 2$, stride $= 2$ were applied to halve the size of the feature maps.

A pixel-wise softmax was applied in the proposed neural network that yielded the predicted probability scores. The range of the predicted score was limited from 0.0 to 1.0 for each class (gangue or the background), whereas the summation of all probabilities was equal to 1. The softmax function can be defined as:

$$\mathbf{y}_p = \Phi_{\text{softmax}}(\mathbf{O}^L) = \frac{e^{\mathbf{O}^L_p}}{e^{\mathbf{O}^L_p} + e^{\mathbf{O}^L_q}} \tag{3}$$

where $\mathbf{O}^L_p$ and $\mathbf{O}^L_q$ denote the feature maps produced by the final layer. $\mathbf{O}^L_p$ corresponding to the probabilities that the pixel belongs to gangue, while $\mathbf{O}^L_q$ denotes the probabilities to the background. The computed result of the softmax function $\mathbf{y}_p$ indicates the probability of a pixel belonging to gangue.

As in the original U-Net architecture, the network used in this study consists of two parts: the encoder progressively down-samples the feature maps and doubles the number of feature maps per layer at the same time. Meanwhile, every step in the decoder consists of an up-sampling of the feature map followed by a 2×2 convolution. The decoder recombined the up-sampled feature maps with the high-resolution features from the encoder via skip connections, which gradually increased the lateral details of the feature maps.

The above architecture is significantly different from the CNN based approach used previously for gangue detection by [4,5,10]. Their architecture contains only the encoder part to extract image features followed by the fully connected layers. Each input image is classified as gangue or non-gangue as

a whole based on features extracted by the encoder that would not provide the position and shape information of gangue.

3.2. Model Training

The proposed U-Net model was trained using raw coal images and the corresponding ground truth. A total of 48 pairs from 60 were randomly selected to train the network. The remaining 12 were equally divided for validation and testing. The network was implemented in Python 3.6 programming language with the use of the Keras library and the TensorFlow backend. The Adam optimizer (learning rate = 0.001, exponential decay rates $\beta_1 = 0.9$ and $\beta_2 = 0.999$) was used to minimize the cross-entropy loss. A dropout rate of 0.2 was utilized to combat overfitting. The network was trained with 600 epochs which took about 5.8 h on a GeForce GTX 1080Ti GPU equipped with 32 GB of DDR5 RAM memory. The segmentation of a single test image by the trained network took 48.2 milliseconds on average.

3.3. Evaluation Metrics

The performance of the trained U-Net model was evaluated with the commonly used evaluation metrics: area under the receiver operating characteristics (AUROC), area under the precision–recall curve (AUPRC), and intersection over union (IoU). The gangue pixels (white pixels in images) were defined as positive instances. According to the combinations of ground-truth case and predicted case, all pixels were divided into four types: false positive (FP), false negative (FN), true positive (TP), and true negative (TN), the definition of which is listed in Table 1. FP, FN, TP, and TN represent the summation of pixels belonging to each type, respectively [22].

Table 1. Definition of the evaluation metrics.

		Ground Truth	
		Gangue (P)	**Background (N)**
Predicted Result	Gangue (P′)	True Positive (TP)	False Positive (FP)
	Background (N′)	False Negative (FN)	True Negative (TN)
ACCURACY $= \frac{TP+TN}{P+N}$	PRECISION $= \frac{TP}{TP+FP}$	RECALL $= \frac{TP}{TP+FN}$	TPR $= \frac{TP}{TP+FN}$	FPR $= \frac{FP}{FP+TN}$	

Figure 5 shows the schematic diagram of IoU, which is defined as the size of the intersection divided by the size of the union of the sample sets. When it was applied to evaluate the similarity between the ground-truth and segmentation results, it can be expressed as follows:

$$\mathrm{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} = \frac{TP}{TP + FP + FN} \tag{4}$$

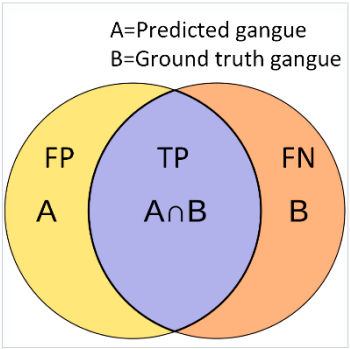


Figure 5. Schematic diagram of IoU.

The computed IoU has a value between 0 and 1, and the closer this value is to 1.0, the better the similarity between the segmentation results and the ground-truth.

4. Results and Discussion

This section presents the results of the gangue image segmentation obtained by the proposed approach and the comparison experiments with other CNN based approaches in [4,5,10]. Meanwhile, a comprehensive experiment for evaluating the effectiveness of data augmentation was conducted, and the impacts of different input image size were evaluated.

4.1. Visual Results

The visual results of the proposed approach are shown in Figure 6. The original raw coal images unseen during the training process are shown in the first column of Figure 6. The second column shows the ground truth images, the third and the fourth columns are the results of the proposed approach and the original images overlaid by the probability maps, respectively.

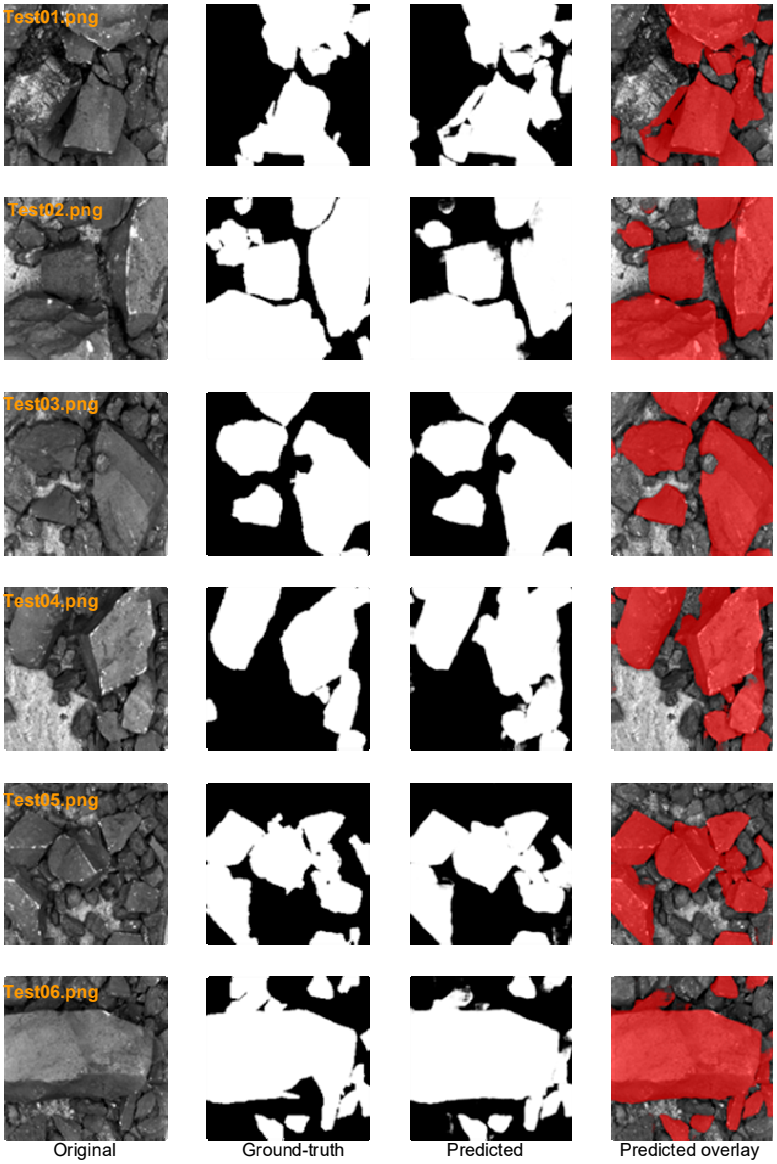


Figure 6. The results of gangue segmentation using the U-Net based approach: first column, testing images; second column, the manually labeled ground truth; third column, the probability maps generated by the trained model; and the last column; results overlaid on the original images.

4.2. Network Performance

A quantitative assessment of the proposed approach was performed threefold. First, the convergence of the proposed U-Net model is shown in Figure 7. In particular, Figure 7a corresponds to the accuracy of the training and validation in each epoch, while Figure 7b presents the loss curves. The validation loss curves of the proposed network converged in around 100 epochs, which shows that the proposed model is effective in training.

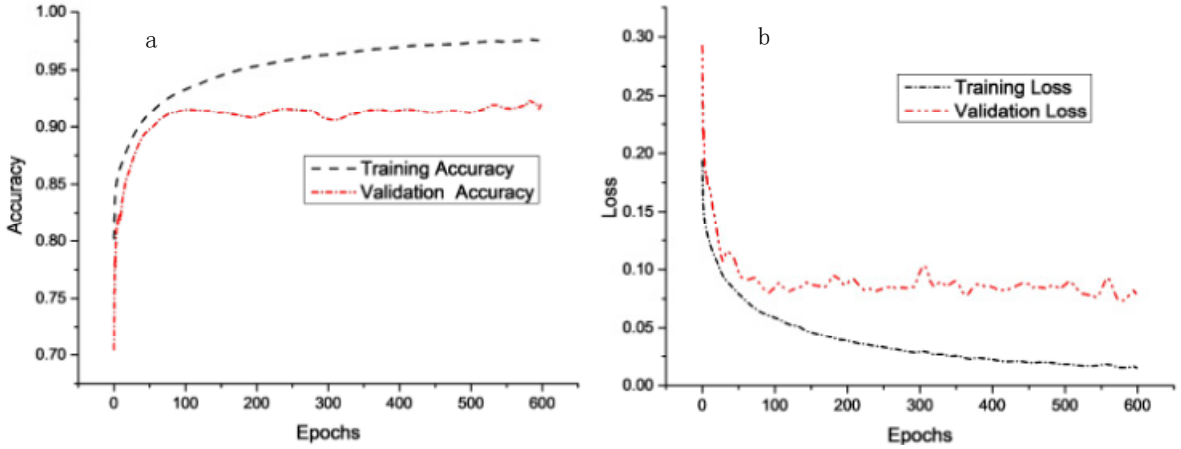


Figure 7. Accuracy and loss over epochs. (**a**) The accuracy curves; (**b**) the loss curves.

Second, the performance of the trained U-Net in segmenting the testing dataset was evaluated. AUROC and AUPRC were used as the performance evaluation metrics in this part. Figure 8a presents the precision–recall curves with the corresponding AUC value equal to 0.96, while Figure 8b shows the ROC curves with the AUC value equals to 0.96. Third, the AUC values obtained for each image within the testing dataset are shown in Table 2. Meanwhile, the accuracy, precision, recall, and IoU values of each testing image are listed. These indicators were calculated through the binarization of the probability map with a threshold of 0.5.

Though it is hard to provide an explicit description on the U-Net model (a common challenge in deep learning), some insights can be obtained by providing visualizations of some intermediate layers that made up the proposed network. Some activation maps of the trained U-Net were visualized when inputting the third test image as seen in Figure 9. The basic mechanism of deep learning is to identify statistical invariance through the model training process. It is convincing that the proposed U-Net model indeed had the capability to detect image features such as edge, brightness, morphology, texture, and other features to distinguish gangue from the background.

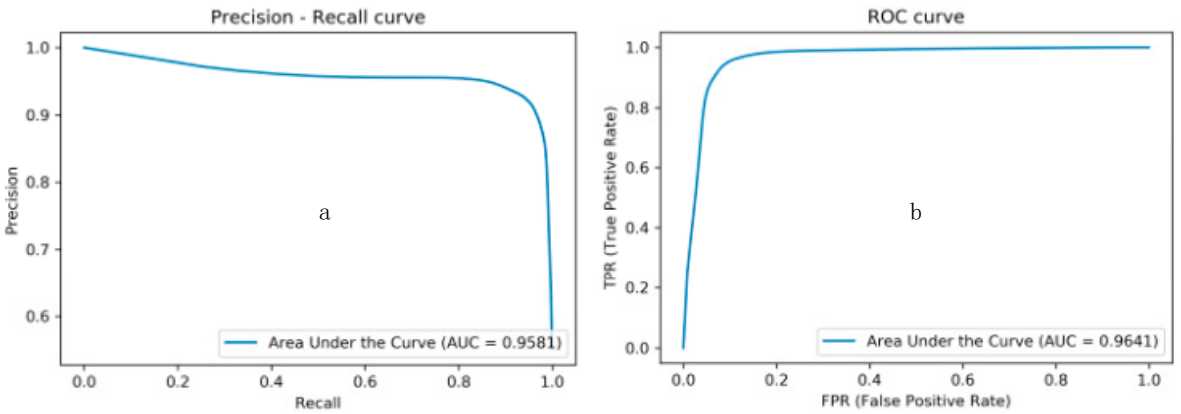


Figure 8. The P-R curves and the ROC curves with their corresponding AUC values. (**a**) The P-R curve; (**b**) the ROC curve.

Table 2. The performance of the trained model.

Images	AUROC	AUPRC	Accuracy	Precision	Recall	IoU
Test01.png	0.94	0.90	0.88	0.83	0.94	0.80
Test02.png	0.98	0.99	0.93	0.97	0.81	0.79
Test03.png	0.99	0.99	0.98	0.98	0.98	0.96
Test04.png	0.94	0.90	0.90	0.83	0.99	0.82
Test05.png	0.99	0.99	0.95	0.95	0.96	0.92
Test06.png	0.96	0.97	0.93	0.90	0.91	0.82
Mean (σ)	0.96 (2.13%)	0.96 (4.07%)	0.93 (3.24%)	0.90 (6.19%)	0.94 (6.04%)	0.86 (6.44%)

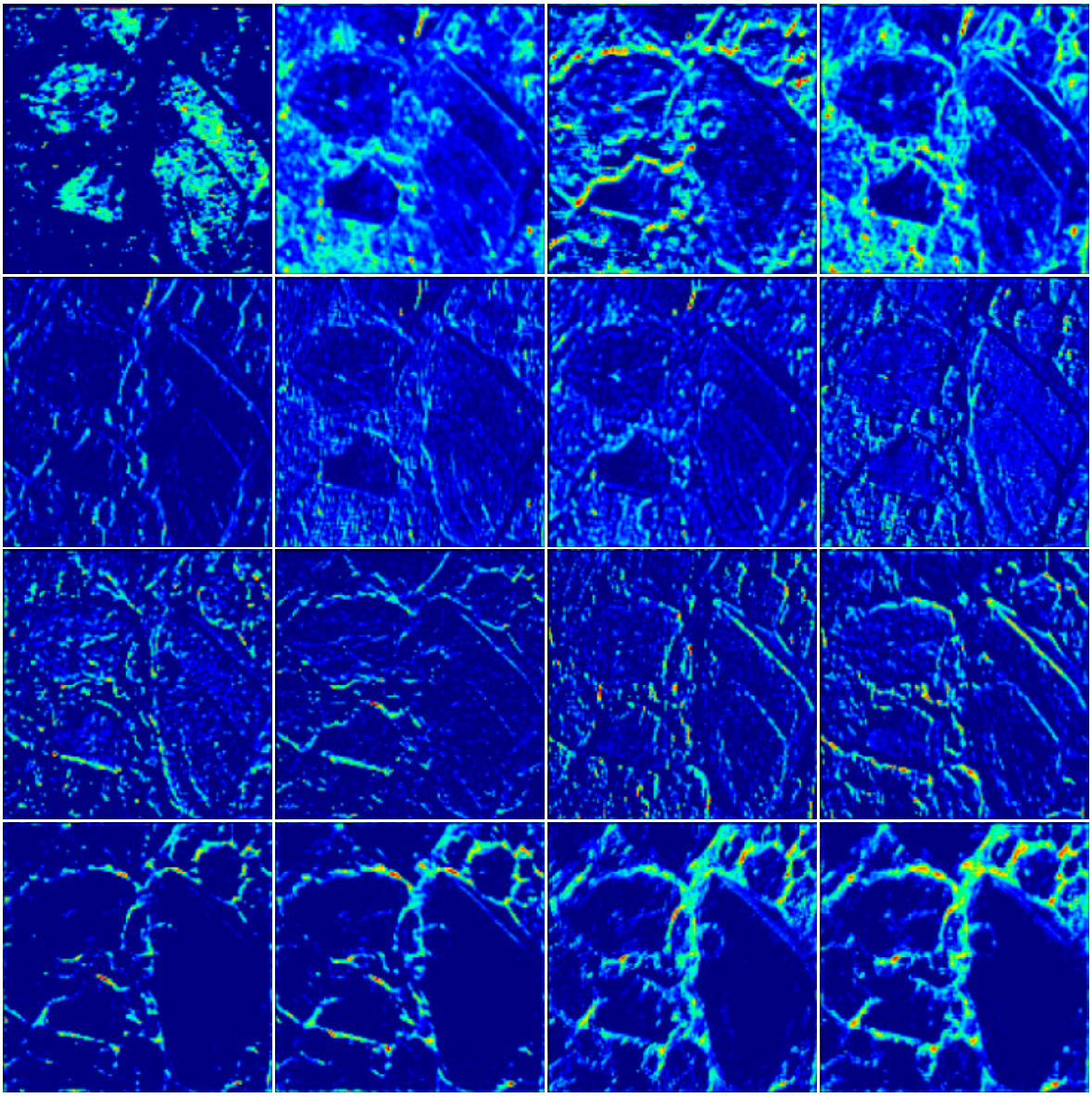


Figure 9. Activation maps of intermediate layers by inputting the third test image.

4.3. Effects of Data Augmentation

In this experiment, the effect of data augmentation on the performance of the model was analyzed. Table 3 shows the confusion matrixes of the trained model. Figure 10 presents the performance of the trained model without data augmentation. Figure 10a shows the precision–recall curves with the corresponding AUC values, while Figure 10b shows the ROC curves with the AUC values. Both models

were trained with the same system configurations (i.e., as described in Section 3.2). By observing the test statistics, it can be seen that the model trained with data augmentation achieved a better performance.

Table 3. Impacts of data augmentation.

	Without Augmentation		**With Augmentation**	
Confusion matrix	654,689 73,321	77,528 767,326	661,849 40,913	70,368 799,734
Accuracy	0.90		0.93	
Precision	0.89		0.90	
Recall	0.90		0.94	
IoU	0.81		0.86	

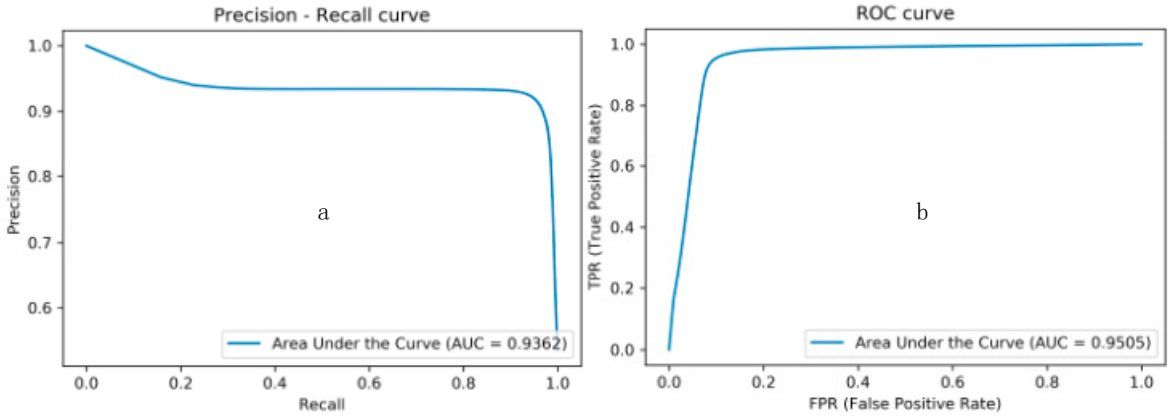


Figure 10. Performance of the trained model without data augmentation. (**a**) The P-R curve; (**b**) the ROC curve.

4.4. Impacts of Input Image Size

The size of the input image was reduced to half of the original ones to train and evaluate the flexibility of the model. Table 4 shows the influences of the size of the input images on the time consumption of training a model and predicting per image. At the same time, the performance difference of the physical size per pixel is shown in the last column. Figure 11a presents the precision–recall curves of the compared methods with the corresponding AUC values, while Figure 11b shows the ROC curves with the AUC values. Comparing the AUC values of Figures 8 and 11, it can be seen that different input image sizes provided equivalent results for both AUPRC and AUROC. However, training models with larger input image size takes a longer time, and the model takes longer to predict each image (see Table 4).

Table 4. Impacts of the input image size.

Task	**Predict Time**	**Train the Model**	**Physical Size Per Pixel**
512 × 512 Input size	48.2 ms per image	5.8 h	1.56 mm
256 × 256 Input size	18.5 ms per image	1.8 h	3.13 mm

It is worth noting that a higher resolution camera can provide more available details in the collected images. The field of view of the camera is the same width as the 800 mm conveyor belt. The physical size of each pixel can be calculated by dividing 800 mm by the width of the input image. A smaller physical pixel resolution can usually provide a more accurate gangue location and shape information to the controller of manipulators, which is beneficial for gangue sorting. However, high quality images should be collected by more advanced devices and processed by higher configured computers, which calls for more capital investments. The size of the input image was carefully selected

according to the balance between the segment accuracy and the time consumption. The physical size of each pixel and the time required for processing of each image (see the second row of Table 4) are sufficient for gangue segmentation while ensuring real-time segmentation.

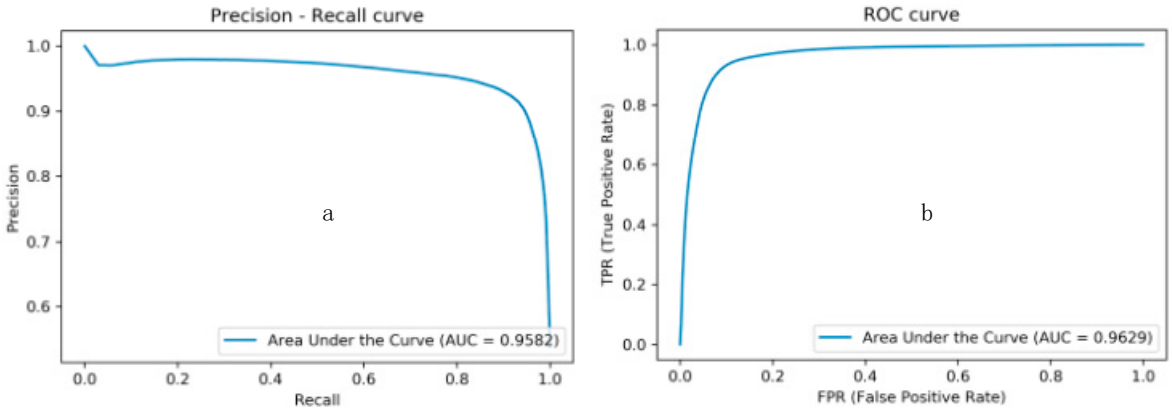


Figure 11. Performance of the trained model with an input image size 256 × 256. (**a**) The P-R curve; (**b**) the ROC curve.

4.5. Comparisons with Other Methods

The proposed method was compared with three other CNN based methods for gangue detection. The sample images used by the previous studies [3,4] contain only a single object. Though several objects were contained in the sample images of [10], the relative positions among them are very sparse. The sample images used in this study were featured the gangue and coal heaped randomly. Table 5 shows the performance for gangue detection using both this method and the other three methods. The size of the dataset used in this study was much smaller than in other studies. This indicates that the proposed method is faster than the three other methods. Additionally, the images used in our study were more complicated, with coal and gangue randomly heaped and contained much ash. The boundaries of gangue are clearly seen in the probability map by the proposed approach, which demonstrates that the proposed method can be generalized to the complex conditions in a coalfield. Finally, as mentioned earlier, the proposed method can output a pixel-level probability map. Therefore, the detection accuracy (see the last column of Table 5) refers to the percentage of correctly identified gangue and background pixels. However, the accuracy in the other three studies refers to the percentage of correctly identified gangue and coal samples.

Table 5. Comparison to the previous studies.

Methods	Dataset Size	Image Attributes	Coal and Gangue Position	Algorithm Category	Detection Accuracy
LeNet [3]	20,000	224 × 224 8-bit	Single	Classification	95.9%
AlexNet [4]	2012	224 × 224 RGB	Single	Classification	96.0%
CG-RPN [10]	2316	1024 × 1024 RGB	Sparse	Classification with box	98.3%
Our method	60	512 × 512 8-bit	Multiple heaped	Pixel-wise segmentation	93.0%

5. Conclusions

The method for the segmentation of gangue images presented in this paper features a novel structure of convolutional neural networks, which have the capability of extracting multiple features of the images. A fully convolutional neural network called U-Net was constructed initially, followed by training the network using the coal image dataset that have been collected under complex environmental

conditions. The trained U-Net is able to segment gangue pixels with an accuracy of a human capability (AUC = 0.96). For some testing images (see the third row in Figure 6), the results were perfect, meaning that almost each gangue pixel within the image was correctly predicted. It is also worth noting that a lump of small coal heaped on the edge of a gangue (see the third row in Figure 6), and the predicted borders around the coal were in good agreement with the contours of the ground truth.

These results demonstrate that U-Net is promising for the industry application of coal image segmentations. However, this method still needs future improving to be used in more areas. It should be noted that the method developed in this research is based on the image data collected from Datong Coalfield, Shanxi, China. The coals obtained in this region are Middle Jurassic coals that are characterized by low ash yield content, low moisture content, low-medium volatile bituminous, and ultra-low sulfur content. Since the coals from the different areas should differ significantly from each other, the method proposed in this research may not be able to obtain similar results for other images, especially for the situation where the geological condition of two coalfields were significantly different from each other. However, the results of this research provide an insight into using a similar method for more complex coal image segmentation. Additionally, the model can be improved by training on more comprehensive datasets, which should give more satisfying results in multiple coal deposits, which are also the future work of this research.

Author Contributions: Conceptualization, W.L. and R.G.; Methodology, R.G.; Software, R.G.; Validation, R.G., W.L. and L.X.; Formal analysis, L.X.; Investigation, W.L.; Resources, L.P.; Data curation, R.G.; Writing—original draft preparation, R.G.; Writing—review and editing, L.X.; Visualization, Y.H.; Supervision, Z.S.; Project administration, Z.S.; Funding acquisition, W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Key Research and Development Program of China (No. 2018YFB1600202); the National Natural Science Foundation of China (No. 51978071); and the Fundamental Research Funds for the Central Universities (No. 300102249301, No. 300102249306).

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

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wu, B.; Yang, M. Analysis on technical and economic policy of comprehensive utilization of coal ash and gangue in China. *Energy China* **2012**, *11*, 8–11.
2. Wang, H.; Wang, Y.; Zhao, X.; Wang, G.; Huang, H.; Zhang, J. Lane Detection of Curving Road for Structural High-way with Straight-curve Model on Vision. *IEEE Trans. Veh. Technol.* **2019**, *68*, 5321–5330. [CrossRef]
3. Tripathy, D.P.; Reddy, K.G.R. Novel Methods for Separation of Gangue from Limestone and Coal using Multispectral and Joint Color-Texture Features. *J. Inst. Eng. Ser. D* **2017**, *98*, 109–117. [CrossRef]
4. 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.
5. 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; IEEE: Piscataway, NJ, USA, 2018.
6. Gao, K.; Du, C.; Wang, H.; Zhang, S. An efficient of coal and gangue recognition algorithm. *Int. J. Signal Process. Image Process. Pattern Recognit.* **2013**, *6*, 345–354.
7. Hou, W. Identification of coal and gangue by feed-forward neural network based on data analysis. *Int. J. Coal Prep. Util.* **2019**, *39*, 33–43. [CrossRef]
8. Liu, K.; Zhang, X.; Chen, Y. Extraction of Coal and Gangue Geometric Features with Multifractal Detrending Fluctuation Analysis. *Appl. Sci.* **2018**, *8*, 463. [CrossRef]
9. Sun, Z.; Lu, W.; Xuan, P.; Li, H.; Zhang, S.; Niu, S.; Jia, R. Separation of gangue from coal based on supplementary texture by morphology. *Int. J. Coal Prep. Util.* **2019**, 1–17. [CrossRef]
10. Li, D.; Zhang, Z.; Xu, Z.; Xu, L.; Meng, G.; Li, Z.; Chen, S. An Image-Based Hierarchical Deep Learning Framework for Coal and Gangue Detection. *IEEE Access* **2019**, *7*, 184686–184699. [CrossRef]

11. Minar, M.; Naher, J. Recent advances in deep learning: An overview. *arXiv* **2018**, arXiv:1807.08169.
12. Shahri, A.A.; Spross, J.; Johansson, F.; Larsson, S. Landslide susceptibility hazard map in southwest Sweden using artificial neural network. *Catena* **2019**, *183*, 104225. [CrossRef]
13. Ghaderi, A.; Shahri, A.A.; Larsson, S. An artificial neural network based model to predict spatial soil type distribution using piezocone penetration test data (CPTu). *Bull. Eng. Geol. Environ.* **2019**, *78*, 4579–4588. [CrossRef]
14. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
15. Buslaev, A.; Seferbekov, S.S.; Iglovikov, V.; Shvets, A. Fully Convolutional Network for Automatic Road Extraction from Satellite Imagery. In Proceedings of the CVPR Workshops, Salt Lake City, UT, USA, 18–22 June 2018.
16. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
17. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015.
18. Wang, J.; Perez, L. The effectiveness of data augmentation in image classification using deep learning. *arXiv* **2017**, arXiv:1712.04621.
19. Fabijańska, A. Segmentation of corneal endothelium images using a U-Net-based convolutional neural network. *Artif. Intell. Med.* **2018**, *88*, 1–13. [CrossRef] [PubMed]
20. Zhang, W.; Itoh, K.; Tanida, J.; Ichioka, Y. Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Appl. Opt.* **1990**, *29*, 4790–4797. [CrossRef] [PubMed]
21. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010.
22. Zhang, Z.; Liu, Q.; Wang, Y. Road extraction by deep residual u-net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).