

Retraction

Retracted: Intelligent Detection of Foreign Matter in Coal Mine Transportation Belt Based on Convolution Neural Network

Scientific Programming

Received 18 July 2023; Accepted 18 July 2023; Published 19 July 2023

Copyright © 2023 Scientific Programming. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] G. Ma, X. Wang, J. Liu et al., "Intelligent Detection of Foreign Matter in Coal Mine Transportation Belt Based on Convolution Neural Network," *Scientific Programming*, vol. 2022, Article ID 9740622, 10 pages, 2022.

Research Article

Intelligent Detection of Foreign Matter in Coal Mine Transportation Belt Based on Convolution Neural Network

Guanchao Ma ¹, Xisheng Wang ², Jianfeng Liu,¹ Weibiao Chen,¹ Qinghe Niu,¹ Yun Liu,² and Xiaocheng Gao¹

¹China Coal Shaanxi Yulin Daze Coal Industry Co., Ltd, Yulin City, Shaanxi 719000, China

²China Coal Information Technology (Beijing) Co., Ltd, Chaoyang District, Beijing 100120, China

Correspondence should be addressed to Xisheng Wang; wangxis@chinacoal.com

Received 10 May 2022; Revised 3 June 2022; Accepted 6 June 2022; Published 5 July 2022

Academic Editor: Lianhui Li

Copyright © 2022 Guanchao Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper applies the CenterNet target detection algorithm to the foreign object detection of coal conveying belts in coal mines. Given the fast running speed of coal conveying belts and the influence of background and light sources on the objects to be inspected, an improved algorithm of CenterNet is proposed. First, the depth separable volume is introduced. The product replaces the standard convolution, which improves the detection efficiency. At the same time, the normalization method is optimized to reduce the consumption of computer memory. Finally, the weighted feature fusion method is added so that the features of each layer are fully utilized, and the detection accuracy is improved. The experimental results show that the improved algorithm has improved speed and accuracy compared with the original CenterNet algorithm. The foreign object detection algorithm proposed in this paper mainly detects coal gangue and can also detect iron tools such as bolts, drill bits, and channel steel. In the experimental environment, the average detection rate is about 20fps, which can meet the needs of real-time detection.

1. Introduction

As the main artery of coal mining and transportation, the working state of coal transportation belt directly affects the mining and transportation volume of coal. Foreign matters on the belt, such as large gangue and anchor bolt, are easy to cause problems such as scratch and tear of the belt and coal stacking and blocking at the coal chute during the high-speed operation of the belt. Therefore, the detection of large blocks, anchor bolts, and other foreign matters on the transportation belt can effectively ensure the safe production of the coal mine [1, 2].

Both target detection and image classification technology can realize the classification and recognition of belt foreign objects. However, target detection needs to locate the foreign object in the image before recognition, which increases the amount of calculation of the network to a certain extent, while image classification technology can directly identify the foreign object without locating the foreign object, and more computing resources can be used in fast foreign object recognition.

The complex environment of the mine makes the application of the existing image classification methods in the classification of foreign objects in coal conveying belt very challenging. Many scholars introduce machine vision technology into the image classification of foreign objects in the mine. Wang et al. [3] identified large foreign bodies of belt conveyor based on interframe difference method, threshold classification, and select-shape operator. He et al. [4] used the classification method of support vector machine to classify foreign objects in combination with the texture and gray features of foreign objects. Zhang [5] used multifeature fusion, combined with k-nearest neighbor algorithm and support vector machine for foreign object recognition. The above methods have achieved good results, but the image processing method combining feature extraction and classification algorithm has some problems, such as poor robustness and easy to be affected by illumination [6–9].

Convolution neural network uses convolution method for feature extraction, which has strong robustness and has

been widely used in many fields [10–13]. Some scholars also study the image classification network of mine foreign bodies. Pu et al. [14] established a foreign object recognition model based on vgg16 network and the idea of transfer learning, but the sample set was small, only 240 pieces. Su et al. [15] designed an improved LeNet-5 network and trained 20000 foreign object pictures in nonproduction environment, with a recognition rate of 95.88%. Ma [16], based on MobileNet network, optimized the network structure, improved the loss function, and further improved the recognition rate according to the characteristics of foreign object image. At this stage, the shortcomings of image classification of mine foreign bodies are as follows: (1). The sample collection is ideal without considering the actual working environment. (2). The network model has high complexity, large amount of parameters, low accuracy, and poor real-time performance.

Based on convolution neural network, in order to improve the adaptability of the algorithm for the detection of foreign matters in coal conveying belt under coal mine, this paper puts forward the corresponding improved method. In view of the low image quality caused by dark and uneven illumination in the underground, firstly, the dataset is preprocessed. Aiming at the problem of complex background of foreign object and large interference by coal block, the training data is correctly marked, the backbone network with good effect is adopted, and the weighted feature fusion of feature layers with different scales is introduced, which speeds up the convergence speed and reduces the amount of parameter calculation. Under the condition of ensuring the detection speed, the detection accuracy has been greatly improved.

2. Methods

2.1. Target Detection Algorithm. The CenterNet target detection method was proposed at the 2019 CVPR (Computer Vision and Pattern Recognition) conference [17–21]. The core idea of the algorithm is to regard the object to be detected as a point, that is, the center point of the target frame. Then, the center is found through the heat map and other attributes of the target object, such as the size information and pose information of the thing. The network structure of the CenterNet algorithm adopts ResNet-18, DLA-34, and Hourglass-104, three backbone networks for design and experiment. The three networks are complete encoding-decoding networks. The resolution of the final output feature map is subsampled 4 times compared to the original image, so that it can adapt to the detection of objects of various scales without multiscale design. The algorithm is based on anchor-free, so there is no need to set anchor boxes in advance, thereby avoiding the selection of related hyperparameters and eliminating the postprocessing process of NMS, which significantly reduces the computational load and training time of the network. The CenterNet object detection algorithm consists of three independent head structures: the center point prediction, the center point offset, and the target box's size, as shown in Figure 1.

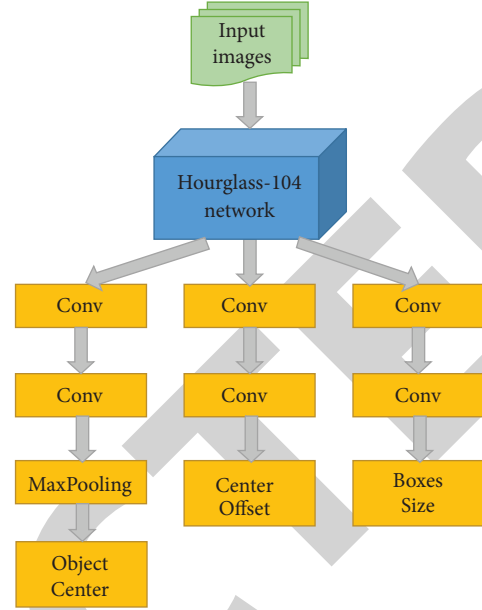


FIGURE 1: Algorithm structure.

Assuming that $I \in R^{W \times H \times 3}$ is the input image, where W and H represent the width and height of the image, respectively. After it is sent to the backbone network, a heat map containing the key will be generated.

$$Y \in [0, 1]^{(W/R) \times (H/R) \times C}, \quad (1)$$

where R is the subsampled factor and C is the number of classes in the detection target.

If class c is detected at the position (x_i, y_i) of the heat map, $\hat{Y}_{x_i, y_i, c} = 1$; on the contrary, if there is no target object at the point, $\hat{Y}_{x_i, y_i, c} = 0$.

In the training phase, for an object, assuming that the coordinates of its real frame in the input image are $(x_1, y_1, x_2, \text{ and } y_2)$, the center point of the object can be expressed as $p = (x_1 + x_2)/2, (y_1 + y_2)/2$. After subsampled, the coordinates of the point are mapped to the heat map as $\tilde{p} = \lfloor p/R \rfloor$ and mapped it to the heat map $Y \in [0, 1]^{(W/R) \times (H/R) \times C}$ through the Gaussian kernel transformation of the following formula:

$$y_{x_i y_i} = \exp\left(-\frac{(x - \tilde{p}_x)^2 + (y - \tilde{p}_y)^2}{2\sigma_p^2}\right). \quad (2)$$

where σ_p is the standard deviation of the object size adaptation if the Gaussian kernels of two objects of the same category overlap, taking the one with the most value.

In the prediction stage, the most considerable eight-neighborhood value is first screened on the heat map, which is equivalent to performing an entire pooling operation with a kernel size of 3 on the heat map, and a total of 100 such values are selected. Assuming that P_c is the detected point, the coordinate of the i -th key point is (x_i, y_i) and using this point to regress, the calibration frame is

$$\left(x_i + \delta x_i - \frac{w_i}{2}, y_i + \delta y_i - \frac{h_i}{2}, x_i + \delta x_i + \frac{w_i}{2}, y_i + \delta y_i + \frac{h_i}{2}\right)s. \quad (3)$$

where $(\hat{x}_i, \hat{y}_i) = \hat{O}_{\hat{x}_i, \hat{y}_i}$ is the offset of the point in the heat map relative to the original image during the subsampled process and $(\hat{w}_i, \hat{h}_i) = \hat{S}_{\hat{x}_i, \hat{y}_i}$ is the length and width of the target corresponding to the current point measured.

$\hat{Y}_{x,y,c} = 0$ is used to represent the confidence of the point, that is, the probability of the existence of an object at the current center point. In this paper, the screening threshold is set to 0.3. For the 100 values selected above, it will be reserved as the final result if the predicted probability is greater than the threshold.

The Loss function of the CenterNet algorithm consists of three parts: the category loss function L_{cls} , the regression loss function L_{size} , and the bias loss function L_{off} , which are linear combinations of the three.

$$L_{det} = L_{cls} + \lambda_{reg} L_{reg} + \lambda_{off} L_{off}. \quad (4)$$

The category loss function is the Focal Loss function.

$$L_{cls} = -\frac{1}{N} \begin{cases} (1 - \hat{Y}_{xyc}) \log(\hat{Y}_{xyc}) Y_{xyc} = 1, \\ (1 - Y_{xyc})^\beta (\hat{Y}_{xyc})^\alpha \log(1 - \hat{Y}_{xyc}) \text{ other.} \end{cases} \quad (5)$$

The target center bias loss function is

$$L_{off} = \frac{1}{N} \sum_p \left| \hat{O}_p - \left(\frac{p}{R} - \tilde{p} \right) \right|. \quad (6)$$

The regression loss function is

$$L_{reg} = \frac{1}{N} \sum_{k=1}^N |s_{pk} - s_k|. \quad (7)$$

The relevant hyperparameters are chosen as $\lambda_{reg} = 0.1$, $\lambda_{off} = 0.1$, $\alpha = 2$, and $\beta = 4$.

2.2. Hourglass Network Improvements. In this paper, the Hourglass-104 network [22] is selected as the backbone network of the CenterNet algorithm to extract features and generate heat maps. The network consists of two Hourglass modules stacked, as shown in Figure 2.

For the first-order module shown in Figure 3, the network adopts a fully convolutional neural network structure in the form of encoding-decoding. The repeated use of bottom-up and top-down supervision mechanisms combined with intermediate results has achieved good results in keypoint detection.

The most basic network unit in the Hourglass-104 network is the residual module, as shown in Figure 4. Almost all the network computation is consumed in the convolution operation of the residual module. For the detection task in this paper, since the datasets are directly derived from video surveillance, the image resolution is vast, the training time is also very long, and the consumption of computer hardware facilities is also relatively high.

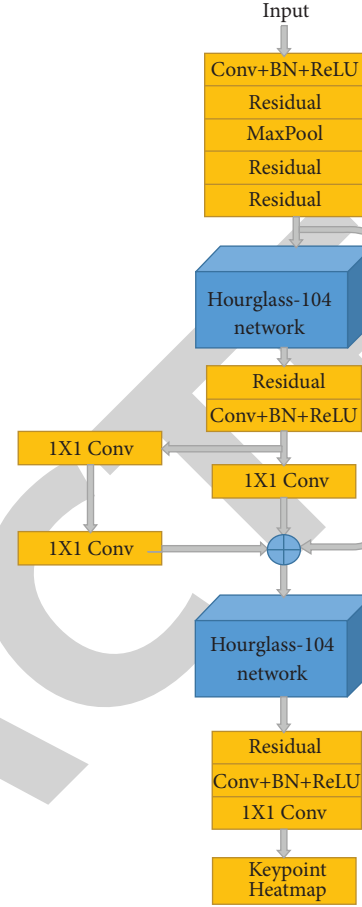


FIGURE 2: The structure of Hourglass-104 network.

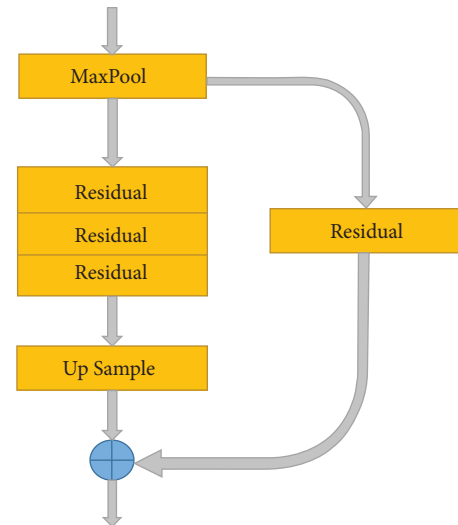


FIGURE 3: The first-order module.

To solve this problem, depthwise separable convolution (DSC) module is introduced. The DSC module is a lightweight network. In the processing process, the depthwise convolution of spatial relationship mapping is performed first, and then the point-by-point convolution of channel relationship mapping is performed. In the process of network

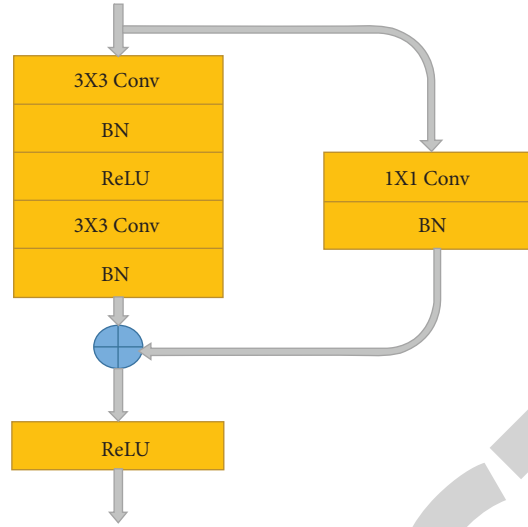


FIGURE 4: The residual module in the Hourglass-104 network.

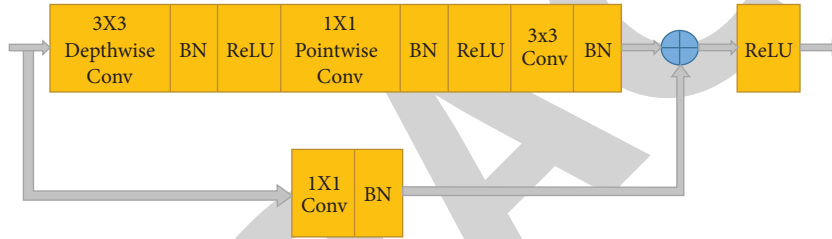


FIGURE 5: The residual module after introducing DSC.

training and detection, the calculation of parameters can be significantly reduced, which dramatically improves the network's work efficiency. The structure diagram of the residual module after introducing DSC is shown in Figure 5.

Assume that the input is (B, C, W, H) data, where $B, C, W,$ and H are the size of the feature image batch size, the number of channels, width, and height, respectively. The convolution operation of 3×3 is performed, and the calculation of standard convolution is as follows:

$$stc = B \times C \times W \times H \times 3 \times 3. \quad (8)$$

Depthwise separable convolution for calculation is used first, and then the depthwise convolution operation is performed, the calculation of the amount is as follows:

$$dw = C \times H \times W \times 3 \times 3. \quad (9)$$

Then, the pointwise convolution operation is performed, and the calculation of the amount is as follows:

$$pw = B \times C \times W \times H. \quad (10)$$

The total computational cost of depthwise separable convolution is the sum of the two.

$$dsc = dw + pw = (B + 3 \times 3) \times C \times W \times H. \quad (11)$$

The computational cost ratio of the standard convolution operation and the depthwise separable convolution operation is as follows:

$$\frac{dsc}{stc} = \frac{(B + 3 \times 3) \times C \times W \times H}{B \times C \times W \times H \times 3 \times 3} = \frac{1}{B} + \frac{1}{3 \times 3}. \quad (12)$$

When a lot of data are processed, the difference between the two calculation methods is more prominent.

2.3. Group Normalization Method. The calculation result of the BN layer depends on the data of the current batch. When the batch size is small, the mean and variance of the batch data are less representative, so the final result is also greatly affected. As shown in Figure 4, as the batch size becomes smaller and smaller, the reliability of the statistical information calculated by the BN layer becomes worse and worse, which will quickly lead to an increase in the final error rate. However, when the batch size is more significant, there is no noticeable difference. In target detection, segmentation, and video-related algorithms, the batch size is generally set relatively small due to large input images, diverse dimensions, and computer performance reasons. To weaken the influence of small-batch training on the detection results and make the CenterNet algorithm more flexibly applied to different hardware configuration environments, this paper improves the BN (batching normalization) in the backbone network to GN (grouping normalization).

Unlike BN, which normalizes channel by channel, GN groups channels and normalizes them in groups. First, the C channels are divided into G groups, and $S_k = \{c | [(c/G) = k], c = 0, 1, \dots, C\}$ is the channel number set composed of the

channel numbers that fall into the k -th channel group, where $k \in \{0, 1, \dots, G-1\}$, let the i -th feature of the j -th sample of the current mini-batch sample set in the c -th channel be $x_i^{(j,c)}$, the result after normalization is $y_i^{(j,c)}$, and the corresponding calculation formula is as follows:

$$x_i^{(j,c)} = \frac{1}{\sigma_i^{(k)}} (x_i^{(j,c)} - \mu_i^{(k)}) \text{ and} \quad (13)$$

$$y_i^{(j,c)} = a\hat{x}_i^{(j,c)} + b, \quad (14)$$

where sum is the mean and standard deviation of the i -th feature of each feature map of the k -th channel, respectively,

$$\mu_i^{(k)} = \frac{1}{m \cdot |S_k|} \sum_{j=1}^m \sum_{c \in S_k} x_i^{(j,c)} \text{ and} \quad (15)$$

$$\sigma_i^{(k)} = \sqrt{\frac{1}{m \cdot |S_k|} \sum_{j=1}^m \sum_{c \in S_k} (x_i^{(j,c)} - \mu_i^{(k)})^2 + \varepsilon}. \quad (16)$$

The group-based normalization groups the feature map channels, making each channel's mean and standard deviation more stable, effectively weakening the influence of the number of samples in the small-batch sample set on the feature normalization, and at the same time, the learning efficiency of the model based on the gradient descent method under low memory capacity is improved.

2.4. Weighted Feature Map Fusion. In the CenterNet network structure, the input image is input into the Hourglass-104 network after being subsampled 4 times. After 8 times, 16 times, 32 times, 64 times, and 128 times of subsampled in turn, and then 2 times of upsampling in turn, only one feature map is output. The output feature map size is subsampled 4 times from the input image. It can be seen that only using the most extensive feature map for target detection will inevitably lose some features of the image.

To make full use of the feature map generated after the convolution operation, improve the detection ability of small-sized targets, and reduce the algorithm's complexity. In this paper, the input image is subsampled 2 times, and then sent to Hourglass, and then subsampled 2 times, 4 times, 8 times, and 16 times in a turn, and output four feature maps, namely, P_1 , P_2 , P_3 , and P_4 , each feature's layer resolution is $(1/2^{i+1})$ times the input image, and multiple output feature maps are fused. Since each feature map contributes differently to the final fusion output, feature fusion cannot be performed directly. This paper uses a weighted method to fuse the features of each feature map; that is, when the feature map is fused, a learnable weight is assigned to each input feature map. During training, the network can learn to change the fusion weight of each feature map to change the importance of each feature map to the final detection result. To reduce the number of parameters and computation, the improved model directly uses short-circuit connections in the bypass convolution and instantly removes the convolution operation. The schematic diagram of feature fusion is

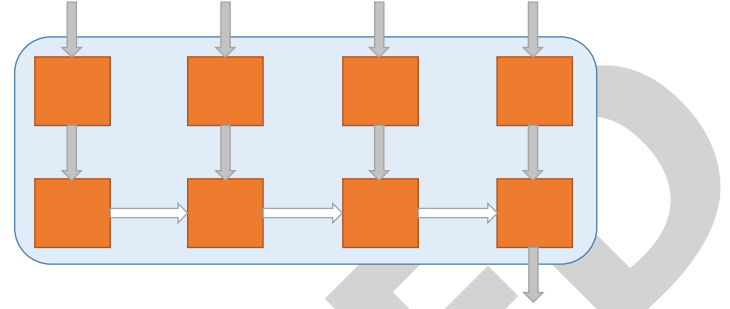


FIGURE 6: Feature fusion.

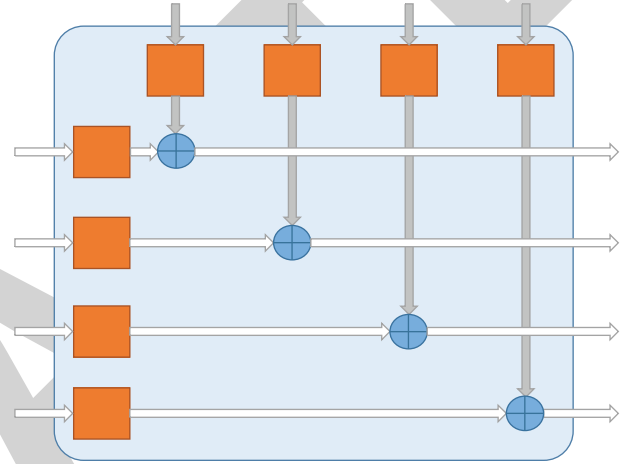


FIGURE 7: Weight fusion.

shown in Figure 6. The feature maps from high-level to low-level are obtained after upsampling, and then added, and the final output has only one fused feature map P_{out} .

The calculation formula of feature fusion is

$$O = \sum_i \frac{\omega_i}{\varepsilon + \sum_j \omega_j} \cdot I_i, \quad (17)$$

where I_i is the input feature map, the fused feature map, and ω_i , $\omega_i \in [0, 1]$ is the corresponding weight. A minimum value $\varepsilon = 1e-4$ is set to avoid the situation where the denominator in the above formula is 0.

The fast normalized fusion method is also used for weight fusion of the feature maps output by the two Hourglasses, and the calculation formula is

$$P_{out} = \frac{\sum \omega_{ij} \cdot P_{ij}}{\sum \omega_{ij} + \varepsilon}. \quad (18)$$

The schematic diagram of weight fusion is shown in Figure 7.

The final improved CenterNet (Improve-CN) algorithm network structure is shown in Figure 8.

3. Experiments and Analysis

3.1. Data Preprocessing and Labeling. The dataset images in this paper include two types of gangue and iron. The gangue is mainly coal road gangue and washing gangue, and the iron

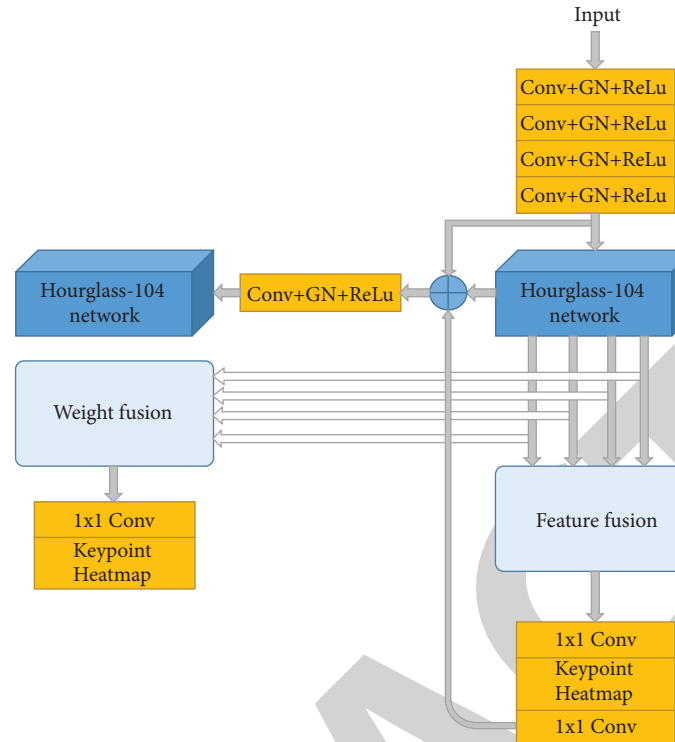


FIGURE 8: Improved CenterNet (Improve-CN) algorithm network structure.

includes bolts, drills, steel bars, and channel steel. The data is obtained through industrial infrared surveillance cameras. The foreign bodies obtained with this mounting type have the largest geometry. Detecting the foreign object early in entering the belt, effectively reducing the missed detection rate. There is no additional light source in the experiment, only the underground lighting source, and the possible miner's head-mounted lamp.

Due to low brightness in the coal mine and the uneven illumination caused by the irritating light source, the image quality obtained is poor. It is necessary to preprocess the directly acquired coal mine images to solve this problem. There is a lot of underground impulse noise in coal mines, so median filtering is performed first to eliminate the influence of noise and slight jitter. Although the effect of noise is eliminated, the contrast of the denoised image is still not high, and it is not easy to distinguish the foreground and background accurately. Therefore, the image continues to be enhanced. The method used is adaptive histogram equalization (AHE). AHE achieves the purpose of adjusting the image contrast by calculating the local histogram of the image and redistributing the brightness. The AHE algorithm improves the local contrast of the original image, makes the dark image brighter, suppresses the overbright area, and obtains more detailed information, which enhances the quality of the image.

Figure 9 compares the original image and the image after median filtering and adaptive histogram equalization.

The dataset in this paper is mainly composed of two classes, namely, "gangue" and "iron" (including channel steel, anchor rod, drill bit, and I-beam), which are marked

with label software. After an image is observed, a file with the suffix "json" will be generated in the directory, which contains the name and ID of the marked image, the category label of the significant object, and the coordinates of the object frame. The coordinates are marked as the coordinates of the upper left corner and the lower right corner of the box. A total of 5,605 images of two types were collected in the dataset of this paper. Since the foreign objects in the coal conveyor belt are mainly gangue, the majority of gangue images are 3,421, and there are a total of 2,184 iron images. The pictures are divided in the dataset, 70% of which are used as the train set, 20% are used as a validation set, and the remaining 10% are used as the test set. The distribution of the datasets is shown in Table 1.

3.2. Network Training. Since the collected dataset pictures are few and cannot meet the training requirements of deep learning, the dataset needs to be augmented. In this paper, the dataset is augmented by random cropping, flipping, rotation, translation, and scaling. There are five types of scaling factors in augmentation (0.5, 0.8, 1, 1.2, and 1.5). During training, Batch size is set to 8, an epoch is 140, and the initial learning rate is 5×10^{-5} . At 90 rounds and 120 rounds, respectively, it is attenuated 10 times, and SGD is used for network optimization. The weight decay rate and momentum factor are designed to be 10^{-4} and 0.9, respectively, and the input image is uniformly scaled to 512×512 sizes. The experiments are all completed on Ubuntu18.04, Nvidia GTX1060 graphics card, E5-2650 v2 CPU, CUDA10.1, cudnn7.6.5, Pytorch1.2.0, and the

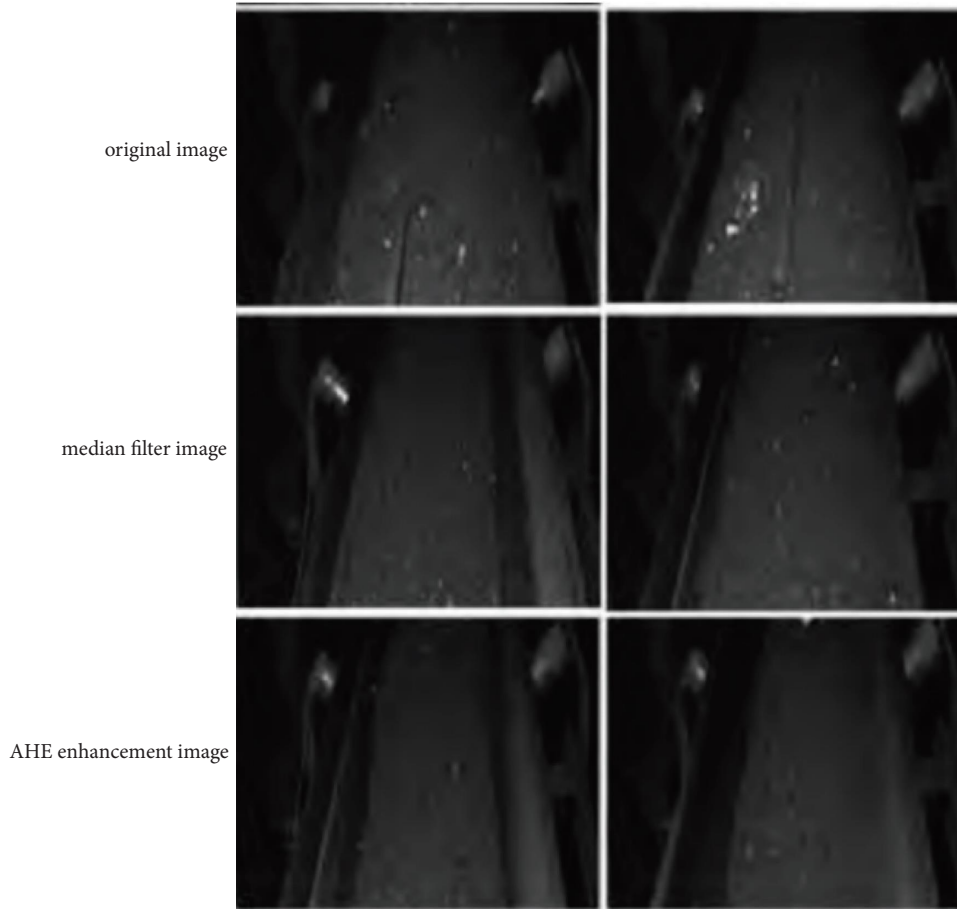


FIGURE 9: Comparison of the three images.

TABLE 1: Foreign object dataset in the coal belt.

Datasets	Gangue/piece	Iron/piece
Train set	2388	1545
Validation set	676	423
Test set	355	223

hyperparameters selection of each algorithm is completely consistent.

3.3. *Experimental Results and Analysis.* After about 100 rounds of training, the network converges. The loss function comparison between the improved algorithm and the original standard CenterNet algorithm is shown in Figures 10–13. The four figures are in order of total loss function curve, center point prediction loss function curve, target box regression loss function curve, and center point bias loss function curve. The abscissa is the number of training rounds, and the ordinate is loss change value. The blue line in the comparison image is the loss curve of the standard CenterNet algorithm and the orange line is the loss curve of the improved algorithm. It can be seen that the loss of the enhanced algorithm is reduced by about 0.3 compared with the original algorithm, and the original algorithm converges in about 120 rounds, that is, the improved

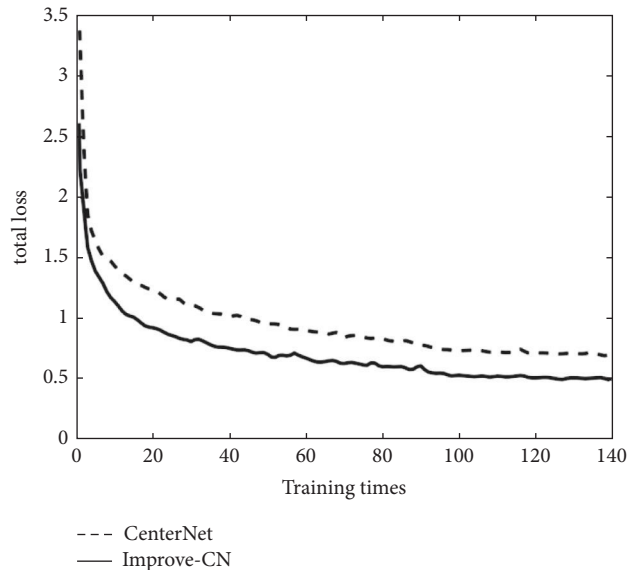


FIGURE 10: Total loss function curve.

algorithm speeds up the convergence speed. It can be seen from the loss function comparison curve that the improved algorithm has a smaller loss value than the original algorithm.

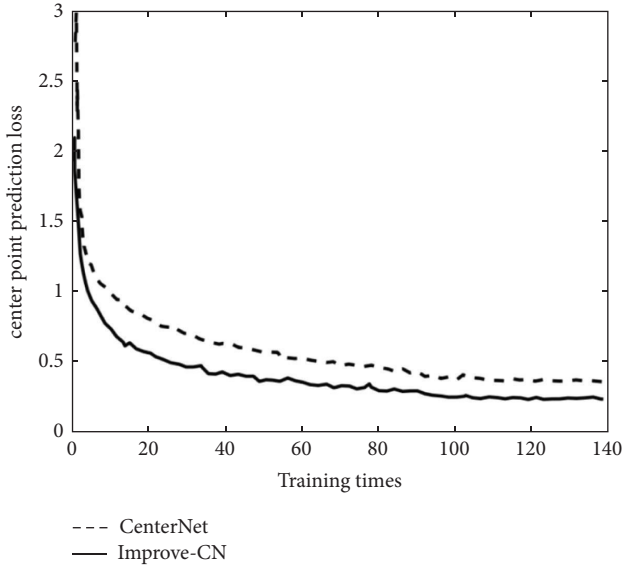


FIGURE 11: Center point prediction loss function curve.

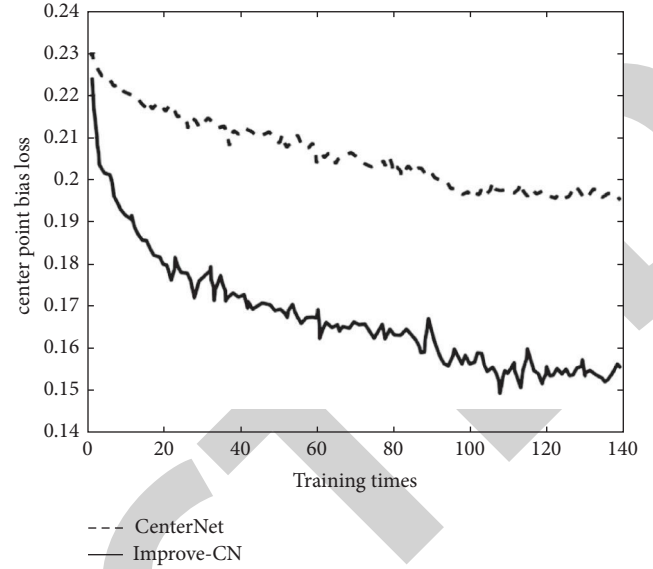


FIGURE 13: Center point bias loss function curve.

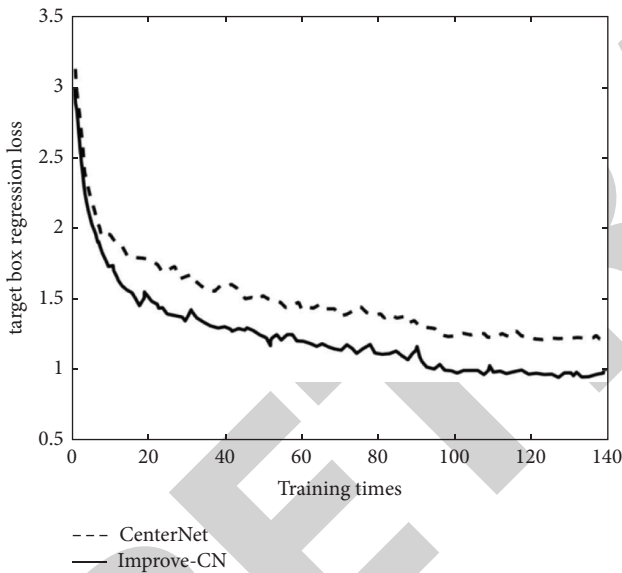


FIGURE 12: Target box regression loss function curve.

The quantitative comparison results of the two algorithms on the test set at 140 rounds are shown in Table 2. The main test indicators are average precision (AP) and average recall (AR), AP reflects the algorithm's accuracy and AR reflects the false detection level of the algorithm. The larger these two indicators are, the better the algorithm's performance is. AP_{50} and AP_{75} in Table 2 are the average accuracy rates when they are $IOU \geq 0.5$ and $IOU \geq 0.75$, respectively. AP_S , AP_M , and AP_L are the average accuracy indicators for small targets, medium targets, and large targets, respectively, and the AR indicators are marked in the same way.

It can be seen from Table 2 that the improved algorithm has a significant improvement in accuracy compared to the original algorithm, and the improved algorithm has

improved target accuracy and recalls for various scales, which is related to feature weighted fusion. By fusing the features at multiple scales, on the one hand, the rich semantic information of the high level can be fully utilized. On the other hand, the spatial location information of the low level can be obtained, which significantly improves the accuracy of the target detection.

Four test images are randomly selected from the test set and tested with the algorithm designed in this paper. The detection results are shown in Figure 14. It can be seen that the proposed algorithm can accurately detect targets of various scales, and the improved algorithm can also accurately detect slender "iron objects," especially those close to the edge of the conveyor belt and along with the movement direction distribution target.

Table 3 compares the detection results of different target detection algorithms in the test set after training under the same dataset. It can be seen that nearly 130 ms reduces the average detection time of the improved algorithm compared with the original algorithm, and the overall accuracy rate is improved by 6.9% or so. Compared with the two-stage faster R-CNN algorithm [23–26], the algorithm before and after the improvement significantly shortens the detection time, but the accuracy rate decreases because the improved algorithm only needs to predict the target's center point compared to the anchor frame. The computational complexity is significantly reduced, so the speed is improved. Still, many anchor boxes provide a greater possibility of correctly detecting the target, so the accuracy rate is higher, and the application can be used to monitor the same transmission belt through multiple cameras to offset this shortcoming. Compared with the one-stage detection algorithm YOLOv3, the detection speed and accuracy of the improved algorithm have been improved. The results show that the enhanced algorithm achieves a good balance between detection speed and accuracy.

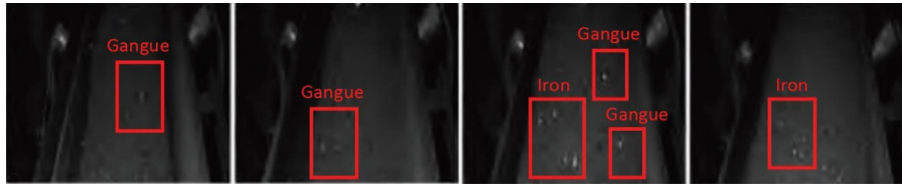


FIGURE 14: The foreign objects detection result with the proposed algorithm.

TABLE 2: Comparison between CenterNet and Improve-CN.

Evaluation indicators		CenterNet	Improve-CN
Average precision (AP)	AP ₅₀	0.4822	0.5522
	AP ₇₅	0.2669	0.3080
	AP _s	0.1056	0.1834
	AP _M	0.2778	0.3422
	AP _L	0.4531	0.4734
Average recall (AR)	AR ₅₀	0.3888	0.5025
	AR ₇₅	0.2522	0.3138
	AR _s	0.1835	0.3221
	AR _M	0.4341	0.5551
	AR _L	0.5012	0.5982

TABLE 3: The comparison among other objects detection algorithm.

Algorithm	Average detection Time/s	MAP
Faster R-CNN	4.1671	0.5877
YOLOv3	0.0943	0.4344
CenterNet	0.0626	0.4828
Improve-CN	0.0508	0.5521

4. Conclusion

As fixed-point monitoring by network cameras is gradually replacing manual inspection and becoming an essential underground monitoring method for coal mining enterprises, the proposed foreign object detection method can be directly deployed on this basis. As long as the host computer that meets the hardware requirements is configured, the noncoal foreign matter can be detected at the early stage when it enters the coal conveying belt, and an alarm response can be made. In the later stage, the robot can automatically sort the detected foreign objects to realize the integration of foreign object detection and grasping, which can improve work efficiency while protecting personal safety and further reducing potential safety hazards.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] J. Kan, P. Wang, and P. Wang, "Influencing factors of disturbance effects of blasting and driving of deep mine roadway groups," *Shock and Vibration*, vol. 2021, Article ID 8873826, 13 pages, 2021.
- [2] K. Zheng, C. Du, J. Li, and B. Qiu, "Coal and gangue underground pneumatic separation effect evaluation influenced by different airflow directions," *Advances in Materials Science and Engineering*, vol. 2016, Article ID 6465983, 13 pages, 2016.
- [3] Y. Wang, X. Guo, X. Liu, and Z. Zhao, "Design of visual detection system for large foreign body in belt conveyor," *Mechanical Science and Technology for Aerospace Engineering*, vol. 12, pp. 1–7, 2020.
- [4] M. He, P. Wang, and H. Jiang, "Recognition of coal and stone based on SVM and texture," *Computer Engineering and Design*, vol. 33, no. 03, pp. 1117–1121, 2012.
- [5] Y. Zhang, *Research on Gangue Identification Based on Video Processing*, pp. 1–99, China University of Mining and Technology, Xuzhou, 2018.
- [6] F. Cen, X. Zhao, W. Li, and G. Wang, "Deep feature augmentation for occluded image classification," *Pattern Recognition*, vol. 111, Article ID 107737, 2021.
- [7] G. Zhang, J. Yang, Y. Zheng, Z. Luo, and J. Zhang, "Optimal discriminative feature and dictionary learning for image set classification," *Information Sciences*, vol. 547, pp. 498–513, 2021.
- [8] J. Fang, H. Liu, L. Zhang, J. Liu, and H. Liu, "Region-edge-based active contours driven by hybrid and local fuzzy region-based energy for image segmentation," *Information Sciences*, vol. 546, pp. 397–419, 2021.
- [9] W. Wang, H. Zhang, Z. Zhang, L. Liu, and L. Shao, "Sparse graph based self-supervised hashing for scalable image retrieval," *Information Sciences*, vol. 547, pp. 622–640, 2021.
- [10] D. Qiu, L. Zheng, J. Zhu, and D. Huang, "Multiple improved residual networks for medical image super-resolution," *Future Generation Computer Systems*, vol. 116, pp. 200–208, 2021.
- [11] Z. Zhang, M. Huang, L. Shang, B. Xiao, and T. S. Durrani, "Fuzzy multilayer clustering and fuzzy label regularization for unsupervised person reidentification," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 07, pp. 1356–1368, 2020.
- [12] D. Fernandes, A. Silva, R. Névoa et al., "Point-cloud based 3D object detection and classification methods for self-driving applications: a survey and taxonomy," *Information Fusion*, vol. 68, pp. 161–191, 2021.
- [13] D. Cheng, L. Chen, Y. Cai, and L. Gao, "Image super-resolution reconstruction based on multi-dictionary and edge fusion," *Journal of China Coal Society*, vol. 43, no. 07, pp. 2084–2090, 2018.
- [14] Y. Pu, D. B. Apel, A. Szmigiel, and J. Chen, "Image recognition of coal and coal gangue using a convolutional neural network and transfer learning," *Energies*, vol. 12, no. 9, pp. 1735–1746, 2019.
- [15] L. Su, X. Cao, H. Ma, and Y. Li, "Research on Coal Gangue Identification by Using Convolutional Neural network," in *Proceedings of the 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation*

- Control Conference (IMCEC)*, pp. 810–814, Xi'an, China, May 2018.
- [16] M. A. Yan, *Research on Deep Learning Algorithm and Key Technology of Coal Gangue Image Recognition*, pp. 1–125, China University of Mining and Technology, (Beijing), Beijing, China, 2019.
- [17] S. Hou and W. Guo, "Fault location method in nonsolid-earthed network based on spatial domain image fusion and convolution neural network," *Journal of Sensors*, vol. 2022, Article ID 8969479, 12 pages, 2022.
- [18] Q. Liang, "Application of Convolution Neural Network (CNN) Model Combined with Pyramid Algorithm in Aerobics Action Recognition," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 6170070, 10 pages, 2021.
- [19] T. Puri, M. Soni, G. Dhiman et al., "Detection of emotion of speech for RAVDESS audio using hybrid convolution neural network," *Journal of Healthcare Engineering*, vol. 2022, Article ID 8472947, 9 pages, 2022.
- [20] S. Sridhar, J. Amutharaj, P. Valsalan et al., "A torn ACL mapping in knee MRI images using deep convolution neural network with inception," *Journal of Healthcare Engineering*, vol. 2022, Article ID 7872500, 9 pages, 2022.
- [21] J. Zhang and S. Ke, "Improved YOLOX Fire Scenario Detection Method. Wireless Communications and Mobile Computing," *Explorations in Pattern Recognition and Computer Vision for Industry 4.0*, vol. 2022, Article ID 9666265, 8 pages, 2022.
- [22] M. Gao, W. Cai, and R. Liu, "AGTH-net: attention-based graph convolution-guided third-order Hourglass network for sports video classification," *Journal of Healthcare Engineering*, vol. 2021, Article ID 8617161, 10 pages, 2021.
- [23] L. Li, B. Lei, and C. Mao, "Digital twin in smart manufacturing," *Journal of Industrial Information Integration*, vol. 26, no. 9, Article ID 100289, 2022.
- [24] S. Meivel, N. Sindhwani, R. Anand et al., "Mask Detection and Social Distance Identification Using Internet of Things and Faster R-CNN Algorithm," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 2103975, 13 pages, 2022.
- [25] L. Li, T. Qu, Y. Liu et al., "Sustainability assessment of intelligent manufacturing supported by digital twin," *IEEE Access*, vol. 8, pp. 174988–175008, 2020.
- [26] J. Zhang, X. Song, J. Feng, and J. Fei, "X-ray image recognition based on improved mask R-CNN algorithm," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6544325, 14 pages, 2021.