

## Article

# Enhanced Detection of Foreign Objects on Molybdenum Conveyor Belt Based on Anchor-Free Image Recognition

Meng Li <sup>1,2,\*</sup>, Caiwu Lu <sup>1,3</sup>, Xuesong Yan <sup>3</sup>, Runfeng He <sup>1</sup> and Xuyang Zhao <sup>4</sup>

<sup>1</sup> School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, China; lucaiwu@126.com (C.L.); hrf101@xauat.edu.cn (R.H.)

<sup>2</sup> Architecture College of Xi'an, Xi'an University of Architecture and Technology, Xi'an 710055, China

<sup>3</sup> School of Resource Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China; acebased@163.com

<sup>4</sup> Luanchuan Longyu Molybdenum Industry Co., Ltd., Nannihu Molybdenum Mine, Luoyang 471500, China; nnhksgsdizhi@163.com

\* Correspondence: limeng@xauat.edu.cn

**Abstract:** During the molybdenite mining process, conveyor belts stretching for miles are used to transport ore between the blasting sites, crushing stations, and the concentrator plant. In order to ensure the safety and stability of the industrial production process, this paper introduces a foreign matter detection method based on deep learning for the belt conveyor. Aiming at the problems of insufficient feature extraction capabilities in existing machine vision-based foreign body detection methods and poor detection accuracy due to imbalanced positive and negative samples, an improved foreign body detection method for anchorless frame-type metal mine belt conveyors is proposed. This method introduces atrous convolution in the pooling layer to increase the receptive field of feature extraction and improve the ability of extracting feature details of foreign objects. By optimizing the ratio of positive and negative samples in the training process, the overall loss function value of the algorithm is reduced to ensure the accuracy of foreign body recognition. Finally, the improved model is trained after enhancing and labeling the sample dataset. The experimental results show that the average mean accuracy of foreign body detection (MAP) is 90.9%, better than existing methods. It can be used as an effective new method for detecting foreign objects on molybdenum mine belt conveyors.



**Citation:** Li, M.; Lu, C.; Yan, X.; He, R.; Zhao, X. Enhanced Detection of Foreign Objects on Molybdenum Conveyor Belt Based on Anchor-Free Image Recognition. *Appl. Sci.* **2024**, *14*, 7061. <https://doi.org/10.3390/app14167061>

Academic Editor: Noel Aquilina

Received: 14 June 2024

Revised: 3 August 2024

Accepted: 5 August 2024

Published: 12 August 2024



**Copyright:** © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** molybdenum mining; mining belt conveyor; foreign body; target detection; anchor-free

## 1. Introduction

Molybdenum is an important additive in the steel industry, primarily used to enhance the strength, toughness, corrosion resistance, and high-temperature performance of steel. It is widely used to produce various alloys. China is one of the major production areas for molybdenum. Most molybdenum is found in the form of molybdenite within granite mountains. In open-pit mining, the extraction primarily involves blasting, shoveling, and belt transportation, with a particular reliance on long-distance conveyor belts. Foreign objects (drill roots, shovel teeth, and I-beams) on metal mining belt conveyors are one of the main causes of belt damage, tearing and grinder wear, seriously threatening the safe and stable operation of mine production. Therefore, metal mining enterprises urgently need to accurately detect foreign objects on the conveyor belt, remove foreign objects in advance or give a timely warning, and prohibit pre-incident accidents, so as to reduce the economic loss and the impact of safety accidents caused by foreign objects.

The current fault protection system of belt conveyors can only perform self-checks on conveyor operation faults, usually during the shutdown after the accident has occurred. The traditional conveyor belt foreign object detection technology mainly includes radar detection, a metal detector and manual detection. The use of the above foreign object detection methods has disadvantages such as high cost, few detection types, and difficulty

regarding equipment installation, operation and maintenance, and it is difficult to widely popularize in the actual production of metal mining enterprises. In recent years, with the blowout development of artificial intelligence, the application of machine vision in various fields has become more and more mature, making machine vision possible as a simple, low-cost and reliable foreign object detection method for belt conveyors. Researchers initially use computer graphics methods to process the collected digital images of conveyor belts, using feature extraction operators to identify simple, unobstructed foreign objects. These methods could be applied to simple environments, such as foreign object detection on product production line conveyors, with the core idea being to detect objects, exclude non-foreign objects, and raise an alarm for foreign objects. Reverse inspection methods could detect unknown-shaped foreign objects, but they were only suitable for standardized production lines. For non-standardized conveyor belts such as coal and metal ores, reverse inspection was not feasible, and detection had to rely on pre-set common foreign objects. However, common foreign objects come in various shapes and can be obstructed. In the early days, scholars used feature extraction combined with machine learning methods to expand the detection capabilities of algorithms. They first used operators to extract local features and then employed machine learning classifiers (such as SVM [1]) for further classification [2]. With the increase in computing power, image deep learning [3] technology has been widely developed. Convolutional neural networks can provide machines with excellent image understanding capabilities, leading to the development of various image recognition and object detection methods based on this technology.

For the problem of foreign object detection on conveyor belts, using object detection methods is a good choice. Object detection [4] is a supervised learning method where, after multiple rounds of annotated data learning, the algorithm can identify different shapes of objects, their sizes, and positions in the image matrix, and provide an estimate of confidence in their classification. Object detection algorithms have evolved into various approaches, mainly categorized into one-stage and two-stage modes. Two-stage object detection methods typically consist of two stages: first, a Region Proposal Network (RPN [5]) generates candidate regions, and then these regions are classified and precisely located. Faster R-CNN is a typical two-stage object detection algorithm that uses RPN to extract candidate regions and classifies and locates them through RoI Pooling and fully connected layers. On the other hand, one-stage object detection algorithms are more concise. They transform the object detection problem into a regression problem and directly output the object's bounding box coordinates and class probabilities through a convolutional neural network. Generally, two-stage algorithms are slower but more accurate, while one-stage algorithms are faster but slightly less accurate. However, with network design and improvements, the accuracy of one-stage algorithms can be effectively enhanced, making the overall model logic simpler with better development prospects.

Therefore, we choose the single-stage object detection algorithm as the foundation of our research. The developed machine vision foreign single-stage object detection modes can be divided into two types: anchor frame type and anchorless frame type. The famous YOLO series algorithm adopts the anchor box mode, which predefines a set of fixed bounding boxes (anchor boxes) with different sizes and aspect ratios to adapt to targets of different sizes and shapes. The algorithm detects targets by adjusting these predefined anchor boxes. This method performs well when dealing with a large number of standard targets but may not perform as well on unconventional or small targets. Anchor-free methods such as Center-Net do not predefine fixed bounding boxes but directly predict the edges or center points of the targets. This method is more flexible when handling irregular shapes or small targets. In terms of the proportion of foreign objects to the main body of the conveyor belt, the anchor-free method shows more potential in terms of detection accuracy.

For example, SSD [6–8], YOLO V3 [9–11], and Faster R-CNN [12] have anchor frames. The anchor frame method often leads to a decrease in model accuracy [13,14] due to the processing of the non-maximum suppression value (NMS). Center-Net [15–17] is an anchor-free method, and the anchor-free method is limited by the receptive field of view of the

pooling layer feature extraction network, and a large number of foreign object details will be lost. At the same time, in the anchor-free method, the detection target predicts the point corresponding to the center position through the heat map, which will lead to the extreme imbalance of the foreign object sample and the problem that the loss function is difficult to converge.

Based on the research mentioned above, we have focused on improving the accuracy of anchor-free object detection [18] on the molybdenum ore conveyor belt. Firstly, we enhanced the algorithm to improve the model's ability to capture multi-scale features in complex backgrounds. Based on the anchor-free object detection method Center-Net, we introduced the dilation rate parameter into atrous convolution [19] to reduce the loss of detailed foreign object information when passing through the pooling layer. By adjusting parameters with positive and negative samples, we reduced the overall loss function value and improved the accuracy of foreign object detection [20].

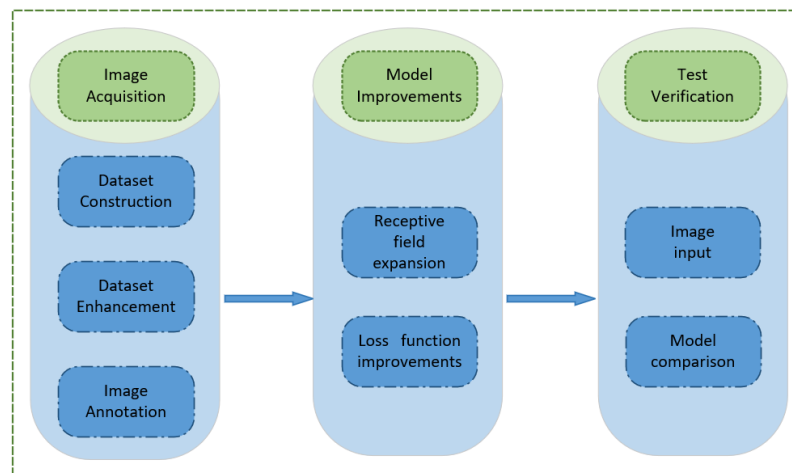
In order to address the existing gaps in research, we propose an enhanced deep learning network for object detection specifically designed for identifying foreign objects on molybdenum mine conveyor belts, building upon the Center-Net framework to achieve improved detection capabilities for small targets in complex backgrounds. First, in the Method section of this paper, we systematically outline the detection process for foreign objects on conveyor belts and discuss the current state of anchor-free object detection networks. We introduce a novel convolutional pattern and redesign the loss function, enabling the model to achieve greater accuracy in detecting small targets against complex backgrounds. Subsequently, we collected data by deploying network cameras at a mining site in Henan, China, followed by image data processing and annotation to construct a dedicated dataset for foreign object detection on molybdenum conveyor belts. After implementing the model improvements and preparing the dataset, we conducted a series of experiments based on the enhanced model. In the result section of this paper, we analyze the training process and compare the experimental outcomes. Ultimately, our findings demonstrate that the proposed improved anchor-free object detection method exhibits a significant accuracy advantage over other networks of comparable scale in the context of detecting foreign objects on molybdenum mine conveyor belts.

## 2. Methods of Research

### 2.1. Basic Process of Foreign Body Detection

Molybdenum ore typically appears as a blue-white hard rock, which distinguishes it from conventional foreign objects. Images for foreign object detection in ore are usually captured using network cameras positioned at critical nodes of the conveyor belt, allowing for the acquisition of clear image data that provides a solid foundation for subsequent image processing and object detection algorithms. For the task scenario addressed in this paper, we collected data from the L mine in Henan Province, China, with the camera fixed at a position located 45° to the left of the conveyor belt. The model of the conveyor belt is YKK4502-4. A total of 4532 foreign body images were collected, with an image resolution of 6000 × 4000 pixels (more detailed data information can be found in Section 2.3). We established a dataset in this scenario as the basis for our research.

The steps of the basic process of foreign body detection in this paper are divided into the following: ① Install the image acquisition device at the belt conveyor to collect the working image of the conveyor. ② Classify the collected foreign objects, eliminate foreign objects that have no effect, and construct a foreign object dataset for metal ore belt conveyors. ③ Perform data enhancement on the foreign body dataset. ④ Use the marking software to mark the foreign objects of the metal mine belt conveyor. ⑤ Improve the anchor-free Center-Net target detection method to increase the robustness of the recognition system. ⑥ Input the collected image data into the improved target detection method to detect foreign objects. The specific detection steps are shown in Figure 1.



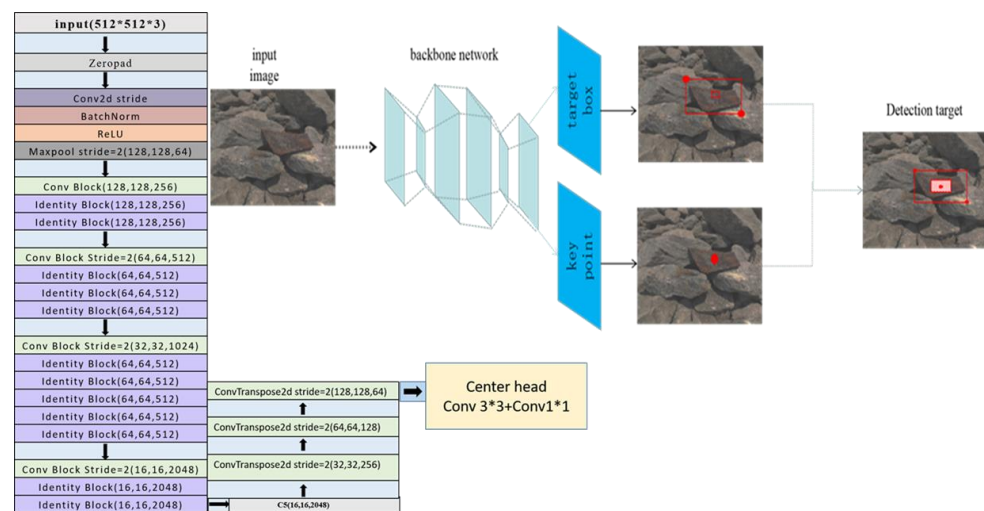
**Figure 1.** Basic procedures for foreign body detection.

### 2.2. Improvements of the Approach

The target detection method with anchor boxes may cause a large number of anchor boxes to be removed by NMS [21]. In order to avoid this phenomenon, the network detection efficiency decreases. Using the anchor-free target detection method, the image recognition is realized by predicting the center point of the target in the detection map and returning to the location of the target. The algorithm greatly improves the detection efficiency of the network. Coner-Net, Coner-Net Lite, Center-Net, FCOS (Fully convolutional one-stage object detection) are longer target detection networks using anchor-free boxes [22].

#### 1. Center-Net Object Detection Algorithm

The Center-Net model is divided into two steps. First, the model needs to predict the center point of the target detection object through the heat map, and then predict the recognition frame of the target object through the generated center point. Figure 2 shows the structure of Center-Net.



**Figure 2.** Center-Net structure.

The image is sent to the backbone network for feature extraction, and then input to the key point generation module and the target box generation module, respectively. In the key point module, the heat map is used as the basic information to predict the center point, the center point offset and the target size, respectively, and, finally, the key points of the predicted object are obtained. In the target box generation module, the target box is

generated by matching the corners belonging to the same pair according to the heat map. This method does not need to use NMS, which greatly improves the detection efficiency of the target detection model.

### 2. Hourglass-104 Network Architecture

In this paper, the Hourglass-104 network structure is used to extract the foreign matter characteristics of the metal mine belt conveyor. The Hourglass-104 network structure is composed of two modules similar to the Hourglass structure. A series of convolutional layers, where residual unit modules, feature fusion modules and upsampling layers form a Hourglass module [23,24]. A single Hourglass structural module is shown in Figure 3.

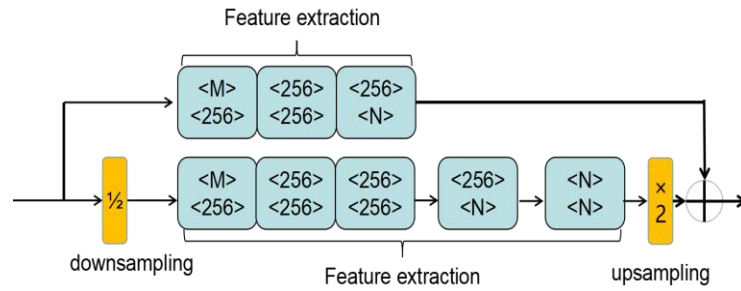


Figure 3. Hourglass structure module.

The specific working steps of Hourglass-104 are as follows: ① Input the original image to the preprocessing module of Hourglass-104 to reduce the resolution of the original image by  $\frac{3}{4}$ . ② Input the residual unit, and perform 5 downsampling processes through convolution with a step size of 2. The downsampling process increases the number of channels of the feature map. ③ Upsampling is performed five times to bring the image resolution back to the pre-processed size, while also reducing the number of feature channels.

### 3. Expansion of the receptive field

Since the characteristics of deep learning often require a lot of computation, traditional convolutional neural networks reduce the size of the input image through the pooling layer, thereby reducing the time required for computation. In order to avoid the loss of foreign body detail information caused by the pooling layer, a dilation rate parameter is introduced, which is added to the original convolution to make the original convolution into an atrous convolution. Taking a  $3 \times 3$  convolution as an example, Figure 4 below shows the difference between the ordinary convolution and the atrous convolution.

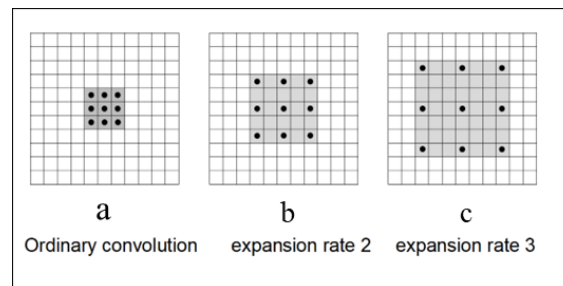
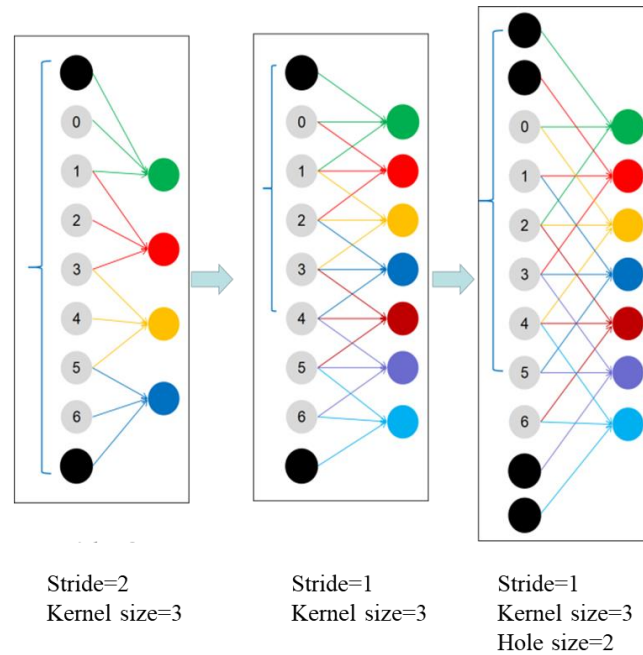


Figure 4. Atrous convolution contrast.

The three large boxes in Figure 5, named a, b, and c represent the input images, respectively; the black dots represent the  $3 \times 3$  convolution, and the shaded gray area in the figure is the recognition area of the convolution kernel. a is an ordinary  $3 \times 3$  convolution, and the receptive field of view is  $3 \times 3$ . Picture b shows atrous convolution with an expansion rate of 2, and the receptive field of view is increased to  $5 \times 5$ . Picture c

shows atrous convolution with an expansion rate of 3, and the perceived field of view is increased to  $8 \times 8$ . Atrous convolution increases the receptive field of view of convolution without increasing the convolution kernel and the image output feature size, and provides a better method for identifying image details. This method is used to make up for the detailed feature information of foreign objects lost by the pooling layer. Figure 5 is a one-dimensional comparison diagram of atrous convolution.



(a) original convolution (b) original convolution (c) atrous convolution

**Figure 5.** One-dimensional contrast.

*a* is the original convolution, and *b* and *c* are atrous convolutions. Suppose the size of the convolution kernel is *k*, the expansion rate is *d*, and the size of the convolution kernel after the dilation of the dilated convolution is *O*, then the size of the dilated convolution can be obtained by the following calculation formula:  $O = k + (k - 1)(d - 1)$  [25]. Atrous convolution [26] with dilation rate 2 is introduced, considering the size of the foreign object.

#### 4. Design of Loss Function

The loss function of the model is calculated by detecting the target recognition frame and the prediction result of the corresponding picture. According to the core steps of the Center-Net algorithm, the loss functions are as follows: ① loss of center point; ② loss of heat map (classification confidence loss); ③ loss of recognition frame width and height; ④ total loss function.

When the image is input to the feature extraction network, the downsampling will lead to the offset of the pixel points and then the accuracy error.

Center point loss  $L_C$ :

$$L_c = \frac{1}{N} \sum_P \hat{O}_{\hat{p}} - \left( \frac{P}{R} - \hat{P} \right) \tag{1}$$

where *N* is the total number of target center points predicted by the heat map, *P* is the real coordinates of the original image,  $\hat{P}$  is the image center coordinate predicted by the heat map, *R* is the multiple set for downsampling in the feature extraction network, and  $\hat{O}_{\hat{p}}$  is the predicted offset.

For the loss of classification confidence  $L_H$ , due to the working principle of the model, a detection target is predicted by only one feature point, which leads to an extreme imbalance of samples. This phenomenon is avoided by the loss function of the target classification confidence.

Classification confidence  $L_H$ :

$$L_H = \alpha_1 L_n + \alpha_2 L_p \quad (2)$$

$$L_n = -\left(1 - \hat{Y}_{xyc}\right)^{\beta_1} \log\left(\hat{Y}_{xyc} + \delta\right) \quad (3)$$

$$L_p = -\left(1 - \hat{Y}_{xyc}\right)^{\beta_2} \log\left(\hat{Y}_{xyc}\right) \quad (4)$$

where  $\hat{Y}_{xyc}$  is the predicted value of the detected image; in order to increase the loss of positive samples and reduce the loss of negative samples, there are three parameters set in the loss function,  $\alpha$ ,  $\beta$ , and  $\delta$ . The specific values of the three parameters were obtained through multiple experimental searches. Negative sample  $L_n$  and positive sample  $L_p$  are set, respectively, as  $\beta_1$ ;  $\delta$  and  $\beta_2$  adjust. The loss of the final positive and negative samples will be reflected on the parameter  $\alpha$  to obtain the classification confidence function  $L_H$ . where  $\hat{Y}_{xyc}$  is the predicted value of the detected image.

In the Center-Net target detection algorithm, the target size frame is generated based on the predicted center point. There will be a certain error in the width and height of the target size box.

The loss of the width and height of the recognition box  $L_S$ :

$$L_s = \frac{1}{N} \sum_{K=1}^N \left| \hat{S}_{PK} - S_K \right| \quad (5)$$

where  $\hat{S}_{PK}$  is the size of the predicted image and  $S_K$  is the size of the real image.

Total loss function  $L_E$ :

$$L_E = L_H + \varphi_C L_C + \varphi_S L_S \quad (6)$$

In the formula,  $\varphi_C$ ,  $\varphi_S$  are the ratio of center point loss and recognition frame width and height loss, respectively.

### 2.3. Preprocessing of Experimental Data

The experimental data in this paper are collected from the L mine in the Henan province of China, taking a fixed position, located at  $45^\circ$  to the left of the conveyor belt; the image acquisition angle is down  $30^\circ$ , and the distance from the ground is 180 cm. The collection of test images is carried out. The acquisition environment is shown in Figure 6.



**Figure 6.** Environment of data collection.

The collection data are collected on the 1# belt conveyor located in the crushing workshop, and the conveyor belt equipment model is SIMO YKK4502-4, made in Xi'an of China. A total of 4532 foreign body images were collected. The image resolution is  $6000 \times 4000$ . The collected images are divided into training image sets, verification image sets and test image sets, respectively. The specific image sets are shown in Figure 7.



Figure 7. Image database acquisition.

#### 1. Sample dataset enhancements

- Histogram equalization

Use histogram equalization for poor lighting conditions to enhance the darker areas of the image. By nonlinearly stretching the image to be processed, the image values are redistributed, so that the pixel information values within a certain grayscale range are roughly consistent [27]. The comparison chart after histogram equalization [28,29] is shown in Figure 8.

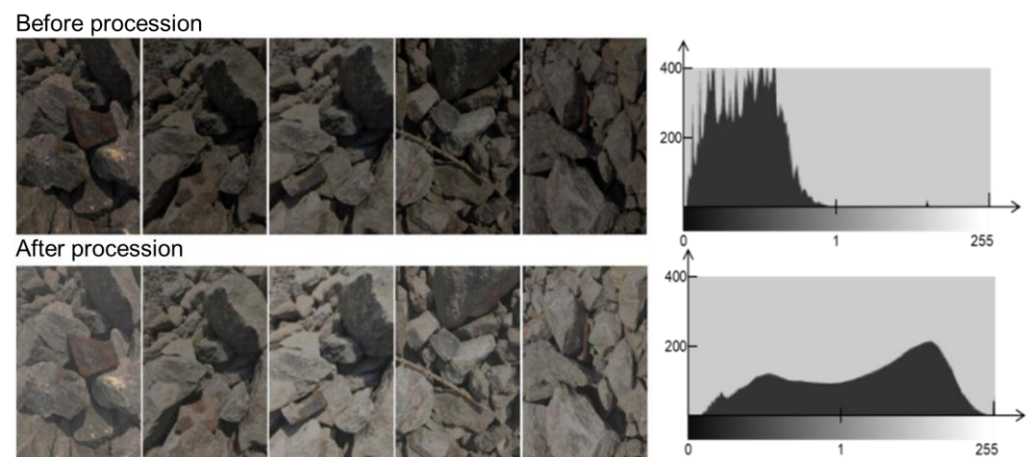
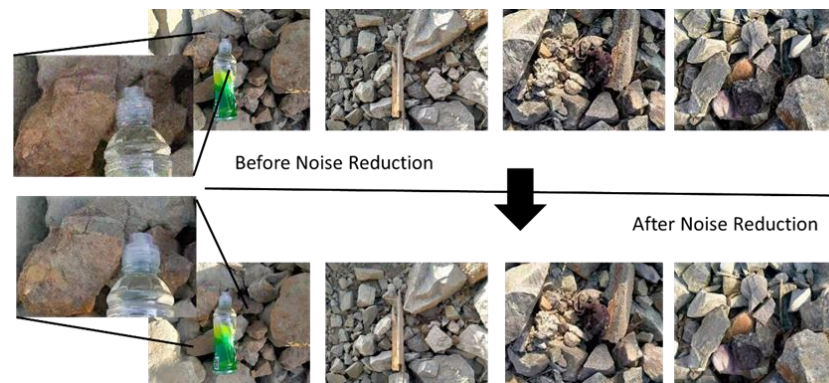


Figure 8. Histogram equalization.

- Noise reduction

For the noise existing in the image, median filtering [30] is used to make the pixel value of the image approach the real value, thereby achieving the effect of noise reduction. The comparison chart before and after noise reduction is shown in Figure 9.

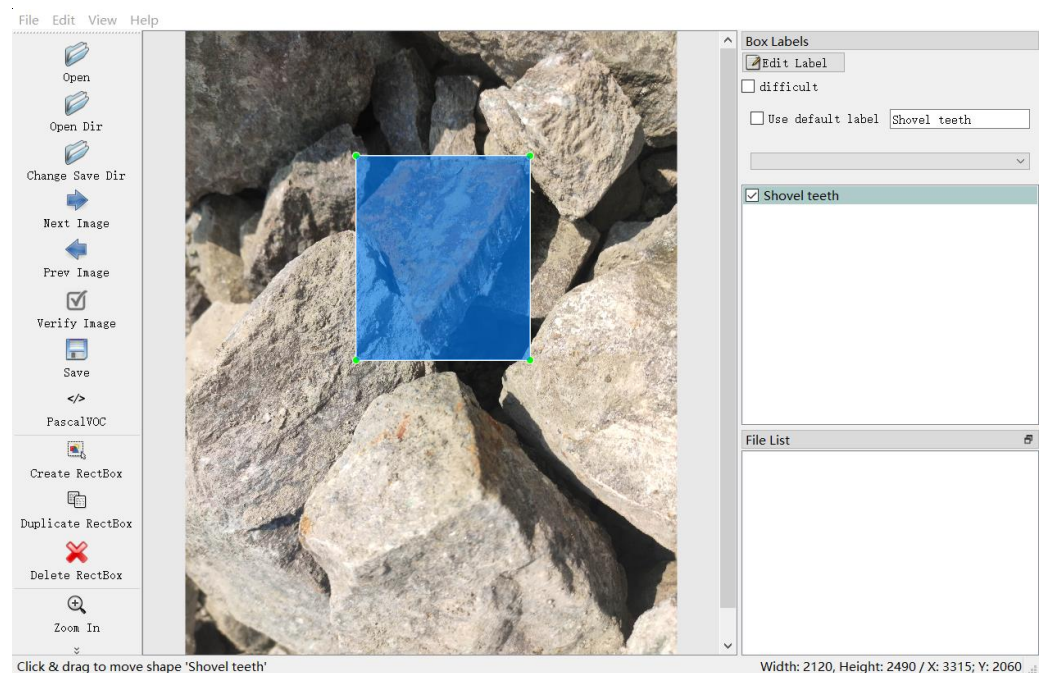




**Figure 9.** Comparison before and after noise reduction.

## 2. Annotated datasets

After completing the above steps, it is necessary to label the obtained data image, which needs to be completed manually. Image annotation is an important part of image foreign object detection. Only through the process of image annotation, when the same kind of foreign object reappears, can the foreign object detection system recognize this kind of foreign object. The quality of image annotation directly affects the accuracy of subsequent training and recognition. Frequently used image annotation tools include LabelImg, labelbox and via. LabelImg is used in this study to label the collected and processed data. The detailed annotation process is illustrated in Figure 10.



**Figure 10.** Annotation process.

A total of 5234 available sample data were obtained through sample expansion, image enhancement, and the elimination of poor images. The dataset comprises 2146 images of elongated foreign objects, 2324 images of polygonal foreign objects, and 764 images of circular foreign objects. The collected foreign object sample images of mining belt conveyor are allocated according to the proportion of 6:2:2 to the training set, verification set and test set, respectively. After being annotated by LabelImg, the image will form a label file in XML format.

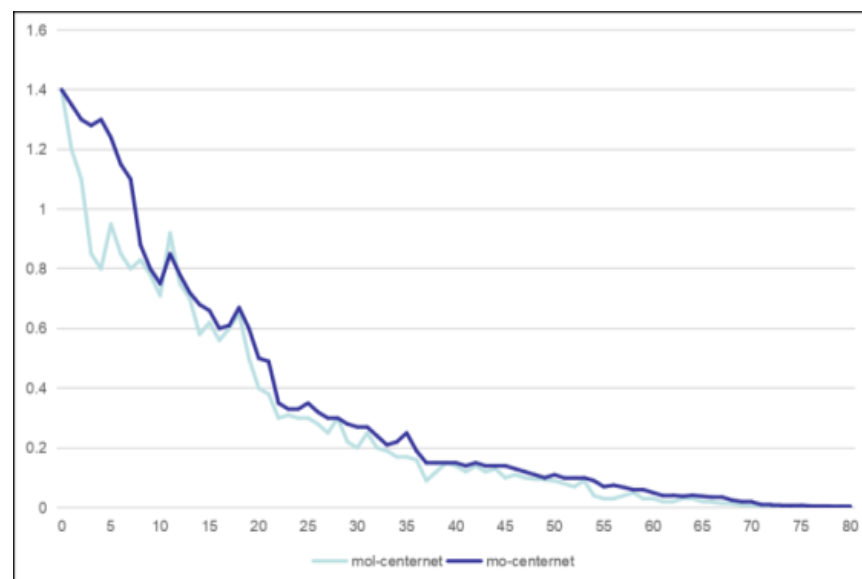
### 3. Results

#### 3.1. Test Environment Configuration

Data training uses the improved Center-Net target detection network training, the transfer learning method, and the pre-trained VOC dataset weights, iterates a total of 80 epochs, sets the batch size to 8, and determines the initial learning rate. The foreign object recognition network uses the pytorch 1.2 framework, and the computer used in the experiment is configured as follows: the processor is intel(R)-Xeon(R)-Silver-4114; the memory is 128 G; the GPU is 2 Nvidia-GeForce-RTX-2080Ti-12G. The training samples used have an input resolution of  $512 \times 512$  and all models have an output resolution of 128.

#### 3.2. Target Set Training

The processed datasets were tested under different algorithms. Experiment 1 is the mine-Center-Net algorithm obtained by improving the receptive field. The second experiment is to modify the loss function based on the mine-Center-Net algorithm. The modified algorithm is called mine-Center-Net. In order to ensure the balance of the model, it is necessary to continuously set  $\alpha_1$ ,  $\alpha_2$ ,  $\beta_1$ ,  $\beta_2$  and  $\delta$ . The following data are obtained through multiple optimization experiments: ① When  $\beta_1 = 3$ ,  $\delta = 0.2$ , a reduction in negative sample loss is achieved, and the corresponding loss function is  $L_n$ . ② When  $\beta_2 = 1.5$ , to achieve the improvement of positive sample loss, the corresponding loss function is  $L_p$ . The  $\alpha$  parameter controls the ratio of positive samples to negative samples; set up  $\alpha_1 = 0.25$ ,  $\alpha_2 = 1$ , the corresponding loss function is  $L_K$ . Figure 11 below shows the loss function curves of the two models.



**Figure 11.** Loss function.

During the experiments, we observed a rapid decrease in the loss curve within the first 50 epochs. Convergence was largely achieved around 80 epochs, with further training providing negligible improvements. Therefore, we set the total number of training rounds to 80. When the iteration reaches 80 rounds, the loss function value of the mine-Center-Net algorithm is 0.0052, and the mine-Center-Net is 0.0024. The minimum value of the model after the modified loss function is reduced by a factor of 2.166, making the loss function of the entire model closer to 0. This shows that it is effective to optimize the loss function by introducing parameters into the loss function to adjust the proportion of positive and negative samples, respectively.

### 3.3. Results of Experiment

Using the mine-Center-Net algorithm to carry out the test. The foreign body detection process and results are shown in Figure 12.



**Figure 12.** Test results.

Then, the method proposed in this paper is compared with the general algorithm, and the improved method is further analyzed. Finally, each method is evaluated using four indicators: precision, average precision (AP), mean average precision (MAP) and model inference speed (FPS). MAP is the average of multiple class APs. The specific calculation formula is as follows:

$$Precision = \frac{Num(TP)}{Num(TP) + Num(FP)} = \frac{Num(TP)}{N} \quad (7)$$

*TP*: number of detection boxes with  $IoU > 0.5$ .

*FP*: number of detection boxes with  $IoU \leq 0.5$ .

$$AP = \frac{\sum Precision}{Num(Totalobjects)} \quad (8)$$

Total objects: total number of objects.

$$MAP = \frac{\sum AP}{Num(Classes)} \quad (9)$$

Classes: number of types.

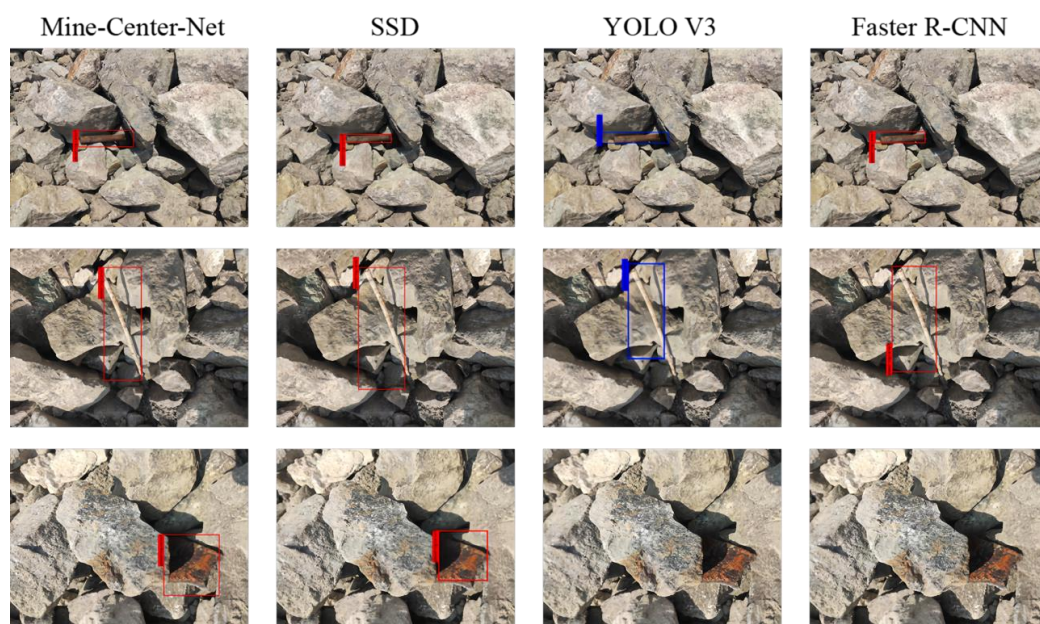
In this paper, five groups of control experiments are designed, which are the mine-Center-Net algorithm obtained by expanding the receptive field and improving the loss function and SSD, YOLO V3, Faster R-CNN, and Center-Net. The specific results are shown in Table 1 below.

Anchor-based object detection algorithms, such as YOLO, require the generation of numerous anchor boxes within the foreign object image during the detection process. This inherently increases model redundancy and significantly reduces detection robustness. In contrast, anchor-free object detection, by predicting the center point of the target object directly from the feature map, greatly enhances network efficiency. Furthermore, the absence of anchor boxes eliminates the need for non-maximum suppression (NMS) processing, thereby improving the system's overall detection accuracy.

**Table 1.** Algorithm contrast.

Model	AP			MAP	Inference Speed
	Rill Root	Shovel Teeth	I-Beam		
SSD	0.652	0.801	0.750	0.734	21.05 FPS
YOLO V3	0.562	0.780	0.722	0.688	22.02 FPS
Faster R-CNN	0.722	0.780	0.802	0.774	20.08 FPS
Center-net	0.750	0.780	0.953	0.828	26.08 FPS
Our method	0.822	0.941	0.963	0.909	26.05 FPS

As demonstrated in the detection results, the proposed mine–Center-Net achieves an average detection accuracy of 91.10% for foreign objects on mining belt conveyors, surpassing the performance of general-purpose algorithms in this specific application. Compared to SSD with VGG16, YOLO V3 with Darknet53, and Faster R-CNN with VGG16, mine–Center-Net exhibits a significant improvement in mean average precision (mAP) of 17.70%, 22.30%, and 13.70%, respectively. In addition to its superior accuracy in identifying various foreign objects, mine–Center-Net achieves a real-time detection speed of 24.05 FPS, representing a speed enhancement of 3.00 FPS, 2.03 FPS, and 3.97 FPS over SSD, YOLO V3, and Faster R-CNN, respectively. The actual detection results are shown in Figure 13.

**Figure 13.** Comparison between mine–Center-Net and common models.

In the comparative experiments, the two-stage object detection algorithm used was Faster R-CNN, while the one-stage anchor-based algorithms included SSD and YOLOv3, and the one-stage anchor-free algorithm was Center-Net. The experimental results indicate that among the anchor-based algorithms, Faster R-CNN demonstrated the precision advantages of the two-stage approach, albeit with a slight sacrifice in speed, which aligns with the characteristics of object detection algorithms. The anchor-free model Center-Net, along with the method proposed in this paper, exhibited superiority for the specific task at hand. Furthermore, the proposed method achieved an additional improvement in accuracy after modifications, while the speed remained largely unchanged. From the results, we observed a significant enhancement in the recognition accuracy of the shovel teeth, which are relatively small and prone to being buried. The improved network has enhanced the detection capability for small objects.

Our object detection network employs an anchor-free one-stage detection method that achieves end-to-end object detection by predicting center points. Since the predictions

are centered around these points, it performs better in detecting small objects against complex backgrounds, specifically resulting in fewer false positives. The introduced dilated convolution method expands the convolutional kernel by adding spaces (zeros) between its elements, which effectively increases the network's receptive field, allowing it to capture broader contextual information and leading to more accurate predictions.

#### 4. Conclusions

The purpose of this paper is to improve the traditional method of foreign body detection without an anchor frame, and to provide a high-stability and high-precision method for the foreign body detection of Molybdenum belt conveyors.

Due to the problems of a too large recognition area, insufficient dark details and too much noise in the images collected on the Molybdenum ore belt conveyor, the collected images were processed with histogram equalization and median filter noise reduction, effectively reducing the impact of environmental factors on the later identification work. The introduction of atrous convolution through space improves the perception field of view of the Center-Net model. In order to solve the problem of unbalanced positive and negative samples in the training process of the metal ore belt conveyor, the parameters  $\alpha_1$ ,  $\alpha_2$ ,  $\beta_1$ ,  $\beta_2$  and  $\delta$  are introduced into the traditional loss function, to optimize the ratio of positive and negative samples in the training process, reduce the overall loss function value of the algorithm, and improve the detection accuracy of the algorithm. According to our experiment, it can be seen that the recognition accuracy of the method proposed in this paper for bar, polygon and circular objects reaches 0.822, 0.941 and 0.909, respectively, and the recognition accuracy of foreign objects on various conveyor belts is significantly improved. The high-precision identification of foreign objects on metal ore belt conveyors is realized. The experimental results show that the mine-Center-Net method proposed in this paper can effectively detect foreign objects on metal ore belt conveyors. Compared with the existing algorithms SSD, YOLO V3, Faster R-CNN and Center-Net, the recognition accuracy is significantly improved. The following work will focus on how to optimize the algorithm and improve the recognition speed of the algorithm.

In the aforementioned work, we conducted research on the intelligent visual detection method for foreign objects on the molybdenum ore conveyor belt. By improving the existing framework network and optimizing the dataset, we achieved the optimization of the accuracy and timeliness of foreign object visual detection on the conveyor belt. However, there is still much room for improvement in the current work. For example, (1) enriching the types of foreign objects and detecting large ore blocks that may affect the stability of ore conveyor belt transportation; (2) innovating algorithms to enhance the stability of visual detection algorithms for possible dust conditions and visual motion blur that may be caused by high-speed conveyor belts, to improve their adaptability to industrial environments; (3) conducting further research on foreign object recognition to achieve a complete solution for the perception, monitoring, and operation control of foreign object recognition, localization, and exclusion. In the future, researchers can combine experiments with engineering projects to conduct follow-up research on the above-mentioned issues.

**Author Contributions:** Methodology, M.L.; software, M.L.; formal analysis, M.L.; investigation, M.L.; resources, C.L.; data curation, X.Z.; writing—original draft preparation, X.Y.; writing—review and editing, X.Y.; visualization, R.H.; supervision, M.L.; project administration, M.L.; funding acquisition, C.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Natural Science Foundation of China (51974223, title: "Research on Intelligent Fusion and situation assessment of multi-source heterogeneous flow data of rock failure in underground metal mines"), and the Natural Science Foundation of Shaanxi Province (2019 JLP-16, title: "Research and development of key technologies for intelligent production control and intelligent decision-making of open pit coal mine under cloud service").

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data is contained within the article.

**Conflicts of Interest:** Author Xuyang Zhao was employed by the company Luanchuan Longyu Molybdenum Industry Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Chandra, M.A.; Bedi, S.S. Survey on SVM and their application in image classification. *Int. J. Inf. Technol.* **2021**, *13*, 1–11. [[CrossRef](#)]
2. Fernando, B.; Fromont, E.; Tuytelaars, T. Mining mid-level features for image classification. *Int. J. Comput. Vis.* **2014**, *108*, 186–203. [[CrossRef](#)]
3. Obaid, K.B.; Zeebaree, S.; Ahmed, O.M. Deep learning models based on image classification: A review. *Int. J. Sci. Bus.* **2020**, *4*, 75–81.
4. Jena, B.; Nayak, G.K.; Saxena, S. Convolutional neural network and its pretrained models for image classification and object detection: A survey. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e6767. [[CrossRef](#)]
5. Fan, Q.; Zhuo, W.; Tang, C.-K.; Tai, Y.-W. Few-shot object detection with attention-RPN and multi-relation detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
6. Jesse, K.; Changying, L. A pulsed thermographic imaging system for detection and identification of cotton foreign matter. *Sensors* **2017**, *17*, 518. [[CrossRef](#)] [[PubMed](#)]
7. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 1904–1916. [[CrossRef](#)] [[PubMed](#)]
8. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)]
9. Zhan, W.; Sun, C.; Wang, M.; She, J.; Zhang, Y.; Zhang, Z.; Sun, Y. An improved yolov5 real-time detection method for small objects captured by uav. *Soft Comput.* **2022**, *26*, 361–373. [[CrossRef](#)]
10. Liu, Z.; Wu, W.; Gu, X.; Li, S.; Wang, L.; Zhang, T. Application of combining yolo models and 3d gpr images in road detection and maintenance. *Remote Sens.* **2021**, *6*, 1081. [[CrossRef](#)]
11. Santos, P.M.; Simeone, M.; Pimentel, M.; Sena, M.M. Non-destructive screening method for detecting the presence of insects in sorghum grains using near infrared spectroscopy and discriminant analysis. *Microchem. J.* **2019**, *149*, 104057–104071. [[CrossRef](#)]
12. Jiang, Y.; Ge, H.; Zhang, Y. Detection of foreign bodies in grain with terahertz reflection imaging. *Opt.—Int. J. Light Electron Optics.* **2018**, *181*, 1130–1138. [[CrossRef](#)]
13. Taufiq, M.K.M.; Sallehuddin, I.; Amri, M.Y.M.; Mahdi, F. Noninvasive techniques for detection of foreign bodies in food: A review. *J. Food Process Eng.* **2018**, *41*, e12808.
14. Mustafic, A.; Jiang, Y.; Li, C.Y. Cotton contamination detection and classification using hyperspectral fluorescence imaging. *Text. Res. J.* **2016**, *86*, 1574–1584. [[CrossRef](#)]
15. Li, J.; Yang, L.; He, Y.; Li, W.; Wu, C. Terahertz nondestructive testing method of oil-paper insulation debonding and foreign matter defects. *IEEE Trans. Dielectr. Electr. Insul.* **2021**, *28*, 1901–1908. [[CrossRef](#)]
16. Li, D.; Wang, R.; Xie, C.; Liu, L.; Liu, W. A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network. *Sensors* **2020**, *20*, 578. [[CrossRef](#)] [[PubMed](#)]
17. Wang, R.; Jiao, L.; Xie, C.; Chen, P.; Du, J.; Li, R. S-rpn: Sampling-balanced region proposal network for small crop pest detection. *Comput. Electron. Agric.* **2021**, *187*, 106290–106299. [[CrossRef](#)]
18. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: A simple and strong anchor-free object detector. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 1922–1933. [[CrossRef](#)]
19. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
20. Luo, B.; Kou, Z.; Han, C.; Wu, J.; Liu, S. A Faster and Lighter Detection Method for Foreign Objects in Coal Mine Belt Conveyors. *Sensors* **2023**, *23*, 6276. [[CrossRef](#)]
21. Bewley, A.; Upcroft, B. Background Appearance Modeling with Applications to Visual Object Detection in an Open-Pit Mine. *J. Field Robot.* **2017**, *34*, 53–73. [[CrossRef](#)]
22. Tanaka, S.; Ohtani, T.; Narita, Y.; Hatsukade, Y.; Suzuki, S. Development of metallic contaminant detection system using rf high-tc squids for food inspection. *IEEE Trans. Appl. Supercond.* **2015**, *25*, 1601004. [[CrossRef](#)]
23. Tanaka, S.; Narita, Y.; Ohtani, T.; Ariyoshi, S.; Suzuki, S. Development of metallic contaminant detection system using rf high-tc squid with cu pickup coil. *IEEE Trans. Appl. Supercond.* **2016**, *12*, 600304. [[CrossRef](#)]
24. Tanaka, S.; Ohtani, T.; Uchida, Y.; Hatsukade, Y.; Suzuki, S. Contaminant detection system using high tc squid for inspection of lithium ion battery cathode sheet. *IEICE Trans. Electron.* **2015**, *98*, 174–177. [[CrossRef](#)]
25. Tanaka, S.; Ohtani, T.; Krause, H.J. Prototype of multi-channel high-tc squid metallic contaminant detector for large sized packaged food. *IEICE Trans. Electron.* **2017**, *E98.C*, 269–273. [[CrossRef](#)]
26. Du, S.; Xing, J.; Li, J.; Zhang, C. Open-pit mine extraction from very high-resolution remote sensing images using OM-DeepLab. *Nat. Resour. Res.* **2022**, *31*, 3173–3194. [[CrossRef](#)]

27. Zhang, X.; Hao, Y.; Shangguan, H.; Zhang, P.; Wang, A. Detection of surface defects on solar cells by fusing multi-channel convolution neural networks. *Infrared Phys. Technol.* **2020**, *108*, 10333–10334. [[CrossRef](#)]
28. Li, C.; Liu, J.; Zhu, J.; Zhang, W.; Bi, L. Mine image enhancement using adaptive bilateral gamma adjustment and double plateaus histogram equalization. *Multimed. Tools Appl.* **2022**, *81*, 12643–12660. [[CrossRef](#)]
29. Bora, D.J.; Gupta, A.K. A new efficient color image segmentation approach based on combination of histogram equalization with watershed algorithm. *Int. J. Comput. Sci. Eng.* **2016**, *4*, 156–167.
30. Zhao, Y. Multi-level denoising and enhancement method based on wavelet transform for mine monitoring. *Int. J. Min. Sci. Technol.* **2013**, *23*, 163–166.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.