Article

Detection of Large Foreign Objects on Coal Mine Belt Conveyor Based on Improved

Kaifeng Huang [1], Shiyan Li [1,*], Feng Cai [2] and Ruihong Zhou [1]

[1]   School of Mechanical and Electrical Engineering, Huainan Normal University, Huainan 232001, China
[2]   State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mines, Huainan 232001, China
*   Correspondence: 19955463478@163.com

Abstract: An algorithm based on the YOLOv5 model is proposed to address safety incidents such as tearing and blockage at transfer points on belt conveyors in coal mines caused by foreign objects mixed in with the coal flow. Given the tough underground conditions and images acquired with low quality, recursive filtering and MSRCR image enhancement algorithms were utilized to preprocess the dynamic images collected by underground monitoring devices, substantially enhancing image quality. The YOLOv5 model has been improved by introducing a multi-scale attention module (MSAM) during the channel map slicing, thereby increasing the model's resistance to interference from redundant image features. Deep separable convolution was utilized in place of conventional convolution to detect, identify, and process large foreign objects on the belt conveyor as well as to increase detection speed. The MSAM-YOLOv5 model was trained before being installed on the NVIDIA Jetson Xavier NX platform and utilized to identify videos gathered from the coal mine belt conveyor. According to the experimental findings, the upgraded MSAM-YOLOv5 model has a greater recognition accuracy than YOLOv5L, with an average recall rate for different foreign objects of 96.27%, an average detection accuracy of 97.35%, and a recognition speed of 44 frames/s. The algorithm assures detection accuracy while increasing detection speed, satisfying the requirements for large foreign object detection on belt conveyors in coal mines.

Keywords: belt conveyor; large foreign object recognition; YOLOv5; MSAM; deep separable convolution

1. Introduction

Industrialization has stimulated an increasing demand for coal resources. In this process, belt conveyors serve a crucial role in the coal industry [1]. However, belt conveyors are vulnerable to damages such as deviation, slipping, and tearing due to overload, lengthy transport distances, and foreign objects [2]. Real-time detection of large foreign objects on belt conveyors in coal mines and timely detection and processing are crucial to minimizing the damage caused by foreign objects to the conveyor. Not only does this improve coal mine safety, but it also contributes significantly to production efficiency.

Ultrasonic, radar, and photoelectric sensing technologies were the methods for detecting foreign objects on belt conveyors initially adopted globally [3,4]. Due to disadvantages such as high detection costs and difficulty in maintenance, these detection approaches were eventually phased out. With the advent of the era of big data and the continuous advancement of GPU processing, deep-learning-based object detection algorithms have become increasingly prevalent. There are two types of these algorithms [5]: two-stage object detection algorithms and single-stage object detection algorithms. Single-stage object detection algorithms include the SSD [6] and You Look Only Once (YOLO) series [7–10]. Two-stage object detection algorithms include R-CNN [11], Fast R-CNN [12], and Faster R-CNN [13].

Due to their lengthy inference time caused by the generation of candidate regions, two-stage algorithms struggle to meet the real-time requirements of foreign object detection.

Regression-based YOLO series models have both high accuracy and real-time performance. You Only Look Once version 5 (YOLOv5) is the fifth generation of the YOLO series of object detection algorithms, which inherit the end-to-end advantages of the YOLO family. It directly takes the raw image as input and outputs the location and category information of the target in one go, characterized by fast speed, high accuracy, and light weight. YOLOv5 is currently widely used in video surveillance, autonomous driving, robotic vision, medical image processing and other fields. Ref. [14] proposes a method for garbage detection and classification using the YOLOv5s network model. The trained YOLOv5s network is used to extract features and location information from images of different types of garbage, achieving classification and detection of garbage. The method was tested in real-world scenarios. Ref. [15] introduces a real-time detection method for apple targets on a picking robot based on an improved YOLOv5 algorithm. The improved YOLOv5 algorithm is used to recognize apple images to improve the accuracy of target detection. These improvements include using more advanced feature extractors and classifiers and adjusting the network structure and parameters. Ref. [16] explores the use of an improved YOLOv5 network for breast tumor detection and classification. The author improves the original YOLOv5 network by adding a convolutional neural network module to enhance feature extraction capabilities, making it more suitable for breast tumor detection and classification tasks. Detection of foreign objects in underground production processes is a challenging task due to the complex and difficult-to-observe nature of the underground environment. However, YOLOv5 as an advanced target detection algorithm can offer certain advantages in this field. Ref. [17] studies the application of an improved YOLOv5 model for coal gangue recognition. Optimizing and improving the model improves its accuracy and performance in the coal gangue recognition task. A YOLOv5 model was used to detect foreign objects on belt conveyors in reference [18], with the introduction of the convolutional block attention module (CBAM) to enhance the recognition accuracy and speed of foreign objects. Pruning based on channels and layers is implemented for the model of foreign object detection on the belt conveyor in reference [19], which increases the model's detection speed.

The attention mechanism can help the target detection algorithm pay more attention to the key areas and features in the image, so as to reduce the interference of redundant information and noise and improve the detection accuracy. By learning the importance of different regions and features, the attention mechanism can adaptively adjust the degree of attention of different regions to better capture target features and distinguish between different targets [20–23]. A feature extraction method for target detection is studied, which gradually refines the feature information from shallow layer to deep layer by constructing a pyramidal feature representation. Based on FPN, a top-down attention mechanism is introduced to enhance the representation ability of low-level features by transferring the feature information from high-level to low-level [24]. A channel-wise attention mechanism is proposed for adaptive weighting of features. By learning the importance of each channel, the channel attention mechanism can improve the accuracy of target detection [25–27]. A spatial attention mechanism is proposed to weight each pixel in a feature map. By learning the importance of each pixel, the spatial attention mechanism can improve the accuracy of target detection, especially for small targets [28]. A method combining the channel attention mechanism and context embedding is proposed for scaling problems in target detection. By learning the importance of feature channels at different scales and embedding contextual information into feature representations, the method can improve the accuracy and robustness of target detection. These object detection algorithms combined with the attention mechanism optimize and improve the object detection problem in different aspects. It should be noted that object detection algorithms combining attention mechanisms also have some challenges and limitations. For example, the design and training of attention mechanisms requires additional computational resources and time, potentially increasing the complexity and development cost of the algorithm. At the same time, the effect of the attention mechanism is also affected by the dataset, network structure, training strategy, and other factors, so it needs to be fully verified and optimized.

Considering the above networks and realizing the complementary advantages of each network, this paper proposes a YOLOv5 model combining image enhancement and the attention mechanism to realize foreign-body detection in a coal mine belt conveyor. Among them, a multiscale attention module (MSAM) combined with the CBAM is proposed. After assembling feature maps of different levels, YOLOv5 uses the MSAM model to weight different feature channels. By learning different weights, the model can select more important feature channels according to task requirements, thus reducing the impact of redundant features. In addition to channel attention, the CBAM in the MSAM model can pay attention to features of different regions and weight different feature maps according to different regions of the image, thus reducing the impact of redundant features. In order to further improve the detection speed, the models in this paper combine depthwise separable convolution (DWConv). The improved model can detect foreign bodies efficiently while ensuring accuracy. The experimental results show that the proposed method has a good balance between accuracy and detection speed. These innovative contributions are summarized as follow:

(1) We propose an enhanced YOLOv5 model that incorporates the attention mechanism module (CBAM) and the multi-scale attention module (MSAM) to augment the model's feature extraction capability and mitigate redundant features inherent in the original YOLOv5 framework.

(2) The efficacy of our model improvement is empirically substantiated through a series of ablation experiments. The results demonstrate a substantial enhancement in the detection accuracy of foreign objects on conveyor belts through the utilization of image processing techniques. Notably, the integration of the MSAM yields a more pronounced improvement in model performance, as evidenced by a 4.78% increase in precision (P) and a 7.86% increase in recall (R) metrics when compared to the initial YOLOv5 architecture.

(3) To enhance computational efficiency, we introduce a lightweight convolutional neural network known as DWconv. By replacing the conventional convolutional network with DWconv, our model achieves an impressive frame rate (FPS) of 46.8 frames per second, with only a marginal reduction in the P and R metrics. This attainment underscores the delicate equilibrium achieved between accuracy and efficiency.

(4) The synergistic integration of the MSAM and DWconv methodologies further elevates the performance of the YOLOv5 model, yielding heightened recognition speed and accuracy and augmenting the detection and recognition capabilities of target objects.

2. Image Processing Algorithm for Coal Transport Image in Underground Mines

The underground conditions of coal mines are unlike those of any other industry. Due to the presence of coal particles and water mist in the air, coal mines require artificial lighting. Moreover, due to the terrible conditions in underground mines, image acquisition could be extremely difficult. Therefore, image preprocessing is necessary for removing or minimizing distracting information and enhancing useful objective information. The preprocessing of images is a crucial stage in the visual detection system for large foreign objects, particularly in the processing of images of coal transport in underground mines. Image noise reduction and enhancement increase the reliability of subsequent identification, selection, and classification of large foreign objects.

2.1. Recursive Filtering Denoising Algorithm

The recursive filtering algorithm, also known as feedback filtering, derives its name from the fact that a portion of the denoised output is fed back into the input stream to influence consecutive denoising iterations. This algorithm is primarily employed for

noise suppression in dynamic images. The calculation formula for recursive filtering in a first-order gradient is represented by the following equation:

$$Y_i = \sum_{i=0}^{n-1}(\alpha_i X_i) - \sum_{j=1}^{n}(\beta_j Y_{i-j})$$

(1)

where X = the input of recursive filtering; Y = the output; and α_i, β_j = the weight.

The pixels along a certain linear distance can affect the output of the recursive filtering algorithm. Assuming that distance is the distance between two adjacent pixels, the previously stated calculation formula can be simplified as follows:

$$Y_i = (1 - \gamma_{i,i-1}\alpha)X_i + \gamma_{i,i-1}\alpha Y_{i-1}$$

(2)

where α = the coefficient and $\gamma_{i,i-1}$ = the weight coefficient between X_i and Y_{i-1}. As shown in the equation above, the simplified recursive filtering algorithm outputs Y_i, which is related only to X_i and Y_{i-1}. The recursive filter is essentially a lowpass filter that is often used to suppress random noise in dynamic images. The filter size is a parameter that needs to be set manually, and its setting has a great impact on the de-noising effect. Too small a filter size may lead to incomplete noise removal, and too large a filter size may smooth out image details, including edge details. Therefore, a suitable filter size was selected according to the actual situation to balance the image noise and the preservation of edge details. Mean square error (MSE) and peak signal-to-noise ratio (PSNR) [29] were employed in this study as image-denoising evaluation indicators to select a suitable filter size. The recursive filtering algorithm customizes parameter settings to denoise the images and calculates MSE and PSNR. These evaluations are presented in Table 1 and Figure 1. According to results in Table 1 and Figure 1, a filter size of 2.5 yields the smallest MSE of 0.24 and highest PSNR of 54.3, suggesting that the edge details of denoised coal transport images remain unaltered while the image quality improves significantly. In light of this, it can be concluded that the recursive filter algorithm with a filter size of 2.5 is the optimum option for denoising coal transport images.

Table 1. MSE and PSNR values of recursive filtering denoised images with different parameters.

Size	MSE	PSNR
1.5	6.74	39.5
2.0	0.57	52.1
2.5	0.24	54.3
3.0	1.51	44.9
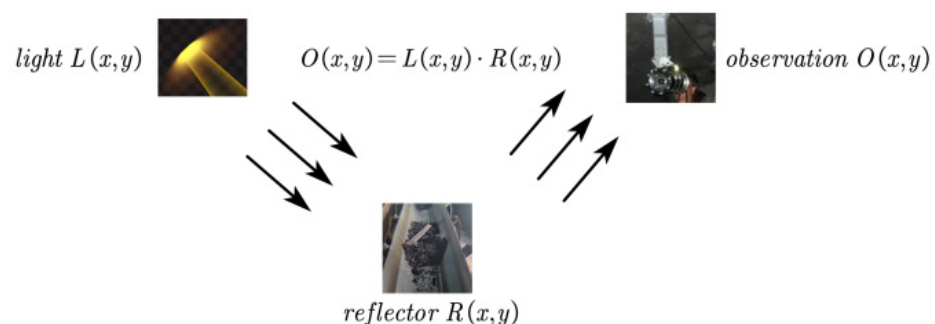
Figure 1. Principle of object imaging.

2.2. MSRCR Image Enhancement Algorithm

Due to coal dust particles in the air, low light, and uneven illumination conditions within coal mines result in low-quality images of coal transport, characterized by low brightness, uneven illumination, and indistinct object outlines. It is crucial to develop effec-

tive image enhancement algorithms for enhancing the quality of images of coal transport to address these issues. The enhanced images of coal transport disclose useful information and features within the images, thereby facilitating subsequent feature extraction and rapid detection of target objects. As for image quality improvement, images are typically subjected to global stretching, brightness and color tone adjustment, and other general image processing operations during image enhancement. In this study, an image enhancement technique termed Multi-Scale Retinex with Color Restoration (MSRCR) was applied to remove undesirable interference from photos. The Retinex image enhancement algorithm operates on the following principle.

According to the Retinex theory, the image obtained by surveillance devices is related to both incident light and the reflected image. The following equation expresses the relationships:

$$O(x, y) = L(x, y) \cdot R(x, y) \tag{3}$$

The Retinex image enhancement algorithm selectively removes or reduces the effects caused by the incident component $L(x, y)$ while preserving the reflectance component $R(x, y)$, which reflects the original colors of the imaged objects. In 1997, Jobson et al. introduced the Single-Scale Retinex (SSR) algorithm. However, SSR is difficult to balance between preserving image color information and detail enhancement. Therefore, academics have proposed the Multi-Scale Retinex (MSR) algorithm that uses multiple scales to enhance images. In addition, the MSRCR algorithm integrates a color restoration factor into MSR to adjust color distortion resulting from contrast enhancement in local image regions. The Retinex algorithm may introduce color distortion when enhancing image details because the original color information of the image may be changed when the algorithm adjusts the brightness, contrast, and color of the image. In order to solve this problem, the MSRCR algorithm is adopted in this paper, which reduces the color distortion and preserves the original color information of the image through multi-scale processing and color recovery technology. Specifically, the MSRCR algorithm uses a multi-scale method to decompose and enhance the image. First, the original image is divided into multiple scales, each of which corresponds to different high and low frequency information. Then, the Retinex algorithm is applied at each scale for image enhancement, including brightness, contrast, and detail enhancement. During the enhancement process, different scales of information are considered simultaneously to preserve the details and color information of the image. This effectively highlights information in relatively darker areas, eliminates color distortion defects in images, and achieves a better balance between color restoration and image detail enhancement. The algorithm can be expressed using the following equation:

$$\log[R_{MSR}(x, y)] = \sum_{j=1}^{K} \omega_j \{ \log[O(x, y)] - \log[F_j(x, y)O(x, y)] \} \tag{4}$$

$$R_{MSRCR_i}(x, y) = C_i(x, y)R_{MSR_i}(x, y) \tag{5}$$

$$R_{MSRCR_i} = G[R_{MSRCR_i}(x, y) - b] \tag{6}$$

$$C_i(x, y) = f\left[I_i'(x, y)\right] = f\left[\frac{I_i(x, y)}{\sum_{i=0}^{N} I_i(x, y)}\right] \tag{7}$$

where $R_{MSR}(x, y)$ = the reflection component after MSR computation; $O(x, y)$ = the image to be enhanced; $F_j(x, y)$ = the Gaussian filter function with a scale parameter of σ_j; K = the number statistics of scale parameter σ, generally equal to 3; ω_j = the weight factor of the j-th filtering function, meeting $\sum_{j=1}^{K} \omega_j = 1$, and generally $\omega_j = \frac{1}{3}$; $I_i(x, y)$ = the image of the i-th channel; and $C_i(x, y)$ = the color recovery factor of the i-th channel.

As shown in Figure 2, there is a clear contrast between the image of a conveyor belt before and after the application of MSRCR image enhancement. It is evident that after

image enhancement, the outline of the conveyor belt becomes clearer with more distinct features upon image enhancement.

Figure 2. Comparison before and after image enhancement.

3. YOLOv5 Algorithmic Improvement

The YOLOv5 model has fewer parameters and requires less memory, making it better suited for underground equipment. It satisfies the requirements for detecting large foreign objects on mining conveyor conveyors due to its high detection accuracy and quick processing speed. YOLOv5 includes YOLOv5s, YOLOv5m, YOLOv5L, and YOLOv5x, which are four models in the YOLOv5 series. The "L" represents "large". YOLOv5L has relatively high accuracy but slower detection speed, making it suitable for scenarios that require higher accuracy. The YOLOv5L model is upgraded in this study by integrating the MSAM mechanism in the neck portion and DWConv in place of regular convolution in the head section. Figure 3 depicts the improved YOLOv5L structure, which consists primarily of the input, backbone, neck, and head sections [30]. The integration of the MSAM mechanism in the neck will increase the complexity of the model. This is because the MSAM mechanism introduces more parameters and computations, which results in higher memory requirements and computational load. On the other hand, the DWConv mechanism reduces the complexity and computational load of the model by dividing the convolution operation into two steps: depthwise convolution and pointwise convolution. The introduction of these new mechanisms may increase the complexity, computational burden, memory requirements, and inference time of the model. However, these mechanisms typically provide better performance, especially on large-scale models and datasets.
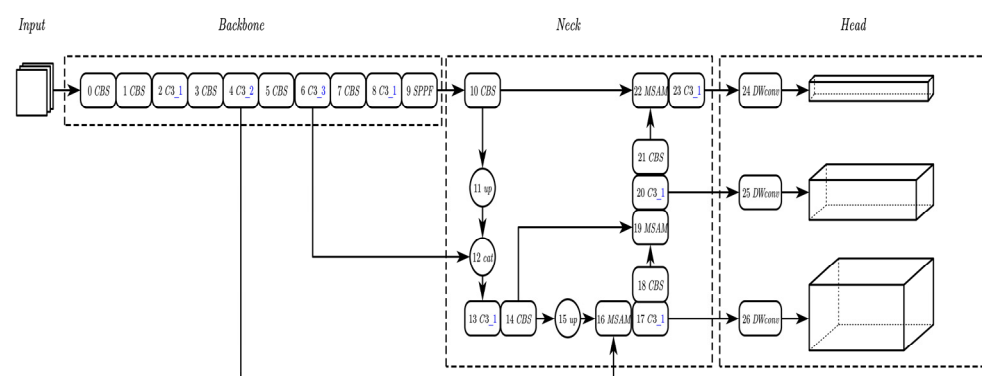
Figure 3. Improved YOLOv5 algorithm mode.

3.1. MSAM

The channel-wise concatenation procedure utilized by the original YOLOv5 network to combine two network layers can lead to duplicate characteristics that are not helpful for object detection. Additionally, the blurriness of mining conveyor belt surveillance films hinders the YOLOv5 model's capacity to extract features from foreign objects in the image. Consequently, it is necessary to enhance the model to better its feature extraction capabilities. Simply extracting superficial features would fail to detect large foreign objects on a conveyor belt. There is still space for improvement in the model that combines multiple feature maps in the channel dimension. To improve the model's ability to recognize objects and

lower the miss detection rate, this study introduces the MSAM. This module instructs the model to focus more of its attention on the regions of interest. Figure 4 depicts the MSAM structure, while the right side of the figure represents the structure of the CBAM [31]. The terms "MaxPool" and "AvgPool" stand for maximum and average pooling, respectively, and "SharedMLP" is short for shared multilayer perceptron. Cat is the concatenation operation, and the initial features are obtained by adding two feature maps, F_1 and F_2, which, respectively, represent shallow and deep feature maps. Following feature addition, feature refinement was carried out by the CBAM using attention processes in both the channel and spatial dimensions, as shown by the following equation:

$$F_a = C_a(F_1 \oplus F_2) \tag{8}$$

where C_a represents the operation after the CBAM, and F_a refers to the output, which is preferably mapped to the range $[0, 1]$ through the sigmoid function. This is used to determine the weights of different feature maps and then to perform concatenation to output a final feature fusion map, F.

$$F = [F_1 \otimes \sigma(F_a); F_2 \otimes (1 - \sigma(F_a))] \tag{9}$$
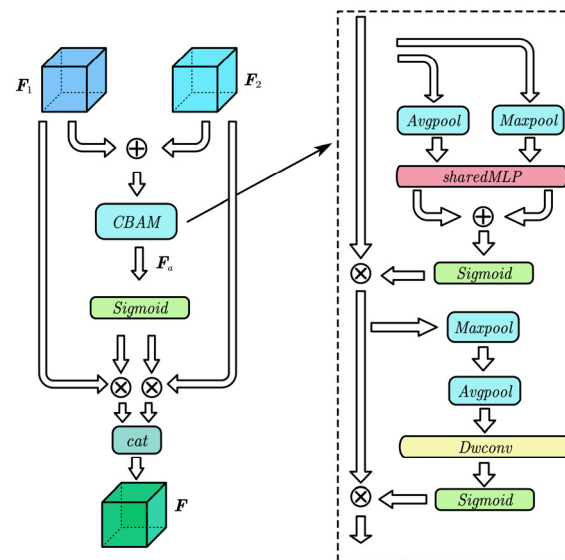
Figure 4. Structure of the MSAM.

3.2. DWconv

DWConv is used to increase the pace of foreign object detection and recognition to meet real-time detection and recognition requirements for large foreign objects on mining conveyor belts. Specifically, the conventional convolution in the head section is replaced with DWConv, resulting in a slight loss of accuracy while increasing the speed of foreign object detection [32–34]. The complexities of DWConv Q_D and ordinary convolution Q_c are shown in the equation below:

$$Q_c = D_K D_K M N D_W D_H \tag{10}$$

$$Q_D = D_K D_K M D_W D_H + M N D_W D_H \tag{11}$$

where D_K is the width of the convolution kernel; M is the number of input feature map channels; and D_H, D_W, and N represent the height, width, and number of output feature

map channels, respectively. The computational complexities of DWConv and ordinary convolution can be compared as follows:

$$\frac{Q_D}{Q_C} = \frac{D_K D_K M D_W D_H + M N D_W D_H}{D_K D_K M N D_W D_H} = \frac{1}{N} + \frac{1}{D_K^2} \tag{12}$$

As shown in Equation (7), DWConv instead of ordinary convolution for feature extraction of foreign objects greatly reduces computation complexity.

4. Results and Analysis

4.1. Experimental Environment and Data

The specific experimental environment is shown in Table 2.

Table 2. Environmental environment.

Configuration Name	Version/Parameter
Operating system	Windows 10 21H2
CPU	Intel(R) core (TM) i7-10875H CPU@2.3GHZ
GPU	NVIDIA GeForce GTX2060
Memory	16 GB
Algorithm framework	PyTorch 1.11.0

The experiment's sample data were gathered by real-time monitoring videos of coal mines with transporting conveyor belts. The labels for large coal gangue, large wood, and large metal objects, etc., were set, and 20,000 images were then annotated using the LabelImg software in accordance with the VOC2007 dataset format. To enhance the model's generalizing capability, the number of images was increased to 50,000, of which 12,230 were images of large coal gangue, 14,390 were images of large wood, and 9470 were images of large metal objects. In addition, 4000 were environmental photographs. The training set consisted of 18,200 randomly selected images, while the testing set consisted of 5000 images. The hyperparameters of a model were set based on prior knowledge. The total epochs for the model was set to 200. The learning rate was set to 0.01 for the first epoch and 0.001 for the last 100 epochs to prevent overfitting.

4.2. Recognition Effect of Improved YOLOv5 Model

This study conducted a comparative analysis of the original YOLOv5 model, YOLOv3 model, YOLOv5 model combined with the CBAM and the proposed improved YOLOv5 model. The training process is shown in Figure 5, where it can be seen that the YOLOv5 loss gradually decreased to around 0.05 after about 30 iterations and ultimately stabilized around 0.047. After introducing the attention mechanism module, both converged at around 20 iterations, and the MSAM made improvements on the basis of the CBAM, with its final loss decreasing to around 0.25, which was overall lower than that of the CBAM-YOLOv5.

For the trained model, the performance is evaluated based on the main evaluation metrics of precision (P), recall (R), and frames per second (FPS). Precision represents the accuracy of the model's detection results, measured as the ratio of true positives (TPs) to the total number of objects recognized as positive. A higher *p* value indicates a lower false positive rate. Recall measures the thoroughness of the model's detections, as it represents the ratio of true positives to the total number of actual objects. A higher R value indicates a lower false negative rate. Table 3 shows the P, R, and recognition speed of three models. Table 3 indicates that both YOLOv5 and its enhanced version have higher P, R, and recognition speed than YOLOv3. Therefore, we have chosen YOLOv5 as the overall model framework to improve its performance. The combination of the CBAM with YOLOv5 and the proposed improved YOLOv5 have both resulted in improvements, with the proposed

improved YOLOv5 performing better, with a P value of 97.35%, an R value of 96.27%, and a recognition speed of 0.022 s, equivalent to 44 FPS. These metrics meet the research scope.
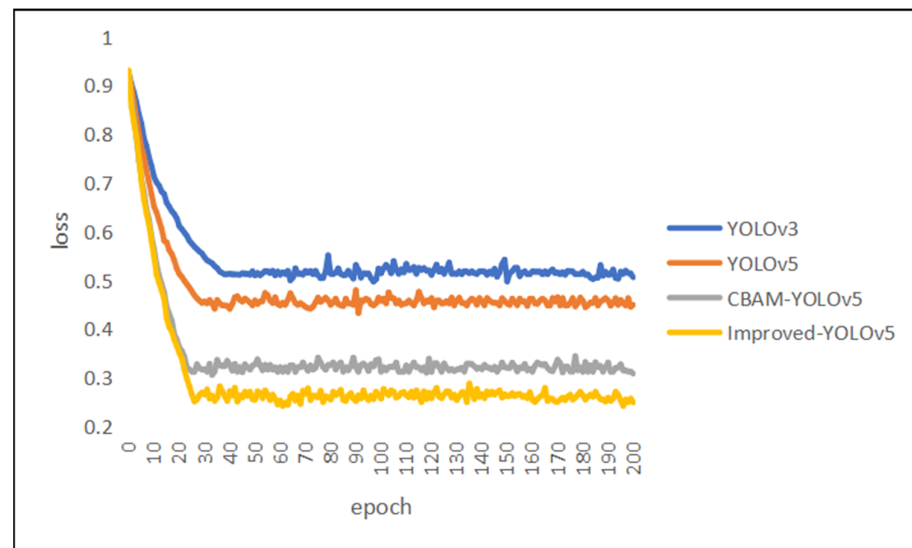

Figure 5. Comparison of loss functions.

Table 3. P, R and recognition speed of different models.

Models	P	R	FPS
YOLOv3	90.89%	86.38%	22
YOLOv5	92.57%	87.41%	41
CBAM-YOLOv5	94.51%	91.61%	44
Improved YOLOv5	97.35%	96.27%	44

Figure 6 contrasts the results before and subsequent to the enhancement. The detection results of the original YOLOv5 are presented in the left column, whereas the results of the enhanced YOLOv5 are in the right column. The confidence score is exhibited atop the bounding box, and the two models ascertain identical categories, which remain unlabeled in the illustration.
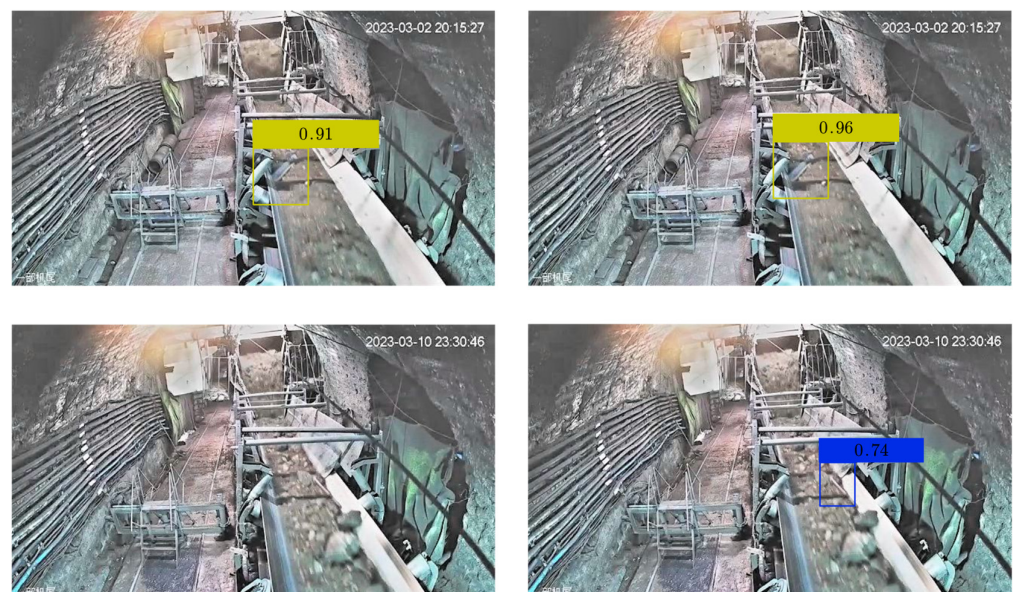
Figure 6. Identification results of different models.

The two images in the first row show the detection results of two algorithms for the detection of wooden debris. It can be seen that the improved YOLOv5 has a higher recognition confidence for large debris. The two images in the second row show the detection results of the two algorithms for anchor bolts. After introducing the MSAM, the improved YOLOv5 enhances the model's ability to extract foreign object features, and the anchor bolts are accurately identified without any missed detections. The recognition effect is better than that of the original YOLOv5.

4.3. Effects of the Ablation Experiments

The efficacy of image augmentation, the MSAM, and DWconv in enhancing YOLOv5 was assessed by conducting a series of ablation experiments. The experiments consisted of five distinct groups, namely the original YOLOv5, image processing, the MSAM, DWconv, and the enhanced YOLOv5 that was optimized with all three factors. The study employed identical equipment and the same dataset for the experiments, and the outcomes of the ablation experiments are presented in Table 4.

Table 4. Results of the ablation experiments.

Factors	P	R	FPS
Original YOLOv5	92.57%	88.41%	41
Adding image processing	93.62%	90.89%	41
MSAM	98.74%	97.53%	40.1
DWconv	91.39%	87.79%	46.8
Improved YOLOv5	97.35%	96.27%	44

The ablation experiments show that image processing significantly increases the detection accuracy of foreign objects on belt conveyors, producing a more well-defined outline and resulting in an increase in the P and R indicators of 1.05% and 2.48%, respectively. The model architecture is left unaltered, and the FPS is unaffected. In addition, the addition of the MSAM to YOLOv5 improves the model's ability to extract features and gives it more protection against disruptive noise brought on by redundant features. It became the most successful ablation experiment group when the P and R indices increased by 6.17% and 9.12%, respectively, in comparison to the initial YOLOv5. Nevertheless, adding this module also leads to an increase in computational complexity, which makes a slowdown in recognition speed inevitable. Further, DWconv, a lightweight convolutional neural network, can produce results that are nearly on par with those of conventional convolution networks while using fewer calculations and parameters. In YOLOv5, the FPS is increased to 46.8 frames/s when convolution networks are replaced with DWconv, with just a slight decrease in P and R of 1.18% and 0.62%, respectively. This is because, compared to traditional convolutions, DWConv has a significantly reduced number of parameters. This may result in a slightly reduced ability to extract features from the model, but, at the same time, it brings about a significant increase in computational efficiency. Given the trade-off between accuracy and efficiency, this kind of accuracy loss is acceptable. By incorporating the MSAM and DWconv modules, the enhanced YOLOv5 model has achieved remarkable advancements in both accuracy and recognition speed. In the conducted experiment, the model demonstrated outstanding performance by attaining an impressive accuracy (P) of 97.35% and a recall rate (R) of 96.27%. Additionally, the model exhibited a recognition speed of 0.022 s, which is equivalent to processing 44 frames per second. These findings provide compelling evidence for the high performance and practicality of our proposed model in object detection tasks. Our analysis reveals a significant improvement in the precision (P) and recall rate (R) indicators, which increased by 4.78% and 7.86%, respectively, when compared to the initial YOLOv5 model.

5. Conclusions

This study aimed to tackle the difficulties of foreign object detection on belt conveyors in coal mines, which are exacerbated by complex background conditions, diverse object classes, unclear images, and the inability to station high-performance servers underground. To this end, we have proposed an improved YOLOv5 model with the aim of enhancing the object detection algorithm. Our model's effectiveness was confirmed by experimental findings, which showed a notable increase in accuracy and recognition speed. The following findings were deduced from this study:

(1) Comparative experiments have revealed that the incorporation of the multi-scale attention module (MSAM) within the YOLOv5 model can significantly enhance its feature extraction ability and resistance to redundant features. The improved model demonstrated an increase in precision (P) and recall rate (R) by 1.94% and 4.2%, respectively, compared to the initial YOLOv5 model. This underscores the crucial role of attention mechanisms in enhancing the performance of object detection tasks.

(2) Introducing lightweight convolutional neural networks (DWconv) to enhance computational efficiency has resulted in a significant increase in the model's frame rate (FPS) to 46.8 frames/s, while only marginally reducing the P and R indicators. These findings indicate that a balance has been achieved between accuracy and efficiency. Although the accuracy loss is acceptable, we posit that the model's recognition speed has improved significantly, thereby enhancing real-time performance in practical applications.

(3) The efficacy of the model improvement was verified through ablation experiments. The results displayed that image processing techniques significantly improved the detection accuracy of foreign objects on conveyor belts, while the introduction of the MSAM had a more notable impact on model performance, increasing the precision (P) and recall rate (R) metrics by 4.78% and 7.86%, respectively, compared to the initial YOLOv5.

(4) By integrating the MSAM and DWconv, the improved YOLOv5 model achieved significant improvements in accuracy and recognition speed. In the experiment, our model achieved a precision (P) of 97.35% and a recall rate (R) of 96.27%, with a recognition speed of 0.022 s, equivalent to a processing capacity of 44 frames per second. These results demonstrate that our proposed model has high performance and practicality in object detection tasks.

Author Contributions: Conceptualization, K.H.; Data curation, F.C.; Formal analysis, S.L. and R.Z.; Funding acquisition, K.H.; Investigation, F.C.; Methodology, S.L.; Validation, R.Z.; Writing—original draft, K.H. and S.L.; Writing—review and editing, K.H., S.L., F.C. and R.Z. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data set of this study comes from the video surveillance during coal mine production. Due to the security and privacy of coal mine production, this study cannot disclose the data set.

Conflicts of Interest: The authors declare that there are no conflict of interest.

References

1.　Lv, Z.Q. Research on Image Recognition of Foreign Bodies in the Process of Coal Mine Belt Transportation in Complex Environment. Master's Thesis, China University of Mining and Technology, Xuzhou, China, 2020; pp. 1–60.

2.　Wang, C.; Hao, B.N.; Zhang, L.Y.; Yang, Z.F. Conveyor belt foreign object detection technology based on RetinaNet in underground coal mine. *Coal Mine Mach.* **2022**, *43*, 180–183. [CrossRef]

3. Yang, R.; Qiao, T.; Pang, Y.; Yang, Y.; Zhang, H.; Yan, G. Infrared spectrum analysis method for detection and early warning of longitudinal tear of mine conveyor belt. *Measurement* **2020**, *165*, 107856. [CrossRef]

4. Yang, Y.; Hou, C.; Qiao, T.; Zhang, H.; Ma, L. Longitudinal tear early-warning method for conveyor belt based on infrared vision. *Measurement* **2019**, *147*, 106817. [CrossRef]

5. Chen, K.Q.; Zhu, Z.L.; Deng, X.M.; Ma, C.X.; Wang, H.A. Deep learning for multi-scale object detection: A survey. *J. Softw.* **2021**, *32*, 1201–1227.

6. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In *Computer Vision—ECCV 2016, Proceedings of the 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.

7. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 779–788.

8. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: New York, NY, USA, 2017; pp. 7263–7271.

9. Redmon, J.; Farhadi, A. YOLOv3: Incremental improvement. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 89–95.

10. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.

11. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; IEEE: New York, NY, USA, 2014; pp. 580–587.

12. Girshick, R. Fast R-CNN. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; IEEE: New York, NY, USA, 2015; pp. 1440–1448.

13. 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]

14. Wang, L.; He, M.; Xu, S.; Yuan, T.; Zhao, T.; Liu, J. Garbage classification and detection based on YOLOv5s network. *Packag. Eng.* **2021**, *42*, 50–56.

15. Yan, B.; Fan, P.; Lei, X.; Liu, Z.; Yang, F. A real-time apple targets detection method for picking robot based on improved YOLOv5. *Remote Sens.* **2021**, *13*, 1619. [CrossRef]

16. Mohiyuddin, A.; Basharat, A.; Ghani, U.; Peter, V.; Abbas, S.; Bin Naeem, O.; Rizwan, M. Breast tumor detection and classification in mammogram images using modified YOLOv5 network. *Comput. Math. Methods Med.* **2022**, *2022*, 1359019. [CrossRef]

17. Zhang, S.R.; Huang, Z.L.; Zhang, Y.H.; Zhang, A.; Ji, L. Research on coal and gangue recognition based on improved YOLOv5. *J. Mine Autom.* **2022**, *48*, 39–44.

18. Hao, S.; Zhang, X.; Ma, X.; Sun, S.Y.; Wen, H.; Wang, J.; Bai, Q. Foreign object detection in coal mine conveyor belt based on CBAM-YOLOv5. *J. China Coal Soc.* **2022**, *47*, 4147–4156.

19. Xiao, D.; Kang, Z.; Yu, H.; Wan, L. Research on belt foreign body detection method based on deep learning. *Trans. Inst. Meas. Control* **2022**, *44*, 2919–2927. [CrossRef]

20. Mi, C.; Huang, S.; Zhang, Y.; Zhang, Z.; Postolache, O. Design and implementation of 3-D measurement method for container handling target. *J. Mar. Sci. Eng.* **2022**, *10*, 1961. [CrossRef]

21. Lu, S.; Ding, Y.; Liu, M.; Yin, Z.; Yin, L.; Zheng, W. Multiscale feature extraction and fusion of image and text in VQA. *Int. J. Comput. Intell. Syst.* **2023**, *16*, 54. [CrossRef]

22. Chen, P.; Liu, H.; Xin, R.; Carval, T.; Zhao, J.; Xia, Y.; Zhao, Z. Effectively detecting operational anomalies in large-scale IoT data infrastructures by using a GAN-based predictive model. *Comput. J.* **2022**, *65*, 2909–2925. [CrossRef]

23. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.

24. Chen, J.; Wan, L.; Zhu, J.; Xu, G.; Deng, M. Multi-scale spatial and channel-wise attention for improving object detection in remote sensing imagery. *IEEE Geosci. Remote. Sens. Lett.* **2020**, *17*, 681–685. [CrossRef]

25. Deng, Y.; Du, S.; Wang, D.; Shao, Y.; Huang, D. A calibration-based hybrid transfer learning framework for RUL prediction of rolling bearing across different machines. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3511015. [CrossRef]

26. Shi, J.; Zhao, B.; He, T.; Tu, L.; Lu, X.; Xu, H. Tribology and dynamic characteristics of textured journal-thrust coupled bearing considering thermal and pressure coupled effects. *Tribol. Int.* **2023**, *180*, 108292. [CrossRef]

27. Singh, S.K.; Srivastava, R. SL-Net: Self-learning and mutual attention-based distinguished window for RGBD complex salient object detection. *Neural Comput. Appl.* **2022**, *35*, 595–609. [CrossRef]

28. Ullah, I.; Jian, M.; Shaheed, K.; Hussain, S.; Ma, Y.; Xu, L.; Muhammad, K. AWANet: Attentive-aware wide-kernels asymmetrical network with blended contour information for salient object detection. *Sensors* **2022**, *22*, 9667. [CrossRef]

29. Wu, X.L.; Xi, B.Q.; Wang, Z.Z.; Meng, F.H.; Zhen, R.; Li, S.K. Ground penetrating radar denoising method based on IRPCA and image enhancement. *Radio Eng.* **2023**, *53*, 899–908.

30. Mao, Q.H.; Li, S.K.; Hu, X.; Xue, X.S.; Yao, L.J. Foreign object recognition of belt conveyor in coal mine based on improved YOLOv7. *J. Mine Autom.* **2022**, *48*, 26–32.

31. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.

32. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

33. Zong, C.; Wan, Z. Container ship cell guide accuracy check technology based on improved 3D point cloud instance segmentation. *Brodogradnja* **2022**, *73*, 23–35. [CrossRef]

34. Deng, Y.; Lv, J.; Huang, D.; Du, S. Combining the theoretical bound and deep adversarial network for machinery open-set diagnosis transfer. *Neurocomputing* **2023**, *548*, 126391. [CrossRef]