

Article

Quantitative Monitoring Method for Conveyor Belt Deviation Status Based on Attention Guidance

Xi Zhang ¹, Zihao Yang ¹, Mengchao Zhang ^{1,*} , Yan Yu ¹, Manshan Zhou ²  and Yuan Zhang ^{1,2}

¹ College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao 266590, China; 13869872704@163.com (X.Z.); yzh1799@163.com (Z.Y.); zhangyuanzms@126.com (Y.Z.)

² Libo Heavy Industries Science and Technology Co., Ltd., Taian 271000, China

* Correspondence: zhangmc1995@sdust.edu.cn

Abstract: The efficient monitoring of the belt deviation state will help to reduce unnecessary abnormal wear and the risk of belt tear. This paper proposes a coupling characterization method involving the prediction box features of the target detection network and the linear features of the conveyor belt edge to achieve the quantitative monitoring of conveyor belt deviations. The impacts of the type, location, and number of attention mechanisms on the detection effect are fully discussed. Compared with traditional image-processing-based methods, the proposed method is more efficient, eliminating the tedious process of threshold setting and improving the detection efficiency. In detail, the improved practice and tests are carried out based on the Yolov5 network, and the Grad-CAM technique is also used to explore the effect of attention mechanisms in improving the detection accuracy. The experiments show that the detection accuracy of the proposed method can reach 99%, with a detection speed of 67.7 FPS on a self-made dataset. It is also proven to have a good anti-interference ability and can effectively resist the influence of the conveying material flow, lighting conditions, and other factors on the detection accuracy. This research is of great significance in improving the intelligent operation and maintenance level of belt conveyors and ensuring their safe operation.

Keywords: belt conveyor; supervised machine learning; intelligent digital monitoring; deviation detection; feature visualization



Citation: Zhang, X.; Yang, Z.; Zhang, M.; Yu, Y.; Zhou, M.; Zhang, Y.

Quantitative Monitoring Method for Conveyor Belt Deviation Status Based on Attention Guidance. *Appl. Sci.* **2024**, *14*, 6916. <https://doi.org/10.3390/app14166916>

Academic Editors: Maria Pia Morigi and Marco Seracini

Received: 27 June 2024

Revised: 30 July 2024

Accepted: 2 August 2024

Published: 7 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/>).

1. Introduction

A belt conveyor is a friction transmission-based machine that transports materials in a continuous manner. It is the preferred equipment for the continuous conveying of bulk materials and is widely used in mines, ports, and wharves; for cement; and in agriculture, the chemical industry, and other fields [1]. In recent years, under the demands of Industry 4.0 [2,3] and “energy conservation and emission reduction” [4], belt conveyors have been developed for use over long distances, at high belt speeds, on a large scale, in intelligence [5,6], and in ways that save energy [7,8]. The research and development of reliable equipment and technology that efficiently monitors operating conditions have gradually become an important component to ensure the reliable operation of belt conveyors.

Belt deviation, as one of the most common operating failures of belt conveyors, is closely related to the processing quality and installation accuracy of idlers and rollers, the balance of material loading, and the quality of conveyor belt joints [9,10]. Deviation will lead to the abnormal wear of the belt edge, which may cause problems such as material spillage and the increased energy consumption of the conveyor [11]. Serious deviation will directly lead to tearing accidents [12], trigger safety accidents, and affect the normal production of enterprises [13]. Therefore, it is of great practical significance to effectively monitor the belt deviation state.

Currently, most of the applied deviation monitoring methods rely on a deviation sensor [13]. Essentially, this is a multi-stage travel switch that outputs a switch signal. In recent years, non-contact measurement methods based on machine vision have also been designed. Specific implementation methods include the traditional image-processing-based straight-line detection method, deep learning-based deviation state classification, target segmentation-based conveyor belt region positioning, and the decision method based on idler number identification and comparison. The traditional image-processing-based straight-line detection method is mainly represented by Canny edge detection and Hough line detection [14–16]. However, it requires a complex image processing procedure due to the influence of the complex working environment and conveying material flow in practical applications, including the acquisition of the region of interest, grayscale transformation, edge detection, connected domain analysis, morphological processing, Hough line detection, etc. [15]. During this period, a large number of threshold determinations are involved, and the overall practicability is poor, especially when the camera is tilted: the larger the field of view, the more complex the processing operations required to avoid the interference of other targets. Then, the classification of the deviation state based on deep learning [17] and the judgment method based on the identification of the number of idlers beside the belt can only be used to realize the qualitative measurement of the deviation state of the belt. This cannot meet the needs of quantitative measurement and is not conducive to determining the deviation rules. The location of the conveyor belt area based on target segmentation [18,19] can easily separate the conveyor belt area from the background of the field of view, but target segmentation is a pixel-level operation. It is difficult to create a dataset in the pre-production stage, and it is also difficult to achieve quantitative measurement. In comparison, it is more feasible to obtain the belt edge line directly through the segmentation method, but if the judgment of the deviation state is to be realized, it is also necessary to further fit the obtained segmentation results.

In view of the above problems, considering the straight-line feature of the belt edge, in this work, we propose a quantitative monitoring method for the deviation state of the conveyor belt based on the intrinsic features of the prediction box. A general target detection network is used first to detect the conveyor belt edge line and obtain its line equation, which is further used to achieve the quantitative monitoring of the deviation state. Then, the Grad-CAM technique [20] is used to discuss and analyze the feature position that the general target detection network uses to detect straight lines in order to ensure a higher-precision result.

The remainder of the paper is organized as follows. Section 2 presents the proposed methodology, including the relationship between the prediction box and the belt edge line and the structure of the used detection network and its improvements. Section 3 introduces the dataset and hardware usage. The results of a case study and the related discussion are presented in Section 4. The conclusions and future work are summarized in Section 5.

2. Methodology

2.1. Relationship between the Prediction Box and the Belt Edge Line

It is well known that the prediction result of a generic target detection network, including the RCNN [21], SSD [22,23], and Yolo [24,25] series, is a rectangular box that is used to frame and position the potential targets in the image and predict the category to which the target belongs. The complete prediction process is shown in Figure 1. The generic target detection network usually uses a backbone feature extraction network to extract features from the input image, uses upsampling and feature splicing techniques to enrich and enhance the extracted features (Neck), and, finally, uses a prediction network to predict the target class and regress the target position [25]. The prediction box is usually determined by four values: the relative center point coordinates of the prediction box (both horizontal and vertical) and the relative width and height of the prediction box.

However, it is clear that it is difficult to characterize the straight line of a conveyor belt edge in the form of a prediction box. Therefore, in light of the actual camera arrangement

and how the conveyor belt is presented in the camera's field of view, this paper proposes methods for the detection of straight belt edges suitable for vertical and inclined shots, respectively. In particular, for vertical camera shots, we propose the use of the longitudinal edges of the prediction box (rectangular box) to characterize the belt edges, with two representative edges, such as the left or the right sides of the prediction box, as shown in Figure 2a. In contrast, for oblique camera shots, we propose a straight-line detection method for belt edges based on the diagonal features of the prediction box, aiming to use the upper right and lower left diagonal lines of the prediction box to characterize the left edge and aiming to use the upper left and lower right diagonal lines to characterize the right edge, as shown in Figure 2b. In general, the essence of the deviation state monitoring task is to effectively identify and locate the two straight edge lines of the conveyor belt under a complex environment and, on this basis, to obtain its spatial state equation for the further judgment of the deviation state.

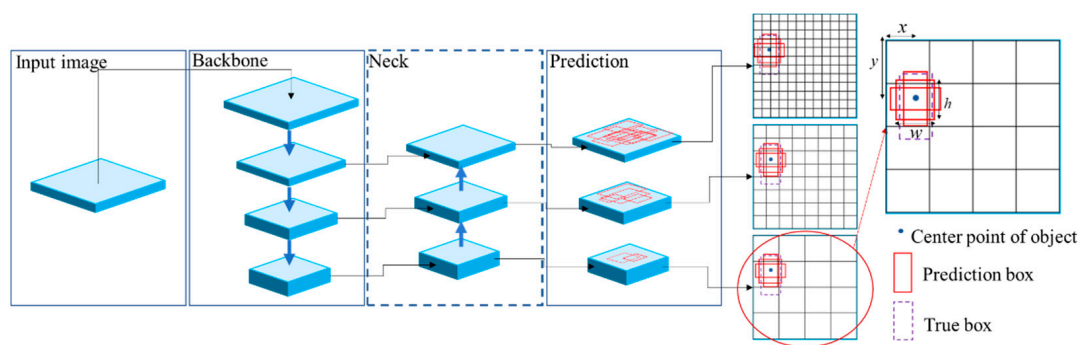


Figure 1. Prediction process for a general target detection network.

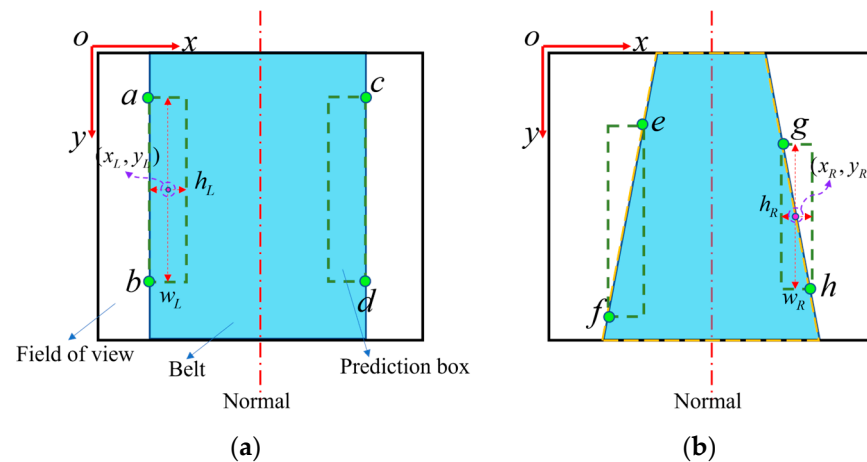


Figure 2. Principle of conveyor belt edge detection for different shooting situations: (a) vertical shooting, (b) oblique shooting.

For a vertical camera shot, the belt edge presented will be parallel to the camera's field of view, as shown in Figure 2a. If we use the detection of the left edge as an example and assume that the relative center point coordinates of the prediction box are (x_L, y_L) , and the relative width and height are w_L and h_L , respectively, then, for Figure 2a, the coordinates of point a , (x_a, y_a) , can be calculated using Equation (1), and the coordinates of point b , (x_b, y_b) , can be calculated using Equation (2). It should be noted that the horizontal coordinates of point a and point b are equal. At this time, the straight-line equation of the belt left edge can be expressed as Equation (3):

$$x_a = x_L - \frac{w_L}{2} \quad y_a = y_L - \frac{h_L}{2} \quad (1)$$

$$x_b = x_L - \frac{w_L}{2} \quad y_b = y_L + \frac{h_L}{2} \quad (2)$$

$$x = x_L - \frac{w_L}{2} \quad (3)$$

Then, whether the belt deviates can be judged by comparing the lateral offset of the center point of the prediction box.

However, since the oblique shooting method can achieve a larger monitoring range, it has wider application in practice, as shown in Figure 2b. Similar to vertical shooting, one should use the detection of the right belt edge as an example, assuming that the relative center point coordinates of the prediction box are (x_R, y_R) and the predicted width and height are w_R and h_R , respectively. Then, for Figure 2b, the coordinates of point g , (x_g, y_g) , can be calculated using Equation (4), and the coordinates of point h , (x_h, y_h) , can be calculated using Equation (5). The straight-line equation of the right belt edge can be expressed as Equation (6):

$$x_g = x_R - \frac{w_R}{2} \quad y_g = y_R - \frac{h_R}{2} \quad (4)$$

$$x_h = x_R + \frac{w_R}{2} \quad y_h = y_R + \frac{h_R}{2} \quad (5)$$

$$\frac{y - y_g}{y_h - y_g} = \frac{x - x_g}{x_h - x_g} \quad (6)$$

In the same way, if we assume that the coordinates of points e and f are (x_e, y_e) and (x_f, y_f) , the linear equation of the left belt edge in the oblique shooting mode can be expressed as follows:

$$\frac{y - y_e}{y_f - y_e} = \frac{x - x_e}{x_f - x_e} \quad (7)$$

In contrast to vertical shooting, certain auxiliary and reference lines are required to determine the deviation state of a conveyor belt in a tilted state, and the rules for determining the deviation state proposed in this paper can be found in Figure 3.

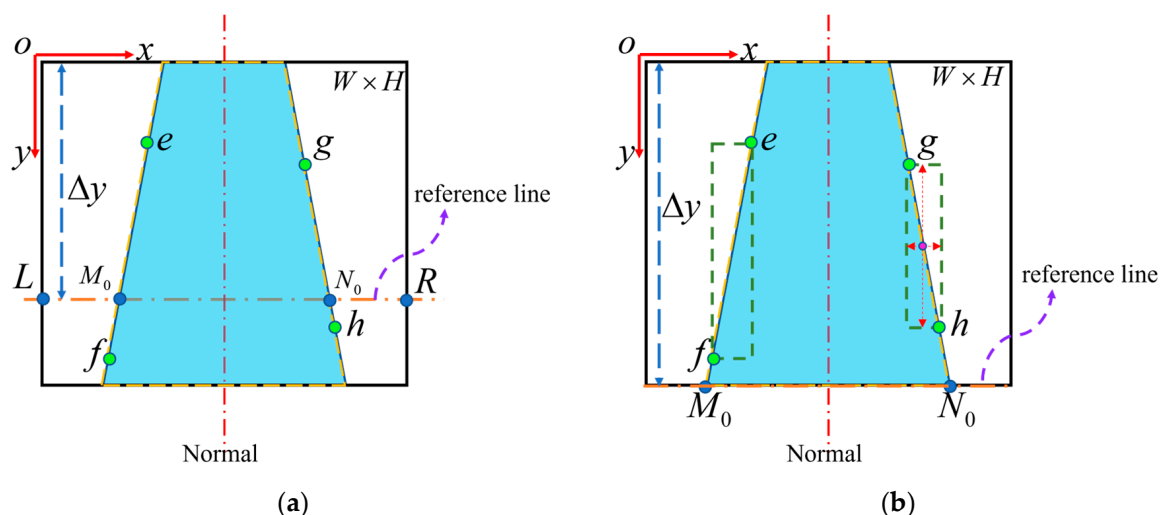


Figure 3. Schematic diagram of the calculation method for deviation: (a) general calculation rules, (b) method used in this paper.

One should assume that there exists a horizontal virtual reference line at distance Δy from the x -axis that intersects the left boundary of the field of view at point L , intersects the right boundary of the field of view at point R , intersects the left belt edge at point M_0 , and intersects the right belt edge at point R . Then, the distance between the conveyor belt

and the edges on both sides of the field of view can be represented by D_{LM_0} and D_{RN_0} , respectively. The calculation method is shown in Equations (8) and (9):

$$D_{LM_0} = \frac{(\Delta y - y_e)(x_f - x_e)}{(y_f - y_e)} + x_e \quad (8)$$

$$D_{RN_0} = W - \frac{(\Delta y - y_g)(x_h - x_g)}{(y_h - y_g)} - x_g \quad (9)$$

Further, the judgment of the deviation state of the belt can be realized by setting the deviation threshold τ , and the judgment rule can be found in Equation (10). The proposed linear equation-based deviation estimation method is suitable for both the belt overall deviation and belt inclination, as it takes the belt slope into account.

$$if \begin{cases} |D_{LM_0} - D_{RN_0}| > \tau & \text{deviation} \\ |D_{LM_0} - D_{RN_0}| \leq \tau & \text{normal} \end{cases} \quad (10)$$

In this paper, the intersection of the belt edge's straight line and the camera's field of view is directly used as a reference, as shown in Figure 3b, which can also be understood as $\Delta y = H$ in Equations (8)–(10).

As for the setting of deviation threshold τ , we recommend the following two methods. (1) First, researchers should manually screen the conveyor belt deviation images in the dataset and then use the method proposed in this article to identify the conveyor belt edge in the screened image, calculate the deviation amount, set the calculation result as the threshold, and use this for subsequent automatic detection. (2) Researchers should randomly select an image in the dataset and then use the drawing tool to draw an inclined straight line parallel to the left/right edge. When assuming that the straight line is the left/right boundary of the conveyor belt, the drawing tool can be used to measure the distance (pixels) between the drawn straight line and the actual edge line of the conveyor belt on the horizontal virtual reference line. Then, the threshold can be calculated.

2.2. Network Structure

In order to make the selected network relatively suitable for edge computing devices, the YOLOv5 network [26] is selected for subsequent algorithm improvement and testing based on previous experience with foreign object detection networks and taking into account the number of parameters, computational power, and model size [27]. The structure of the YOLOv5 network is clearly shown in Figure 4.

In combination with that shown in Figure 1, CSPDarkNet is used as the backbone. The cross-stage partial (CSP) module is used repeatedly to deepen the network depth, thus enhancing the feature extraction ability. It contains two structures, namely CSP 1-X and CSP 2-X, which are shown in Figure 4b. The “X” indicates that the structure in the dotted box is repeated X times. In detail, in the CSP 1-X module, the network inside the dotted box contains the residual structure, which effectively increases the gradient value of backpropagation between each convolutional layer, thus reducing the possibility of gradient disappearance caused by network deepening.

It should be mentioned that the YOLOv5 network can achieve different complexity levels by adjusting the value of X. Specifically, for networks of YOLOv5 s-m-l-x with different complexity, the values of X in the network can be found in Table 1. To explain the network more clearly, we use B1–B3 and N1–N5 in Figure 4a and Table 1 to indicate the locations of the corresponding CSP modules, in which B1–B3 represent positions 1–3 in the backbone part, and N1–N5 represent positions 1–5 in the neck part.

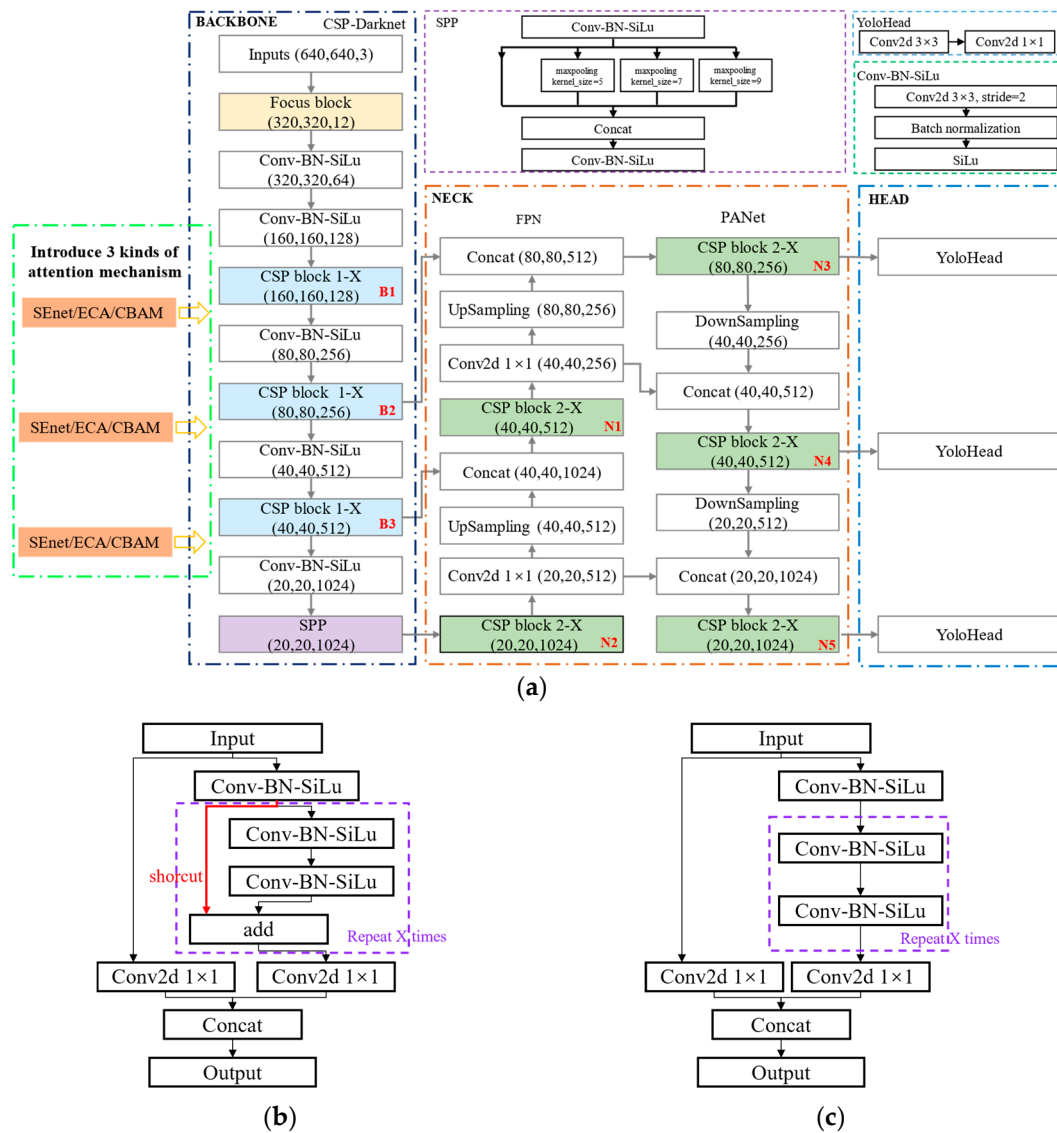


Figure 4. The detection network structure used in this paper: (a) main body, (b) CSP 1-X, (c) CSP 2-X.

Table 1. The statistics of the number of X in the CSP block.

Model \ Position	B1	B2	B3	N1	N2	N3	N4	N5
Yolov5-s	1	3	3	1	1	1	1	1
Yolov5-m	2	6	6	2	2	2	2	2
Yolov5-l	3	9	9	3	3	3	3	3
Yolov5-x	4	12	12	4	4	4	4	4

As for the feature enhancement part, the feature pyramid network (FPN) and path aggregation network (PANet) [28] are used to enhance the obtained features. Through up-sampling and feature layer splicing, target semantic information and location information at different levels can be fully fused or compensated, which has positive significance in improving the accuracy of detection.

In the head part, no changes are made to the prediction head; it is still the same and uses the original Yolov5, which is used to predict the left or right belt edge region. Its prediction results contain the confidence of the predicted targets, the center point coordinates, the width and height information of the prediction box, etc.

In addition, considering that the conveyor belt edge straight-line detection method proposed in this paper relies heavily on the detection results for the belt edge region, we incorporate three attention mechanism models—the squeeze-and-excitation network (SENet) [29], efficient channel attention (ECA) [30], and the convolutional block attention module (CBAM) [31]—into the backbone feature extraction network. The added positions are shown in Figure 4, and the network structure of the three attention mechanisms is shown in Figure 5.

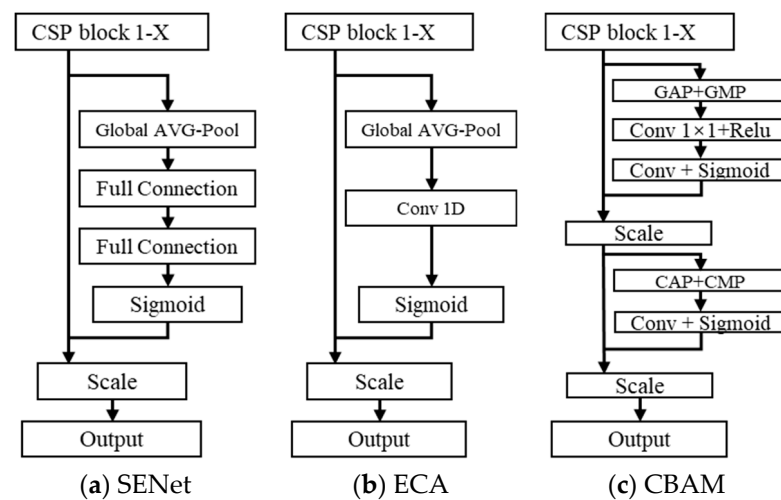


Figure 5. Network structure of the attention mechanism module: (a) SENet, (b) ECA, (c) CBAM.

It should be mentioned that each attention mechanism is used separately. As shown in Figure 4, there are three places where attention is added, but a corresponding attention is not added to each place.

In Figure 5, Global AVG-Pool (GAP) represents global average pooling, GMP (Global Max-Pool) represents global max pooling, CAP (Channel AVG-Pool) represents the average pooling per channel, and CMP (Channel Max-Pool) represents the max pooling per channel. Through pooling and full connection, the weight of the space occupied by each channel or location can be obtained, i.e., the influence of this channel or location on feature extraction. A large weight means that the channel or location has a greater influence on feature extraction and, therefore, a greater impact on the final output, while a small weight means that the channel or location has a smaller impact on the final output. Then, the weight distribution can be added to the original input feature information through scaling.

The SENet and ECA models are typical channel attention mechanisms. The ECA model uses convolution operations to replace the two fully connected layers in SENet to achieve smaller parameters and calculations and a faster speed. At the same time, the CBAM module integrates spatial attention and channel attention, enabling the two to complement each other and making it easier to achieve the effective prediction of the weight distribution. In this work, we fully discuss the effect of the position and number of the three attention mechanisms in improving the detection accuracy.

3. Hardware and Data Usage

As for a supervised learning-based target detection network, the data, computing power, and algorithms simultaneously determine the final performance of the network [32].

Due to the particularity of the monitoring targets, in this study, we constructed a new conveyor belt deviation dataset, based on the on-site collection with 5000 images, and complete the data labeling using the labelImg software. All images that we used originated from video surveillance data in three scenarios, namely the Material Handling and Control Research Institute in Shandong University of Science and Technology, Qingdao Port, and Ma'anshan Power Plant. All data were obtained using a camera under tilted shooting

conditions, but the parameters of the camera and lens, the parameters of the conveyor, etc., are unknown.

Since the diagonal feature is used to detect the straight line of the belt edge, in the labeling work, it is also strictly required that the diagonal line of the marked rectangular box coincides with the straight line of the belt edge, as shown in Figure 6. More sample data can be found in our public datasets at <https://github.com/zhangzhangzhang1618/images-for-belt-deviation-detection/tree/main>, accessed on 1 August 2024.

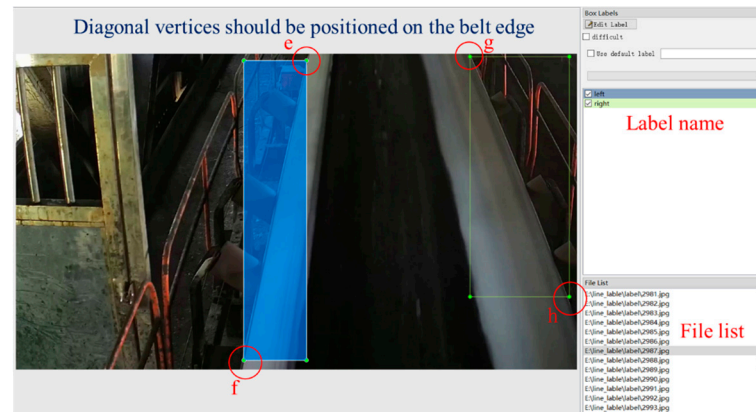


Figure 6. Data annotation.

Based on the completed network improvements and data construction, we conducted network training and testing of the corresponding improved algorithms on the hardware platform shown in Table 2. During the network training, the batch size was set as eight and the total number of training steps was set as 100 epochs. The CutMix and mosaic data enhancement techniques [25] were used during the training process to enrich the background information of the target, thus helping to improve the detection accuracy.

Table 2. Algorithm operating platform.

OS	CPU	GPU	Pytorch	Python
Windows 10	E5-2620v3*2	RTX2060-6G	1.7.0	3.6.13

* = means two CPUs.

4. Results and Discussion

4.1. Evaluation Metrics of Network Performance

In this paper, the loss changes during network training, the mean average precision (mAP), the number of images that the algorithm can process per second (FPS), and the number of parameters and calculations are used as metrics to evaluate the improved Yolov5-s-m-l-x network. Among them, the changes in the loss and average precision are shown in Figure 7, and the statistical results of the parameters, calculations, and processing speed of the algorithm are shown in Table 3.

It can be seen that the Yolov5-s-m-l-x network model with only the prediction network improvement shows good convergence performance without overfitting, as the changes in the training loss and validation loss during the training process are close. The accuracy evaluation index mAP0.5 reached a value of about 0.99 after about six epochs of training and stabilized with training, while the mAP0.5:0.95 reached accuracy of about 0.88 after about 30 epochs of training and grew slowly, to about 0.9, with training.

The high accuracy of the prediction results can be attributed to the specificity of the linear detection method in this paper, i.e., the specificity of the detection target objects and the specificity of the scene: (1) there are only two detection target objects, and their shapes and sizes are relatively fixed; (2) the backgrounds of the targets are relatively singular and they do not involve complex and variable backgrounds.

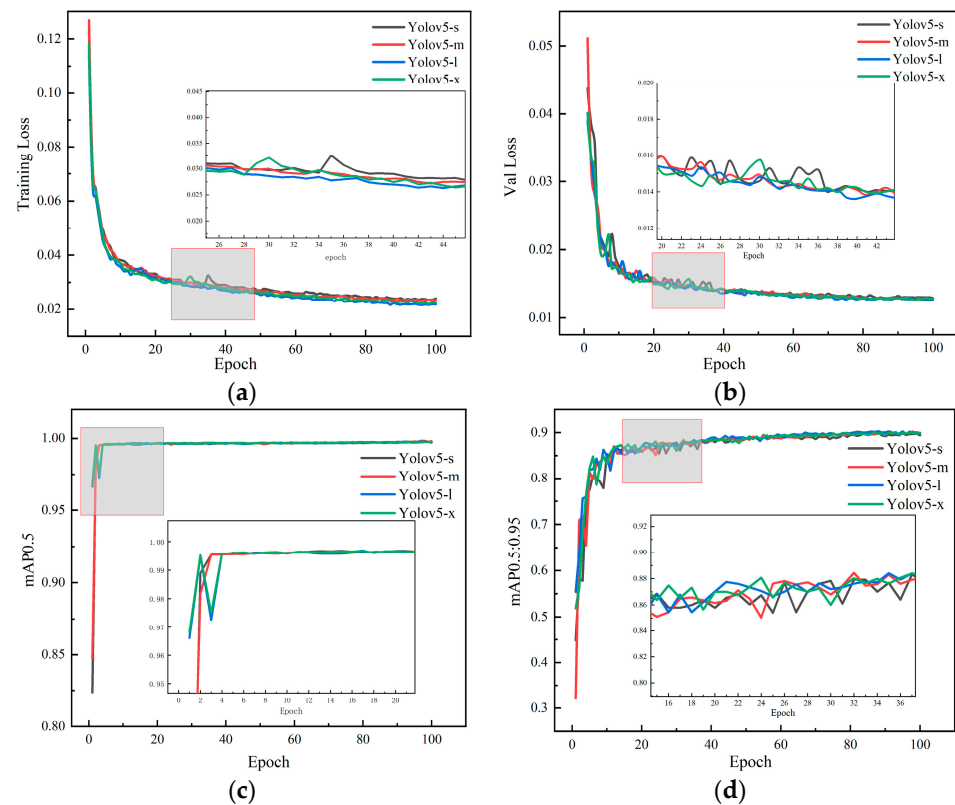


Figure 7. Loss and accuracy changes during network training: (a) loss change on the training set, (b) loss change on the validation set, (c) change in mAP0.5, (d) change in mAP0.5:0.95.

Table 3. Comparison of parameters, calculations, and speed of networks with different complexity.

Model	Parameters	Calculations/(GFLOPs)	Inference Speed/(FPS)
Yolov5-s	7,066,239	16.4	67.7
Yolov5-m	21,060,447	50.4	56
Yolov5-l	46,636,735	114.3	41.52
Yolov5-x	87,251,103	217.3	33.11

Table 3 shows the differences in the number of parameters, calculations, and inference speed of the Yolov5-s-m-l-x network model with only prediction network improvements. In particular, Yolov5-s achieves the highest prediction speed of 67 FPS with the smallest number of parameters and the smallest amount of computation, while the number of parameters and the amount of computation gradually increase as the number of layers and the complexity of the network increase, leading to a reduction in the prediction speed of Yolov5-x to around 33 FPS.

Combined with the prediction accuracy shown in Figure 7, it can be seen that the difference in prediction speed between the four Yolov5-s-m-l-x networks with only prediction network improvements is much greater than the difference in prediction accuracy. Therefore, to meet the needs of real-time detection, the corresponding algorithm improvements and results were implemented only for the Yolov5-s network in the subsequent research.

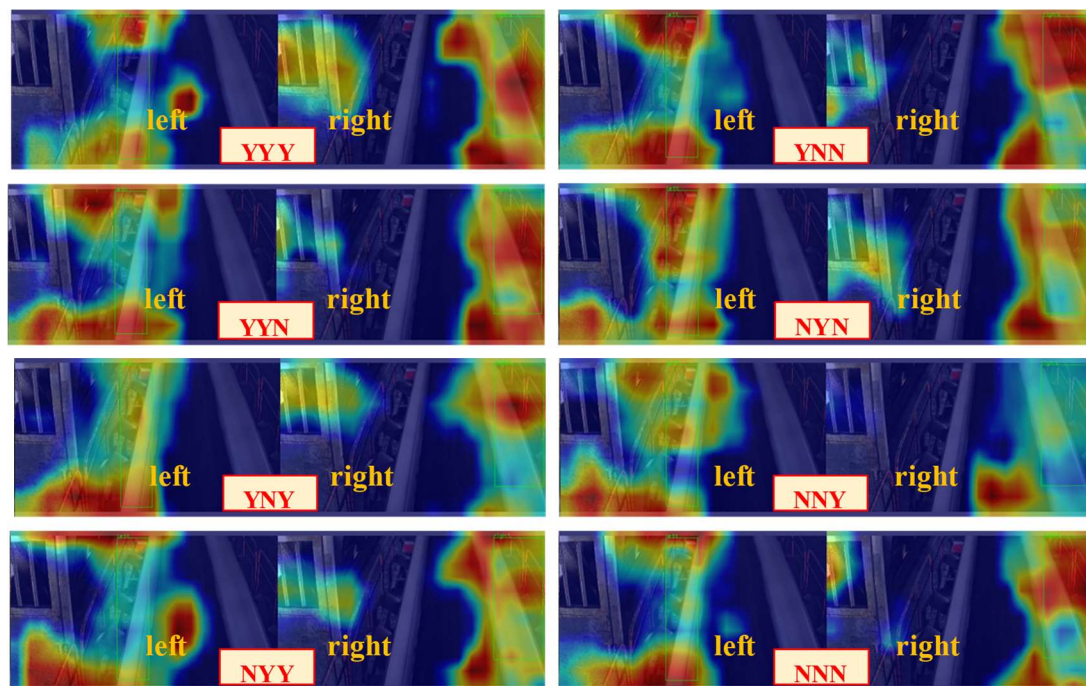
4.2. Ablation Experiment

The detection result of the belt edge line depends heavily on the detection result of the belt edge region. In order to ensure that the diagonal vertices of the prediction box always fall on the belt edge, we explored the effect of the attention mechanism in improving the detection accuracy. As mentioned earlier, we introduced three different types of attention mechanisms (SENet, ECA, and CBAM) at three locations in the backbone, as shown in Figure 4.

In this section, we explore in depth the effect of the type, number, and addition position of the attention mechanisms on the backbone feature extraction network through ablation experiments and the Grad-CAM technique, and the experimental results are shown in Table 4 and Figure 8.

Table 4. Effect of attention mechanism on network performance.

Type of Attention Mechanism	Position			Parameters	Calculations/(GFLOPs)	FPS	mAP0.5	mAP0.5:0.95
	B1	B2	B3					
SENet	Y	Y	Y	7,088,243	16.4	61.6	0.997	0.8982
	Y	Y	N	7,071,579	16.4	64.6	0.997	0.8973
	Y	N	Y	7,084,003	16.4	65	0.9969	0.8979
	N	Y	Y	7,087,143	16.4	64.8	0.9978	0.9004
	Y	N	N	7,067,339	16.4	64	0.9971	0.8997
	N	Y	N	7,070,479	16.4	66.7	0.9978	0.8992
	N	N	Y	7,082,903	16.4	62	0.9976	0.9014
ECA	Y	Y	Y	7,066,248	16.4	59.5	0.9976	0.9008
	Y	Y	N	7,066,245	16.4	65.1	0.9972	0.8984
	Y	N	Y	7,066,245	16.4	60.9	0.9971	0.8967
	N	Y	Y	7,066,245	16.4	64.4	0.998	0.9
	Y	N	N	7,066,242	16.4	62.9	0.9971	0.9009
	N	Y	N	7,066,242	16.4	61.3	0.9974	0.8998
	N	N	Y	7,066,242	16.4	66.7	0.9974	0.8984
CBAM	Y	Y	Y	7,221,370	19.2	46.7	0.9964	0.8885
	Y	Y	N	7,213,080	19.2	50.8	0.9965	0.8881
	Y	N	Y	7,219,224	19.2	48.35	0.9965	0.8905
	N	Y	Y	7,220,760	19.2	49.7	0.9961	0.8904
	Y	N	N	7,210,934	19.2	53.7	0.9966	0.8891
	N	Y	N	7,212,470	19.2	50.6	0.9969	0.89
	N	N	Y	7,218,614	19.2	51.4	0.9966	0.8924



(a) SENet

Figure 8. Cont.

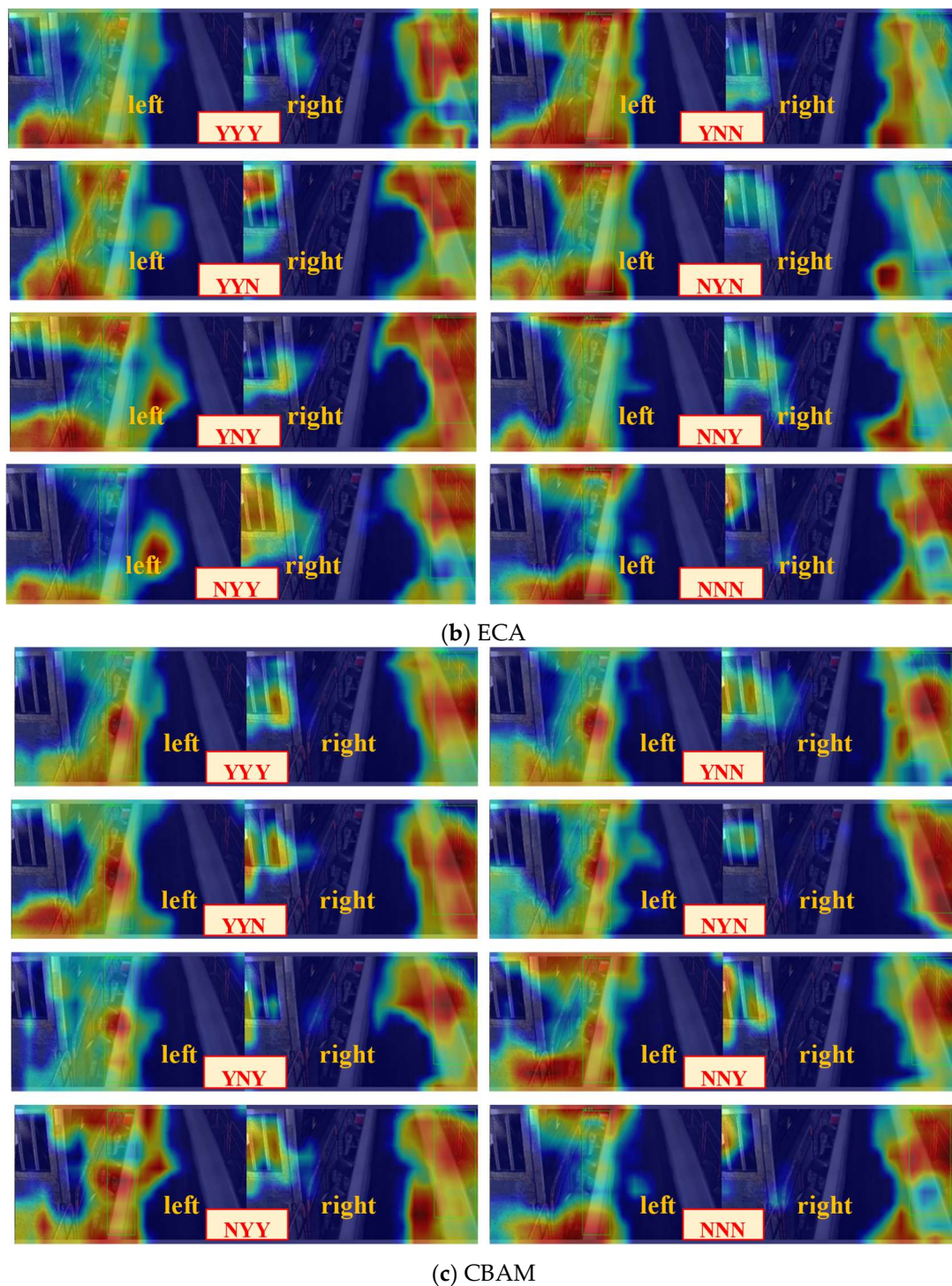


Figure 8. Visualization of the effect of the attention mechanism on network feature extraction.

A “Y” in Table 4 indicates that the appropriate attention mechanism has been added after the CSP block at that location, while an “N” indicates that the attention mechanism has not been added. Since there are three candidate positions (B1–B3) and three attention mechanisms (SENet, ECA, and CBAM), there are a total of 18 possibilities, as shown in Table 4.

Comparing the data in Tables 3 and 4, it can be seen that all Yolov5-s networks with the addition of the attention mechanism have a small increase in the number of parameters, among which the CBAM attention mechanism has the most significant increase, followed

by SENet and, finally, ECA. At the same time, in the comparison of the other performance parameters, CBAM has a more significant increase in the number of parameters, which also leads to a decrease in the inference speed, with the fastest prediction speed of only around 51 FPS. However, the prediction accuracy does not change significantly, and the amount of computation and the processing speed of the SENet and ECA attention mechanisms do not change significantly because the number of parameters does not change significantly.

Comparing the influence of the attention mechanisms at different numbers and locations, it can be seen that introducing three attention mechanisms into the backbone feature extraction network produces the largest processing speed loss compared to introducing one or two attention mechanisms; at the same time, separately introducing the attention mechanism at position 2 or position 3 has relatively little impact on the inference speed.

The results presented in Figure 8 visualize the effect of the attention mechanism using the Grad-CAM technique, with the red part of the figure indicating where the network's attention lies, i.e., where the network prefers to use the image features at that location for target detection. In the image, "YYY" indicates that the attention mechanism was added to B1, B2, and B3, while "NNN" indicates that no attention mechanism was added to backbone B1, B2, or B3.

The feature distribution that the network tends to use without adding the attention mechanism is shown in the image of "NNN". In this case, to detect the left belt edge, the feature information that the network tends to use is mostly concentrated at the top and bottom of the annotation box but seldom distributed in the interior of the annotation box. To detect the right belt edge, the feature information used is mainly distributed in the upper right corner of the annotation box. Some of it is also distributed in the window area in the upper left corner of the image, which means that when the left edge of the conveyor belt is in different backgrounds, especially when the top and bottom scenes of the left edge region of the conveyor belt in the image undergo obvious changes, the belt deviation monitoring network is very likely to fail, thus not succeeding in identifying the left edge. Comparatively speaking, environmental changes have little impact on the identification effect of the right edge.

Further, by observing the feature distribution after adding the attention mechanism, it can be seen that, in the improved network proposed in this paper, for the detection of the left belt edge, the SENet and ECA attention mechanisms do not produce obvious effects. The feature distribution of the attention paid by the network is not shifted significantly and is still concentrated at the top and bottom of the annotation box. Moreover, when the attention mechanism is added to some positions, the feature distribution is partially transferred to the material flow, as in the case of "YYY" and "NYY" in SENet and "YNY" and "NYY" in ECA. At this time, the change in the material flow distribution may immediately affect the conveyor belt edge detection result. The CBAM attention mechanism transfers more network attention to the interior of the annotation box, which is conducive to improving the robustness of the network. Similarly, for the detection of the right belt edge, the effect of the SENet and ECA attention mechanisms is not ideal, and some of them even deviate from the annotation box, such as "NNY" in SENet and "YNN, NYN, NNY" in ECA. Comparatively, CBAM shifts the center of gravity of the feature distribution from the upper right corner center of the annotation box to the inside of the annotation box, which will also improve the anti-interference ability of the network.

Based on the above visualization of the feature distribution, it can be seen that for the scene where the test image is located in Figure 8, during the actual installation and arrangement of the system, large changes should be avoided in the top and bottom areas within the left visual field so as not to affect the influence of the network on the left edge.

4.3. Verification of the Proposed Method

The above evaluation quantitatively demonstrated the performance of the deviation monitoring network under different improvement strategies from an objective level, proving that the improved network model described in this paper performs well on the dataset.

In this section, the effects of the proposed network are shown more intuitively from a subjective point of view. A part of the image data of the test set is assigned priority to be used in the generalization ability test of the network model (Yolov5-s). The respective influences of different conveying materials and their shapes, the image brightness and darkness, random noise, etc., on the detection results are discussed. The deep Hough transform (DHT) [33] linear detection algorithm is also used to compare the results.

(1) The influence of the material flow

Some types of conveying materials are similar to the conveyor belt in color, appearance, and form, especially a coal conveyor under the condition of low illumination. The presence of certain materials may interfere with the identification and detection results of the belt edge and may lead to the incorrect judgment of the deviation state.

Figure 9 compares the detection effects of the method proposed in this paper and the DHT algorithm when different materials are transported and where the deviation threshold τ is set as 80. Figure 9a is the original image used for network testing, collected from the power plant and the port. Figure 9b shows the DHT algorithm's detection effect on straight conveyor belt edge lines. It can be seen that the algorithm detects the belt edge and the boundary of the material flow and, in addition, the frame edge, indicating that the shape of the material flow and other potential straight lines indeed affect the belt edge detection to a certain extent. Figure 9c shows the detection results of the method proposed in this paper.

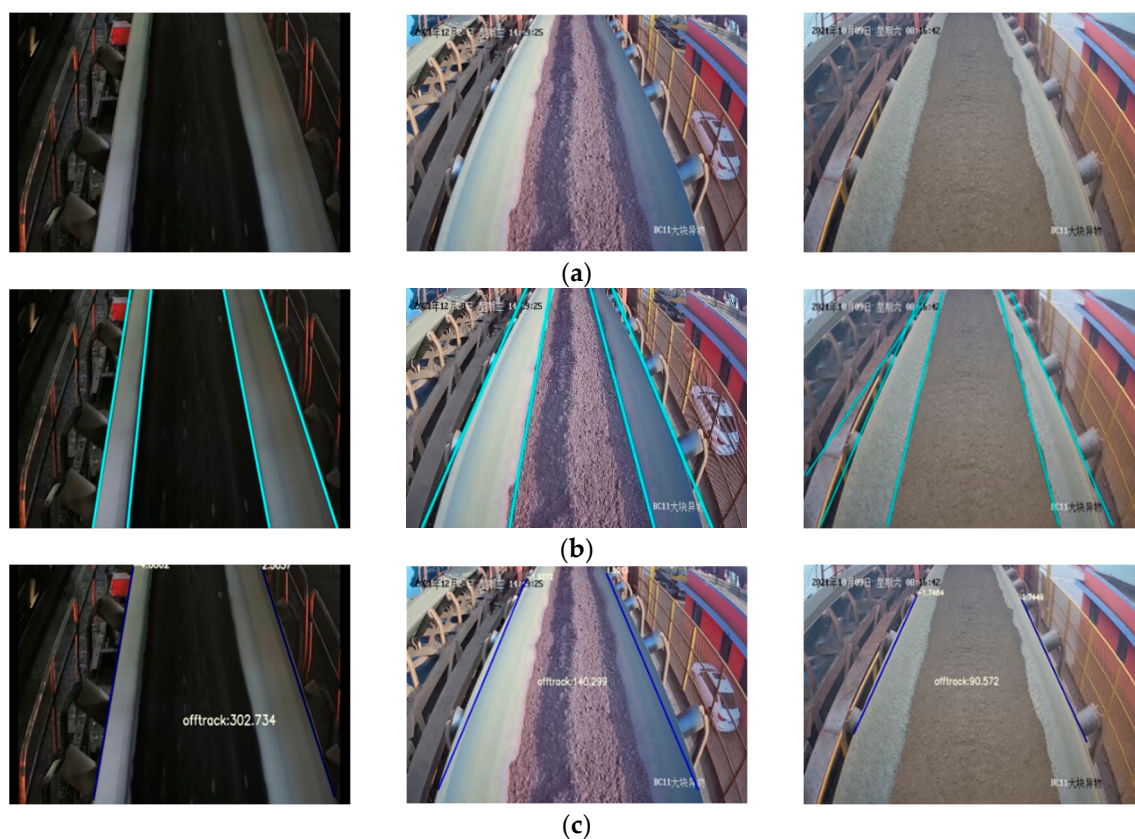


Figure 9. Influence of conveying material flow on conveyor belt edge detection: (a) original image, (b) detection result of DHT, (c) detection result of the method proposed in this paper.

Compared with the DHT algorithm, the proposed method can not only detect the left and right belt edges but also effectively avoids the influence of other potential lines in the field of vision on the detection results. Further, it provides the corresponding deviation amount and the judgment results of the deviation state. For example, “offtrack:302.734” means that the state of the conveyor belt is “off-track”, and the deviation amount is

302.734 pixels, which is calculated according to Equations (8) and (9). The deviation state is judged according to the comparison between the calculated amount of deviation and the threshold value in Equation (6). Similarly, “no_offtrack:42.466” means that the conveyor belt’s calculated deviation amount is 42.466 pixels, and the deviation state is determined to be “no_offtrack”. In addition, there are still two values in the image shown in Figure 9c (“−1.6195” and “1.5164”), which are the slope of the identified conveyor belt edge line. When the camera is installed in line with the width of the conveyor belt, this value can be directly used in the auxiliary judgment of the belt deviation state, especially the belt inclined state.

(2) The effect of the brightness and shade of the image

In the open environment, at different periods, due to uneven illumination, the images obtained by the visual perception system of the belt conveyor may have different brightness degrees. Especially in the case of low illumination, the difference between the material flow and the conveyor belt will be less obvious, which easily leads to the monitoring method’s failure. Therefore, in this section, we detail our simulation of conditions of low illumination. This was achieved by reducing the brightness of the image shown in Figure 9a to further verify the network’s generalization ability. The results are shown in Figure 10.

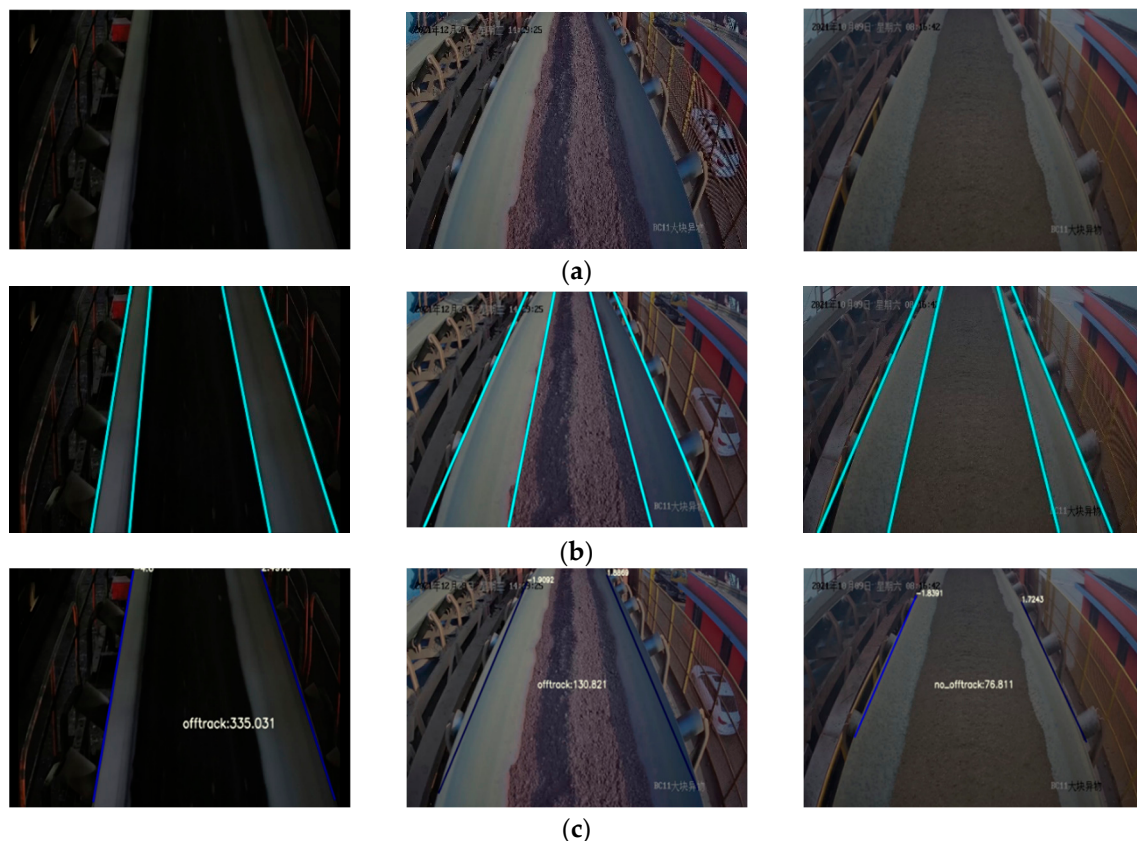


Figure 10. Influence of low illumination on conveyor belt edge detection: (a) image with low illumination, (b) detection result of DHT, (c) detection result of the method proposed in this paper.

Among them, Figure 10a shows the result of reducing the image brightness in Figure 9a, while Figure 10b shows the detection effect of the DHT algorithm. By comparing Figures 9b and 10b, it is found that the influence of the material flow on the detection results still exists under low illumination. The influence of other potential lines in the field of vision (such as the frame) on the detection results of the belt edge of the conveyor decreases gradually. Figure 10c shows the detection results of low-brightness images for the network model proposed in this paper. It can be seen that the method proposed in this paper is not

affected by the image brightness and can still generate a correct judgment on the state of the conveyor belt deviation, but the calculated amount of deviation is slightly different from the result shown in Figure 9c. This difference is caused by the different detection results for the detected edge line. It can be seen from the comparison that the slopes of the belt edge detected by the network are different. Further, when the deviation amount is calculated according to Equations (8)–(10), different deviation amounts are obtained, and different deviation states may be determined. Thus, the threshold value should be set reasonably according to the site conditions to accurately determine the running state of the conveyor belt.

(3) Effects of random vibration and noise from rain and snow

In the daily use of the camera, some unavoidable external forces (camera base vibration, strong wind, rain and snow, etc.) will inevitably lead to image dithering, blur, and random noise in the camera's field of vision. This will affect the detection results. In this section, global motion blur and global rain or snow noise are added to the original image shown in Figure 9a using image processing technology to simulate the interference of the camera shaking and of rainy or snowy weather, and the respective influence of these types of interference on the algorithm's detection results is explored in Figures 11 and 12.

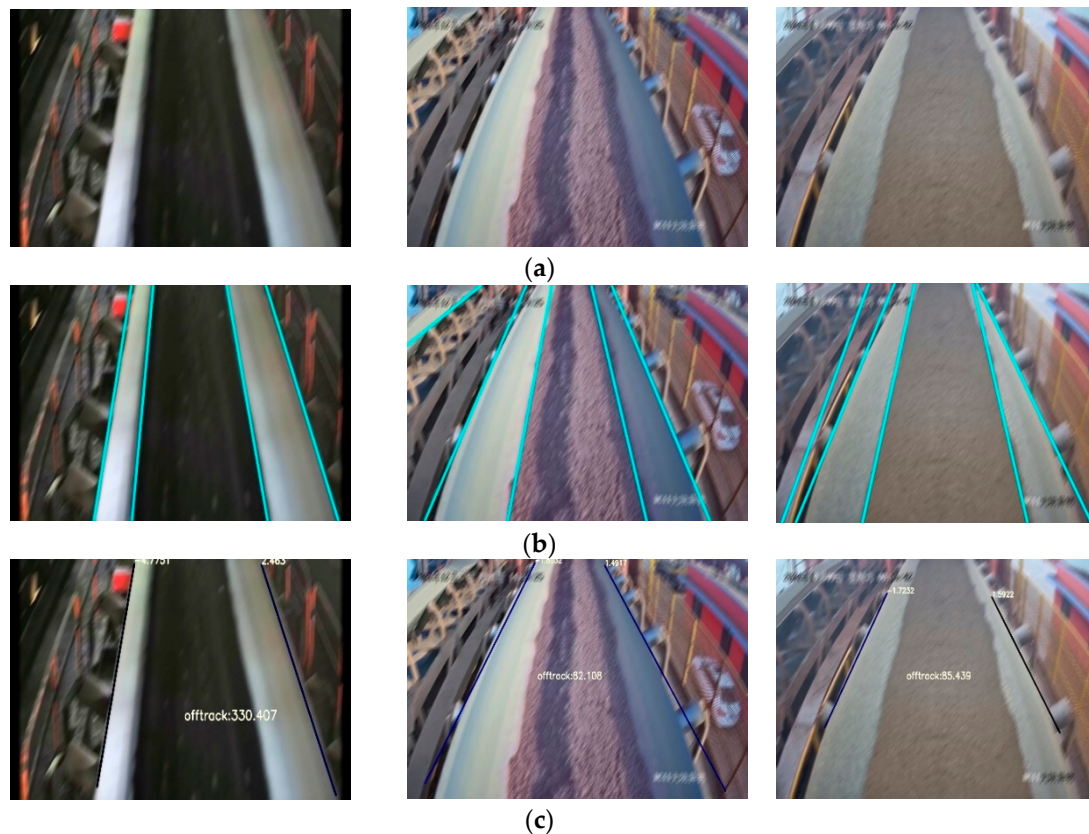


Figure 11. Influence of image jitter on conveyor belt edge detection: (a) image with motion blur, (b) detection result of DHT, (c) detection result of the method proposed in this paper.

According to the influence of image dithering on the detection results shown in Figure 11, compared with the DHT algorithm, the proposed method in this paper still has better stability under the condition of camera (image) dithering and is not disturbed by other “potential” lines in the field of vision. However, by comparing the calculated value of the amount of deviation and the judgment results, it can be seen that, in this case, camera (image) dithering greatly interferes with the proposed method. This leads to the unreliability of the calculations regarding deviation and poor accuracy. By analyzing the

influence of rain or snow noise on the detection results in Figure 12, it can be seen that rain and snow noise greatly impact both the DHT algorithm and the proposed method, which leads to both misses and false detection in the scenario of low-illumination power plants, but the method proposed in this paper still maintains a good detection effect in other scenarios.

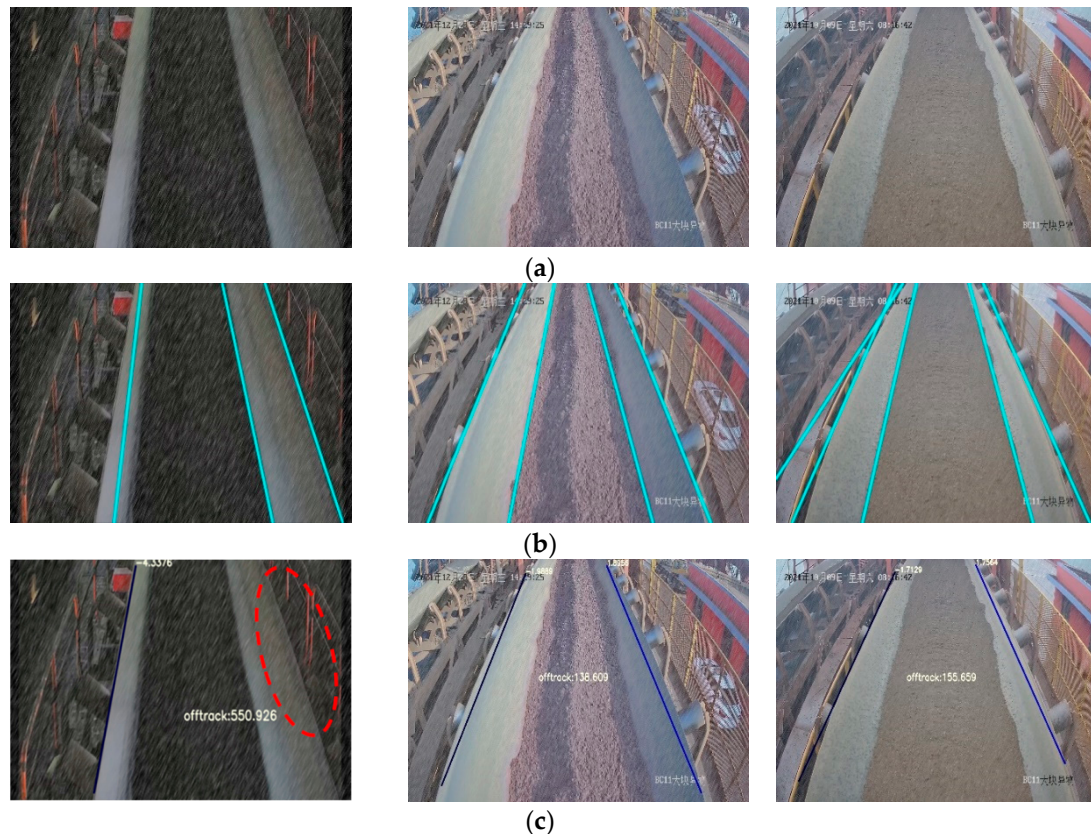


Figure 12. Influence of rain or snow noise on conveyor belt edge detection: (a) image with rain or snow noise, (b) detection result of DHT, (c) detection result of the method proposed in this paper.

4.4. Discussion

This paper treats the issue of conveyor belt deviation as a problem of straight-line detection in specific scenarios. By leveraging the unique characteristics of the monitoring scene and the object being monitored, a method is proposed to represent the conveyor belt edge lines using the diagonal features of the prediction boxes from the object detection network. Additionally, a quantitative determination method for the conveyor belt deviation status is introduced. End-to-end conveyor belt edge detection is achieved using the YOLOv5 network on a custom dataset.

Similarly to object detection in other scenarios, this method eliminates the need for the threshold setting processes inherent in traditional image processing methods. For instance, traditional image processing involves steps such as Canny edge detection and Hough line detection, which require setting numerous thresholds, such as grayscale transformation thresholds, edge detection gradient thresholds, Hough transform accumulator thresholds, minimum line segment length parameters, and gap parameters. These thresholds need to be readjusted for different monitoring scenarios or conveyor belt system scenarios. Therefore, the method proposed in this paper is comparatively simpler and more straightforward and the algorithm possesses stronger adaptability. Moreover, thanks to the end-to-end detection approach, the proposed method significantly enhances the detection speed. Compared to methods based on conveyor belt edge segmentation or conveyor belt region segmentation,

the object detection-based monitoring method greatly reduces the difficulty and intensity of dataset creation.

However, the method proposed in this paper still faces some issues, such as ensuring that the diagonal vertices of the prediction boxes always fall on the edge of the conveyor belt. This study used Grad-CAM technology to explore the effectiveness of different categories' attention mechanisms in improving the detection accuracy and identifying the feature locations utilized by the current network predictions. This investigation revealed feature locations that may affect the detection results, providing a reference to ensure the detection accuracy and reliability in future work. This is also one of the main differences between the work described in this paper and our previous versions. We are attempting to find a method that always maintains a high degree of coupling. We believe that the ideal feature distribution should have all utilized features located within the prediction box, with greater weight at the relevant diagonal vertex positions.

In future work, we will attempt to use the detection approach of the CornerNet network, representing the conveyor belt edge by directly predicting the vertex coordinates of the left and right edges.

5. Conclusions and Future Work

Targeting the methods for the monitoring of conveyor belt deviation states, this paper proposes a belt edge straight-line detection method based on diagonal features, and the Yolov5 network was used to complete the tests. The proposed method was fully combined with the prediction box results of the general target detection network. The main contributions of this paper are as follows.

- (1) A new dataset containing 5000 images and a labeling method for conveyor belt deviation detection were presented.
- (2) We designed a quantitative characterization method and a judgment method for the determination of deviation states. These methods effectively balance the detection speed and accuracy, achieving detection accuracy of 99% on the self-made dataset and a detection speed of 67.7 FPS on the RTX2060.
- (3) Through the Grad-CAM technique, the effects of the types, positions, and quantities of the three attention mechanisms on the improvement in the deviation detection accuracy were explored, and the characteristic positions that affected the detection results were identified.
- (4) The proposed method has a strong anti-interference ability and can effectively resist environmental interference such as low illumination and motion blur and achieve the stable and reliable monitoring of operational states.

This work is of great significance in ensuring the safety and efficiency of the transportation system. In future work, we will continue to expand the corresponding dataset and continuously enhance the anti-interference ability of the network.

Author Contributions: Conceptualization, M.Z. (Mengchao Zhang) and Y.Y.; methodology, M.Z. (Manshan Zhou); software, X.Z. and Z.Y.; validation, M.Z. (Mengchao Zhang); investigation, Z.Y.; writing—original draft preparation, M.Z. (Mengchao Zhang) and X.Z.; writing—review and editing, M.Z. (Mengchao Zhang) and Y.Z.; visualization, X.Z. and Z.Y.; supervision, M.Z. (Manshan Zhou) and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data in this study are available on request from the corresponding author.

Conflicts of Interest: Author Manshan Zhou and Yuan Zhang was employed by the company Libo Heavy Industries Science and Technology 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.

Nomenclature

CAP	channel average pooling
CBAM	convolutional block attention module
CMP	channel max pooling
CSP	cross-stage partial network
DHT	deep Hough transform
ECA	efficient channel attention
FPN	feature pyramid networks
FPS	frames per second
GAP	global average pooling
GMP	global max pooling
mAP	mean average precision
PANet	path aggregation network
SEnet	squeeze-and-excitation network
W, H	horizontal and vertical resolution of the image or horizontal and vertical length of the image
$a\ b\ c\ d$	four vertices of the prediction box
$e\ f\ g\ h$	diagonal vertices of the prediction box
(x_a, y_a)	the coordinates of point a
(x_b, y_b)	the coordinates of point b
(x_g, y_g)	the coordinates of point g
(x_h, y_h)	the coordinates of point h
(x_L, y_L)	center point coordinates of the left prediction box
(x_R, y_R)	center point coordinates of the right prediction box
h_L	height of the left prediction box
h_R	height of the right prediction box
w_L	width of the left prediction box
w_R	width of the right prediction box
D_{LM_0}	distance between the left edge of the conveyor belt and the left edge of the field of view
D_{RN_0}	distance between the right edge of the conveyor belt and the right edge of the field of view
L	intersection of the virtual reference line and the left side of the camera's field of view
M_0	intersection of the virtual reference line and the left side of the conveyor belt area
N_0	intersection of the virtual reference line and the right side of the conveyor belt area
R	intersection of the virtual reference line and the right side of the camera's field of view
Δy	distance between reference line and x-axis
τ	threshold for deviation determination

References

1. He, D.; Pang, Y.; Lodewijks, G.; Liu, X. Healthy speed control of belt conveyors on conveying bulk materials. *Powder Technol.* **2018**, *327*, 408–419. [\[CrossRef\]](#)
2. Mendes, D.; Gaspar, P.D.; Charrua-Santos, F.; Navas, H. Enhanced Real-Time Maintenance Management Model—A Step toward Industry 4.0 through Lean: Conveyor Belt Operation Case Study. *Electronics* **2023**, *12*, 3872. [\[CrossRef\]](#)
3. Jan, Z.; Ahamed, F.; Mayer, W.; Patel, N.; Grossmann, G.; Stumptner, M.; Kuusk, A. Artificial intelligence for industry 4.0: Systematic review of applications, challenges, and opportunities. *Expert Syst. Appl.* **2023**, *216*, 119456. [\[CrossRef\]](#)
4. Soori, M.; Arezoo, B.; Dastres, R. Internet of things for smart factories in industry 4.0, a review. *Internet Things Cyber-Phys. Syst.* **2023**, *3*, 192–204. [\[CrossRef\]](#)
5. Zhang, M.; Jiang, K.; Cao, Y.; Li, M.; Wang, Q.; Li, D.; Zhang, Y. A new paradigm for intelligent status detection of belt conveyors based on deep learning. *Measurement* **2023**, *213*, 112735. [\[CrossRef\]](#)
6. Semrád, K.; Draganová, K. Implementation of Magnetic Markers for the Diagnostics of Conveyor Belt Transportation Systems. *Sustainability* **2023**, *15*, 8705. [\[CrossRef\]](#)
7. Zhang, S.; Xia, X. Optimal control of operation efficiency of belt conveyor systems. *Appl. Energy* **2010**, *87*, 1929–1937. [\[CrossRef\]](#)
8. He, D.; Pang, Y.; Lodewijks, G. Green operations of belt conveyors by means of speed control. *Appl. Energy* **2017**, *188*, 330–341. [\[CrossRef\]](#)
9. Zhang, M.; Shi, H.; Zhang, Y.; Yu, Y.; Zhou, M. Deep learning-based damage detection of mining conveyor belt. *Measurement* **2021**, *175*, 109130. [\[CrossRef\]](#)
10. Bortnowski, P.; Kawalec, W.; Król, R.; Ozdoba, M. Types and causes of damage to the conveyor belt-review, classification and mutual relations. *Eng. Fail. Anal.* **2022**, *140*, 106520. [\[CrossRef\]](#)

11. Sun, X.; Wang, Y.; Meng, W. Evaluation System of Curved Conveyor Belt Deviation State Based on the ARIMA–LSTM Combined Prediction Model. *Machines* **2022**, *10*, 1042. [\[CrossRef\]](#)
12. Qiao, W.; Lan, Y.; Dong, H.; Xiong, X.; Qiao, T. Dual-field measurement system for real-time material flow on conveyor belt. *Flow Meas. Instrum.* **2022**, *83*, 102082. [\[CrossRef\]](#)
13. Zhang, M.; Jiang, K.; Cao, Y.; Li, M.; Hao, N.; Zhang, Y. A deep learning-based method for deviation status detection in intelligent conveyor belt system. *J. Clean Prod.* **2022**, *363*, 132575. [\[CrossRef\]](#)
14. Zhang, M.; Shi, H.; Yu, Y.; Zhou, M. A computer vision based conveyor deviation detection system. *Appl. Sci.* **2020**, *10*, 2402. [\[CrossRef\]](#)
15. Wang, J.; Liu, Q.; Dai, M. Belt vision localization algorithm based on machine vision and belt conveyor deviation detection. In Proceedings of the 34rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), Jinzhou, China, 6–8 June 2019; IEEE: Piscataway, NJ, USA, 2019.
16. Liu, Y.; Miao, C.; Li, X.; Xu, G. Research on deviation detection of belt conveyor based on inspection robot and deep learning. *Complexity* **2021**, *2021*, 3734560. [\[CrossRef\]](#)
17. Wang, W.; Xu, S.; Teng, Y. Design of Belt Sprinkler Monitoring System Based on Image Processing Technology. *IAENG Int. J. Comput. Sci.* **2022**, *49*, 94–100.
18. Zeng, C.; Zheng, J.; Li, J. Real-time conveyor belt deviation detection algorithm based on multi-scale feature fusion network. *Algorithms* **2019**, *12*, 205. [\[CrossRef\]](#)
19. Liu, Y.; Wang, Y.; Zeng, C.; Zhang, W.; Li, J. Edge detection for conveyor belt based on the deep convolutional network. In *Proceedings of the 2018 Chinese Intelligent Systems Conference, Wenzhou, China*; Springer: Berlin/Heidelberg, Germany, 2019.
20. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
21. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015.
22. Wang, Y.; Wang, Y.; Dang, L. Video detection of foreign objects on the surface of belt conveyor underground coal mine based on improved SSD. *J. Ambient Intell. Humaniz. Comput.* **2020**, *14*, 5507–5516. [\[CrossRef\]](#)
23. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016.
24. Redmon, J.; Farhadi, A. Yolo3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
25. Bochkovskiy, A.; Wang, C.; Liao, H.M. Yolo4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
26. Ultralytics. Yolo5. Available online: <https://github.com/ultralytics/yolov5> (accessed on 1 August 2024).
27. Zhang, M.; Yue, Y.; Jiang, K.; Li, M.; Zhang, Y.; Zhou, M. Hybrid Compression Optimization Based Rapid Detection Method for Non-Coal Conveying Foreign Objects. *Micromachines* **2022**, *13*, 2085. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
29. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
30. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
31. Woo, S.; Park, J.; Lee, J.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
32. Zhang, M.; Cao, Y.; Jiang, K.; Li, M.; Liu, L.; Yu, Y.; Zhou, M.; Zhang, Y. Proactive measures to prevent conveyor belt Failures: Deep Learning-based faster foreign object detection. *Eng. Fail. Anal.* **2022**, *141*, 106653. [\[CrossRef\]](#)
33. Zhao, K.; Han, Q.; Zhang, C.; Xu, J.; Cheng, M. Deep hough transform for semantic line detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 4793–4806. [\[CrossRef\]](#) [\[PubMed\]](#)

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.