technologies

MDPI

Article

A Foreign Object Detection Method for Belt Conveyors Based on an Improved YOLOX Model

Rongbin Yao [1], Peng Qi [2], Dezheng Hua [2], Xu Zhang [2], He Lu [1] and Xinhua Liu [2,*]

[1] Lianyungang Normal College, Lianyungang 222006, China; rby1972@126.com (R.Y.); luhe119879@126.com (H.L.)
[2] School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 221116, China; ts22050044a31@cumt.edu.cn (P.Q.); hua_dezheng@cumt.edu.cn (D.H.); ts20050081a31@cumt.edu.cn (X.Z.)
* Correspondence: liuxinhua@cumt.edu.cn

Abstract: As one of the main pieces of equipment in coal transportation, the belt conveyor with its detection system is an important area of research for the development of intelligent mines. Occurrences of non-coal foreign objects making contact with belts are common in complex production environments and with improper human operation. In order to avoid major safety accidents caused by scratches, deviation, and the breakage of belts, a foreign object detection method is proposed for belt conveyors in this work. Firstly, a foreign object image dataset is collected and established, and an IAT image enhancement module and an attention mechanism for CBAM are introduced to enhance the image data sample. Moreover, to predict the angle information of foreign objects with large aspect ratios, a rotating decoupling head is designed and a MO-YOLOX network structure is constructed. Some experiments are carried out with the belt conveyor in the mine's intelligent mining equipment laboratory, and different foreign objects are analyzed. The experimental results show that the accuracy, recall, and mAP^{50} of the proposed rotating frame foreign object detection method reach 93.87%, 93.69%, and 93.68%, respectively, and the average inference time for foreign object detection is 25 ms.

Keywords: belt conveyor; foreign object detection; YOLOX; image enhancement; rotation detection

check for updates

Citation: Yao, R.; Qi, P.; Hua, D.; Zhang, X.; Lu, H.; Liu, X. A Foreign Object Detection Method for Belt Conveyors Based on an Improved YOLOX Model. *Technologies* **2023**, *11*, 114. https://doi.org/10.3390/technologies11050114

Academic Editor: Valeri Mladenov

Received: 20 July 2023
Revised: 7 August 2023
Accepted: 8 August 2023
Published: 26 August 2023

Copyright: © 2023 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

With the development of intelligent mines and the improvement of machine vision technology, real-time detection systems for belt conveyors have become an important research topic in recent years [1]. The transportation belt plays a pivotal role and is prone to severe accidents such as deviation, slipping, belt breakage, and longitudinal belt tearing during production [2]. An in-depth analysis revealed that non-coal foreign objects entering the belt conveyor system are accountable for 61% of belt tearing and breakage incidents, amounting to a total of 21 cases [3]. The accurate and rapid identification of foreign objects in the belt transportation system, followed by removal, can substantially mitigate damage to the belt system and ensure the safe and stable operation of the belt transportation system [4].

Intelligent detection for the production of safe underground coal mines has become a hot research topic [5,6]. By utilizing video surveillance images, combined with image processing and machine-vision-related technologies, the theory of mine image monitoring has been applied to multiple aspects of automatic safety detection in coal mines, such as the automatic identification of spontaneous combustion fires [7], coal production monitoring [8], face detection and recognition methods for underground miners [9], and the automatic recognition of coal rock interfaces in coal faces [10]. However, traditional belt conveyor foreign object systems still rely on cameras to transmit collected video data to the central control room, where the staff can monitor the coal transportation area and the surrounding environment in real time. This practice is associated with significant

drawbacks, including the duplication of work and fatigue-induced misjudgments. At the same time, staff cannot address foreign objects in a timely manner, which can easily cause foreign objects to block the transportation belt or sharp parts to scratch the belt, resulting in belt tearing and causing major safety accidents.

Nowadays, many systems have been proposed for the detection of foreign objects on the belt, and some results have been achieved. With an understanding of the characteristics of remote sensing targets with different dimensions, including their dense distribution and complex background, Xu et al. [11] applied YOLO-V3 to the field of remote sensing to detect remote sensing targets at different scales. The lightweight, fast Illumination Adaptive Transformer (IAT) was proposed by Cui et al. [12] to restore a normally lit sRGB image from either low-light or under-exposed conditions. Wang et al. [13] introduced the attention mechanism into YOLO-V5 to detect Solanum rostratum Dunal seeds, and it was found that the CBAM attention mechanism can effectively improve accuracy during model recognition. However, for the underground environment of a coal mine, there is no special model for the detection of foreign bodies on a coal flow belt, and the performance of other models still needs to be further improved.

In this work, a foreign object detection method for belt conveyors is proposed, and the remainder of this paper is organized as follows. In Section 2, the status and deficiencies of current foreign object detection methods are introduced. In Section 3, the improved algorithm and model architecture are proposed. In Section 4, comparative experiments are carried out on the proposed model, and the effectiveness of the improved algorithm is verified using a self-made foreign object dataset. The conclusions and future works are summarized in Section 5.

2. Literature Review

2.1. Foreign Object Detection Methods Based on Image Processing

The foreign object detection methods for belt conveyors based on images work with images of coal and non-coal foreign objects by obtaining the shallow or deep abstract features of the object, and use image processing to detect foreign objects. This approach has the advantages of simple installation and maintenance processes and low application costs, and has become one of the research focuses of foreign object detection for belt conveyors. Due to the diversity of the types of foreign objects (such as anchor rods and wood) causing belt tearing, many scholars have begun to extract the color, texture, shape, spatial relationship, and other features of foreign objects from the image features, achieving the automatic detection of foreign objects through image processing. Jiang et al. [14] used extreme median filtering to perform image noise processing and improve the traditional Canny edge detection algorithm to obtain an improved edge detection method for a Canny operator. The algorithm is used to perform image edge detection, and the image gray histogram is used for enhanced foreign object image processing. Zhang et al. [15] proposed a new image segmentation algorithm for belt conveying. A multi-scale linear filter composed of a Hessian matrix and Gaussian function forms the core of the algorithm, which can effectively obtain the edge intensity image, form a good seed area for watershed segmentation, and segment the background between the coal pile and foreign objects. Saran et al. [16] developed an image-processing-based foreign object detection solution to detect foreign objects such as concrete boulders and iron bars that often occur in the conveyor belts used for G furnace raw coal. The solution uses a multi-mode imaging (polarization camera)-based system to distinguish foreign objects. Tu et al. [17] proposed a new moving target detection method to solve the difficulties caused by the intermittent motions, temperatures, and dynamic background sequences of moving targets. By further comparing the similarities of edge images, ghosts and real static objects can be classified. Lins et al. [18] developed a system based on the concept of machine vision, which aims to realize the automation of the crack measurement process. Using the above method, a series of images can be processed and the crack size can be estimated as long as a camera is installed on a truck or robot.

2.2. Foreign Object Detection Method Based on Deep Learning

With the rapid development of deep learning, using this data learning method to learn image data features and perceive the surrounding environment has good research value in foreign object detection, to obtain a foreign object detection model that is more adaptable to a complex and changeable environment [19,20]. Deep learning is achieved by establishing and simulating the information processing neural structure of the human brain to extract low-level to high-level features from external input data, enabling machines to understand the learning data and obtain useful information [21]. Pu et al. [22] used CNN to identify coal and gangue images and to help separate coal and gangue, and introduced transfer learning to solve the problems of massive trainable parameters and limited computing power faced by the model. In order to apply CNN to the field of target detection, Ren et al. [23] put forward the RCNN method, which uses a selective search to obtain pre-selected regions, and completes image recognition through CNN combined with SVM. Because the multi-stage implementation of the algorithm led to its huge time cost, Girshick et al. [24] further put forward the concept of a ROI (region of interest) pooling layer, and replaced SVM with fully connected neural network, and proposed the Fast-RCNN algorithm. In order to solve the problem of foreign objects on the belt conveyor in the coal mine damaging the belt conveyor, Wang et al. [25] proposed a video detection method of foreign objects on the surface of the belt conveyor based on SSD. Firstly, the deep separable convolution method was adopted to reduce the number of parameters of the SSD algorithm and improve the calculation speed. Then, the GIOU loss function was used to replace the position loss function in the original SSD, which improves detection accuracy. Finally, the extraction position of the feature map and the proportion of the default frame were optimized, which improves the detection accuracy. Considering the fast-running speed of the belt and the influence of background and light source on foreign object targets, Ma et al. [26] proposed an improved Center-Net algorithm, which improved detection efficiency. The normalization method was optimized to reduce computer memory consumption, and a weighted feature fusion method was added to fully utilize the features of each layer, improving detection accuracy. In the experimental environment, the average detection rate was about 20fps, and met the demand for the real-time detection of foreign objects. Xiao et al. [27] used a median filtering method to preprocess images with foreign objects, removed the influence of dust, improved the clarity of ore edges, and established a dataset to train the YOLOv3 belt foreign object detection algorithm. Finally, after sparse training based on the BN layer, the YOLOv3 model was lightweight, and its parameters were fine tuned. Compared with the original YOLOv3 model, the model achieved smaller calculations, faster processing, and a smaller size.

2.3. Discussion

However, although many approaches for detecting foreign objects have been developed in the above literature, some common disadvantages of them are summarized as follows. Firstly, due to the specific coal mine environment, with a lot of dust, noise, and a complex background for foreign objects, it is difficult to achieve accurate detection of foreign objects on the belt conveyor. Therefore, the general target detection algorithms cannot be easily migrated to the coal mine environment. At the same time, the robustness of the traditional foreign object detection algorithms is poor, and the extraction of foreign object features requires a wealth of experience. Finally, the current public foreign object detection dataset lacks coal-mine-belt foreign object detection data, so it cannot flexibly adapt to different scenarios in different mining areas.

In this paper, the image dataset of belt foreign bodies in the coal mine environment is collected and established, and a target detection algorithm based on improved YOLO-V5 is used to detect non-coal foreign bodies on the coal belt.

3. The Proposed Foreign Object Detection Method

3.1. Target Detection of YOLO Model

The YOLO series target detection algorithm is a supervised learning target detection algorithm [28]. Its basic principle is to divide the input image into several grids, then extract the features of each part of the image through the convolutional neural network, and finally output the predicted bounding box, which is the center coordinates of the predicted object; the length and width of the detected object; and the confidence of the object category.

As shown in Figure 1, the input foreign object image is divided into S × S squares, and features are extracted from each grid through the convolutional neural network, then, features are fused and analyzed to output the confidence degree of the foreign object target, the boundary box coordinate information, and the foreign object category. In order to improve the accuracy of foreign object detection, a fixed number of anchor boxes are used for each grid to assist in learning position information. Clustering analysis is performed on the known labels of the target detection object in the image to obtain the initial size of the anchor box.

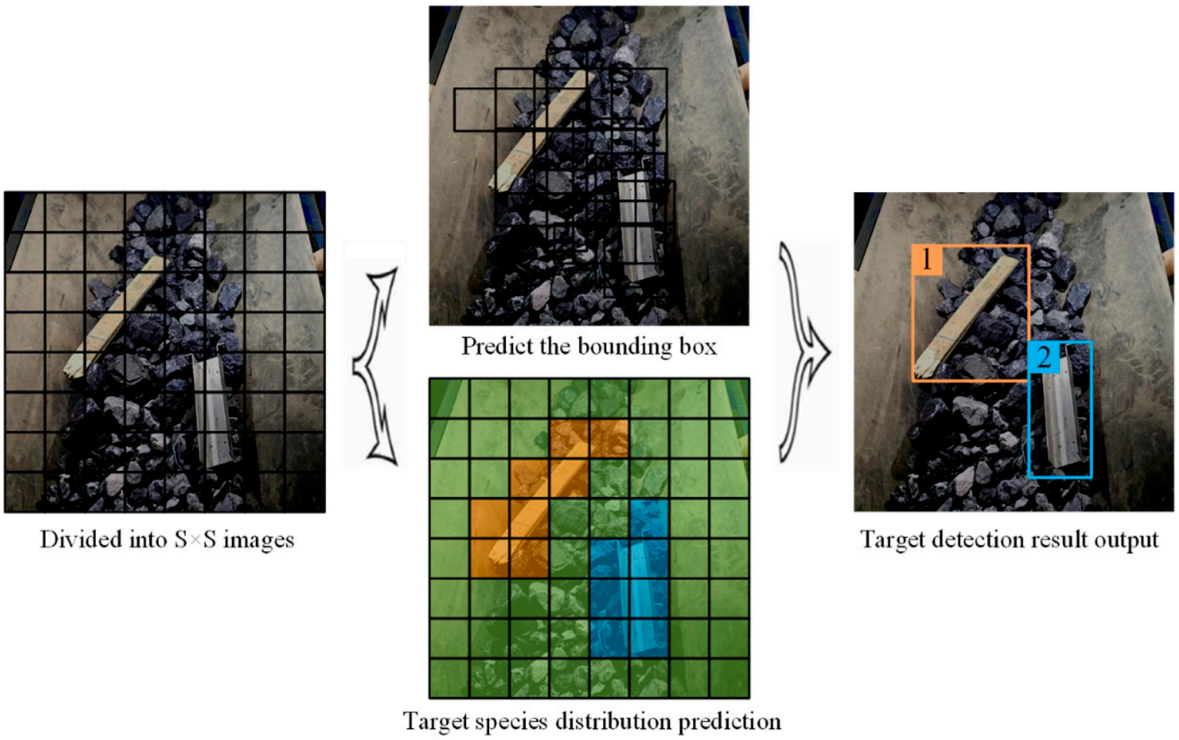


Figure 1. YOLO target detection process.

The framework of the YOLO series detection models has always been composed of three parts: the backbone feature extraction network, feature fusion layer, and detection decoupling head, as shown in Figure 2. The feature extraction network mainly extracts features from the input image data, then the feature fusion layer fuses the low-dimensional and high-dimensional features of the image to provide richer image information. Finally, the detection decoupling head outputs and predicts the position and category information of objects of interest. The YOLO series of algorithms all use a three branch detection head algorithm to predict objects of different scales, such as large, medium, and small.

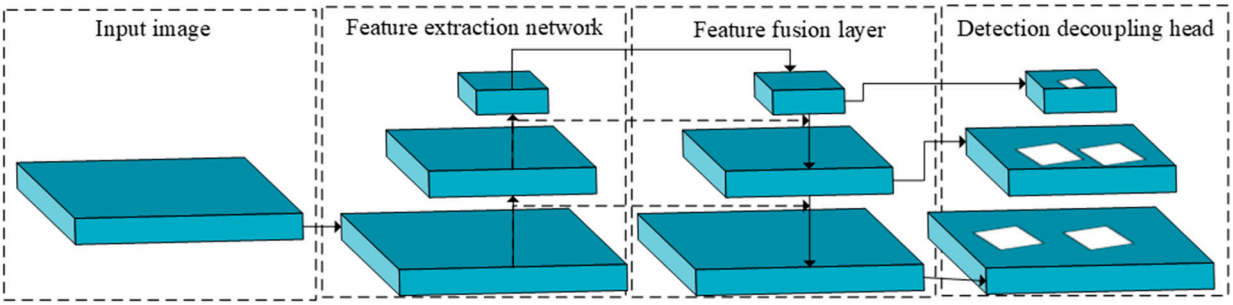


Figure 2. YOLO series network architecture.

In the actual target detection process, directly predicting the central coordinates, width, and height of the bounding box will result in too large a solution space for the predicted target, which will seriously waste computing resources. Therefore, an anchor frame mechanism is designed to accelerate the convergence of the model and improve the target detection accuracy, and the prediction principle of the bounding box is shown in Figure 3.

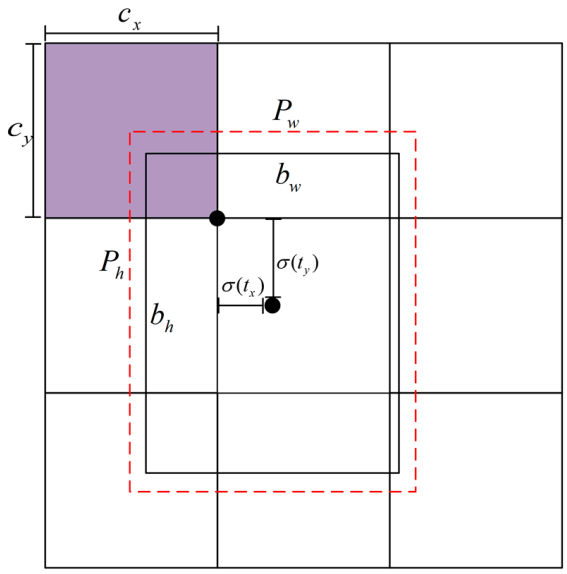


Figure 3. The prediction principle of YOLO's bounding box.

$$b_x = \sigma(t_x) + c_x \tag{1}$$

$$b_y = \sigma(t_y) + c_y \tag{2}$$

$$b_w = p_w e^{t_w} \tag{3}$$

$$b_h = p_h e^{t_h} \tag{4}$$

where p_w and p_h are the width and height of the anchor frame; b_w and b_h are the width and height of the prediction box; t_x and t_y are the offset from the anchor frame to the center of the prediction box; c_x and c_y are the coordinates of the upper left corner of the bounding box; $\sigma()$ is the normalized function.

3.2. Established Foreign Object Image Dataset

At present, the publicly available large-scale datasets do not include non-coal foreign objects. Therefore, it is necessary to establish an actual foreign object engineering dataset for belt conveyors to solve the problem of foreign object detection in practical engineering. This self-made dataset is named the belt conveyor foreign object detection dataset, and the sample categories of the dataset mainly include the following three types of foreign objects: iron, wood, and large gangue, as shown in Figure 4.



Figure 4. Different kinds of foreign object samples.

This work selected laboratory and belt-conveyor work scenarios for foreign object image collection. At the same time, foreign object image datasets were captured under different natural light conditions, and directional foreign objects such as iron and wood were offset to increase the information of image angles. In the laboratory environment, for the same foreign object, a foreign object dataset can be established that includes images of areas without coal flow, areas with coal flow, and areas obstructed by coal flow. Photos of foreign objects in different directions were collected in the laboratory environment to increase the diversity of foreign object dataset samples, as shown in Figure 5.

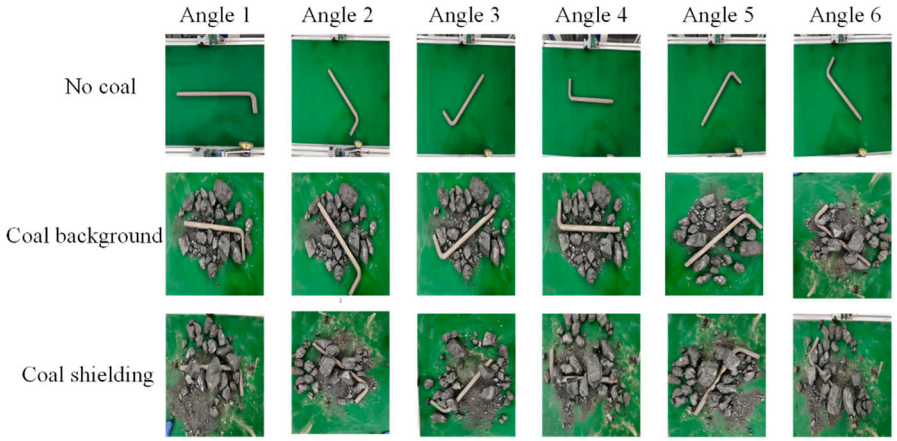


Figure 5. Single foreign material data samples.

In order to ensure the diversity of perspective in the collected dataset and better simulate the different shooting angles of cameras installed in actual working conditions, the top view images were collected by using a DJI drone with a pan tilt camera. The heights from the ground during the collection weare 1 m, 2 m, and 4 m, respectively, to ensure the diversity of perspective in the collected data, as shown in Figure 6.

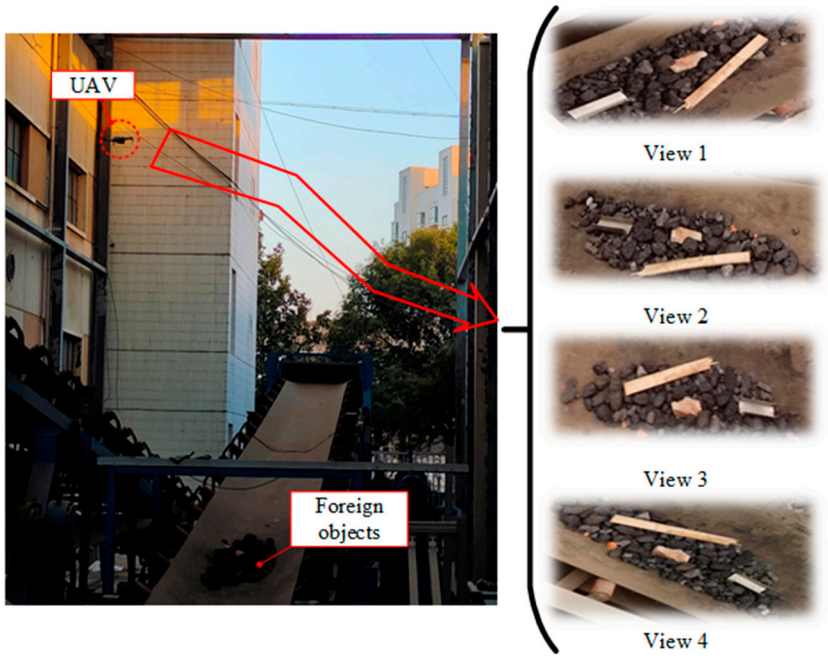


Figure 6. Multi-view image acquisition by UAV.

In order to improve the robustness and generalization of the model, the Mosaic multi-samples data augmentation method proposed by YOLOv4 was adopted. During the training process, four images in the training set were randomly selected, and the images were randomly scaled, cropped, and arranged for image combination. The sample size of the images during the training process was expanded, as shown in Figure 7.

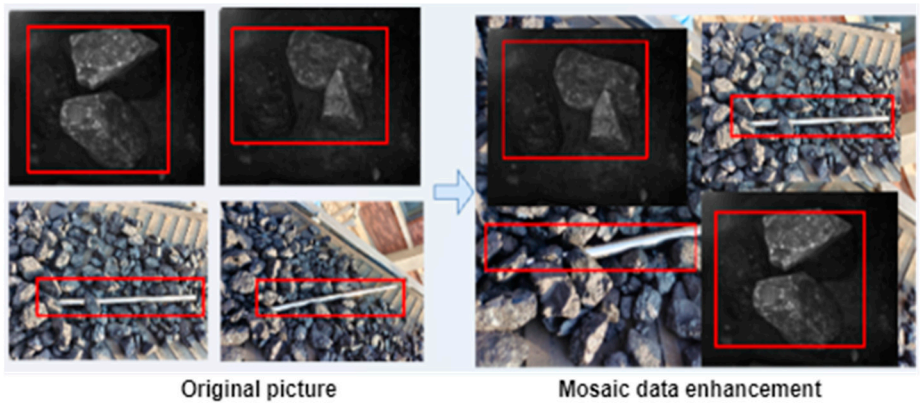


Figure 7. Mosaic multi-sample data enhancement.

As shown in Figure 8, the images were expanded by means of horizontal flipping, random occlusion, random scaling, motion blur, random scaling and filling, and salt and pepper noise. A total of 1105 foreign object image datasets were collected for the belt conveyor foreign object image dataset, including 303 large gangue datasets, 401 iron tools datasets, 301 wood datasets, and 100 mixed target images. The belt conveyor foreign object image dataset was labeled with horizontal and rotating boxes, and the horizontal and

rotating box foreign object detection datasets were constructed. Finally, the dataset was expanded to 8100 datasets through geometric expansion, and a complete dataset of foreign object images for belt conveyors had been constructed.

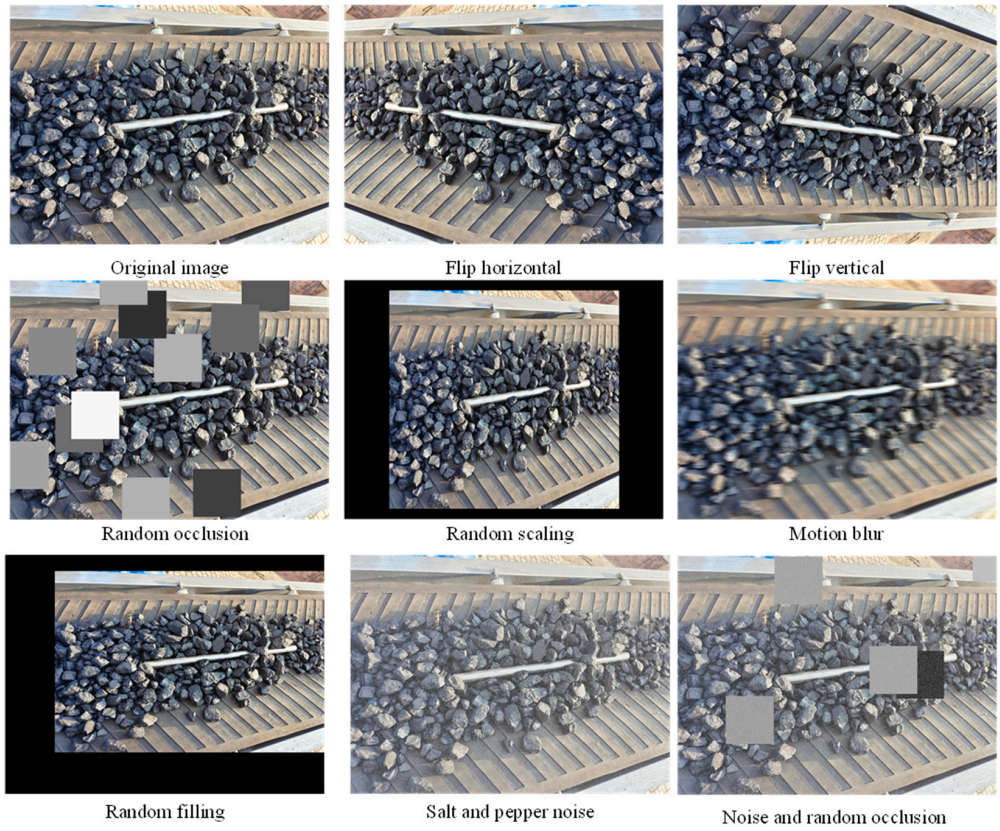


Figure 8. Geometry data enhancement.

3.3. Improved Depthwise Separable Convolution Block

Depthwise separable convolution breaks down the operations of standard convolution into depthwise convolution and point by point convolution [29]. Depthwise convolution performs separate spatial convolutions on each input channel, while point by point convolution combines the convolution results of each channel, which can greatly reduce the size and complexity of the model while maintaining high accuracy. The specific operations are shown in Figure 9.

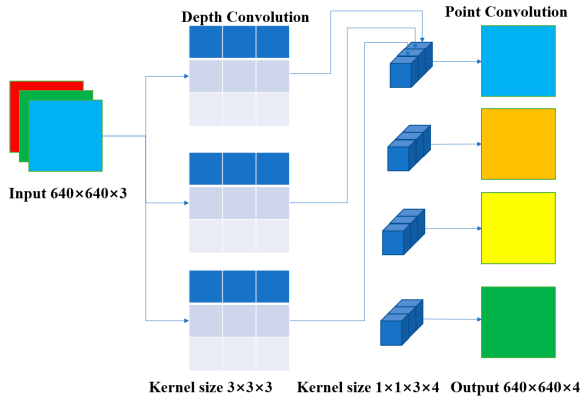


Figure 9. Depthwise separable convolution.

Assuming that the input image size is 640 × 640 × 3 and the expected output size is 640 × 640 × 4, and the ordinary convolution uses a convolution kernel of 3 × 3 × 3 × 4, then the parameter *stc* of the ordinary convolution is:

$$stc = H \times W \times chanel_{in} \times chanel_{out} = 3 \times 3 \times 3 \times 4 \tag{5}$$

Depthwise separable convolution is used for depth-by-depth convolution, and then point-by-point convolution of the channel relationship is carried out. First, convolution is performed by depth, and the number of parameters is as follows:

$$d_w = H \times W \times chanel_{in} = 3 \times 3 \times 3 \tag{6}$$

Then, through the point-by-point convolution operation, the total number of parameters for the depth-separable convolution is:

$$p_w = 1 \times 1 \times chanel_{in} \times chanel_{out} = 1 \times 1 \times 3 \times 4 \tag{7}$$

$$dsc = d_w + p_w = 3 \times 3 \times 3 + 1 \times 1 \times 3 \times 4 \tag{8}$$

The ratio of the number of parameters for the two convolution operations is:

$$\frac{dsc}{stc} = \frac{H \times W \times chanel_{in} + 1 \times 1 \times chanel_{in} \times chanel_{out}}{H \times W \times chanel_{in} \times chanel_{out}} \approx 0.36 \tag{9}$$

Using deep separable convolution for convolution operations can effectively reduce the number of parameters in the model, ensuring the feature extraction ability of the convolution and facilitating the light weight of the model. In addition, the Hard-Swish activation function was selected as the activation function of the belt conveyor foreign object detection model, as shown in Figure 10.

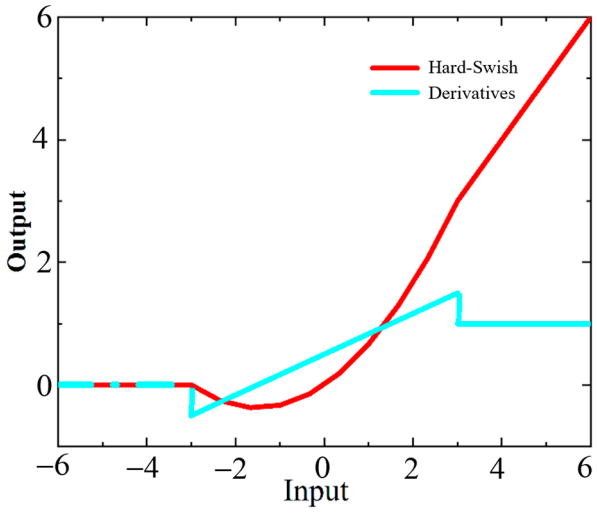


Figure 10. Hard-Swish activation function and its derivative.

The Hard-Swish activation function is a smooth function with no upper bound or lower bound. The activation function makes the model non-linear, which can effectively reduce the calculation cost in the embedded environment, and the expression is as follows:

$$\text{Hard-Swish(x)} = \begin{cases} 0 & if\ x \leq -3 \\ x & if\ x \geq +3 \\ x \times (x+3)/6 & otherwise \end{cases} \tag{10}$$

The basic module of the improved foreign object detection model is shown in Figure 11. Replacing the ordinary convolution at the end of the merge channel in the CSP1_ X and CSP2_X module with depthwise separatable convolution reduces the number of parameters in the convolution process and accelerates the inference speed of the model.

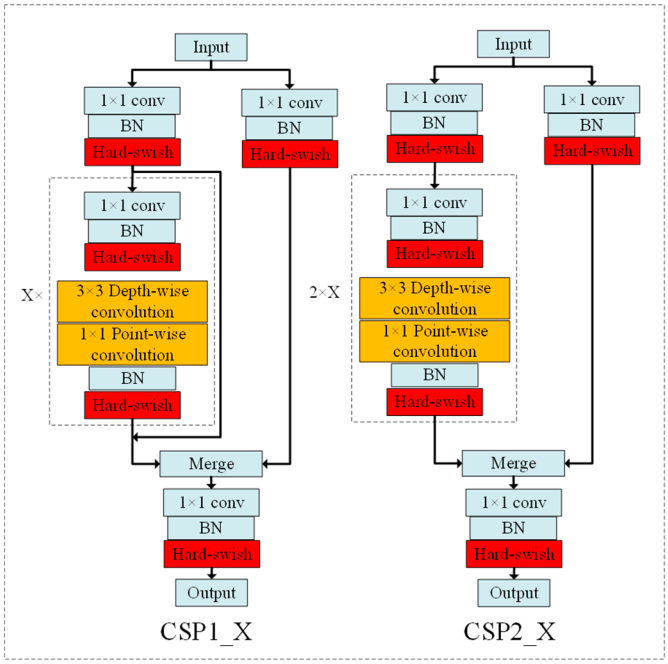


Figure 11. Improved CSP1_ X and CSP2_ X Structural blocks.

3.4. IAT Image Enhancement Module

In order to ensure the end-to-end output characteristics of deep learning, the IAT image enhancement module [30] is introduced to achieve image enhancement, and the network structure is shown in Figure 12. The color matrix in the IAT architecture represents the pixel weight weighted by a self-attention mechanism, in which the different colors are used to distinguish different patches from the original image. The IAT module can enhance the brightness of the image, restore the relevant details, improve the image quality, reduce the noise, and enhance the image contrast.

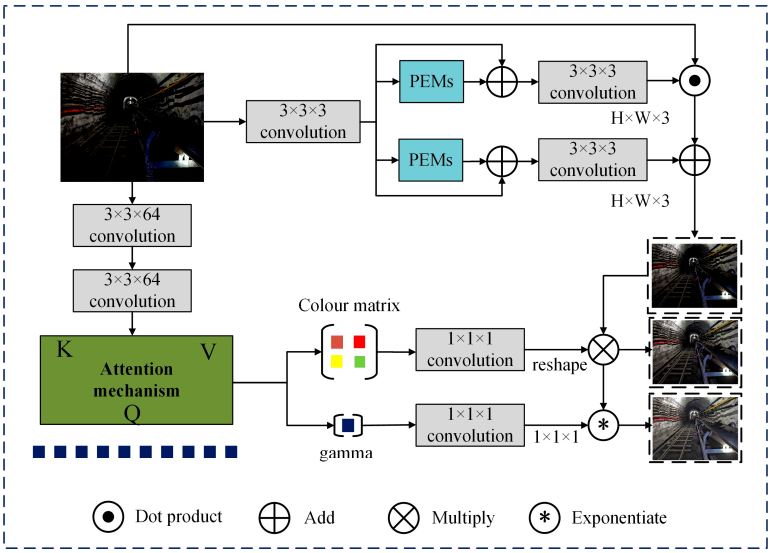


Figure 12. IAT image enhancement structure.

At the same time, the objective evaluation index Peak Signal to Noise Ratio (PSNR) for image enhancement is used as the specific evaluation index for image enhancement, and the formula is as follows:

$$PSNR = 10\log_{10}\left[\frac{(2^n-1)^2}{MSE}\right] \tag{11}$$

$$MSE = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}(X(i,j) - Y(i,j))^2 \tag{12}$$

where $X(i, j)$ are the pixel values of the original image, $Y(i, j)$ are the pixel values of the enhanced image, and H and W are the length and width of the image, respectively.

3.5. Improved CBAM Attention Block

CBAM [31] is a convolutional neural network module based on an attention mechanism, which is used to improve the overall performance of the model. Its essence is to inhibit the expression of redundant features by increasing the weight of non-redundant features. It is composed of a channel attention module (CAM) and a spatial attention module (SAM), and the specific network structure is shown in Figure 13.

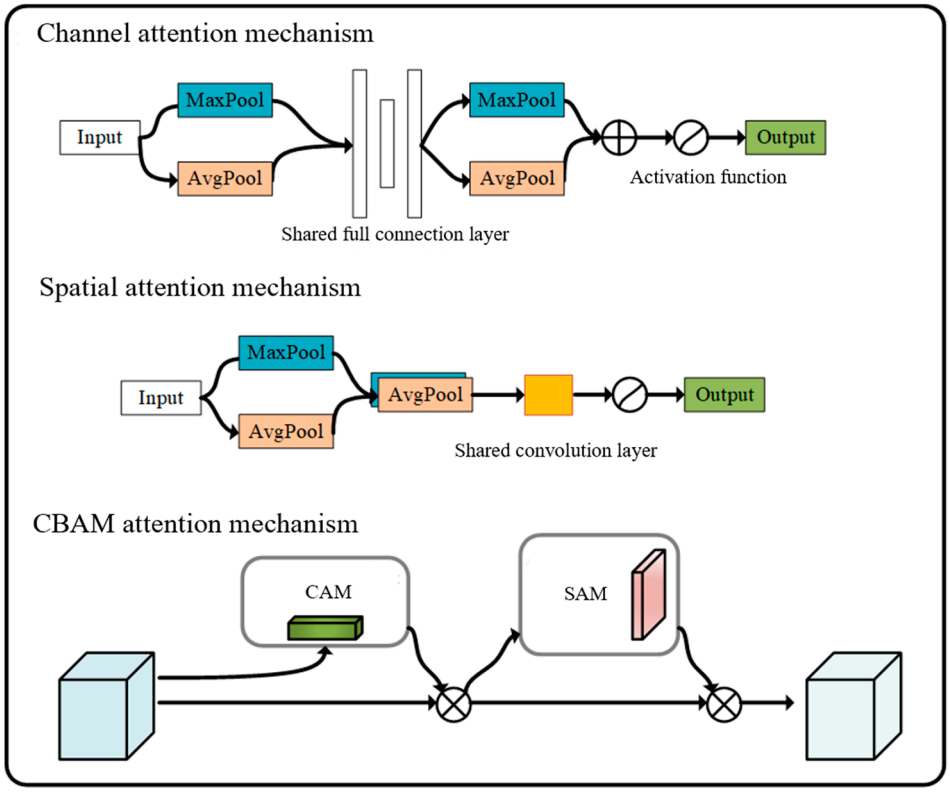


Figure 13. CBAM attention mechanism.

Therefore, in order to suppress redundant features and obtain attention feature maps that pay more attention to channels and spaces, the CBAM attention mechanism was introduced into the network structure, and the specific location of the addition is shown in Figure 14.

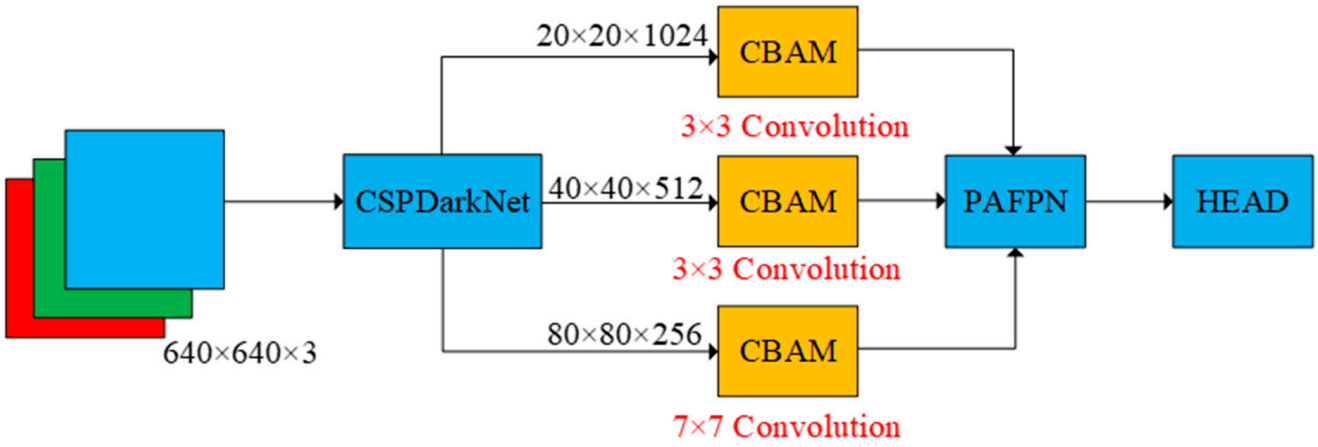


Figure 14. Improved addition location of CBAM in the network.

3.6. Designed Rotating Decoupling Head and MO-YOLOX Network

The detection boxes of the YOLO series object detection algorithms are all horizontal boxes, which is not conducive to the detection of foreign objects with diverse distribution directions such as ironware. Therefore, angle regression prediction was added to the head network of YOLOX, and a branch decoupling head based on the angle regression was constructed to accurately locate directional foreign objects. The structure of the rotary decoupling head is shown in Figure 15, where CBS*2 is an acronym for having two CBS modules. The overall network structure of MO-YOLOX is shown in Figure 16.

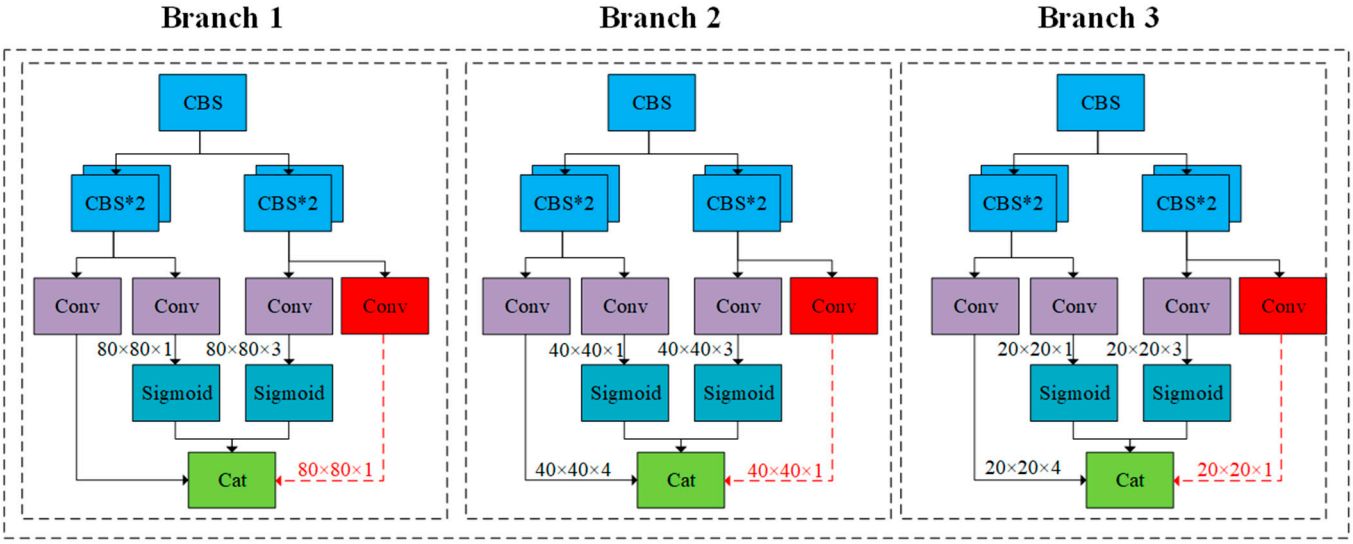


Figure 15. MO-YOLOX rotary decoupling head structure.

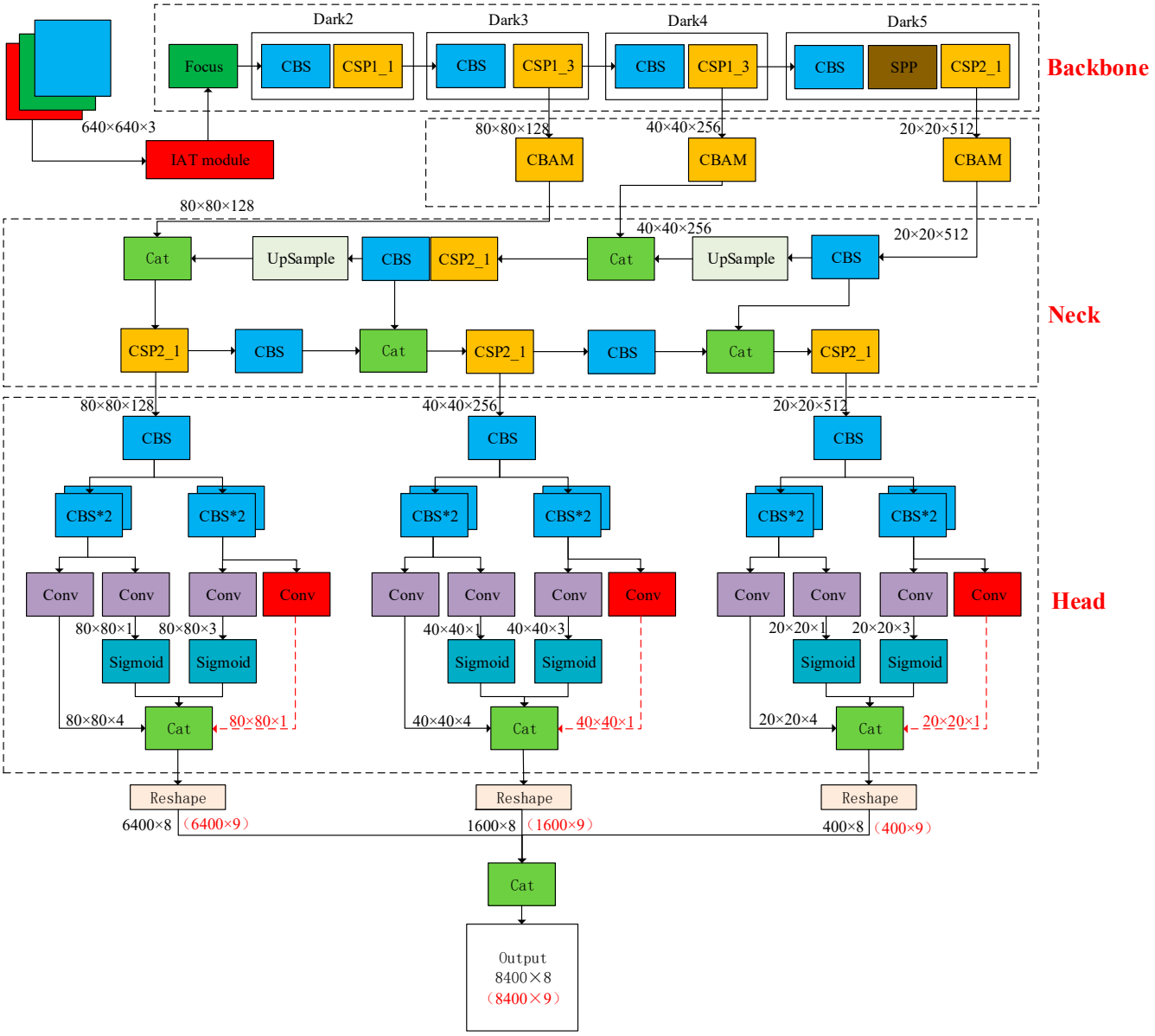


Figure 16. MO-YOLOX network model.

4. Experimental Example and Analysis

4.1. Experimental Platform

The proposed MO-YOLOX network model was trained in the GPU environment, and the environment configuration is shown in Table 1 below.

Table 1. Model training environment configuration.

Name	Parameter
CPU	Intel Core i9-10980XE
Hard disk	2 T
GPU	NVIDIA RTX A4000
Memory	16 G
Deep learning framework	Pytorch1.8.0
OS	Window10
Programming Language	Python3.8
CUDA	11.2

4.2. Experimental Comparisons

In order to verify the comprehensive performance of the proposed MO-YOLOX, comparison experiments of horizontal target detection and rotating target detection were carried out on the PASCALVOC dataset and the DOTA dataset, respectively. The PASCALVOC dataset is marked with horizontal boxes, including Bird, Dog, Cat, Person, Soft, Car, Bottle, and House. There are 312 image data, including a total of 1623 objects. The mainstream horizontal object detection models YOLOX-small and SSD300 [32] were selected for the comparative experiments. The experimental results are shown in Table 2, where excellent results are shown in bold.

Table 2. VOC dataset detection accuracy test results.

Metrics \ Model	MO-YOLOX	YOLOX-Small	SSD300
mAP^{50}	**71.9%**	70.6%	68.7%
Bird AP^{50}	**74.2%**	74.2%	71.2%
Dog AP^{50}	**78.7%**	75.2%	72.8%
Cat AP^{50}	78.5%	**80.3%**	73.9%
Person AP^{50}	**72.8%**	72.5%	69.8%
Sofa AP^{50}	71.1%	71.8%	**71.9%**
Car AP^{50}	**79.1%**	70.2%	70.8%
Bottle AP^{50}	49.2%	**50.9%**	49.9%
Average inference time	**21 ms**	28 ms	27 ms

In addition, in order to verify the effectiveness of MO-YOLOX in rotating target detection, the DOTA dataset marked with the rotating frame was selected, including plane (PL), ship (SH), large vehicle (LV), harbor (HA), small vehicle (SV), and baseball diamond (BD). Mainstream rotating object detection models, such as S2A-Net [33] and CFA [34], were selected for comparative experiments. The experimental results are shown in Table 3.

Table 3. DOTA dataset detection accuracy test results.

Metrics \ Model	MO-YOLOX	S2ANet	CFA
mAP^{50}	79.44%	79.26%	**79.57%**
PL AP^{50}	85.61%	**86.12%**	85.21%
SH AP^{50}	83.23%	82.23%	**83.82%**
LV AP^{50}	75.22%	76.32%	**80.91%**
HA AP^{50}	**76.51%**	75.41%	73.21%
SV AP^{50}	**78.29%**	75.23%	76.25%
BD AP^{50}	77.78%	79.25%	78.00%
Average inference time	**27 ms**	65 ms	66 ms

Compared with YOLOX-Small, the detection accuracy and reasoning speed of the MO-YOLOX target detection model with its attention mechanism and depthwise separable convolution are better than the original model. The average detection accuracy of the proposed model in the VOC test data set is higher than that of the original YOLOX-small and SSD300 model, and its detection accuracy in the DOTA data set is the same as that of S2ANet and CFA, but the reasoning time of MO-YOLOX is better than the above comparison algorithms. Therefore, the proposed foreign body detection network can meet the requirements of both detection accuracy and reasoning speed in the target detection task.

4.3. Experimental Testing and Analysis

The training dataset is the foreign object detection dataset of the belt conveyor, including the horizontal frame labeling dataset and the rotating frame marking dataset, and the relevant parameters are shown in Table 4. After 300 rounds of model training iterations, the proposed model can converge to relatively stable positions, and the loss values during the training process are shown in Figure 17. The models obtained from the above training were used as the optimal model for experimental comparison, and comparison experiments were conducted.

Table 4. Model training parameters.

Training Parameters	Setting Values
Activation function	Hard-Swish
Pooling method	Max-Pooling
Optimization algorithm	Adams, Batch-size = 8,
Loss function	Cross-entropy Loss function, KLD
Epoch	300
Data enhancement	Mosaic
Learning rate	Initial Learning rate $\alpha_0 = 0.01$, Nature Index attenuation
Dataset partitioning ratio	Training set:Verification set:Test set = 0.6:0.3:0.1

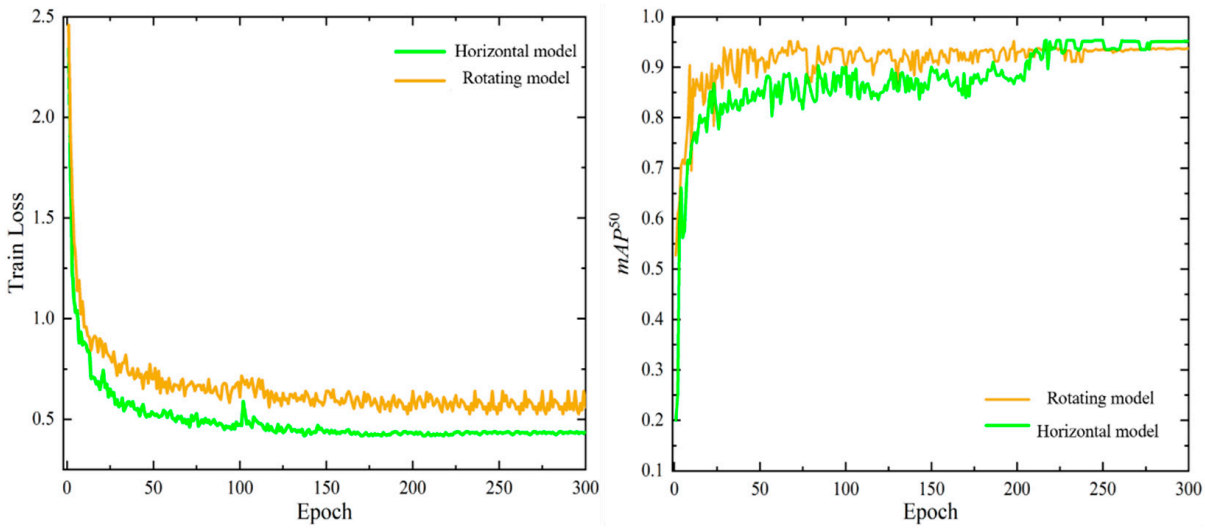
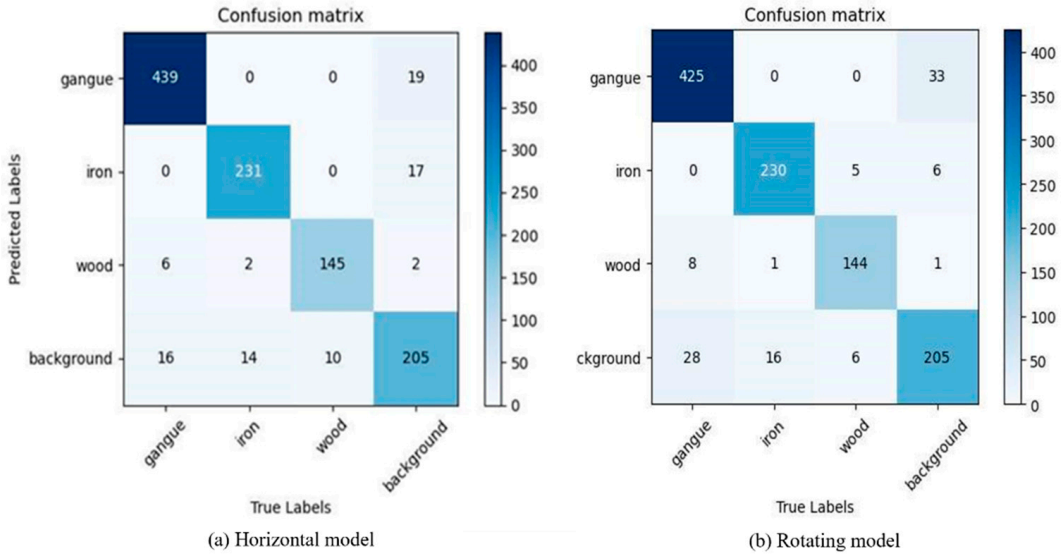


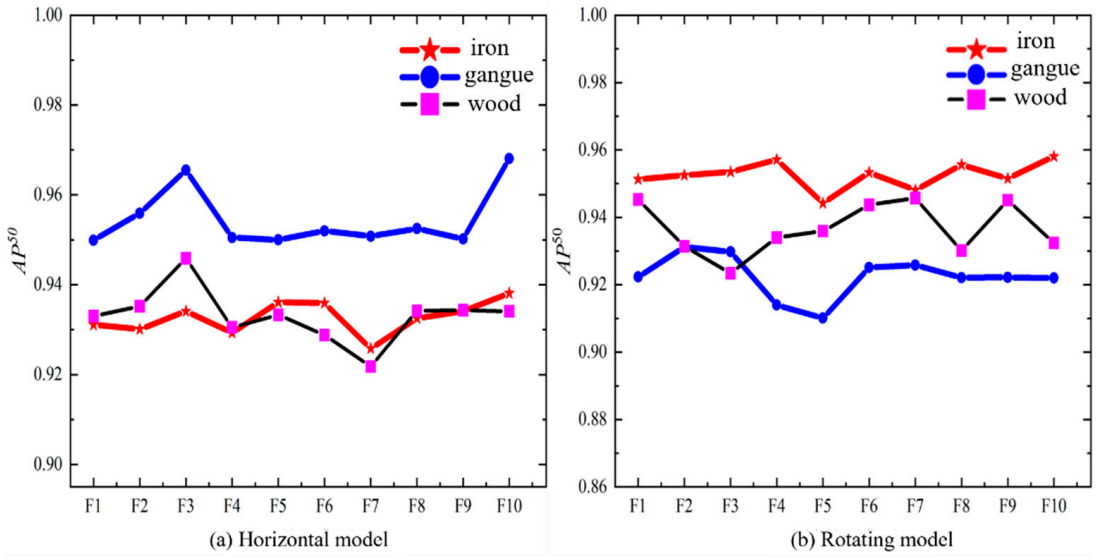
Figure 17. Model training curve.

In order to verify the detection performance of our foreign object detection model on the dataset, the self-made belt conveyor foreign object detection dataset was used for testing, with a total of 1070 images (including 253 background images). The confusion matrixes during the foreign object detection process were obtained, as shown in Figure 18. Ten-fold cross-validation was used to comprehensively evaluate the performance of the model, and the results of the cross-validation are shown in Figure 19.



(a) Horizontal model (b) Rotating model

Figure 18. Confusion matrix results of the self-made dataset.



(a) Horizontal model (b) Rotating model

Figure 19. Cross-validation results of foreign object detection models.

It can be seen from Figures 20 and 21 that the proposed foreign object detection model can effectively detect foreign objects in the case of background coal flow. The rectangle in the figures is the target result predicted by the foreign object detection model, and different colors represent different categories. In Figure 21, the predicted angle information is represented by the long side of the rotating rectangular box, with angle values of 36.8, −30.3, and 65.1, which can verify the effectiveness of the rotation decoupling head in angle regression prediction. Figures 22 and 23 show the results of foreign object detection under coal-flow occlusion and the multi-angle detection results of the same foreign object, respectively. The proposed model can locate the foreign object in the image more accurately, and the performance indicators of the foreign object detection model are shown in Tables 5 and 6.

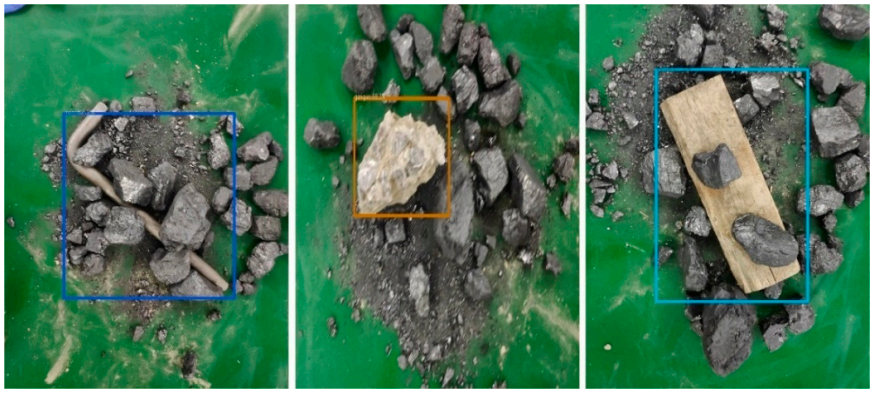
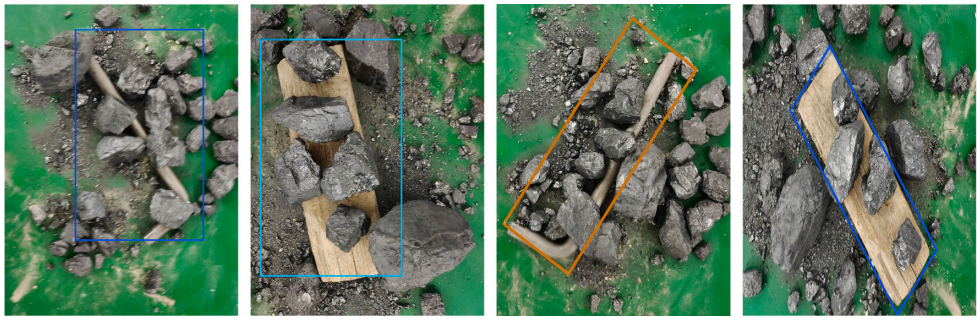


Figure 20. Test results of the foreign object detection of horizontal frames.



Figure 21. Test results of the foreign object detection of rotating frames.



(**a**) Horizontal box detection (**b**) Rotation box detection

Figure 22. Detection of foreign objects under the shelter of coal flow.

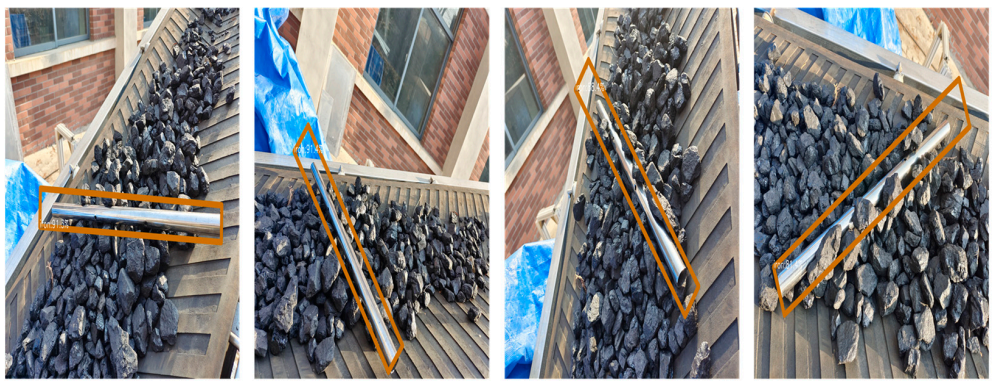


Figure 23. Multi-angle foreign object detection with the same foreign object sample.

Table 5. MO-YOLOX horizontal foreign object detection performance index parameters.

	Precision	Recall	AP^{50}	F2-Score	Inference Time/ms
iron	93.71%	93.20%	93.27%	93.30%	21
wood	93.12%	93.62%	93.30%	95.80%	23
large gangue	95.32%	95.92%	95.45%	93.52%	22
average value	94.05%	94.25%	94.01%	94.20%	22

Table 6. MO-YOLOX rotating frame foreign object detection performance index parameters.

	Precision	Recall	AP^{50}	F2-Score	Inference Time/ms
iron	95.11%	95.32%	95.25%	95.28%	28
wood	92.17%	92.51%	92.23%	92.44%	26
large gangue	94.32%	93.25%	93.56%	93.46%	29
average value	93.87%	93.69%	93.68%	93.73%	27.7

From the experimental results, it can be seen that the performance of the proposed foreign object detection model of the belt conveyor is superior to similar mainstream algorithms on both horizontal foreign object datasets and rotating foreign object datasets. Specifically, when the target foreign object and the coal mine stone have obvious differences in shape and texture, such as large gangue, the proposed horizontal frame foreign object detection model has very excellent performance. However, for slender ironware, the detection effect of the proposed model is slightly poor. It is exciting that the proposed rotating frame foreign body detection model has good detection sensitivity for targets with large length and width, such as slender iron bars, and the cost of angle prediction is a 5.7 ms increase in the reasoning time. There is an obvious difference in length and width between iron and wood, and the proposed model can effectively predict the angle of a foreign body. However, with the irregular gangue, its characteristics are quite different, and the angle information of the data label is irregular, which causes great difficulties in the angle regression prediction of the network. In addition, in the case of slight occlusion from the coal background, both horizontal frame foreign body detection and rotating foreign body detection can accurately detect foreign bodies and determine the types of foreign bodies. Therefore, the experimental results show that the proposed model meets the design requirements.

5. Conclusions and Future Works

In this paper, a foreign object image dataset for the belt conveyor is collected and established, and the IAT image enhancement module and CBAM attention mechanism are introduced. Secondly, a novel rotating decoupling head is designed to predict the angle information of foreign objects, and a MO-YOLOX network structure is constructed. The experimental results show that the proposed algorithm has a performance of 71.9% and 73.2% on the VOC and DOTA test datasets, respectively, with an average inference time of around 26 ms, which can meet the requirements of real-time inference. Ten-fold cross-validation is conducted on the self-built foreign object dataset of the belt conveyor, and the accuracy, recall, and mAP^{50} of horizontal frame foreign object detection are 94.05%, 94.25%, and 94.01%, respectively. Moreover, the accuracy, recall, and mAP^{50} of the rotating frame foreign object detection reaches 93.87%, 93.69%, and 93.68%, and the average inference time of foreign object detection is 25 ms.

However, the proposed foreign object detection method for belt conveyors we have designed has not yet considered embedded deployment as part of the industrial experiment. In the future, further research is needed on the pruning optimization of the model and embedded deployment.

Author Contributions: Conceptualization, R.Y. and P.Q.; methodology, D.H. and X.Z.; validation, H.L.; formal analysis, X.L.; data curation, P.Q.; writing—original draft preparation, D.H. and X.Z.; writing—review and editing, H.L.; funding acquisition, X.L. and D.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Lianyungang 521 High-level Talent Training Project grant number LYG06521202203, the Qing Lan Project for Excellent Teaching Team of Jiangsu Province grant number 2022, the National Natural Science Foundation of China grant number 51975568, the Natural Science Foundation of Jiangsu Province grant number BK20191341 and the Jiangsu Funding Program for Excellent Postdoctoral Talent grant number 2022ZB519.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data that produce the results in this work can be requested from the corresponding author.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this article.

References

1. Petrikova, I.; Marvalova, B.; Samal, S.; Cadek, M. Digital image correlation as a measurement tool for large deformations of a Conveyor Belt. *Appl. Mech. Mater.* **2015**, *732*, 77–80. [CrossRef]
2. Zimroz, R.; Stefaniak, P.K.; Bartelmus, W.; Hardygora, M. Novel techniques of diagnostic data processing for belt conveyor maintenance. In Proceedings of the 12th International Symposium Continuous Surface Mining—Aachen 2014; Springer International Publishing: Cham, Switzerland, 2015; pp. 31–40.
3. Cao, H.Q. Study and analysis on tear belt and break belt of belt conveyor in coal mine. *Coal Sci. Technol.* **2015**, *S2*, 130–134.
4. Hu, J.H.; Gao, Y.; Zhang, H.J. Recognition method of non-coal foreign matter in belt conveyor based on deep learning. *J. Mine Autom.* **2021**, *47*, 106653.
5. Chen, W.; Wang, X. Coal mine safety intelligent monitoring based on wireless sensor network. *IEEE Sens. J.* **2020**, *21*, 25465–25471. [CrossRef]
6. Gao, D.; Li, W.; Dai, K. Design of coal mine intelligent monitoring system based on ZigBee wireless sensor network. In Proceedings of the 2016 International Conference on Mechanics, Materials and Structural Engineering (ICMMSE 2016), Jeju Island, Republic of Korea, 18–20 March 2016; Atlantis Press: Amsterdam, The Netherlands, 2016; pp. 182–187.
7. Saydirasulovich, S.N.; Abdusalomov, A.; Jamil, M.K.; Nasimov, R.; Kozhamzharova, D.; Cho, Y.I. A YOLOv6-based improved fire detection approach for smart city environments. *Sensors* **2023**, *23*, 3161. [CrossRef]
8. Xu, P.; Zhou, Z.; Geng, Z. Safety monitoring method of moving target in underground coal mine based on computer vision processing. *Sci. Rep.* **2022**, *12*, 17899. [CrossRef]
9. Dai, L.; Qi, P.; Lu, H.; Liu, X.; Hua, D.; Guo, X. Image enhancement method in underground coal mines based on an improved particle swarm optimization algorithm. *Appl. Sci.* **2023**, *13*, 3254. [CrossRef]
10. Wei, W.; Li, L.; Shi, W.; Liu, J.P. Ultrasonic imaging recognition of coal-rock interface based on the improved variational mode decomposition. *Measurement* **2021**, *170*, 108728. [CrossRef]
11. Xu, D.; Wu, Y. Improved YOLO-V3 with DenseNet for multi-scale remote sensing target detection. *Sensors* **2020**, *20*, 4276. [CrossRef]
12. Cui, Z.; Li, K.; Gu, L.; Su, S.; Gao, P.; Jiang, Z.; Qiao, Y.; Harada, T. You Only Need 90K Parameters to Adapt Light: A Light Weight Transformer for Image Enhancement and Exposure Correction. *arXiv* **2022**, arXiv:2205.14871.
13. Wang, Q.; Cheng, M.; Huang, S.; Cai, Z.; Zhang, J.; Yuan, H. A deep learning approach incorporating YOLO v5 and attention mechanisms for field real-time detection of the invasive weed Solanum rostra-tum Dunal seedlings. *Comput. Electron. Agric.* **2022**, *199*, 107194. [CrossRef]
14. Jiang, L.; Peng, G.; Xu, B.; Lu, Y.; Wang, W. Foreign object recognition technology for port transportation channel based on automatic image recognition. *EURASIP J. Image Video Process.* **2018**, *2018*, 147. [CrossRef]
15. Zhang, Z.; Su, X.; Ding, L.; Wang, Y. Multi-scale image segmentation of coal piles on a belt based on the Hessian matrix. *Particuology* **2013**, *11*, 549–555. [CrossRef]
16. Saran, G.; Ganguly, A.; Tripathi, V.; Kumar, A.A.; Gigie, A.; Bhaumik, C.; Chakravarty, T. Multi-modal imaging-based foreign particle detection system on coal conveyor belt. *Trans. Indian Inst. Met.* **2022**, *75*, 2231–2240. [CrossRef]
17. Tu, L.; Zhong, S.; Peng, Q. Moving object detection method based on complementary multi resolution background models. *J. Cent. South Univ.* **2014**, *21*, 2306–2314. [CrossRef]
18. Lins, R.G.; Givigi, S.N. Automatic crack detection and measurement based on image analysis. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 583–590. [CrossRef]

19. Ghasemi, Y.; Jeong, H.; Choi, S.H.; Park, K.B.; Lee, J.Y. Deep learning-based object detection in augmented reality: A systematic review. *Comput. Ind.* **2022**, *139*, 103661. [CrossRef]
20. Kaur, J.; Singh, W. Tools, techniques, datasets and application areas for object detection in an image: A review. *Multimed. Tools Appl.* **2022**, *81*, 38297–38351. [CrossRef] [PubMed]
21. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
22. Pu, Y.; Apel, D.B.; Szmigiel, A.; Chen, J. Image recognition of coal and coal gangue using a convolutional neural network and transfer learning. *Energies* **2019**, *12*, 1735. [CrossRef]
23. 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, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
24. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 91–99.
25. 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]
26. Ma, G.; Wang, X.; Liu, J.; Chen, W.; Niu, Q.; Liu, Y.; Gao, X. Intelligent detection of foreign matter in coal mine transportation belt based on convolution neural network. *Sci. Program.* **2022**, *2022*, 9740622. [CrossRef]
27. 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]
28. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2017.08430.
29. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
30. Cui, Z.; Li, K.; Gu, L.; Su, S.; Gao, P.; Jiang, Z.; Qiao, Y.; Harada, T. Illumination adaptive transformer. *arXiv* **2022**, arXiv:2205.14871.
31. Yang, X.; Zhang, G.; Yang, X.; Tang, J.; He, T.; Yan, J. Detecting rotated objects as gaussian distributions and its 3-d generalization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 4335–4354. [CrossRef]
32. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer International Publishing: Cham, The Netherlands; pp. 21–37.
33. Han, J.; Ding, J.; Li, J.; Xia, G.S. Align deep features for oriented object detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5602511. [CrossRef]
34. Guo, Z.; Liu, C.; Zhang, X.; Jiao, J.; Ji, X.; Ye, Q. Beyond Bounding-Box: Convex-hull Feature Adaptation for Oriented and Densely Packed Object Detection. In Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 8792–8801.

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.