algorithms

MDPI

Article

Research on Gangue Detection Algorithm Based on Cross-Scale Feature Fusion and Dynamic Pruning

Haojie Wang, Pingqing Fan *, Xipei Ma and Yansong Wang

School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China; wanghaojieedu@163.com (H.W.); m310121433@sues.edu.cn (X.M.); m310121411@sues.edu.cn (Y.W.)

* Correspondence: fpq@sues.edu.cn

Abstract: The intelligent identification of coal gangue on industrial conveyor belts is a crucial technology for the precise sorting of coal gangue. To address the issues in coal gangue detection algorithms, such as high false negative rates, complex network structures, and substantial model weights, an optimized coal gangue detection algorithm based on YOLOv5s is proposed. In the backbone network, a feature refinement module is employed for feature extraction, enhancing the capability to extract features for coal and gangue. The improved BIFPN structure is employed as the feature pyramid, augmenting the model's capability for cross-scale feature fusion. In the prediction layer, the ESIOU is utilized as the bounding box regression loss function to rectify the misalignment issue between predicted and actual box angles. This approach expedites the convergence speed of the network while concurrently enhancing the accuracy of coal gangue detection. Channel pruning is implemented on the network to diminish model computational complexity and weight, consequently augmenting detection speed. The experimental results demonstrate that the refined YOLOv5s coal gangue detection algorithm outperforms the original YOLOv5s algorithm, achieving a notable accuracy enhancement of 2.2% to reach 93.8%. Concurrently, a substantial reduction in model weight by 38.8% is observed, resulting in a notable 56.2% increase in inference speed. These advancements meet the detection requirements for scenarios involving mixed coal gangue.

Keywords: gangue detection; YOLOv5s; feature extraction; loss function; channel pruning

Citation: Wang, H.; Fan, P.; Ma, X.; Wang, Y. Research on Gangue Detection Algorithm Based on Cross-Scale Feature Fusion and Dynamic Pruning. *Algorithms* **2024**, *17*, 79. https://doi.org/10.3390/a17020079

Academic Editors: Xiao Huang and Frank Werner

Received: 11 December 2023
Revised: 17 January 2024
Accepted: 7 February 2024
Published: 13 February 2024

1. Introduction

In the course of coal mining operations, there is an admixture of solid waste, represented in the form of gangue. Belt conveyors, serving as indispensable apparatus for coal transportation, necessitate efficient methodologies for the separation of coal gangue [1,2]. In complex environments characterized by low illumination, the presence of stacked coal gangue and uneven shapes and sizes, the expeditious and precise identification of coal gangue on conveyor belts is imperative for effective problem resolution.

The conventional image recognition algorithms [3–6] are limited in robustness and generality, coupled with the complexity of classifiers, rendering them inadequate for meeting the demands of swift and efficient target detection. With the advancement of artificial intelligence, coal gangue recognition technology based on deep learning has emerged as the predominant research methodology. Within the domain of deep learning methodologies, the YOLO (You Only Look Once) series algorithms [7–10] represent a class of regression-based single stage approaches, distinguished by their elevated detection performance. Zhang et al. [11] adopted mosaic data enhancement, cosine annealing, and label smoothing to optimize the YOLOv4 algorithm, thereby enhancing the efficiency of coal gangue detection. Guo et al. [12] employed YOLOv5 for coal gangue identification, integrated channel attention within the network, utilized the Acon activation function to enhance the network's adaptability in the nonlinear layer, and applied the refined algorithm on mobile devices. Shang et al. [13] proposed the adoption of the SimAM

attention mechanism in the YOLOv5 network header. They replaced the original backbone network with GhostNet to reduce the network parameters and enhance the detection accuracy of coal gangue. Xue et al. [14] proposed the utilization of ResNet18 as the backbone feature network for YOLOv3 to detect coal gangue, thereby reducing the feature size and enhancing the detection speed. Wang et al. [15] proposed the integration of decoupling heads into the YOLOv5 network to segment the detection task into a classification task and a regression task, with the aim of enhancing the accuracy of coal gangue detection.

Attributed to the similarity between the foreground of coal gangue on the conveyor belt under low illumination conditions and the background, the recognition difficulty is pronounced. Existing coal gangue detection methods still exhibit issues such as missed detections, false negatives, slow detection speeds and large model sizes, hindering their deployment on robotic arms for subsequent grasping and sorting operations. This paper proposes an enhanced coal gangue recognition algorithm based on the YOLOv5s architecture. Firstly, the ConvNext module is employed to optimize the feature extraction network, enhancing its capability to perceive small targets. Secondly, the fusion method of cross-scale features is refined to enhance the network's detection capability for coal gangue. Thirdly, a novel localization loss function is formulated to enhance the precision of coal gangue detection in localization. Finally, the dynamic multitask channel pruning algorithm is employed and redundant channels are removed to lightweight the model, thereby improving the speed of coal gangue detection.

2. Optimized Detection Algorithm for Coal Gangue Based on YOLOv5s

The YOLOv5 network primarily consists of four parts: Input, Backbone, Neck, and Prediction. Among various versions of YOLOv5, the YOLOv5s model is characterized by its smaller model weight and faster inference speed. In this study, YOLOv5s version 7.0 is selected as the foundational model.

2.1. Feature Extraction Optimisation for Enhanced Small Target Detection

Given the prevalence of small and stacked targets within the coal gangue dataset, the feature extraction process in YOLOv5s tends to result in the loss of features related to these small target instances. To address the issue of insufficient feature extraction for such smaller targets, we introduced a refinement in the Backbone section by incorporating a dedicated pure convolutional module referred to as ConvNext [16]. The ConvNext network structure is shown in Figure 1.

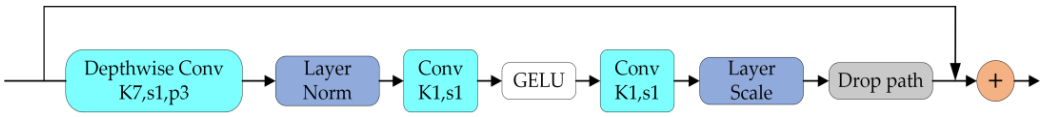


Figure 1. Structure of ConvNext.

The ConvNext network structure has a convolutional kernel in an inverted bottleneck layer configuration. It leverages a 7×7 large convolutional kernel along with depthwise separable convolution operations to extract features from each channel, thereby enhancing computational efficiency. Regularization is only applied once in the second layer, reducing the overall usage of the regularization functions. Additionally, BatchNorm is replaced with Layer Normalization, leading to an improvement in the model's accuracy. Expanding the number of channels to four times the original using 1×1 convolutional kernels and the Gaussian Error Linear Unit (GELU) activation function, the network is better able to capture nonlinear relationships, thereby enhancing the model's expressive power. Subsequently, employing 1×1 convolutional kernel operations and Layer Scale normalization, followed by Drop Path processing, involves randomly removing some connections within the neural network to enhance generalization capabilities. Finally, the processed features are concatenated with the untreated input feature map to obtain the output feature map.

The feature extraction structure enables more comprehensive feature extraction, reducing the impact of occlusion and background interference, effectively minimizing the loss of feature information for small target coal and gangue.

2.2. Cross-Scale Feature Fusion Based on Lightweighted Weighted Feature Pyramids

In the process of feature fusion, the significance of different features varies. YOLOv5s employs a multi-scale feature fusion approach with Feature Pyramid Network (FPN [17]) and Path Aggregation Network (PAN [18]). Constructing a bottom-to-top Path Aggregation Network allows the network to fuse semantic information from both low-level and high-level features. However, during the feature concatenation process, the consideration of feature weight is not taken into account. In the Weighted Bidirectional Feature Pyramid Network (BiFPN [19]), cross-scale connections and weighted feature fusion are employed, assigning appropriate weight values to features at different levels. Figure 2a illustrates the structure of the BiFPN.

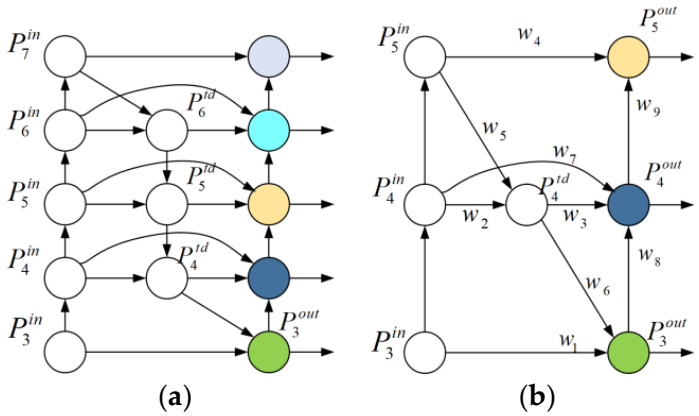


Figure 2. BiFPN network diagram before and after improvement. (**a**) Structure of BiFPN. (**b**) Improved BiFPN-tiny.

This paper further optimizes the BiFPN network as follows. Firstly, YOLOv5s has only three distinct scale output nodes, whereas BiFPN has five output nodes. To integrate the BiFPN network into YOLOv5s, it is imperative to reduce the number of BiFPN network output nodes to match the three nodes of YOLOv5s. Secondly, during the process of feature fusion, nodes that do not contribute to the fused features have a negligible impact on the output results. By reducing the number of nodes, the parameter count is decreased. Therefore, nodes between P_3 and P_5 were excluded. Finally, by adding connections between input nodes and output nodes within the same size dimension, the network achieves more comprehensive feature fusion without significantly increasing computational load. Therefore, the cross-scale connection for the P_4 layer is retained.

During the feature fusion process, the resolution of each layer's input features varies, and they correspond to different weights. Moreover, the magnitude of weight values is unbounded, which may lead to gradient explosions, preventing model convergence. To address this, Softmax normalization is employed to ensure weight values lie within the range of [0–1]. This enhances computational efficiency and enables rapid normalization fusion for bidirectional cross-scale connections. The weight normalization formula is represented as Equation (1).

$$O = \sum_I \frac{w_i}{\epsilon + \sum_j w_j} \cdot I_i, \tag{1}$$

where $\epsilon = 0.001$, the ReLU activation function, is employed to ensure that $w_i \geq 0$, I_i denotes the input features.

The improved Weighted Bidirectional Feature Pyramid Network is illustrated in Figure 2b.

In Figure 2b, the formulas for each output node are shown in Equations (2)–(5).

$$P_3^{out} = Conv\left(\frac{w_1 \cdot P_3^{in} + w_6 \cdot Resize\left(P_6^{td}\right)}{w_1 + w_6 + \epsilon}\right) \tag{2}$$

$$P_4^{td} = Conv\left(\frac{w_2 \cdot P_4^{in} + w_5 \cdot Resize\left(P_5^{in}\right)}{w_1 + w_2 + \epsilon}\right) \tag{3}$$

$$P_4^{out} = Conv\left(\frac{w_7 \cdot P_4^{in} + w_3 \cdot P_4^{td} + w_8 \cdot Resize\left(P_3^{out}\right)}{w_7 + w_3 + w_8 + \epsilon}\right) \tag{4}$$

$$P_5^{out} = Conv\left(\frac{w_4 \cdot P_5^{in} + w_9 \cdot Resize\left(P_4^{out}\right)}{w_4 + w_9 + \epsilon}\right) \tag{5}$$

where P_i^{in} is the input feature of the layer, P_i^{td} is the middle feature of the i layer, P_i^{out} is the output feature of the i layer after feature fusion, Resize represents either upsampling or downsampling operations, and Conv denotes a convolution operation.

2.3. Optimization of the Loss Function

The foreground of coal gangue images on the conveyor belt bears a resemblance to the background, resulting in increased difficulty of detection. The quality of anchor boxes significantly affects the detection performance. The image prediction within the head layer, the IoU (Intersection over Union), in the bounding box localization loss function represents the ratio of the intersection area to the union area between the predicted box and the ground truth box. The loss function in YOLOv5s is based on the CIoU (Complete IoU) loss function, which takes into account three aspects: the loss of overlap between the predicted box and the ground truth box, the loss of distance between the centers of the predicted box and ground truth box, and the loss of aspect ratio between the predicted box and ground truth box. The loss function is formulated as shown in Equations (6)–(8).

$$L_{CIoU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha\nu \tag{6}$$

$$\nu = \frac{4}{\pi^2}\left(\arctan\frac{\omega^{gt}}{h^{gt}} - \arctan\frac{\omega}{h}\right)^2 \tag{7}$$

$$\alpha = \frac{\nu}{(1 - IoU) + \nu} \tag{8}$$

where b and b^{gt} represent the center points of the prediction box and the ground truth box, respectively. c is the length of the minimum diagonal of the external rectangle, ρ^2 represents the Euclidean distance between the centroids of the prediction box and the ground truth box, α is the loss balance factor, ν is the normalization of the difference between the predicted and ground truth box aspect ratios, and w, h, w^{gt}, h^{gt} are the widths and heights of the predicted and ground truth boxes, respectively.

The CIoU Loss does not account for the angular loss between the predicted and ground truth boxes, leading to slow convergence in anchor box regression. To enhance detection accuracy and accelerate convergence speed, this paper builds upon this by introducing the ESIoU (Enhanced Structured IoU) loss function [20,21], the schematic diagram of the loss function is illustrated in Figure 3.

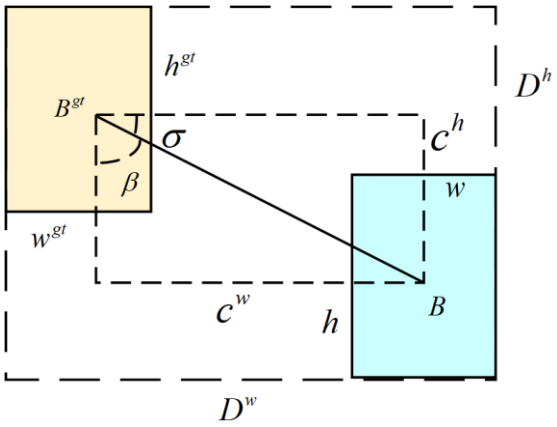


Figure 3. Schematic of the ESIoU Loss.

The ESIoU loss function takes into account the IoU loss, angle loss, distance loss, shape loss, width loss, and height loss between the predicted and ground truth boxes. The ESIoU loss function is formulated as shown in Equation (9).

$$L_{ESIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} + \frac{\rho^2(w, w^{gt})}{(D^w)^2} + \frac{\rho^2(h, h^{gt})}{(D^h)^2} \tag{9}$$

The distance loss Δ is formulated as shown in Equation (10).

$$\Delta = \sum_{t=x,y} \left(1 - e^{-(2-\Lambda)\rho_t}\right) \tag{10}$$

where $\rho_x = \left(\frac{B^{gt} - B}{C_w}\right)^2$, $\rho_y = \left(\frac{B^{gt} - B}{C_h}\right)^2$, C_w, and C_h, respectively, denote the width and height of the minimum bounding rectangle of the centroids for the predicted box and the ground truth box.

The angle loss Λ is formulated as shown in Equation (11).

$$\Lambda = \cos(2(\arcsin(\frac{c_h}{\sigma}) - \frac{\pi}{4})) \tag{11}$$

where c_h represents the difference in height between the predicted box and the ground truth box centroids, and σ denotes the separation between the centroids of the predicted box and ground truth box.

The shape loss Ω is expressed as shown in Equation (12).

$$\Omega = \sum_{t=w,h} \left(1 - e^{-\omega_t}\right)^{\theta} \tag{12}$$

where $\omega_w = \frac{|w - w^{gt}|}{max(w, w^{gt})}$, $\omega_h = \frac{|h - h^{gt}|}{max(h, h^{gt})}$, (w, w^{gt}), and (h, h^{gt}) represent the width and height of the predicted box and the ground truth box, respectively. θ represent the attention level on the shape difference between the predicted and ground truth boxes; θ is set to 4.

In the width and height losses, D^w and D^h, respectively, denote the width and height of the minimum bounding rectangle for the predicted and ground truth boxes.

Based on the optimized ESIoU loss function described above, the convergence speed of the predicted boxes was accelerated. In the coal and gangue detection task, the optimized approach alleviated missed detection cases attributed to unclear features, leading to an enhancement in the detection accuracy of coal and gangue.

3. Model Lightweighting Based on Global Dynamic Channel Pruning

3.1. Dynamic Model Pruning

In order to better apply the coal and gangue detection algorithm to hardware devices, considerations must be given to the detection speed, computational load, and model weight. Therefore, this paper employs the global channel pruning algorithm to lightweight the detection model. The Performance-Aware Global Channel Pruning (PAGCP) [22], based on performance-aware approximation for multitask models, achieves model compression by pruning globally redundant filters. During the pruning process, the establishment of a channel saliency indicator model analyzes the correlation and interaction of different channel-layer convolution kernels, measuring the importance of filters within and between layers. In the network architecture, taking the i-th layer and the j-th convolutional kernel as an example, the saliency index is expressed by Equation (13).

$$S(x;\theta_{ij}) = \left\| f(x;\theta_{ij}) - f(x;\theta_{ij}=0) \right\|_r \tag{13}$$

where θ_{ij} is a binary mask, and the convolutional kernel is deleted only when $\theta_{ij} = 0$; r is the order of the norm. The x represents the input of the saliency indicator and calculates the Euclidean distance in the convolutional kernel's Euclidean space. A larger Euclidean distance indicates greater importance for that convolutional kernel.

Currently, global channel pruning primarily adopts a static pruning order, as illustrated in Figure 4a. The outcomes of static pruning depend on the chosen pruning strategy, introducing increased challenges to model optimization. This paper employs dynamic pruning by calculating the floating-point operations per second (FLOPs) for each layer of the model: First, the FLOPs contributions are dynamically sorted, and layers with significant FLOPs contributions but minimal impact on model performance are selected for pruning, retaining layers with smaller FLOPs contributions but significant improvements in model performance. Then, we determined the pruning ratios for all layers and reevaluated the contribution of each layer's pruning ratio to FLOPs, continuously adjusting the pruning order dynamically, as illustrated in Figure 4b. Finally, after pruning, the model's computational load and weight were reduced. While maintaining detection accuracy, these enhancements resulted in an improved inference speed and reduced deployment challenges on hardware devices.

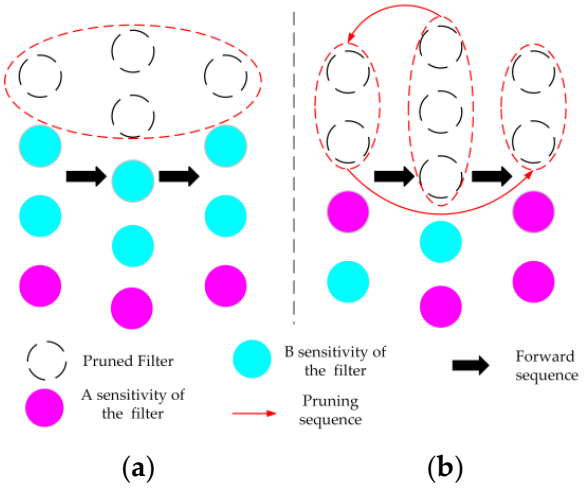


Figure 4. Comparing static and dynamic trimming. (**a**) Static pruning. (**b**) PAGCP dynamic pruning.

3.2. Improved Network Model for Gangue Detection Algorithm

This study refines the coal and gangue detection algorithm based on the YOLOv5s network model. The modified network structure of the improved gangue detection algorithm is illustrated in Figure 5.

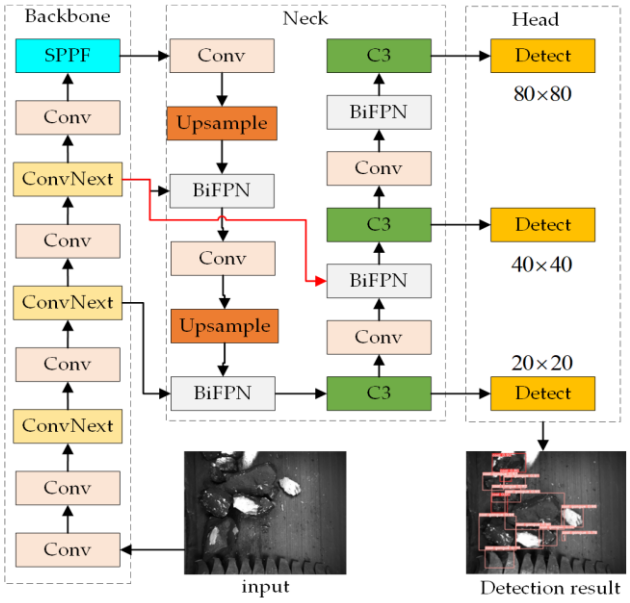


Figure 5. Improved network model of the algorithm for the detection of coal gangue.

Firstly, to address the issue of suboptimal detection for small targets in coal and gangue, the feature extraction in the Backbone network is replaced with the ConvNext module, enhancing the network's ability to perceive small targets. This replacement enhances the network's ability to perceive small targets, and, furthermore, aims to incorporate more semantic and low-level information into the feature fusion process. This is achieved by assigning appropriate weights to the features, facilitating rapid and efficient cross-scale fusion to enhance detection performance for coal and gangue, thereby addressing instances of missed detections. Subsequently, to address the issue of mismatched angles between predicted and ground truth bounding boxes, the localization loss function adopts the ESIOU loss function to expedite network convergence. This approach enhances localization precision, and improves the overall accuracy of the model in detecting coal and gangue. Finally, global channel pruning is applied to eliminate redundant channels, leading to model lightweighting, and thereby improving the inference speed.

4. Experimental Validation and Analysis

4.1. Parameter Settings and Experimentaldata

In this study, the experimental setup included an Intel® Core™ i9-10900X CPU @ 3.70 GHz, NVIDIA GeForce RTX 3090 GPU, and the Ubuntu 20.04.3 operating system. The deep learning framework used was PyTorch 1.11.3, programming language was Python 3.8, and CUDA version was 11.6. The experimental parameters included a batch size of 80, image size of 640, training epochs of 200, and a learning rate of 0.01. To ensure the comparability of training results, all algorithm comparisons were tested under the same set of parameters mentioned above.

The dataset was acquired from a video recorded by a conveyor belt at a coal factory. Non-identifying images were removed after image processing. The dataset consisted of 5000 images, with two identification targets: coal and coal gangue, respectively. The images were annotated using Labelme, which generates a JSON-format file. Subsequently, the JSON file was converted into a textual label. The datasets are partitioned into training, validation, and test sets in a ratio of 7:2:1, consisting of 3500 training samples, 1000 validation samples, and 500 test samples. The training process of the coal gangue detection network is depicted in Figure 6.

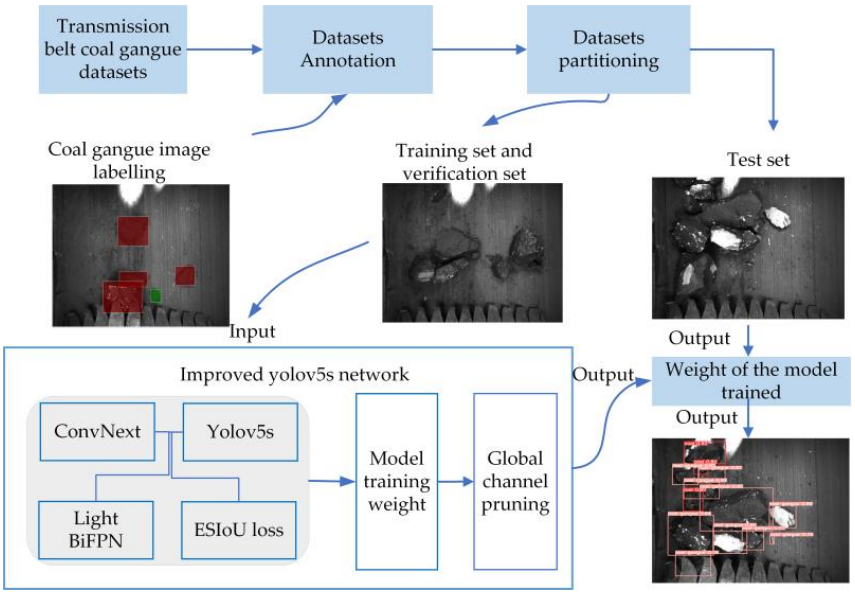


Figure 6. Network training flowchart for detecting coal gangue.

4.2. Evaluation Indicators

To validate the improved coal gangue detection algorithm's capability in detecting coal and coal gangue, this study primarily evaluates and analyzes precision (P), recall (R), mean Average Precision (mAP), model weight, and detection speed.

P refers to the proportion of coal or gangue correctly detected in the overall detection results, as shown in Equation (14).

$$P = \frac{TP}{TP + FP} \tag{14}$$

where *TP* represents the number of correctly detected coal or gangue instances, and *FP* represents the number of instances falsely identified as the target by the model.

R refers to the proportion of correctly predicted coal or gangue in the overall detection results, as shown in Equation (15).

$$R = \frac{TP}{TP + FN} \tag{15}$$

where *FN* represents the quantity of coal or gangue not detected by the model.

The *mAP* is the average precision across multiple categories, and the calculation formula is provided in Equations (16) and (17).

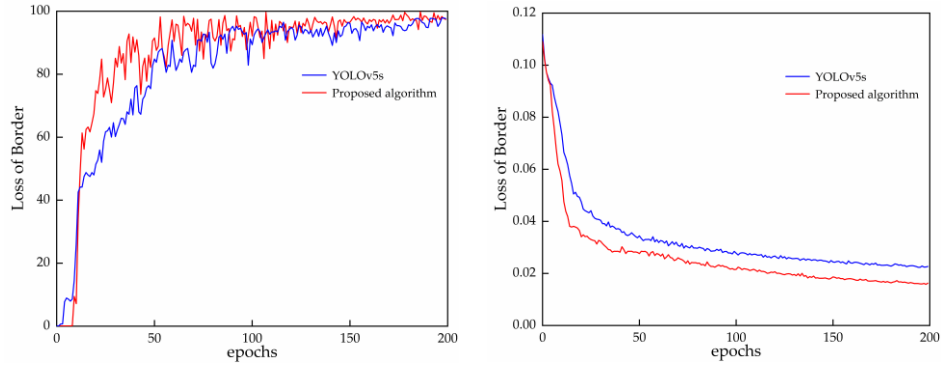
$$AP = \int_0^1 P(\mathrm{r})dr \tag{16}$$

$$mAP = \frac{\sum_{q=1}^{Q} AP(q)}{Q} \tag{17}$$

where *AP* is the average, and *Q* is the number of categories.

4.3. Detection Accuracy and Anchor Box Localization Loss Curves

Training was conducted on coal gangue data using the YOLOv5s model and the algorithm proposed in this paper. The results of comparing detection accuracy and anchor box localization loss values are shown in Figure 7.



(**a**) (**b**)

Figure 7. Comparison of detection accuracy and border loss. (**a**) Comparison chart for model accuracy. (**b**) Comparison diagram of model losses.

From Figure 7a, it can be observed that the precision of the proposed algorithm reaches above 80% after around 30 iterations, which is approximately 20% higher than the YOLOv5s during the same period. The convergence is achieved gradually, and by around 50 iterations, the algorithm converges rapidly. Figure 7b illustrates that the bounding box loss of our algorithm shows a significant decrease in loss value after 20 iterations, followed by a more gradual decline. The loss value is lower than YOLOv5s, indicating a faster convergence speed. The effectiveness of the improved algorithm was verified through the evaluation of detection accuracy and bounding box loss. The effectiveness of the improved algorithm was verified through the evaluation of detection accuracy and bounding box loss.

4.4. Model Pruning Experiments

Based on the proposed improved pruning algorithm in the above-mentioned study, pruning experiments were conducted on the detection model, and the results of channel comparison before and after pruning are shown in Figure 8. The model pruned channels are mainly concentrated in the feature fusion part, and the number of feature extraction channels is pruned less, so that the redundant channels with low impact on model performance are eliminated, and the number of parameters is reduced from 7.02 M to 4.24 M on the basis of maintaining the model performance, reducing the complexity of the model. It makes the model detection more efficient.

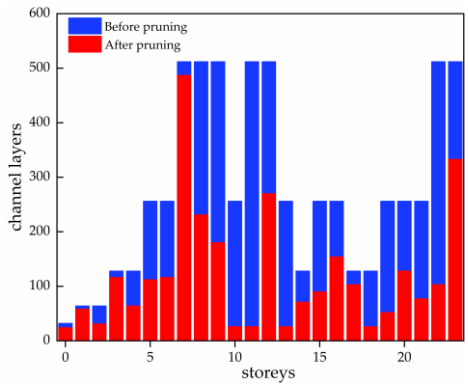


Figure 8. Diagram of model pruning channel number change.

4.5. Ablation Experiments

In order to validate the effectiveness of the improved coal and gangue detection model based on YOLOv5s and ensure the validity of the training results, various ablation experiments were conducted for performance analysis. The results of the ablation experiments are shown in Table 1.

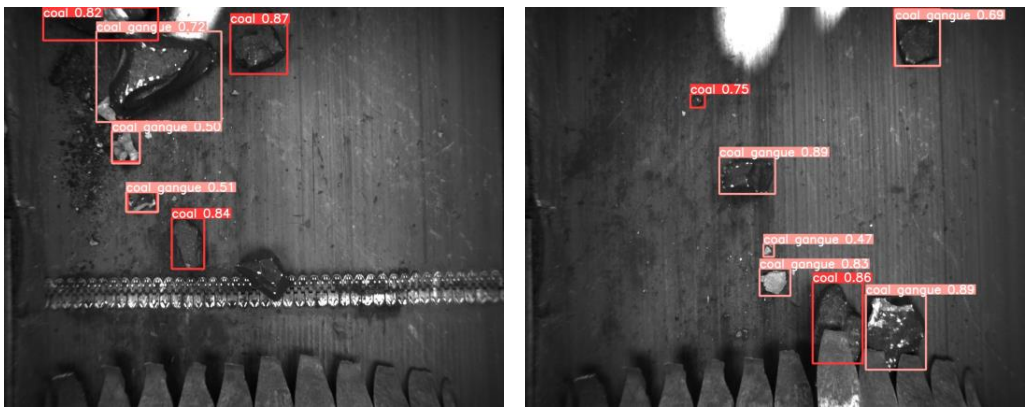
Table 1. Ablation experiments.

Experiment	ConvNext	BiFPN-Tiny	ESIoU	PAGCP	P/%	R/%	mAP/%	Weight/Mb	FLOPS/G
A	×	×	×	×	91.3	91.6	83.7	14.4	16.0
B	√	×	×	×	91.8	93.3	84.6	14.6	16.7
C	×	√	×	×	92.4	93.1	83.8	14.8	16.5
D	×	×	√	×	93.3	91.7	84.2	14.7	16.1
E	√	√	×	×	93.6	93.2	84.8	14.8	17.2
F	√	√	√	×	93.8	93.0	85.6	14.9	17.3
G	√	√	√	√	93.5	91.9	85.5	8.8	9.7

Experiment A represents the original YOLOv5s algorithm, with a model accuracy of 91.3%, model weight of 14.4 MB, and computational complexity of 16.0 G. Experiment B introduces the ConvNext Block to YOLOv5s, resulting in a 1.7% increase in model recall and a 0.5% increase in accuracy, effectively reducing the false-negative rate for small targets. In Experiment C, after improving BiFPN, the model's accuracy and recall increased by 1.1% and 1.5%, respectively. In Experiment D, by changing the loss function to ESIOU, the model's accuracy increased by 2%. In Experiments E and F, the improvements were combined, resulting in a slight increase in model weight. However, accuracy, recall rate, and mAP all showed improvement. In Experiment G, compared to the original YOLOv5s model, the accuracy was improved and the model weight was reduced by 38.8%, with a decrease in parameter count by 39.3%. Pruning redundant channels in the model ensured performance while reducing its size. The experiments demonstrated that the improved coal gangue detection model based on YOLOv5s proposed in this paper effectively enhances the detection results of coal gangue in harsh environments.

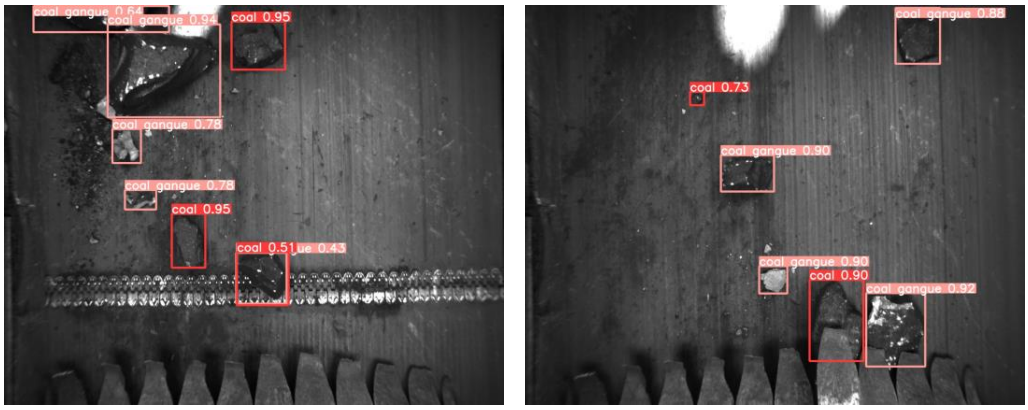
4.6. Visualisation and Analysis of Test Results

To validate the feasibility of the improved YOLOv5s model, a subset of images from the test set was selected for testing. The test results are illustrated in Figure 9. Figure 9a depicts the detection results before the improvement, while Figure 9b showcases the enhanced detection performance after the improvements. In Figure 9a, there are instances of missed detections and lower accuracy in target identification. However, in Figure 9b, smaller targets are successfully identified, and the detection accuracy of most targets has significantly improved. Therefore, the improved YOLOv5s model demonstrates low false-negative rates and high accuracy, meeting the detection requirements for coal and gangue.



(**a**)

Figure 9. *Cont.*



(**b**)

Figure 9. Visual comparison of model detected results. (**a**) Detection results of yolov5s. (**b**) Improved detection results of yolov5s.

4.7. Comparative Experiments

To further validate the improved algorithm's detection performance on coal and gangue, comparative experiments were conducted using a self-created dataset specifically designed for coal and gangue detection. The experiments involved comparing the algorithm proposed in this paper with current mainstream object detection algorithms. Analyzing Table 2 reveals that, compared to SSD, Faster R-CNN, YOLOv3, and YOLOv5s, the algorithm proposed in this paper shows an improvement in mAP by 20.4%, 3.7%, 0.7%, and 1.8%, respectively. The two-stage object detection algorithm, Faster R-CNN, has the highest model weight, and the detection time for a single image is the longest, reaching 76.4 ms. Therefore, it exhibits a slower inference speed. For the remaining single-stage object detection algorithms, the shortest detection speed was 15.3 ms. Despite being faster than Faster R-CNN, there is still a need for improvement to meet the real-time detection requirements for coal and gangue. The algorithm proposed in this paper achieves a detection time of 9.8 ms per single image, showing a 56.2% improvement in inference speed. The model weight is reduced by 35.7% to 8.8 Mb. Therefore, the improved model exhibits significant advantages in both detection speed and accuracy, showcasing good performance while also reducing subsequent hardware deployment costs.

Table 2. Comparison of results of different algorithms.

Algorithm	Weight/Mb	mAP/%	Inference Time/ms
SSD	94.8	65.1	34.9
Faster-RCNN	206.7	81.8	67.4
YOLOv3	123.5	84.8	40.8
YOLOv5s	13.7	83.7	15.3
Textual algorithm	8.8	85.5	9.8

5. Conclusions

This paper proposes an improved algorithm, YOLOv5s, for coal gangue detection on a conveyor belt. The aim is to address challenges such as low contrast in coal gangue, low recognition accuracy, susceptibility to false negatives, and difficulties in hardware deployment due to the large weight of existing models.

In the YOLOv5s backbone network, ConvNext is utilized to ensure more comprehensive feature extraction for small targets. Optimizing the feature pyramid network, the improved BiFPN structure is tailored to better align with the YOLOv5s algorithm, thereby enhancing the multiscale feature fusion in the network. During anchor box detection, the ESIOU loss function is employed to expedite the convergence speed of the model. Finally, dynamic channel pruning is applied to eliminate redundant channels, reducing both the

model's weight and computational load, thereby achieving network lightweighting. The experimental results indicate a notable improvement in the proposed coal and gangue detection algorithm compared to the original one. The model weight is 8.8 Mb, reduced by 38.8%, achieving a detection accuracy of 93.5%. The detection time for a single image is 9.8 ms, with a 56.2% increase in inference speed. The algorithm meets the recognition and detection requirements for coal and gangue on industrial conveyor belts, facilitating the subsequent deployment for the application of robotic arms to grab and sort.

Author Contributions: Conceptualization, H.W. and P.F.; methodology, H.W. and P.F.; software, H.W.; validation, H.W.; formal analysis, H.W.; investigation, H.W.; resources, P.F.; data curation, H.W.; writing—original draft preparation, H.W.; writing—review and editing, H.W.; visualization, H.W.; supervision, X.M.; project administration, X.M.; funding acquisition, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shanghai Nature Foundation, grant number 21ZR1425800, and the National Nature Foundation, grant number 5217237.

Data Availability Statement: Data available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, G.; Ren, H.; Zhao, G.; Zhang, D.; Wen, Z.; Meng, L.; Gong, S. Research and practice of intelligent coal mine technology systems in China. *Int. J. Coal Sci. Technol.* **2022**, *9*, 24. [CrossRef]
2. Ma, H.; Wei, X.; Wang, P.; Zhang, Y.; Cao, X. Multi-arm global cooperative coal gangue sorting method based on improved Hungarian algorithm. *Sensors* **2022**, *22*, 79–87. [CrossRef] [PubMed]
3. Xie, Y.; Yu, S.; Huang, Z. Foreign matter detection of coal conveying belt based on machine vision. In Proceedings of the 2021 2nd International Conference on Computer Science and Management Technology (ICCSMT), Shanghai, China, 12–14 November 2021; pp. 293–296.
4. Zhao, X.; Li, X.; Yin, L.; Feng, W.; Zhang, N.; Zhang, X. Foreign body recognition for coal mine conveyor based on improved PCANSet. In Proceedings of the 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP), Xi'an, China, 23–25 October 2019; pp. 1–6.
5. Ye, T.; Zheng, Z.; Li, Y.; Zhang, X.; Deng, X.; Ouyang, Y.; Zhao, Z. An adaptive focused target feature fusion network for detection of foreign bodies in coal flow. *Int. J. Mach. Learn. Cybern.* **2023**, *14*, 2777–2791. [CrossRef]
6. Luo, Q.; Wang, S.; Li, X.; He, L. Recognition of coal and gangue based on multi-dimensional gray gradient feature fusion. *Energy Sources Part A Recovery Util. Environ. Eff.* **2022**, *44*, 8060–8076. [CrossRef]
7. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
8. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
9. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804. 027 67.
10. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
11. Luo, B.; Kou, Z.; Han, C.; Han, C.; Wu, J. A Faster and Lighter Detection Method for Foreign Objects in Coal Mine Belt Conveyors. *Sensors* **2023**, *23*, 62–76. [CrossRef] [PubMed]
12. Guo, Y.; Zhang, Y.; Li, F.; Cheng, G. Research of coal and gangue identification and positioning method at mobile device. *Int. J. Coal Prep. Util.* **2023**, *43*, 691–707. [CrossRef]
13. Shang, D.; Yang, Z.; Lv, Z. Recognition of coal and gangue under low illumination based on SG-YOLO model. *Int. J. Coal Prep. Util.* **2023**, 1–16. [CrossRef]
14. Xue, G.; Li, S.; Hou, P.; Gao, S.; Tan, R. Research on lightweight Yolo coal gangue detection algorithm based on resnet18 backbone feature network. *Internet Things* **2023**, *22*, 100762. [CrossRef]
15. Wang, S.; Zhu, J.; Li, Z.; Sun, X.; Wang, G. Coal Gangue Target Detection Based on Improved YOLOv5s. *Appl. Sci.* **2023**, *13*, 11220. [CrossRef]
16. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986.
17. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
18. 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; pp. 8759–8768.
19. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.

20. Zhang, Y.F.; Ren, W.; Zhang, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *arXiv* **2021**, arXiv:2101.08158. [CrossRef]
21. Gevorgyan, Z. SIoU loss: More powerful learning for bounding box regression. *arXiv* **2022**, arXiv:2205.12740.
22. Ye, H.; Zhang, B.; Chen, T.; Fan, J.; Wang, B. Performance-aware Approximation of Global Channel Pruning for Multitask CNNs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 10267–10284. [CrossRef] [PubMed]