Engineering and Technology Journal e-ISSN: 2456-3358
Volume 09 Issue 04 April-2024, Page No.- 3791-3802
DOI: 10.47191/etj/v9i04.18, I.F. – 7.136

Algorithm For Coal Mine Rock Foreign Object Detection Based on Enhanced Yolov8

Shuai Yang[1], Qunpo Liu[2]
[1]School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo, Henan, China; &
[2]School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo, Henan, China; &

ABSTRACT: In addressing the challenge of suboptimal detection precision in coal mine rock foreign object detection due to complex environments and variable object scales, this paper proposes a coal mine rock foreign object detection algorithm based on YOLOv8. Initially, the incorporation of the BiFormer attention mechanism is advocated to refine the backbone network, augmenting the model's attention towards pivotal information regions, consequently enhancing localization and feature extraction capabilities. Secondly, a lightweight Content-Aware Recurrent Affine Feature Extraction (CARAFE) operator is utilized within the neck architecture to effectively capture and preserve intricate features at lower hierarchical levels. Finally, Wise-IoU v3 is adopted as the bounding box regression loss for the proposed algorithm, coupled with a prudent gradient allocation approach, thereby enhancing the model's localization capabilities. Empirical findings illustrate that compared to baseline algorithms, the proposed algorithm has fewer parameters, with an average mAP improvement of 2.5%, and a detection speed increase of 2fps/s.

KEYWORDS: Foreign object detection; YOLOv8; BiFormer attention mechanism; CARAFE; Wise-IoU v3;

I. INTRODUCTION

The advancement of science and technology has shifted attention towards the establishment of intelligent coal mines, signifying a pivotal aspect of the high-quality development paradigm within China's coal industry in the contemporary era[1,2].At present, the detection of foreign bodies in coal mines in China is mainly artificial and mechanical. In the detection of artificial coal mine, several challenges persist, including elevated workloads and suboptimal working conditions, low detection rate and harm to personal safety and health. In mechanical coal mine foreign body detection, there are problems such as environmental pollution, difficult maintenance of large equipment, and high investment cost[3]. Therefore, wise detection of foreign objects in coal mines has emerged as a focal point in contemporary research endeavours.

In recent years, the rapid development of target detection technology, it can not only achieve the classification of the object, but also obtain the relevant position of the object, and complete the accurate detection of the location. At present, object detection algorithms have two major development processes. The first stage is to generate candidate frames, then obtain features through CNN, and then combine classifiers with boundary box regression to complete the classification and positioning of objects of interest, such as regional convolutional neural network (Region-CNN, R-CNN)[4]. Further developed algorithms include SPPNet[5], FastR-CNN[6], FasterR-CNN[7], etc. Since it costs a lot of computation to generate many anchor frames, some scholars proposed to directly use CNN regression to predict candidate frames. Therefore, a series of detection models with low computational cost, such as YOLOv1[8], YOLO9000[9], YOLOv3[10], YOLOv4[11], SSD[12] and CenterNet[13], were born in the second stage of algorithm development.

Gao Han et al.[14] designed a low-level feature enhancement and Transformer mechanism foreign object detection algorithm based on Transformer model for detecting foreign matter on a coal mine conveyor belt. Cao Xiangang[15] et al. proposed a detection method for coal foreign bodies based on cross-modal attention fusion to address the issue of insufficient feature extraction in image detection of foreign bodies caused by low contrast and mutual occlusion in raw coal flow during intelligent washing process. Cao Zhengyuan et al.[16] proposed an advanced detection technique for foreign objects within coal streams, utilizing a dual attention generation adversarial network. This method aims to enhance model classification accuracy significantly. Tang Jun et al.[17] proposed a real-time detection algorithm for foreign objects on belt conveyors, leveraging the Faster-YOLOv7 framework to enhance foreign object detection accuracy. Hao Shuai et al.[18] introduced a target detection algorithm based on YOLOv5, which incorporates a convolution block attention model. This method addresses the challenge of accurately detecting images of foreign objects on conveyor belts, influenced by factors such as coal dust interference, high-speed conveyor belt movement, and uneven illumination. Du Jingyi et al.[19] introduced an enhanced YOLOv3 model specifically designed for detecting foreign objects on coal mine belt conveyors. This improvement addresses the slower detection speeds observed in existing deep learning-based methods for foreign object detection on belt conveyors. Ren Zhiling et al.[20] proposed an improved CenterNet foreign body detection algorithm for coal belt to realize rapid and accurate identification

of foreign objects entering the coal belt during operation and prevent belt tearing.

To sum up, the deep learning-based coal mine foreign body detection technology has been developed and mature, and is widely used in various production and transportation scenarios for foreign body detection, so it can be used to build a coal mine foreign body detection model. However, during the production and transportation phases in coal mining operations, the environment is complicated and the types of foreign bodies are diverse, and the defect forms of different types are greatly different. Hence, there is a need to investigate and enhance the defect detection algorithm based on machine learning methods, so that it can detect foreign bodies of different sizes and shapes in coal mine with high accuracy and fast speed, and can replace manual sampling inspection, which has important practical significance.

II. COAL MINE FOREIGN BODY DETECTION ALGORITHM BASED ON IMPROVED YOLOV8

Foreign object detection represents a significant aspect of target detection, employing computer vision technology to classify and localize foreign objects within images or video streams. Considering that in the actual detection process, the scales of different metal foreign bodies vary greatly and the background is very similar to that of coal mine, which affects the algorithm's detection efficacy, an improved YOLOv8 based coal mine foreign bodies detection algorithm is proposed.

A. YOLOv8 network model

YOLOv8 represents the most recent iteration within the YOLO algorithm series, offering five fundamental variants: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x, depending on the operating environment. YOLOv8n is the fastest model with the smallest number of parameters, while YOLOv8x is the slowest but most accurate model. Compared with other algorithms in the YOLO family, the detection principle of YOLOv8 is most similar to that of YOLOv5 and YOLOv7, which are composed of the backbone network, the neck network and the first three main parts of the detection. extraction. Then the neck network strengthens feature fusion on the extracted features to obtain feature maps of three varying dimensions: large, medium, and small. Ultimately, the combined features are forwarded to the detection head for object recognition and localization. Test and output the final result. Fig1 illustrates the architecture of the YOLOv8 algorithm.

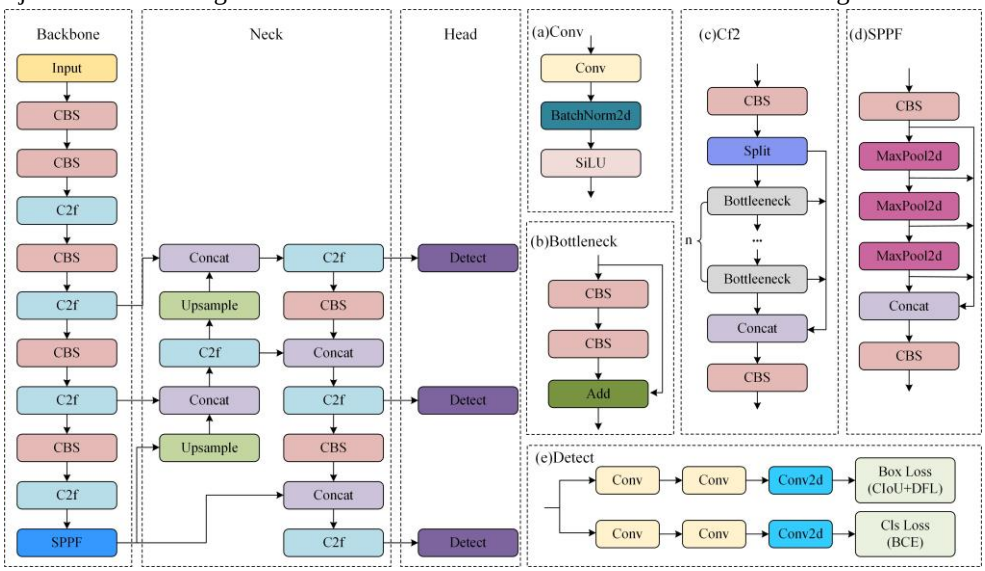


Fig1. Schematic diagram of YOLOv8 algorithm structure

B. Introducing the backbone network of BiFormer attention mechanisms

The user discusses integrating a dynamic sparse attention mechanism, the BiFormer module, into the model's backbone network to improve detection in coal mines, where foreign bodies are often obscured by ore dust, making backgrounds similar and challenging for detection models[21]. BiFormer efficiently identifies key and value pairs with significant relevance, reducing computational and storage overhead while enhancing the model's understanding of input content. The network architecture is illustrated in Fig2.

We incorporate the BiFormer dynamic sparse attention module into the YOLOv8 network model to improve foreign object detection's feature extraction capabilities. They highlighted the limitations of CNNs' local processing and discussed the Transformer's attention mechanism for global perception[22]. The BiFormer dynamic sparse attention force consists of BiFormer Block modules, with the BRA module as its core component, which filters irrelevant key-value pairs within the coarse feature map region using region-level directed graphs and fine-grained token-to-token attention during region association. The structures of the BiFormer Block module and BRA module are shown in Fig3.

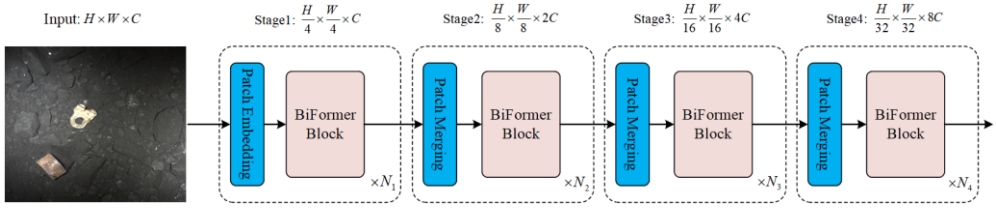


Fig 2. Structure diagram of Biformer dynamic sparse attention module

The BiFormer Block module employs a two-level routing approach, integrating DWConv for deep separable convolution to reduce model parameters and computational workload. Layer normalization (LN) contributes to faster training and improved model generalization, while the multi-layer perceptron (MLP) fine-tunes weights for specific areas, enhancing the model's attention to diverse features.

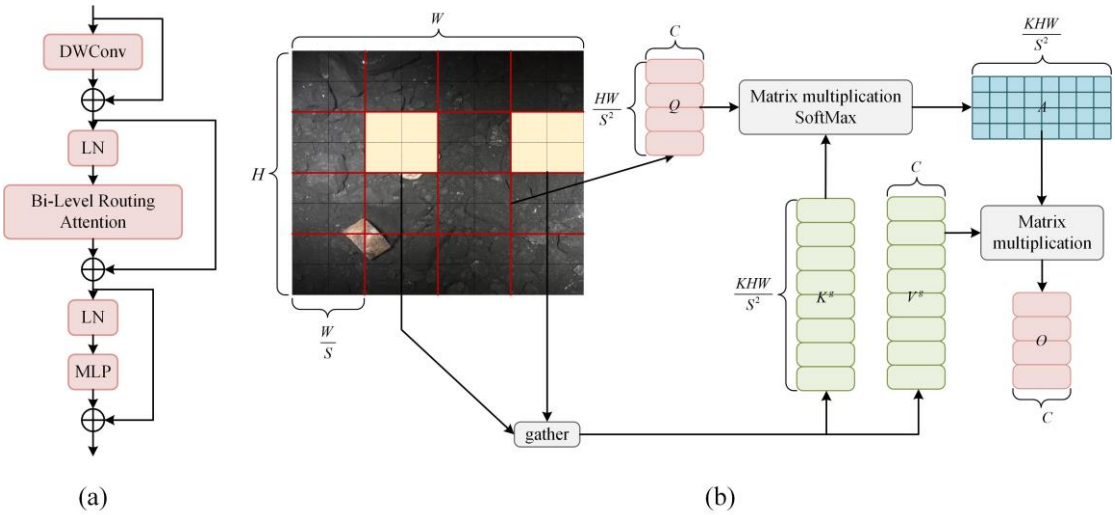


Fig3 (a) Structure diagram of BiFormer Block module (b) Structure diagram of BRA module

As depicted in Fig3-3(b), the initial feature illustration in Figure $X \in R^{H \times W \times C}$ comprises two sub-regions, with each encompassing three feature vectors. By modifying the configuration of X, we obtain $X^r \in R^{S^2 \times \frac{HW}{S^2} \times C}$. Subsequently, the eigenvectors undergo a linear transformation to yield the matrices Q, K, and V. The computational formula is expressed as follows:

$$Q = X^r W^Q \quad (1)$$

$$K = X^r W^k \quad (2)$$

$$V = X^r W^V \quad (3)$$

Next, by establishing a directed graph to pinpoint relevant regions within a given area, the attentional connection between these regions is established. The specific implementation procedure is as follows: region-wise processing of Q and V is conducted to acquire the region levels Q^r and $K^r \in R^{S^2 \times C}$. Subsequently, the dot product of Q^r and K^r is computed to yield the adjacency matrix $A^r \in R^{S^2 \times S^2}$, used for assessing inter-regional correlation as outlined in Eq (4):

$$A^r = Q^r (K^r)^T \quad (4)$$

The route index matrix $I^r \in N^{S^2 \times k}$ is derived by excluding the least pertinent tokens in A^r at the coarse-grained level and preserving only the top A^r most pertinent regions in A^r. This calculation is detailed in Eq (5):

$$I^r = topkIndex(A^r) \quad (5)$$

Following that, fine-grained token-to-token attention is applied. When considering a query within region i, attention is restricted to the k routing regions specified by $I^r_{(i,1)}, I^r_{(i,2)}, \ldots, I^r_{(i,k)}$. The process involves gathering all the Q and V tensors within these k regions to derive K^g and V^g. The computational formula is expressed as:

$$K^g = gather(K, I^r) \quad (6)$$

$$V^g = gather(V, I^r) \quad (7)$$

In the final step, the gathered K^g and V^g undergo attention processing, incorporating the local context enhancer $LCE(V)$ to produce the output tensor O. This formula is represented as Eq (8):

$$O = Attention(Q, K^g, V^g) + LCE(V) \quad (8)$$

We incorporated the BiFormer Block module into the backbone network for foreign object detection. This integration serves two main purposes: firstly, it considers the hardware platform's constraints such as limited computing power and storage resources during the detection process. Secondly, the dynamic attention mechanism within this block enhances the model's focus on crucial target information, thereby optimizing the model's detection performance. To fully leverage the module's efficient focus mechanism, we strategically placed BiFormer at the end of the backbone network to bolster the overall model's feature extraction capability.

C. Lightweight *CARAFE* upsampling operator

To enhance the detection capability for small and medium-sized foreign bodies in foreign object detection, we integrated the lightweight CARAFE upsampling operator[23] into the original YOLOv8, replacing the conventional upsampling operation. The role of the upsampling module is to expand low-resolution images or feature maps into high-resolution ones, enabling better display on high-resolution devices or enhancing subsequent task performance. This module can function as an intermediate layer in convolution, enlarging the feature map size and aiding tensor concatenation. Various methods exist for upsampling, including nearest neighbor, bilinear, bicubic, trilinear interpolation, deconvolution, and transposed convolution. Upsampling modules are commonly used in tasks such as image segmentation, super resolution, and style transfer. Most upsampling methods involve interpolation, where new elements are inserted between pixel points using suitable algorithms based on existing image pixels.

CARAFE, or content-aware recombination of features, is a lightweight general upsampling operator designed to guide the upsampling process based on semantic information within the input feature graph. Its primary approach involves creating an adaptive upsampled kernel via a small convolutional network. This kernel is then used to compute the dot product with corresponding adjacent pixels in the input feature map, resulting in the upsampled feature map. Compared to traditional nearest neighbor or bilinear interpolation methods, CARAFE offers a larger receptive field, better semantic adaptability, and introduces fewer parameters, thus minimally increasing computational costs. The CARAFE network structure comprises two key components illustrated in Fig4: the kernel prediction module, responsible for generating weights for the reorganized calculation kernel, and the content-aware reorganization module, which rearranges features based on these calculated weights.

In Fig.4, the feature graph X with a size of $C \times H \times W$ is up-sampled using CARAFE by a factor of σ. At each position, predicted nuclei are used for recombination. Initially, the channel compression module reduces the channel dimension to C_m to streamline subsequent calculations and enable the utilization of a larger kernel during upsampling. Subsequently, based on the compressed feature map's size, a convolutional layer with a size of $k_{encoder}$ generates the recombination kernel, expanding the receptive field with a larger $k_{encoder}$ while adjusting the channel dimension to $\sigma^2 \times k_{up}^2$. The resulting feature map is then reconstructed into a size $k_{up}^2 \times \sigma H \times \sigma W$ feature map, followed by the application of the softmax function to normalize all channels at each location.

$$W_{l'} = \mu(N(X_l, k_{encoder})) \quad (9)$$

$$X_{l'} = \phi(N(X_l, k_{up}), W_{l'}) \quad (10)$$

At any given position in the output X', there exists a corresponding source position $l = (i, j)$ in the input X, where $i = \left(\frac{i'}{\sigma}\right)$ and $j = \left(\frac{j'}{\sigma}\right)$ hold true. Consider $N(X_l, k_{up})$ as the $k_{up} \times k_{up}$ subregion of X centered at position l. The prediction kernel module μ predicts the location kernel X_l for each position l' based on the subregions of X_l, as illustrated in Eq (9). In Eq (10), the perceptual recombination module ϕ combines the subregion of X_l with the location kernel $W_{l'}$ to yield $X_{l'}$.

The integration of CARAFE into the YOLOv8 network architecture enables the dynamic generation of diverse upsampling cores across various locations in the input feature map. This adaptability caters to targets of varying scales and shapes within different scenes. Through computing the inner product of the input feature map and the local domain, the enhanced feature map exhibits higher resolution and richer information, thereby improving the detection and localization capabilities in foreign object detection tasks. Moreover, compared to alternative upsampling methods like nearest neighbor interpolation and deconvolution, the CARAFE model introduces minimal parameters, boasts low computational costs, occupies less space, and runs faster. These attributes align with the real-time and high-efficiency demands of target detection tasks.

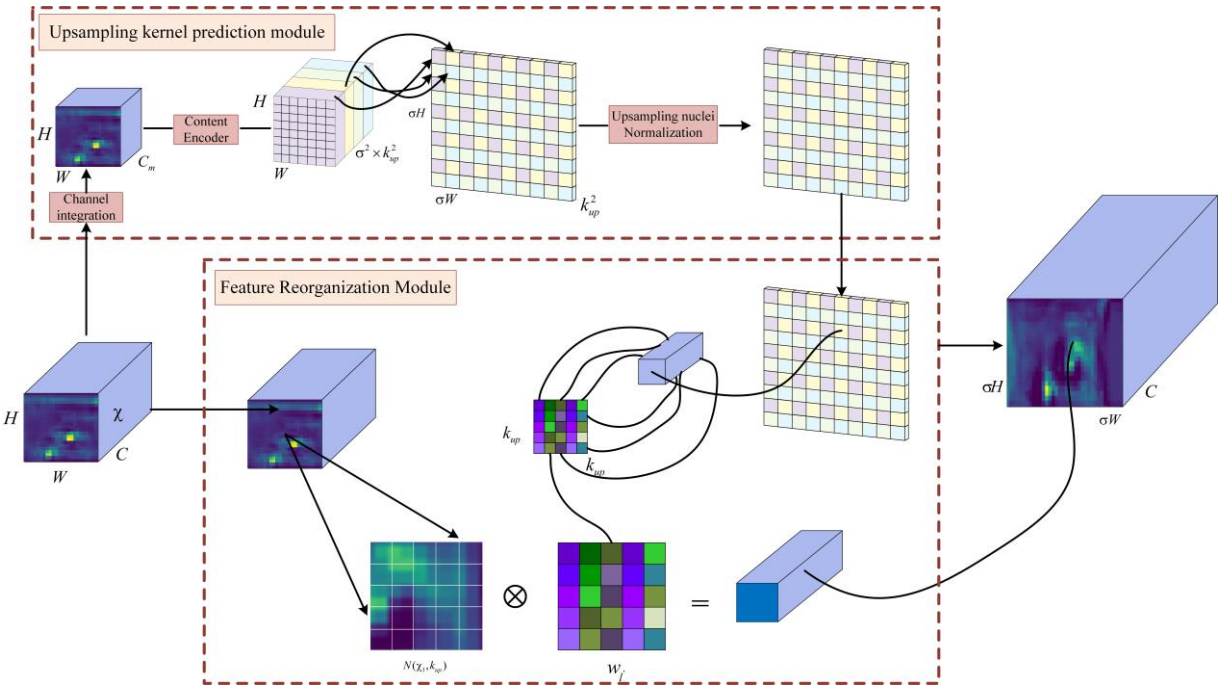


Fig 4. Schematic diagram of CARAFE network structure

D. Loss function optimization

In coal mine foreign object detection tasks, detecting small objects can pose a challenge during the detection process. A well-designed loss function can significantly enhance the model's detection performance. In object detection, the Intersection over Union (IoU) metric is commonly used to gauge the similarity between the model's predicted bounding box and the actual ground truth box. A higher IoU value indicates a closer match between the model's prediction and the ground truth label. During training, IoU is often incorporated into the loss function to aid the network in effectively learning the target detection task. In the context of YOLOv8, Wise-IoU serves as the bounding box regression loss function, aiming to improve the model's generalization capability and speed up convergence. Wise-IoU implements a dynamic non-monotonic focus mechanism for bounding box regression, dynamically adjusting gradient gain based on the bounding box's outlier value. This non-monotonic characteristic arises from the gradient gain's non-monotonic change with the loss value's gain. This dynamic mechanism enhances the model's ability to focus on challenging bounding box regression tasks, leading to improved detection performance.

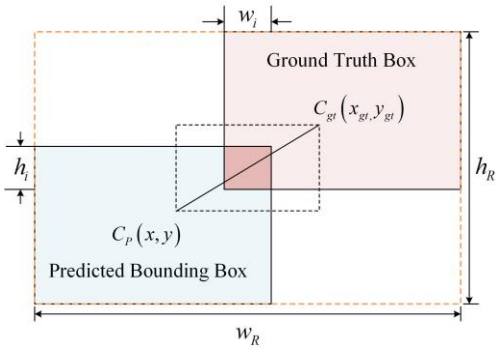


Fig 5. Schematic diagram of loss function parameters

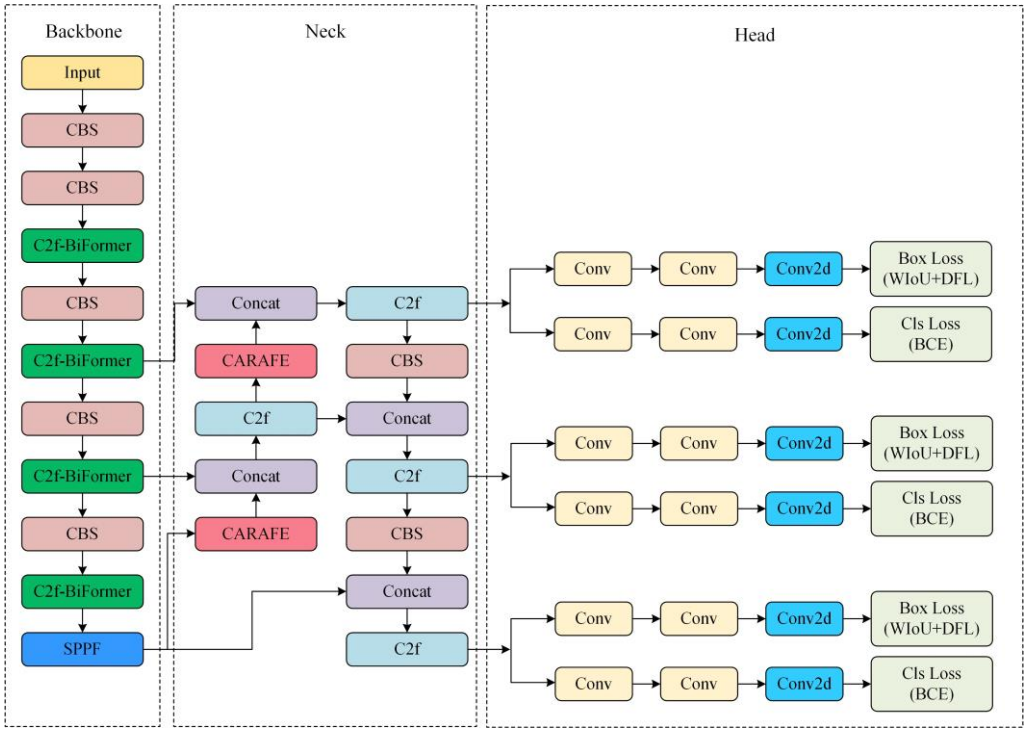


Fig 6. Improved network structure

Certainly, in the context of target detection, we can denote the predicted bounding box as $B_p = [x, y, w, h]$ and the ground truth bounding box as $B_{gt} = [x_{gt}, y_{gt}, w_{gt}, h_{gt}]$, as depicted in Fig5.

Intersection (IoU) losses on a union are defined as follows:

$$L_{IoU} = 1 - IoU = 1 - \frac{w_i h_i}{wh + w_{gt} h_{gt} - w_i h_i} \tag{11}$$

The limitation with L_{IoU} is notable: when there's no overlap between bounding boxes ($w_i = 0$ or $h_i = 0$), the gradient vanishes during backpropagation. Consequently, the width of the overlap area w_i remains static throughout training.

$$\frac{\partial L_{IoU}}{\partial w_i} = \begin{cases} -h_i \dfrac{1 + IoU}{wh + w_{gt} h_{gt} - w_i h_i}, w_i > 0 \\ 0, w_i = 0 \end{cases} \tag{12}$$

Wise-IoU[24] adopts a dynamic method to compute IoU losses in category prediction tasks. Its definition is as follows:

$$L_{WIoU} = R_{WIoU} L_{IoU} \tag{13}$$

Where, $R_{WIoU} = \exp(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(w_R^2 + h_R^2)}) \in [1, e)$ denotes the distance metric, significantly amplifying the value of $L_{IoU} \in [0,1]$ for a prediction bounding box of average quality. To circumvent gradient obstacles that impede convergence, the variables w_R^2 and h_R^2 are excluded from the calculation graph (denoted by superscript *). Furthermore, w_g and h_g signify the width and height of the minimum enclosing box, respectively.

$$box\_loss = \gamma L_{WIoU} \tag{14}$$

The loss function L_{WIoU} incorporates two attention mechanisms. Within this framework, $\gamma = \frac{\beta}{\delta \alpha^{\beta - \delta}}$ signifies the non-monotonic focusing coefficient, $\beta = \frac{L_{IoU}^*}{\overline{L}_{IoU}} \in [0, +\infty]$ characterizes the anomaly degree that predicts the quality of the surrounding frame, and L_{IoU}^* represents the monotonic focusing

coefficient. Additionally, $\overline{L}_{IoU}$ denotes the exponential running average momentum m.

The loss function L_{WIoU} utilizes a distance metric-based attention mechanism that reduces penalties for geometric discrepancies when there is substantial overlap between the predicted and actual bounding boxes. This approach minimally disrupts model training. As $\overline{L}_{IoU}$ is dynamic, the criteria for categorizing frame quality also adjust dynamically. This enables box_loss to dynamically distribute gradient gain based on the current scenario. A lower degree of outliers suggests a higher box mass, leading to a smaller allocation of gradient gain that focuses boundary regression on average-mass boxes.

E. The network structure of the improved algorithm

Our algorithm is enhanced from YOLOv8 to improve the detection capability of foreign bodies in coal mines. Enhancements are made to the convolutional module of the backbone network, the upsampling module of the neck network, and the IoU aspect of the head. These improvements significantly boosted the network's detection performance. Fig6 depicts the enhanced network structure.

To boost feature extraction and optimize the detection of foreign bodies, BiFormer is integrated into the backbone network's end. The neck network utilizes CARAFE for up-sampling, enhancing detailed feature and structural retention for improved quality and accuracy. Furthermore, Wise-IoU is adopted as YOLOv8's bounding box regression loss in the detection head's loss function, accelerating convergence and enhancing model generalization.

III. EXPERIMENTAL PART

In this paper, experiments are carried out on the coal mine foreign body dataset constructed by ourselves to prove the effectiveness of the improved coal mine foreign body detection algorithm based on YOLOv8.

A. Evaluation index and parameter setting

To comprehensively assess the foreign body detection's recognition efficacy, four standard evaluation metrics in target detection are chosen: average accuracy (AP), average average accuracy (mAP), reference number Params, floating-point operation FLOPs, and detection frames per second (FPS). The experiments are conducted in the environment detailed in Table 1, with specific training hyperparameters outlined in Table 2.

TABLE 1. EXPERIMENTAL ENVIRONMENT

Experimental configuration	Version parameter
Operating system	Windows10
Video memory	11GB
CPU	Intel(R) Core (TM) i9-9900KF
GPU	NVIDIA GeForce RTX 2080 Ti
CUDA	10.1
Python version	Python3.7

The entire training process incorporates a learning rate decay method, where the initial learning rate regulates how fast model parameters are updated, and the coefficient of the initial learning rate controls the rate of decay during training to achieve the final learning rate. The final learning rate is determined by multiplying the initial learning rate coefficients. To ensure training stability, a total of 200 epochs are completed, gradually reducing the learning rate throughout training. This approach facilitates smooth model convergence to the optimal solution and prevents abrupt fluctuations.

TABLE 2. MODEL TRAINING HYPERPARAMETER SETTINGS

Hyperparameter	Parameter setting
Input picture size	640×640
Batch size	8.0
NMS IoU	0.75
Initial learning rate	0.01
Final learning rate	0.0001
Optimizer	SGD
Momentum parameter	0.937
Training cycle	200

B. Ablation experiment

During the ablation experiment, we meticulously assessed each module's function, with the experimental outcomes detailed in Table 3. BiFormer, CARAFE, and WIoU modules are individually integrated into the original YOLOv8, while the original algorithm serves as the control group. Table 3 outlines the specific experimental setup and test results. Model 1 represents the base YOLOv8 algorithm, model 8 stands for the enhanced algorithm proposed in this study, and models 2-7 correspond to comparative algorithms from the ablation experiment. The presence of "√" in the table signifies the module's inclusion, whereas an empty space indicates its exclusion.

Based on Table 3, the BiFormer Block module is integrated into the backbone network of Model 2, replacing the C2f module in the original setup. The efficient attention mechanism within BiFormer enhances attention towards crucial areas in the feature mAP, resulting in a 1.6% increase in mAP. The BiFormer module's structure is relatively straightforward. It reduces the model's parameter count by 0.4 million, increases floating-point computations by 1.3 billion, and decreases detection speed by 0.8 frames per second. Table 4 illustrates the experimental outcomes of various attention modules in YOLOv8, demonstrating that among CBAM, SE, ECA, CA, and BiFormer, the BiFormer module exhibits the most effective performance in YOLOv8.

In Model 3, the lightweight CARAFE upsampling operator is integrated into the enhanced feature extraction process of the neck network, replacing the nearest neighbor interpolation method used in the original setup. The utilization of CARAFE upsampling boosts the mAP by 0.6%, increases the parameter count by 0.2 million, and improves detection speed by 0.4 frames per second. Table 5 presents the impact of different upsampling methods on network detection outcomes, clearly demonstrating that the CARAFE upsampling operator delivers the best performance in YOLOv8.

In Model 4, the WIoUv3 loss function is introduced into the prediction box regression loss, which enhances the generalization ability of the model and speeds up the convergence rate. The mAP of the model is improved by 1.1%, and the detection speed is improved by 1.1 frames per second. Table 6 shows the detection results of YOLOv8 algorithm under different loss functions, and WIoUv3 loss function has the best performance.

TABLE 3. RESULTS OF ABLATION EXPERIMENT

Model	BiFormer	CARAFE	WIoU	mAP/%	Params/M	GFLOPs	FPS
1				95.2	11.1	28.8	15.4
2	√			96.8	10.7	28.6	14.6
3		√		95.8	11.3	30.3	15.8
4			√	96.3	11.1	30.3	16.5
5	√	√		97.5	11.2	32.5	15.3
6	√		√	97.3	10.7	29.8	15.8
7		√	√	97.2	11.3	30.3	18.2
8	√	√	√	97.7	10.9	32.5	17.4

TABLE 4. RESULTS OF DIFFERENT ATTENTION EXPERIMENTS

Attention mechanism	mAP	Params/M	GFLOPs	FPS
CBAM	95.6	11.3	107.5	15.3
SE	95.8	11.2	107.4	16.5
ECA	96.1	11.1	107.3	17.6
CA	96.4	11.2	108.2	16.2
BiFormer	96.8	10.7	108.6	14.6

TABLE 5. EXPERIMENTAL RESULTS OF DIFFERENT UPSAMPLING OPERATORS

Upsampling operator	mAP	Params	GFLOPs	FPS
Bilinear interpolation	94.8	11.1	28.6	15.6
Nearest neighbor interpolation	95.2	11.1	28.8	15.4
Trilinear interpolation	95.3	11.2	29.3	15.1
CARAFE	95.8	11.3	28.5	15.8

TABLE 6. EXPERIMENTAL RESULTS OF DIFFERENT LOSS FUNCTIONS

Loss function	mAP	Params	GFLOPs	FPS
CIoU	95.2	11.1	28.8	15.4
DIoU	95.4	11.1	28.8	15.6
GIoU	95.7	11.1	28.8	15.1
EIoU	95.8	11.0	28.6	15.0
SIoU	96.0	11.1	28.8	16.2
WIoU	96.3	11.1	28.9	16.5

Based on BiFormer, several enhancements are made in different YOLOv8 models:

1.Model 5: Integrates the lightweight CARAFE upsampling operator, improving performance on high-resolution input data while controlling model complexity and computational costs.

2.Model 6: Introduces the WIoU loss function based on BiFormer, enhancing the model's ability to handle IoU-sensitive tasks more effectively.

3.Model 7: Combines the lightweight CARAFE operator with the WIoU loss function. During training, the WIoU loss guides the model to generate accurate predictions, while the CARAFE upsampling operator maintains spatial detail and accuracy in prediction results.

In the comparison of ablation experiments, Model 8 in YOLOv8 stands out for achieving a balanced improvement across detection performance, parameter count, and detection speed. With the addition of various improved modules, its mAP surged to 97.7%. Notably, the model effectively reduced the number of parameters while enhancing detection speed, ensuring faster and accurate detection capabilities.

C. Contrast experiment

Fig6 shows the change curves of some important evaluation indicators in the training process of our proposed algorithm. Through these curves, we can intuitively gain insight into the training progress and performance changes of the model. In the early stages of training, the mAP curve rose sharply, showing the model's ability to quickly learn and improve detection performance when exposed to new data. However, after the 30th round of training, the curve became relatively stable, indicating that the performance improvement speed of the model began to slow down and gradually became stable. On the whole, the model has shown good training performance and fitting state.

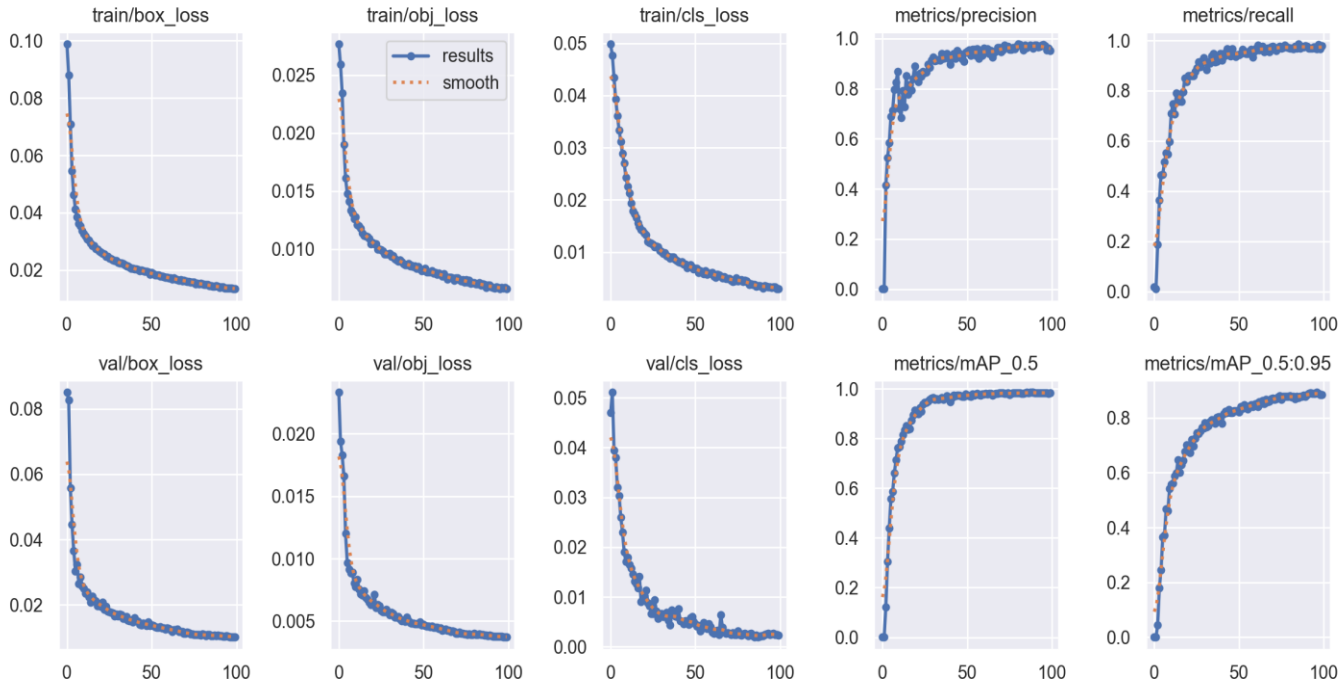
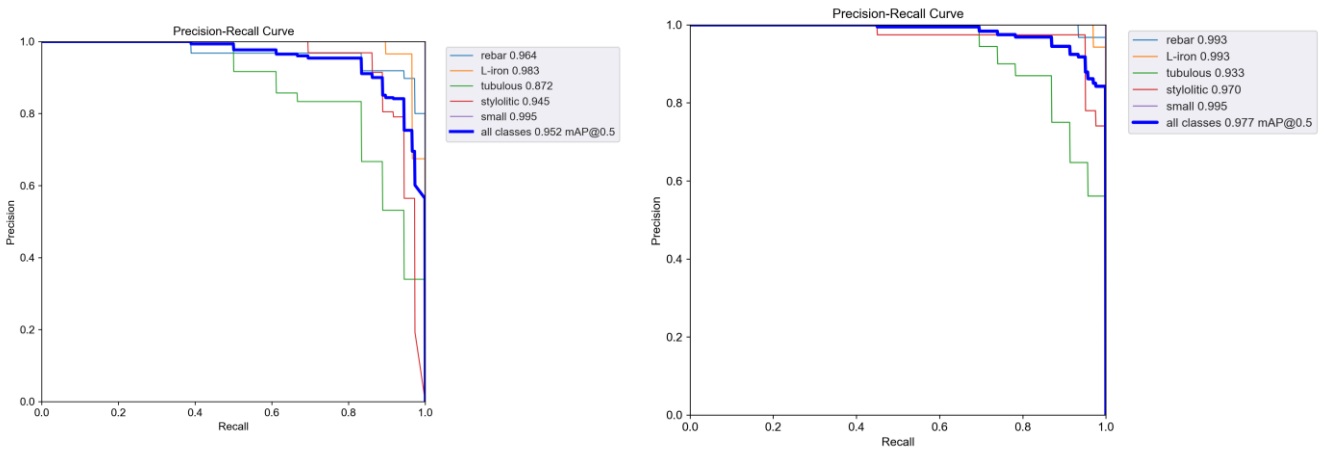


Fig 7. Change curve of evaluation index in the training process of improved algorithm



(a) YOLOv8 (b) Ours

Fig 8.P-R curve of the algorithm before and after improvement

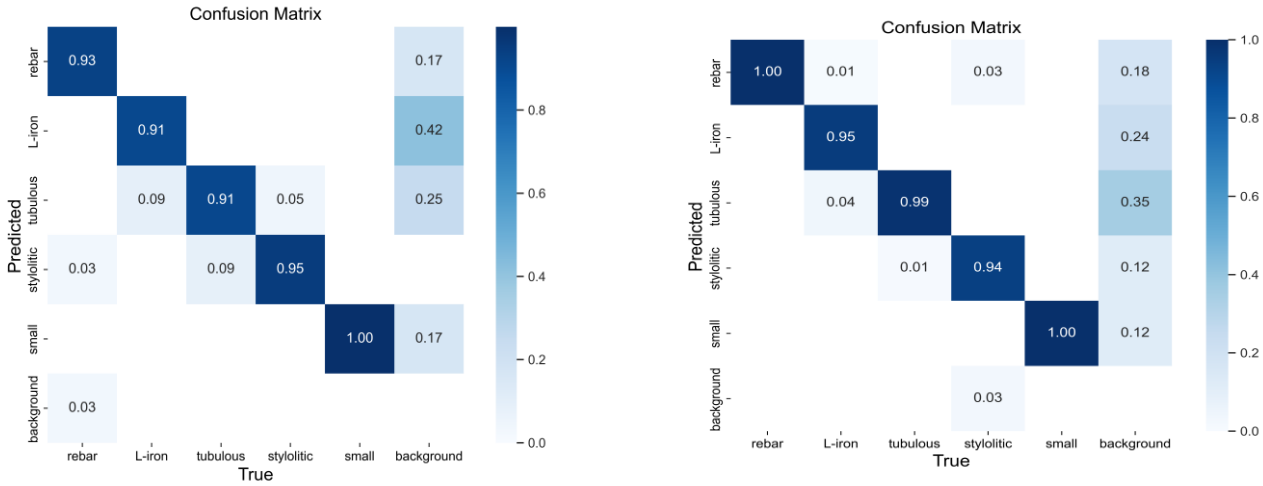
To comprehensively assess the algorithm's detection performance improvements in this study, we've plotted the P-R curve, showcasing accuracy across various recall rates. The area under this curve quantifies the algorithm's enhanced performance. Fig8 illustrates that the improved algorithm achieves higher average detection accuracy across different recall rates, indicating enhanced detection capabilities.

Additionally, Fig9 depicts the normalized confusion matrix for foreign body detection categories, with rows and columns representing actual and predicted categories, respectively. Diagonal values indicate the predicted percentage for each category. This comparison reveals that our enhanced detection algorithm significantly improves average detection accuracy, particularly for steel bars, angle iron, and pipes

.

TABLE 7.EXPERIMENTAL RESULTS OF DIFFERENT ALGORITHMS

Model	Input size	mAP	Params	GFLOPs	FPS
SSD	600×600	92.8	26.3	62.8	54.4
YOLOv3	640×640	89.0	61.5	155.3	32.7
YOLOv4	640×640	83.9	52.5	119.8	44.3
YOLOv5	640×640	93.4	7.3	16.1	42.6
YOLOX	640×640	95.0	9.06	26.8	48.4
YOLOv7	640×640	89.4	7.1	15.8	22.6
YOLOv8	640×640	95.2	11.1	28.8	15.4
Ours	640×640	97.7	10.9	32.5	17.4



(a) YOLOv8 (b) Ours

Fig 9. Confusion matrix of the algorithm before and after improvement

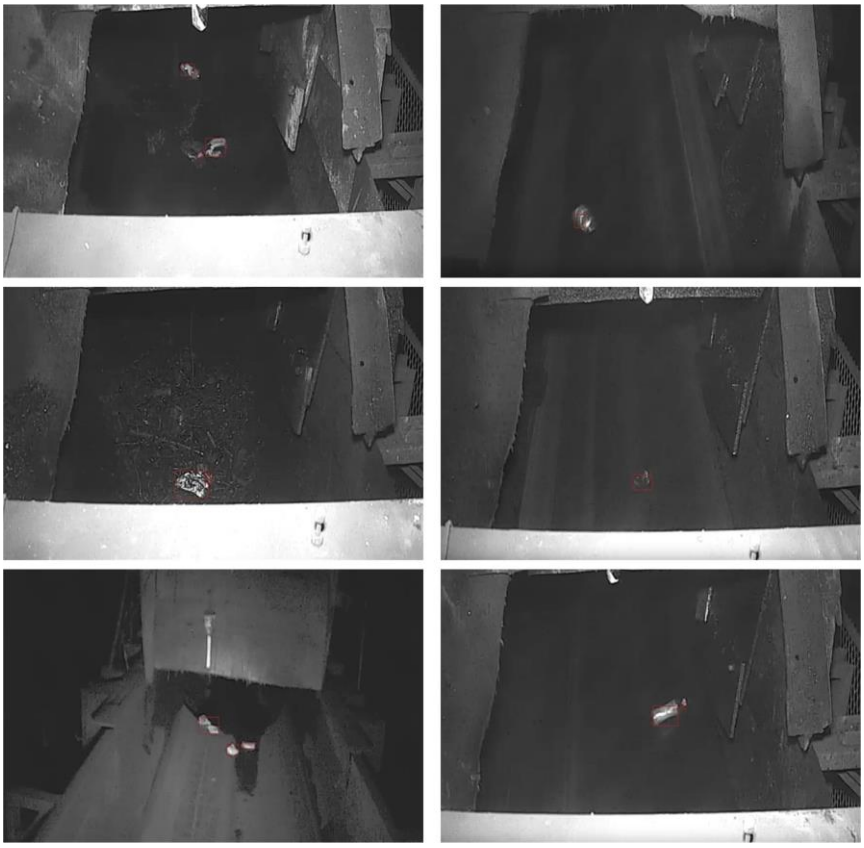


Fig10 . Detection results of the algorithm in this paper

The user compared the improved algorithm model with classical target detection models like SSD, YOLOv3, YOLOv4, YOLOv5, YOLOX, YOLOv7, and YOLOv8, maintaining consistent training hyperparameters and strategies. The results in Table 7 demonstrate that the proposed algorithm surpasses other models in detection accuracy. SSD, YOLOv3, and YOLOv4 have overly large model weights, parameter counts, and GFLOPs, hindering full deployment. YOLOv5, YOLOX, and YOLOv7 have comparable parameter counts and GFLOPs but lower detection accuracy than YOLOv8 and the improved algorithm. Specifically, the proposed algorithm's mAP exceeds YOLOv5, YOLOX, YOLOv7, and YOLOv8 by 4.3%, 2.7%, 8.3%, and 2.5%, respectively.

Fig10 displays randomly chosen images from our test set containing five types of metal foreign objects: steel bars, pipes, columns, angle irons, and small miscellaneous metal foreign objects, along with the detection outcomes using the algorithm described in this paper. The figure illustrates that our algorithm effectively detects and accurately identifies foreign objects.

IV. CONCLUSIONS

In this paper, we present an enhanced algorithm based on YOLOv8 to address issues related to low detection accuracy and imprecise object positioning in coal mine foreign object recognition tasks. The approach incorporates BiFormer dynamic sparse attention into the backbone network to enhance feature characterization. Additionally, it utilizes the CARAFE lightweight upsampling operator for more detailed feature retention during upsampling, thereby improving upsampling quality and accuracy. The detection head of the network is enhanced by incorporating IoU prediction and using Wise-IoU as the bounding box regression loss function in YOLOv8, enhancing model generalization and accelerating convergence. Experimental validation is conducted on a custom coal mine foreign object dataset, comparing objective evaluation metrics and visual results with mainstream general target detection methods. The experimental findings demonstrate superior detection performance and efficacy of the enhanced method compared to other object detection approaches.

REFERENCES

1. Guofa W, Feng L, Yihui P, Ren. (2019). Coal mine intelligence: Core technical support for high-quality development of the coal industry. Journal of China Coal Society, 44(2), 349-357.
2. Guofa W. (2022). Latest Technological Advanceme-nts an d Issues in Coal Mine Intelligence. Coal Scien-ce & Techn ology (0253-2336), 50(1).
3. Xiangang C, Siyin L, Peng W. (2022). Research on Coal Gangue Recognition and Localization System for Coal Gangue Sorting Robots. Coal Science & Technology (0253-2336), 50(1).
4. Zhang N, Donahue J, Girshick R, et al. Part-Based R-CNN s for Fine-Grained Category Detection[J]. Lect-ure Notes i n Computer Science, 2014,8689(1):834-849.
5. He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J].

IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014,37(9):1904-1916.
6. Girshick R. Fast R-CNN[C]. IEEE International Conference on Computer Vision. Santiago, Chile, 2015.
7. Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017,39(6):1137-1149.
8. Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA. 2016:779-788.
9. Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017:6517-6525.
10. Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018.
11. Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, USA, 2020.
12. LIU W, ANGUELOV D, D E, et al. SSD: Single Shot MultiBox Detector[J]. European Conference on Computer vision, 2016:21-37.
13. Duan K, Bai S, Xie L, et al. CenterNet: Keypoint Triplets for Object Detection. arXiv,2019.
14. Han G, Peipei Z, Zheng Y. (Year). Coal Mine Conveyor Belt Foreign Object Detection Based on Feature Enhancement and Transformer. Coal Science & Technology, 1-11.
15. Xiangang C, Hu L, Peng W. (2024). Coal Foreign O-bject Detection Method Based on Cross-Modal Atte-ntion Fusio n. Industrial and Mining Automation (01), 57-65. doi:10.13 272/j.issn.1671-251x.2023110035.
16. Zhengyuan C, Wei J, & Chenghui F. (2023). Intelligent Detection Method for Coal Flow Foreign Objects Based on Dual Attention Generative Adversarial Networks. Industrial and Mining Automation (12), 56-62. doi:10.13272/j.issn.1671-251x.18094.
17. Jun T, Jingzhao L, Qing S. (2023). Real-time Detecti-on of Foreign Objects on Belt Conveyor Based on F-aster-YOL Ov7. Industrial and Mining Automation (11), 46-52+66. do i:10.13272/j.issn.1671-251x.2023020037.
18. Shuai H, Xu Z, Xu M (2022). Foreign Object Detection on Coal Mine Conveyor Belt Based on CBAM-YOLOv5. Journal of China Coal Society (11), 4147-4156. doi:10.13225/j.cnki.jccs.2021.1644.
19. Jingyi D, Rui C, Le H. Foreign Object Detection on Coal Mine Belt Conveyor. Industrial and Mining Automation (08), 77-83. doi:10.13272/j.issn.1671-251x.2021040026.
20. Zhiling R & Yancun Z. (2023). Research on Foreign Objec t Recognition in Coal Mine Belt Transportatio-n Based on Improved CenterNet Algorithm. Contro-l Engineering (04) , 703-711. doi:10.14107/j.cnki.kzgc.20200792.
21. Zhu, L., Wang, X., Ke, Z., Zhang, W., & Lau, R. W. (2023). Biformer: Vision transformer with bi-level routing attention. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10323-10333).
22. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.
23. Wang, J., Chen, K., Xu, R., Liu, Z., Loy, C. C., & Lin, D. (2019). Carafe: Content-aware reassembly of features. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 3007-3016).
24. Tong, Z., Chen, Y., Xu, Z., & Yu, R. (2023). Wise-I-oU: b ounding box regression loss with dynamic foc-using mecha nism. arXiv preprint arXiv:2301.10051.