Advances in Computer, Signals and Systems (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2022.060512
ISSN 2371-8838 Vol. 6 Num. 5

Coal Gangue Identification Based on Improved YOLOv5s Algorithm

Shunan Jia

School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, Henan, China

Keywords: YOLOv5s, gangue, image recognition

Abstract: For coal gangue sorting task, most of them are still in the traditional gangue sorting method, for this situation, based on the analysis and summary of previous research work, a coal For coal gangue sorting task, most of them are still in the traditional gangue sorting method, for this situation, based on the analysis and summary of previous research work, a coal gangue identification method based on improved YOLOv5s is proposed. On the basis of the original network model to make improvements in the backbone module to add non-local attention mechanism for multi-scale network prediction of gangue, improve the feature extraction ability of the model; and design adaptive anchor frame to predict gangue location information efficiently; improve the loss function of the original model to reduce the target miss detection, so as to effectively improve the accuracy of gangue identification. Finally, through experimental comparison, compared with the original network, the improved YOLOv5s model has an mAP value of 8.7%, and the detection speed has been greatly improved.

1. Introduction

In today's society, deep learning is quietly growing with the high speed of artificial intelligence high-tech obsolescence and iterative updates [1][2]. The concept of deep learning has led to a leap forward in the whole field of image recognition, abandoning the traditional algorithms of image pre-processing, image segmentation, and artificially designing the image features of objects for recognition, and using neural networks and big data to achieve automatic learning of the features of images and thus object recognition and localization [2][5].

Currently, the existing target detection based on non-candidate regions can be divided into two categories, one based on region selection and the other on regression learning. Although the detection capability based on region selection is quite good, its detection speed seriously drags it down and is poor in this aspect of real-time. In the regression learning-based target recognition method, it is easier to achieve real-time detection and recognition because the candidate region extraction is omitted as a major step, and all recognition and detection steps are integrated for processing.

For the gangue sorting task, it is a study to solve the real time sorting of gangue in coal mines want to use machine instead of manual, so it will need to complete the sorting work in a short time in the actual work. In this paper, we decided to use YOLOv5 for gangue target detection, but

YOLOv5 is not an independent model, but a family model architecture, respectively v5s, v5m, v5I and v5x, overall there is no big difference, the network structure is basically the same, the difference between the number of modules and the number of convolutional kernels are different. Among them, YOLOv5s is the network with the two smallest depths and graph widths, and the other three can be considered as the basis on which the deepening and widening are carried out 错误!未找到引用源。[6], and finally the YOLOv5s network will be chosen as the detection model.

2. YOLOv5s Algorithm

The entire structure of YOLOv5s is shown in Figure 1.

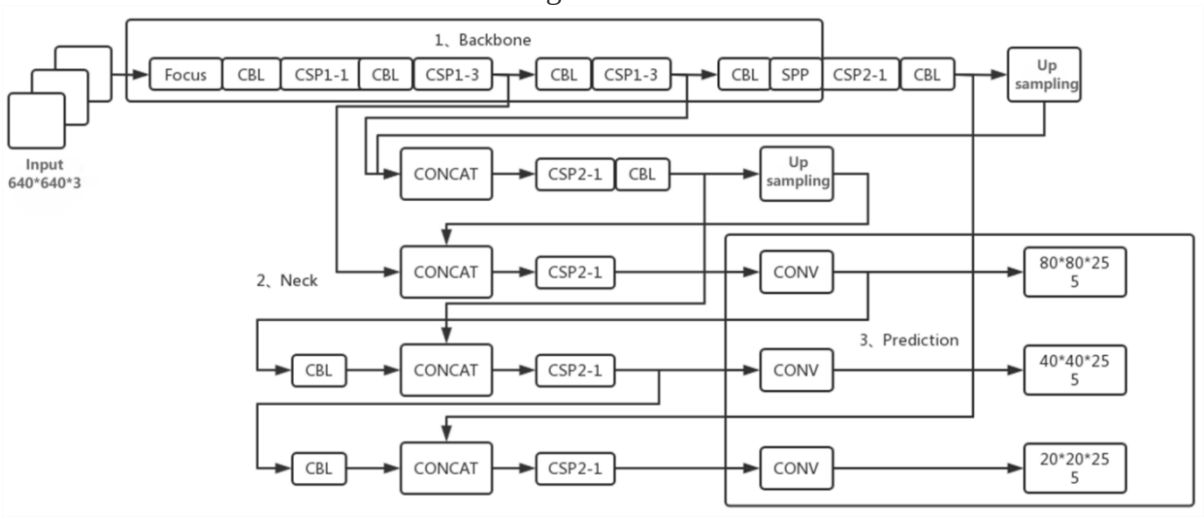


Figure 1: YOLOv5s network structure

Firstly, at the very beginning of the model, the first step is to perform Mosaic enhancement on the images of the input model at the input side. This operation can directly calculate the data of four images, which are stitched using random scaling, random cropping, and random arrangement, enriching the detection dataset and enabling better robustness of the network.

In the second step, the processed image is then propagated forward to reach the Backbone module, and the whole module contains four structural blocks: Focus, CBL, CSP and SPP. Among them, the Focus block will perform a block-cutting operation on the image, similar to downsampling, and the significance of this step is to ensure that the feature information of the image will not be lost. The CBL block refers to the convolution of the image, batch normalization and Leaky ReLU function activation operation. CSP block is designed in the whole model a total of two, in which CSP1_X is stored in Backbone and CSP2_X is stored in the next stage of Neck, CSP1_X first divides the feature mapping of the base layer into two parts, and then merges them through the cross-stage hierarchy, which can guarantee the accuracy while reducing the computational accuracy. SPP block is short for spatial pyramid pooling, which first halves the input channels by a standard convolution module and then does max pooling with kernel-size of 5, 9, and 13, respectively, and the padding is adaptive for different kernel sizes. The result of the three max-pooling is concat with the data without pooling operation, and the final number of channels after merging is twice the original number. Figures 2 to 6 show the schematic diagrams of the individual modules [8][9].

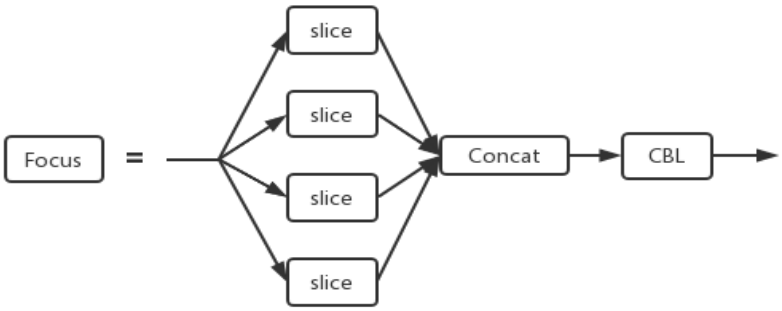


Figure 2: Schematic diagram of the Focus module

Figure 3: Schematic diagram of CBL module

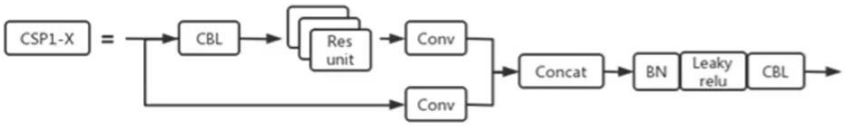


Figure 4: Schematic diagram of CSP1-X module

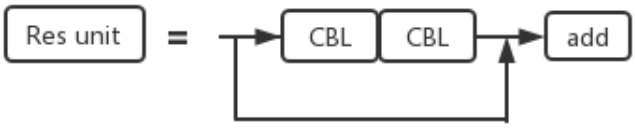


Figure 5: Res unit schematic diagram

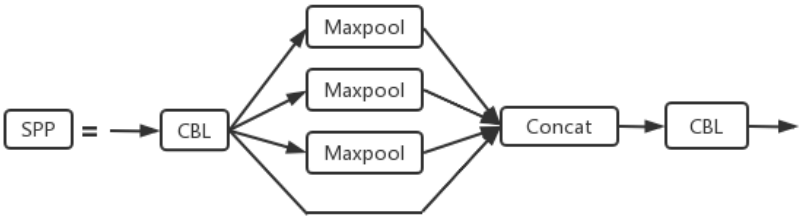


Figure 6: Schematic diagram of SPP module

The third step then goes through the Neck module, the whole module consists of FPN+PAN, this module will integrate all the feature information extracted in the previous step for feature fusion using the CSP2 structure designed by CSPnet to strengthen the network feature fusion capability. Figure 7 shows the schematic diagram of Neck module [10]错误!未找到引用源。.

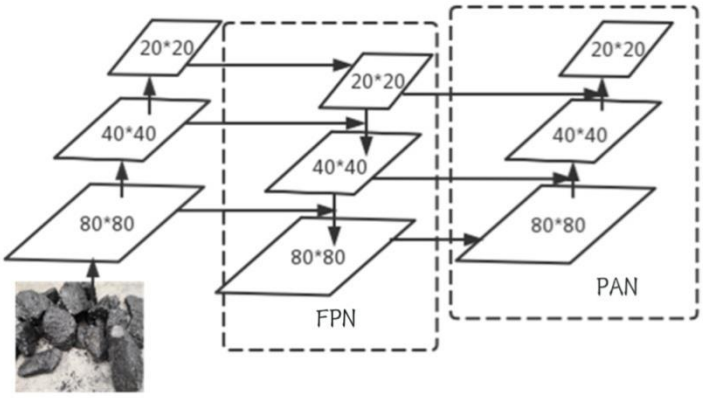


Figure 7: Schematic diagram of Neck module

Next comes the Prediction module, which is the final step in the operation of the entire model ——detection process, specifically the final detection and identification of the integrated feature maps obtained earlier.

The process of gangue identification based on YOLOv5s can be described as follows.

(1) Divide the home-made gangue dataset into training and test sets in the ratio of 4:1.

(2) Inputting the coal gangue images in the training set, enhancing them by Mosaic, and then following a set batch order as the input.

(3) Obtain the position and scale size information of the prediction frame of the coal gangue by forward propagation through the three modules Backbone, Neck and Prediction in turn.

(4) Then calculate the difference in loss function values between the prediction frame and the real frame in the annotated file.

(5) Calculate the weight coefficient and bias value when the loss function takes the minimum value under the set number of iterations.

(6) Then the weight coefficients and bias values are used as the parameters for the model to start forward propagation to obtain the prediction information of the gangue in the gangue images in the test set.

3. YOLOv5s Model Improvements

3.1 Add Attention Mechanism

The backbone module region of YOLOv5s is mainly formed by stacking multiple residual modules on top of each other. However, the disadvantage of residual modules is that they cannot fully fuse multi-scale feature information. The solution for this is to add non-local in the backbone module, non-local module compared to the constant pair of stacked convolutional layers and RNN operators, and non-local operation directly calculates the relationship between two locations to quickly capture the long range dependence, which can be temporal location, spatial location and spatio-temporal location, but will ignore their Euclidean distance, this calculation method is actually to find the autocorrelation matrix, which is a generalized autocorrelation matrix. Moreover, the nonlocal operation is computationally efficient, and only fewer stacking layers are needed to achieve the same effect. In addition, the non-local operation can ensure that the input scale and

output scale remain unchanged, and this design can be easily incorporated into the current network architecture for use. Figure 8 shows the schematic diagram of the non-local module.

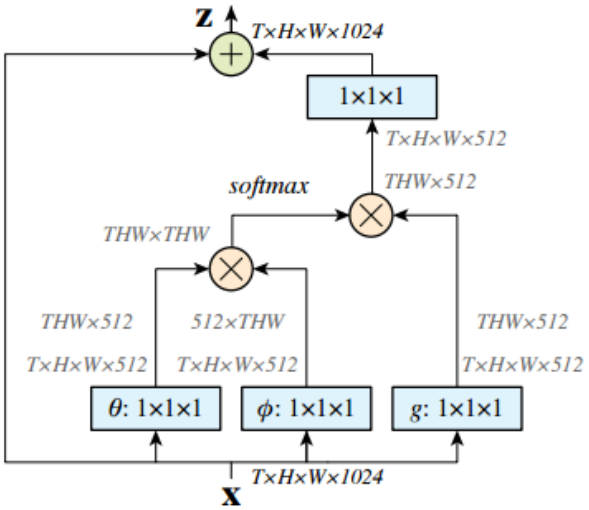


Figure 8: Schematic diagram of Non-local module

The generic formula for Non-local is represented by equation 1.

$$y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_j) \qquad (1)$$

Where x is the input signal, I represents the output location, such as spatial, temporal or spatio-temporal index, its response should be enumerated for j and then calculated, the f function calculates the similarity between i and j, the g function calculates the representation of the feature map at position j, and the final y is obtained after normalization by the response factor $C(x)$.

For simplicity in the experiment, $g(\cdot)$ takes the form of equation 2 by default.

$$g(x_j) = W_g x_j \qquad (2)$$

In this paper, the non-local structure is embedded into the backbone module of the yolov5s model, as shown in Figure 9 to improve the feature extraction ability of the backbone region without significantly increasing the complexity of the yolov5s model.

Figure 9: Improved backbone structure

3.2 Designing Adaptive Anchor Frames

In the target detection algorithm, the final candidate box (anchor box) is obtained by training with predefined boxes, and a reasonable candidate box can not only improve the detection speed of the network but also the detection accuracy. In the original YOLOv5s is obtained by k-means clustering on coco dataset, but these sizes are not applicable to the homemade gangue images, and the clustering centers obtained by k-means clustering are randomly generated, and the different initial values of clustering points will produce different effects, which will have a greater impact on the prediction results in the chase. Therefore, the candidate frames should be re-clustered.

The prior frame is obtained in YOLOv5s by training with the K-Means++ clustering algorithm,

and a sample is randomly selected from the dataset as the center of the initial clusteringc_1. First, the shortest distance between each sample in the training set and the initial clustering center of the selected sample, denoted by *D(x)*, is calculated, and then the probability of each sample being selected as the next clustering center, P_{x_i}, and finally, the next clustering center is selected according to the roulette wheel method. This is repeatedly done to select K clustering centers, and then for each sample in the training set, the distance from it to the K clustering centers is calculated until the clustering centers no longer change, the clustering is finished, and the final prior frame is obtained. This can effectively reduce the clustering bias caused by the original algorithm at the initial clustering point, and get a better sized prior frame and match it to the corresponding feature map, which can effectively improve the detection accuracy and recall rate. The related equations are given in Equations 3 and 4.

$$c_1 = \frac{1}{c_1}\sum_{x \in c_1} x \tag{3}$$

$$P_{x_i} = \frac{D(\mathrm{X})^2}{\sum_{x \in X} D(\mathrm{X})^2} \tag{4}$$

In this paper, the method of pre-setting the initial clustering centers is used to analyze the characteristics of coal gangue, and finally 12 initial clustering frames are obtained, which are (3, 4), (4, 8), (7, 6), (7, 12), (10, 13), (12, 18), (15, 9), (16, 30), (27, 15), (30, 61), (33, 23) and (46, 36).

3.3 Improved Loss Function

YOLOv5 uses the GIOU loss function, the advantage is that GIOU increases the loss of detection frame scale on the basis of DIOU, and increases the loss of length and width, so that the final prediction frame will be more close to the real frame. However, there are disadvantages that should not be underestimated. Here the aspect ratio of the bounding box describes the relative value, and there is a certain fuzzy concept, there is also the problem of not considering the balance between simple and complex samples. Figure 10 shows the schematic diagram of GIOU detection, and the corresponding L_GIOU values are the same at this time.

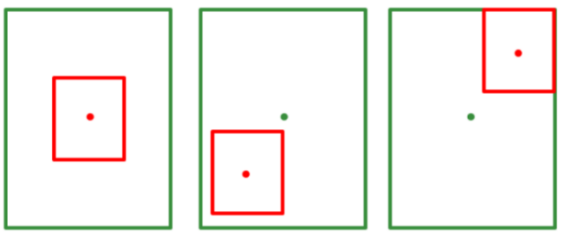


Figure 10: GIOU testing schematic

From Figure 10, it can be seen that the first one has a better effect, the second one has a worse effect, and the third one has the worst effect. So for this situation, this paper decided to adopt the method of EIOU. GIOU Loss though considers the overlapping area, centroid distance, and aspect ratio of the bounding box regression. But the difference of aspect ratio reflected by v in its formula, rather than the real difference of width and height respectively with its confidence, so it sometimes hinders the model to optimize the similarity effectively.

The penalty term of EIOU is based on the penalty term of GIOU to split the influence factor of aspect ratio to calculate the length and width of the target frame and anchor frame separately. The loss function contains three parts: overlap loss, center distance loss, and width and height loss. The first two parts continue the method in GIOU, but the width and height loss directly minimizes the

difference between the width and height of the target frame and anchor frame, which makes the convergence faster. The penalty term formula is represented by equation 5.

$$L_{EIOU} = L_{IOU} + L_{dis} + L_{asp}$$
$$= 1 - \mathrm{IOU} + \frac{\rho^2(b,\ b^{gt})}{(w^c)^2+(h^c)^2} + \frac{\rho^2(w,\ w^{gt})}{(w^c)^2} + \frac{\rho^2(h,\ h^{gt})}{(h^c)^2} \quad (5)$$

Where b and b^{gt} denote the respective centroid positions and $\rho(\cdot)^2$ denotes the Euclidean distance.

3.4 Analysis of Results

3.4.1 Model Performance Evaluation Metrics

In the target detection task, the common metrics used to evaluate the model performance are IOU, precision, Recall, AP, mAP, and so on. The mAP is one of the most important metrics in the performance evaluation of target detection algorithms, and is the average value of AP for multiple target categories. mAP is between 0 and 1, and the larger the mAP value is, the higher the detection accuracy of the model in all categories, which means the better the target detection effect.

3.4.2 Analysis of Experimental Results

The model is trained and run to get a series of visual result graphs, as well as the record files generated during the training process, all of which are kept for subsequent analysis, and some specific information about the network training process can be further understood from the result graphs of the experiments.

The model training environment in this paper is Ubuntu 16.04 system, Intel Core i7-7700HQ processor, 16GB RAM, NNVIDIA GeForce GTX 1660 graphics card, platform development based on python, pytorch deep learning framework.

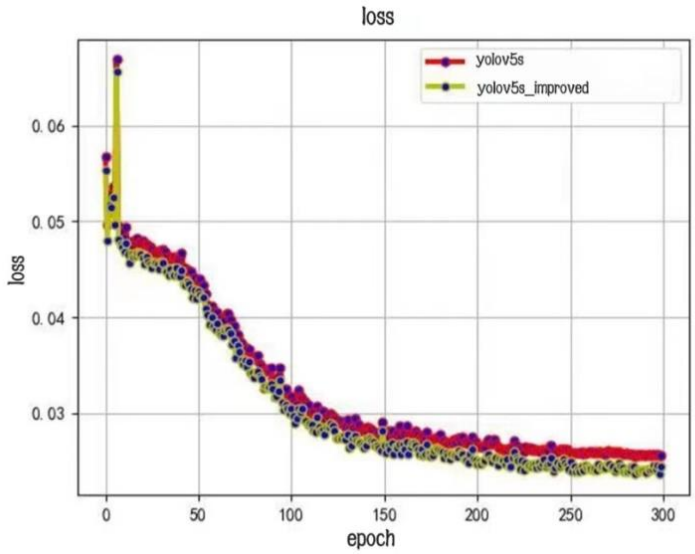


Figure 11: Comparison Chart of Confidence Loss Function

The comparison of the confidence loss function between yolov5s and the improved yolov5s_improved model in this paper is shown in Figure 11 for the loss function of confidence and Figure 12 for the loss function of localization when the loss function values are recorded once every

50 iterations during training.

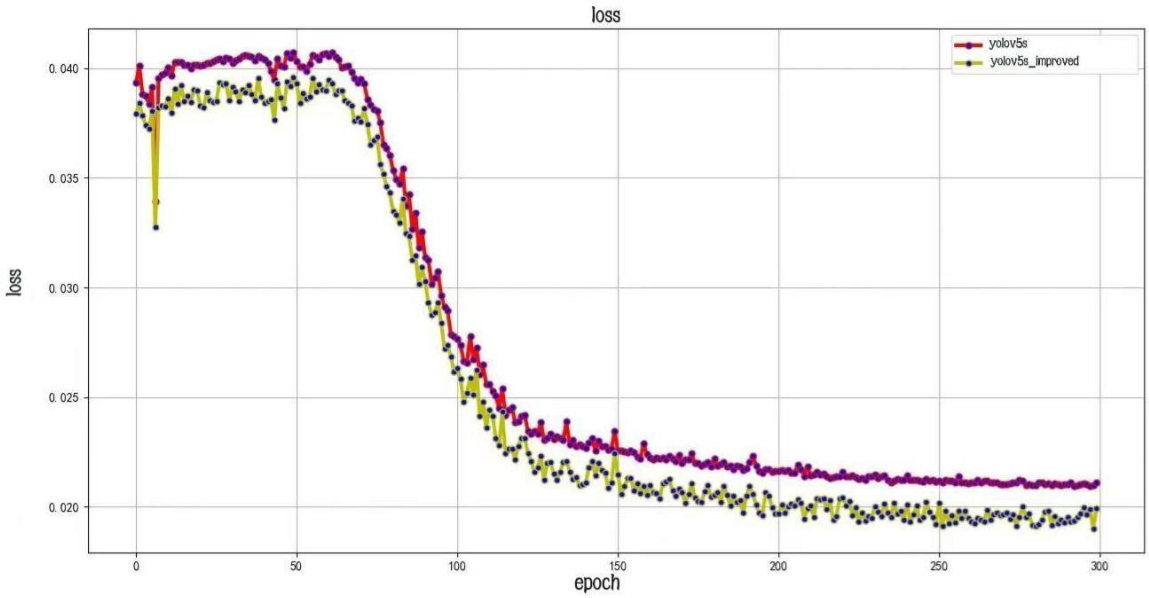


Figure 12: Comparison Diagram of Positioning Loss Function

According to the two loss function plots, it can be seen that the loss of the improved yolov5s_impored model decreases faster than that of the original yolov5s model, and both loss function values of the improved yolov5s_impored model are lower than that of the original yolov5s model, which is enough to show that the improved model works better. And according to Figure 12, it is easy to find that the loss value of the improved yolov5s_impored model is lower than 0.02 when the number of iterations is 200, which means that the error value between the prediction frame and the calibration frame of the model is smaller and the positioning is more accurate.

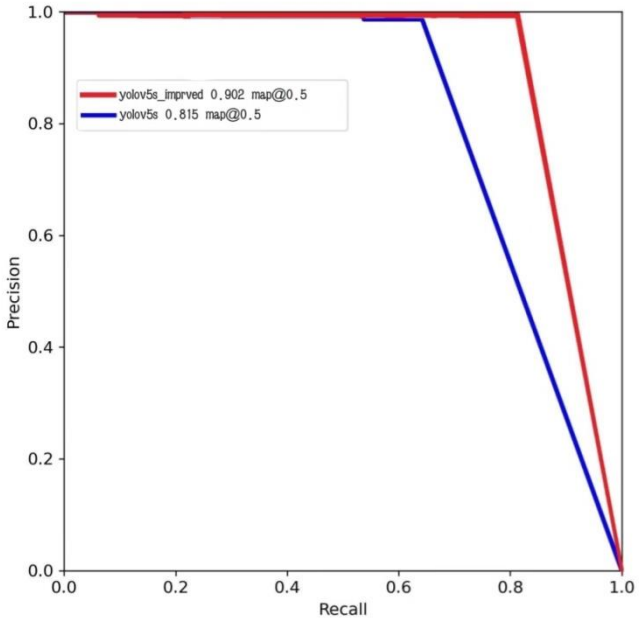


Figure 13: P-R comparison chart

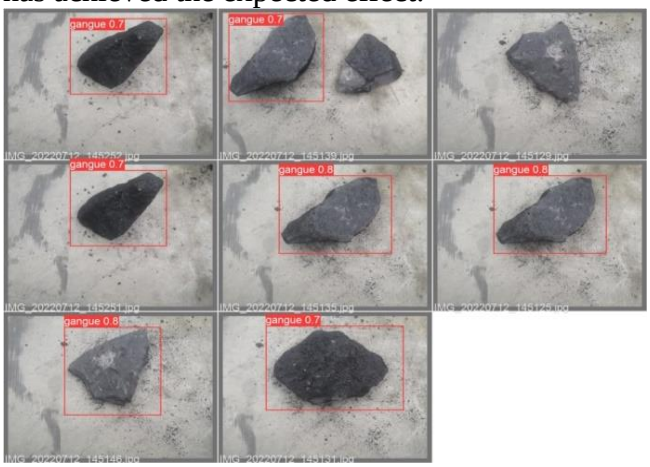
Figure 13 shows the P-R comparison before and after the model improvement, with only one category option for the specificity of the research object in this paper. Therefore the AP value is the

same as the mAP value, which is the size of the area of the enclosed closed graph in the resultant plot. The area formed by the improved yolov5s_impored model is significantly larger than that of the original yolov5s model.

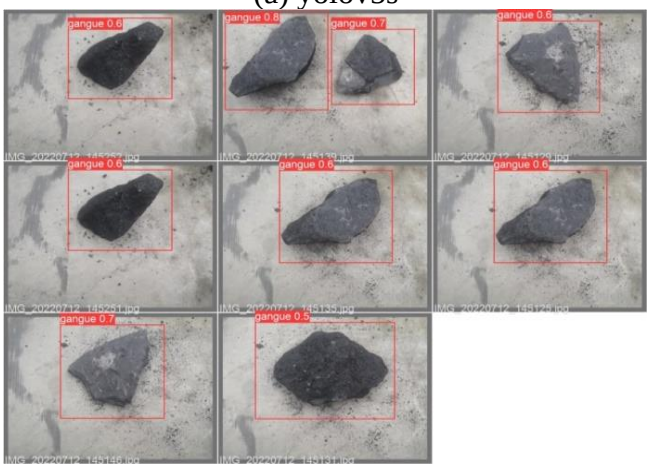
Table 1: Training results before and after the improvement of the YOLOv5s algorithm

Models	Precision/%	Recall/%	mAP/%	Frame Rate/fps
YOLOv5s	78.8	64.0	81.5	39.4
YOLOv5s_ipproved	80.7	81.0	90.2	43.6

According to Table 1, the values of the improved yolov5s_impored model have been improved, and the Precision, Recall and mAP values have increased by 1.9%, 17% and 8.7% in turn. The target detection algorithm generally has an fps value greater than 24 to meet the real-time detection, and the improved yolov5s_impored model has improved the detection speed compared with the original model, and the fps value has reached 43.6, so it can be seen that the improved yolov5s_impored model has achieved the expected effect.



(a) yolov5s



(b) yolov5s impored

Figure 14: Comparison Chart of Test Results

And, the original yolov5s model in the detection of target omission phenomenon, the improved yolov5s_impored model also better to complete the task, can accurately detect each gangue target. Figure 14 model detection results comparison chart.

4. Conclusion

In this paper, we make improvements based on YOLOv5s model, add non-local attention mechanism for multi-scale network prediction of coal gangue, and design adaptive anchor frame for efficient prediction of coal gangue location information, improve the loss function of the original model, and reduce the case of target miss detection. The final experimental results show that the mAP value of the improved yolov5s_impored model increases by 8.7% and the detection speed reaches 43.6/fps, and the improved model can better complete the identification and detection of coal gangue.

References

[1] Zhang T. Multivariate industrial order forecasting based on feature expansion and LSTM model. System Simulation Technology, 2021, 17 (4): 221-225.

[2] Cao, Xian-Gang, Xue, Zhen-Ye. Migration learning based GoogLenet coal gangue image recognition. Software Guide, 2019,18 (12): 183-186.

[3]Ma Jun, Yu Zhongming, Shu Shihai, et al. Environmental hazards of coal gangue in mining areas and management measures. Coal Engineering, 2015, 47 (10): 70-73.

[4] Zhang Zhenhong. Development status and application prospect of dry coal beneficiation technology in China. Coal Processing Technology, 2019 (01): 43-47+52.

[5] Yu D., Zou S. W., Qin C. Research on the application of image grayscale information in automatic coal gangue sorting. Industrial and mining automation, 2012, 38 (02): 36-39.

[6] YAN B, FAN P, LEI X Y, et al. A real-time apple targets detection method for picking robot based on improved YOLOv5. Remote Sensing, 2021, 13 (9): 1619.

[7] Liao Yenna, Li Wan. Bridge Crack Detection Method Based on Convolutional Neural Network. Computer Engineering and Design, 2021, 42 (08): 2366-2372.

[8] YAN B, FAN P, LEI X Y, et al. A real-time apple targets detection method for picking robot based on improved YOLOv5. Remote Sensing, 2021, 13 (9): 1619.

[9] LIN Sen, LIU Meiyi, TAO Zhiyong. Detection of underwater treasures using attention mechanism and improved YOLOv5. Transactions of the CSAE, 2021, 37 (18): 307-14.

[10] YU J, LUO S.Detection method of illegal building based on YOLOv5. Computer Engineering and Applications, 2021, 57 (20): 236-244.

[11] ZHAO Rui, LIU Hui, LIU Peilin, et al. Research on safety helmet detection algorithm based on improved YOLOv5s. Journal of Beijing University of Aeronautics and Astronautics, 2021, 1-16.