GOSPODARKA SUROWCAMI MINERALNYMI – MINERAL RESOURCES MANAGEMENT

2025 Volume 41 Issue 2 Pages 177–196

DOI: 10.24425/gsm.2025.154546

GAIPIN CAI[1], HUI LUO[2], JINRUO HUANG[3], GUOTAO YI[4]

Ore image target detection based on improved YOLOv5 network

Introduction

The environment of collecting ores in mines is more complex; with a wide variety of collected ores and different ore location morphologies, coupled with the varying particle sizes of crushed ores, detecting ores with small targets is particularly challenging. The study of fast and accurate target recognition method of ore on the conveyor belt after crushing constitutes an important portion of the research domain of realizing intelligent mines,

✉ Corresponding Author: Gaipin Cai; e-mail: 1982473318@qq.com

[1] School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, China; ORCID iD: 0009-0007-2026-1725; e-mail: 1982473318@qq.com

[2] School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, Jiangxi, China; Jiangxi Province Engineering Research Center for Mechanical and Electrical of Mining and Metallurgy, Ganzhou 341000, Jiangxi, China; e-mail: 1473926147@qq.com

[3] School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, China; e-mail: 1277237512@qq.com

[4] School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, China; e-mail: 2468298330@qq.com

and it is also a necessary part of the measurement of ore particle size. Therefore, it is of great significance to propose a target recognition algorithm with a low leakage detection rate and high precision for more images.

In recent years, deep learning algorithms have become more and more mature in artificial intelligence technology through the advantages of neural networks to learn and remember relevant features, and neural networks can choose different structures according to the data features such as text, sound, image, video, etc. (Zhang et al. 2021). In addition, it can also establish connections between data and achieve data integration by analyzing and processing input data information (Dawson et al. 2020; Liu et al. 2021), and nowadays, many researchers combine deep learning algorithms with artificial intelligence technology to efficiently complete target detection tasks. Target detection algorithms can be mainly divided into two types; one is the standard two-stage target detection algorithms that generate pre-selected box regions and then classify and predict the targets by classifiers, such as R-CNN (Xie et al. 2024), Fast R-CNN (Zhang et al. 2024), Faster R-CNN (Li et al. 2024), Mask R-CNN (He et al. 2020) and so on. These detection algorithms are characterized by relatively high detection precision. However, they have a large number of parameters in the detection process, and the slow processing speed limits their use in practical engineering environments. (Fu et al. 2018) enhanced the Faster R-CNN model, attaining a target detection precision of approximately 90% and a target detection time of 5 milliseconds. (Yang et al. 2023) proposed a target detection performance optimization algorithm taking Mask R-CNN as the basis.

First of all, the network structure of FPN is improved and optimized by adding feature fusion paths to improve the network's full utilization. Secondly, an attention mechanism is incorporated to enhance the efficiency of model feature extraction. The other category consists of one-stage target detection algorithms that do not generate candidate frames and predict the target result directly through forward reasoning, such as YOLO (Redmon et al. 2015), FCOS (Wu et al. 2023), SSD (Yan et al. 2022) and other series of algorithms. These algorithms are more applicable to real engineering situations due to their smaller number of parameters, faster processing speed, and simplified processing flow. Among them is the YOLO algorithm, an end-to-end deep learning model that acquires and extracts feature information from the input image through learning, which not only simplifies the process of detection but also makes the model more intuitive and easy to understand. (Shen et al. 2024) introduced the spatial pyramid pooling module into the YOLOv5 model to increase the precision of target detection by improving the loss function, and the convergence speed of the model is enhanced. Moreover, the average precision of its detection reaches 97.3%. As (Zhang et al. 2023), addressing the issues of complex image backgrounds and high resolution, and a large number of small targets, proposed a lightweight network based on the YOLOv7 architecture, the hybrid attention mechanism is incorporated into the feature extraction network, fuses the feature information between the shallow and deep layers. Moreover, it enhances the learning capability of the model and its ability to recognize small targets. The YOLOv8 algorithm is characterized by high precision, but its detection

of small targets still has some limitations in complex scenarios such as target adhesion and occlusion (Fan et al. 2025).

In response to the complex situations of ore stacking, edge adhesion, uneven particle size distribution, and minor feature differences on the conveyor belt after the crushing process, to further enhance the adaptability of the model and improve precision while keeping the operating speed unchanged, this paper proposes targeted improvements to the algorithm based on YOLOv5. The CA attention mechanism is added to the original network, which realizes the feature recognition visualization and boosts the network's ability to extract features and recognize more images; the edge regression loss function SIoU (Scale-Invariant Intersection over Union) is introduced based on the IOU characteristics (Sun et al. 2023), which considers the vectors corresponding to the required regression angle, improve the model performance, improve the positioning precision of target recognition, enhance the detection speed and detection precision of ore recognition, and provide the basis for subsequent ore particle size calculation.

1. Dataset introduction

As there is a shortage of publicly accessible ore image datasets, this experiment constructed the ore dataset required for model training. In order to obtain high-quality

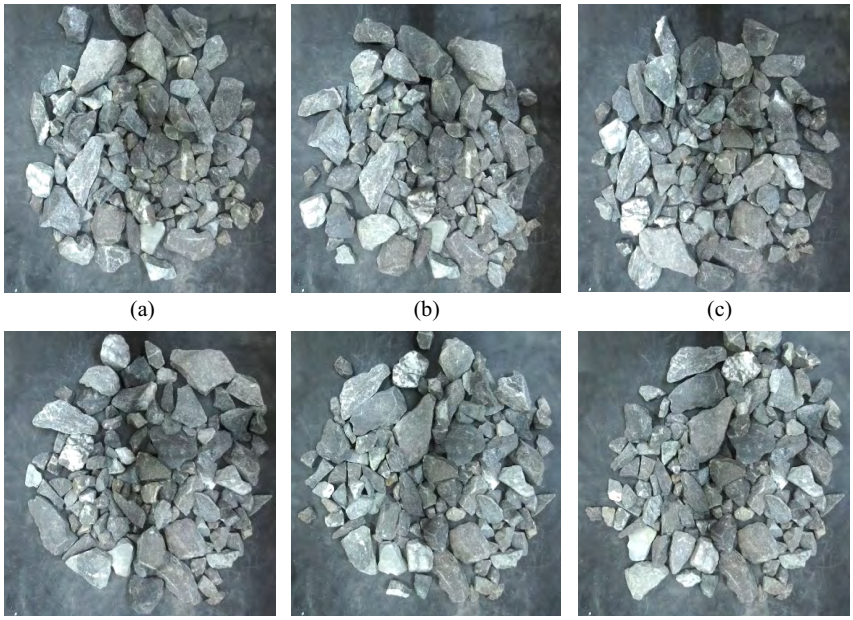


Fig. 1. Dataset section ore images

Rys. 1. Sekcja zbioru danych obrazów rudy

datasets, the ore was randomly dispersed on the conveyor belt in the laboratory to simulate an on-site environment, and the ore images were collected. To enhance the universality of the trained network model and enable it to handle the detection tasks of ore stacking and adhesion better, 300 simulated on-site ore images were used in this experiment. The ore position and morphology of each image were different, and some ores had adhesion and overlap between them. The dataset was expanded by random rotation and mirror-image data augmentation methods to generate corresponding label files; the expanded dataset has a total of 1,000 samples. The dataset is partitioned into training, validation, and testing sets at a ratio of 4:3:3. Figure 1 shows the partial ore images of the dataset generated in this experiment.

2. YOLOv5 Network improvements

YOLOv5 is an object detection model, and its developers have developed four versions, namely YOLOv5s, YOLOv5m, YOLOv5x, and YOLOv5l, the key distinction among these four versions lies in the depth and width of the network. This section focuses on the architecture of YOLOv5s, which is the network in the YOLOv5 series that has the least depth and the narrowest feature map width. The subsequent versions of m, l, and x are all extended and deepened based on the YOLOv5s basic model (Dai et al. 2023).

The YOLOv5 network can be mainly segmented into four components: input section, Backbone, Neck, and Prediction. Figure 2 depicts the network structure of YOLOv5.

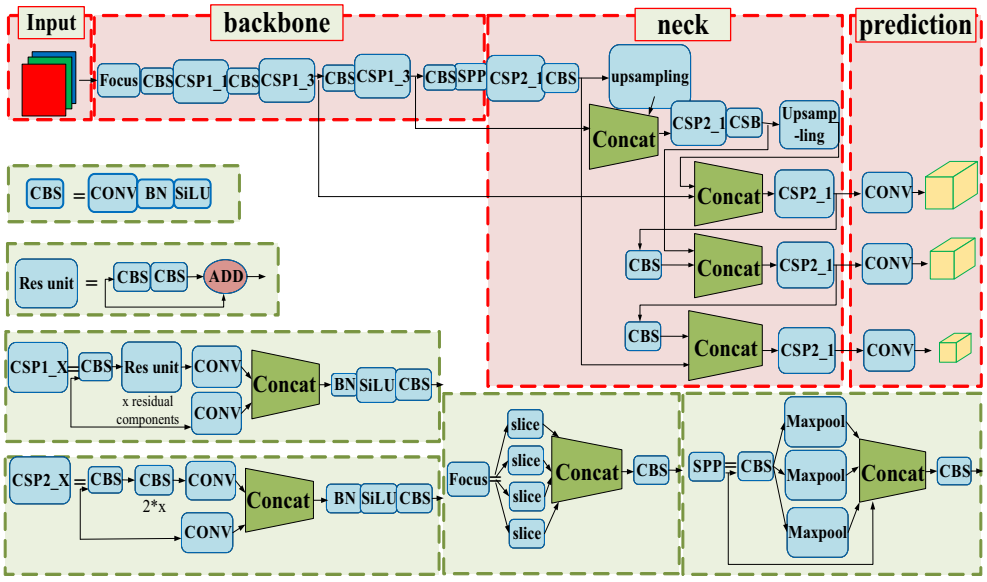


Fig. 2. Network structure diagram of YOLOv5

Rys. 2. Schemat struktury sieci YOLOv5

Due to the pooling layer of the C3 module in the network, the convolutional neural network experiences a certain loss of image information when increasing the receptive field of the upper convolutional kernel and the background of the aggregated image, resulting in incomplete extracted target features and inability to fully fuse features of different scales. This leads to false positives and missed detections, and the balance between precision and speed cannot be achieved well.

2.1. Improvement of attention mechanism

Although the YOLOv5 model has good detection speed, there are still shortcomings in target feature extraction, resulting in insufficient target detection precision. Coordinate attention mechanism (CA) (Guo et al. 2022) (Coordinate attention for effective mobile network design) can achieve more effective extraction of key features of detection targets. Its advantages are predominantly manifested in taking into account not only channel information but also directional position information. Moreover, it is lightweight to such an extent that it can be readily inserted into the core modules of lightweight networks. This article considers that the recognition target is ore. In order to meet the characteristics of stacking, adhesion, similarity, and irregularity between ores in target recognition, as well as the requirements for network speed, based on YOLOv5, an attention mechanism is incorporated. The attention mechanism module is used for feature screening to improve the precision of the detection network.

The CA attention mechanism can be regarded as a computational unit that augments the feature expression capability of mobile networks, accepting intermediate feature $X = [x_1, x_2, \ldots x_C] \in R^{C \times H \times W}$ as inputs and outputting enhanced features $Y = [y_1, y_2, \ldots y_C]$ of the same size as X.

The algorithm flowchart and network structure diagram of CA attention mechanism are shown in Figures 3 and 4.

The CA attention mechanism enables networks to have the ability to distinguish key area information, invest greater weight in specific areas, highlight and strengthen useful features, suppress and ignore irrelevant features, and reduce feature data loss caused by two-dimensional pooling operations (Shi et al. 2023). Firstly, this mechanism transforms the feature vector into a non-one-dimensional global pooling single vector while decomposing it into two independent spatial structures to achieve more effective data transmission and capture long-distance dependencies in a spatial direction while maintaining positional information. Secondly, The generated feature maps are transformed into direction-aware and position-sensitive ones and then applied to the input feature maps to strengthen the representation of the object of interest (Wen et al. 2023). Specifically, for the input dimension $C \times H \times W$ of the feature map X, first pooled using pooling kernels of sizes (H, 1) and (1, W) to average the coordinates in the X and Y directions in each channel, and acquire the output value of channel c having a height of h. as shown in Equation (1).

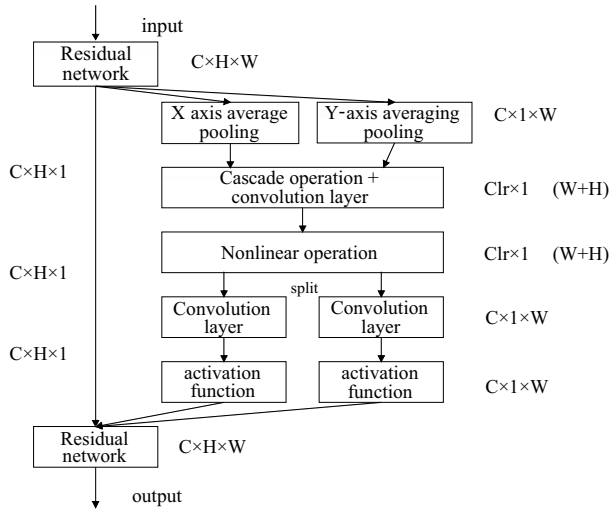


Fig. 3. Flowchart of CA algorithm

Rys. 3. Schemat blokowy algorytmu CA

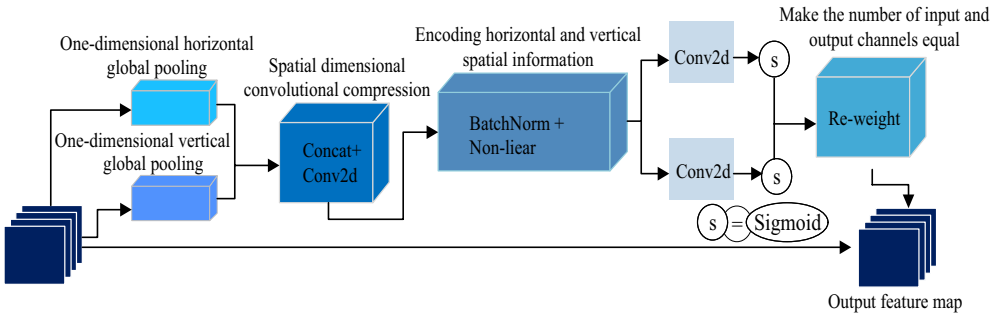


Fig. 4. Network structure diagram of CA

Rys. 4. Schemat struktury sieci CA

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \le i < w} x_c(h,i) \tag{1}$$

The output in channel c with width W is shown in Equation (2).

$$Z_c^w(w) = \frac{1}{H} \sum_{0 \le i < H} x_c(j,w) \tag{2}$$

Through the CA attention mechanism, coordinate data from the global receptive field can be implanted into the code to obtain accurate localization data. To achieve coordinate attention, it is necessary to cascade the feature maps obtained from two pooling operations and utilize a 1 × 1 communal convolutional transformation function F_1 is used to transform it, as shown in Equation (3).

$$k = \delta\left(F_1\left(\left[Z^{h_1} Z^w\right]\right)\right) \tag{3}$$

In the Equation, $Z^{h_1}Z^w$ represents cascading the feature maps obtained from two pooling operations, δ denotes the nonlinear activation function, and k is an intermediate feature map encoded in the horizontal and vertical directions. k is divided into two separate tensors k^h and k^w along the horizontal and vertical directions, and then two 1 × 1 convolutional transformation functions are utilized, which transforms into the same number of channels as the output X, as shown in Equations (4) and (5).

$$g^h = \sigma\left(F_n\left(k^h\right)\right) \tag{4}$$

$$g^w = \sigma\left(F_w\left(k^w\right)\right) \tag{5}$$

g^h and g^w respectively denote the weights in terms of dimensions for the feature maps in the horizontal and vertical directions. Finally, as shown in Equation (6), the final output after adding CA attention mechanism is obtained.

$$y_c\left(m,n\right) = x_c\left(m,n\right) \cdot g_c^h\left(m\right) \cdot g_c^w\left(n\right) \tag{6}$$

$x_c(m,n)$ and $y_c(m,n)$ denote the results for channel c at coordinate (m,n) in the input and output feature maps, respectively.

The improvement and optimization was carried out by adding CA modules to the initial network structure, and Table 1 shows the comparison of the test results before and after the model improvement.

Table 1. Added CA attention mechanism network test results

Tabela 1. Dodane wyniki testu sieci mechanizmu uwagi CA

Network structure	AP (precision)	F1	FPS (speed)
YOLOv5	0.968	0.94	120 f/s
CA+YOLOv5	0.975	0.95	100 s/s

By comparing the evaluation indicators before and after the model improvement, it can be concluded that the performance of the improved model has better results

2.2. Improvement of loss function

GIoU, as a loss function and insensitive to scale, is a good distance metric; compared with IOU, GIoU can reflect more accurately the overlap between the predicted and real frames. However, when the two boxes are not aligned accurately, the area of the smallest bounding box will increase, which will reduce the value of GIoU. When the two boxes belong to the inclusion relationship, GIoU will degenerate into IOU, and its relative position relationship cannot be distinguished (Song et al. 2022).

To solve this problem, by improving the inadequacy of the loss function of the target detection network, SIoU is adopted as the improved loss function. Because IoU and GIoU do not consider the direction between the real box and the prediction box, the convergence speed is slow, and the relative position relationship cannot be distinguished. Therefore, SIoU takes into account the vector angle between the actual box and the predicted box to redefine the loss function, which includes three-parts.

2.2.1. Angle loss

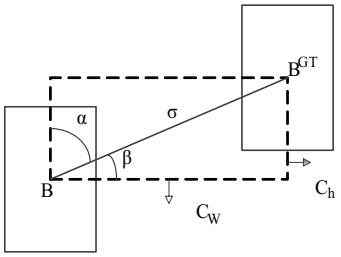


Fig. 5. Schematic diagram of angular loss

Rys. 5. Schematyczny schemat strat kątowych

$$\varepsilon = 1 - 2 \cdot \sin^2\left(\arcsin\left(\frac{C_h}{\sigma}\right) - \frac{\pi}{4}\right) \tag{7}$$

$$\text{å} = \cos\left(2 \cdot \left(\arcsin\left(\frac{C_h}{\sigma}\right) - \frac{\pi}{4}\right)\right)$$

Where, C_h epresents the height disparity between the center points of the real bounding box and the predicted bounding box. Additionally, σ indicates the separation between the center points of the real frame and the prediction frame.

$$\frac{C_h}{\sigma} = \sin(\alpha) \tag{8}$$

$$\sigma = \sqrt{\left(b_{C_x}^{gt} - b_{C_x}\right)^2 + \left(b_{C_y}^{gt} - b_{C_y}\right)^2} \tag{9}$$

$$C_h = \max\left(b_{C_y}^{gt},\, b_{C_y}\right) - \min\left(b_{C_y}^{gt} - b_{C_y}\right) \tag{10}$$

Where, $\left(b_{C_x}^{gt},\, b_{C_y}\right)$ represents the center coordinate of the actual frame while $\left(b_{C_x},\, b_{C_y}\right)$ stands for the center coordinate of the prediction frame.

2.2.2. Distance loss

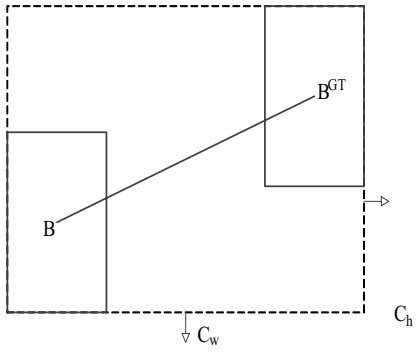


Fig. 6. Schematic diagram of distance loss

Rys. 6. Schematyczny schemat strat odległości

$$\Delta = \sum_{t=x,y} \left(1 - e^{-\gamma\rho_t}\right) = 2 - e^{-\gamma\rho_x} - e^{-\gamma\rho_y} \tag{11}$$

Where, $\rho_x = \left(\frac{b_{C_x}^{gt} - b_{C_x}}{C_w}\right)^2$, $\rho_y = \left(\frac{b_{C_y}^{gt} - b_{C_y}}{C_h}\right)^2$, $\gamma = 2 - \varepsilon$.

2.2.3. Shape loss

$$\tau = \sum_{t=w,h} \left(1 - e^{-w_t}\right)^{\theta} = \left(1 - e^{-w_w}\right)^{\theta} + \left(1 - e^{-w_h}\right)^{\theta} \tag{12}$$

Where, $w_w = \frac{\left|W - W^{gt}\right|}{\max\left(W, W^{gt}\right)}$, $w_w = \frac{\left|h - h^{gt}\right|}{\max\left(h, h^{gt}\right)}$. (w,h) and (w^{gt},h^{gt}) respectively denote the width and height of the predicted and actual frames, the value is an empirical one that indicates the degree of attention paid to shape loss.

In conclusion, the SIoU loss function's definition is presented in Equation (13).

$$Loss_{\mathrm{SIoU}} = 1 - \mathrm{IoU} + \frac{\Delta + \tau}{2} \tag{13}$$

The results of adding CA attention mechanism and improving the loss function in YOLOv5 network structure compared with other target detection algorithms are shown in Table 2.

Table. 2. Parameter change table before and after improvement

Tabela 2. Tabela zmian parametrów przed i po ulepszeniu

Network Structure	AP (precision)	FPS (speed)
YOLOv5	0.968	120 f/s
YOLOv7	0.970	76.9 f/s
YOLOv8	0.983	39.8 f/s
CA+YOLOv5	0.975	100 f/s
SIoU+CA+YOLOv5	0.980	100 f/s

3. Network evaluation index

3.1. Network training parameters

To address complex computing issues during network training and boost the speed of network training and operation, this paper selects the dynamic Pytorch framework as the

network framework of deep learning neural networks. Due to its high flexibility and easy debugging, data parameters can be easily transferred between CPU and GPU. By using parallel computing architecture CUDA and GPU acceleration, fast network training can be realized. Therefore, the system shown in Table 3 was built.

In this paper, Recall and Precision are regarded as the evaluation metrics of the YOLOv5 network model target detection algorithm. The recall indicates the ratio of all ore targets that are correctly predicted as ore, thus, it is possible to assess the comprehensiveness of the detection model performance tests. The precision reflects the actual percentage of ore in the target that is predicted to be ore. By contrasting the disparity between the model's predicted outcomes and the actual results, two evaluation indices, namely recall rate and precision, were computed.

Table. 3. Software and hardware system parameter table

Tabela 3. Tabela parametrów systemu oprogramowania i sprzętu

Hardware name	Parameter setting
GPU	Nvidia GeForce RTX 20
CPU	Intel Core i7-10700@2.90ghz Octanuclear
Graphics card	NVIDIA GeForce RTX 3060
Video memory	12G
Computer system	Win 10
Anaconda	2020.07
Keras	2.1.5
Python	3.8
Opencv-python	4.4.0.46
Pytorch	1.8
CUDA	10.0
CUDNN	7.6.1

Table. 4. Confusion matrix table

Tabela 4. Tabela macierzy pomyłek

Real situation	Forecast result	
	positive example	counterexample
Positive example	TP	FN
Counterexample	FP	TN

As shown in Table 4, TN (True Negatives) represents the quantity of correctly identified nonores; FP (False Positions) are the nonore quantities that have been incorrectly identified as ores; FN (False Negatives) stands for the count of ores that have been wrongly classified as nonores. Equations (14) and (15) are the expressions for the calculation of precision and recall rates.

$$\text{precision} = \frac{TP}{TP + FN} \tag{14}$$

$$\text{recall} = \frac{TP}{TP + FP} \tag{15}$$

In order to comprehensively evaluate the precision and recall values, the F_1 score is employed to assess the superiority of the model. The evaluation scale of F_1 is between [0,1], and a single value can be used to evaluate the given model to make its results more intuitive. The nearer it is to 1, the more excellent the model's performance, and vice versa. The calculation formula is shown in Equation (16).

$$\mathrm{F} = 2 \cdot \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{16}$$

3.2. Analysis of Network Evaluation Indicators

The changes in various evaluation indicators of the improved YOLOV5 model trained on a self-made ore dataset are shown in Figures 7, 8, 9, and 10. The overall loss value of the improved YOLOV5 loss function curve is lower than that before the improvement, and the oscillation amplitude of the loss function curve is significantly reduced, indicating that the improved training process is smoother and the training effect is also improved to a certain extent; The area encompassed by the PR curves and the axes signifies the model's checking completeness and precision, and additionally reflects the variation in precision before and after improvement; The F_1 value change curve shows that the improved model's F_1 value is generally greater than the original F_1 value during the confidence change process, indicating a significant improvement in the detection precision of the improved model; The vibration amplitude of the precision curve is significantly reduced compared to before improvement, and it can converge quickly to stabilize. Through the analysis of evaluation indicators, the AP value increased from 96.8% to 98%, an increase of 1.2%. The F_1 value increased from 0.94 to 0.95, an increase of 0.01. The precision has increased from 96% to 98%, an increase of 2%, indicating that the enhanced model has superior detection performance.

3.3. Analysis of ore image recognition results

Randomly select four different ore images from the test set and use the improved model to recognize the ore particles in each image. The recognition images of the ore before and after improvement, as well as the corresponding local enlarged images, are shown in Figures 11, 12, 13, and 14.

From the results of ore identification on the conveyor belt by YOLOv5 before improvement in Figure 11 to Figure 14, it can be observed that most of the different kinds

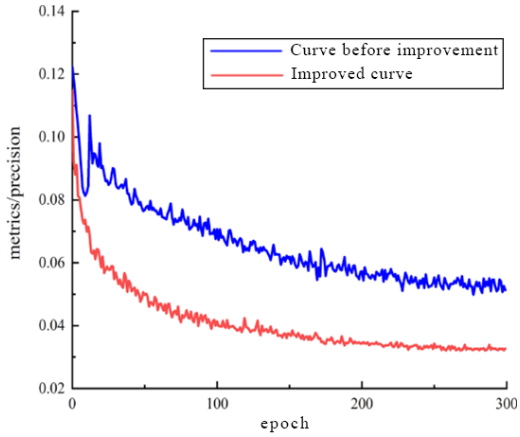


Fig. 7. Curve of loss function before and after improvement

Rys. 7. Krzywa funkcji strat przed i po poprawie

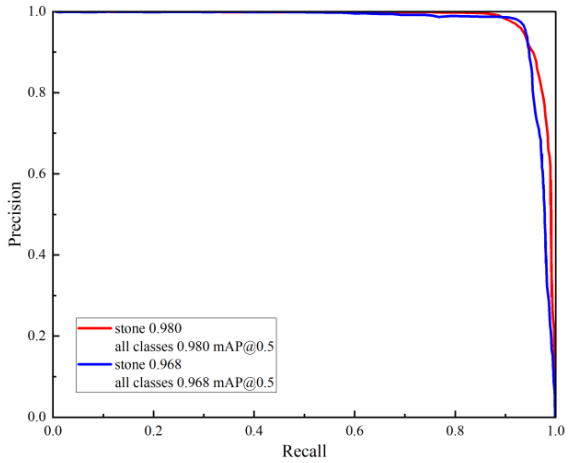


Fig. 8. PR curve before and after improvement

Rys. 8. Krzywa PR przed i po poprawie

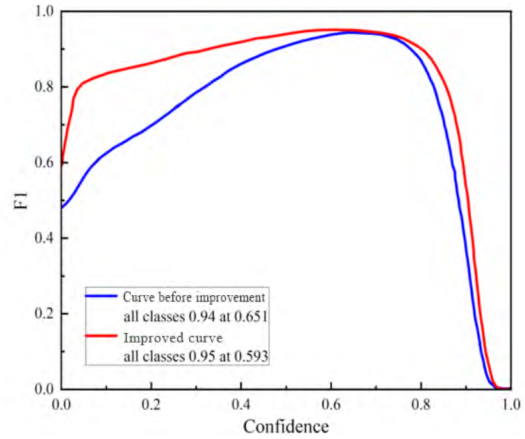


Fig. 9. F1 value curve before and after improvement

Rys. 9. Krzywa wartości F1 przed i po poprawie

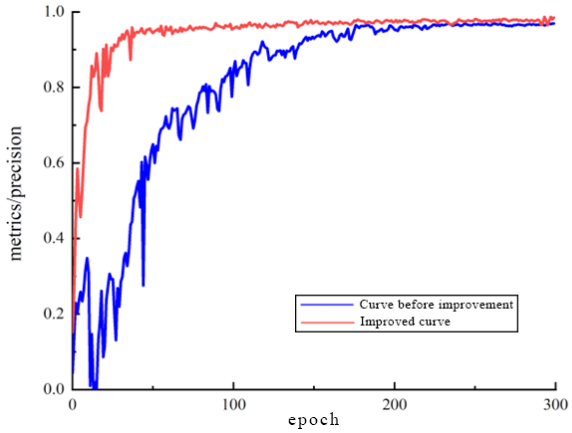
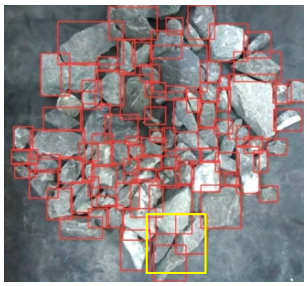


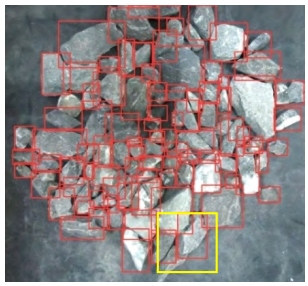
Fig. 10. Precision rate curve before and after improvement

Rys. 10. Krzywa współczynnika precyzji przed i po poprawie

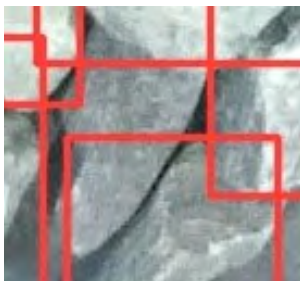
of ore particles can be accurately identified, but due to overlap and adhesion among ore particles, there will be cases of missing identification and misidentification of individual ores in the identification results of YOLOV5 model. On the basis of the YOLOv5 model, the CA attention mechanism is added, and the GIoU function is replaced by the SIoU function in the improved CA-YOLOV5 network model. Based on the model, the sensitivity of ore characteristics identification is improved, and the precision is improved to reduce the identification error. The visualization of feature extraction was incorporated to study the ore feature extraction process so as to enhance the recognition precision,



(a) Pre-improved ore identification map



(b) Improved ore identification map

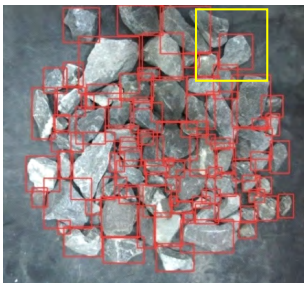


(c) Partial magnification of ore before improvement

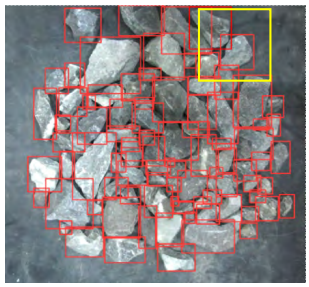
(d) Partial magnification of ore after improvement

Fig. 11. Comparison of the first group of ore image recognition

Rys. 11. Porównanie pierwszej grupy rozpoznawania obrazu rudy



(a) Pre-improved ore identification map



(b) Improved ore identification map

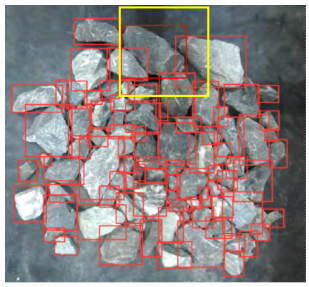
(c) Partial magnification of ore before improvement

(d) Partial magnification of ore after improvement

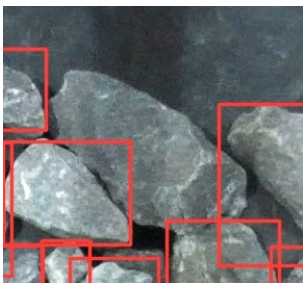
Fig. 12. Comparison of the second group of ore image recognition

Rys. 12. Porównanie drugiej grupy rozpoznawania obrazu rudy

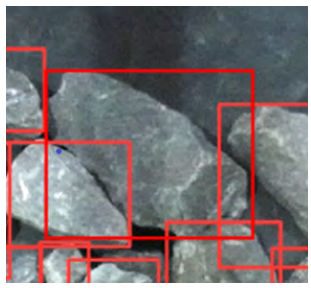
(a) Pre-improved ore identification map



(b) Improved ore identification map



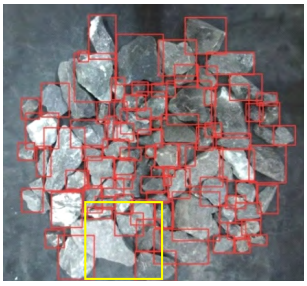
(c) Partial magnification of ore before improvement



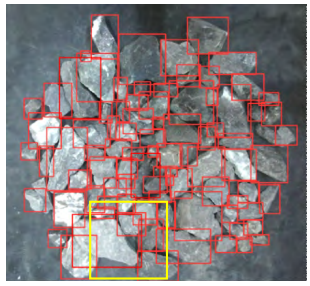
(d) Partial magnification of ore after improvement

Fig. 13. Comparison of the third group of ore image recognition

Rys. 13. Porównanie trzeciej grupy rozpoznawania obrazu rudy

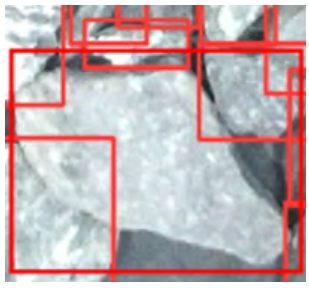


(a) Pre-improved ore identification map



(b) Improved ore identification map

(c) Partial magnification of ore before improvement



(d) Partial magnification of ore after improvement

Fig. 14. Comparison of the fourth group of ore image recognition

Rys. 14. Porównanie czwartej grupy rozpoznawania obrazu rudy

and the GIoU function made up for the failure to consider the angular difference and the disparity between the predicted bounding box and the actual bounding box, further improving the recognition precision. Compared with the recognition effect before the improvement, the problems of missing recognition and misidentification in the ore image have been basically solved, which indicates that the enhanced approach adopted in this paper has a remarkable result on the recognition of stacked ore and cohesive ore particle images on the conveyor belt.

Conclusions

1. To detect and identify the ore on the conveyor belt more accurately, this paper improved the YOLOv5 model by adding a CA module to extract the multi-scale characteristic information of the ore and enhance the network feature extraction capability. The improved SIoU loss function accelerates the model's convergence rate and enhances the detection precision.
2. This study produced an image data set for the training, verification, and testing of the model. The experimental contrast between the improved SIoU+CA+YOLOv5 model and the YOLOv5, YOLOv7, YOLOv8, and CA+YOLOv5 models demonstrated that the enhanced model presented in this paper can all boost the precision, F1 value and detection precision of ore image recognition, which are 1.2, 1 and 2 percentage points higher than the base model, respectively. With the improvement of the complexity of the upgraded model, the detection speed also drops to 100 f/s. Nevertheless, it is still capable of meeting the demands of real-time recognition. The model has good performance for ore identification and detection in the case of stacking and adhesion and can be well applied to the ore detection task on the conveyor belt.

Supported by the National Natural Science Foundation of China under the project "Study on the damage evolution and dissociation mechanism of encapsulated mineral interface under multi-point symmetric ultrasonic loading" (52364025).

The Authors have no conflict of interest to declare.

REFERENCES

Dai et al. 2023 – Dai, M., Dorjoy, M.H., Miao, H. and Zhang, S. 2023. A New Pest Detection Method Based on Improved YOLOv5m. *Insects* 14(1), DOI: 10.3390/insects14010054.

Dawson et al. 2020 – Dawson, M., Perez, A. and Sylvestre, S. 2020. Artificial Neural Networks Solve Musical Problems With Fourier Phase Spaces. *Scientific Reports* 10(1), DOI: 10.1038/s41598-020-64229-4.

Fan et al. 2025 – Fan, Q., Li, Y., Deveci, M., Zhong, K. and Kadry, S. 2025. LUD-YOLO: A novel lightweight object detection network for unmanned aerial vehicle. *Information Sciences* 686, DOI: 10.1016/j.ins.2024.121366.

Fu et al. 2018 – Fu, L., Feng, Y., Majeed, Y., Zhang, X., Zhang, J., Karkee, M. and Zhang, Q. 2018. Kiwifruit detection in field images using Faster R-CNN with ZFNet. *IFAC – Papers Online* 51(17), DOI: 10.1016/j.ifacol.2018.08.059.

Guo et al. 2022 – Guo, R., Zuo, Z., Su, S., Sun, B. and Lakshmanna, K. 2022. A Surface Target Recognition Algorithm Based on Coordinate Attention and Double-Layer Cascade. *Wireless Communications and Mobile Computing*, DOI: 10.1155/2022/6317691.

He et al. 2020 – He, K., Gkioxari, G., Dollár, P., Girshick, R. 2020. Mask R-CNN. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 42(2), pp. 386–397, DOI: 10.1109/TPAMI.2018.2844175.

Li et al. 2024 – Li, R., Sun, T., Dong, P., Wang, Q., Li, Y. and Sun, C. 2024. MSF-CSPNet: A Specially Designed Backbone Network for Faster R-CNN. *IEEE Access* 12, DOI: 10.1109/ACCESS.2024.3386788.

Liu et al. 2021 – Liu, C., Feng, L., Liu, G., Wang, H. and Liu, S. 2021. Bottom-up broadcast neural network for music genre classification. *Multimedia Tools and Applications* 80(5), DOI: 10.1007/s11042-020-09643-6.

Redmon et al. 2015 – Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. 2015. You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, DOI: 10.1109/CVPR.2016.91.

Shen et al. 2024 – Shen, M., Liu, Y., Chen, J., Ye, K., Gao, H., Che, J., Wang, Q., He, H., Liu, J., Wang, Y. and Jiang, H.J. 2024. Defect detection of printed circuit board assembly based on YOLOv5. *Scientific reports* 14(1), DOI: 10.1038/s41598-024-70176-1.

Shi et al. 2023 – Shi, H., Xiao, W., Zhu, S., Li, L. and Zhang, J. 2023. Ca-yolov5: Detection model for healthy and diseased silkworms in mixed conditions based on improved yolov5. *International Journal of Agricultural and Biological Engineering* 16(6), DOI: 10.25165/j.ijabe.20231606.7854.

Song et al. 2022 – Song, Y., Zhang, P., Huang, W., Zha, Y., You, T. and Zhang, Y. 2022. Object detection based on cortex hierarchical activation in border sensitive mechanism and classification-GIou joint representation. *Pattern Recognition* 137(2), DOI: 10.1016/j.patcog.2022.109278.

Sun et al. 2023 – Sun, D., Zhang, L., Wang, J., Liu, X., Wang, Z., Hui, Z. and Wang, J. 2023. Efficient and accurate detection of herd pigs based on Ghost-YOLOv7-SIoU. *Neural Computing and Applications* 36(5), DOI: 10.1007/s00521-023-09093-9.

Wen et al. 2023 – Wen, G., Li, S., Liu, F., Luo, X., Er, M.J., Mahmud, M. and Wu, T. 2023. YOLOv5s-CA: A Modified YOLOv5s Network with Coordinate Attention for Underwater Target Detection. *Sensors* 23(7), DOI: 10.3390/s23073367.

Wu, Z. 2023. Detect small object based on FCOS and adaptive feature fusion. *Journal of Physics: Conference Series* 2580, DOI: 10.1088/1742-6596/2580/1/012005.

Xie et al. 2024 – Xie, X., Cheng, G., Wang, J., Li, K., Yao, X. and Han, J. 2024. Oriented R-CNN and Beyond. *International Journal of Computer Vision* 132(7), DOI: 10.1007/s11263-024-01989-w.

Yan et al. 2022 – Yan, C., Zhang, H., Li, X. and Yuan, D. 2022. R-SSD: refined single shot multibox detector for pedestrian detection. *Applied Intelligence* 52(9), DOI: 10.1007/s10489-021-02798-1.

Yang et al. 2023 – Yang, F., Huo, J., Cheng, X., Chen, H. and Shi, Y. 2023. An Improved Mask R-CNN Micro-Crack Detection Model for the Surface of Metal Structural Parts. *Sensors* 24(1), DOI: 10.3390/s24010062.

Zhang, Y. 2024. Research on fabric yarn detection based on improved fast R-CNN algorithm. *Applied Mathematics and Nonlinear Sciences* 9(1), DOI: 10.2478/amns.2023.2.00449.

Zhang et al. 2021 – Zhang, S., Tuo, H., Zhong, H. and Jing, Z. 2021. Aerial image detection and recognition system based on deep neural network. *Aerospace Systems* 4(2), DOI: 10.1007/s42401-020-00077-4.

Zhang et al. 2023 – Zhang, S., Chen, Y., Gao, Y., Wang, H. 2023. Lightweight Target Detection Algorithm for Aerial Images. *IEEE Access* 11, DOI: 10.1109/ACCESS.2023.3337157.

ORE IMAGE TARGET DETECTION BASED ON IMPROVED YOLOV5 NETWORK

Keywords

YOLOv5, object detection, coordinate attention for efficient mobile network design, feature extraction

Abstract

The existing target detection algorithms detect the ore on the conveyor belt after the crushing process with low precision and slow detection speed. This leads to challenges in achieving a balance between precision and speed, to enhance the detection precision and speed of ore, and in view of the problems of leakage, misdetection, and insufficient feature extraction of YOLOv5 in the task of ore image detection; this study presents a target detection approach relying on the CA attention mechanism (Coordinate attention for efficient mobile network design), the SIoU loss function and the target detection algorithm YOLOv5 combination of ore image particle target detection method. Integrating the CA attention mechanism into the YOLOv5 backbone feature network enhances the feature learning and extraction of ore images, thereby improving the precision of the detection model; the SIoU loss function is refined to boost the recognition precision of the network on ore images and address the shortcomings of the original loss function that fails to take angular loss, distance loss, and shape loss into account, thereby further improving the precision and speed of ore image detection. The experimental findings demonstrate that the AP value, value, and precision rate are improved compared with the pre-improved algorithm. The CA-YOLOv5 method is verified to be fast, effective, and advanced and provides a foundation for real-time target detection of ores on conveyor belts in subsequent intelligent mine production.

WYKRYWANIE CELU OBRAZU RUDY W OPARCIU O ULEPSZONĄ SIEĆ YOLOV5

Słowa kluczowe

YOLOv5, wykrywanie obiektów, koordynacja uwagi dla wydajnego projektowania sieci mobilnych, ekstrakcja cech

Streszczenie

Istniejące algorytmy wykrywania celu wykrywają rudę na taśmie przenośnika po procesie kruszenia z niską precyzją i niską szybkością wykrywania. Prowadzi to do wyzwań związanych z osiągnięciem równowagi między precyzją i szybkością, w celu zwiększenia precyzji i szybkości wykrywania rudy, a także ze względu na problemy z wyciekami, błędnym wykrywaniem i niewystarczającą ekstrakcją cech YOLOv5 w zadaniu wykrywania obrazu rudy; niniejsze badanie przedstawia podejście do wykrywania celu polegające na mechanizmie uwagi CA (*Coordinate attention for efficient mobile network design*), funkcji straty SIoU i kombinacji algorytmu wykrywania

celu YOLOv5 w połączeniu z metodą wykrywania celu cząstek obrazu rudy. Zintegrowanie mechanizmu uwagi CA z siecią funkcji szkieletowych YOLOv5 usprawnia uczenie się funkcji i ekstrakcję obrazów rudy, tym samym zwiększając precyzję modelu wykrywania; funkcja straty SIoU została udoskonalona w celu zwiększenia precyzji rozpoznawania sieci na obrazach rudy i usunięcia niedociągnięć oryginalnej funkcji straty, która nie uwzględnia strat kątowych, strat odległości i strat kształtu, co jeszcze bardziej poprawia precyzję i szybkość wykrywania obrazów rudy. Wyniki eksperymentów pokazują, że wartość AP, wartość i wskaźnik precyzji są lepsze w porównaniu z wcześniej ulepszonym algorytmem. Metoda CA-YOLOv5 została zweryfikowana jako szybka, skuteczna i zaawansowana oraz stanowi podstawę do wykrywania celów rud na taśmach przenośnikowych w czasie rzeczywistym w późniejszej inteligentnej produkcji kopalnianej.