GOSPODARKA SUROWCAMI MINERALNYMI – MINERAL RESOURCES MANAGEMENT

2025 Volume 41 Issue 3 Pages 73–91

DOI: 10.24425/gsm.2025.155335

LIRONG YANG[1], HUI YANG[2], XIAOLONG ZHU[3]

Stacked fine ore detection method based on SPDM-YOLO network

Introduction

Mineral resources serve as the material foundation for national economic development. With the continuous advancement of science and technology, the mining industry is gradually evolving toward intelligence, automation, and informatization, making the digital transformation of traditional mining an irreversible trend (Shao et al. 2024). Consequently, the demand for image recognition technologies in key processes such as ore particle size detection and mineral identification is steadily increasing. In ore processing, accurate identification of ore particle size not only optimizes the crushing and screening processes but also reduces equipment wear and energy consumption (Tian et al. 2021).

✉ Corresponding Author: Lirong Yang; e-mail: 3397205923@qq.com

[1] Jiangxi Mining and Metallurgical Engineering Research Center, China; Jiangxi University of Science and Technology, School of Electromechanical Engineering, China; e-mail: 3397205923@qq.com

[2] Jiangxi University of Science and Technology, School of Electromechanical Engineering, China; ORCID iD: 0009-0001-6736-7638; e-mail: 1850881582@qq.com

[3] Jiangxi University of Science and Technology, School of Electromechanical Engineering, China; e-mail:2774159670@qq.com

In recent years, image processing techniques based on deep learning have achieved remarkable results in industrial inspection, particularly convolutional neural networks (CNNs), which have demonstrated outstanding performance in object detection tasks (Xiaoyan et al. 2022). Currently, mainstream object detection algorithms can be categorized into three types: two-stage algorithms, one-stage algorithms, and Transformer-based DETR models (Zhu et al. 2024). Two-stage methods, such as R-CNN (Usha et al. 2022) and Mask R-CNN (Shunling et al. 2022), offer high detection accuracy but often make predictions only on high-level feature maps, resulting in a high miss rate for small ore particles. Transformer-based DETR models (Li et al. 2024) introduce global attention mechanisms to enhance end-to-end detection capabilities but exhibit limited adaptability in domains with small sample sizes and complex backgrounds, such as those in image datasets. One-stage methods, such as YOLO (Longzhen et al. 2022) and SSD (Yuhuan et al. 2023), maintain relatively high accuracy while achieving faster detection speeds. However, while SSD improves small-object detection through multi-scale feature maps, its default box design lacks robustness for covering irregularly shaped ore particles. The YOLO series strikes a better balance between speed and accuracy, but still requires multiple module enhancements to effectively address feature interference caused by stacked ores. Overall, the YOLO family remains the most widely used object detection solution due to its excellent performance and strong engineering foundation.

As one of the latest models in the YOLO series, YOLOv9 incorporates dynamic pyramid convolution (DPConv) and a task-decoupled architecture, resulting in further improvements in both detection accuracy and inference speed (Chien et al. 2024). These innovations offer a new and promising approach for detecting fine ore particles that are stacked.

For the detection of stacked fine-grained ores (such as tungsten ores with particle sizes ranging from 6 mm to 50 mm), existing object detection algorithms still face challenges in terms of accuracy. To improve the detection accuracy of fine-grained stacked ores in mineral images, this paper proposes a novel architecture called SPDM-YOLO. The proposed method has been trained and evaluated on a bespoke dataset of ores. The findings demonstrate that the proposed method in this paper enhances the detection of fine ores in images compared to currently available state-of-the-art methods. The principal contributions of this paper are as follows:

1. We introduced a novel convolutional neural network (CNN) building block called SPD-Conv to replace the conventional stepwise convolution and pooling layers commonly used in traditional CNNs. The design of a space-to-depth layer and a non-step-length convolutional layer addresses the challenges associated with low-resolution images and small object detection.
2. We introduced a lightweight attention module (MLCA) to improve the extraction of channel and spatial information. This module is capable of effectively combining local and overall features, thereby ensuring the preservation of important feature regions of small objects and enhancing the expressive power of the network.

We proposed a new bounding box regression loss function named Inner-FocalerIoU to replace the traditional CIoU loss function. By assisting the bounding box to accelerate the

training convergence as well as focusing on the complex samples by using linear interval mapping, the problem of tiny objects with different sensitivities to position and scale deviations is solved to enhance the detection performance.

1. The basic architecture of YOLOv9

1.1. Input

The input layer of the YOLOv9 model accepts fixed-size images, typically in RGB format with three channels. It is also subject to normalization, which aims to enhance the comparability of images under varying lighting conditions and scales, thereby contributing to the stability and accuracy of subsequent feature extraction. This stage is designed so that the model can handle images of different sizes while maintaining efficient computational performance.

1.2. Backbone

YOLOv9 uses CSPDarknet as its backbone network. CSPDarknet is a highly optimized deep convolutional neural network comprising multiple convolutional and pooling layers (Pan et al. 2024). The network performs deep learning through a series of residual blocks, each of which contains multiple convolutional layers inside. Through this hierarchical convolution and pooling operation, the network can efficiently extract detailed information and high-level features from the image, capturing different scales and shapes of the target object. Meanwhile, the backbone network introduces a new lightweight network structure, GELAN, which combines the advantages of CSPDarknet's enriched gradient, reduced redundancy, and lower computation and the strengths of ELAN's highly efficient feature aggregation to extract the image features better and lay an accurate foundation for the subsequent detection.

1.3. Neck and head

In YOLOv9, the Neck section is mainly responsible for connecting the backbone network to the detection head. This part utilizes a feature pyramid network (FPN) structure to enhance the model's detection capability through multi-scale feature fusion. It also utilizes CSP (Cross Stage Partial) convolutional blocks for channel segmentation and feature map connection, enabling the efficient integration of features from different layers. This design enables the network to acquire feature information from multiple layers, which significantly

improves the accuracy and robustness of target detection. The Head part of YOLOv9 adopts a decoupled head design, which transforms the regression task into a classification task in the bounding box branch. A multi-scale prediction strategy is applied; this strategy can perform target detection at different scales to improve the adaptability and detection ability of the model.

1.4. Auxiliary section

The additions to YOLOv9 relative to previous versions are used to improve training reliability. It contains a main branch responsible for the final target detection output, an auxiliary reversible branch to solve the problem of information loss due to the depth of the network, multiple levels of auxiliary information to solve the problem of error accumulation in deep supervision, and the computation of the auxiliary head will fuse multiple feature maps.

2. Improved YOLOv9 model

To improve the accuracy and robustness of ore detection, this chapter proposes an enhanced YOLOv9-based algorithm, named SPDM-YOLOv9. Considering the characteristics of stacked fine ore particles, such as dense distribution, irregular shapes and frequent occlusion, the algorithm incorporates space-to-depth convolution (SPD-Conv). This helps preserve fine-grained spatial details while increasing the feature depth. Meanwhile, the hybrid local channel attention module (MLCA) is introduced to strengthen the model's focus on subtle and locally relevant features, which is crucial for distinguishing between overlapping or adhesive ores. Additionally, to further tailor the model to the ore detection task, an Inner-FocalerIoU loss function is proposed, combining the advantages of Inner-IoU and Focaler-IoU to improve localization accuracy in complex scenes. Compared with traditional object detection tasks, ore images present unique challenges, and these targeted improvements address the specific demands of this domain. The SPDM-YOLOv9 model is shown in Figure 1.

2.1. SPD-Conv module

In traditional convolutional neural network (CNN) architectures, stepwise convolution and pooling layers gradually reduce the spatial resolution of feature maps as the network deepens. This often results in the loss of fine-grained details, particularly for small objects, thereby reducing detection accuracy. To address this, the SPD-Conv module is introduced to replace standard convolution and pooling layers, helping preserve spatial information and enhance detection accuracy for such complex targets. The schematic diagram of SPD-Conv is shown in Figure 2.

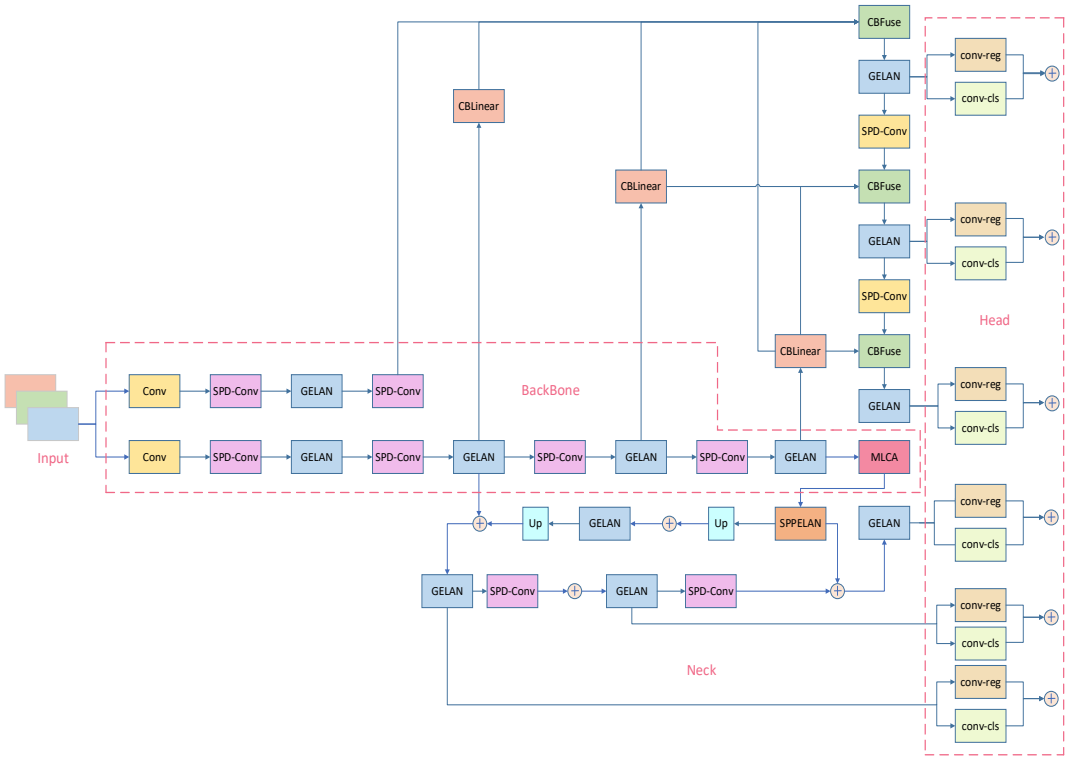


Fig. 1. SPDM-YOLOv9 network model

Rys. 1. Model sieci SPDM-YOLOv9

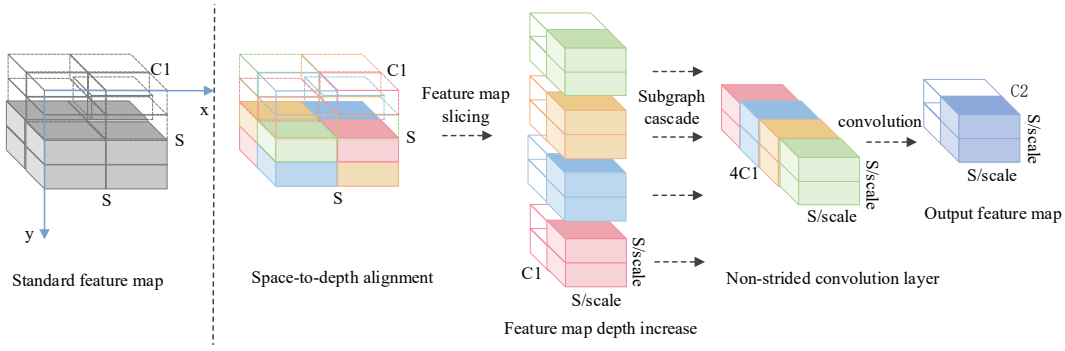


Fig. 2. Schematic diagram of SPD-Conv

Rys. 2. Schematyczny diagram SPD-Conv

SPD-Conv (Yang et al. 2023) consists of a Space-to-depth (SPD) layer and a non-step-length convolutional layer. The SPD layer transforms spatial information into the depth dimension, increasing the feature map depth without losing important details. Following this, a non-step-length convolutional layer is applied, which performs feature extraction without reducing the spatial size of the feature map. This design helps retain fine-grained details, which are crucial for detecting small and dense targets. The specific steps are as follows.

1. Feature map slicing: downsampling using a *scale* factor to divide an $scale^2$ subfeature map from the central feature map (X) of size $S \cdot S \cdot C1$.
2. Sub-feature map concatenation: these sub-feature maps are spliced along the channel dimension, resulting in a new feature map of size $\frac{S}{scale} \cdot \frac{S}{scale} \cdot scale^2 C1$ (X1). That is, the spatial dimension of the new feature map is the original $\frac{1}{scale}$, and the channel dimension is increased by a factor of $scale^2$, which maintains all the information in the original feature map and reduces its spatial resolution.
3. Non-step-length convolution: a convolution layer with a step size of 1 is applied to the new feature map formed, and a filter is used to extract important features and reduce the dimensionality of the channels, resulting in an output feature map of size $\frac{S}{scale} \cdot \frac{S}{scale} \cdot C2$ (X2).

2.2. MLCA module

The attention mechanism can filter out key information in the target task from a large amount of data, thereby overcoming the limitations of traditional models in long-distance dependent processing. In this study, the MLCA attention mechanism is introduced for the stacked ore target recognition task, which enables the model to focus on analyzing the region where the stacked ore is located.

MLCA (Mixed Local Channel Attention) (Cheng et al. 2022) is a lightweight attention module designed to enhance both local and global feature representation. When integrated into the backbone network, it enables the model to effectively differentiate channel-wise priorities at multiple levels, thereby capturing more detailed and comprehensive features and improving its ability to process fine-grained information. MLCA schematic diagram as shown in Figure 3.

Traditional attention mechanisms such as SE (Squeeze and Excitation) and ECA (Efficient Channel Attention), which both compress the entire channel feature map into a single value, ignore the spatial information within each channel. While spatial feature information is crucial for constructing spatial attention maps and contains only feature channel information, most of the traditional attention mechanisms may result in the loss of important feature information.

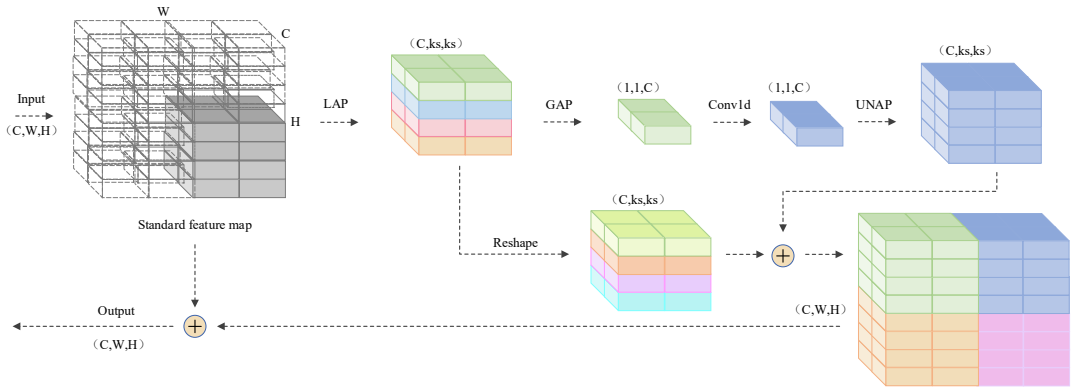


Fig. 3. Schematic diagram of MLCA

Rys. 3. Schematyczny diagram MLCA

MLCA incorporates spatial information, unlike these traditional channel attention mechanisms. It integrates channel information, spatial information, local channel information, and global channel information into a single piece. It improves model representability without adding undue computational complexity. The specific steps are as follows.

1. Division of the input feature map: the MLCA input feature map is divided into small blocks to capture local information through adaptive pooling, ensuring that spatial features within each block are considered and preventing the loss of important spatial information.
2. Local information branch: this branch contains local spatial information and applies a local SE attention mechanism to each parcel, preserving important local spatial features within each patch. The original resolution of the two vectors is then recovered by one-dimensional convolution as well as inverse pooling. The convolution kernel size k is proportional to the channel dimension C with the following formula:

$$k = \Phi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{1}$$

   - γ and b – hyperparameters with default value of 2,
   - *odd* – means k is *odd*, and *add* 1 if k is even.
3. Global information branch: this branch obtains global information through convolution and reshaping, considering the entire feature map to capture global dependencies and interactions between different regions of the image.
4. Combination of local and global attention: the two parts of the information are fused after acquiring the information through the above two branches to achieve the goal of hybrid attention and enhance the model's representational ability.

2.3. Inner-FocalerIoU loss function

The IoU (Intersection over Union) loss function is widely used in computer vision tasks, particularly for evaluating regression performance and accelerating convergence during bounding box prediction. In YOLOv9, the bounding box regression adopts the CIoU (Complete IoU) loss function. However, the ore images in this study contain particles with irregular shapes, including small and densely stacked ores, which often lead to a high degree of bounding box overlap. In such scenarios, CIoU may fail to accurately measure the overlap, resulting in suboptimal regression performance for small, overlapping objects.

In order to improve the accuracy of the overlap calculation of the bounding box, the Inner-FocalerIoU loss function replaces the CIoU. Inner-FocalerIoU combines the advantages of Inner-IoU and Focaler-IoU. Inner-IoU (Zhang et al. 2023) is a proposed method that utilizes IoU to calculate loss by incorporating auxiliary boundaries. The size of the auxiliary boundary can be controlled by the scale factor ratio to calculate the loss to speed up the convergence. Its calculation expression is as follows.

$$b_l^{gt} = x_c^{gt} - \frac{w^{gt} \cdot ratio}{2}, \quad b_r^{gt} = x_c^{gt} + \frac{w^{gt} \cdot ratio}{2} \tag{2}$$

$$b_t^{gt} = y_c^{gt} - \frac{h^{gt} \cdot ratio}{2}, \quad b_b^{gt} = y_c^{gt} + \frac{h^{gt} \cdot ratio}{2} \tag{3}$$

$$b_l = x_c - \frac{w \cdot ratio}{2}, \quad b_r = x_c + \frac{w \cdot ratio}{2} \tag{4}$$

$$b_t = y_c - \frac{h \cdot ratio}{2}, \quad b_b = y_c + \frac{h \cdot ratio}{2} \tag{5}$$

$$inter = \left(\min\left(b_\tau^{gt}, b_r\right) - \max\left(b_l^{gt}, b_l\right)\right) \cdot \left(\min\left(b_b^{gt}, b_b\right) - \max\left(b_t^{gt}, b_t\right)\right) \tag{6}$$

$$union = \left(w^{gt} \cdot h^{gt}\right) \cdot (ratio)^2 + \left(w \cdot h\right) \cdot (ratio)^2 - inter \tag{7}$$

$$IoU^{inner} = \frac{inter}{union} \tag{8}$$

Focaler-IoU (Jia et al. 2022) employs a linear interval mapping approach to reconstruct the IoU loss, enhancing bounding box edge regression. This method enables the model to focus on different regression samples across various detection tasks, thereby improving detection

performance, particularly on challenging or hard-to-localize targets. Its computational expression is shown below.

$$Focaler-IoU=\begin{cases}0 & if\ IoU<d\\ \dfrac{IoU-d}{u-d} & if\ d\ll IoU\ll u\\ 1 & if\ IoU>u\end{cases} \tag{9}$$

where $[d,u] \in [0,1]$ can adjust the values of d and u thus prompting Focaler-IoU to focus on different regression samples. Integrating Inner-IoU into Focaler-IoU edge loss not only ensures higher recognition accuracy for small ore particle targets and more tolerance for large ore particle targets when calculating overlap, but also effectively accelerates the process of edge regression. The Inner-FocalerIoU-based border regression loss function is as follows.

$$L_{Focaler-IoU}=1-Focaler-IoU \tag{10}$$

$$L_{Inner-FocalerIoU}=L_{Focaler-IoU}+IoU-IoU^{inner} \tag{11}$$

3. Model training and experimentation

3.1. Model training environment platform

We chose PyTorch as the platform for developing the YOLOv9 network due to its outstanding flexibility, intuitive debugging features, and efficient migration of data parameters between CPU and GPU. The hardware and software parameters in Table 1 are applied for both model training and experimental conditions.

We trained and validated the model using the network training parameters specified in Table 2. The training and experimental platform was developed in Visual Studio Code to enable effective runtime execution.

3.2. Improvement of modelling effects

Due to the absence of publicly available datasets tailored to ore detection tasks, this study constructs a custom dataset comprising ore images collected under laboratory conditions that simulate a mining environment. To reflect the complexity of real-world scenarios,

Table 1. Hardware and software system parameters

Tabela 1. Parametry systemu sprzętowego i programowego

Hardware Name	Parameter Configuration
Graphics Card	NVIDIA GeForce RTX 4060
CPU	Gen Intel Core i5-12490F
Computer System	Windows 10
Anaconda	23.1.0
CUDA	12.4
CUDNN	8.4.1
Python	3.8.19
Pytorch	2.2.0
Opencv-python	4.9.0.80
Torchvision	0.17.0

Table 2. Network training parameters

Tabela 2. Parametry treningu sieci

Parameter Name	Parameter Configuration
lr0	0.01
epochs	300
patience	0
batch	8
workers	0
optimizer	auto

ores of varying particle sizes (ranging from 6 mm to 50 mm) were randomly scattered and stacked in a controlled environment, simulating practical production conditions such as adhesion, overlap, and irregular morphology. A total of 300 ore images were captured, and Mosaic (Li et al. 2023) data augmentation was applied to expand the dataset and improve generalization. While this controlled setup allows for targeted analysis of stacked and scattered ores, it is acknowledged that the simplified laboratory environment cannot fully replicate the complexity of actual mining operations, such as variable lighting, dust interference, and continuous ore flow. However, the experimental design provides

a foundational step for model validation under partially controlled conditions, offering valuable insights for future field deployments.

In this paper, the improved SPDM-YOLO network adopts the SPD-Conv module instead of the traditional convolutional pooling layer, adds the MLCA intentional force mechanism module, and replaces the CIoU loss function in the bounding box with the Inner-FocalerIoU loss function. This will enable the model to learn the semantic information of different layers, thereby improving its perceptual ability, reducing redundant information in the feature map, and enhancing computational efficiency. It also enables the model to focus more on ore targets with small particles, thereby improving recognition accuracy. The recognition effect is shown in Figure 4.

(a) Original image (b) Ore recognition image

Fig. 4. Identification diagram of the model

Rys. 4. Diagram identyfikacyjny modelu

3.3. Analysis of the experimental results

To validate the effectiveness of the improved model (SPDM-YOLO), the model was evaluated from both qualitative and quantitative perspectives. From a qualitative evaluation perspective, the detection differences between the SPDM-YOLO model and other models were compared to assess model performance. The evaluation indexes selected for comparative analysis included accuracy (P), detection precision (PR), and mean average precision (mAP). From a quantitative evaluation perspective, the detection differences before and after model improvement were compared to assess the model's performance, consistent with the selection of evaluation indexes in the

qualitative evaluation perspective. Additionally, ablation experiments were conducted for comparative analysis.

(1) Qualitative evaluation

Accuracy P is the proportion of samples that actually belong to the positive class among those judged as a positive class by the model. Generally speaking, the higher the accuracy rate, the more accurate the model is in predicting the positive class and the better its performance is. The PR curve can be used to describe the comprehensive performance of the model, which calculates the P value and R value under different thresholds by setting different confidence thresholds. The expression is as follows:

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{12}$$

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{13}$$

- TN – the number of correctly identified non-ore,
  FP – the number of non-ore incorrectly identified as ore,
  FN – the number of ores incorrectly identified as non-ore.

The larger the PR curve and the area enclosed by the *x* and *y* axes of the model, the higher the detection accuracy and completeness rate will be indicated.

AP (average precision) is the calculation of the average accuracy of the model for a single category. A numerical form of the assessment metric AP can be obtained by calculating the average of the precision values corresponding to each recall value based on the PR curve. According to the standard, the AP calculation can be defined as the area of the interpolated PR curve enveloped with the x-axis. This approach is known as AUC (area under the curve). Its calculation formula is:

$$AP = \sum_{i=1}^{n-1} \left(r_{i+1} - r_i \right) p_{interp} \left(r_{i+1} \right) \tag{14}$$

- r_i – the recall value corresponding to the first interpolation at the first interpolation of the precision interpolation segment in ascending order.

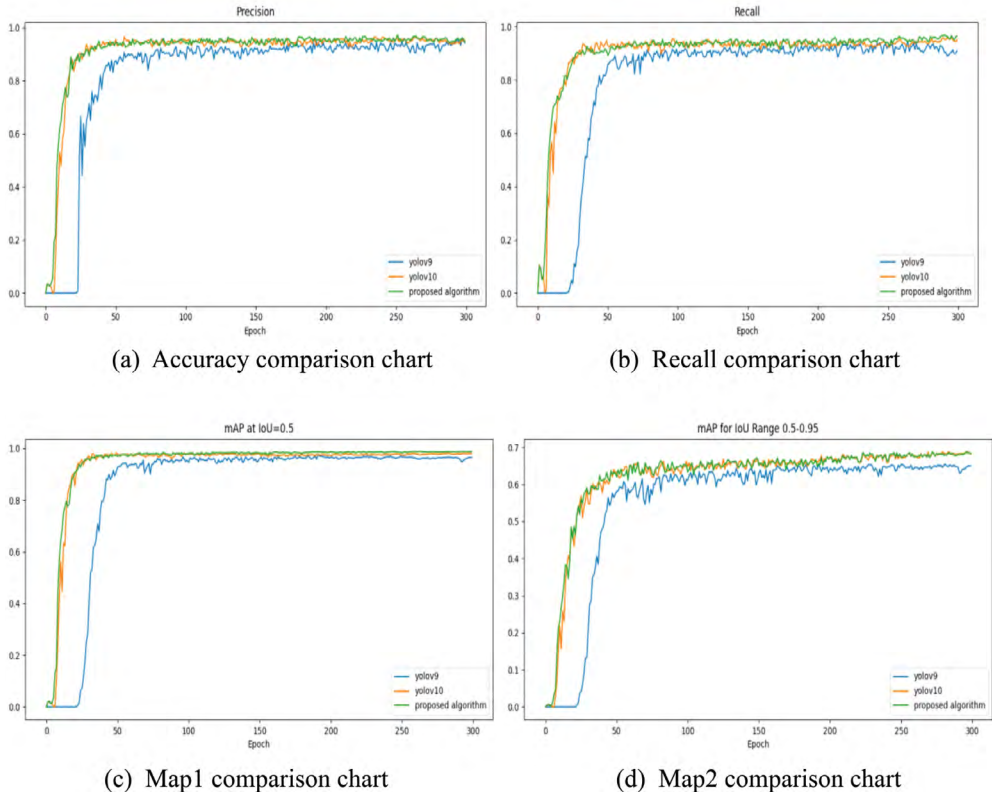
The mAP is the AP calculated for all categories and then averaged. Its can be used to weigh the goodness of the detection ability of the trained model on all categories. Assuming that there are m categories and $m > 1$, then the formula for mAP is:

$$mAP = \frac{\sum_{i=1}^{m} AP_i}{m} \tag{15}$$

- there is only one category, i.e. $m = 1$, the value of mAP should be equal to the value of AP.

The study carries out comparative experiments on a homemade stacked ore dataset to test the P, PR and mAP metrics of the three models, YOLOv9, YOLOv10 and SPDM-YOLO, which are trained on the ore dataset. The experimental results are presented in Table 3.

From the comparison experiment results in Table 3, it was found that the improved model increased accuracy by 3.09%, detection accuracy by 2.41%, and mean average accuracy by 2.18% compared to the original YOLOv9 model. The improved model improved the accuracy by 0.52%, detection accuracy by 0.63%, and average precision mean by 0.75% compared to the YOLOv10 model. The comparison curves of the models are shown in Figure 5.



(a) Accuracy comparison chart (b) Recall comparison chart

(c) Map1 comparison chart (d) Map2 comparison chart

Fig. 5. Model comparison curves

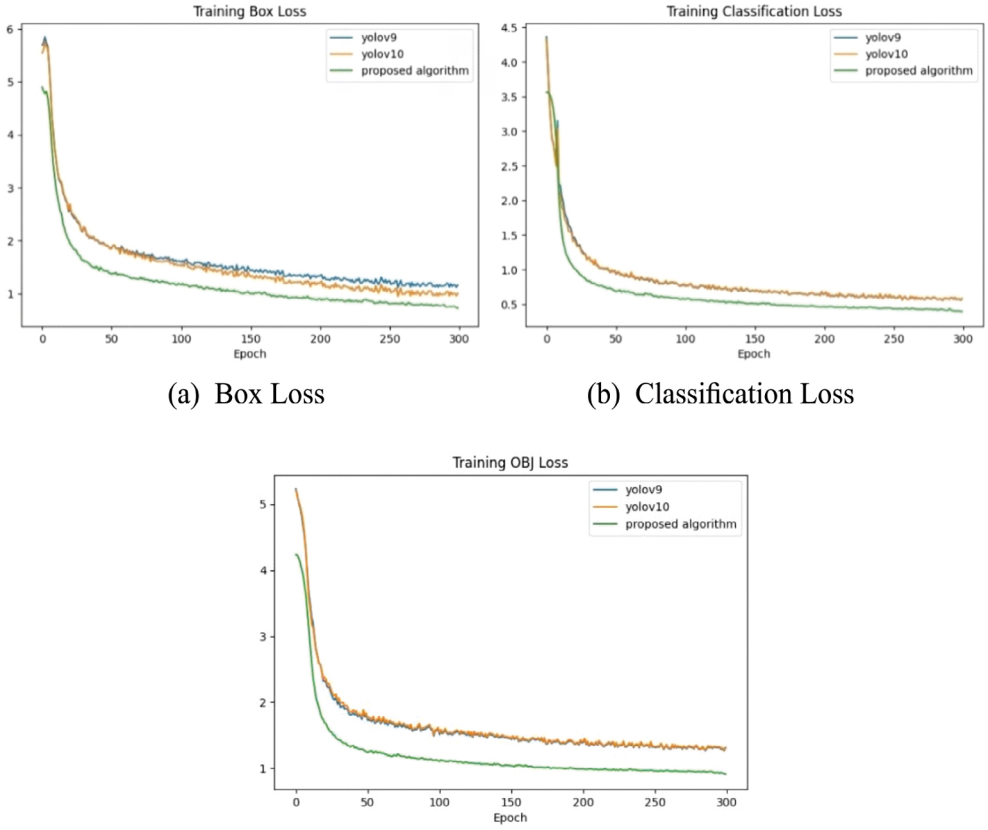
Rys. 5. Krzywe porównania modeli

Table 3. Results of model comparison

Tabela 3. Wyniki porównania modeli

Model type	P/%	PR/%	mAP/%
YOLOv9	91.37	96.56	96.64
YOLOv10	93.94	98.34	98.07
SPDM-YOLO	94.46	98.97	98.82

From Figure 5, it can be seen that the improved model achieves stability faster in terms of Precision, Recall, and mAP compared to the pre-improvement curves, resulting in improved performance. The newly released YOLOv10 model is also introduced for comparison.



(a) Box Loss (b) Classification Loss

(c) OBJ Loss

Fig. 6. Comparison of loss functions

Rys. 6. Porównanie funkcji strat

However, the improved model does not exhibit significant performance improvement compared to YOLOv10 in these three aspects, but the performance on the loss function has been improved to a certain extent. The variable values of the loss function of the improved model trained on the homemade ore dataset are shown in Figure 6.

Figure 6 loss function in the loss function by comparing the YOLOv9 YOLOv10 original model in the early iteration, the loss value briefly appeared to rise in the state, may be at the beginning of the iteration of the model's weight is random initialization, may be in a not very good state; with the increase in the number of iterations, the model adapts to the learning of the sample, the loss value shows a downward trend, iteration 270 times or so tends to converge. The overall loss value of the improved model is significantly lower than that of the other two models, and the oscillation amplitude of the loss value during the decrease is smaller compared to the other two. This further indicates that the training effect is ideal.

(2) Quantitative evaluation

The ablation experiments are conducted to verify the effect of the improved model by gradually introducing the SPD-Conv module, the MLCA attention mechanism, and the Inner-FocalerIoU loss function, based on the original YOLOv9 model. The results of the ablation experiments are shown in Table 4, where "+" indicates that this item is adopted and "–" indicates that this item is not adopted.

Table 4. Results of ablation experiments

Tabela 4. Wyniki eksperymentów ablacji

SPD-ConV	MLCA	Inner-FocalerIoU	P/%	PR/%	mAP/%
–	–	–	91.37	96.56	96.64
+	–	–	92.28	97.20	97.52
+	–	–	93.54	98.15	98.44
+	+	+	94.46	98.97	98.82

From the results of the ablation experiments in Table 4, it is found that the introduction of the SPD-Conv module improves the accuracy by 0.91%, the detection precision by 0.64%, and the average precision mean by 0.88%. The simultaneous introduction of the SPD-Conv module and MLCA attention mechanism resulted in a 2.17% increase in accuracy, a 1.59% increase in detection precision, and a 1.80% increase in the mean average precision. After the simultaneous introduction of the SPD-Conv module, the MLCA attention mechanism, and the Inner-FocalerIoU loss function, the accuracy is improved by 3.09%, the detection precision is improved by 2.41%, and the average precision mean is improved by 2.18%.

Conclusions. Advantages and contributions

1. The backbone part of the YOLOv9 network structure is enhanced by introducing the SPD-Conv building block, replacing the original convolution and pooling layers. This modification transforms the spatial dimension information of the input image into the depth dimension, thereby enhancing the depth of the feature map while preserving crucial details. It also enables feature extraction without reducing the resolution, thereby preserving fine-grained information of small and stacked ore particles.
2. Introducing the MLCA attention mechanism to help the model effectively differentiate between local and global priorities on the feature maps of different channels so as to obtain more detailed and comprehensive feature information and improve the model's micro-detail processing capability.
3. The introduction of the Inner-FocalerIoU loss function pays more attention to the calculation of the overlap degree of ore small targets and stacked ores, as well as accelerates the edge regression process to improve the accuracy of ore recognition further.
4. The mean values of stacked ore recognition accuracy, detection accuracy, and average accuracy of the improved SPDM-YOLO model are 94.46%, 98.97%, and 98.82%, which improve the model performance compared to the pre-improvement model, and the recognition effect is better.

Although the proposed SPDM-YOLOv9 demonstrates improved performance in detecting dense and overlapping ore particles, several limitations remain.

1. The current experimental setup is based on a simplified laboratory environment, which may not fully reflect the complexities of real mining production. The dataset used, although enhanced through Mosaic data augmentation, is relatively small. These limitations may affect the generalizability of the model in real-world scenarios.
2. While the improved model outperforms both YOLOv9 and YOLOv10, the performance gains are relatively modest when compared to the latest YOLOv10 architecture. Experimental results indicate that although the proposed modifications are effective, there is still room for further enhancement.

In future work, we plan to expand the dataset with real on-site ore images, explore domain adaptation techniques to improve generalization and investigate the integration of other lightweight modules or attention mechanisms to improve performance in complex environments further.

The Authors have no conflict of interest to declare.

REFERENCES

Chien et al. 2024 – Chien, C.T., Ju, R.Y., Chou, K.Y. and Chiang, J.S. 2024. YOLOv9 for fracture detection in pediatric wrist trauma X-ray images. *Electronics Letters* 60(11), DOI: 10.1049/ell2.13248.

Cheng et al. 2022 – Cheng, Q., Huang, H., Xu, Y., Zhou, Y., Li, H. and Wang, Z. 2022. NWPU-Captions Dataset and MLCA-Net for Remote Sensing Image Captioning. *IEEE Transactions on Geoscience & Remote Sensing* 60, pp. 1–19, DOI: 10.1109/TGRS.2022.3201474.

Deng, T. and Yu, Y. 2021. Research on ore identification and separation based on improved PSO-Faster R-CNN algorithm. *Mining Research and Development* 41(2), pp. 178–182 (*in Chinese*).

Li et al. 2023 – Li, Y., Cheng, R., Zhang, C., Chen, M.,Liang, H. and Wang, Z. 2023. Dynamic Mosaic algorithm for data augmentation. *Mathematical Biosciences and Engineering* 20(4), DOI: 10.3934/mbe.2023311.

Li et al. 2024 – Li, D., Jiang, C. and Liang, T. 2024. REDef-DETR: real-time and efficient DETR for industrial surface defect detection. *Measurement Science and Technology* 35(10), DOI: 10.1088/1361-6501/ad60ea.

Longzhen et al. 2022 – Longzhen, Y., Zhu, J., Zhao, Q. and Wang, Z. 2022. An Efficient YOLO Algorithm with an Attention Mechanism for Vision-Based Defect Inspection Deployed on FPGA. *Micromachines* 13(7), DOI: 10.3390/mi13071058.

Luo et al. 2022 – Luo, X., Liu, S., Tang, W. and Wang, X. 2022. Research on identification and location of blocked ore at ore bin inlet based on Mask RCNN. *Nonferrous Metals Science and Engineering* 13(1), pp. 101–107, DOI: 10.13264/j.cnki.ysjskx.2022.01.013.

Pan et al. 2024 – Pan, W., Chen, J., Lv, B. and Peng, L. 2024. Optimization and Application of Improved YOLOv9s-UI for Underwater Object Detection. *Applied Sciences* 14(16), DOI: 10.3390/app14167162.

Ruan et al. 2022 – Ruan, S., Liu, D., Gu, Q. and Jing, Y. 2022. An Intelligent Detection Method for Open-Pit Slope Fracture Based on the Improved Mask R-CNN. *Journal of Mining Science* 58(3), DOI: 10.1134/S1062739122030176.

Shao et al. 2024 – Shao, A., Liu, C., Yue, X. et al. 2024. High-quality development path of China's mineral resources industry in the new period. *Metal Mining* 1, pp. 2–6 (*in Chinese*).

Usha et al. 2022 – Usha, M., Priyanka, C. and Rajeev, T. 2022. EnsembleNet: a hybrid approach for vehicle detection and estimation of traffic density based on faster R-CNN and YOLO models. *Neural Computing and Applications* 35(6), DOI: 10.1007/s00521-022-07940-9.

Yang et al. 2023 – Yang, Z., Wu, Q., Zhang, F., Zhang, X., Chen, X. and Gao, Y. 2023. A New Semantic Segmentation Method for Remote Sensing Images Integrating Coordinate Attention and SPD-Conv. *Symmetry* 15(5), DOI: 10.3390/sym15051037.

Yuhuan et al. 2023 – Li, Y., Yang, F., Li, Y., Tan, C. and Liu, Z. 2023. Circuit Breaker Identification Based on SSD. *Journal of Physics: Conference Series* 2418(1), DOI: 10.1088/1742-6596/2418/1/012080.

Zhang et al. 2022 – Zhang, Y.F., Ren, W., Zhang, Z., Jia, Z., Wang, L. and Tan, T. 2022. Focal and efficient IOU loss for accurate bounding box regression.*Neurocomputing* 506, pp. 146–157, DOI: 10.1016/j.neucom.2022.07.042.

Zhang et al. 2023 – Zhang, H., Xu, C. and Zhang, S. 2023. Inner-IoU: More Effective Intersection over Union Loss with Auxiliary Bounding Box.

Zhu et al. 2024 – Zhu, G., Wang, Z., Zhu, F. et al. 2024. Small target detection algorithm based on HCAC and MWFN. *Electro-Optics and Control* 31(9), pp. 31–37.

STACKED FINE ORE DETECTION METHOD BASED ON SPDM-YOLO NETWORK

Keywords

MLCA, YOLOv9, ore detection, loss function

Abstract

To address the issue of low accuracy in detecting stacked fine ore particles using existing target detection algorithms, which subsequently affect the crusher's ore crushing and screening process, a new ore detection method based on the SPDM-YOLO deep learning network is proposed. The study focuses on detecting stacked fine tungsten ore particles ranging in size from 6 mm to 50 mm. The initial step involves incorporating SPD-Conv into YOLOv9, whereby it replaces the conventional CNN's step-length convolutional layer and pooling layer. This is done to reduce the loss of ore target detail information. Secondly, a lightweight hybrid local channel attention module (MLCA) is integrated into the backbone to improve focus on target features and enhance the model's ability to process intricate details. Finally, the Inner-FocalerIoU loss function is proposed by combining the advantages of Inner-IoU and Focaler-IoU. Replacing the original CIoU loss function with it as the edge loss function of the algorithm not only improves the accuracy of calculating the overlap of the bounding box but also effectively speeds up the edge regression process. The Mosaic data enhancement method was employed to augment the dataset for experimental purposes. The findings indicate that the enhanced YOLOv9 network model improves the accuracy (P) by 3.09%, the detection precision (PR) by 2.41%, and the mean average precision (mAP) by 2.18% compared to the original model. This substantiates the assertion that the augmented algorithm is more efficacious in recognition.

METODA WYKRYWANIA UŁOŻONYCH DROBNYCH RUD OPARTA NA SIECI SPDM-YOLO

Słowa kluczowe

MLCA, YOLOv9, wykrywanie rudy, funkcja strat

Streszczenie

Aby rozwiązać problem niskiej dokładności wykrywania ułożonych drobnych cząstek rudy za pomocą istniejących algorytmów wykrywania celu, co następnie wpływa na proces kruszenia i przesiewania rudy kruszarki, zaproponowano nową metodę wykrywania rudy opartą na głębokiej sieci uczenia SPDM-YOLO. Badanie koncentruje się na wykrywaniu ułożonych drobnych cząstek rudy wolframu o rozmiarze od 6 mm do 50 mm. Początkowy krok obejmuje włączenie SPD-Conv do YOLOv9, dzięki czemu zastępuje on konwencjonalną warstwę splotową CNN o długości kroku i warstwę pulowania. Ma to na celu zmniejszenie utraty szczegółowych informacji o docelowej

rudzie. Ponadto lekki hybrydowy moduł uwagi lokalnego kanału (MLCA) jest zintegrowany z kręgosłupem, aby poprawić koncentrację na cechach docelowych i zwiększyć zdolność modelu do przetwarzania skomplikowanych szczegółów. Na koniec zaproponowano funkcję strat Inner--FocalerIoU, łącząc zalety Inner-IoU i Focaler-IoU. Zastąpienie oryginalnej funkcji strat CIoU nią jako funkcją strat krawędzi algorytmu nie tylko poprawia dokładność obliczeń nakładania się pola ograniczającego, lecz także skutecznie przyspiesza proces regresji krawędzi. Do rozszerzenia zestawu danych do celów eksperymentalnych zastosowano metodę wzbogacania danych Mosaic. Wyniki wskazują, że ulepszony model sieci YOLOv9 zwiększa dokładność (P) o 3,09%, precyzję wykrywania (PR) o 2,41% i średnią precyzję (mAP) o 2,18% w porównaniu z oryginalnym modelem. Potwierdza to twierdzenie, że rozszerzony algorytm jest skuteczniejszy w rozpoznawaniu.