

## LFOS-YOLO: A lighter and faster YOLO model for intelligent ore sorting in edge device environment

Decan Zeng <sup>1,2</sup>, Chunrong Pan <sup>1</sup>, Kai Feng <sup>1</sup>

<sup>1</sup> School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China

<sup>2</sup> Department of Electromechanical Manufacturing, Heyuan Institute of Technology, Heyuan 517000, China

Corresponding author: [crpan@jxust.edu.cn](mailto:crpan@jxust.edu.cn) (Chunrong Pan)

**Abstract:** Minerals are non-renewable resources that are indispensable for contemporary industrial production. The advent of intelligent ore sorting technologies has been pivotal in enhancing mineral utilization, a phenomenon that has been further propelled by the advent of YOLO (You Only Look Once) object detection models for end-to-end detection. However, intelligent ore sorting equipment deployed at mining sites has more exacting requirements regarding the parameters, computational complexity, and inference speed of deep learning networks. Accordingly, this paper proposes a lighter and faster ore sorting YOLO model, namely LFOS-YOLO, derived from the configuration of the YOLO model. The model incorporates partial convolution, parameter-free attention mechanisms, and redundant channel pruning strategies with the objective of achieving an accuracy-lightweight tradeoff that meets the requirements for deployment in a mining site. The experimental results demonstrate that LFOS-YOLO is capable of accomplishing the task of sorting ore samples from a mine in Liaoning, China, with fewer parameters (1.46M), lower GFLOPs (Giga Floating-Point Operations per Second) (3.4), and higher FPS (Frames per Second) (74), achieving the highest mAP (mean Average Precision) (95.4%), which outperforms other models in the YOLO series.

**Keywords:** ore sorting, YOLO, attention mechanism, lightweight network, channel pruning

### 1. Introduction

Ore resources occupy a central position in modern industrial production as a consequence of their status as a precious, non-renewable resource. It is therefore imperative that the intelligence level of ore sorting equipment be upgraded in order to ensure the efficient utilization of resources. The implementation of this intelligent upgrade will result in a notable reduction in the generation of tailings, while simultaneously leading to a substantial decrease in energy consumption during the processes of ore crushing and grinding. This will ultimately yield a dual benefit, namely the reduction of emissions and the enhancement of efficiency.

Currently, two principal categories of sophisticated ore sorting machinery are in existence. One is based on ray sensors, the other on machine vision. Robben et al. investigated the application of X-ray lithography (XRL), X-ray transmission (XRT), and X-ray fluorescence (XRF) techniques in ore sorting, which pointed out the direction for academic research and engineering applications of ore sorting (Robben and Wotruba, 2019). Several researchers have studied the sorting of different ores. For instance, Bellusci et al. (2022) successfully sorted trona ore using the XRT technique. Robben et al. (2020) also achieved satisfactory results in sorting tin ore using radiometric sensors. Abd Halim et al. (2023) conducted a study on the sorting of iron ore using the XRF technique, which provided valuable insights into the mining of iron ores. The implementation of ore sorting equipment based on ray sensors has demonstrated a high degree of accuracy in the sorting process. Nevertheless, the considerable expense of the apparatus and the inherent danger of radiation render it unfeasible for extensive implementation.

In recent years, machine learning and deep learning methods have brought new developments to numerous industries (Keshun et al., 2024), as well as new ideas for intelligent ore sorting (Yang et al., 2023). The use of machine vision-based equipment involves the utilisation of industrial cameras for the

capture of images of the ore, which are then subjected to machine vision algorithms for the purpose of sorting the ore. The initial generation of intelligent ore sorting equipment predominantly employs traditional machine learning algorithms, such as the SVM algorithm, which is a type of linear classifier in supervised learning. This approach has been demonstrated to perform exceptionally well in the task of ore image classification, thereby substantially enhancing the efficiency of ore utilization (Chen et al., 2018; Dong et al., 2021). Moreover, the Decision Trees, K-Nearest Neighbours (KNN), and Artificial Neural Network (ANN) methods also demonstrate exceptional efficacy in certain specific ore sorting tasks (Izadi et al., 2017; Sun et al., 2022; Khorram et al. 2017; Chen et al., 2024).

In comparison to the conventional machine learning algorithms, the deep learning networks demonstrate superior capacity for extracting image features. The advancement of mineral sorting technology based on deep learning has generally proceeded through three stages. The initial stage involved the construction of a self-built deep convolutional model. Xiao et al. constructed an 18-layer CNN model to address the traditional iron ore sorting problem (Xiao et al., 2021). The 25-layer CNNs model proposed by Lei et al. was successfully deployed in a robot to complete the coal sorting task (Lei et al., 2020). Okada et al. constructed their own deep convolutional model to achieve the classification task of arsenic mineral images (Okada et al., 2021). The second stage is that of representative deep convolutional models. The most common representative deep convolutional models are AlexNet (Krizhevsky et al., 2012), VGG (Simonyan and Zisserman, 2014), and ResNet (He et al., 2016), which have achieved excellent results in image classification tasks. Scholars have also employed these models in mineral sorting. For instance, Liu et al. developed a coal image classification model based on AlexNet and VGG models, which served as a reference for intelligent sorting tasks in coal mines (Liu et al., 2021), Liu et al. (2019) employed deep learning to categorize 12 types of fine rock images, thereby offering a novel approach to rock recognition. Nizinski et al. (2022) validated the efficacy of a ResNet-based deep learning model in uranium ore image recognition. The third stage is that of end-to-end object detection models, including the two-stage model Faster R-CNN (Ren et al., 2016) and the one-stage YOLO (You Only Look Once) series of models. These models demonstrate enhanced capabilities for image localization and classification. Accordingly, Jiang et al. (2022) employed the two-stage algorithm of Faster R-CNN for mineral classification and obtained favorable outcomes. The YOLO object detection model directly and uniformly addresses the image localization and classification issues. For instance, He et al. (2024) employed the YOLOv8 algorithm to address the challenge of classifying seven prevalent minerals, achieving favorable outcomes. Liu et al. (2024) proposed a combination of a convolutional neural network with a transformer architecture and the addition of a self-attention mechanism, thereby obtaining a new sorting model, which showed excellent performance in gas coal, coking coal and anthracite coal. Zeng et al. (2025) developed a deep learning model to propose a new detection and classification method for magnetite sorting, achieving excellent results.

Despite the potential benefits, there are two key challenges to the application of deep learning for ore sorting. One challenge is that deep learning models are often characterized by a large number of parameters and high computational complexity, which requires significant storage and computational resources and is difficult to deploy at the mining site. Additionally, the relatively slow inference speeds of deep learning models present a challenge in meeting the requirements of industrial mining. In order to address these two issues, this paper proposes a new ore sorting model, LFOS-YOLO (A Lighter and Faster YOLO Model for Intelligent Ore Sorting), which draws on the end-to-end real-time detection of the YOLO model. Furthermore, the LFOS-YOLO model has a smaller model size, lower computational complexity, and faster inferencing, which makes it more suitable for the mining industry. Specifically, to address the issue of deep neural network convolutional stacking causing a large number of parameters and computational complexity, partial convolution is introduced to lighten the model architecture. A parameter-free attention module is introduced to improve the feature extraction capability of the model without increasing the model size. A new detection head is designed to improve the inference speed of the model. Finally, a trade-off is made between model accuracy and efficiency, and the overall network is pruned with redundant channels to further lighten the model. Furthermore, the performance of LFOS-YOLO is compared and analyzed with eight conventional YOLO object detection networks (YOLOv3-YOLOv10) in the context of ore sorting in a mine in Liaoning, China. The experiments were analyzed in terms of accuracy, model size, computational complexity, and inference

speed. LFOS-YOLO demonstrated the highest accuracy (95.5%) and exhibited optimal performance in terms of computational complexity, model size, and inference speed.

## 2. Related works

### 2.1 YOLO object detection model

The You Only Look Once (YOLO) algorithm was first proposed in 2016 (Redmon et al., 2016). The model has gained considerable popularity as a tool for object detection and image segmentation in the field of computer vision. It has attracted considerable interest from both industry and academia. YOLOv2 facilitated enhanced speed and precision (Redmon and Farhadi, 2016). YOLOv3 further enhanced the algorithm with more efficient networks, multi-scale anchors, and spatial pyramid pooling (Redmon and Farhadi, 2016). YOLOv4, which was launched in 2020, represents a significant advancement in performance, incorporating innovations such as Mosaic data enhancement, anchorless detection heads, and novel loss functions (Bochkovskiy et al., 2020). YOLOv5 facilitates the use of the model for users with features such as hyperparameter optimization, experiment integration tracking, and automated model format conversion (Zhu et al., 2021). The release of YOLOv6 in 2022 has seen its implementation in autonomous delivery robots (Li et al., 2022). YOLOv7 is capable of supporting pose estimation for the COCO keypoint dataset (Wang, et al., 2022). YOLOv8 is a highly versatile system that is capable of supporting a wide range of vision AI tasks. YOLOv9 introduces advanced technologies (Wang et al., 2024). Aggregation Networks (GELAN) enhance performance and flexibility. The innovative network architecture and NMS-free post-processing of YOLOv10 facilitate enhanced performance and efficiency in real-time target detection. The model's widespread adoption in real-world application scenarios is driven by its efficiency-accuracy-driven design strategy and simplified deployment process (Wang et al., 2024). In September 2024, Ultralytics released the latest iteration of the YOLO11, which features enhanced feature extraction capabilities, higher accuracy, and faster inference. Consequently, the YOLO model was rapidly adopted by researchers in a multitude of industries. For instance, Liu et al. utilized the YOLO model for the detection of weld defects in the manufacturing industry (Liu et al., 2023). Similarly, Wu et al. (2023) employed the YOLO model for the identification of cracks in bridges and dams within the construction industry. Additionally, Xie et al. (2023) applied the YOLO model for the detection of tea trees in the agricultural sector. The objective of this paper is to examine the potential applications of the YOLO model in the context of intelligent ore sorting.

### 2.2 Attention mechanism

In the ore sorting task, the model has high requirements for image feature extraction due to the close similarity in shape features among ores, which presents a significant challenge in distinguishing between them. Furthermore, the majority of raw ores contain non-mineral materials, such as plastics and metals, which further complicates the process. Consequently, a considerable number of studies have been conducted with the objective of optimizing the feature extraction process for the model. Among the proposed solutions, the introduction of an attention mechanism has emerged as a promising approach. For example, Liu et al. incorporated various attentional modules (CA block, SA block, MA block, etc.) into the model, which yielded favorable outcomes in a multi-type coal classification task (Liu et al., 2021). Ma et al. (2021) enhanced the model's capacity for feature extraction by incorporating SE blocks into an ore image classification model. Zheng et al. (2023) have investigated the application of the simAM attention block in the context of mining tasks. In view of the above, the present study will also incorporate an attention mechanism into the intelligent ore sorting model.

### 2.3 Partial convolution

Ensuring the real-time performance of deep models is challenging due to the considerable number of conventional convolutional stacks, which result in a substantial number of model parameters, high computational complexity, and prolonged inference times. The substitution of these with depth-wise separable convolutions (Chollet, 2017) has been shown to be an effective method of reducing the overall number of parameters (Korban and Li, 2023; Tu et al., 2023; Xiang et al., 2024). Nevertheless, this methodology will result in a considerable number of memory accesses, which will consequently result

in a reduction in detection speed. For this reason Chen et al. (2023) proposed partial convolution. The fundamental principle is illustrated in Fig. 1. Partial convolution (PConv) is a technique that extracts specific features from a designated portion of the input channel feature map, while leaving the remaining portion unchanged. This approach is both rapid and efficient, and it can reduce the number of parameters and memory accesses in comparison to depth-wise separable convolution. The following equations (1) and (2) illustrate the FLOPs and memory accesses, respectively. The number of floating-point operations (FLOPs) for a PConv is as follows:

$$S_F = h * w * k^2 * c_p^2 \quad (1)$$

When  $r$  is equal to  $\frac{c_p}{c}$ , which is equivalent to  $1/4$ , the FLOPs associated with PConv is only  $1/16$  of what would be expected from a regular convolution operation. Moreover, the PConv requires less memory access, i.e.,

$$S_{acc} = h * w * 2c_p + k^2 * c_p^2 \approx h * w * 2c_p \quad (2)$$

This is merely  $1/4$  of the standard convolution for  $r=1/4$ .

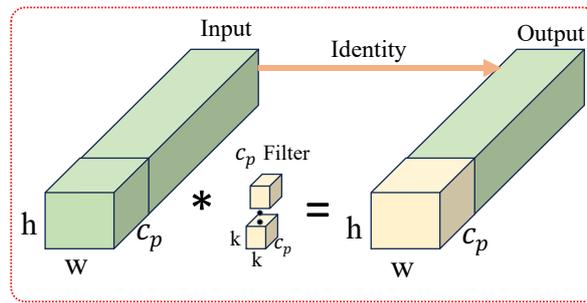


Fig. 1. Partial Convolution ('PConv')

## 2.4. Intelligent ore sorting

Mineral processing encompasses a series of key steps, including but not limited to crushing, sorting, grinding, and flotation. Effective sorting technology can markedly enhance the quality of the ore grade, diminish subsequent processing costs and the production of tailings. Presently, the prevailing ore sorting technologies are classified into two principal categories: ray sensor-based technologies and machine vision-based technologies (Luo et al., 2022). While ray sensors have been demonstrated to be effective for the sorting of certain ores (Robben et al., 2020; Zhang et al., 2021), the use of such equipment is often associated with significant radiation and energy consumption issues. With the advancement of deep learning technology, there has been a growing emergence of computer vision-based ore sorting techniques. In this field, researchers have gradually progressed from the initial exploration of self-built models to the adoption of representative models, and finally to the end-to-end object detection model represented by YOLO, which has demonstrated satisfactory results in terms of mineral sorting accuracy. For example, Pratama et al. (2023) successfully demonstrated the use of YOLOV4 for the sorting of rock minerals, and Wan et al. (2023) employed YOLOV5 for ore sorting, achieving high accuracy rates. However, as with any new technology, there are also challenges associated with deep learning-based computer vision detection models. Deep models are often distinguished by a considerable number of parameters, considerable computational complexity, and a relatively slow inference speed, which collectively present challenges to their deployment in edge devices at the mining site. In response to these challenges, this paper proposes LFOS-YOLO, which replaces a substantial number of stacked conventional convolutions with partial convolutions, employs a parameter-free attention mechanism to enhance the model's feature extraction capabilities, and utilizes pruning techniques to further lighten the model.

## 2. Methodology

Fig. 2 depicts the research methodology employed in this paper, which is divided into four phases. Initially, a batch of ores from a mining site in Liaoning, China is gathered for analysis. Subsequently, each category of the sample is labelled and divided into a training dataset, a test dataset, and a

validation dataset. The third stage of the process is the input of the dataset into the LFOS-YOLO deep model, which is then trained and tested. Ultimately, the model is further simplified through the process of pruning.

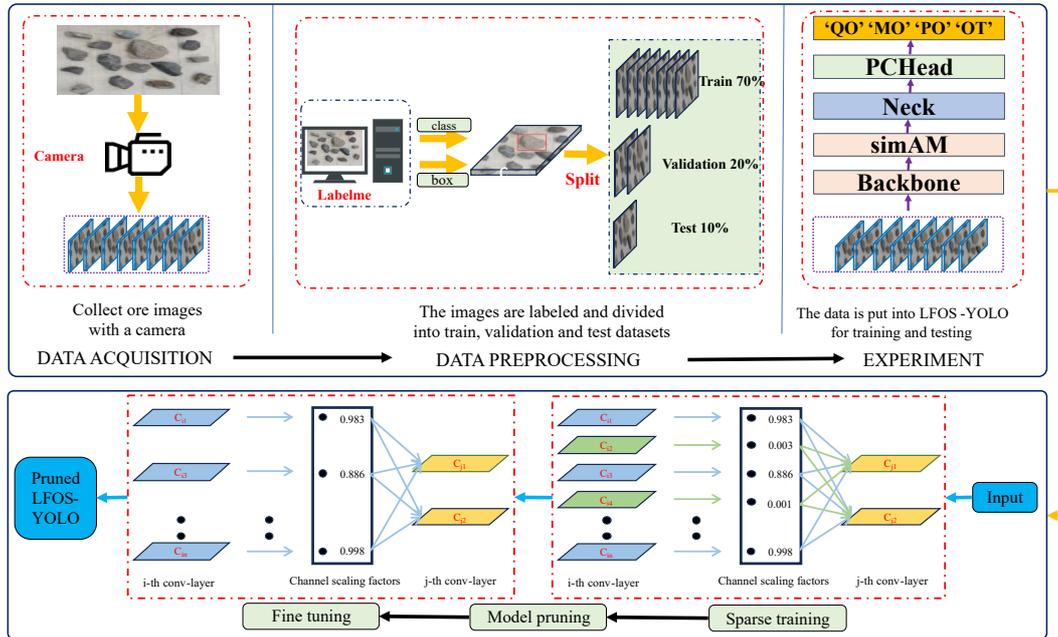


Fig. 2. The schematic diagram of the study, including creating datasets, algorithm experiments, pruning experiments

### 3.1. Intelligent ore sorting platform

As illustrated in Fig. 3, the intelligent ore sorting platform is comprised of five principal components: a feeding device, an ore conveyor belt, a camera, a computer, and a pneumatic injection device. The raw ore is then conveyed from the feeding device to the ore conveyor belt, where it is imaged by the industrial camera. Subsequently, the captured images are transmitted in real time to the computer control system, where LFOS-YOLO performs the classification of the mineral images. Subsequently, the classification data is transmitted to the pneumatic injection unit, whereupon the minerals are injected into the designated area.

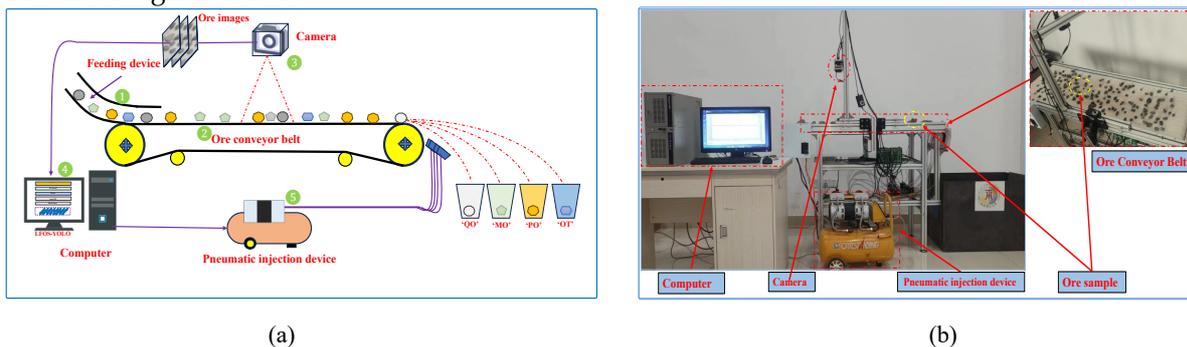


Fig. 3. Intelligent ore sorting platform, including feeding device, ore conveyor belt, camera, computer, pneumatic injection device. (a) A schematic diagram of the platform. (b) is a physical image of the laboratory.

### 3.2. A lighter and faster ore sorting model

Fig. 4 illustrates the design specifications of LFOS-YOLO, with the entire network divided into four principal modules: the backbone module, the attention module, the neck module, and the head module. The backbone module is tasked with the extraction of features from the input image. The parameter-free attention mechanism module can enhance the model's feature extraction capabilities without

increasing its parameters. The neck module is responsible for fusing the extracted feature maps at varying scales, while the head module is utilized for ore localization and classification.

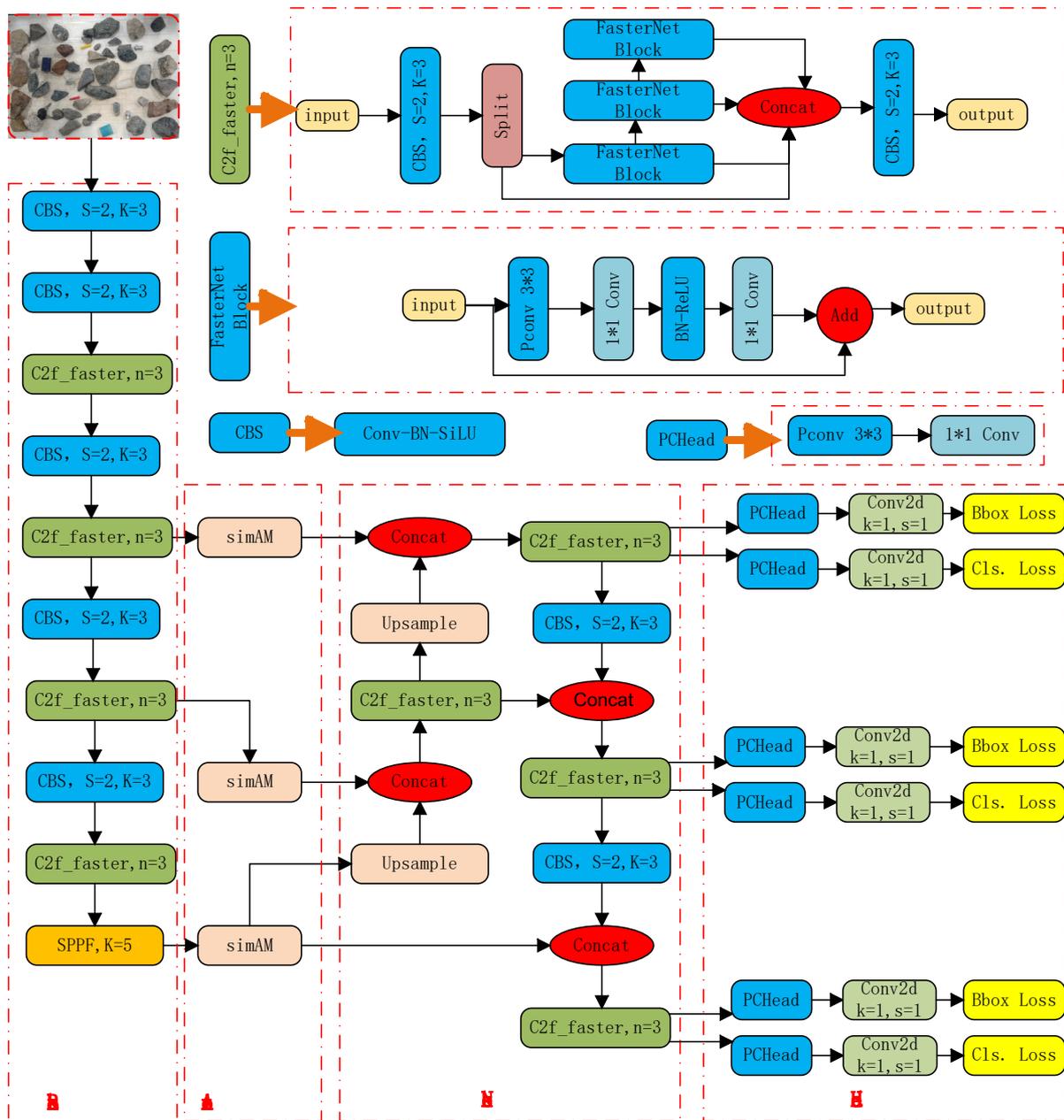


Fig. 4. The architecture of LFOS-YOLO, including backbone module, attention module, neck module, and head module

### 3.2.1. Lighter backbone and neck modules

The backbone module of the model is responsible for the extraction of features from the input image, while the neck module employs a feature pyramid structure to facilitate the extraction of multi-scale feature maps for the detection of objects of varying sizes. The backbone and neck modules in the YOLO model comprise multiple C2f modules, as illustrated in Fig. 5. Each bottleneck layer incorporates two regular convolutional layers, resulting in a considerable increase in model parameters and computational complexity. Accordingly, this paper proposes an optimization of the C2f module through a modification of the bottleneck to a FasterNet Block (FBlock), as illustrated in Fig. 6. The original stack of two CBS modules is modified to a stack of one PConv module and two pointwise convolutions (PWConv) (Chen et al., 2018), which significantly reduces the number of parameters and computational complexity of the model.

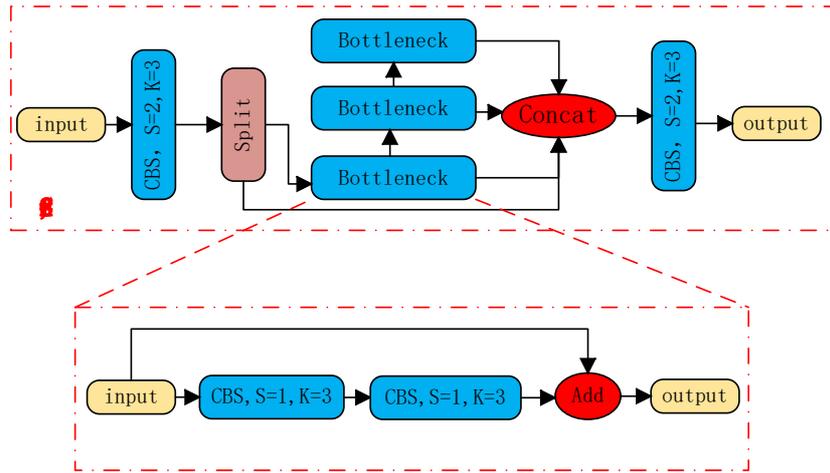


Fig. 5 Structure of C2f, CBS = Conv + BN + SiLU

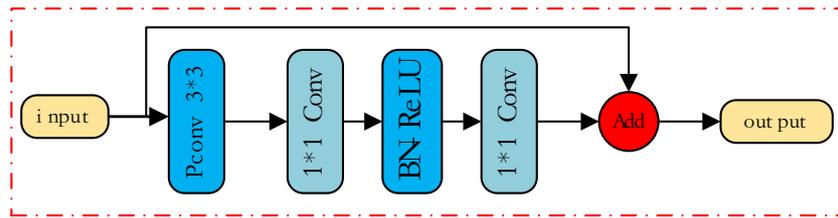


Fig. 6. Structure of FasterNet Block (FBlock)

### 3.2.2. Parameter-free attention mechanism

In order to enhance the network's capability to extract features from the image, it is essential to integrate attention mechanisms into the model. To circumvent the issue of parameter inflation, this paper introduces the simAM module, a parameter-free attention mechanism (Yang et al., 2021). In contrast to the conventional approach of performing convolution operations internally, the simAM module derives 3D attention weights of feature maps by channel weights computed from the energy function, as illustrated in Fig. 7. In this figure, the input feature map size is represented by  $H * W$ , and the number of channels is  $C$ . To generate the 3D weights using the energy function, the weights are first normalized by the sigmoid function and then fused with the original feature map to obtain the final output.

The simAM module is based on the neuroscientific theory that neurons activate other neurons through the emission of firing patterns that differ from those of their surrounding neurons. The activated neurons typically exert an inhibitory effect on the surrounding neurons, a phenomenon known as null spatial inhibition. In other words, neurons exhibiting null suppression should be afforded greater significance. Defining the energy function  $e_t^*$  on each channel can be expressed as follows:

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{u})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (3)$$

where  $\hat{u}$  is the mean value,  $\hat{\sigma}$  is the variance value, and  $\lambda$  is the hyperparameter, the lower the energy, the more different the neuron  $t$  is from the surrounding neurons and the higher the importance. Therefore, the importance of a neuron can be expressed by  $\frac{1}{e_t^*}$ . The final output feature map can be expressed as follows:

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{E}\right)\Theta X \quad (4)$$

### 3.2.3 A faster PCHead module

As illustrated in Fig. 8, the conventional YOLOv8 network comprises three groups of detector heads, each designed to detect objects of varying sizes, including small, medium, and large. Each group employs a two-branch detector head that is decoupled from classification and localization. Two regular  $3*3$  convolutional layers are stacked on each detector head branch. Consequently, the total number of conventional  $3*3$  convolutional layers for the three groups is 12, which results in a considerable number

of model parameters and computational complexity. This paper presents a novel detection head design. In order to reduce the model parameters and computational complexity, an alternative approach was taken whereby the classification and localization were not decoupled immediately upon receipt of the features from the neck. Instead, a new module, designated as a PConv and PWconv, was introduced. This module consisted of a PCHead, which replaced the original two 3\*3 convolutional layers. The rationale behind this design choice is illustrated in Fig. 9. Subsequently, the classification and localization are decoupled, which can significantly reduce the number of model parameters and computational costs, as well as enhance the detection accuracy through training.

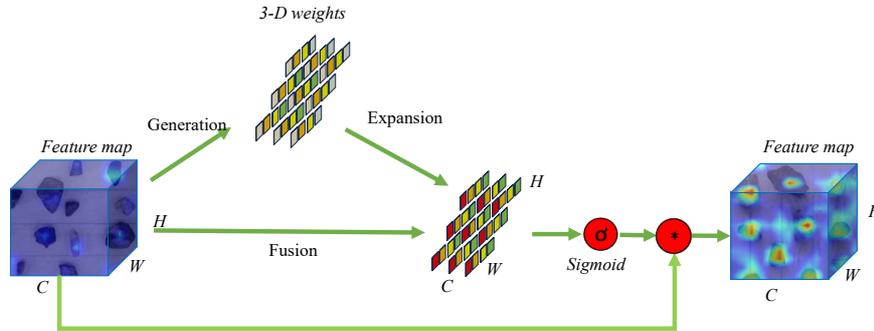


Fig. 7. The architecture of SimAM

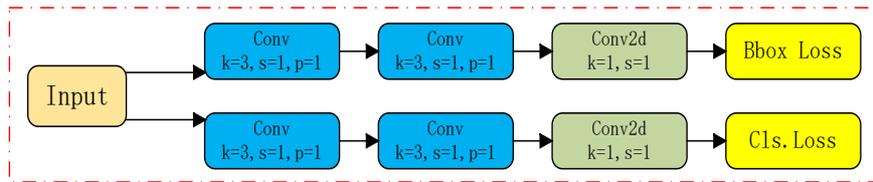


Fig. 8. One set of detection heads for the YOLOV8 model

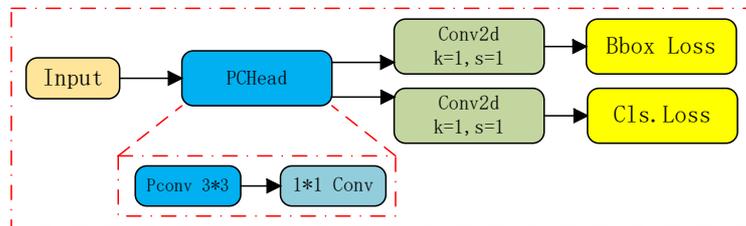


Fig. 9. A new set of improved detection heads

### 3.2.4. Channel pruning

Fig. 4 illustrates the modules utilized in the network, with the exception of the already optimized C2f\_faster module and the PCHead module. The remaining modules, designated as CBS, each contain a 3\*3 convolution network. The objective is twofold: to reduce the model size and GFLOPs of the network while pruning redundant channels. The specific steps of pruning are as follows (Liu et al., 2017), as illustrated in Fig. 10. In order to achieve structural sparsity in the LFOS-YOLO ore sorting model, we have applied a regularity constraint to the coefficients of the BN layer. Once training has been completed, the channels can be pruned to a specific rate in order to produce a more streamlined model using less memory. The objective is to address the issue of significant loss of accuracy that can occur as a result of channel pruning, and to recover the accuracy that has been lost.

## 4. Experiment

This section outlines the experimental process used to train the model. Fig. 11 illustrates the three main steps of the experiment. The first step is the acquisition of the dataset, capturing images of the ore on the conveyor belt inside the ore sorter as it passes through the industrial camera. The second step is data preprocessing, where the ore images obtained in the first step are labeled, the dataset is expanded, and

finally divided into training, validation, and testing sets. The third step is model training experiments, where the dataset processed in the previous step is fed into the model for training, and algorithm ablation experiments and model comparison experiments are performed.

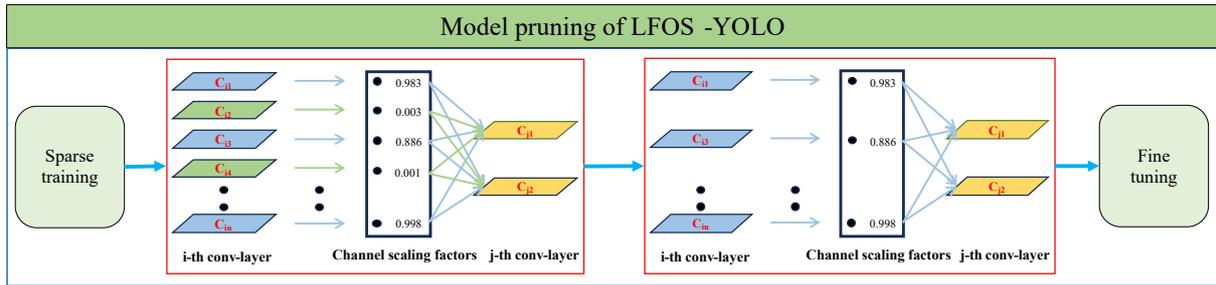


Fig. 10. Redundant channel pruning algorithm

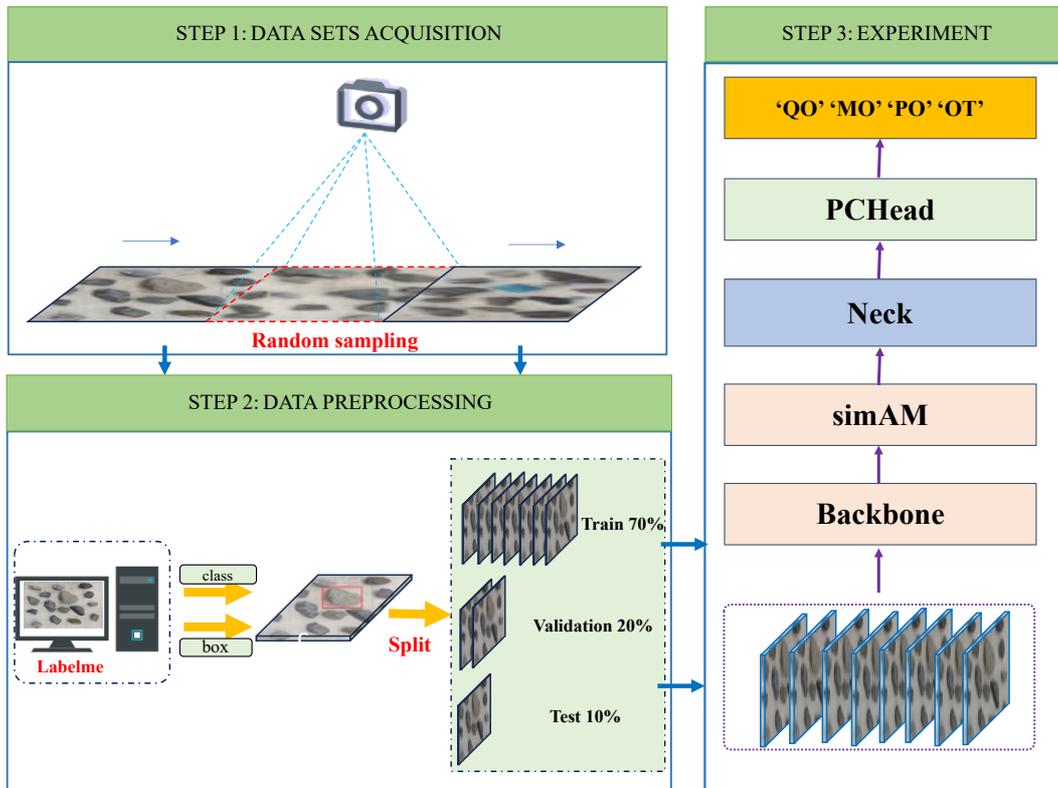


Fig. 11 The main experimental procedures, step 1 is the datasets acquisition, step 2 is the data preprocessing, step 3 is the experiment

## 4.1. Experimental preparation

### 4.1.1. Datasets acquisition

The samples employed in the experiment were derived from raw ores extracted from an industrial mine located in Liaoning Province, China. The mine, which is primarily engaged in the mining of magnetite ore, belongs to the sedimentary type of deposit. This was formed around the Pliocene period, predominantly through sedimentation, with subsequent hydrothermal modification. In addition to a substantial quantity of magnetite ore, the mine is also associated with quartz, pyrite, and other minerals. The ore was crushed into particles between 10 and 50 mm in size and was then classified into four categories based on compositional testing: quartz ores ("QO"), magnetite ores ("MO"), pyrite ores ("PO"), and other ("OT"). Furthermore, any ore that does not fall into one of the first three categories will be designated as 'OT'. To reflect the complexity of real mine conditions and to improve the generalisation performance of the model, two other non-mineral samples (metal materials, e.g., keys and plastic

materials, e.g., pen caps) were added to the experiments and labelled as "OT". In the experiment, the ore samples were conveyed on the conveyor belt inside the sorter at a speed of 1m/s, which was not easily affected by the external operating environment. The conveyor belt was lighted to ensure that the camera was bright enough to take pictures. The camera extracted the ore images at intervals of 500ms per frame, and a total of 7200 ore images were extracted as a dataset in 60 minutes, and the pictures of the ores were fed into the visual inspection model at a size of 640\*640. The resultant image data is presented in Fig. 12.

Upon examination of the samples, the following difficulties were identified in this ore sorting task. (1) Some of the ore samples exhibit no discernible differences in terms of morphology, texture, or colour, which can lead to misdetection. (2) The criteria for categorizing ores as "OT" lack a clear definition, which can result in interference and misdetection. (3) The number of ore samples visible to the camera is not limited, and an excess of samples can result in missed detection. To address these challenges, the model will be augmented with a parameter-free attention module, enhancing its feature extraction capability and improving sorting accuracy.

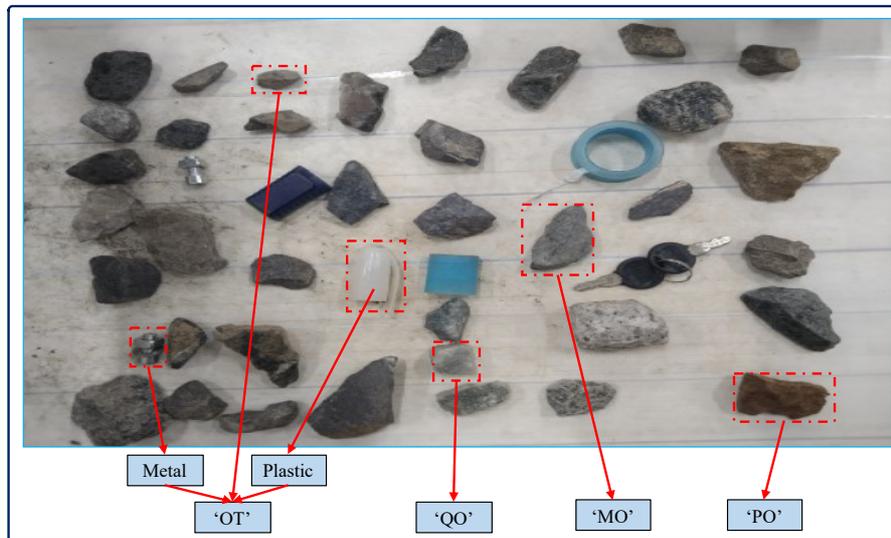


Fig. 12. The acquired image data

#### 4.1.2. Dataset preprocessing

In the preceding step, 7,200 ore images were obtained and labeled with Labelme for each type of ore instance, with the objective of determining the location and category of each instance. Furthermore, the total number of images was augmented to 28,800 through the application of data augmentation techniques, namely Mosaic and Mixup. Subsequently, the data were divided into three distinct datasets, comprising a 7:2:1 ratio of training, validation, and test sets, respectively. The training set is employed for model training, the validation set is utilized to refine the model hyperparameters, and the test set is utilized to assess the model performance. Fig. 13 illustrates the instance and labeling information of the dataset. The Fig.13(a) depicts the number of instances in each category, demonstrating that the number of samples in each category is largely balanced. The Fig.13(b) depicts the positional distribution of the instances, with x and y representing the normalized coordinates. The darker the position, the greater the number of instances, and it can be observed that the majority of instances are concentrated in the central region of the conveyor belt. The Fig.13(c) illustrates the size of the instances, with width and height denoting the normalized dimensions. The darker regions indicate a greater number of instances, indicating that the majority of instances are relatively small in size.

#### 4.1.3. Experimental setups

The Pytorch framework is employed for the training of ore sorting images. The image input size is 640 × 640. The optimal weight model is obtained after 300 epochs of training. The configuration of the experimental platform is presented in Table 1.

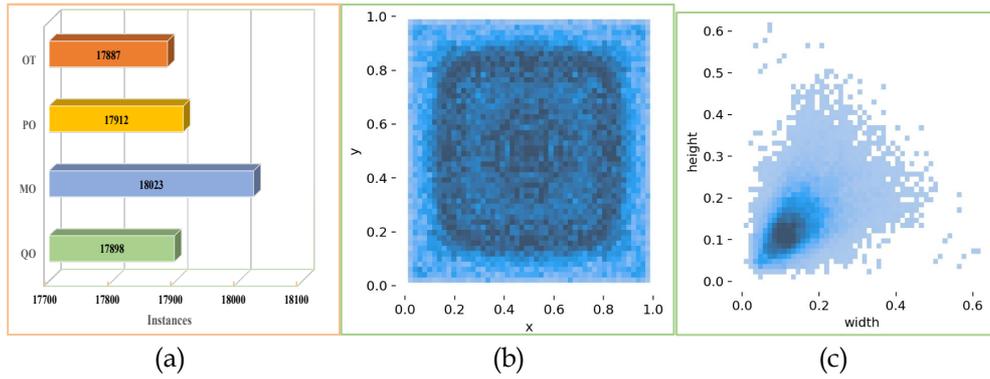


Fig. 13. Number of instances and labeling information in the Dataset, (a) is the number of instances, (b) is the location information of the instances, (c) is the size information of the instances

Table 1. Experimental configuration

configuration	Details
GPU	Tesla V100
Operating systems	Ubuntu 18.04
Memory	32G
GPU acceleration tools	CUDA 11.1

In assessing the efficiency of the LFOS-YOLO model, the mean average precision (mAP) serves as an assessment of the accuracy of the model. The number of model parameters (Para.) serves as a measure of model size. The computational complexity of the model is evaluated using Giga Floating-Point Operations per second (GFLOPs). Whereas the speed of model inference is evaluated using Frames per Second (FPS). mAP is calculated based on the following equations (5) to (7) where equation (5) is used for calculating precision, equation (6) is used for determining recall, and equation (7) is used for calculating the mAP. While the FPS is calculated based on equation (8).

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$mAP = \frac{1}{n} \int_0^1 P(R) dR \quad (7)$$

where TP represents the number of accurately identified objects, FP denotes the number of incorrectly identified objects, and FN corresponds to the number of objects not detected by the system. The variable "n" represents the total number of categories, while "P" and "R" represent precision and recall, respectively.

$$FPS = \frac{1000}{T_i+T_j+T_k} \quad (8)$$

where for each image  $T_i$  is the preprocessing time,  $T_j$  is the inference time and  $T_k$  is the post-processing time.

## 4.2. Analysis and verification

### 4.2.1. Ablation study

The objective of the ablation experiments was to compare and contrast the impact of the FBlock module, the PCHead module, and the simAM module on the model detection accuracy, the number of parameters, the computational complexity, and inference speed in order to identify the optimal model configuration. In this study, YOLOv8n was employed as the baseline model, and the mAP, Para., GFLOPs, and FPS values were utilized as metrics for ablation studies. All ablation experiments were performed multiple times under the same conditions as the experimental setups previously described.

As illustrated in Table 2, Model A represents the baseline configuration without additional techniques, exhibiting a mAP of 92.3%, a Para. of 3.15M, GFLOPs of 8.9, and an FPS of 59. Models B, C, D, and E are all derived from Model A and serve as the baseline reference.

Model B uses the FBlock module to improve the C2f in backbone and neck based on the benchmark model, resulting in an increase of 1.4% in mAP of 93.7%, a decrease of 22.5% in Para. value of 2.44M, a decrease of 21.3% in GFLOPs of 7.0, and an increase of 8.4% in FPS of 64. It can be seen that for the addition of FBlock there is a large reduction in the model's Para. and GFLOPs, as well as an increase in mAP and FPS, mainly because FBlock replaces the regular convolution with PConv, thus facilitating the lightweighting of the model without decreasing the accuracy.

Model C adds the parameter-free attention mechanism to the baseline model, and the results obtained are 95.1% for mAP, a 2.8% improvement, 3.15M for Para. with no increase in parameters, 8.8 for GFLOPs, a 1.1% decrease, and 69 for FPS, an 16.9% improvement. It can be seen that the addition of the simAM attention mechanism has a better improvement on the mAP and FPS of the model, which is mainly because simAM can enhance the feature extraction ability of the model, leading to an increase in the accuracy and detection speed of the ore sorting task.

Model D improves the detection head with PCHead, and the results obtained are mAP of 94.6%, a 2.3% improvement, Para. of 2.45M, a 22.2% reduction, GFLOPs of 5.6, a 37.1% reduction, and an FPS of 77, a 30.5% improvement, which is mainly due to the fact that PCHead replaces PConv and PWConv with multiple stacked 3\*3 regular convolutions, leading to a reduction in parameters, operations and memory accesses of the model, and thus a dramatic increase in the speed of detection.

In order to further investigate the impact of FBlock, SimAM and PCHead modules on the improvement of the algorithm, models E (FBlock+SimAM), F (FBlock+PCHead) and G (SimAM+PCHead) were designed for experiments. Finally, FBlock, SimAM and PCHead were integrated to obtain model H. This process involved the lightening and enhancement of the model's backbone, neck and head, and the incorporation of the parameter-free attention mechanism simAM. The results obtained were as follows: 95.5% for mAP, marking a 3.2% improvement, 1.75M for Para., indicating a 44.4% reduction, 3.8 for GFLOPs, representing a 57.3% decrease, and FPS of 72 is 22.0% higher than the baseline model.

Table 2. The ablation comparison

Model	YOLOv8	FBlock	SimAM	PCHead	mAP(%)	Para. (M)	GFLOPs	FPS
A	√				92.3	3.15	8.9	59
B	√	√			93.7 (+1.4)	2.44 (-22.5%)	7.0 (-21.3%)	64 (+8.4%)
C	√		√		95.1 (+2.8)	3.15 (0)	8.8 (-1.1%)	69 (+16.9%)
D	√			√	94.6 (+2.3)	2.45 (-22.2%)	5.6 (-37.1%)	77 (+30.5%)
E	√	√	√		95.2(+2.9)	2.44(-22.5%)	7.1(-20.2%)	68(+15.2%)
F	√	√		√	94.5(+2.2)	1.75(-44.4%)	3.7(-58.4%)	71(+20.3%)
G	√		√	√	95.1(+2.8)	2.45(-22.2%)	5.9(-33.7%)	74(+25.4%)
H	√	√	√	√	95.5 (+3.2)	1.75 (-44.4%)	3.8 (-57.3%)	72 (+22.0%)

The outcomes of the ablation experiments corroborate the rationality and efficacy of the LFOS-YOLO model. In particular, the incorporation of lightweight modules and a parameter-free attention mechanism can mitigate the storage and computational complexity, while also enhancing the accuracy and speed of detection.

#### 4.2.2. Pruning study

The pruning experiments were conducted in three phases: initial training with a sparsification method, subsequent channel pruning, and finally fine-tuning. To ascertain the efficacy of the pruning algorithm, a comparison is made between the metrics of the LFOS-YOLO model before and after pruning. Table 3 illustrates the metrics for pruning rates between 10% and 70%. It can be observed that the mAP value declines, reaching a minimum of 91.0% at a pruning rate of 70%. Similarly, the para. and GFLOPs demonstrate a consistent decline. The FPS value exhibits slight fluctuations, which can be attributed to

the influence of  $T_i$  and  $T_k$  on this metric, in addition to  $T_j$ . After weighing the various indexes, the model with a pruning rate of 50% was selected as the final model. The comparison results with the YOLOv8 benchmark model are presented in Table 4. The mAP value is 95.4%, representing an improvement of 3.1%. The Para value is 1.46M, indicating a reduction of 53.7%. The GFLOPs value is 3.4, reflecting a decrease of 61.8%. The FPS value is 74, demonstrating an improvement of 25.4%.

Table 3. Pruning experiment

method	mAP (%)	Para. (M)	GFLOPs	FPS
Before prune	95.5	1.75	3.8	72
Prune rate=10%	95.4	1.74	3.8	72
Prune rate=20%	95.5	1.74	3.7	71
Prune rate=30%	95.4	1.72	3.7	72
Prune rate=40%	95.4	1.62	3.6	71
<b>Prune rate=50%</b>	<b>95.4</b>	<b>1.46</b>	<b>3.4</b>	<b>74</b>
Prune rate=60%	93.0	1.36	3.3	73
Prune rate=70%	91.0	1.29	3.2	74

Table 4. Final model vs. baseline model

method	mAP(%)	Para. (M)	GFLOPs	FPS
YOLOv8n	92.3	3.15	8.9	59
Final	95.4(+3.1)	1.46(-53.7%)	3.4(-61.8%)	74(+25.4%)

#### 4.2.3. Model comparison

In order to demonstrate the effectiveness of LFOS-YOLO in the ore sorting task, 8 lightweight versions of the popular classical YOLO models were built for comparative experiments. These include YOLOv3-tiny (Yi et al., 2019), YOLOv4-tiny (Cui et al., 2023), YOLOv5 (You and Liu, 2023), YOLOv6n (Song et al., 2023), YOLOv7-tiny (Li et al., 2023), YOLOv8n (Wu et al., 2023), YOLOv9 (Wang et al., 2024) and YOLOv10 (Wang et al., 2024), YOLO11. Results from experiments of these models in this paper's dataset are shown in Table 5. Specifically, LFOS-YOLO achieves 95.4 % accuracy on this paper's dataset, 2.1 % higher than the sub-optimal Yolov6n (93.3%), Para. of 1.46M, 1.19M less than the sub-optimal Yolov5n (2.65M), GFLOPs of 3.4, 3.0 less than the sub-optimal Yolo11n (6.4), the FPS of 74 is slightly lower than the optimal Yolov10n (75) by 1.0%. The outcomes indicate that LFOS-YOLO offers an optimal compromise between detection precision, detection speed, and model size, rendering it a suitable candidate for implementation at mining sites.

Table 5. Comparison with other YOLO algorithms

method	mAP(%)	Para.(M)	GFLOPs	FPS
YOLOv3-tiny	92.0	12.2	19.1	53
YOLOv4-tiny	91.8	11.5	9.6	64
YOLOv5n	93.0	<b>2.65</b>	7.8	63
YOLOv6n	<b>93.3</b>	4.5	13.1	70
YOLOv7-tiny	91.1	12.3	19.5	59
YOLOv8n	92.3	3.15	8.9	59
YOLOv9T	92.1	2.61	10.7	71
YOLOv10n	91.7	2.69	8.2	<b>75</b>
YOLO11n	93.1	2.59	<b>6.4</b>	73
LFOS-YOLO	95.4	1.46	3.4	74

### 4.3. Detection performance visualization

To evaluate the performance changes of the model during the training process, the convergence trend curve of the mAP values, as shown in Fig.14, was plotted. In this graph, epochs are on the x-axis and mAP values on the y-axis, providing a clear visualization of the overall improvement in model accuracy. The mAP values show a rapid increase in the early stages of training, followed by a gradual stabilization, indicating that the model possesses good learning ability and generalization performance.

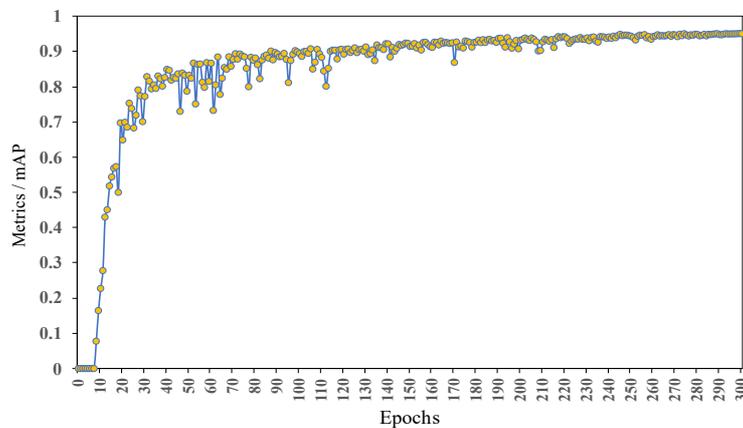


Fig. 14. Convergence trend plot of mAP for the training process of LFOS-YOLO model

As illustrated in Fig. 15, the LFOS-YOLO model demonstrates favorable outcomes in ore sorting experiments. The predicted bounding box effectively encompasses the outlines of the targets, and the model exhibits an aptitude for accurately identifying the per-target species. The visualization outcomes indicate that LFOS-YOLO is capable of accurately locating and classifying targets with high confidence, both in optimal conditions with complete ore samples and in complex conditions with non-minerals. Moreover, the Grad-CAM technique (Selvaraju et al., 2020) was employed in this experiment to facilitate the visualization of the model weights. Specifically, the focus of attention of a given feature map in the model is presented in the form of a heat map. Such a map indicates the degree of attention paid to a given feature, with red representing a higher level of attention, or higher network weight, and blue indicating a lower level of attention, or lower weight. Fig. 16 illustrates the heat map generated by the network layer of the LFOS-YOLO model after incorporating the simAM attention mechanism. The left image depicts the detection results in an all-ore scenario, demonstrating that the model accurately focuses on the target ore. The right image depicts a scenario containing metal and plastic interference elements, indicating that the model demonstrates robust attention for non-ore objects as well.

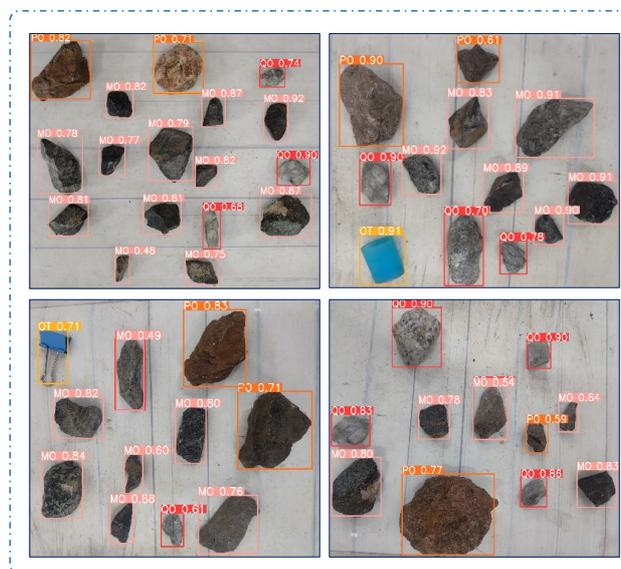


Fig. 15. The visualization of the LFOS-YOLO ore sorting process in a real-world operational environment

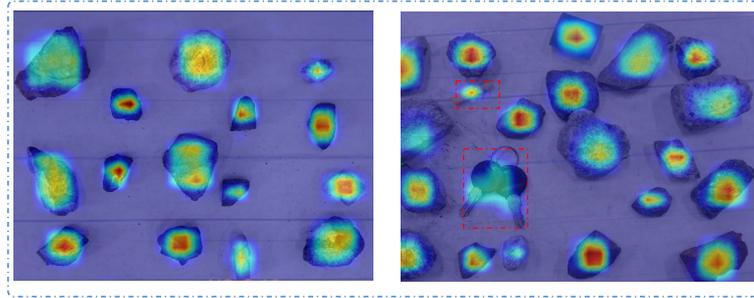


Fig. 16. The Grad-CAM visualization of the LFOS-YOLO model illustrates the activation map

## 5. Conclusions

The advancement of intelligent ore sorting research is of paramount importance to enhance the efficiency of mineral resource mining. In this study, we construct a novel dataset using raw ore samples from a mine in Liaoning Province, China. In light of the limitations of existing smart ore sorting models, which are characterized by a large number of parameters, high computational complexity, and slow detection speed, this study proposes a novel ore sorting network, LFOS-YOLO, which is both lighter and faster. The model draws inspiration from the YOLO object detection algorithm and employs a partial convolution to design lightweight backbone, neck, and head modules. Furthermore, to enhance the model's feature extraction capability, we incorporate a parameter-free attention mechanism and optimize the network structure by eliminating redundant channels, thus achieving a lean and efficient model. The experimental results demonstrate that the mAP value of LFOS-YOLO reaches 95.4, representing a 3.1% improvement in comparison to the benchmark model. Concurrently, the number of model parameters is markedly diminished by 53.7%, reaching 1.46 M, while the GFLOPs are considerably reduced by 61.8%, reaching 3.4. Additionally, the FPS is markedly elevated by 25.4%, reaching 74. In the comparative experiments, LFOS-YOLO was evaluated in relation to the traditional YOLO family of algorithms. The results demonstrated that LFOS-YOLO exhibited notable advantages in performance. The proposed model is lighter, faster and well suited for application in the field of ore sorting. Although this study is based on ore from a specific mine, the model is novel, reasonable and has the ability to be migrated and generalized. The model can be applied to the ore sorting process at other mines. If the ore type of the new mine is exactly the same as the ore type defined in this paper, the model in this paper can be used directly. If the new mine has new ore types in addition to the ore types defined in this paper, only incremental learning training is required for the new ore types. However, if the ore type is completely different from the ore type defined in this paper, the dataset of the new ore needs to be collected for the transfer learning of the model.

## Acknowledgment

This work was supported in part by the Science and Technology Research Project GJJ2203618, 2022, Department of Education of Jiangxi Province and the Special Project for Postgraduate Innovation of Jiangxi Province (YC2021-S577).

## References

- ABD HALIM, M.H., et al., 2023. *X-Ray Fluorescence (XRF) Analysis of Iron Ore at Ancient Kedah Iron Smelting Site, Sungai Batu Archaeological Complex, Bujang Valley, Kedah, Malaysia*. *Heliyon*, 9(4), e14850.
- BELLUSCI, N., et al., 2022. *Coarse Beneficiation of Trona Ore by Sensor-Based Sorting*. *Mining, Metallurgy & Exploration*, 39(5), 2179–2185.
- BOCHKOVSKIY, A., et al., 2020. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. *arXiv:2004.10934*.
- CHEN, J., et al., 2023. *Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks*. *arXiv:2303.03667*.
- CHEN, Y., et al., 2024. *Multispectral LiDAR-Based Underwater Ore Classification Using a Tunable Laser Source*. *Optics Communications*, 551, 129903.
- CHEN, Y., et al., 2018. *Feasibility Study of Ore Classification Using Active Hyperspectral LiDAR*. *IEEE Geoscience and Remote Sensing Letters*, 15(11), 1785–1789.

- CHOLLET, F., 2017. *Xception: Deep Learning with Depthwise Separable Convolutions*. CVPR 2017, 1800–1807.
- CUI, M., et al., 2023. *LES-YOLO: A Lightweight Pinecone Detection Algorithm Based on Improved YOLOv4-Tiny Network*. *Computers and Electronics in Agriculture*, 205, 107613.
- DONG, L., et al., 2021. *Prediction and Analysis of Time Series with the Aid of Granular Parallel Support Vector Machine*. *IEEE ICNSC*, 1, 1–5.
- HE, K., et al., 2016. *Deep Residual Learning for Image Recognition*. CVPR 2016, 770–778.
- HE, L., et al., 2024. *Application of Target Detection Based on Deep Learning in Intelligent Mineral Identification*. *Minerals*, 14(9), 873.
- IZADI, H., et al., 2017. *An Intelligent System for Mineral Identification in Thin Sections Based on a Cascade Approach*. *Computers & Geosciences*, 99, 37–49.
- JIANG, Q., et al., 2022. *Development of a Core Feature Identification Application Based on the Faster R-CNN Algorithm*. *Engineering Applications of Artificial Intelligence*, 115, 105200.
- YO, K., et al., 2024. *Optimizing Prior Distribution Parameters for Probabilistic Prediction of Remaining Useful Life Using Deep Learning*. *Reliability Engineering & System Safety*, 242, 109793.
- KHORRAM, F., et al., 2017. *Lithological Classification and Chemical Component Estimation Based on the Visual Features of Crushed Rock Samples*. *Arabian Journal of Geosciences*, 10(15), 324.
- KORBAN, M., LI, X., 2023. *Semantics-Enhanced Early Action Detection Using Dynamic Dilated Convolution*. *Pattern Recognition*, 140, 109595.
- KRIZHEVSKY, A., et al., 2012. *ImageNet Classification with Deep Convolutional Neural Networks*. NeurIPS 2012.
- LEI, S., et al., 2020. *Visual Classification Method Based on CNN for Coal-Gangue Sorting Robots*. CACRE 2020, 543–547.
- LI, C., et al., 2022. *YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications*. arXiv:2209.02976.
- LI, J., et al., 2023. *Real-Time Pineapple Detection for Agricultural Robot via Lightweight YOLOv7-Tiny Model*. *Procedia Computer Science*, 226, 92–98.
- LIU, C., et al., 2019. *An Enhanced Rock Mineral Recognition Method Integrating a Deep Learning Model and Clustering Algorithm*. *Minerals*, 9(9), 516.
- LIU, M., et al., 2023. *LF-YOLO: A Lighter and Faster YOLO for Weld Defect Detection of X-Ray Image*. *IEEE Sensors Journal*, 23(7), 7430–7439.
- LIU, Y., et al., 2021. *Deep Learning Based Mineral Image Classification Combined With Visual Attention Mechanism*. *IEEE Access*, 9, 98091–98109.
- LIU, Y., et al., 2021. *Ore Image Classification Based on Small Deep Learning Model: Evaluation and Optimization of Model Depth, Model Structure and Data Size*. *Minerals Engineering*, 172, 107020.
- LIU, Y., et al., 2024. *OreFormer: Ore Sorting Transformer Based on ConvNet and Visual Attention*. *Natural Resources Research*, 33(2), 521–538.
- LIU, Z., et al., 2017. *Learning Efficient Convolutional Networks through Network Slimming*. arXiv:1708.06519.
- LUO, X., et al., 2022. *A Review of Intelligent Ore Sorting Technology and Equipment Development*. *International Journal of Minerals, Metallurgy and Materials*, 29(9), 1647–1655.
- MA, H., et al., 2021. *Rock Thin Sections Identification Based on Improved Squeeze-and-Excitation Networks Model*. *Computers & Geosciences*, 152, 104780.
- NIZINSKI, C.A., et al., 2022. *Characterization of Uncertainties and Model Generalizability for CNN Predictions of Uranium Ore Concentrate Morphology*. *Chemometrics and Intelligent Laboratory Systems*, 225, 104556.
- OKADA, N., et al., 2021. *Classification of Arsenic Bearing Minerals Using Hyperspectral Imaging and Deep Learning for Mineral Processing*. *Journal of MMIJ*, 137(1), 1–9.
- PRATAMA, B.G., et al., 2023. *Building YoloV4 Models for Identification of Rock Minerals in Thin Section*. *IOP Conference Series: Earth and Environmental Science*, 1151(1), 012046.
- REDMON, J., et al., 2016. *You Only Look Once: Unified, Real-Time Object Detection*. CVPR 2016, 779–788.
- REDMON, J., FARHADI, A., 2016. *YOLO9000: Better, Faster, Stronger*. arXiv:1612.08242.
- REDMON, J., FARHADI, A., 2018. *YOLOv3: An Incremental Improvement*. arXiv:1804.02767.
- REN, S., et al., 2016. *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. arXiv:1506.01497.
- ROBBEN, C., et al., 2020. *X-Ray-Transmission Based Ore Sorting at the San Rafael Tin Mine*. *Minerals Engineering*, 145, 105870.
- ROBBEN, C., WOTRUBA, H., 2019. *Sensor-Based Ore Sorting Technology in Mining – Past, Present and Future*. *Minerals*, 9(9), 523.

- SELVARAJU, R.R., et al., 2020. *Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization*. *International Journal of Computer Vision*, 128(2), 336–359.
- SIMONYAN, K., ZISSERMAN, A., 2014. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv:1409.1556.
- SONG, T., et al., 2023. *Lightweight Detection Network Based on Receptive-Field Feature Enhancement Convolution and Three Dimensions Attention for Images Captured by UAVs*. *Image and Vision Computing*, 140, 104855.
- SUN, G., et al., 2022. *Machine Learning Coupled with Mineral Geochemistry Reveals the Origin of Ore Deposits*. *Ore Geology Reviews*, 142, 104753.
- TU, C., et al., 2023. *Hyperspectral Image Classification Based on Residual Dense and Dilated Convolution*. *Infrared Physics & Technology*, 131, 104706.
- WAN, D., et al., 2023. *YOLO-HR: Improved YOLOv5 for Object Detection in High-Resolution Optical Remote Sensing Images*. *Remote Sensing*, 15(3), 614.
- WANG, A., et al., 2024. *YOLOv10: Real-Time End-to-End Object Detection*. arXiv:2405.14458.
- WANG, C.-Y., BOCHKOVSKIY, A., et al., 2022. *YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors*. arXiv:2207.02696.
- WANG, C.-Y., YEH, I.-H., et al., 2024. *YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information*. arXiv:2402.13616.
- WANG, Z., et al., 2023. *An Efficient Detection of Non-Standard Miner Behavior Using Improved YOLOv8*. *Computers and Electrical Engineering*, 112, 109021.
- WU, Y., et al., 2023. *LCA-YOLOv8-Seg: An Improved Lightweight YOLOv8-Seg for Real-Time Pixel-Level Crack Detection of Dams and Bridges*. *Applied Sciences*, 13(19), 10583.
- XIANG, Q., et al., 2024. *Quadruplet Depth-Wise Separable Fusion Convolution Neural Network for Ballistic Target Recognition with Limited Samples*. *Expert Systems with Applications*, 235, 121182.
- XIAO, D., et al., 2021. *Iron Ore Identification Method Using Reflectance Spectrometer and a Deep Neural Network Framework*. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 248, 119168.
- XIE, S., SUN, H., 2023. *Tea-YOLOv8s: A Tea Bud Detection Model Based on Deep Learning and Computer Vision*. *Sensors*, 23(14), 6576.
- YANG, L., et al., 2021. *SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks*. *Proceedings of the 38th International Conference on Machine Learning*, 11863–11874.
- YI, Z., et al., 2019. *An Improved Tiny-Yolov3 Pedestrian Detection Algorithm*. *Optik*, 183, 17–23.
- YOU, K., LIU, H., 2023. *Feature Detection of Mineral Zoning in Spiral Slope Flow under Complex Conditions Based on Improved YOLOv5 Algorithm*. *Physica Scripta*, 2023.
- ZENG, D., et al., 2025. *Nowatorska Metoda Sortowania Rafinowanej Rudy Magnetytu Oparta Na Indukcji Magnetycznej i Sieci CNN-SK-BiLSTM*. *Gospodarka Surowcami Mineralnymi - Mineral Resources Management*, 2025, 123–123.
- ZHANG, Y., et al., 2023. *Faster OreFSDet: A Lightweight and Effective Few-Shot Object Detector for Ore Images*. *Pattern Recognition*, 141, 109664.
- ZHANG, Y.R., et al., 2021. *Assessment of Sortability Using a Dual-Energy X-Ray Transmission System for Studied Sulphide Ore*. *Minerals*, 11(5), 490.
- ZHU, X., et al., 2021. *TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-Captured Scenarios*. *ICCVW 2021*, 2778–2788.