

Article

Not peer-reviewed version

---

# CM-YOLOv8: Lightweight YOLO for Coal Mine Fully Mechanized Mining Face

---

[Yingbo Fan](#) , [Shanjun Mao](#) <sup>\*</sup> , [Mei Li](#) , [Zheng Wu](#) , [Jitong Kang](#)

Posted Date: 25 January 2024

doi: [10.20944/preprints202401.1814.v1](https://doi.org/10.20944/preprints202401.1814.v1)

Keywords: object detection; coal mine; YOLOv8; lightweight purning



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Article*

# CM-YOLOv8: Lightweight YOLO for Coal Mine Fully Mechanized Mining Face

Yingbo Fan, Shanjun Mao \*, Mei Li, Zheng Wu and Jitong Kang

Institute of Remote Sensing and Geographic Information Systems, Peking University, No. 5 Summer Palace Road, 100000, Beijing, China; ybfan@stu.pku.edu.cn(Y.F.); mli@pku.edu.cn(M.L.); zheng\_wu@pku.edu.cn(Z.W.); 2101210061@stu.pku.edu.cn(J.K.)

\* Correspondence: sjmao\_pku@163.com

**Abstract:** With the continuous development of deep learning, the application of object detection based on deep neural network in the coal mine has been expanding. Simultaneously, as the production applications demand higher recognition accuracy, most research chooses to enlarge the depth and parameters of the network to improve accuracy. However, due to the limited computing resources in the coal mining face, it is challenging to meet the computation demands of a large number of hardware resources. Therefore, this paper proposes a lightweight object detection algorithm designed specifically for coal mining face, referred to as CM-YOLOv8. The algorithm introduces adaptive predefined anchor boxes tailored to the coal mining face dataset to enhance the detection performance of various targets. Simultaneously, a pruning method based on the L1 norm is designed, significantly compressing the model's computation and parameter volume without compromising accuracy. The proposed algorithm is validated on the coal mining dataset DsLMF+, achieving a compression rate of 40% on the model volume with less than a 1% drop in accuracy. Comparative analysis with other existing algorithms demonstrates its efficiency and practicality in coal mining scenarios. The experiments confirm that CM-YOLOv8 significantly reduces the model's computational requirements and volume while maintaining high accuracy.

**Keywords:** object detection; coal mine; YOLOv8; lightweight pruning

## 1. Introduction

In recent years, with the continuous growth of global energy demand, coal, as a crucial energy resource that has served industrial production for centuries, still plays a vital role in the modern energy system [1–5]. It is currently widely used in electricity generation and industrial manufacturing, providing reliable energy support for societal development [6–8]. However, despite the undeniable significance of coal in the energy industry, the process of coal mining still faces a series of technical difficulties and challenges. Coal mining methods are typically associated with highly hazardous working environments, such as the risks of gas explosions, roof collapses, and other safety hazards, posing a threat to the life and safety of miners [9,10]. To enhance the safety of coal mining, it is necessary to introduce advanced safety identification technologies and systems to achieve more effective monitoring and prevention. Simultaneously, the widespread application of comprehensive intelligent mining technologies has also brought about new challenges [11,12]. In these highly automated work areas, the rapid and precise identification of various production processes and surrounding environments has become a pivotal factor in ensuring production efficiency. Traditional methods of manual observation and monitoring are inadequate to meet the demands of modern coal mining. Therefore, there is a need to rely on advanced image recognition and computer vision technologies to achieve real-time monitoring of the working face's status [13–17].

Currently, common image recognition methods for coal mine fully-mechanized working faces mainly include traditional methods based on feature extraction and deep learning-based methods. Traditional approaches employ image processing techniques such as edge detection, texture analysis,

and shape descriptors. These techniques extract manually designed features from images, which are subsequently utilized for target detection and classification [18–20]. Presently, machine learning-based methods, such as Support Vector Machine (SVM) [21] or Random Forest [22], are more widely applied. These methods combine manually extracted features for the classification and detection of coal mine images. Alternatively, Convolutional Neural Networks (CNNs), such as LeNet [23], AlexNet [24], VGG [25], or deeper networks, are employed for end-to-end feature learning and target detection of images from fully-mechanized coal mine working faces [26]. Further advancements in deep learning-based object detection algorithms, such as Faster R-CNN [27], YOLO (You Only Look Once) [28], SSD (Single Shot MultiBox Detector) [29], also enable efficient and precise localization and classification of targets in coal mine images.

However, for the paramount safety requirements of coal mine working faces, real-time and rapid identification of abnormal situations to minimize accident risks is an extremely crucial demand [30,31]. Timely detection of geological structural changes, roof collapses, hazardous personnel movements, or gas leaks is essential for taking prompt measures [32–34]. Therefore, in coal mine working faces, real-time capability is a key requirement, particularly for automation and safety monitoring systems [35,36]. Some deep learning algorithms may face speed challenges due to computational resource limitations. Additionally, coal mine working faces are often constrained by computational resources, necessitating lightweight algorithms for rapid image processing and recognition in resource-constrained environments [37–39].

To address the aforementioned issues, this paper proposes a lightweight object detection algorithm tailored for fully-mechanized coal mine working faces. The algorithm is based on the YOLOv8 [40] object detection algorithm. Considering the typically fixed sizes and proportions of various identifiable targets within coal mine working face scenes, this paper designs predefined Anchor Boxes to predict the sizes of target boxes. Simultaneously, a lightweight optimization is applied to the network structure of YOLOv8. This optimization involves pruning operations based on the L1 norm for most CBS (Convolution, Batch Normalization, and Leaky ReLU) convolution modules [41] and some convolution operations in the network structure. This ensures that the object detection algorithm achieves more efficient hardware deployment and faster recognition speed when facing scenarios like fully-mechanized coal mine working faces for target identification. In the design of predefined Anchor Boxes, this paper initially analyzes the dataset of fully-mechanized coal mine working faces, gathering information on target width, height, and other aspects. Subsequently, the K-means clustering algorithm [42] is employed to cluster the target boxes in the dataset, thereby determining the sizes of the predefined Anchor Boxes. During clustering, the width and height of the target boxes can be utilized as features.

After performing K-means clustering, manual adjustments to the sizes of Anchor Boxes can be made to ensure their suitability for the shapes and proportions of targets in fully-mechanized coal mine working faces. If there are changes in the distribution of targets in the working faces, the clustering algorithm can be periodically rerun to adapt to the new data distribution, ensuring that the Anchor Boxes can still effectively capture the sizes of the targets. In designing pruning and quantization strategies, this paper adopts L1 norm-based pruning operations for the majority of convolution kernels. The pruned portions are convolved with zero-filled kernels. Additionally, different compression ratios can be dynamically adjusted according to the specific requirements of different practical scenarios, accommodating the actual needs of fully-mechanized coal mine working faces. This approach supports the efficient operation of the object detection algorithm on resource-constrained embedded systems and mobile devices.

## 2. Related Work

In the field of computer vision, object detection is a fundamental task that plays a crucial role in various human life and production applications, including areas such as autonomous driving, robotics, and intelligent security [43–46]. It has evolved from the extraction of handcrafted local invariant features, such as SIFT (Scale-Invariant Feature Transform) [47], HOG (Histogram of Oriented Gradients) [48], and LBP (Local Binary Patterns) [49]. The process then involves the

aggregation of local features, achieved through simple cascading or encoders, such as the Bag-of-Words model [50], SPM (Spatial Pyramid Matching) [51], and Fisher Vectors [52]. For many years, the dominant paradigm in computer vision relied on handcrafted local descriptors and discriminative classifiers, forming a multi-stage process. The landscape changed with the unprecedented success of DCNNs (Deep Convolutional Neural Networks) in image classification [53]. The success of DCNNs in image classification propelled a paradigm shift that extended into the field of object detection [54].

Until recent years, the representative two-stage object detector, RCNN, was proposed. It initially extracts candidate boxes based on the image and then refines the detection results by making a second correction based on these candidate regions. Subsequently, Faster RCNN introduced a fully convolutional network as the RPN (Region Proposal Network), introducing the concept of Anchors for classification and bounding box regression, further improving the accuracy of object detection [27]. Following this, algorithms like FPN [55] and Mask RCNN [56] enriched the components of Faster RCNN, enhancing its performance by adding a branch to parallelly conduct pixel-level object instance segmentation. While two-stage object detectors offer high accuracy, their detection speed is relatively slow, and their complex detection model workflow limits their development on small terminal devices. Therefore, one-stage detectors, represented by YOLO, were introduced. YOLO restructures the detection problem, treating it as a regression problem and directly predicting image pixels as the target and its bounding box properties [57].

Subsequently, based on the improved YOLOv4 [58], techniques such as data augmentation, regularization methods, class label smoothing, CIOU-loss, CmBN (Cross mini-Batch Normalization), self-adversarial training, and cosine annealing learning rate scheduling were employed to enhance training. It utilized the "bag of freebies" approach, which only increases training time without affecting inference time. The recent YOLOv8, by referencing designs from algorithms like YOLOX[59]、YOLOv6[60]、YOLOv7[61] and PPYOLOE[62], offers a new SOTA (state-of-the-art) model. In terms of loss function design, it incorporates the Task Aligned Assigner positive sample allocation strategy and introduces Distribution Focal Loss to further reduce precision loss [63]. Despite YOLOv8's outstanding performance in object detection tasks, it may face some challenges when dealing with complex scenarios like fully-mechanized coal mine working faces. For instance, its large parameter model may not be effectively deployed for real-time object detection on terminal devices. Moreover, YOLOv8 typically requires extensive and diverse training data to achieve optimal performance. In certain specific domains or for certain target types, additional data may be needed for effective training.

In the field of coal mining production, the application of object detection technology continues to evolve. Yang et al. proposed the use of sensor-based technologies such as LiDAR and radar to detect coal mine obstacles, monitor geological structures, and ensure worker safety [64]. However, sensor performance is often influenced by environmental conditions, and underground extreme conditions may lead to a decline or failure in sensor performance. Additionally, the manufacturing and maintenance costs of high-performance sensors are relatively high [65].

In the realm of image-based object detection, Pan et al. introduced an improved fast recognition model based on YOLO-v3 for rapid identification of coal and gangue [66]. Zhang et al. presented a YOLOv4 algorithm based on deep learning for coal gangue detection. The detection algorithm with optimization methods showed higher accuracy, recall rate, and real-time performance compared to SSD and Faster R-CNN detection algorithms [67]. Fan et al. proposed a coal particle morphology, particle size, liberation feature, and density separation process based on a CNN and an improved U-Net network model [68]. Wang et al. developed the Var-Con-Sin-GAN model and constructed a sample generation and feature transfer framework to address the issue of insufficient coal-rock image data [69]. The methods mentioned above have made some significant progress in the field of coal mine object detection, but there are still some challenges and areas for improvement. For instance, the application of some larger models in coal mine scenarios with limited computational resources may be restricted. Additionally, some models may face the challenge of reduced recognition accuracy in situations with inadequate lighting underground.



3. Materials and Methods

This paper introduces a lightweight object recognition algorithm tailored for fully-mechanized coal mine working faces. The algorithm employs predefined Anchor Boxes to predict the sizes of target boxes, enhancing the speed and accuracy of object detection. Simultaneously, it optimizes the network structure through pruning operations based on the L1 norm, significantly reducing model computational operations and size while maintaining nearly undiminished object recognition accuracy. Consequently, CM-YOLOv8 (Coal Mining-YOLOv8) achieves streamlined and optimized deep learning models with reduced parameter count and computational complexity. It provides a more real-time, efficient, and adaptable image recognition solution for fully-mechanized coal mine working faces, meeting the urgent demands of the mining industry for safety and production efficiency.

3.1. Predefined Anchor Box

This paper proposes a method for adaptive predefined anchor box generation using a genetic algorithm-based K-Means clustering analysis on the fully-mechanized coal mine working face dataset to enhance object detection performance. The method begins by reading training set images and extracting the width ( $w$ ), height ( $h$ ), and bounding box sizes ( $w_{bbox}$ ,  $h_{bbox}$ ) for each image. Different ratios of  $w$  and  $h$  are proportionally scaled to a specified size while maintaining aspect ratios. Each image and its bounding box are then scaled proportionally, ensuring that the relative coordinates of the bounding box remain unchanged. The scaled width and height of the corresponding image are multiplied, converting the bounding box coordinates from relative to absolute coordinates. Bounding boxes with widths and heights less than two pixels are filtered out because smaller objects may contribute more significantly to the training process. Next, K-means clustering is applied to the remaining bounding boxes to obtain  $k$  initial anchor boxes. This step helps initialize the genetic algorithm with diversified anchor box sizes. The genetic algorithm is then implemented to iteratively optimize the anchor box sizes. The algorithm involves randomly mutating the width and height of each anchor box and evaluating the fitness of the obtained set of anchor boxes using a custom fitness function. The mutation process is repeated up to 1000 times, adopting the new size if the mutation improves fitness; otherwise, the mutation is discarded. Finally, the anchor boxes are sorted based on their area, and the optimized anchor box sizes are returned for use in the object detection framework. The schematic diagram of the process is shown in Figure 1.

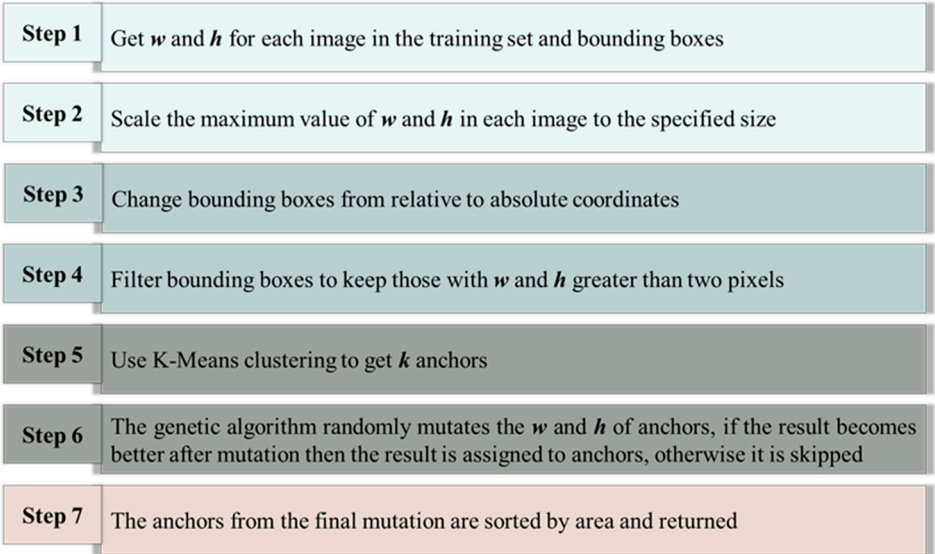


Figure 1. The steps of predefined Anchor Box generation method based on genetic algorithm.

The K-Means algorithm used in this method first initializes  $K$  cluster centers  $\{C_1, C_2, C_3, \dots, C_k\}$ ,  $1 < j \leq n$ , and then calculates the Euclidean distance from each data set sample  $X = \{X_1, X_2, X_3, \dots, X_n\}$  to each cluster center. The calculation formula is as follows:

$$dis(X_i, C_j) = \sqrt{\sum_{t=1}^m (X_{it} - C_{jt})^2} \quad (1)$$

Where  $X_{it}$  represents the attribute  $t$  of the data sample  $i$ , and  $C_{jt}$  represents the attribute  $t$  of the cluster center  $j$ . Subsequently, based on the distances, the data set samples are partitioned into  $K$  clusters, and the mean is calculated for each cluster to update the cluster centers. This iterative process continues until the specified number of iterations is reached. The mean calculation is represented by:

$$Center_k = \frac{1}{|C_k|} \sum_{X_i \in C_k} X_i \quad (2)$$

### 3.2. Pruning Based on L1 Norm

This paper addresses the issue of redundant network structures and computational complexity by proposing an L1 norm-based pruning method. The method divides the weight parameters of the convolutional layer to be pruned into  $n$  groups in the channel direction based on the parameter *group*, and calculates the L1 norm of weights on different channel numbers within each group. Pruning is then applied to the group's weight values exceeding a certain threshold, while retaining and participating in retraining for the group's weight values below the threshold. The weight parameters reaching the target accuracy after retraining are obtained, and the grouping and pruning operation is repeated until the network converges.

The calculation method for the intra-group channel L1 norm of each weight group in each convolutional layer to be pruned is shown in Equation (3). In the formula,  $w$  and  $h$  represent the maximum width and height of the current channel in the convolutional layer,  $c$  is the number of channels in the current convolutional kernel,  $n$  is the group after weight grouping in the convolutional layer to be pruned,  $i$  and  $j$  represent the current position of the weight value in the horizontal and vertical directions,  $w_{i,j}$  is the weight value at the current position,  $all$  is the total number of weights in the convolutional layer, and  $\|w_c\|_1$  is the L1 norm within channel  $c$ . Subsequently, pruning is determined based on comparing the L1 norm with the threshold within each group after segmentation.

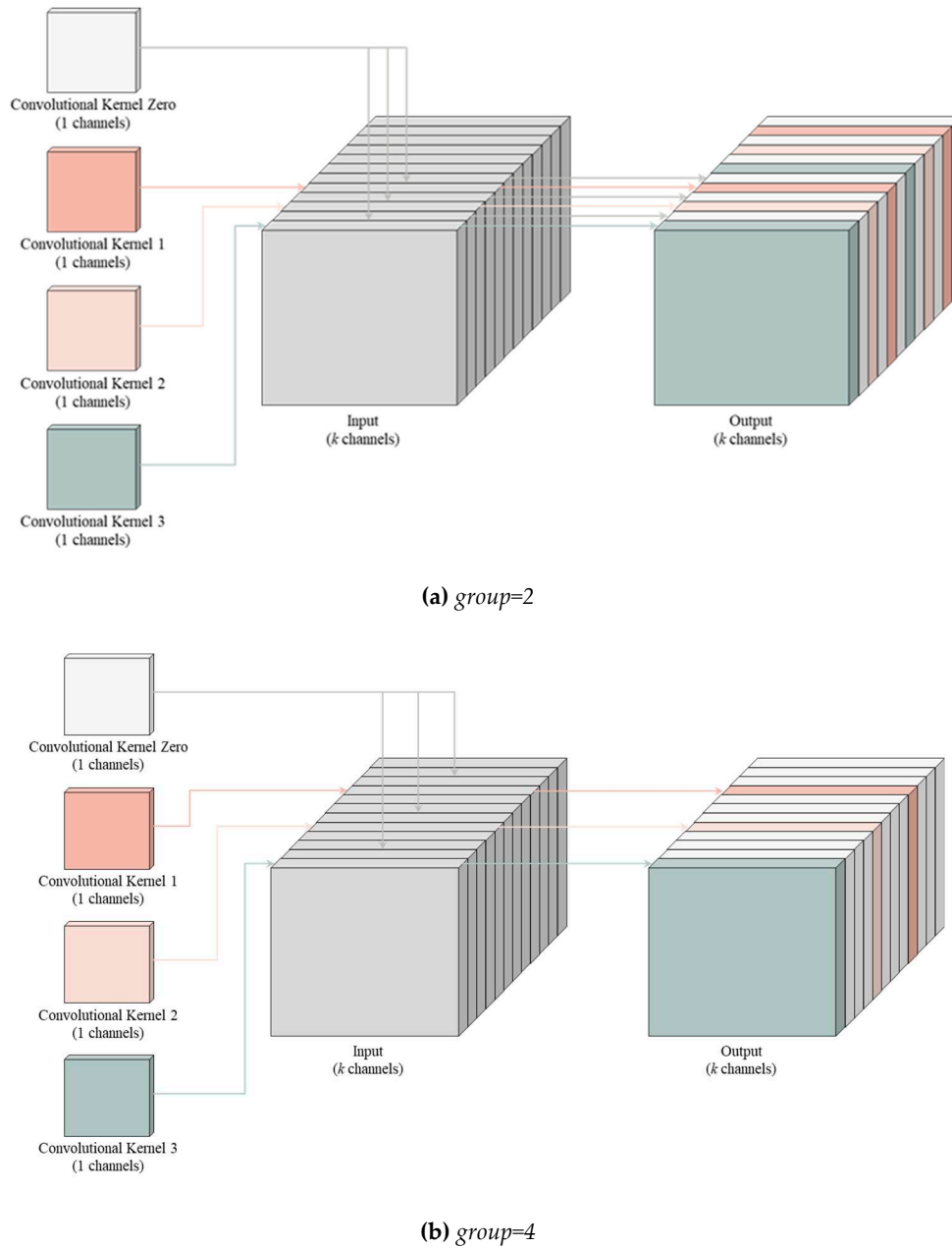
$$\|w_c\|_1 = \sum_{i=1, j=1}^{w, h} |w_{i,j}| \quad (3)$$

The calculation method for the threshold  $\theta_n$  within the group  $n$ , used to determine whether pruning should be applied, is shown in Equation (4). In the formula,  $w_j$  represents the  $j$  weight in the convolutional layer, and *group* indicates the pruning parameter set for the current convolutional layer.

$$\theta_n = \sum_{j=1}^{group} |w_j| / group \quad (4)$$

To provide a more illustrative description of how the L1 norm-based pruning method is applied to the CBS convolutional block, the paper designs the schematic diagram shown in Figure 2 for further clarification. Figure 2 (a) demonstrates the pruning approach when the parameter *group* is set to 2. Each pair of channels in the convolutional layer is grouped, and pruning is applied to channels with values below the threshold, effectively multiplying them by a zero matrix. This approach has the advantage of significantly reducing computation in the output and subsequent network calculations, allowing the feature map size to remain unchanged without affecting further pruning and retraining. Channels with values above the threshold continue with the depth-wise separable convolution calculation. Figure 2 (b) illustrates the pruning approach when the parameter *group* is

set to 4, following a similar process where each group comprises four channels in the convolutional layer. Channels with values below the threshold undergo pruning, while those above the threshold continue with the depth-wise separable convolution calculation.



**Figure 2.** L1-paradigm based operation of pruned CBS convolutional block.

### 3.3. Network Structure

The lightweight pruning-enhanced algorithm network structure based on YOLOv8 for coal mining face is illustrated in Figure 3. The input size of this network structure remains at  $640 \times 640 \times 3$  and is divided into three main parts: Backbone, Neck, and Head. The PCBS in the network structure represents the module pruned based on L1 norm using depth-wise separable convolution, and its process schematic is depicted in Figure 3 (d). Depth-wise separable convolution is a specialized convolution operation in convolutional neural networks aimed at reducing the number of parameters and computational complexity while maintaining model expressiveness. This type of convolution operation is widely used in resource-constrained environments such as mobile devices and embedded systems to improve model lightweighting and operational efficiency. However, the

conventional YOLOv8 network structure involves numerous CBS, leading to parameter redundancy and repetitive computations. Hence, this paper performs L1 norm-based pruning on the convolutional part of CBS, reducing parameter and computation overhead while striving to maintain recognition accuracy.

Meanwhile, YOLOv8 replaces the commonly used C3 module in the YOLO series algorithms with the C2f module, introducing more skip connections and additional Split operations [70]. The C3 module inserts a CSP (Cross Stage Partial) connection between each branch in the CSP structure, dividing the feature map into two parts. One part undergoes multiple layers of convolution before merging with the other part, aiding in information integration at different levels. The C2f module is designed based on the ideas of the C3 module and ELAN, as shown in Figure 3 (e). This design allows YOLOv8 to obtain richer gradient flow information while ensuring lightweight of network. Figure 3 (f) illustrates the detailed structure of the SPPF module. Spatial Pyramid Pooling (SPP) is commonly used in deep learning for spatial pyramid pooling methods [71]. This module is typically used to handle variations in input sizes to adapt to objects or scenes of different sizes. The SPPF module further optimizes the operation sequence and size of Maxpooling based on SPP, maintaining consistent sizes while accelerating computation speed. It addresses the problem of convolutional neural networks extracting redundant features related to the image, significantly improving the speed of generating candidate boxes.

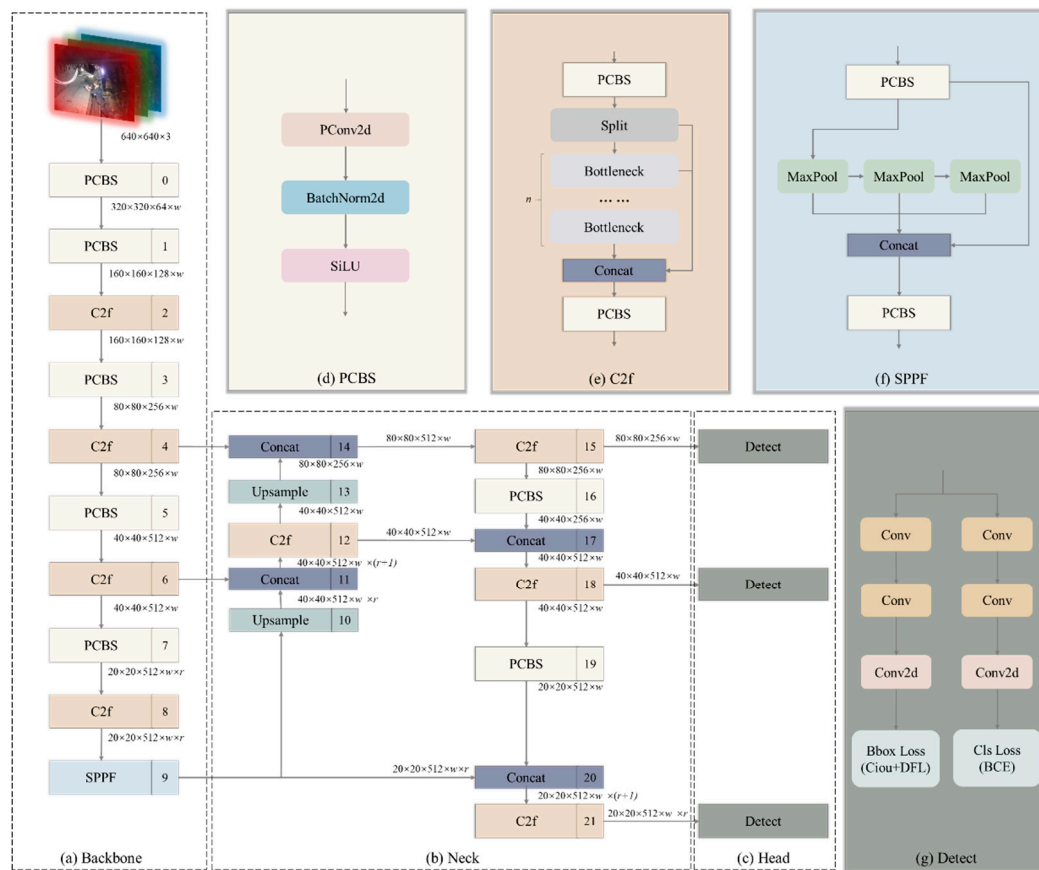


Figure 3. L1-paradigm based operation of pruned CBS convolutional block.

## 4. Results

### 4.1. Experiment Introduction

This section begins by introducing the dataset used in the experimental methodology, followed by descriptions of the experimental environment and training strategies. Finally, evaluation metrics related to the experimental results are presented.



4.1.1. Dataset

The dataset utilized in this study is the Underground Longwall Mining Face (DsLMF+) image dataset [72], comprising 138004 images annotated with six classes: coal miner, hydraulic support guard plate, large coal, mine safety helmet, miner behaviors, and towline. The dataset incorporates diverse angles, scenes, and tasks, providing a comprehensive representation with varied object categories (monotonous and diverse), object quantities (few and abundant), and object distributions (sparse and dense). Some representative images from the dataset are illustrated in Figure 4. All labels in the dataset are openly available in both YOLO and COCO formats, and domain experts in the mining field have assessed the dataset's utility and accuracy.



**Figure 4.** Samples of the DsLMF+ dataset.

Through the genetic algorithm-based K-Means clustering analysis on the coal mining face dataset, this study discovered that to achieve better recovery of pruning accuracy, the predefined aspect ratios for various target anchors should be as follows: Coal miner: [2.6, 2.9], Hydraulic support guard plate: [1.9, 2.3], Large coal: [0.8, 1.3], Mine safety helmet: [0.9, 1.2], Miner behaviors: [2.1, 2.5], Towline: [4.1, 4.5].

4.1.2. Experimental Environment and Training Strategies

The experimental setup for the CM-YOLOv8 algorithm in this study is composed of the components listed in Table 1. This includes the hardware platform and the deep learning framework employed in the experiments.

**Table 1.** Composition of experimental algorithm configuration workstation.

Parameters	Configuration
CPU	Intel Core i9-13900HX (2.2 GHz, 6 cores, 12 threads)
RAM	16GB DDR4 2400MHz
GPU	NVIDIA GeForce GTX 4070 Ti × 2
GPU memory size	16GB
cuDNN	8.9.6
CUDA	12.1
Deep learning framework	python 3.8.18 + pytorch 2.1.2

The YOLOv8m model was employed as the backbone network for improvement training in this study. This model follows all the experimental ideas of the YOLOv8 series, with the only modification being the scaling of network width and depth. The crucial parameter settings for the training process are outlined in Table 2.

**Table 2.** Important parameter settings during experimental training.

Parameters	Value
Epochs	300
Initial learning rate	$1 \times 10^{-2}$
Final learning rate	$1 \times 10^{-4}$
Momentum	0.937
Weight-Decay	$5 \times 10^{-4}$
Batch size	4
Mosaic	1.0
$\alpha$ (Wise-IoU)	1.9
$\delta$ (Wise-IoU)	3
Input image size	$640 \times 640$
Number of images	138004
Optimizer	SGD
Close Mosaic	Last 10 epochs

Following the partition rules of the DsLMF+ coal mine dataset, the dataset was divided into a training set (110403 images), a test set (16800 images), and a validation set (10801 images).

#### 4.1.3. Evaluation Indicators

To provide a detailed and accurate description of the excellent performance of the proposed CM-YOLOv8 improved model in this study, various evaluation metrics were introduced, including precision, recall, mAP0.5, mAP0.5:0.95, model parameter count, and model size.

where precision is how many of the samples predicted by the model to be in the positive category are truly in the positive category. It is calculated as shown in Equation (5):

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

Where TP (True Positives) represents the number of samples correctly predicted as the positive class by the model, and FP (False Positives) represents the number of samples incorrectly predicted as the positive class by the model.

Recall refers to how many actual positive class samples are correctly predicted as the positive class by the model. The calculation is shown in Equation (6):

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

Where FN (False Negatives) represents the number of samples incorrectly predicted as the negative class by the model.

Average Precision (AP) is equal to the area under the Precision-Recall curve, and the calculation Equation is shown in Equation (7):

$$AP = \int_0^1 Precision(Recall) d(Recall) \quad (7)$$

mAP (Mean Average Precision) is the weighted average of the AP values for all sample categories, used to measure the detection performance of the model across all categories. The Equation is shown in Equation (8):

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (8)$$

Where  $AP_i$  represents the AP value for the category with index  $i$ , and  $N$  is the number of categories in the training dataset. Additionally, mAP0.5 is the mean average precision at an IoU (Intersection over Union) threshold of 0.5. In object detection, IoU is used to measure the overlap

between the predicted bounding box and the ground truth bounding box [73]. mAP0.5:0.95 is an extension of mAP over a broader IoU range, similar to mAP0.5 but considering a wider IoU range from 0.5 to 0.95.

## 4.2. Experiment Results

This paper takes the DsLMF+ dataset as an example and conducts thorough validation on the coal mining dataset. To ensure that the test results of different network models are not influenced by factors other than model differences, consistent parameter settings are maintained for all network models during the experiments.

### 4.2.1. Quantitative Comparison of Different Models

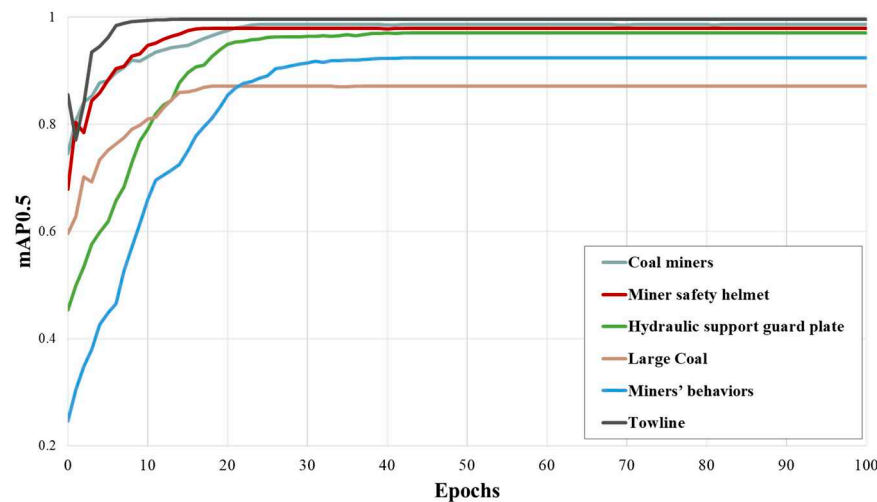
This paper compares the performance of YOLOv7 [61], DETA [74], ViT-Adapter [75], YOLOv8 [40], and the proposed CM-YOLOv8 algorithm on the DsLMF+ dataset, as shown in Table 3. From the table, it can be observed that compared to YOLOv7, YOLOv8 shows improved recognition accuracy due to its deeper network architecture. Additionally, YOLOv8 enhances detection speed by replacing the traditional C3 module with the C2f module, partially addressing its speed limitation. DETA, built on the Deformable DETR two-stage architecture, adopts a one-to-many matching strategy with a traditional IOU-based matching policy in CNN. Although DETA performs well overall, it is slightly inferior to YOLOv8. ViT (Vision Transformer), a vision processing model based on the Transformer architecture, shows relatively mediocre performance on the coal mining dataset. This could be attributed to ViT's emphasis on global information over local features, resulting in poorer recognition of details and local structures in the coal mining face scenario. The proposed CM-YOLOv8 achieves a slight decrease in average recognition accuracy compared to the YOLOv8m model (0.1%), while significantly reducing computational and parameter overhead. From the table, it is evident that in the recognition accuracy of Coal miners, Miners' behaviors, and Towline targets, the proposed algorithm maintains accuracy even after removing network redundancy.

**Table 3.** Comparison of quantitative experimental accuracy of different models.

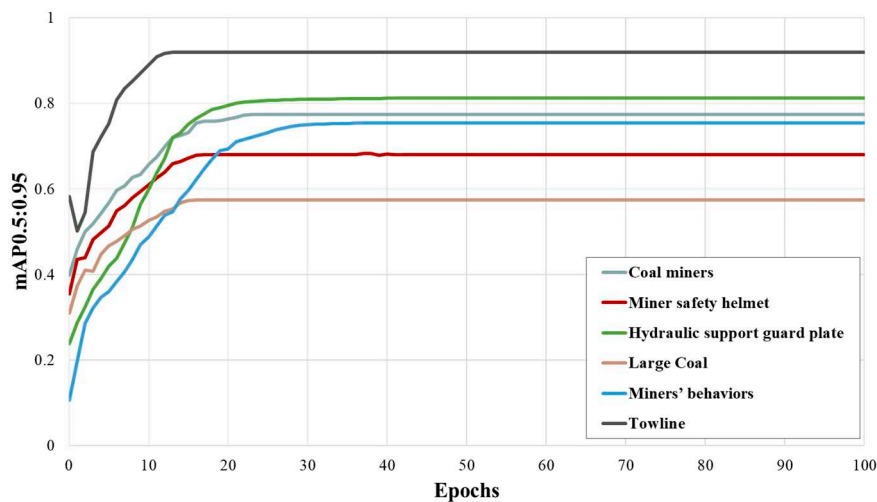
Method	YOLOv7				DETA				ViT-Adapter-L				YOLOv8m				Ours			
Metrics	Precision	Recall	mAP 0.5	mAP 0.5:0.95	Precision	Recall	mAP 0.5	mAP 0.5:0.95	Precision	Recall	mAP 0.5	mAP 0.5:0.95	Precision	Recall	mAP 0.5	mAP 0.5:0.95	Precision	Recall	mAP 0.5	mAP 0.5:0.95
Coal miners	0.965	0.968	0.986	0.773	0.967	0.970	0.976	0.684	0.961	0.965	0.966	0.702	0.968	0.971	0.988	0.775	0.968	0.970	0.986	0.774
Mine safety helmet	0.942	0.958	0.976	0.679	0.948	0.965	0.960	0.601	0.945	0.951	0.961	0.624	0.946	0.962	0.980	0.680	0.943	0.959	0.979	0.680
Hydraulic support guard plate	0.972	0.927	0.978	0.813	0.971	0.932	0.958	0.762	0.963	0.928	0.963	0.753	0.978	0.927	0.974	0.817	0.972	0.925	0.971	0.812
Large Coal	0.814	0.776	0.868	0.572	0.820	0.771	0.815	0.549	0.811	0.776	0.854	0.532	0.815	0.780	0.873	0.580	0.814	0.786	0.871	0.574
Miners' behaviors	0.880	0.880	0.913	0.752	0.884	0.886	0.914	0.718	0.879	0.874	0.928	0.714	0.882	0.882	0.926	0.754	0.883	0.879	0.924	0.754
Towline	0.995	0.997	0.997	0.916	0.996	0.998	0.989	0.915	0.995	0.992	0.989	0.871	0.997	0.997	0.996	0.920	0.997	0.996	0.996	0.919
Average	0.928	0.918	0.953	0.751	0.931	0.920	0.935	0.705	0.926	0.914	0.944	0.699	0.931	0.920	0.956	0.754	0.930	0.920	0.955	0.752

#### 4.2.2. Comparison of Loss Function Changes during Training for Different Kinds of Targets

This paper initially sets the training epochs to 300. However, after exceeding 100 epochs, the change in mAP becomes minimal and lacks visual value. Therefore, this paper illustrates the mAP variation curve for the first 100 epochs. As shown in Figure 5, the convergence of the training process for six types of target objects, including Coal miners, Miner safety helmet, Hydraulic support guard plate, Large Coal, Miners' behaviors, and Towline, is depicted. From the figure, it can be observed that Towline achieves the highest mAP0.5 recognition accuracy among all target objects, reaching 99.6%. In contrast, Large Coal has the lowest mAP0.5 recognition accuracy, which is 87.1%. This may be attributed to the significant diversity and complexity in the shape, color, and texture of large coal blocks, making feature extraction and matching more challenging. On the other hand, cable groove appearances are relatively consistent, making it easier to recognize and extract features. Each type of target generally converges before reaching 20 epochs, but Miners' behaviors require a relatively longer convergence time. This is because the appearance features of individuals are relatively easy to extract, typically involving static features such as faces and bodies. In contrast, behavioral features may include dynamic movements, poses, and other information, making their extraction more challenging and requiring a longer convergence time.



(a)mAP0.5



(b)mAP0.5:0.95

**Figure 5.** Comparison of mAP changes during training for different objects.



#### 4.2.3. Comparison of Computational and Model Parametric Quantities after Lightweight Pruning

This paper addresses the practical hardware deployment requirements for scenarios such as coal mining comprehensive mining faces and proposes a lightweight pruning method based on L1 norm. The comparative results of various methods, including recognition accuracy, parameter quantity, and computational load on the coal mining dataset, are presented in Table 4. From the table, it can be observed that early target detection algorithms, such as YOLOv3 [76] and HTC [77], are not conducive to deployment on small-terminal devices in coal mines due to their complex model structure and large parameter quantities, and their detection accuracy is relatively low. Although YOLOv7 achieves extreme compression of parameter and computational loads, its detection accuracy also significantly decreases, making it challenging to meet the safety requirements of actual coal mining production. Compared to other algorithms, YOLOv7 and YOLOv8m achieve a reduction in parameter quantity and computational load while ensuring high recognition accuracy. However, in the context of limited computing resources in coal mine production, these models are still challenging to deploy on small-terminal devices on a large scale. Moreover, both models adopt a three-scale detection network structure, which cannot meet the detection requirements for high-proportion small targets in the coal mining production scene, resulting in deficiencies in detection accuracy. In contrast, CM-YOLOv8, proposed in this paper, significantly compresses the required parameter quantity and computational load for model operation while ensuring that the recognition accuracy does not decrease or decreases minimally. Taking the network model with parameter setting  $group=2$  as an example, compared to YOLOv8m, the proposed model reduces the parameter quantity by 39.8% and the computational load by 44.6% while only sacrificing 0.1% in recognition accuracy. This enables the proposed network model to provide technical support for real-time deployment on small devices in coal mining comprehensive mining face scenarios.

**Table 4.** Comparison of quantitative experimental accuracy of different models.

Method	mAP:0.5(%)	Params(M)	GFLOPs
YOLOv3[76]	85.9	61.5	193.9
YOLOv5m[78]	87.7	21.8	39.4
YOLOv7[61]	95.3	36.5	103.5
YOLOv7-tiny[79]	78.4	6.0	13.1
YOLOv8m[40]	95.6	25.9	78.9
Faster-RCNN[27]	86.2	40.2	207.3
DETR[80]	79.4	41.9	225.7
DAB-DETR[81]	90.3	44.8	256.1
HTC[77]	82.3	80.5	441.3
<b>Ours(<math>group=2</math>)</b>	<b>95.5</b>	<b>15.6</b>	<b>43.7</b>
<b>Ours(<math>group=4</math>)</b>	<b>94.1</b>	<b>9.7</b>	<b>23.1</b>

#### 4.2.4. Comparison of Recognition Results for Visualization of Different Kinds of Targets

The performance of the proposed CM-YOLOv8 algorithm in real coal mining comprehensive mining face scenarios is depicted in Figure 6. The figure illustrates the visual results of the algorithm in recognizing targets such as Coal miners, Mine safety helmet, Hydraulic support guard plate, Large Coal, Miners' behaviors, and Towline. It is evident that the model proposed in this paper performs well on the coal mining dataset, successfully identifying targets in coal mine images. The accurate identification boxes and confidence scores enable high-precision localization of target objects.

Compared to traditional YOLO series algorithms, the proposed model benefits from fixed-size bounding boxes, making it more precise in recognizing specific target objects. This avoids the decrease in the model's recognition performance for targets of different scales in the image due to factors such as distance, angle, or image resolution. Thanks to the multi-scale design adopted by YOLOv8 series algorithms, the proposed algorithm can detect targets at different scales

simultaneously. This allows the algorithm to improve its recognition capability for small or low-light targets in dark conditions underground in coal mines.

Moreover, the lightweight YOLOv8 reduces the computational requirements of the model through pruning, enhancing the model's inference speed. This is especially important for scenarios such as coal mining face that require real-time monitoring, allowing for more accurate and timely identification of targets and the adoption of appropriate measures. Additionally, using lightweight models on embedded or mobile devices can reduce power consumption, extending the device's battery life—a critical aspect for coal mine monitoring systems that need to operate for extended periods.

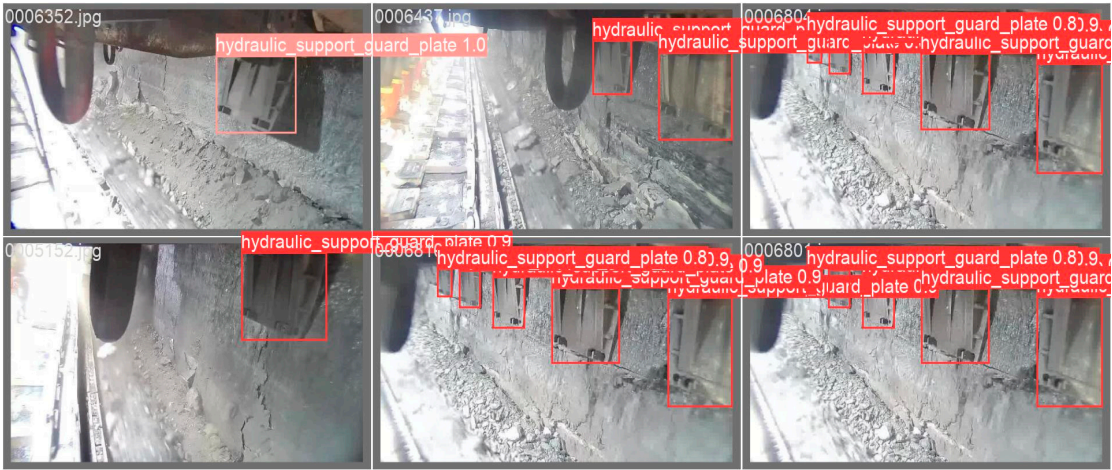


(a) Coal miners

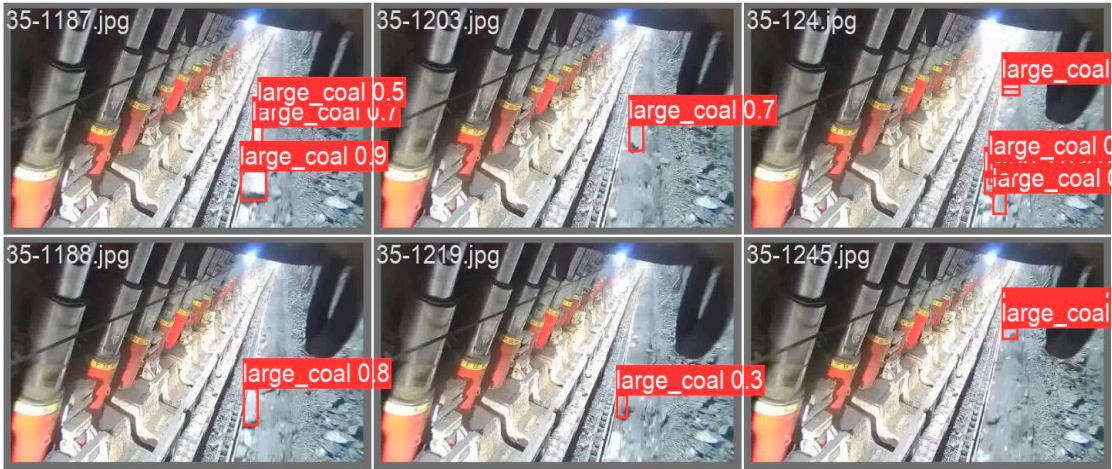


(b) Mine safety helmet





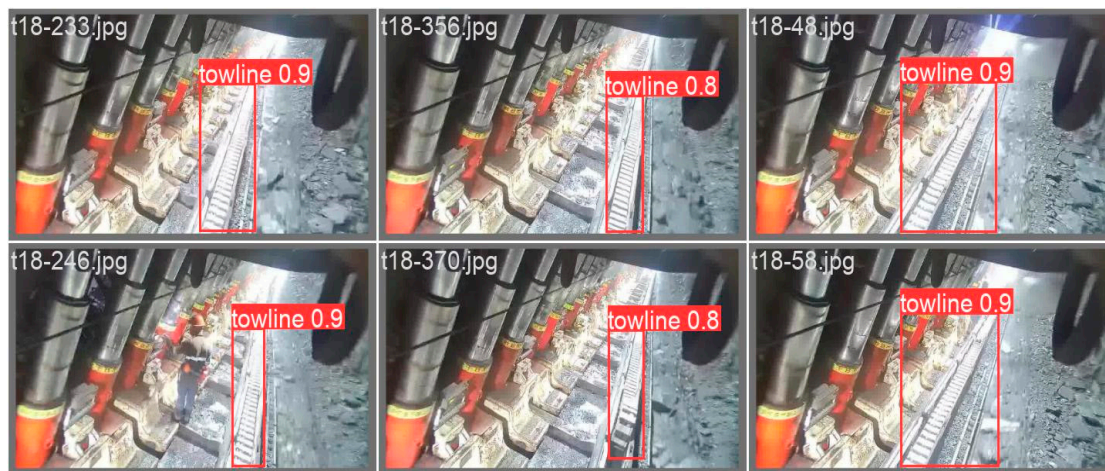
(c) Hydraulic support guard plate



(d) Large Coal



(e) Miners' behaviors



(f) Towline

**Figure 6.** Comparison of the effect of visualization identification of various types of targets in coal mine working face.

## 5. Conclusions

This study addresses the challenges of insufficient image data, high model complexity, and limited computational resources in coal mining face scenarios. It proposes a lightweight object detection algorithm tailored for coal mining face applications. The algorithm enhances detection performance for various targets within a coal mine by generating adaptive predefined anchor boxes suitable for the coal mining face dataset. Additionally, a pruning method based on L1 norm is designed to significantly reduce model computation and parameter complexity while maintaining accuracy. The experimental results of CM-YOLOv8 on the coal mining dataset are compared with other algorithms, demonstrating its efficiency and practicality in coal mining scenarios. However, there is room for further improvement in this research. Due to the strong confidentiality of coal mining datasets, open-source datasets are limited, making large-scale model training challenging. Future work could explore the use of transfer learning techniques, leveraging models trained in other domains to enhance object detection performance in coal mining scenarios. Furthermore, investigating corresponding weakly supervised learning methods may reduce the dependency on labeled data.

**Author Contributions:** Conceptualization, S.M. and Y.F.; methodology, Y.F.; software, Y.F.; validation, Y.F, L.M. and W.Z.; formal analysis, J.K.; investigation, S.M.; resources, M.L.; data curation, Y.F.; writing—original draft preparation, Y.F.; writing—review and editing, S.M.; visualization, Y.F.; supervision, S.M.; project administration, W.Z.; funding acquisition, S.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key R&D Program for the 14th Five-Year Plan (Prevention and Control of Major Natural Disasters and Public Security), grant number SQ2022YFC3000083.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We would like to express our gratitude to many colleagues at Beijing LongRuan Technology Co., Ltd. for their extensive assistance in providing data and hardware support for this experiment.

**Conflicts of Interest:** Not applicable.

## References

1. Oparin, V.N. and A.A. Ordin, *HUBBERT'S THEORY AND THE ULTIMATE COAL PRODUCTION IN TEMRS OF THE KUZNETSK COAL BASIN*. Journal of Mining Science, 2011. 47(2): p. 254-266.



2. Tang, E. and C. Peng, *A macro- and microeconomic analysis of coal production in China*. Resources Policy, 2017. **51**: p. 234-242.
3. Li, C.H., et al., *Life-cycle assessment for coal-based methanol production in China*. Journal of Cleaner Production, 2018. **188**: p. 1004-1017.
4. Gao, Z.X., et al., *Controlling mechanism of coal chemical structure on biological gas production characteristics*. International Journal of Energy Research, 2020. **44**(6): p. 5008-5016.
5. Zhang, Y., et al., *Remarkable Spatial Disparity of Life Cycle Inventory for Coal Production in China*. Environmental Science & Technology, 2023. **57**(41): p. 15443-15453.
6. Ma, T.J. and C.J. Chi, *Spatial configuration and technology strategy of China's green coal-electricity system*. Journal of Renewable and Sustainable Energy, 2012. **4**(3).
7. Zhao, X.L., et al., *Why do electricity utilities cooperate with coal suppliers? A theoretical and empirical analysis from China*. Energy Policy, 2012. **46**: p. 520-529.
8. Wang, W., C. Zhang, and W.Y. Liu, *Impact of Coal-Electricity Integration on China's Power Grid Development Strategy*. in *3rd International Conference on Energy, Environment and Sustainable Development (EESD 2013)*. 2013. Shanghai, PEOPLES R CHINA.
9. Wei, J., H. Chen, and H. Qi, *Who reports low safety commitment levels? An investigation based on Chinese coal miners*. Safety Science, 2015. **80**: p. 178-188.
10. Wu, B., et al., *Development, effectiveness, and deficiency of China's Coal Mine Safety Supervision System*. Resources Policy, 2023. **82**.
11. Li, X.C. and X.Y. Tao, *Study on decision-making system of coal mines environment management base on integrated artificial intelligence method*. in *5th China-Japan International Symposium on Industrial Management*. 2000. Beijing, Peoples R China.
12. Qin, Z., et al. *Research on Key Technologies and System Construction of Smart Mine*. in *5th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*. 2020. Electr Network.
13. Zhang, S.R., M.Y. Zhang, and Ieee, *On Identification of Coal and Rock Images*. in *4th International Symposium on Computer, Consumer and Control (IS3C)*. 2018. Natl Chin Yi Univ Technol, Taichung, TAIWAN.
14. Si, L., et al., *A Novel Coal-Rock Recognition Method for Coal Mining Working Face Based on Laser Point Cloud Data*. Ieee Transactions on Instrumentation and Measurement, 2021. **70**.
15. Wang, D.X., J.X. Ni, and T.Y. Du, *An Image Recognition Method for Coal Gangue Based on ASGS-CWOA and BP Neural Network*. Symmetry-Basel, 2022. **14**(5).
16. Yu, X.C. and X.W. Li, *Sound Recognition Method of Coal Mine Gas and Coal Dust Explosion Based on GoogLeNet*. Entropy, 2023. **25**(3).
17. Zhang, Y.X., Y.R. Tao, and S.H. Li, *An efficient method for recognition of coal/gangue with thermal imaging technique*. International Journal of Coal Preparation and Utilization, 2023. **43**(10): p. 1665-1678.
18. Opiso, E.M., et al., *Synthesis and characterization of coal fly ash and palm oil fuel ash modified artisanal and small-scale gold mine (ASGM) tailings based geopolymer using sugar mill lime sludge as Ca-based activator*. Heliyon, 2021. **7**(4).
19. Singh, S.K., S. Raval, and B. Banerjee, *Roof bolt identification in underground coal mines from 3D point cloud data using local point descriptors and artificial neural network*. International Journal of Remote Sensing, 2021. **42**(1): p. 367-377.
20. Yang, J.J., et al., *Establishment of a Coal Mine Roadway Model Based on Point Cloud Feature Matching*. Mathematical Problems in Engineering, 2022. **2022**.
21. Ding, S.F., et al., *Wavelet twin support vector machines based on glowworm swarm optimization*. Neurocomputing, 2017. **225**: p. 157-163.
22. Breiman, L., *Random forests*. Machine Learning, 2001. **45**(1): p. 5-32.
23. Wen, L., et al., *A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method*. Ieee Transactions on Industrial Electronics, 2018. **65**(7): p. 5990-5998.
24. Long, J., et al. *Fully Convolutional Networks for Semantic Segmentation*. in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015. Boston, MA.
25. Badrinarayanan, V., A. Kendall, and R. Cipolla, *SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation*. Ieee Transactions on Pattern Analysis and Machine Intelligence, 2017. **39**(12): p. 2481-2495.
26. Dai, Y.M., et al., *Attentional Local Contrast Networks for Infrared Small Target Detection*. Ieee Transactions on Geoscience and Remote Sensing, 2021. **59**(11): p. 9813-9824.



27. Girshick, R. and Ieee. *Fast R-CNN*. in *IEEE International Conference on Computer Vision*. 2015. Santiago, CHILE.
28. Redmon, J., et al. *You Only Look Once: Unified, Real-Time Object Detection*. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. Seattle, WA.
29. Liu, W., et al. *SSD: Single Shot MultiBox Detector*. in *14th European Conference on Computer Vision (ECCV)*. 2016. Amsterdam, NETHERLANDS.
30. Lu, G., et al. *Resolution to DM Technology for Coal Mine Safety Data*. in *2nd International Conference on Information Management, Innovation Management and Industrial Engineering*. 2009. Xian, PEOPLES R CHINA.
31. Chen, H., H. Qi, and Q. Feng, *Characteristics of direct causes and human factors in major gas explosion accidents in Chinese coal mines: Case study spanning the years 1980-2010*. *Journal of Loss Prevention in the Process Industries*, 2013. **26**(1): p. 38-44.
32. Schatzel, S.J., et al., *A study of leakage rates through mine seals in underground coal mines*. *International Journal of Mining Reclamation and Environment*, 2016. **30**(2): p. 165-179.
33. Chen, D.X., C. Sun, and L.G. Wang, *Collapse behavior and control of hard roofs in steeply inclined coal seams*. *Bulletin of Engineering Geology and the Environment*, 2021. **80**(2): p. 1489-1505.
34. Wang, K., et al., *Numerical study on the mechanism of air leakage in drainage boreholes: A fully coupled gas-air flow model considering elastic-plastic deformation of coal and its validation*. *Process Safety and Environmental Protection*, 2022. **158**.
35. An, W.P. and M. Li. *The Research of Coal-mining Control Configuration Software's Real-Time Database*. in *3rd International Symposium on Computer Science and Computational Technology (ISCST)*. 2010. Jiaozuo, PEOPLES R CHINA.
36. Agioutantis, Z., et al., *Development of an atmospheric data-management system for underground coal mines*. *Journal of the Southern African Institute of Mining and Metallurgy*, 2014. **114**(12): p. 1059-1063.
37. Sang, P.D. and Iop. *Research and Application of Green Filling Mining Technology for Short Wall Mining in Aging Mine*. in *4th International Workshop on Renewable Energy and Development (IWRED)*. 2020. Electr Network.
38. Bing, Z., et al., *A novel edge computing architecture for intelligent coal mining system*. *Wireless Networks*, 2023. **29**(4): p. 1545-1554.
39. Zhang, L.Y., et al., *Edge Computing Resource Allocation Method for Mining 5G Communication System*. *Ieee Access*, 2023. **11**: p. 49730-49737.
40. Li, Y.T., et al., *A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition*. *Drones*, 2023. **7**(5).
41. Jang, J.G., et al., *Falcon: lightweight and accurate convolution based on depthwise separable convolution*. *Knowledge and Information Systems*, 2023. **65**(5): p. 2225-2249.
42. Jain, A.K., *Data clustering: 50 years beyond K-means*. *Pattern Recognition Letters*, 2010. **31**(8): p. 651-666.
43. Erfani, S., M. Ahmadi, and L. Chen. *The Internet of Things for Smart Homes: An Example*. in *8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*. 2017. Bangkok, THAILAND.
44. Acemoglu, D. and P. Restrepo, *Robots and Jobs: Evidence from US Labor Markets*. *Journal of Political Economy*, 2020. **128**(6): p. 2188-2244.
45. Dong, Z., et al. *Collaborative Autonomous Driving: Vision and Challenges*. in *International Conference on Connected and Autonomous Driving (MetroCAD)*. 2020. Detroit, MI.
46. Sharma, V.K. and R.N. Mir, *A comprehensive and systematic look up into deep learning based object detection techniques: A review*. *Computer Science Review*, 2020. **38**.
47. Ng, P.C. and S. Henikoff, *SIFT: predicting amino acid changes that affect protein function*. *Nucleic Acids Research*, 2003. **31**(13): p. 3812-3814.
48. Wang, X.Y., et al. *An HOG-LBP Human Detector with Partial Occlusion Handling*. in *12th IEEE International Conference on Computer Vision*. 2009. Kyoto, JAPAN.
49. Davarzani, R., S. Mozaffari, and K. Yaghmaie, *Perceptual image hashing using center-symmetric local binary patterns*. *Multimedia Tools and Applications*, 2016. **75**(8): p. 4639-4667.
50. Zhang, Y., R. Jin, and Z.H. Zhou, *Understanding bag-of-words model: a statistical framework*. *International Journal of Machine Learning and Cybernetics*, 2010. **1**(1-4): p. 43-52.
51. Yang, J.C., et al. *Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification*. in *IEEE-Computer-Society Conference on Computer Vision and Pattern Recognition Workshops*. 2009. Miami Beach, FL.
52. Chatfield, K., et al. *The devil is in the details: an evaluation of recent feature encoding methods*. in *22nd British Machine Vision Conference*. 2011. Univ Dundee, Dundee, SCOTLAND.
53. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning*. *Nature*, 2015. **521**(7553): p. 436-444.

54. Tong, K. and Y.Q. Wu, *Deep learning-based detection from the perspective of small or tiny objects: A survey*. Image and Vision Computing, 2022. **123**.
55. Zhang, T.W., X.L. Zhang, and X. Ke, *Quad-FPN: A Novel Quad Feature Pyramid Network for SAR Ship Detection*. Remote Sensing, 2021. **13**(14).
56. Shi, J., et al. *Target Detection Based on Improved Mask Rcnm in Service Robot*. in 38th Chinese Control Conference (CCC). 2019. Guangzhou, PEOPLES R CHINA.
57. Redmon, J., A. Farhadi, and Ieee. *YOLO9000: Better, Faster, Stronger*. in 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2017. Honolulu, HI.
58. Li, P., et al., *Improved YOLOv4-tiny based on attention mechanism for skin detection*. Peerj Computer Science, 2023. **9**.
59. Liao, S.D., et al., *Solder Joint Defect Inspection Method Based on ConvNeXt-YOLOX*. Ieee Transactions on Components Packaging and Manufacturing Technology, 2022. **12**(11): p. 1890-1898.
60. Saydirasulovich, S.N., et al., *A YOLOv6-Based Improved Fire Detection Approach for Smart City Environments*. Sensors, 2023. **23**(6).
61. Wu, D.L., et al., *Detection of *Camellia oleifera* Fruit in Complex Scenes by Using YOLOv7 and Data Augmentation*. Applied Sciences-Basel, 2022. **12**(22).
62. Liu, B., et al. *DAC-PPYOLOE plus : A Lightweight Real-time Detection Model for Early Apple Leaf Pests and Diseases under Complex Background*. in 47th IEEE-Computer-Society Annual International Conference on Computers, Software, and Applications (COMPSAC). 2023. Univ Torino, Torino, ITALY.
63. Hussain, M., *YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection*. Machines, 2023. **11**(7).
64. Yang, X., et al., *A Robust LiDAR SLAM Method for Underground Coal Mine Robot with Degenerated Scene Compensation*. Remote Sensing, 2023. **15**(1).
65. Kirkwood, L., et al. *Challenges in cost analysis of innovative maintenance of distributed high-value assets*. in 3rd International Conference on Through-life Engineering Services (TESConf). 2014. Cranfield Univ, Cranfield, ENGLAND.
66. Pan, H.G., et al., *Fast identification model for coal and gangue based on the improved tiny YOLO v3*. Journal of Real-Time Image Processing, 2022. **19**(3): p. 687-701.
67. Zhang, Y.C., et al., *Research on intelligent detection of coal gangue based on deep learning*. Measurement, 2022. **198**.
68. Fan, J.W., et al., *Macerals particle characteristics analysis of tar-rich coal in northern Shaanxi based on image segmentation models via the U-Net variants and image feature extraction*. Fuel, 2023. **341**.
69. Liang, B., et al., *A Novel Pressure Relief Hole Recognition Method of Drilling Robot Based on SinGAN and Improved Faster R-CNN*. Applied Sciences-Basel, 2023. **13**(1).
70. Lou, H.T., et al., *DC-YOLOv8: Small-Size Object Detection Algorithm Based on Camera Sensor*. Electronics, 2023. **12**(10).
71. He, K.M., et al., *Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition*. Ieee Transactions on Pattern Analysis and Machine Intelligence, 2015. **37**(9): p. 1904-1916.
72. Yang, W.J., et al., *An open dataset for intelligent recognition and classification of abnormal condition in longwall mining*. Scientific Data, 2023. **10**(1).
73. Liu, L.S., et al., *Visual detection on posture transformation characteristics of sows in late gestation based on Libra R-CNN*. Biosystems Engineering, 2022. **223**: p. 219-231.
74. Wang, D.S., et al., *Farmland Obstacle Detection from the Perspective of UAVs Based on Non-local Deformable DETR*. Agriculture-Basel, 2022. **12**(12).
75. Chen, Z., et al., *Vision Transformer Adapter for Dense Predictions*. 2022.
76. Tian, Y.N., et al., *Apple detection during different growth stages in orchards using the improved YOLO-V3 model*. Computers and Electronics in Agriculture, 2019. **157**: p. 417-426.
77. Chen, K., et al. *Hybrid Task Cascade for Instance Segmentation*. in 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. Long Beach, CA.
78. Yang, R.J., et al., *KPE-YOLOv5: An Improved Small Target Detection Algorithm Based on YOLOv5*. Electronics, 2023. **12**(4).
79. Yang, Z.J., et al., *Tea Tree Pest Detection Algorithm Based on Improved Yolov7-Tiny*. Agriculture-Basel, 2023. **13**(5).

80. Sun, H.Y., et al., *Pruning DETR: efficient end-to-end object detection with sparse structured pruning*. Signal Image and Video Processing, 2023.
81. Liu, S., et al., *DAB-DETR: Dynamic Anchor Boxes are Better Queries for DETR*. 2022.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.