



DEYONG LI¹, GUOFA WANG², SHUANG WANG³, WENSHAN WANG⁴, MING DU⁵

Research on Coal Gangue Detection and Recognition Based on Lightweight Network MS-YOLOV3

Introduction

The intellectualization of coal mine is the only way for the high-quality development of coal, and the online detection and separation of coal and gangue are necessary links of coal mine intellectualisation (Wang et al. 2019a; Wang et al. 2019b; Cong et al. 2021; Kłojzy-Karczmarczyk et al. 2016). The rapid and accurate detection and identification of coal gangue is one of the premises and key technologies of the intelligent sorting of coal gangue. In recent years, the traditional manual separation of gangue has been gradually replaced by automatic separation equipment (Zhao et al. 2014; Gupta 2016; Sahu and Dey 2017). At present, the mark-

✉ Corresponding Author: Guofa Wang; e-mail: guofawang@126.com

¹ State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mines, Anhui University of Science and Technology, Huainan 232001, China; e-mail: ldyziyou@126.com

² Collaborative Innovation Center for Mine Intelligent Technology and Equipment, Anhui University of Science and Technology, Huainan 232001, China; ORCID iD: 0000-0002-1772-6049; e-mail: guofawang@126.com

³ China Coal Technology Engineering Group Coal Mining Research Institute, Beijing 100013, China; e-mail: shuangw094@126.com

⁴ China Coal Technology Engineering Group China Coal Research Institute, Beijing 100013, China; e-mail: wwsaust@126.com

⁵ China Coal Technology Engineering Group China Coal Research Institute, Beijing 100013, China; e-mail: dm_pyz@163.com



© 2022. The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-ShareAlike International License (CC BY-SA 4.0, <http://creativecommons.org/licenses/by-sa/4.0/>), which permits use, distribution, and reproduction in any medium, provided that the Article is properly cited.

edly old automatic separation equipment of coal gangue both at home and abroad adopts ray methods, such as X-ray and γ ray (Zhang 2015; Guo et al. 2019; Zhao et al. 2022). However, this kind of equipment has high requirements with regard to system execution speed, radiation risk and environmental protection. Therefore, studying the detection and recognition method of coal gangue is necessary to improve the quality and separation efficiency of gangue, and realize less humanization, high efficiency and the intelligent separation in coal mines. With the advancement of science and technology, the coal and gangue image recognition method has rapidly developed and has gradually become a key research subject in the field of coal and gangue real-time identification (Alfarzaei et al. 2020; Zhang et al. 2020; Li et al. 2022).

At present, the coal gangue image recognition technology based on machine learning mainly uses images to extract the grey scale and texture of coal gangue (McCoy and Auret 2019; Wang et al. 2021). The eigenvalue model was established by different algorithms, and the segmentation parameters were analyzed and optimized to realize the identification of coal and gangue. However, the feature selection and the threshold determination must be manually selected and identified, and problems such as single detection dimension and slow detection speed are observed (Liu et al. 2019; Hou 2017; Wang et al. 2018; He et al. 2022). In recent years, deep learning technology has been gradually applied in coal and gangue identification. Deep learning technology automatically acquires and learns image features through convolutional neural networks (CNN), which can quickly extract and detect feature information in coal and gangue images (Li et al. 2020, 2021; Li 2020). Guo et al. (Guo et al. 2022) proposed an optimization method for the coal gangue depth identification model which combines the idea of transfer learning and structure optimization. The performance of the improved shallow Im Alexnet model has been improved, but the detection speed needs to be further improved. Pu et al. (Pu et al. 2022) used CNN and transfer learning technology to improve the VGG16 (visual geometry group) network to identify the coal gangue. The training accuracy was 100%, but the data accuracy of verification was only 82.5%. Cao Xiangang et al. (Cao et al. 2022) used transfer learning to improve the AlexNet feature extraction network and obtained pixel coordinates and classification information of coal gangue by combining the RPN (region proposal network). However, the detection accuracy is only 90.17% and memory cutting occupies a large proportion. The deep learning coal and gangue image recognition method has considerable prospects for intelligent coal gangue identification (Pan et al. 2022; Lv et al. 2022). The above studies have improved the performance of coal and gangue identification algorithms from different angles. However, the accuracy, efficiency and overall performance of detection must be further improved.

Overall, a lightweight, fast detection method for coal gangue is proposed in this paper to reduce network complexity and improve detection speed. The MobileNetv2 lightweight feature extraction network is used to replace the Darknet-53 main network layer in the YOLOv3 detection algorithm, and SPP is introduced after the backbone network to obtain a fast and accurate lightweight algorithm for the detection of coal gangue MS-YOLOV3. The test results reveal that the method proposed in this paper can effectively detect coal and gangue, has high accuracy and speed, and occupies less memory.

1. Model construction

1.1. Target detection algorithm YOLOv3

The YOLOv3 (Redmon and Farhadi 2018) algorithm is an improvement on the YOLOv2 algorithm, which draws on the idea of ResNet. This algorithm mainly includes network Darknet53 feature extraction, adds shortcut connections, and fuses the feature pyramid network (FPN) (Lin et al. 2017). Compared with the previous algorithm, YOLOv3 improves the detection of small targets. The processing flow of the YOLOv3 algorithm is shown in Figure 1 (Li et al. 2021).

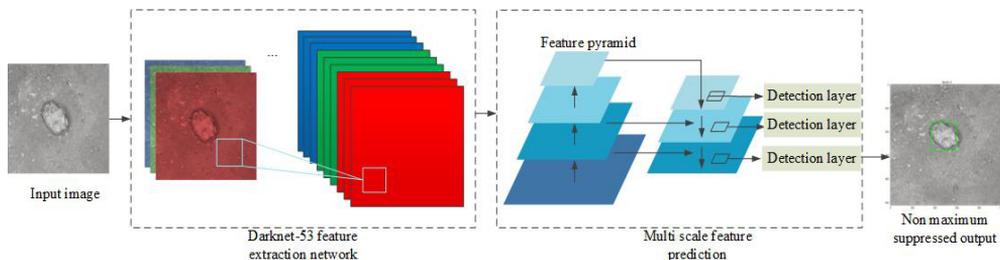


Fig. 1. Processing flow of YOLOv3 algorithm

Rys. 1. Przebieg przetwarzania algorytmu YOLOv3

YOLOv3, an end-to-end network structure, uses a 53-layer convolution layer as the backbone, which is also known as Darknet-53. Extracting the corresponding features from different scales can reduce the computation amount and minimize the loss of low-level features in the pooling layer. L2 regularization is conducted during each convolution, and batch normalization (BN) and the leaky ReLU activation function are performed after the convolution.

1.2. Construction of the feature extraction network MobileNetv2

MobileNetv2 (Sandler et al. 2018) is an improvement of the lightweight convolutional neural network MobileNetv1 (Howard et al. 2017). In contrast to MobileNetv1, MobileNetv2 firstly introduces a different residual structure from ResNet. The residual block of ResNet initially reduces, convolutes and then raises the dimension whilst MobileNetv2 is only the opposite, which is called an inverted residual block. Secondly, a 1×1 convolution is performed before depthwise separable convolution to increase the number of feature graph channels. Finally, ReLU was abandoned after the end of pointwise and replaced with a linear activation function to prevent ReLU from destroying features.

1.2.1. Depthwise separable convolution

In lightweight network structures, depthwise separable convolution (Haase and Amthor 2020) mainly plays a role in reducing network parameters and accelerating operational speed. Depthwise separable convolution uses depth and point-by-point convolution operations instead of the standard convolution operation. F represents the input feature map, G represents the output feature map, and K represents the standard convolution. The convolution formula of standard convolution is (Howard et al. 2017):

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} \cdot F_{k+i-1,l+j-1,m} \quad (1)$$

The convolution formula of depth convolution is (Howard et al. 2017):

$$\hat{G}_{k,l,n} = \sum_{i,j} \hat{K}_{i,j,m,n} \cdot F_{k+i-1,l+j-1,m} \quad (2)$$

where k and L represent the size of the output feature graph, N represents the number of output channels, and M represents the number of input channels.

Figure 2 shows the comparison between depthwise separable and standard convolutions. For standard convolution: Assume that the input layer is a three-channel and 64×64 pixel color image. After a convolution layer containing four filters, four feature maps are then outputted with the same sizes as the input layer. The number of parameters in the convolution layer is computed as $4 \times 3 \times 3 = 108$. In depthwise convolution (filtering), a single filter is used for each input channel. Using the above example, the number of the filters herein is at the same depth as the upper layer. Therefore, the number of parameters is $3 \times 3 \times 3 = 27$.

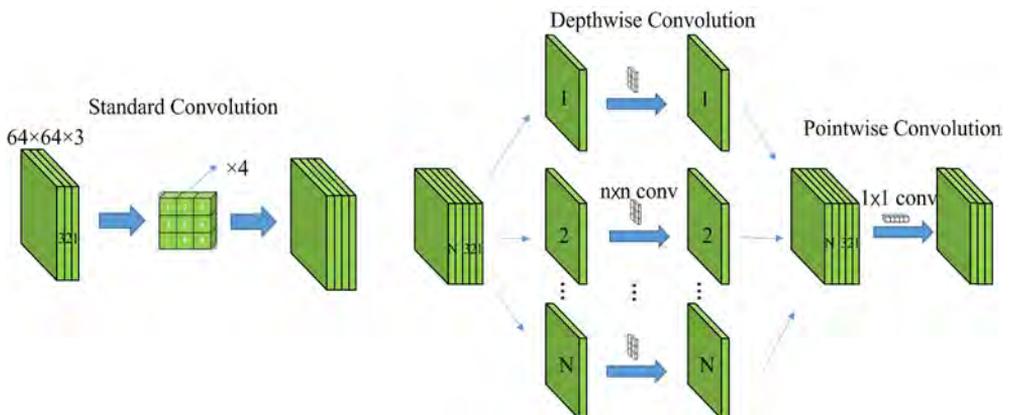


Fig. 2. Comparison between standard convolution and depth separable convolution

Rys. 2. Porównanie splotu standardowego i splotu separowanego na głębokość

Pointwise convolution (combination) is when a set of new outputs is obtained by combining the convolution outputs at different depths with a 1×1 convolution. The convolution kernel is $1 \times 1 \times M$, where M is the upper depth. Multiple feature maps with multiple filters are available, and the number of calculated parameters is $1 \times 1 \times 3 \times 4 = 12$.

Overall, the number of standard convolution parameters is 108. The number of the depth-wise separable convolution parameters is 39, which is approximately one-third of that of the standard convolution. Therefore, depthwise separable convolution can achieve a superior recognition effect by using fewer parameters and operations.

1.2.2. Inverted residual block

As the core of MobileNetV2, the inverted residual block (Sandler et al. 2018) introduces a shortcut structure to realize the reuse of features and reduce the volume and calculation of the model. Referring to the residual block of ResNet, its main function is to convert input samples into high-dimensional data and then send it into deep decomposition convolution for feature extraction. MobileNetV2 is firstly raised (six times) and then reduced after convolution, as shown in Figure 3.

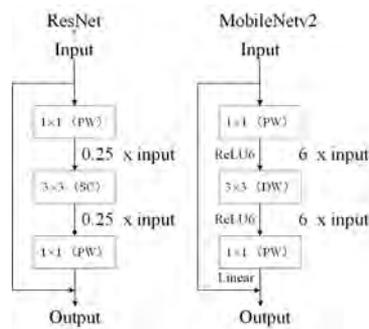


Fig. 3. Comparison of standard residual and inverted residual

Rys. 3. Porównanie reszty standardowej i reszty odwróconej

1.3. Spatial Pyramid Pooling

In standard CNN structures, the convolutional layer is usually connected to the full connection layer behind the convolutional layer and the number of features in the full connection layer is fixed. Therefore, the input image size will be fixed during network input. In training, the input image sizes cannot meet the requirements and need cutting and stretching, causing image distortion. Therefore, He et al. proposed the Spatial Pyramid Pooling (SPP) module (He et al. 2015). The SPP mainly aims to solve the fixed input image size of

CNN; therefore, the aspect ratio and the input image size can be arbitrary. SPP attempts to map image pixels to the receptive field center of feature images through enhanced image processing to improve the receptive field and improve the accuracy of processing results.

The mapping formula in the upper left corner of the picture is:

$$x' = [x/S] + 1 \quad (3)$$

The mapping formula in the lower right corner of the picture is:

$$x' = [x/S] - 1 \quad (4)$$

The SSP also solves the problem of repeated feature extraction in neural networks, which can increase the speed of the waiting box and markedly reduce the cost.

The SPP module, which is used to perform the maximum pooling operation on the local area of the feature maps, is shown in Figure 4 in this paper.

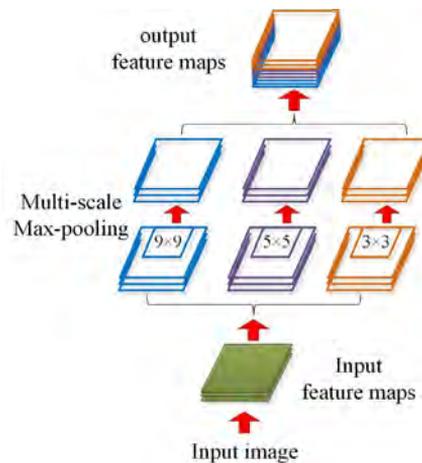


Fig. 4. The SPP structure

Rys. 4. Struktura SPP

2. Network improvement

The backbone of YOLOv3 is Darknet53, which is a complex convolutional network with twenty-three residual units and strong feature extraction capability. However, with the increase in the number of residual units and network channels, the number of network parameters will sharply increase, thus affecting the speed of model detection. The network

is required to achieve rapid detection in the detection of coal and gangue on the conveyor belt; thus, such a complex feature extraction network is necessary. The MobileNet series lightweight network designed by Google for mobile devices is proposed to replace standard convolution with depthwise separable convolution, which can effectively solve the redundancy problem in the number of model parameters. Therefore, this paper changes the feature extraction network of YOLOv3 from Darknet53 to MobileNetv2.

Table 1 shows the selected MobileNetv2 network structure, which comprises sixty-three layers. Layers 39 ($52 \times 52 \times 384$), 54 ($26 \times 26 \times 576$) and 63 ($13 \times 13 \times 160$) were selected as three output layers of the feature extraction network. Figure 5 shows Block1 and Block2 in the MobileNetv2 network structure.

Figure 6 shows the improved MS-YOLOv3 network structure. Take YOLOv3 as the basic network structure. The MobileNetv2 is selected to replace Darknet53 as the feature extrac-

Table 1. MobileNetv2 network structure

Tabela 1. Struktura sieci MobileNetv2

Input size	Type	Output channel number	Stride
$416 \times 416 \times 3$	Conv 3×3	32	2
$208 \times 208 \times 32$	Block1	16	1
$208 \times 208 \times 16$	Block2	24	2
$104 \times 104 \times 24$	Block1	24	1
$104 \times 104 \times 24$	Block2	32	2
$52 \times 52 \times 32$	Block1	32	1
$52 \times 52 \times 32$	Block1	32	1
$52 \times 52 \times 32$	Block1	64	1
$52 \times 52 \times 64$	Block1	64	1
$52 \times 52 \times 64$	Block1	64	1
$52 \times 52 \times 64$	Block1	64	1
$52 \times 52 \times 64$	Block2	96	2
$26 \times 26 \times 96$	Block1	96	1
$26 \times 26 \times 96$	Block1	96	1
$26 \times 26 \times 96$	Block1	96	1
$26 \times 26 \times 96$	Block2	160	2
$13 \times 13 \times 160$	Block1	160	1
$13 \times 13 \times 160$	Block1	160	1

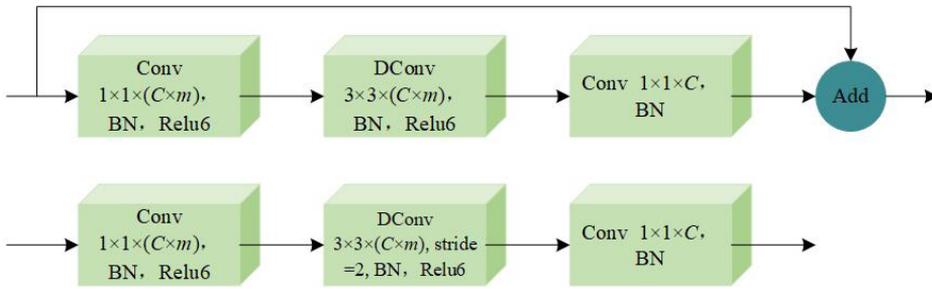


Fig. 5. Block1 and Block2 in MobileNetV2 network structure

Rys. 5. Blok1 i Blok2 w strukturze sieci MobileNetV2

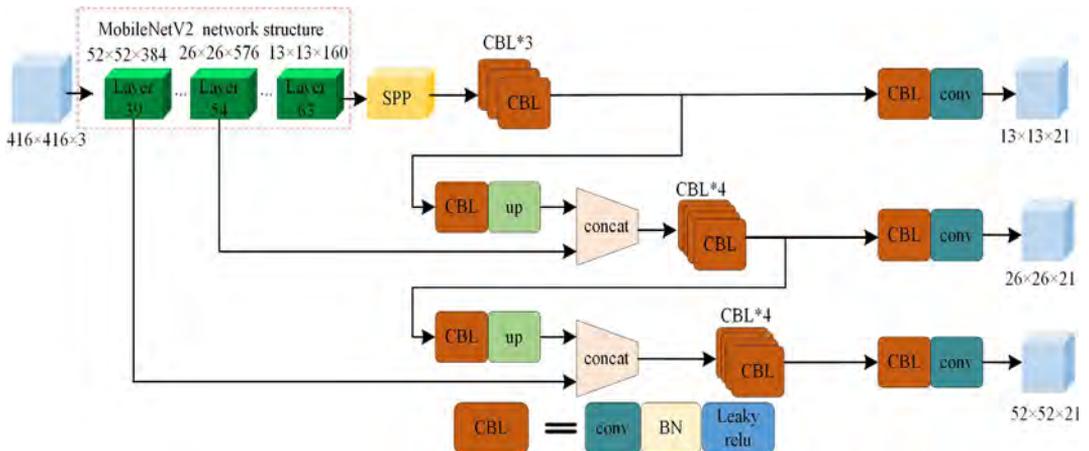


Fig. 6. MS-YOLOv3 network structure

Rys. 6. Struktura sieci MS-YOLOv3

tion network of the algorithm to improve the detection speed of the algorithm. The SPP is added after the backbone network to convert different feature maps into fixed feature maps to improve the positioning accuracy and detection capability of the algorithm. Finally, the network structure is slightly adjusted and the redundant layers are deleted, further reducing the model capacity and the parameter number and improving the model detection speed. The input image is sent to the enhanced feature extraction network to feature aggregation after passing through the MobileNetV2 backbone feature extraction network. The output layer outputs three different scales of prediction anchor frames (YOLO heads).

3. Experiment

3.1. Online detection and identification device for coal gangue

An online recognition method of coal gangue based on MS-YOLOv3 is studied in this paper. The experimental device collects images of coal gangue on the conveyor belt by using a camera set up perpendicularly to the belt. The collected images are then sent to the trained lightweight model for detection and identification. The classification, location information and confidence of the current coal gangue are outputted, and real-time classification and labelling are performed, providing reliable position signals for the final separation of coal gangue. An experimental device was set up in the laboratory according to the proposed method, as shown in Figure 7.



Fig. 7. Coal and gangue identification and location device

Rys. 7. Urządzenie do identyfikacji i lokalizacji węgla i skały płonnej

The experimental device mainly comprises a computer, a camera (Logitech C922), coal and gangue samples, adjustable light sources (2 KM-BRD36030-white lights) and a conveyor belt. The conveyor belt used is a steel wire rope conveyor belt in coal mines, which has a width of 0.7 m, a thickness of 14 mm and a maximum speed of 3.5 m/s. The light controller used was DSC2.0-2C030W-24PS (60 W), which can adjust the brightness of the light source. The experimental steps are shown in Figure 8.

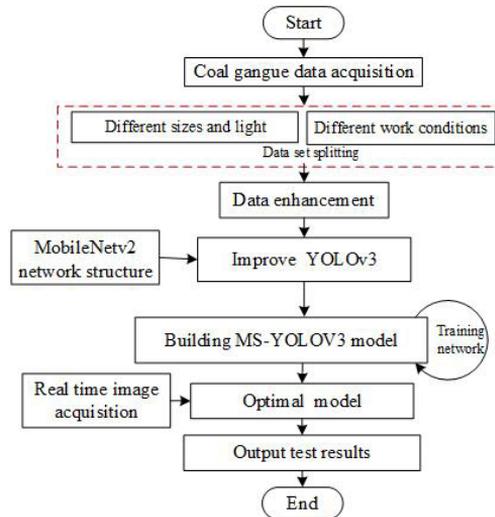


Fig. 8. Experimental steps of the proposed method

Rys. 8. Etapy doświadczalne proponowanej metody

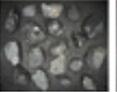
3.2. Multi-condition data collection and processing of coal gangue

Datasets are important in deep learning, and their quality has an important impact on the detection of coal gangue in the actual production environment. The experimental device shown in Fig. 7 was used in this paper to obtain images of coal gangue as a dataset. The experimental data were from Korshintiota Coal Mine, Shenmu City, Yulin City, Shaanxi Province. A total of 500 images of coal and gangue at five light intensities (including 202, 399, 548, 728 and 936 lux) as well as in three sizes (40–60, 70–90 and 100–120 mm) and different working conditions, which include stacking each other and tiling in large areas, were collected in the experiment, as presented in Table 2. Small sample datasets are prone to over-fitting when training in the deep learning network. The mosaic data enhancement was used in this paper to expand the coal and gangue datasets collected under multiple conditions to further improve the environmental robustness of the model.

The dataset was tagged with the labelImg tool to generate a one-for-one label file, which contained all coal and gangue target categories and border location information within each image. The data are pre-processed and normalized, and the pre-processed data are limited to between 0 and 1 by dividing the target marker box data by the height and width of the images, accelerating the data reading of the training process and the model convergence.

Table 2. Coal and gangue samples under different conditions

Tabela 2. Próbki węgla i skały płonnej w różnych warunkach

Different conditions	Light intensity /lux				
	202	399	548	728	936
40–60 mm					
70–0 mm					
100–120 mm					
Stacking each other					
Tiling in large areas					

3.3. Model training and evaluation

3.3.1. Training

The experiment is conducted under Ubuntu 18.04.4 LTS system and uses the Pytorch1.8.1 deep learning framework. Training and verification are performed in this system and the detailed parameters are shown in Table 3.

Table 3. Experimental environment

Tabela 3. Środowisko doświadczalne

Item	Content
GPU	NVIDIA GeForce RTX2060
CPU	Intel(R) Core(TM) i 7 @ 2.90GHz
Language	Python3.7
System	Ubuntu 18.04.4
Drive	CUDA10.1, Cudnn7.3

The training parameters of this paper are shown in Table 4. In each iteration, thirty-two images were loaded and divided into sixteen batches to complete forward propagation (FP). Back propagation was performed to update the parameters after the FP of all thirty-two images was completed. The weight decay regular term (decay) was introduced to prevent the over-fitting phenomenon in the model training. A multi-scale training strategy was adopted to select the image input size from {320,352,384,416,448,480,512,544,576,608} randomly every ten iterations. The model was trained 5020 times, taking a total of one hour.

Table 4. Training parameters

Tabela 4. Parametry uczenia

Parameter	Value
Batch	32
Subdivisions	16
Learning rate	0.001
Iterations	5000
Decay	0.0005
Momentum	0.9
Steps	4000,4500
Scales	0.1,0.1

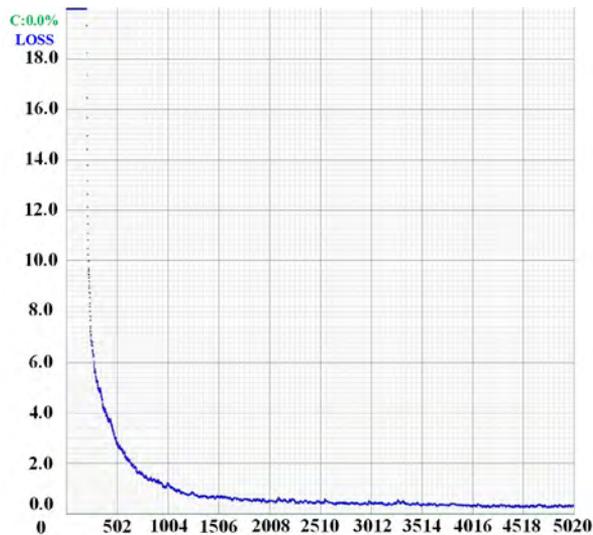


Fig. 9. The loss function curve of MS-YOLOv3 algorithm training

Rys. 9. Krzywa funkcji strat uczenia algorytmu MS-YOLOv3

Figure 9 shows the curve of the training loss function of the MS-YOLOv3 target detection algorithm. The figure reveals that before the 1300 training iterations, the value of the loss function is remarkably large, but the range is substantially small. After 2300 training iterations, the value of the loss function gradually stabilizes, remaining stable at around 0.01. This finding indicates that the algorithm model has been well trained and that the hyperparameter setting of the model is reasonable.

3.3.2 Evaluation

In this experiment, the performance of the model trained by the loss function is evaluated by using precision, recall, mean average precision (mAP), average precision (AP), frame per second (fps) and memory occupied.

The multi-objective mAP is commonly used to evaluate detection accuracy. A high mAP leads to high accuracy, indicating that the currently trained model has superior detection effects. The mAP is defined as:

$$mAP = \frac{1}{n} \sum_n AP_i \quad (5)$$

In Equation, AP represents the area under the composite curve representing recall and precision below the i -th grade, $1 \leq i \leq n$ and the recall and precision are respectively defined as:

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

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

where TP, FN and FP represent the numbers of right detected targets, missed detection targets and false detected targets, respectively.

Formula 8 is the calculation formula of fps

$$fps = \frac{\text{Num-Figure}}{\text{Total-Time}} \quad (8)$$

where Num-Figure is the total quantity of detected images and Total-Time is the total detection time.

4. Result analysis

4.1. Analysis of MS-YOLOv3 experimental results

The evaluation index value of the algorithm is obtained through the training of the improved MS-YOLOv3 algorithm, as shown in Table 5. The table reveals that the recognition accuracy of the improved model for coal and gangue is more than 99%, the detection speed is 139 fps, and the memory occupation is only 9.2 M.

Table 5. Evaluation index values of MS-Yolov3 model of the improved algorithm

Tabela 5. Wartości indeksu oceny modelu MS-Yolov3 udoskonalonego algorytmu

Model	AP/(%coal)	AP/(%gangue)	mAP/%	fps	Memory size/M
MS-YOLOV3	99.10	99.06	99.08	139	9.2

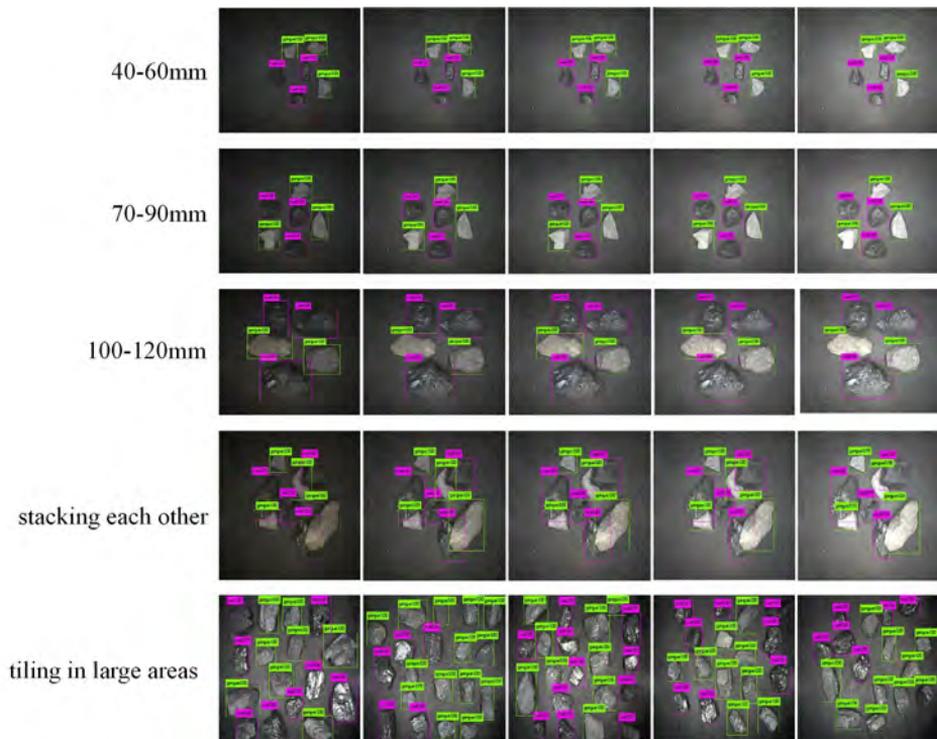


Fig. 10. Detection and identification results of coal gangue under different conditions

Rys. 10. Wyniki detekcji i identyfikacji skały płonnej w różnych warunkach

Using the experimental equipment of the recognition and location of coal gangue, the experiment of coal gangue identification and detection is conducted on the trained improved MS-YOLOV3 algorithm network model. Figure 10 shows the detection and identification results of coal gangue under different conditions. A total of coal and gangue samples with diameters of 40–60, 70–90 and 100–120 mm and stacking each other and large area tiling are tested. The light intensity from left to right is 202, 399, 548, 728 and 936 lux. The figure shows that the MS-YOLOV3 can detect and identify coal and gangue quickly and accurately under different light conditions in the case of large, medium and small sizes stacked with each other. This finding verifies the effectiveness of the algorithm and maintains high robustness.

4.2. Ablation experiment

Ablation experiments are conducted in this paper to verify the effectiveness of each improved part. Table 6 shows the results of the ablation experiment, which reveals that the improvement strategy of each module in YOLOv3 is helpful in improving detection performance.

Table 6. Results of ablation experiments of MS-YOLOv3

Tabela 6. Wyniki doświadczeń ablacyjnych MS-YOLOv3

Model	Improvements	AP/%		Memory size/M	mAP/%	fps
		Coal	Gangue			
A	YOLOv3	98.90	246.3	246.3	99.15	61
B	Model A+Mobilev2	98.34	9.0	9.0	98.29	137
C	Model B + SPP	99.42	9.6	9.6	99.12	134
D	Model C + redundancy elimination	99.10	9.2	9.2	99.08	139

In Table 6, Model A is YOLOv3, Model B is Model A lightweight feature extraction network Mobilev2 added, Model C is Model B with added SPP module, and Model D is a redundant model based on Model C.

Model A→ Model B: YOLOV3 feature extraction network Darknet53 replaced by Mobilev2. Experimental results showed that mAP decreased to 98.29%, but FPS reached 137 and improved by 124.6% and memory dropped to 9 M.

Model B→ Model C: the SPP is added after the backbone network to convert different feature maps into fixed feature maps. The experimental results show that the memory increased by 0.6 M whilst the mAP increased from 98.29% to 99.12%.

Model C→ Model D: the network structure is slightly adjusted, and redundant layers are deleted, further reducing the model capacity and the parameter number and improving model detection speed. With the accuracy remaining the same, the fps was 139, and the memory was reduced to 9.2 M.

4.3. Comparative experiment of different algorithms

The dataset constructed in this paper is trained and tested by the YOLOv3, the YOLOv3-Tiny, the YOLOv4 (Bochkovskiy et al. 2020) and the YOLOv4-Tiny and results are compared with MS-YOLOV3. The results of the detection performance comparison are shown in Figure 11. The figure reveals that the mAP of the proposed MS-YOLOv3 is 99.08%, which is 3.28% and 1.33% higher than that of YOLOv3-Tiny and YOLOv4-Tiny, respectively, and almost the same as that of YOLOv3 and YOLOv4. In addition, the detection speed of the MS-YOLOv3 algorithm is 139 fps, which is higher than 78, 7, 73 and 1 fps compared with the YOLOv3, the YOLOv3-Tiny, the YOLOv4 and the YOLOv4-Tiny, respectively. The memory size of the MS-YOLOv3 is 9.2 M, which is 237.1, 25.5, 247 and 14.3 M less than the YOLOv3, the YOLOv3-Tiny, the YOLOv4 and the YOLOv4-Tiny, respectively.

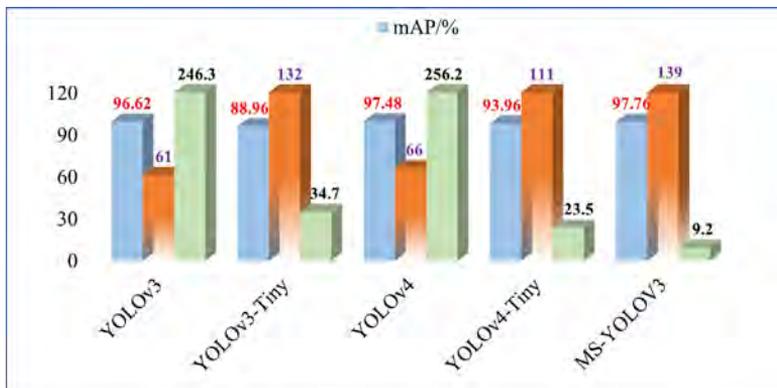


Fig. 11. Experimental results of different detection algorithms

Rys. 11. Wyniki doświadczalne różnych algorytmów detekcji

Conclusion

Focusing on the problems in the current deep learning coal gangue detection and recognition algorithms, such as large model memory and slow detection speed, a rapid detection method for lightweight coal gangue is proposed. YOLOv3 is taken and improved as the

basic structure. The MobileNetv2 lightweight feature extraction network is selected to replace Darknet53 as the main network of the detection algorithm. The SPP module is added after the backbone network, thereby obtaining the MS-YOLOV3 lightweight network. The experimental equipment was set up, and multi-condition coal and gangue data sets were constructed. The model was trained, and the identification and positioning results of the model were tested under different sizes, illumination intensities and various working conditions and compared with other algorithms. The following conclusions can be drawn.

1. The proposed MS-YOLOV3 algorithm can detect coal gangue quickly and accurately, with an mAP of 99.08%, a speed of 139 fps and a memory occupation of only 9.2 M.
2. In addition, the MS-YOLOV3 can effectively detect different lights, different sizes, mutual stacking and multiple quantities of coal and gangue with high confidence and certain environmental robustness and practicability.
3. Compared with YOLOv3, the performance of the proposed algorithm is significantly improved. Under the premise that the accuracy is unchanged, the FPS increases by 127.9% and the memory decreases by 96.2%.

The results of this paper will provide theoretical reference and technical support for coal and gangue separation. The next step will be to improve the algorithm according to the field environment, such as dust interference, to ensure the effectiveness of the algorithm.

This research was funded by Anhui University Graduate scientific research project (No. YJS20210371), the National Natural Science Fund Project (No. 51974004), Collaborative Innovation Project of Collaborative Tackling of Universities in Anhui Province (No. GXXT-2020-060), China Postdoctoral Science Foundation (No. 2019M662133), Open Foundation of State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mine (No. SKLMRDPC20KF10), Collaborative Innovation Project of Collaborative Tackling of Universities in Anhui Province (No. GXXT-2020-054). Science and Technology Innovation Fund of Coal Mining and Designing Department of Tiandi Science and Technology Co., Ltd (KJ-2021-KCMS-04).

REFERENCES

- Alfarzaei et al. 2020 – Alfarzaei, M., Niu, Q., Zhao, J., Eshaq, R. and Hu, E. 2020. Coal/Gangue Recognition Using Convolutional Neural Networks and Thermal Images. *IEEE Access* 8, pp. 76780–76789.
- Bochkovskiy et al. 2020 – Bochkovskiy, A., Wang, C. and Liao, H. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv preprint*, arXiv:2004.10934.
- Cao et al. 2022 – Cao, X., Liu, S., Wang, P., Xu, G. and Wu, D. 2022. Research on coal gangue identification and positioning system based on coal-gangue sorting robot. *Coal Science and Technology* 50(1), pp. 237–246 (in Chinese).
- Cong et al. 2021 – Cong, L. and Xingru, W. 2021. *Research and Practice of China's Intelligent Coal Mines*. [In:] *In 2021 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI)*, pp. 391–399.
- Guo et al. 2019 – Guo, Y., Yu Z. and Lu, Y. 2019. Research on photoelectric intelligent separation technology of coal and gangue based on NP-FSVM with the PSO algorithm. *Coal Science and Technology* 47(4), pp. 13–19 (in Chinese).

- Guo et al. 2022 – Guo, Y., Wang, X., He, L. and Liu P. 2022. Research on Coal and Gangue Recognition Method Based on TW-RN Optimized CNN. *Coal Science and Technology* 50(1), pp. 228–236 (in Chinese).
- Gupta, N. 2016. Evaluation of pneumatic inclined deck separator for high-ash Indian coals. *International Journal of Coal Science – Technology* 3(2), pp. 198–205, DOI: 10.1007/s40789-016-0125-2.
- Haase et al. 2020 – Haase, D. and Amthor, M. 2020. Rethinking Depthwise Separable Convolutions: How Intra-Kernel Correlations Lead to Improved MobileNets. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- He et al. 2015 – He, K., Zhang, X., Ren, S. and Sun, J. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37(9), pp.1904–1916.
- He et al. 2022 – He, L., Wang, S., Guo, Y., Cheng, G., Hu, K., Zhao, Y. and Wang, X. 2022. Multi-scale coal and gangue dual-energy X-ray image concave point detection and segmentation algorithm. *Measurement* 196, DOI: 10.1016/j.measurement.2022.111041.
- Hou, W. 2017. Identification of coal and gangue by feed-forward neural network based on data analysis. *International Journal of Coal Preparation and Utilization* 39(1), pp. 33–43, DOI: 10.1080/19392699.2017.1290609.
- Howard et al. 2017 – Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, DOI: 10.48550/arXiv.1704.04861.
- Klojzy-Karczmarczyk et al.2016 – Klojzy-Karczmarczyk, B., Mazurek, J., and Paw, K. 2016. Possibilities of Utilization of Aggregates and Extractive Waste from hard Coal Mining at Janina Mine in the Process of Reclamation of Open-pit Mines. *Gospodarka Surowcami Mineralnymi – Mineral Resources Management* 32(3), pp. 111–133, DOI: 10.1515/gospo-2016-0030.
- Li et al. 2020 – Li, M., Duan, Y., Cao, X., Liu, C., Sun, K. and Liu, H. 2019. Image recognition method and system of coal gangue sorting robot. *Journal of China Coal Society* 46(9), pp. 1305–1315, DOI: 10.1080/19392699.2020.1760855 (in Chinese).
- Li, Y. 2020. Research on coal gangue detection based on deep learning. *Xi'an University of Science and Technology (in Chinese)*.
- Li et al. 2021 – Li, D., Wang, G., Zhang, Y. and Wang, S.2021. Coal gangue detection and recognition algorithm based on deformable convolution YOLOv3. *IET Image Process* 16(1), pp. 134–144, DOI: 10.1049/ipr2.12339.
- Li et al. 2022 – Li, D., Wang, G., Guo, Y., Zhang, Y. and Wang, S.2022. An identification and positioning method for coal gangue based on lightweight mixed domain attention. *International Journal of Coal Preparation and Utilization*, DOI: 10.1080/19392699.2022.2119561.
- Lin et al. 2017 – Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. 2017. Feature Pyramid Networks for Object Detection. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society.
- Liu et al. 2019 – Liu, C., Li, M., Zhang, Y., Han, S., and Zhu, Y. 2019. An Enhanced Rock Mineral Recognition Method Integrating a Deep Learning Model and Clustering Algorithm. *Minerals* 9(9), DOI: 10.3390/min9090516.
- Lv et al. 2022 – Lv, Z., Wang, W., Zhang, K., Li, W., Feng, J. and Xu, Z.2022. A synchronous detection-segmentation method for oversized gangue on a coal preparation plant based on multi-task learning. *Minerals Engineering* 187, DOI: 10.1016/j.mineng.2022.107806.
- Mccooy, J. and Auret, L. 2019. Machine learning applications in minerals processing: A review. *Minerals Engineering* 132, pp. 95–109, DOI: 10.1016/j.mineng.2018.12.004.
- Pan et al. 2022 – Pan, H., Shi, Y., Lei, X., Wang, Z. and Xin, F., 2022. Fast identification model for coal and gangue based on the improved tiny YOLOv3. *J. Real-Time Image Process* 19(3), pp. 687–701, DOI: 10.1007/s11554-022-01215-1.
- Pu et al. 2019 – Pu, Y., Apel, D., Szmigiel, A. and Chen, J. 2019. Image Recognition of Coal and Coal Gangue Using a Convolutional Neural Network and Transfer Learning. *Energies* 12(9), DOI: 10.3390/en12091735.
- Redmon, J. and Farhadi, A. 2018. YOLOv3: An Incremental Improvement. *arXiv e-prints*. DOI: 10.48550/arXiv.1804.02767.
- Sahu et al. 2017 – Sahu, L. and Dey, S. 2017. Enrichment of carbon recovery of high ash coal fines using air fluidized vibratory deck separator. *International Journal of Coal Science – Technology* 4(3), pp. 262–273, DOI: 10.1007/s40789-017-0172-3.

- Sandler et al. 2018 – Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L. 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Wang et al. 2018 – Wang, J., Li, L. and Yang, S. 2018. Experimental study on gray and texture features extraction of coal and gangue image under different illuminance. *Journal of China Coal Society* 43(11), pp. 3051–3061, DOI: 10.13225/j.cnki.jccs.2018.0866 (in Chinese).
- Wang et al. 2019a – Wang, G., Liu, F., Meng, X., Fan, J., Wu, Q., Ren, H., Pang, Y., Xu, Y., Zhao, G., Zhang, D., Cao, X., Du, Y., Zhang, J., Chen, H., Ma, Y. and Zhang, K. 2019a. Research and practice on intelligent coal mine construction (primary stage). *Coal Science and Technology* 47(8), pp. 1–36 (in Chinese).
- Wang et al. 2019b – Wang, G., Liu, F., Pang, Y., Ren, H. and Ma, Y. 2019b. Coal mine intellectualization: The core technology of high quality development. *Journal of China Coal Society* 44(2), pp. 349–357 (in Chinese).
- Wang et al. 2021 – Wang, X., Wang, S., Guo, Y., Hu, K. and Wang, W. 2021. Dielectric and geometric feature extraction and recognition method of coal and gangue based on VMD-SVM. *Powder Technology* 392, pp. 241–250, DOI: 10.1016/j.powtec.2021.06.057.
- Zhang, N. 2015. *Detection and radiation law of natural gamma ray from coal and roof-rock in the fully mechanized top coal caving mining*. Xuzhou: China University of Mining – Technology (in Chinese).
- Zhang et al. 2020 – Zhang, Z., Liu, Y., Hu, Q., Zhang, Z., Wang, L., Liu, X. and Xia, X. 2020. Multi-information online detection of coal quality based on machine vision. *Powder Technology* 374, pp. 250–262, DOI: 10.1016/j.powtec.2020.07.040.
- Zhao et al. 2014 – Zhao, Y., Yang, X., Luo, Z., Duan, Z. and Song, S. 2014. Progress in developments of dry coal beneficiation. *International Journal of Coal Science – Technology* 1(1), pp. 103–112, DOI: 10.1007/s40789-014-0014-5.
- Zhao et al. 2022 – Zhao, Y., Wang, S., Guo, Y., Cheng, G., He, L., and Wang, W. 2022. The identification of coal and gangue and the prediction of the degree of coal metamorphism based on the EDXRD principle and the PSO-SVM model. *Gospodarka Surowcami Mineralnymi – Mineral Resources Management* 38(2), pp. 113–129, DOI: 10.24425/gsm.2022.141663.

RESEARCH ON COAL GANGUE DETECTION AND RECOGNITION BASED ON LIGHTWEIGHT NETWORK MS-YOLOV3

Keywords

detection and recognition, coal and gangue, MobileNetv2, SPP, YOLOv3

Abstract

The rapid and accurate detection and identification of coal gangue is one of the premises and key technologies of the intelligent separation of coal gangue, which is of considerable importance for the separation of coal gangue. Focusing on the problems in the current deep learning algorithms for the detection and recognition of coal gangue, such as large model memory and slow detection speed, a rapid detection method for lightweight coal gangue is proposed. YOLOv3 is taken as the basic structure and improved. The MobileNetv2 lightweight feature extraction network is selected to replace Darknet53 as the main network of the detection algorithm to improve the detection speed. Spatial pyramid pooling (SPP) is added after the backbone network to convert different feature maps into fixed feature maps in order to improve the positioning accuracy and detection capability of the algorithm, thereby obtaining the lightweight network MS-YOLOV3. The experimental equipment was set up and

multi-condition coal and gangue datasets were constructed. The model was trained and the identification and positioning results of the model were tested under different sizes, illumination intensities and various working conditions, and compared with other algorithms. Experimental results show that the proposed algorithm can detect the coal gangue quickly and accurately, with an mAP of 99.08%, a speed of 139 fps and a memory occupation of only 9.2 M. In addition, the algorithm can effectively detect mutually stacking coal and gangue of different quantities and sizes under different lights with high confidence and with a certain degree of environmental robustness and practicability. Compared with the YOLOv3, the performance of the proposed algorithm is significantly improved. Under the premise that the accuracy is unchanged, the FPS increases by 127.9% and the memory decreases by 96.2%. Therefore, the MS-YOLOv3 algorithm has the advantages of small memory, high accuracy and fast speed, which can provide online technical support for the detection and identification of coal and gangue.

BADANIA NAD WYKRYWANIEM I ROZPOZNAWANIEM SKAŁY PŁONNEJ W OPARCIU O LEKKĄ SIEĆ MS-YOLOV3

Słowa kluczowe

wykrywanie i rozpoznawanie, węgiel i skała płonna, MobileNetv2, SPP, YOLOv3

Streszczenie

Szybkie i dokładne wykrywanie oraz identyfikacja skały płonnej jest jedną z przesłanek i kluczowych technologii inteligentnej separacji skały płonnej. Koncentrując się na problemach związanych z obecnymi algorytmami wykrywania i rozpoznawania skały płonnej z głębokim uczeniem, takimi jak duża pamięć modelu i niska prędkość wykrywania, zaproponowano metodę szybkiego wykrywania lekkiej skały płonnej. YOLOv3 jest traktowany jako struktura podstawowa i ulepszony. Lekka sieć ekstrakcji funkcji Mobilenetv2 została wybrana w celu zastąpienia Darknet53 jako głównej sieci algorytmu wykrywania w celu poprawy szybkości wykrywania. Spatial Pyramid Pooling (SPP) jest dodawany po sieci szkieletowej w celu konwersji różnych map obiektów na mapy stałych funkcji, aby poprawić dokładność pozycjonowania i zdolność wykrywania algorytmu, uzyskując w ten sposób lekką sieć MS-YOLOV3. Ustawiono sprzęt eksperymentalny i skonstruowano wielowarunkowe zbiory danych dotyczące węgla i skały płonnej. Model został przeszkolony, a wyniki identyfikacji i pozycjonowania modelu zostały przetestowane przy różnych rozmiarach, natężeniu oświetlenia i różnych warunkach pracy oraz porównane z innymi algorytmami. Wyniki eksperymentu pokazują, że zaproponowany algorytm jest w stanie szybko i dokładnie wykryć skałę węglową, z mAP na poziomie 99,08%, szybkością 139 fps i zajęciem pamięci zaledwie 9,2 MB. Ponadto może skutecznie wykrywać różne światła, różne rozmiary, wzajemne układanie w stosy oraz wielokrotną ilość węgla i skały płonnej, z dużą pewnością i pewną odpornością środowiskową i wykonalnością. W porównaniu z YOLOv3 wydajność proponowanego algorytmu jest znacznie lepsza. Przy założeniu, że dokładność pozostaje w zasadzie niezmienną, FPS wzrasta o 127,9%, a pamięć spada o 96,2%. Dlatego algorytm MS-YOLOv3 ma zalety małej pamięci, wysokiej dokładności i dużej szybkości, co może zapewnić wsparcie techniczne dla wykrywania i identyfikacji węgla i skały płonnej online.