Research on Engineering Structures & Materials
www.jresm.org
Research Group

Research Article

Mining control method for small coal pillar working faces based on the integration of improved DeepLabV3+ and digital twin technology

Jianjie Wang [1,a], Yue Dong [1,b], Qiming Wei [1,c], Yajun Hu [1,d], Yude Liu [2,e], Nan Li [*,3,f], Yingfu Ma [1,g], Xiaohong Han [1,h]

[1]*Shicaocun Coal Mine, Ningxia Coal Industry Co., Ltd., CHN Energy, Lingwu 750411, China*
[2]*School of Mine Safety, University of Emergency Management, Sanhe 065201, China*
[3]*School of Safety Engineering, University of Emergency Management, Sanhe 065201, China*

Article Info

Article History:

Received 07 Sep 2026
Accepted 08 Oct 2026

Keywords:

DeepLabV3+;
Digital twin;
Small coal pillar;
Mining control;
Convolutional block attention

Abstract

To address the problems of low accuracy in coal-rock interface identification, weak anti-interference ability, and insufficient intelligence in mining control under challenging underground coal mine conditions, a mining control method for small coal pillar working faces integrating improved DeepLabV3+ and digital twin technology is proposed. For coal-rock identification, an improved DeepLabV3+ network is utilized, combined with the convolutional block attention module (CBAM) to enhance feature expression. Meanwhile, the Canny edge detection operator is fused to strengthen the extraction of coal-rock boundary information. A virtual interactive control system for the working face based on digital twin is constructed. Simulation results show that, compared with traditional mainstream algorithms, the proposed method achieves the best recognition performance among the compared methods across the evaluated metrics: accuracy of 98.73%, F1-score of 98.32%, and AUC of 0.996. In addition, the FPS reaches 102.8 and the inference time is reduced to 9.73 ms, demonstrating better real-time performance and a lightweight advantage. Therefore, in the simulation experiments, the proposed method improves coal-rock interface identification accuracy and virtual mining control performance.

© 2026 MIM Research Group. All rights reserved.

1. Introduction

Small coal pillar mining faces involve special working conditions. Under these conditions, mining control methods mainly focus on actively controlling the cutting trajectory of the coal mining machine. Traditional coal mining machines rely on memory cutting and belong to open-loop or semi-open-loop control, which cannot adapt to the sudden changes in coal seam thickness and rock properties under small coal pillar conditions. The adaptive cutting technology is the key to solving these problems, and its implementation must rely on a pre-sensing link - the coal mining machine needs to know the exact position of the Coal-Rock (CR) interface in real time. Otherwise, all adjustments are "blind adjustments". In the small coal pillar scenario, a centimeter-level deviation in interface positioning may lead to roof damage or resource loss. Therefore, CR identification is not an "add-on module" of the control method, but the pre-sensing layer for the closed-loop operation of the cutting control. Currently, CR interface identification mainly relies on manual experience or a single sensor, which has problems of low efficiency, insufficient real-time performance, and being susceptible to environmental interference such as dust, illumination, and

*Corresponding author: linanzn@126.com
[a]orcid.org/0009-0001-1332-4106; [b]orcid.org/0009-0009-7104-7862; [c]orcid.org/0009-0009-4345-5092; [d]orcid.org/0009-0008-5762-6879; [e]orcid.org/0009-0004-4421-2897; [f]orcid.org/0009-0006-7845-934X; [g]orcid.org/0009-0005-6987-1984; [h]orcid.org/0009-0004-0986-6107
DOI: http://dx.doi.org/10.17515/resm2026-2312am0907rs

vibration [1,2]. Deep learning has been widely employed within the realm of CR identification, and convolutional neural networks can automatically extract CR texture and edge features, which has higher recognition accuracy compared to traditional machine learning methods. However, attributed to the challenging underground conditions, the blurred CR boundaries, drastic changes in illumination, and dust obstruction, factors still cause problems such as inaccurate boundary positioning, insufficient Feature Extraction (FE), and poor robustness in complex scenarios in the model [3-5].

Most studies mainly focus on improving the recognition accuracy, while there is relatively little research on the collaborative application of CR recognition outcomes and the control system of the coal mining machine. This makes it difficult to fully utilize the practical value of intelligent recognition technology in intelligent mining. Zhang H et al. [6] suggested a linear filling Transformer network, by introducing the linear multi-head self-Attention Mechanism (AM) to improve the global FE ability of the model, achieving high-precision semantic segmentation of Coal Gangue (CG) images. Experimental outcomes implied that this model outperformed traditional convolutional neural networks in both accuracy and speed, but the computational complexity of the Transformer structure was relatively high, and there were still certain limitations in real-time deployment in underground Coal Mines (CM). Li X et al. [7] introduced a CG detection method grounded in DeepLab V3 Plus (DeepLabV3+) due to the dense distribution of CG particles and the large differences in particle size. This method utilized dilated convolution and encoder-decoder structure to achieve fine segmentation of complex targets. The experimental outcomes implied that the model could improve the accuracy of CG detection. However, this method was mainly for CG target detection and the robustness research for complex CR boundaries and underground complex environments was still relatively insufficient. Zhang Y et al. [8] combined multispectral imaging technology and introduced a CG intelligent detection method grounded in Real-Time Detection Transformer (RT-DETR). The research outcomes implied that multispectral information could effectively improve the expression ability of target features and improve the recognition accuracy and detection stability of CG in complex backgrounds. However, this method relied on multispectral sensors for data collection, had a high system cost, and had strict requirements for on-site hardware, and there were certain limitations in practical engineering applications. Wang H et al. [9] addressed the problem of high noise in laser point cloud data caused by the underground dust environment, proposing a Deep Learning (DL) model integrating graph neural networks, achieving high-precision semantic segmentation of CM laser point cloud data. Experimental outcomes implied that this method could effectively improve the segmentation accuracy and environmental perception ability of point cloud data in dusty environments. However, this study mainly focused on three-dimensional point cloud data and there was still relatively little research on semantic segmentation of two-dimensional CR images and CR recognition tasks for coal mining machines.

Digital twin technology, by establishing a real-time mapping relationship between the physical working face and the virtual working face, can achieve equipment operation status monitoring, production process simulation analysis, and control strategy optimization, providing reliable data support and decision-making basis for intelligent cutting of coal mining machines. Grounded in this, an innovative method for intelligent mining control of small coal pillar working faces, integrating improved DeepLabV3+ and digital twin technology, is introduced. In terms of CR identification, using DeepLabV3+ as the base network, the Convolutional Block Attention Module (CBAM) AM is suggested to improve the model, and combined with Canny edge detection to extract CR boundary information. A digital twin working face virtual interaction platform is constructed, combined with the fuzzy Proportional-Integral-Differential (PID) control strategy, to achieve real-time monitoring, dynamic adjustment, and intelligent control of the operation status of the coal mining machine. The research aims to explore a simulation-validated technical solution with potential. The high stress in the small coal pillar mining area can easily lead to the fragmentation of coal rock mass, the intensification of debris and dust, resulting in the fragmentation of coal rock interface texture, decreased visual contrast, and blurred boundaries. Narrow spaces limit the installation distance, viewing angle, and number of cameras, resulting in limited field of view, severe occlusion, and difficult calibration. Therefore, sensors need to be miniaturized, explosion-proof, wide-angle, anti vibration, dust-proof, and supplemented with active lighting. Rapid deformation causes the

position and shape of the coal rock interface to change rapidly over time, requiring visual sensors to have high frame rates, low latency, and high dynamic response capabilities. The control cycle needs to match the deformation rate, and if necessary, multimodal perception such as visible light, infrared, laser, or vibration can be introduced.

The innovation of the research lies in the construction of an integrated technology link for small coal pillar working faces, rather than a simple superposition of existing algorithms and control technologies. The study introduces CBAM to enhance key channels and spatial features based on DeepLabV3+ and integrates Canny edge information to enhance the structural features of coal rock interfaces. On this basis, the coal rock interface recognition results are further connected to the control link of the coal mining machine, and fuzzy PID is introduced to dynamically adjust the PID parameters based on the control deviation and its change rate. Then, the study realizes the extension from simple visual recognition to dynamic control of the working face. At the same time, a digital twin virtual interactive environment is constructed through TIA Portal V16, S7-PLCSIM Advanced, and NX MCD to achieve linkage verification between control programs, virtual workspace models, and device status feedback. Therefore, compared to existing independent research mainly focused on coal rock recognition accuracy or digital twin control, the specific contribution of this study lies in establishing a complete technical chain from coal rock interface visual perception, adaptive adjustment of control parameters to digital twin virtual interactive verification.

2. Methods

2.1 CR Recognition Method Based on Improved DeepLabV3+

In the traditional CR boundary recognition process, there are common problems such as fuzzy boundaries, high texture similarity, strong interference from light and dust, and significant scale changes. However, traditional CR recognition methods mainly rely on machine learning methods grounded in threshold segmentation, edge detection, or manual features, which are difficult to achieve stable and accurate boundary extraction. For this purpose, DeepLabV3+ is utilized to achieve boundary recognition and analysis of CR. DeepLabV3+ utilizes dilated convolution and dilated spatial pyramid pooling structure to expand the receptive field without reducing feature resolution, achieving multi-scale fusion expression of local texture and global structural information in CR images. Thereby, it improves FE ability and boundary recognition accuracy in complex industrial environments [10,11]. Figure 1 shows the DeepLabV3+structure.

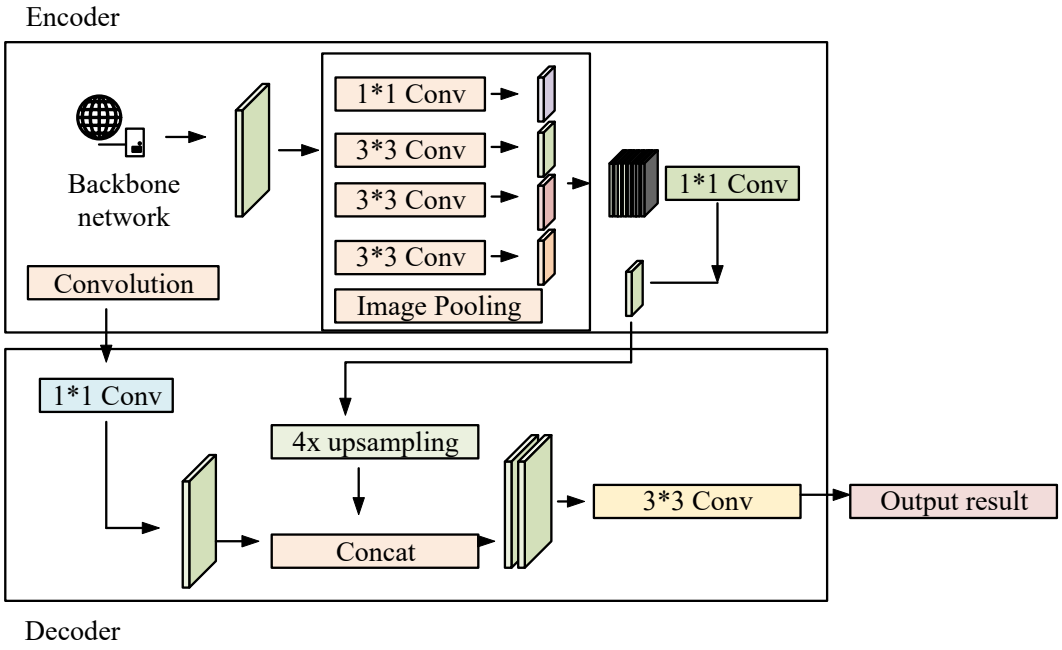


Fig. 1. DeepLabV3+ structure

From Figure 1, the DeepLabV3+ network consists of dilated convolution and dilated spatial pyramid pooling modules. The DeepLabV3+network is composed of two parts: an encoder and a decoder. The encoder first extracts preliminary features through a backbone network and then sends these features to multiple parallel branches for processing, including 1 × 1 convolution. The decoder upsamples the feature map output by the encoder by a factor of 4, concatenates it with the corresponding low-level detail features in the backbone network, and further fuses semantic and spatial information through a 3x3 convolution. Finally, it up samples again by a factor of 4 to output pixel level prediction outcomes of the same size as the input.

Traditional DeepLabV3+ has difficulty processing CR information of various scales and shapes in the recognition process. Therefore, CBAM is introduced to improve the network, and the attention module mainly improves the feature expression ability and improves the perception ability of key regions. CBAM uses a mechanism of channel attention and spatial attention concatenation to weight feature channels and further focus on key regions in the spatial dimension. Thereby, it improves the model's ability to express target boundaries and fine-grained structures, effectively enhancing the discriminative power of ASPP output features, and advancing the accuracy and robustness of semantic segmentation in complex scenes. Figure 2 shows the improved DeepLabV3+ network structure.

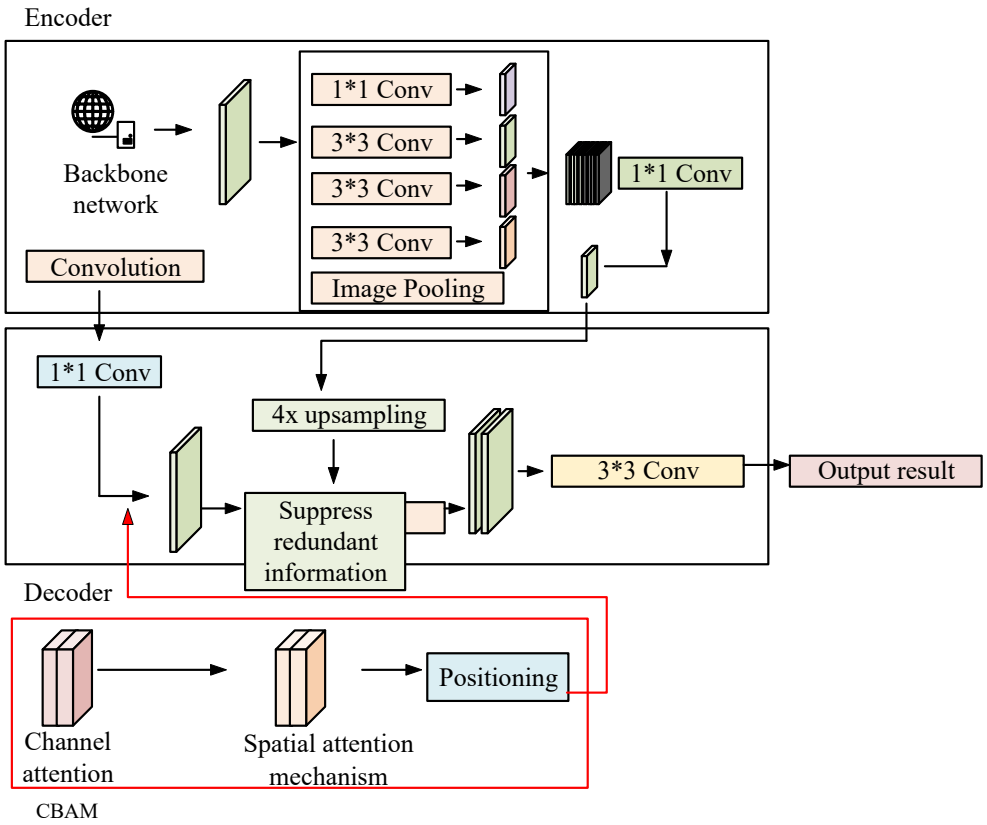


Fig. 2. Improved DeepLabV3+ network structure

From Figure 2, the attention module is mainly embedded after the ASPP module, allowing the network to perform channel and spatial dual weighting optimization on the output features after multi-scale feature fusion is completed. CBAM first adaptively adjusts the importance of each feature channel through channel AM, enhancing the response of key features related to target semantics and suppressing redundant information. Subsequently, the spatial AM is utilized to locate key regions in the feature map, making the model more focused on the target boundary and salient areas. Therefore, the network can transform from uniform feature processing to adaptive feature selection, thereby enhancing feature expression and discrimination capabilities.

The study employs the improved DeepLabV3+ network to conduct semantic segmentation on the working face images. Subsequently, the segmentation outcomes are post-processed using the Canny edge detection operator to further extract and improve the edge information at the CR

junction, ultimately achieving the extraction of clear and continuous CR boundary lines. The filtering and denoising formula for the Canny edge detection operator segmentation process is exhibited in equation (1) [11,12].

$$G(x,y)=\frac{1}{2\pi\sigma^2}e^{\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

In Equation (1), $G(x,y)$ means a two-dimensional Gaussian function, while x and y mean the center point offset distance, σ stands for the standard deviation of the Gaussian function, and $\frac{1}{2\pi\sigma^2}$ stands for the normalization coefficient. The larger the standard deviation value of the Gaussian function, the wider the smoothing range of the Gaussian kernel, and the stronger the noise suppression ability. However, the image details will become more blurred. The smaller σ is, the more details are retained, but the noise reduction effect becomes weaker. The Gaussian image convolution operation process is exhibited in Equation (2) [13,14].

$$D(x,y)=G(x,y)*s(x,y) \tag{2}$$

In Equation (2), $D(x,y)$ means the image convolution operation, and $s(x,y)$ means the initial pixel values of the input image. The calculation formulas for the first-order horizontal and vertical derivatives of the detection operator are exhibited in Equation (3).

$$A(x,y)=\frac{D(x,y+1)-D(x,y)+D(x+1,y+1)-D(x+1,y)}{2} \tag{3}$$

In Equation (3), $A(x,y)$ stands for the first-order derivative value in the horizontal direction. The first-order derivative value in the vertical direction is exhibited in Equation (4) [15,16].

$$B(x,y)=\frac{D(x,y)-D(x+1,y)+D(x,y+1)-D(x+1,y+1)}{2} \tag{4}$$

In Equation (4), $B(x,y)$ means the first-order derivative value in the vertical direction, and the formula for the change in gradient amplitude is exhibited in Equation (5).

$$\begin{cases} C(x,y)=\sqrt{A(x,y)^2+B(x,y)^2} \\ \rho(x,y)=\arctan(\frac{B(x,y)}{A(x,y)}) \end{cases} \tag{5}$$

In Equation (5), $C(x,y)$ means the gradient magnitude and $\rho(x,y)$ stands for the gradient direction angle. Based on the visual Coordinate System (CS), the image CS is constructed, as exhibited in Equation (6), which is the transformation formula between the image CS and the camera CS.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}=\begin{bmatrix} Q & t \\ 0^t & 1 \end{bmatrix}\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{6}$$

In Equation (6), Q stands for the rotation matrix, t stands for the translation vector, 0^t stands for the zero vector, (X_c,Y_c,Z_c) stands for the camera coordinate, and (X_w,Y_w,Z_w) stands for the world coordinate. The coordinate transformation formula for the camera projection relationship is exhibited in Equation (7) [17,18].

$$\begin{cases} x = r\dfrac{X_c}{Z_c} \\ y = r\dfrac{Y_c}{Z_c} \end{cases} \tag{7}$$

In Equation (7), r stands for the camera focal length. The conversion relationship between pixel coordinates and image coordinates is exhibited in Equation (8).

$$\begin{cases} v = \dfrac{x}{dx} + v_0 \\ n = \dfrac{y}{dy} + n_0 \end{cases} \tag{8}$$

In Equation (8), v and n represent the coordinate reference points in the pixel CS. Through camera coordinate transformation, the positioning and recognition assessment of CR images can be achieved.

2.2 Mining Control for Small Coal Pillar Working Faces Based on Digital Twin Technology

After achieving the identification of CR, to improve the efficiency of mining control in the CR working face, a working face mining control module is built through digital twin technology. The control analysis of the working face control module is studied using the PID control algorithm. In PID control, proportional control responds quickly grounded in the current deviation, but it can easily cause oscillation if the value is too large. Integral control accumulates historical deviations to eliminate steady-state errors. Differential control adjusts in improve grounded in the rate of change to improve stability [19]. The three components work together to achieve precise system control. In Figure 3, the structure of the PID control algorithm is presented.

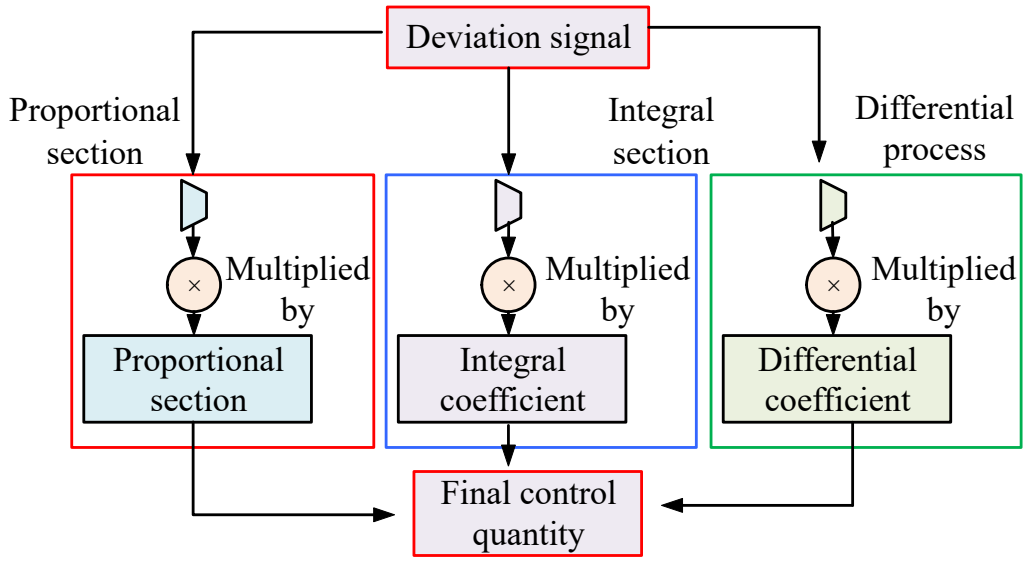


Fig. 3. PID control algorithm structure

From Figure 3, the PID control process system first compares the given target value with the actual feedback value to obtain the deviation signal. This deviation signal is then sent to three parallel processing sections. The proportional section multiplies the deviation by a proportional coefficient. The integral section multiplies the cumulative outcome of the deviation over time by an integral coefficient to eliminate the long-term steady-state error. The differential section calculates the rate of change of the deviation signal and multiplies it by a differential coefficient to predict the future trend of the deviation. The outputs of the three sections are added at the summation point to form the final control quantity, which acts on the controlled object and constitutes a closed-loop feedback.

To ensure better module control performance, the research has adopted a fuzzy controller to improve the PID algorithm through fuzzy control. In Figure 4, this is the structure of the fuzzy control algorithm.

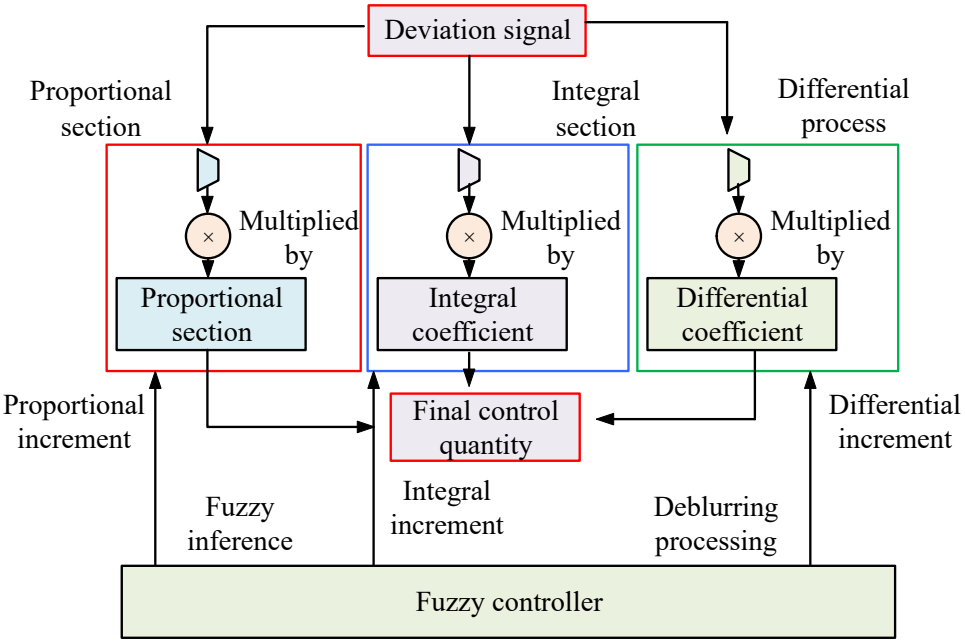


Fig. 4. Fuzzy control algorithm structure

From Figure 4, first, the input deviation value and deviation rate are sent to the fuzzy controller. The fuzzy controller performs fuzzy reasoning and defuzzification grounded in the pre-defined fuzzy rules, and outputs three correction quantities: the proportional Coefficient Increment (CI), the integral CI, and the differential CI. These three increments are respectively utilized to adjust the three control coefficients of the PID controller in real time. After parameter correction, the PID controller applies proportional, integral, and differential coefficient to the error signal, generating control quantities that act on the controlled object, and simultaneously feeding back the output outcomes to the input end to form a closed-loop control.

To ensure that the fuzzy controller can achieve the key control process of the module, the research involves using the Modelica software to simulate the system and build the fuzzy PID control module, as exhibited in Figure 5.

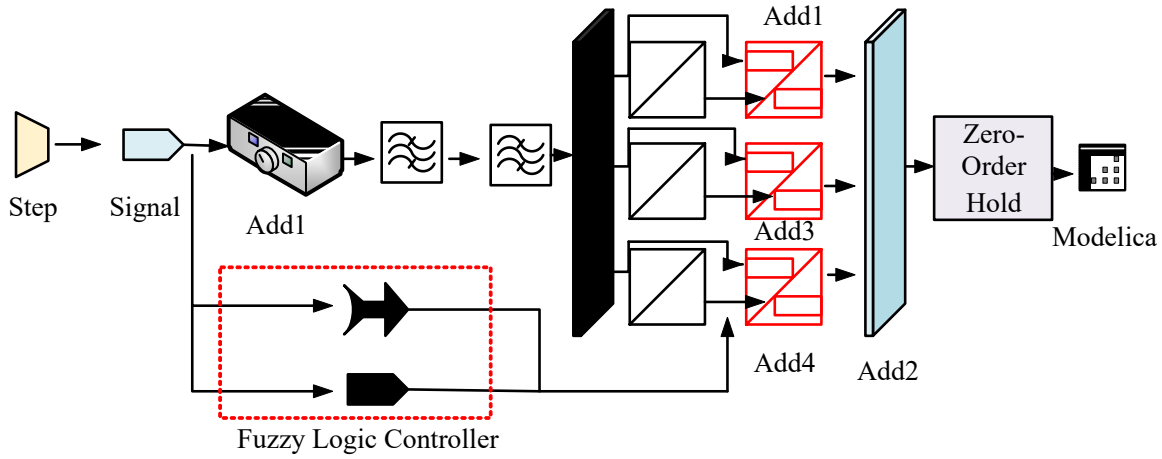


Fig. 5. Fuzzy PID control module

From Figure 5, the module first inputs a step command through the "Step" signal source. The signal is compared with the feedback signal through "Add1" and generates a deviation. The deviation signal is simultaneously sent to two paths. One path is multiplied by the three gain coefficients

(proportional, integral, and derivative) and then summed through "Add2" to form the conventional PID control quantity. The other path is sent to the "Fuzzy Logic Controller", where three correction quantities are output, which are utilized to adjust the three gain coefficients (proportional, integral, and derivative). The two control signals are superimposed through "Add3" and "Add4" and then discretized through "Zero-Order Hold". Finally, they are simulated through the "Modelica" interface and transmitted to the controlled object to form a complete closed-loop control and parameter adaptive adjustment system. The final output is observed in real time through the step oscilloscope and forms a complete closed-loop control and parameter adaptive adjustment system.

The control module uses virtual interaction to simulate the control of the small coal pillar working face. In Figure 6, it is the virtual interaction control framework.

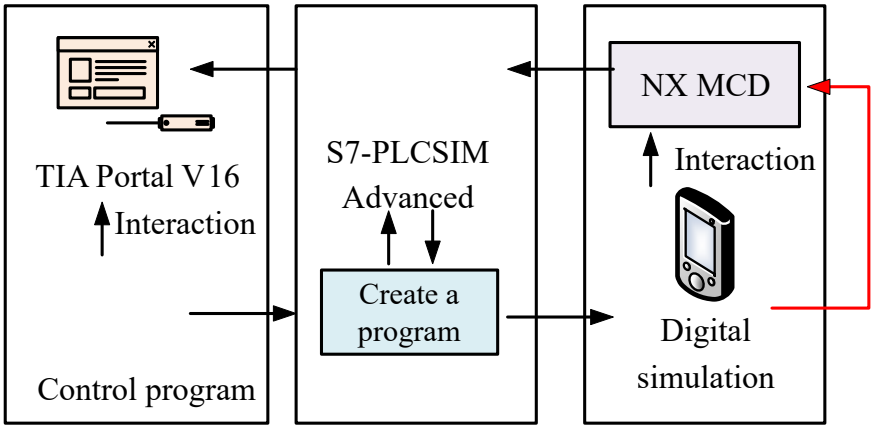


Fig. 6. Virtual interaction control framework

From Figure 6, the virtual interactive control process begins by writing the PLC control program in TIA Portal V16, and then a virtual PLC is created through S7-PLCSIM Advanced. The virtual PLC establishes a communication connection with the digital simulation model in NX MCD. During the system operation, the PLC program sends control instructions to the NX MCD simulation model, and after the model performs the corresponding action, it feeds back the status signal to the virtual PLC. Finally, the virtual interactive control is achieved through the PLC.

In the process of digital twin virtual interactive control, a closed-loop link is established between the physical working face and the virtual working face. The equipment operating status, coal mining machine operating parameters, and coal rock interface recognition results are input into the virtual PLC and transmitted to the NX MCD virtual working face model to achieve dynamic updates of equipment position, operating status, and coal rock interface. NX MCD drives the virtual coal mining machine to operate according to control instructions and feeds back the operating status to S7-PLCSIM Advanced before entering the next control cycle. The synchronization mechanism is driven by the control program running cycle and simulation step size, and the fuzzy PID dynamically adjusts the control parameters based on the coal rock interface recognition results and deviation changes. The coal mining machine, control logic, and operating status in the virtual working face are mapped to the corresponding equipment, PLC program, and operating status in the actual working face, and their fidelity is mainly reflected in the consistency of control logic, equipment motion relationships, and state feedback. Due to the lack of on-site real-time data collection on frequency, communication delay, and virtual real state errors in the current experiment, the system should be positioned as a digital twin simulation verification environment rather than a complete digital twin system.

3. Results

3.1 Experimental Setup

To analyze and test the recognition effects of different algorithms on CRs, the research utilized an experimental platform equipped with an Intel Core i7-12700K processor (with a frequency of 3.60 GHz), 32 GB DDR4 memory, and a NVIDIA GeForce RTX 3080 Ti (with 12 GB video memory) GPU.

The storage was 1 TB NVMe SSD. In terms of software, the operating system was Windows 11 Professional 64-bit version, and the DL framework was grounded in Python 3.9, using PyTorch 2.0 as the main modeling tool, and combined with CUDA 11.8 and cuDNN 8.6 to achieve GPU acceleration. Data preprocessing and algorithm comparison were completed on MATLAB R2022b, and OpenCV 4.7 was utilized to implement image improvement and FE processing. S7-PLCSIM Advanced created a virtual PLC instance and established a connection with NX MCD through S7 communication or OPC UA. The data exchange cycle was set to 50 ms, and the communication delay was controlled within 10-20 ms. The fuzzy PID sampling cycle, zero order keeper cycle, and control cycle were all set to 50 ms, and the NX MCD simulation step size was set to 10 ms. The control program running cycle was consistent with the simulation step size, and the virtual coal mining machine status was fed back to the virtual PLC at the end of each control cycle.

All experiments were carried out under the same data set division and parameter settings. All DL models were trained with uniform parameters, where the Batch Size was set to 32, the Epoch was set to 200, the optimizer was Adam, the initial Learning Rate (LR) was set to 0.001, the weight decay was 1×10^{-4}, the LR was dynamically adjusted using the cosine annealing strategy, and the minimum LR was set to 1×10^{-6}. The input image size was uniformly 224×224, Dropout was set to 0.5, the loss function was cross-entropy loss, and the early stopping strategy (Patience=20) was utilized to prevent overfitting. Traditional machine learning algorithms were optimized using 5-fold cross-validation, where SVM was set with C=10 and γ=0.01, the number of random forest trees was 200, the number of Extra Trees trees was 300, the number of KNN neighbors was 5, and all experiments fixed the random seed to 42.

To ensure fairness in the efficiency testing of different models, all compared models were tested in the same Intel Core i7-12700K, 32 GB DDR4 memory, and NVIDIA RTX 3080 Ti 12 GB GPU environment. Batch Size=1 was uniformly used in the inference stage, and the input image was uniformly adjusted to 224 × 224 pixels. Before the official timing, the study performed 10 GPU warm-up cycles, which were not included in the formal testing process. Subsequently, 100 consecutive inferences were performed on the test set and the average was taken to calculate the inference time and FPS. During the testing process, GPU utilization remained at a similar level. The FPS and inference time of YOLO11, RT-DETR, EfficientViT, ConvNeXt, and improved DeepLabV3+were obtained under identical hardware, input resolution, batch size, warm-up, and timing conditions.

The dataset included three types of labeled areas: coal seam, rock stratum and coal rock interface. Coal seam and rock stratum were the main semantic categories, and coal rock interface was the boundary labeled area. Labelme was used to manually label the original image at the pixel level. Two individuals with experience in CM field recognition independently completed the initial labeling, and a third person reviewed and corrected the boundary areas, ultimately generating PNG format semantic segmentation labels. After removing blurry, duplicated, and annotated images with poor quality, 2856 valid images were obtained, including 998 in the training set, 428 in the validation set, and 430 in the test set, with a division ratio of approximately 7:1.5:1.5. To avoid data leakage, the dataset was not divided into completely random image by image partitions, but was grouped according to acquisition batches and continuous acquisition sequences. Images from the same acquisition site, the same continuous acquisition sequence, and adjacent frames were assigned to only one subset—training, validation, or test—to ensure that test set images did not directly appear during model training. The original image resolution was uniformly adjusted to 224 × 224 pixels after preprocessing, and the same data enhancement and normalization strategy was used for model training. The final test set was used to evaluate the generalization performance of the model on coal rock interface images that were not trained, and the recognition results reported in this article were obtained based on this evaluation.

To avoid data leakage caused by adjacent frames or highly similar images entering both the training and testing sets simultaneously during the continuous acquisition process in the mining area, the study used the acquisition batch and continuous acquisition sequence as the basic units in data partitioning, rather than completely randomly partitioning a single image. Firstly, the data was grouped according to the image acquisition time and continuous acquisition process. The same

acquisition sequence and its adjacent frame images were only allocated to a subset of the training set, validation set, or test set, and images from the same sequence were not allowed to appear across different datasets. Subsequently, the training set, validation set, and testing set were divided among each independent collection group.

The coal rock boundary annotation was completed by professionals with experience in identifying CM sites. The mining area image was annotated pixel by pixel according to a unified coal rock boundary judgment standard. The actual contact area between the coal seam and the rock layer was used as the coal rock boundary, and the areas with blurred boundaries, dust obstruction, and lighting changes were judged based on adjacent textures and spatial continuity. After the initial annotation was completed, another annotator will independently review and verify the areas with differences item by item. Experts with relevant professional experience will confirm the disputed areas and form the final annotation results. During the quality control process, the focus was on checking the integrity of the marked area, the continuity of the coal rock boundary, the consistency of the category, and the accuracy of the boundary position. If missing, incorrect, or boundary offset was found, it would be corrected again. For samples with significant differences between annotators, expert review results were used as the final labels, and a unified data quality check was completed before model training.

3.2 Comparative Analysis of Improved Recognition Effect of DeepLabV3+

To ensure the effectiveness of comparing different models, YOLO11, RT-DETR, EfficientViT, and ConvNeXt were selected as comparison methods, mainly considering their representation of different technical routes such as current object detection, real-time detection, lightweight visual network, and general visual backbone network. They could be used to evaluate the comprehensive performance of improved DeepLabV3+in coal rock interface recognition tasks from the aspects of detection performance, real-time performance, and feature extraction ability. The study compared You Only Look Once 11 (YOLO11), RT-DETR, Efficient Vision Transformer (EfficientViT), and the Convolutional Network Optimization Model (ConvNeXt) separately. The comparison indicators were: Mean Intersection over Union (mIoU) measured the overlap between predicted segmentation regions and real annotated regions, with higher values indicating better consistency and more accurate segmentation. Mean Pixel Accuracy (MPA) was utilized to measure the average accuracy of correctly classifying pixels of each category, and the higher the MPA value, the higher the classification accuracy of the model for each category of pixels. The CR recognition effects of different algorithms were compared, as exhibited in Figure 7.

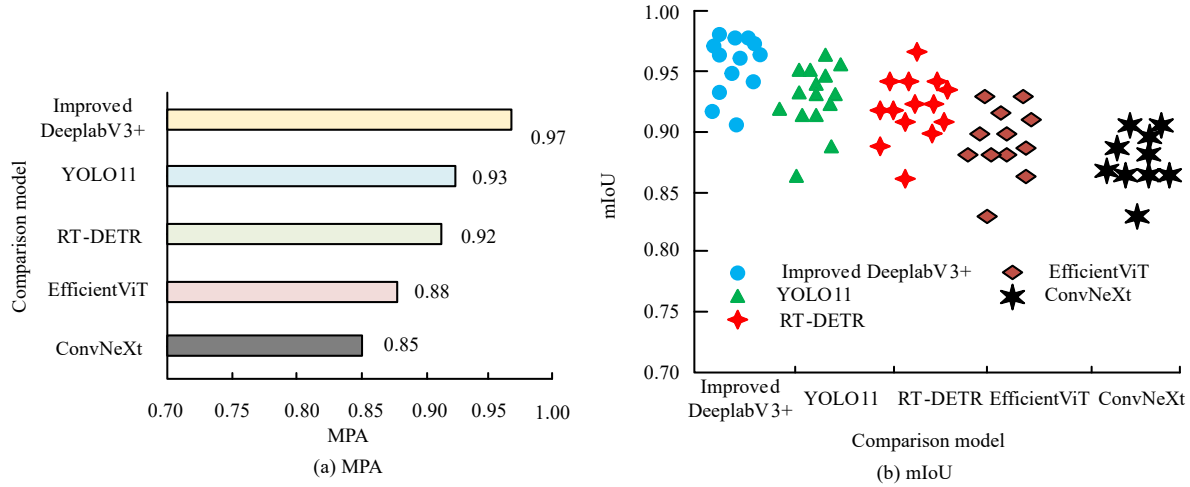


Fig. 7. Comparison outcomes of recognition effects of different algorithms

From Figure 7(a), in the comparison of different algorithms, the improved DeepLabV3+ algorithm utilized in the research had a better CR recognition effect. The MPA value could reach up to 0.97 at its maximum, which was approximately 0.12 higher than that of ConvNeXt. This indicated that in the comparison of CR recognition effects of different algorithms, the improved DeepLabV3+ algorithm utilized in the research had a better recognition effect and higher recognition accuracy.

From Figure 7(b), in the mIoU value comparison analysis, the changes in mIoU values of different algorithms were relatively large. The improved DeepLabV3+ algorithm utilized in the research had a better mIoU value compared to other algorithms, while the mIoU value of ConvNeXt was lower among the several algorithms, with its maximum value being only 0.91, which was 0.07 lower than that of the improved DeepLabV3+ algorithm. Thus, in the comparison of recognition effects of different algorithms, the improved DeepLabV3+ algorithm utilized in the research had better CR recognition accuracy and image segmentation ability.The study conducted a comparison of the running tests for different algorithms, and the outcomes are exhibited in Table 1.

Table 1. Comparison outcomes of recognition effects of different algorithms

Indicator	YOLO11	RT-DETR	EfficientViT	ConvNeXt	Improved DeepLabV3+
Accuracy (%)	96.35	96.82	97.24	97.56	98.73
Precision (%)	95.92	96.41	96.95	97.18	98.46
Recall (%)	95.68	96.07	96.58	96.92	98.18
F1-score (%)	95.8	96.24	96.76	97.05	98.32
AUC	0.978	0.982	0.986	0.989	0.996
FPS	88.7	53.9	96.5	71.8	102.8
Inference Time (ms)	11.28	18.55	10.36	13.93	9.73
Parameters (M)	9.4	20.3	8.2	28.6	7.5
FLOPs (G)	21.6	59.7	17.4	45.9	15.8
Model Size (MB)	19.8	42.5	17.6	56.8	15.4
mIoU(%)	92.84	93.36	94.21	94.86	96.42
MPA(%)	94.17	94.72	95.36	95.83	97.18
BIoU(%)	87.63	88.41	90.12	91.37	94.35
BF-score(%)	91.28	91.96	93.05	93.76	96.08

From Table 1, the performance of different algorithms varied significantly in the CR identification task. Among them, the improved DeepLabV3+ had better model performance in various indicators. The Accuracy, Precision, Recall, F1-score and AUC of the improved DeepLabV3+ algorithm reached 98.73%, 98.46%, 98.18%, 98.32% and 0.996, respectively. Compared with traditional algorithms, it had higher index values, indicating that the improved DeepLabV3+ algorithm had higher CR classification accuracy and stronger comprehensive recognition ability. In the model performance test comparison, the FPS of the improved DeepLabV3+ could reach 102.8, and the inference time was only 9.73 ms. Moreover, the parameter size, computational complexity and model size declined to 7.5 M, 15.8 GFLOPs and 15.4 MB, which were higher than those of other traditional algorithms. This indicated that the model could achieve high recognition accuracy while having faster inference speed, lower computational complexity and better lightweight characteristics, which could meet the actual demand for timely processing and deployment efficiency in the intelligent mining scenarios of CMs. The study would conduct ablation experiments by comparing the algorithm models before and after introducing different modules, and the outcomes are exhibited in Figure 8.

From Figure 8(a), in the comparison of performance tests for different modules, after introducing different modules, the overall performance of the model underwent significant changes compared to before the introduction. Among them, after applying the Canny operator processing, the overall MPA value of the algorithm increased by 0.02 compared to before the improvement. Moreover, after introducing CBAM, the overall MPA value of the algorithm increased by approximately 0.04 compared to before the improvement. The overall MPA value of the algorithm significantly improved after introducing different modules. From Figure 8(b), the mIoU value of the algorithm also significantly improved after introducing different modules. Among them, after applying the Canny operator processing, the overall mIoU value of the algorithm increased by 0.07 compared to before the improvement. After introducing CBAM, the overall mIoU value of the algorithm increased by approximately 0.05 compared to before the improvement. The overall image recognition ability of the algorithm also improved after introducing different modules. The study

would conduct a comparison of the operational test outcomes of the model after introducing different modules, and the comparison outcomes are exhibited in Table 2.

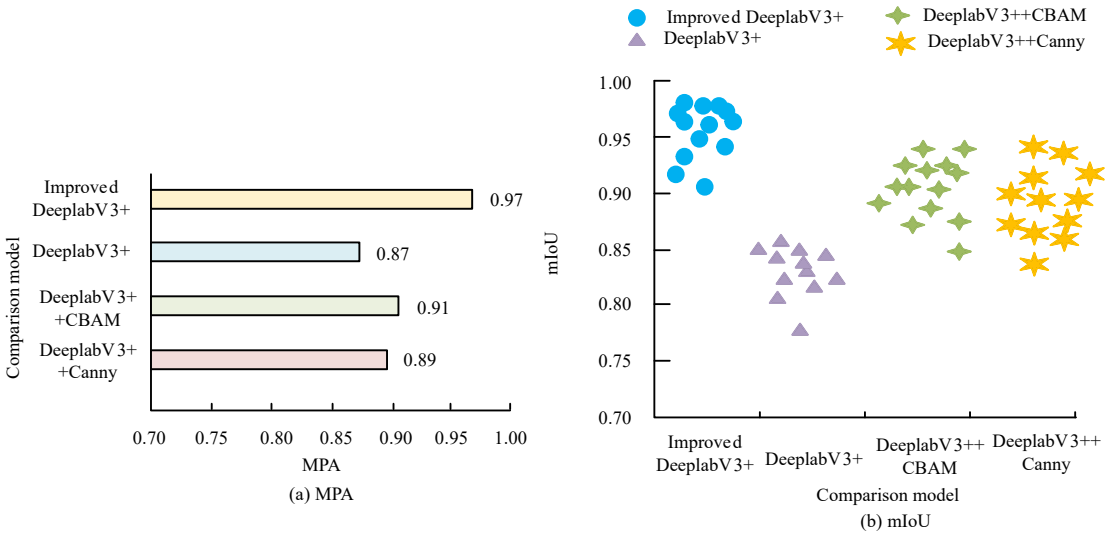


Fig. 8. Comparison of ablation test outcomes for different modules

Table 2. Comparison outcomes of running tests on different module tables

Indicator	DeepLabV3+	DeepLabV3++Canny	DeepLabV3++CBAM	Improve DeepLabV3+
Accuracy (%)	96.84	97.42	97.86	98.73
Precision (%)	96.35	97.01	97.68	98.46
Recall (%)	96.08	96.75	97.34	98.18
F1-score (%)	96.21	96.88	97.51	98.32
AUC	0.983	0.987	0.991	0.996
FPS	91.5	89.8	87.6	102.8
Inference Time (ms)	10.93	11.14	11.42	9.73
Parameters (M)	8.4	8.5	8.8	7.5
FLOPs (G)	18.7	18.9	19.4	15.8
Model Size (MB)	17.2	17.5	18.1	15.4

From Table 2, compared with DeepLabV3+, after adding Canny edge features, the Accuracy of the model grew from 96.84% to 97.42%, while Precision, Recall, F1-score and AUC increased to 97.01%, 96.75%, 96.88% and 0.987. This indicated that the introduction of edge information could improve the model's ability to obtain the boundary features of CR. After presenting the CBAM-AM, the test indicators of the model were significantly improved, with AUC reaching 0.991. Thus, CBAM could effectively strengthen the expression of key features and improve the model's attention to the texture and detail information of CR. The improvement of DeepLabV3+ in terms of model performance also had more obvious operational effects. Its FPS reached 102.8, and the inference time was as low as 9.73 ms. Compared with DeepLabV3+, DeepLabV3++Canny, and DeepLabV3++CBAM models, the parameters, computational volume and model size of the improved DeepLabV3+ had better test performance. The improved DeepLabV3+ not only effectively improved the accuracy of CR recognition but also achieved model lightweighting and real-time performance optimization while maintaining high accuracy recognition. After adding Canny separately, the model's parameter counts and FLOPs became 8.5 M and 18.9 G, respectively. Canny only acted on the network output and did not introduce trainable parameters, but its edge processing added a small amount of computational overhead. After adding CBAM separately, due to the addition of channel attention and spatial attention branches, the parameter counts and FLOPs further increased to 8.8 M and 19.4 G. Finally, the parameter counts of the improved model

decreased to 7.5 M and FLOPs decreased to 15.8 G, which was not directly caused by Canny or CBAM, but was due to the lightweight adjustment of the network structure used simultaneously in the improved model.

After completing the semantic segmentation of coal and rock regions in the improved DeepLabV3+, the obtained segmentation results were input into the Canny edge detection operator for boundary enhancement processing. Canny first performed gradient calculation and non-maximum suppression on the segmentation results and obtained continuous edge information through dual threshold detection and edge connection. Subsequently, the detected edge positions would be used to refine the coal rock boundary area. Canny did not participate in the internal learnable feature extraction and fusion of DeepLabV3+in the research but acted on the network output to refine the boundary of the initial semantic segmentation results. Subsequent research would further use statistical methods such as multiple random seeds, multiple batches of repeated experiments, and confidence intervals to systematically evaluate the performance of the model and expand the testing scope under different working faces and mining conditions to quantify the random variability and cross scenario stability of the model performance.

3.3 Workface Mining Control and CR Identification Outcomes

The study compared the identification outcomes obtained by different methods and the outcomes are exhibited in Figure 9.

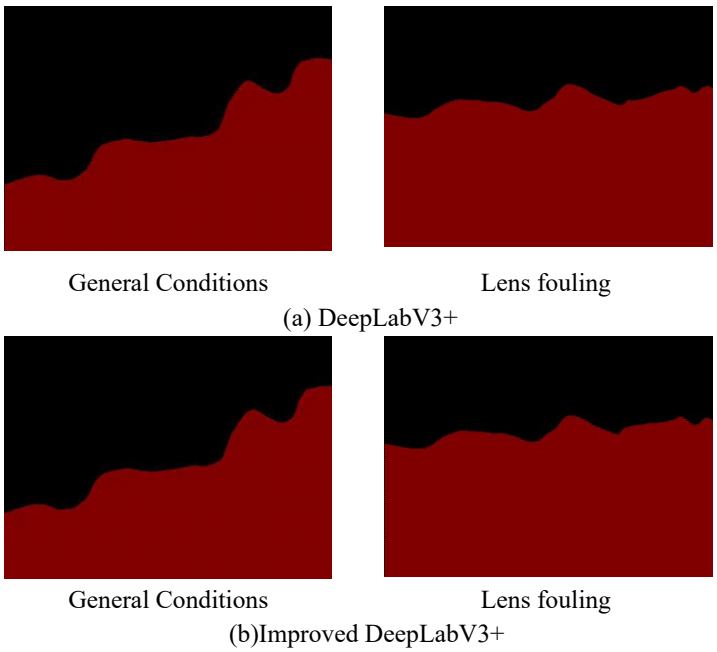


Fig. 9. Comparison of CR recognition effects

In Figure 9, the improved and unimproved DeepLabV3+ models' performance in identifying the CR interface under different conditions was able to successfully complete the segmentation of the target area. The improved model showed a more stable and accurate overall performance. Under normal conditions, the CR boundaries predicted by the improved DeepLabV3+ were more consistent with the real boundaries, with smoother and more continuous edge contours and more complete local details. In the case of lens dirt contamination, the improved model still maintained good segmentation accuracy, with clearer predicted boundaries, more consistent local fluctuations with the real CR interface, and significantly reduced errors in segmentation and omission. Compared with the original DeepLabV3+, the improved DeepLabV3+ showed more stable feature extraction and boundary segmentation under the tested normal and lens-contamination conditions, with more continuous CR boundaries and fewer local segmentation errors. However, the current experiments mainly provided qualitative evidence under these two conditions, and the robustness of the model under dust, illumination variation, vibration, blur, occlusion, lens contamination, and coal seam variation required further quantitative validation. Figure 9

qualitatively demonstrated the segmentation performance of the improved DeepLabV3+ under normal conditions and lens contamination. The results showed that the improved model maintained relatively clear and continuous CR boundaries under the tested interference condition.

Fuzzy PID inputted deviation and change rate, outputted proportional, integral, and derivative three gain correction. The linguistic variable has 7 levels, with a domain of -3 to 3, triangular membership functions, and 50% adjacent overlap. The scaling factor was determined according to the actual range, such as 0.03 for a deviation of 100 mm and 0.06 for a change rate of 50 mm/s. The correction upper limits of 0.3, 0.02, and 0.02 corresponded to output factors of 0.1, 0.0067, and 0.0067. The rule library was of the Mamdani type, with a weight of 1. The smaller the weight, the larger the weight. When the deviation was large, the integral decreases proportionally. When the deviation was small, the integral increases. When the rate of change was large, the derivative increased. Near zero, the correction tended towards zero. The center of gravity method was used to blur and multiply by the output factor. The initial PID values were 1.20, 0.08, and 0.05, and were adjusted using Z-N or Modelica. The online gain was initially adjusted and limited. The tuning program consisted of offline quantitative factors, output factors, and initial gains. The deviation and rate of change, quantization, inference, deblurring, PID correction, and amplitude limiting were calculated on a weekly basis. The control variables were sent to S7-PLCSIM Advanced and NX MCD, and fed back to the virtual PLC. The sampling period was consistent with the PLC scanning and simulation step size. This study analyzed the control effect of simulated working face. The results are shown in Figure 10.

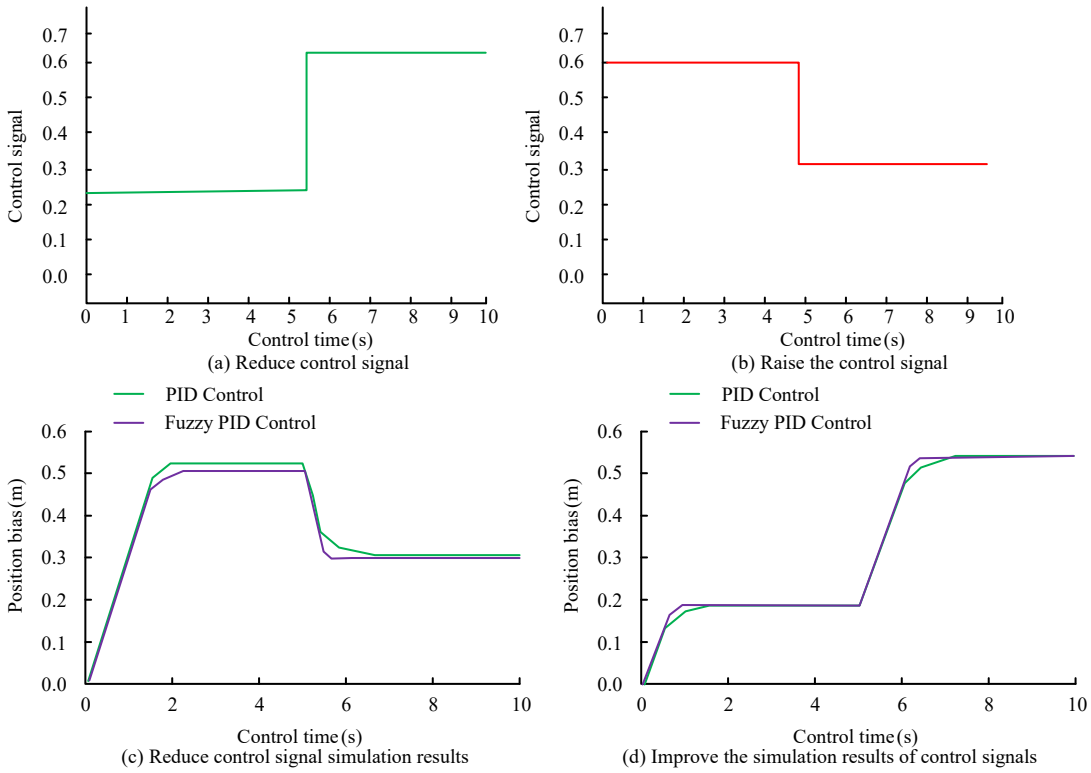


Fig. 10. Comparison of control outcomes for different signals

From Figure 10, there were differences in the fitting degree of different models at the change nodes. Among them, the curve of the fuzzy PID control algorithm was closer to the actual change trend at the key turning points, with a smoother response and smaller deviation. The traditional model had certain fluctuation phenomena in the local area, while the overall curve of the fuzzy PID control algorithm was more stable and had a higher coincidence with the reference curve. Especially near the sudden change points, it showed better tracking ability and robustness. Overall, the improved model outperformed the traditional model in terms of fitting accuracy of the change trend, stability, and local response ability. This indicated that the fuzzy PID had a better control effect on the control signal of the working face mining.

The study selected typical coal and rock images from the test set and inputted them into the improved DeepLabV3+. It outputted the coal and rock interface segmentation results and extracted continuous coal and rock boundary lines through Canny post-processing. The study compared them with the target cutting trajectory to obtain the interface position deviation and deviation change rate, which were used as inputs for the fuzzy PID controller. The fuzzy PID output control quantity was then processed by a PLC program written in TIA Portal V16. The S7-PLCSIM Advanced virtual PLC was fed into the NX MCD virtual working face model to drive the virtual coal mining machine to adjust the cutting arm or drum position. The NX MCD then fed back the actual position and operating status of the virtual coal mining machine to the virtual PLC to form a closed-loop control. At the same time, the recognition results, control variables, and virtual coal mining machine trajectory changes were synchronously recorded, and the trajectory tracking error, maximum deviation, adjustment time, overshoot, steady-state error, and control variable fluctuation were used as evaluation indicators. Compared with traditional PID control, it was expected that the improved DeepLabV3+could trigger real-time adjustment of fuzzy PID parameters to change the virtual coal mining machine trajectory, making the trajectory closer to the target coal rock interface. Fuzzy PID was superior to traditional PID in response speed, stability, and tracking accuracy.

4. Conclusions

To address the issues of insufficient accuracy in identifying the boundary between CR during intelligent coal mining and the low level of intelligence in the control system, a new mining control method for small coal pillar working faces was introduced by integrating the improved DeepLabV3+ with digital twin. The new method utilized DeepLabV3+ as the base network, introduced the CBAM-AM to improve the model's power to express key channels and spatial features, and combined the Canny edge detection operator to strengthen the extraction of edge information in the CR junction area. Thereby, it improved the segmentation accuracy in complex working conditions. The research outcomes implied that in different algorithm comparisons, the improved model delivered better Accuracy, Precision, Recall, F1-score, and AUC, among others. The Accuracy could reach 98.73%, which was higher than that of the ConvNeXt model. The mIoU and MPA indicators were higher than those of traditional algorithms in terms of pixel classification ability. In terms of model performance, the FPS of the improved DeepLabV3+ could reach 102.8, and the inference time was only 9.73ms. The parameter size and computational cost were only 7.5M and 15.8GFLOPs, which were significantly superior to traditional models.

In practical applications of the research model, the use of S7-PLCSIM Advanced could simulate controller and TCP/IP communication functions, and support collaborative simulation with machine simulation. However, the communication process in the simulation environment could not fully represent the industrial Ethernet or wireless communication link of the real coal mining machine site. During actual deployment, the communication between the visual computing unit and the digital twin platform was also affected by factors such as network topology, transmission distance, switching equipment, wireless signal attenuation, on-site electromagnetic interference, packet retransmission, and network load, resulting in varying degrees of transmission and processing latency.

In the control system, a virtual interactive control framework grounded in digital twin was constructed, and the fuzzy PID algorithm was introduced to achieve dynamic optimization control of the coal mining machine operation process. The research outcomes implied that compared with traditional control methods, the improved control strategy had better performance in terms of response speed, stability and tracking accuracy. The improved DeepLabV3+ model introduced in the research combined with digital twin technology achieved good outcomes in CR recognition and intelligent mining control, but still had certain deficiencies. The research mainly proposed an improved simulation verification framework that integrated DeepLabV3+with digital twins, which enhanced the accuracy of coal rock interface recognition and fuzzy PID virtual control performance in a given dataset and virtual working face. However, current research still has limitations, such as evidence only coming from computation and virtual simulation, without actual underground deployment, and not including complete digital twin elements such as on-site data acquisition frequency, communication delay, and virtual real state errors. The effectiveness boundary is

limited to specific datasets, device parameters, sampling periods, and simulation step sizes. The dust, light, vibration, communication delay, and long-term stability under actual underground mining conditions still need to be verified on site. After modification, duplicate values will be removed, focusing on contributions, limitations, and issues to be verified, and clearly labeled as simulation-based research results. Therefore, in the future, different mining areas and complex working conditions can be further expanded to collect CR data sets. Meanwhile, in the future, advanced technologies such as visual Transformer and multimodal fusion will be combined to further advance the accuracy and robustness of CR boundary recognition in complex environments.

Funding

The research is supported by the Natural Science Foundation of HeBei Province of China (Grant No. E2022508046): Optical and Spectroscopic Diagnostics of Flow Fields Involving Gas Explosion and Flame Retardant Effects, and the CHN Energy Ningxia Coal Industry-Group CO. LTD scientific research project 'Research and Application of Mining Key Technology in Small Coal Pillar Working Face with Complex Conditions Multi-Seam' (ShiCaoCun Colliery-Technology (2025) No. 003).

References

[1] Saputra MRU, Bhaswara ID, Nasution BI, Ern MAL, Husna NLR, Witra T. Multi-modal deep learning approaches to semantic segmentation of mining footprints with multispectral satellite imagery. Remote Sensing of Environment. 2025;318(2):114584. https://doi.org/10.1016/j.rse.2024.114584

[2] Kumar A, Singh P, Sharma R, Patel V. Digital twin-driven intelligent monitoring and control framework for underground mining operations. Computers & Industrial Engineering. 2025;201(3):110871. https://doi.org/10.1016/j.cie.2025.110871

[3] Jones S, Williams D, Brown T, Smith K. A framework to define, design and construct digital twins in the mining industry. Computers & Industrial Engineering. 2025;200(3):110805-110806. https://doi.org/10.1016/j.cie.2024.110805

[4] Oliveira G, Silva R, Costa J, Ferreira L. Digital transformation and intelligent automation technologies in the mining industry: A systematic review. Resources Policy. 2024;91(9):104812.

[5] Xiao H, Ling L, Shao H, Sun X, Wan L, Zhou Y, et al. A Fine-Grained Remote Sensing Classification Approach for Mine Development Land Types Based on the Integration of HRNet and DeepLabV3+. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2025;10(10):25159-25176. https://doi.org/10.1109/JSTARS.2025.3608448

[6] Zhang H, Li Y, Wang X, Liu Z, Zhao T, Chen J. LPT-Net: A Line-Pad Transformer Network for efficient coal gangue segmentation with linear multi-head self-attention mechanism. Measurement. 2024;226(10):114043-114044. https://doi.org/10.1016/j.measurement.2023.114043

[7] Li X, Zhao H, Wang Y, Sun P, Zhang J, Liu C. A detection method for densely distributed coal and gangue with uneven particle based on DeepLabV3+. Measurement. 2024;235(6):114756-114757. https://doi.org/10.1016/j.measurement.2024.114756

[8] Zhang Y, Liu X, Wang C, Li Z, Zhao H, Chen Q. Intelligent detection of coal gangue in mining operations using multispectral imaging and enhanced RT-DETR algorithm for efficient sorting. Microchemical Journal. 2024;207(6):111789-111790. https://doi.org/10.1016/j.microc.2024.111789

[9] Wang H, Li P, Zhao J, Sun L, Chen X, Zhang Y. Coal resources under carbon peak: Segmentation of massive laser point clouds for coal mining in underground dusty environments using integrated graph deep learning model. Energy. 2023;285(4):128771-128772. https://doi.org/10.1016/j.energy.2023.128771

[10] Smith K, Johnson R, Miller P, Davis A. Intelligent mine monitoring based on cyber-physical systems and digital twin technology. Journal of Cleaner Production. 2023;417(1):137912-137913.

[11] Patel S, Kumar R, Singh A, Wilson J. Autonomous underground mining systems enabled by artificial intelligence and advanced sensing technologies. International Journal of Mining Science and Technology. 2024;34(2):245-257.

[12] Martin P, Garcia F, Lopez J, Hernandez M. Real-time semantic segmentation for underground mine environment perception using deep neural networks. Engineering Applications of Artificial Intelligence. 2024;132(6):107938-107939.

[13] Wilson J, Brown A, Taylor M, Clark R. Deep learning-based visual perception system for autonomous mining vehicles under complex underground conditions. IEEE Transactions on Industrial Informatics. 2024;20(6):7345-7356.

[14] Turner D, Edwards K, Smith R, Green P. Intelligent robotic mining systems based on digital twins and reinforcement learning control. IEEE Transactions on Automation Science and Engineering. 2025;22(1):512-524.

[15] Anderson J, Williams P, Clark S, Roberts M. Multi-source data fusion framework for intelligent underground mine safety monitoring. Safety Science. 2023;164(2):106164-106165.
[16] Taylor J, Wilson B, Edwards R, Brown C. Artificial intelligence-based geological condition recognition for automated mining operations. Computers & Geosciences. 2024;184(6):105529-105530. https://doi.org/10.1016/j.cageo.2024.105529
[17] Evans D, Parker J, Mitchell A, Collins R. Digital twin technology for real-time decision support in complex industrial environments. Advanced Engineering Informatics. 2023;57(3):102103-102104.
[18] Thompson R, Martin S, Davis L, White P. Intelligent control architecture for autonomous mining equipment using edge computing and digital twins. IEEE Access. 2023;11(3):128456-128468.
[19] Wilson M, Harris J, Cooper D, Adams P. Deep learning-based semantic segmentation for industrial inspection and autonomous robotic perception. Pattern Recognition Letters. 2025;197(1):325-331. https://doi.org/10.1016/j.patrec.2025.08.019