IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 July 2023, accepted 26 July 2023, date of publication 31 July 2023, date of current version 7 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3300243

RESEARCH ARTICLE

Coal-Rock Image Recognition Method for Complex and Harsh Environment in Coal Mine Using Deep Learning Models

SUN CHUANMENG[1,2], **LI XINYU**[1,2], **CHEN JIAXIN**[1,2], **WU ZHIBO**[1,2], **AND LI YONG**[3]
[1]State Key Laboratory of Dynamic Measurement Technology, North University of China, Taiyuan 030051, China
[2]School of Electrical and Control Engineering, North University of China, Taiyuan 030051, China
[3]State Key Laboratory of Coal Mine Disaster Dynamics and Control, Chongqing University, Chongqing 400044, China

Corresponding authors: Sun Chuanmeng (suncm@nuc.edu.cn) and Li Yong (yong.li@cqu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFC2905700, in part by the Fundamental Research Program of Shanxi Province under Grant 202203021221106, and in part by the Shanxi Science and Technology Achievement Transformation Guidance Special Project under Grant 202104021301061.

ABSTRACT The unfavorable factors of underground coal such as dark light, uneven illumination, band shadowing greatly make it difficult to recognize the coal rock at the mining workface accurately. To solve this problem, this paper proposes the fuse attention mechanism's coal rock full-scale network (FAM-CRFSN) model. The deep extraction of coal rock semantic features is achieved by a multi-channel residual attention mechanism and a full-scale connection structure. Meanwhile, the balance between "deep" stacking and error back propagation is achieved by structures such as dilated convolution and Res2Block. Besides, a multi-dimensional loss function consisting of the cross-entropy loss, intersection over union, and multiscale structure similarity loss with pixel-level, area-level, and image-level expressions is established. Finally, the performance of the FAM-CRFSN network is tested with RGB coal rock images collected from an underground coal mining workface and superimposed with different proportions of gaussian noise and salt & pepper noise. The experimental results show that the FAM-CRFSN model can segment the coal rock regions accurately; at a noise intensity of 0.09, it achieves an MIOU of 85.77% and an MPA of 92.12%. Also, it achieves better accuracy and generalization performance than the mainstream semantic segmentation models. This study provides an important theoretical basis for promotes the unmanned and intelligent mining workface.

INDEX TERMS Intelligent mining, automatic recognition of coal rock, semantic segmentation, low-illuminance image segmentation.

I. INTRODUCTION

Intelligent, unmanned mining is a trend in underground coal mining [1]. Automatic recognition of coal rock is the key to achieving intelligent and unmanned mining and excavation, and great efforts have been made based on image processing technologies. Zhang et al [2] studied coal/rock interface recognition by using infrared detection technology. Junli et al [3] investigated coal/rock interface detection and height measurement based on machine vision. Wu and Tian [4] developed coal rock classification and recognition methods based on K-SVD dictionary learning and curvelet domain compressive sensing. Sun et al [5], [6], [7], [8] proposed the coal rock images feature extraction and recognition method that combines wavelet transform and GLCM, GLCM significant clustering features, sparse representation, and BCDTM statistical features. Since coal rock images are generally characterized by weak edges, inhomogeneity, noise pollution, and low contrast, segmentation methods have been proposed by introducing image processing technology based on partial differential equations. The methods include the improved C-V model and improved LBF

The associate editor coordinating the review of this manuscript and approving it for publication was Amin Zehtabian.

model, as well as microscopic damage description methods based on texture features of image gray level co-occurrence matrix [9], [10], [11], [12].

Recently, deep learning techniques have achieved great progress and have shown excellent performance in computer vision applications. Coal rock recognition technologies based on deep learning have been emphasized. Hua et al [13] conducted a preliminary study on the recognition of coal rock using the convolutional neural network (CNN). Tongxing et al [14] investigated the recognition and positioning of coal seam and rock by using the Faster R-CNN method. Bin et al [15] combined the target detection algorithm YOLOv2 with a linear imaging model to recognize and position coal rocks, and the method achieved a recognition success rate of 78%. Lei et al [16], [17], [18] proposed a coal rock image recognition method based on CNN and improved U-Net. Xin et al [19] established a sample generation and feature migration framework based on the Var-Con Sin GAN model. Feng et al [20] proposed an improved lightweight coal gangue recognition method based on the MobileNetV3-largemodule structure. Gao et al [21] improved the coal rock segmentation method based on the tower pooling structure with mixed dilated convolution. Rukundo [22], [23] shows the influence of different sizes of images on the final segmentation accuracy by inputting images of different sizes into the deep learning network model. Considering the continuous and penetrating characteristics of the coal/rock interface, novel indicators were established for the recognition accuracy of the coal/rock interface, and a coal/rock interface recognition method integrating improved YOLOv3 and cubic spline interpolation was proposed to obtain a near-realistic coal/rock interface curve [24]. In the previous study, an intelligent coal rock recognition model was developed by integrating the improved CLBP and receptive field theory [25].

However, the practical mining workface environment is complex due to dark and reflective light, uneven illumination, shading, and shadow conditions, coal dust, mechanical vibrations, and other combined unfavorable factors. This results in low-quality mining workface coal rock images, thus making automatic recognition of coal rocks at the workface highly difficult. By taking encoder-decoder as the basic architecture, this study establishes a model called the fuse attention mechanism's coal rock full-scale network (FAM-CRFSN). The core elements of the proposed model include the full-scale connection structure, multi-channel residual attention module with fused dilated convolution, Res2Block, and multi-dimensional loss function, which help to achieve pixel-level segmentation of low-quality coal rock images of underground coal mines. The model roadmap is shown in the Figure 1.

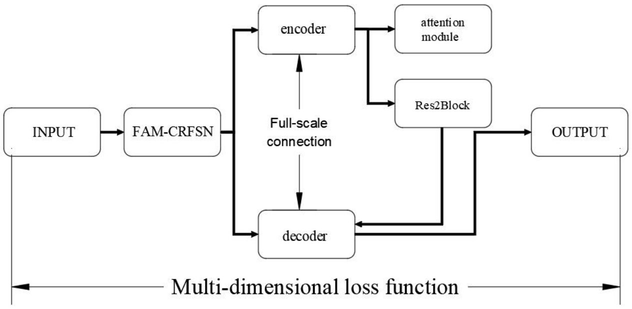


FIGURE 1. FAM-CRFSN road map.

II. THEORY

A. COAL ROCK IMAGES FEATURE ANALYSIS

The practical mining workface environment is complex, with unfavorable lighting conditions such as darkness, reflection, uneven lighting, shading, and shadowing. Meanwhile, coal dust and mechanical vibrations cause image blurring and noise interference, and water dripping changes the characteristics of coal rock areas in the images. Due to these unfavorable factors, mining workface coal rock images often show low-quality features such as low illuminance, weak edges, uneven illumination, low contrast, and severe noise interference [9], [10], [11], which poses a great challenge to the automatic recognition of coal rocks at the mining workface. Fig. 2 shows the image of a typical coal rock. Faced with the above unfavorable factors in underground coal mines, the existing technical solutions often misjudge the coal seam and perform poorly in complex scenarios such as blurred coal rock edge contours and dirt bands, which restricts the intelligent construction of coal mines.

B. STRUCTURE OF THE FAM-CRFSN MODEL

The conventional image processing methods and the current popular deep learning methods usually use some type of "feature representation" [26] to distinguish between coal and rock. The difference between these methods is whether the feature is designed by humans (conventional image processing), or learned automatically by a model (deep learning method).

For coal rock images of underground coal mines, the semantic features of coal and rock are stable, no matter how complex the collecting environment is. This indicates that there must be a deep learning network that can automatically learn semantic features of "coal" and "rock" by gradually abstracting and conceptualizing coal rock features from the bottom-up, thus realizing automatic recognition of coal rock. However, the existing technical solutions cannot effectively handle the unfavorable factors of underground coal mines because their feature representations are either shallow and not semantic features (for conventional image processing measures), or the automatically learned semantic features of "coal" and "rock" are not accurate enough (for deep learning method).

To overcome the difficulty of automatic recognition of coal rock at the mining workface, it is crucial to building a deep learning model that can effectively characterize semantic features of "coal" and "rock" and achieve a balance between

FIGURE 2. Typical coal rock images of underground coal mines.

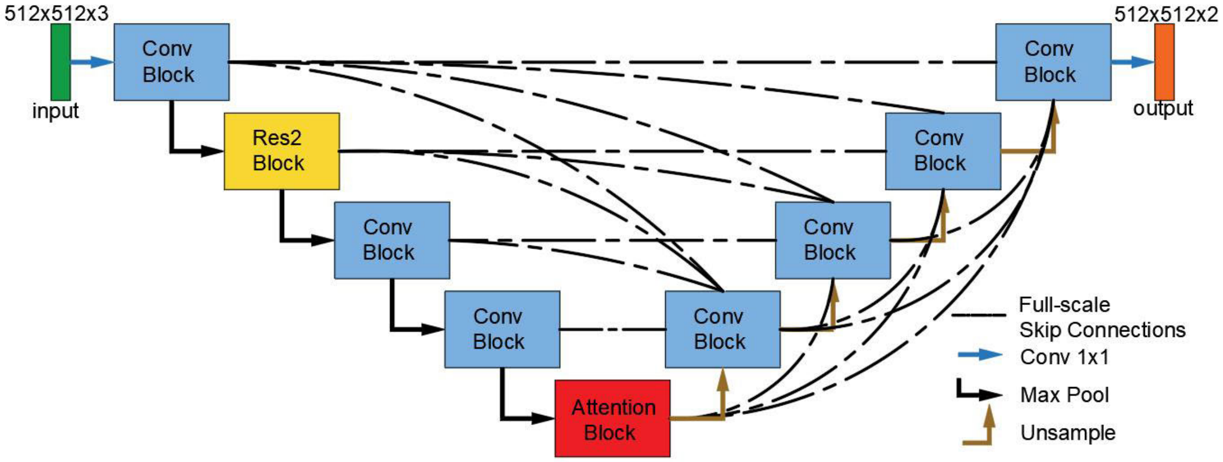


FIGURE 3. The main structure of the FAM-CRFSN mode.

"deep" stacking and error back-propagation. It should be noted that the location features are shallow features, and the stacking "depth" often causes the loss of location feature information although it is beneficial to the learning of coal rock semantic features. How to balance the "deep" stacking with the learning of location features is a problem that must be considered for pixel-level coal rock semantic segmentation. Therefore, Based on Unet3+ [27] model, the FAM-CRFSN model is proposed in this study. The deep extraction of coal rock semantic features is achieved by a multi-channel residual attention mechanism and a full-scale connection structure. Meanwhile, the balance between "deep" stacking and error back-propagation is achieved by structures such as dilated convolution and residual convolution. Besides, automatic learning of deep semantic feature representation and pixel-level segmentation of low-illuminance coal rocks of underground coal mines are achieved through supervised training of deep supervised patterns. As shown in Fig. 3, the FAM-CRFSN model comprises the encoder and the decoder:

(1) The encoder implements deep semantic feature extraction by the convolutional pooling module, the Res2 module, and the multi-channel residual attention module fused with dilated convolution. The Res2 module performs parameter computation and feature extraction at different scales by splitting the feature map channels and dividing them into new dimensions, which effectively increases the feature expression and extraction capability of the network and increases the receptive field range. Meanwhile, the attention module increases the weight of the extracted features at different scales by using multi-channel dilated convolution with different dilation rates, thus better retaining effective features. By encoder, the network obtains a feature map whose size is 1/16 of that of the original image at the highest level.

(2) The decoder fuses and analyzes features of different scales by performing bilinear interpolation upsampling with the convolution module and the full-scale connection structure. The full-scale connection structure fuses features at various scales and acquires semantic information at different depths, thus effectively preventing feature loss in the encoding process. The feature map output by the decoder is finally classified by Softmax and reassociated with each pixel of the original image to achieve pixel-level segmentation on low-illuminance coal rock images.

The core technologies of the proposed FAM-CRFSN model are: (1) adopting a full-scale connection architecture to enhance the extraction of underlying features; (2) adopting a multi-channel residual attention module fused with dilated convolution to reduce noise interference in different receptive fields; (3) adopting the Res2 structure to enhance the extraction of effective features; (4) adopting a multi-dimensional loss function to enhance the accuracy of network training. The above four core technologies will be explained in detail below.

C. FULL-SCALE CONNECTION STRUCTURE

To avoid the loss of shallow features such as location features caused by the increase of network depth, Unet [28] splices the encoder's output feature map of each scale to

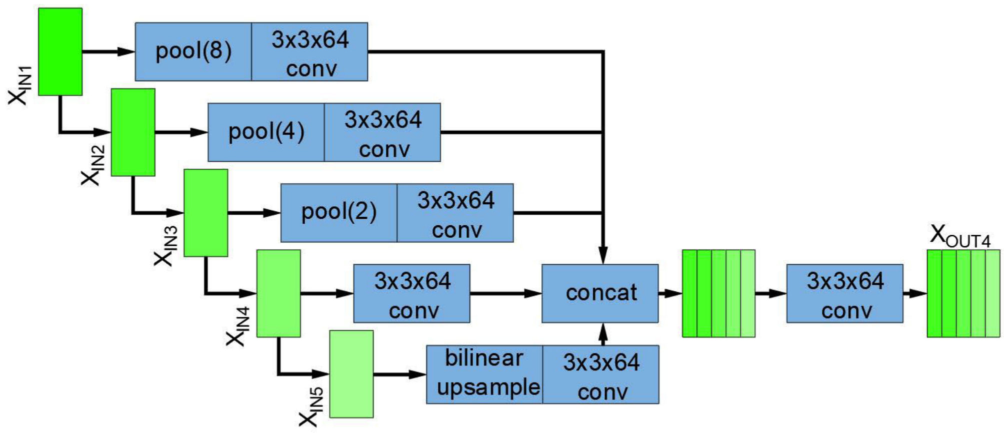


FIGURE 4. Full-scale connection structure.

the corresponding structure of the decoder. This helps to obtain effective high-level features by stacking the depth and preserve the shallow features that would be lost, thus increasing the semantic segmentation accuracy. Since this design spans more network structures, it is called a skip connection structure.

However, the simple skip connection structure cannot realize a full-scale collection of shallow information by the network, which is not conducive to obtaining the precise location and boundary of the coal rock area. For this reason, the proposed FAM-CRFSN model uses a full-scale connection structure, as shown in Fig. 4. In this structure, each module layer of the decoder incorporates the full-scale feature map extracted by the encoder, as well as the feature map from the lower layers of the decoder. In this approach, the decoder can capture full-scale shallow feature information and deep information during decoding operations.

The encoder has five levels of output, which are denoted as X_{INi}, $i \in [1, 5]$. The corresponding five levels of the decoder output are denoted as X_{OUTi}, $i \in [1, 5]$. X_{OUTi} can be calculated by:

$$X_{OUTi} = \begin{cases} X_{INi}, \quad i = 5 \\ H\left(\left[\underbrace{C(D(X_{INk}))_{K=1}^{i-1}, C(X_{INi})}_{Scales:1^{th}\sim i^{th}}, \underbrace{C(U(X_{OUTi}))_{K=i+1}^{N}}_{Scales:i+1^{th}\sim N^{th}}\right]\right), \\ \quad i = 1, \ldots, 4 \end{cases} \tag{1}$$

where $C(\cdot)$ refers to regular convolution operation; $H(\cdot)$ refers to feature aggregation achieved by convolution and splicing; $D(\cdot)$ and $U(\cdot)$ denote upsampling and downsampling operations, respectively.

For X_{OUTi}, the feature maps of each layer are stitched together after the following operations are performed by the skip connection structure:

(1) The feature maps $X_{IN1} \sim X_{IN(i-1)}$ with a size larger than X_{INi} are reduced to the same size as X_{OUTi} by pooling operations;

(2) The feature maps $X_{IN(i+1)} \sim X_{IN5}$ with a size less than X_{INi} are up-dimensioned to the same size as X_{OUTi} by bilinear interpolation upsampling;

(3) The features X_{INi} are further extracted by a 64-channel 3×3 convolution operation.

Due to a large number of feature map channels after splicing, there will be much redundant information. To address this issue, a 3×3 convolution operation with 64 channels is performed on the stitched feature map to achieve feature aggregation, reduce redundancy and make the output feature map size consistent with the number of channels.

D. FUSION DILATED CONVOLUTION'S MULTI-CHANNEL RESIDUAL ATTENTION MODULE

The human visual system tends to focus on the important part of the image and ignore irrelevant information. The attention mechanism help to extract key feature information and reduce the interference of useless information in a way similar to human vision. The Squeeze-and-Excitation Block (SEBlock) [29] automatically obtains the importance of each channel and modifies the weight of each channel during network training, thus improving the extraction capability of valid information and suppressing invalid features.

First, SEBlock uses the global pooling operation to compress each channel of the feature map into a real number $X_c \in R^C$ that represents global feature information:

$$X_c = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} x_{(i,j)} \tag{2}$$

where W and H denote the width and height of the feature map; $x_{(i,j)}$ denotes the grayscale value of the corresponding position in the feature map.

Then, the obtained global features are activated by the ReLU activation function, and the weight $S_C \in [0, 1]$ of global features is adjusted by a gating mechanism in a sigmoid form:

$$S_c = \sigma(W_2\delta(W_1X_c)) \tag{3}$$

where $\delta(\cdot)$ denotes the ReLU activation function, as shown in Eq. (4); $\sigma(\cdot)$ denotes the Sigmoid activation function (see Eq. (5)); W_1 and W_2 denote the fully connected layers added before the activation function. Especially, the two fully connected layers adjust the number of global feature channels to 1/16 of the original number of channels and then revert to the original number of feature channels. The advantage of this scaling approach is that it can reduce the computational effort of network parameters while fusing features between channels.

$$\mathrm{ReLU}(x) = \max(0, x) \tag{4}$$

$$Sigmoid(x) = \frac{1}{1+e^{-x}} \tag{5}$$

In the attention mechanism, the feature vector is multiplied by the original feature map in the channel dimension to obtain the feature map. Meanwhile, to prevent the attention mechanism from losing part of the feature information of the original map, a residual structure is adopted to summarize the original feature map with the obtained result, as shown in Eq. (6).

$$F_c = X_c(1 + S_c) \tag{6}$$

In this way, a residual attention structure is established, as shown in Fig. 5, GAP refers to global average pooling; FC refers to the fully connected layer.

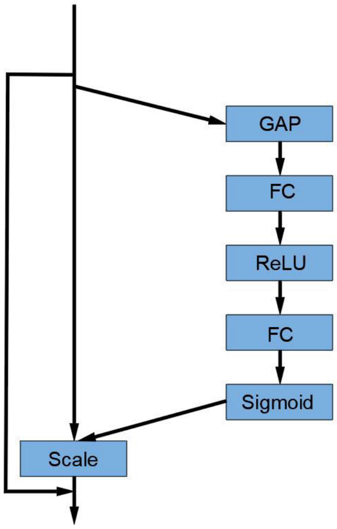


FIGURE 5. Residual attention structure.

SEBlock extracts globally valid features, but it does not suppress local noise interferences well. Meanwhile, the global noise suppression mechanism of SEBlock may lead to the loss of some feature information that should be preserved. To this end, this paper proposes a multi-channel residual attention module fused with dilated convolution to suppress local noise interferences and retain more scale key feature information. The structure of the proposed multi-channel residual attention module is shown in Fig. 6.

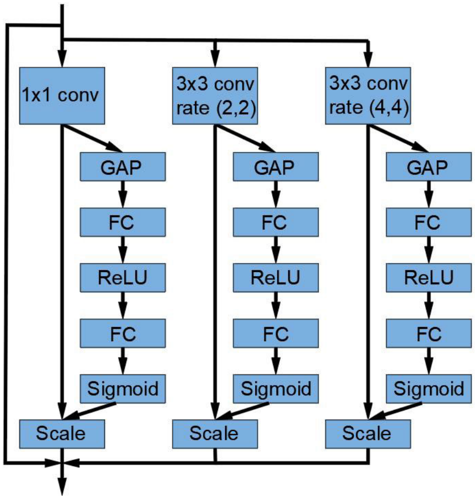


FIGURE 6. Multi-channel residual attention module fused with dilated convolution.

The dilated convolution increases the range of the receptive field without changing the feature map size and thus obtains the semantic information of features on a larger scale. The dilated convolution process with a dilation rate of 2 is shown in Fig. 7.

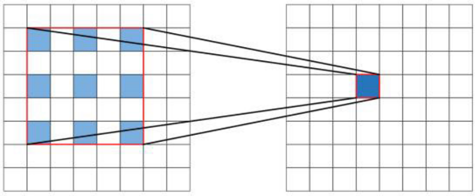


FIGURE 7. Dilated convolution.

Using dilated convolution instead of the normal convolution kernel, the receptive field range of the feature map can be represented as:

$$K_A = K + (K-1) \times (r-1) \tag{7}$$

where K_A denotes the receptive field range of the dilated convolution, K denotes the size of the convolution kernel, and r denotes the dilation rate of the dilated convolution.

In this study, 1×1 convolution, 3×3 dilated convolution with a dilation rate of 2, and 3×3 dilated convolution with a dilation rate of 4 are employed to obtain three sets of attention mechanisms with different observation scales in three parallel channels. In this way, the multi-channel residual attention module can suppress the interference of invalid information at different scales and strengthen effective features by using the dilated convolution with different dilation rates. Meanwhile, the residual structure enhances the validity of shallow features, prevents gradient disappearance of network training, and improves the accuracy and robustness of the overall network model.

E. Res2Block

The residual module [30] can effectively solve the problems such as gradient disappearance and gradient explosion caused by stacking convolution layers to enhance network depth. The structure of the residual module is shown in Fig. 8. It consists of three convolution layers: channel adjustment of the input feature map by 1 × 1 convolution, feature extraction of the feature map by 3 × 3 convolution, and residual connection at the output location and the input location, thus realizing deep feature extraction and preserving shallow features.

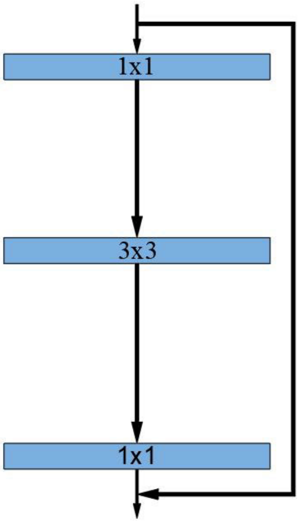


FIGURE 8. The structure of the residual module.

Res2Block is a structure that replaces the 3 × 3 convolution layer in the residual module with multiscale feature extraction, as shown in Fig. 9. It introduces a new dimension called Scale that splits the original channel dimension, and the new feature map after splitting is called 1 Scale. Then, each set of feature maps except the first set of feature maps is convolved, and each set of feature maps after the second set is concatenated with the results after convolution of the previous set of feature maps before convolution:

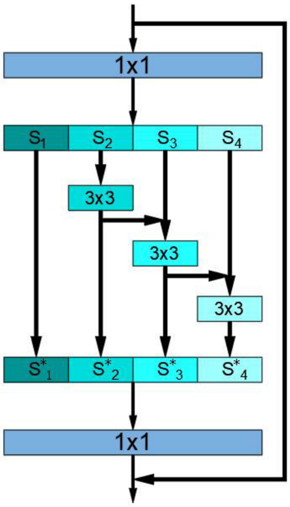


FIGURE 9. Res2 module.

$$S_i^* = \begin{cases} S_i, & i = 1 \\ C_i(S_i), & i = 2 \\ C_i(S_i + S_{i-1}^*), & 2 < i \leq s \end{cases} \tag{8}$$

where S_i denotes the feature map after splitting; S_i^* denotes the feature map result after the convolution operation; $C(\cdot)$ denotes the 3 × 3 convolution layer.

The feature groups output by each scale dimension of Res2Block are convoluted one more time than the output feature groups of the previous scale dimension. Finally, these feature maps with different numbers of convolution operations are spliced and fused to extract semantic information at different scales, which helps the encoder structure to extract coal rock feature information of different depths and significantly increase the receptive field range of the network.

F. MULTI-DIMENSIONAL LOSS FUNCTION

The FAM-CRFSN model proposed in this paper aims to achieve pixel-level segmentation of coal rock images through a network model based on a decoder and encoder. In the segmentation process, the classification of each pixel and the regionality of the coal rock is considered. Due to the characteristics of low illumination, weak edge, uneven illumination, low contrast and serious noise interference in coal-rock images, it is difficult to accurately and effectively realize the recognition of coal-rock images by the simple loss function designed for them. Therefore, the design of loss function is to consider the characteristics of coal-rock images from a multi-dimensional perspective, and establish a multi-dimensional loss function including pixel level, region level and image level. Therefore, this paper uses the sum of the cross-entropy loss (CE loss), intersection over union loss (IOU loss), and multiscale structural similarity loss (MS-SSIM loss) as the loss function of the FAM-CRFSN model.

The cross-entropy loss characterizes the overall error of the image, and it can be expressed as:

$$l_{CE} = -\sum_{i=1}^{N} p(x_i) \log_a q(x_i) \tag{9}$$

where N is the number of categories; i is the category sequence number; $p(x_i)$ is the classification target of the actual real value. It equals to 1 if it corresponds to the target classification, or 0 otherwise; $q(x_i)$ is the predicted probability value; the logarithmic base a can be taken as e without special assertion.

Since the segmentation target of coal rock images is only "coal" and "rock", the cross-entropy loss can be simplified as:

$$l_{CE} = -|p(x_n)\log_a q(x_n) + (1-p(x_n))\ln(1-q(x_n))| \qquad (10)$$

The IOU loss characterizes pixel-level image errors, and it can be expressed as:

$$l_{Iou} = 1 - \sum_{i=1}^{N} \frac{p_{ii}}{\sum_{j=0}^{N} p_{ij} + \sum_{j=0}^{N} p_{ji} - p_{ii}} \qquad (11)$$

where p_{ii} denotes a correctly classified pixel; p_{ij} denotes a pixel that belongs to class i but is classified in class j; p_{ji} denotes a pixel that belongs to class j but is classified in class i.

The MS-SSIM loss characterizes the errors within different regions, and it can be expressed as:

$$l_{MS-SSIM} = 1 - \prod_{m-1}^{M}\left(\frac{2\mu_p\mu_g + C_1}{\mu_p^2 + \mu_g^2 + C_1}\right)^{\beta_m} \times \left(\frac{2\sigma_{pg} + C_2}{\sigma_p^2 + \sigma_g^2 + C_2}\right)^{\gamma_m} \qquad (12)$$

where M denotes the total number of scales of the network model, and the FAM-CRFSN model proposed in this paper contains five scales of feature representation; μ_p, μ_g and σ_p, σ_g denote the mean and variance of the predicted and true values, respectively; σ_{pg} denotes their covariance; β_m and γ_m denotes the importance of each scale; C_1 and C_2 are two constant quantities to prevent division by zero in the equation, and they are usually set to 0.01 and 0.03, respectively.

In this way, the loss function of the FAM-CRFSN model is a multi-dimensional loss with pixel-level, area-level, and image-level expressions:

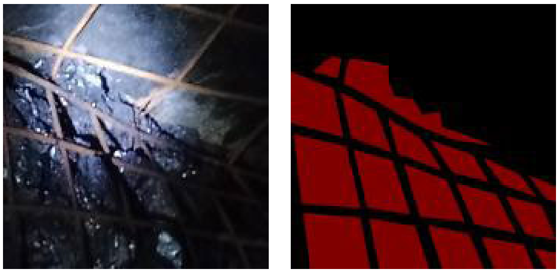
$$l_{oss} = l_{CE} + l_{Iou} + l_{MS-SSIM} \qquad (13)$$

III. EXPERIMENTAL RESULTS AND ANALYSIS

The validity and superiority of the proposed FAM-CRFSN model are verified by using coal rock images with low illuminance, weak edges, and other characteristics collected from the underground coal mining workface. The models taken for comparison are Deeplab [31], PSPNet [32], Unet, SegNet [33], hrnet [34]and other major semantic segmentation network models. To ensure that the test results of different network models are not affected by factors other than model differences, the following experiments are conducted for all network models, and the parameter settings are consistent.

A. DATASETS OF COAL ROCK IMAGES OF UNDERGROUND COAL MINES

The 500 coal rock original images used in this study were collected from underground coal mines in Shanxi, Chongqing, Sichuan, and Yunnan. To facilitate the training and testing of the deep learning models, the original images were labeled by the Labelme software, and labeled images of the same size as the original images were obtained, as shown in Fig. 10.



(a) original images; (b) labeled images

FIGURE 10. Dataset label.

To prevent over-fitting during the training of the deep learning model and improve the generalization performance of the model, it can be expanded by expanding the dataset, the above dataset was enhanced by adding salt-and-pepper noise and Gaussian noise, rotating the image, mirroring and flipping the image, and adjusting the contrast and brightness of the image (see Fig. 11) [35], [36], [37]. The original coal rock images are rotated by 60 degrees, 120 degrees, 180 degrees, 240 degrees and 300 degrees to obtain a total of 2500 images. The raw coal rock images are added with Gaussian noise and salt and pepper noise to obtain a total of 1000 images. The raw coal rock images are mirrored to obtain 500 images. The images are enhanced by brightness and contrast to obtain a total of 1000 images. Finally, a total of 5500 images are obtained after adding 500 original images to form the original image of the data set. The original coal and rock image's masks are rotated by 60 degrees, 120 degrees, 180 degrees, 240 degrees and 300 degrees respectively to obtain a total of 2500 coal and rock images corresponding to the rotated coal and rock images. The original coal and rock image masks are mirrored to obtain 500 coal and rock images masks corresponding to the mirrored coal and rock images. Other modifications are marked for replication but do not add modifications to obtain a total of 2000 annotations. Finally, a total of 5500 masks are obtained after adding 500 original image annotations to form a data set mask.

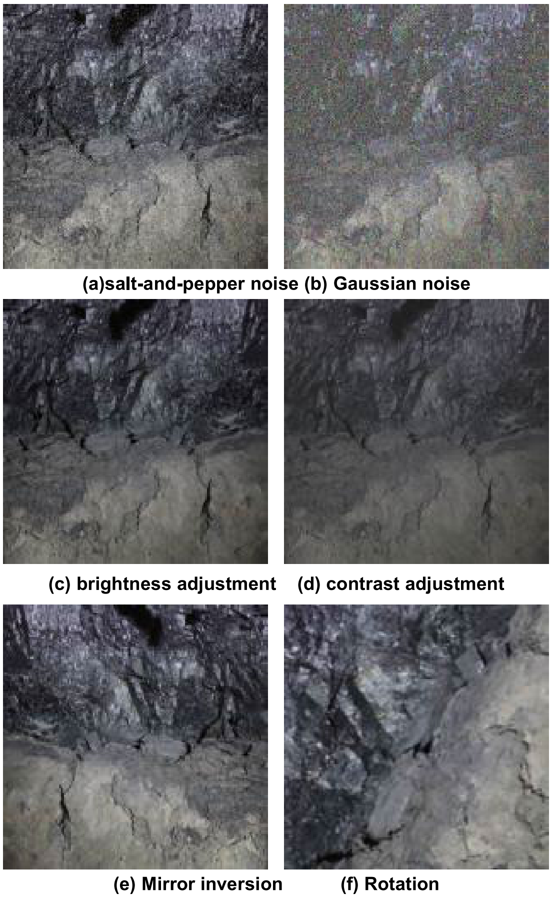
B. MODEL LEARNING AND TRAINING

1) TRAINING ENVIRONMENT

In the hardware equipment used in this paper, the CPU is AMD Ryzen 7 4800H, the GPU is GeForce GTX 1650 Ti, the computer memory size is 4GB, and the computer memory size is 16GB. In the software environment, the Cuda version number is 10.1, the Cudnn version number is 7.4.1, the deep learning framework uses Tenserflow-GPU 2.2.0 version, the programming language is Python 3.6 version, the compiler environment is Pycharm, the related libraries also include Numpy 1.92.2 and Pillow 8.2.0, etc.

2) OPTIMIZED ALGORITHM DESIGN

To increase the training accuracy without increasing the computation time significantly, this study selected the adaptive



(a)salt-and-pepper noise (b) Gaussian noise

(c) brightness adjustment (d) contrast adjustment

(e) Mirror inversion (f) Rotation

FIGURE 11. Dataset expansion effect.

moment estimation algorithm Adam [38] as the optimization algorithm. Adam assigns an adaptive learning rate to different parameters by gradient first-order moment estimation mean and second-order moment estimation mean.

3) DEEP SUPERVISION

The problem of gradient disappearance or gradient explosion caused by too deep layers makes the model difficult to train and converge. The avoid this problem, this paper adopts the deep supervision method [39] to guide the training of the deep learning network. Different from the conventional methods that only supervise at the output location, the deep supervision method direct supervises the middle layer and performs error reversal. Meanwhile, the loss function can be considered as a soft constraint added to the deep learning process.

A 3 × 3 convolution layer is added to the output layer of each decoder of the FAM-CRFSN model; then, a feature map of the same size as the original input image can be obtained by bilinear interpolation upsampling; finally, the segmentation results are obtained by Softmax. Based on the above training requirements, the loss function values of the FAM-CRFSN model are trained 50 iterations with the Adam optimization algorithm, and the results are shown in Fig. 12.

```
Adam
Require: the learning rate $\eta$ (Suggested default:$10^{-4}$)
Require: the minuscule constant parameter $\varepsilon$ (Suggested default:$10^{-8}$)
Require: the first-order moments $\widehat{m_t}$; the second-order moments $\widehat{v_t}$
  (Suggested default:0 and 0)
Require: the decay rates $\beta_1$ and $\beta_2$ (Suggested default:0.9 and 0.999)
Require: network parameter $\theta$;
  the time step $t$ (Suggested default:0)
  while stopping criterion not met do
    Sampling $m$ data $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$ from the training set and its
    corresponding label $\{y^{(1)}, y^{(2)}, \ldots, y^{(m)}\}$.
    Calculating gradient values: $g(\theta) \leftarrow \frac{\delta(\frac{1}{m}\sum_{i=1}^{m} Lf((x^{(1)}),y^{(1)}))}{\delta\theta}$.
    $t \leftarrow t+1$.
    Update first-order moments: $m_t \leftarrow \beta_1 m_{t-1} + (1-\beta_1)g(\theta)$.
    Update second-order moments: $v_t \leftarrow \beta_2 v_{t-1} + (1-\beta_2)g^2(\theta)$.
    Correct first-order moments: $\widehat{m_t} \leftarrow \frac{m_t}{1-\beta_1^t}$.
    Correct second-order moments: $\widehat{v_t} \leftarrow \frac{v_t}{1-\beta_2^t}$.
    Calculate parameter update amount: $\Delta\theta \leftarrow -\frac{\eta}{\sqrt{\widehat{v_t}}+\varepsilon} \odot \widehat{m_t}$.
    Update parameters: $\theta_{t+1} \leftarrow \theta_t + \Delta\theta$.
  end while
```

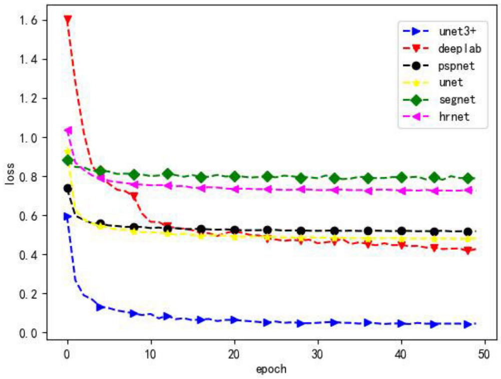


FIGURE 12. Training effect of network model.

From Figure 11, it can be seen that under the loss function mentioned above, the FAM-CRFSN network model proposed in this paper converges faster than other more common network models, and the loss function value converges to a lower range, indicating that the FAM-CRFSN network model proposed in this paper under this loss function is more accurate in segmenting coal and rock images.

C. ABLATION EXPERIMENT

In order to prove the improvement effect of the different improvement schemes mentioned in this paper on the original semantic segmentation model, this paper conducts an ablation experiment, and the experimental results are shown in Table 1.

It can be seen from Table 1 that compared with experimental group 1, experimental group 2 replaces a structure in the encoder with Res2Block, which improves the segmentation accuracy of the network on the premise of a small increase in the number of overall parameters. Compared with the experimental group 2, on the basis of the improvement of the experimental group 2, the convolution layer is replaced with a multi-channel residual attention module that

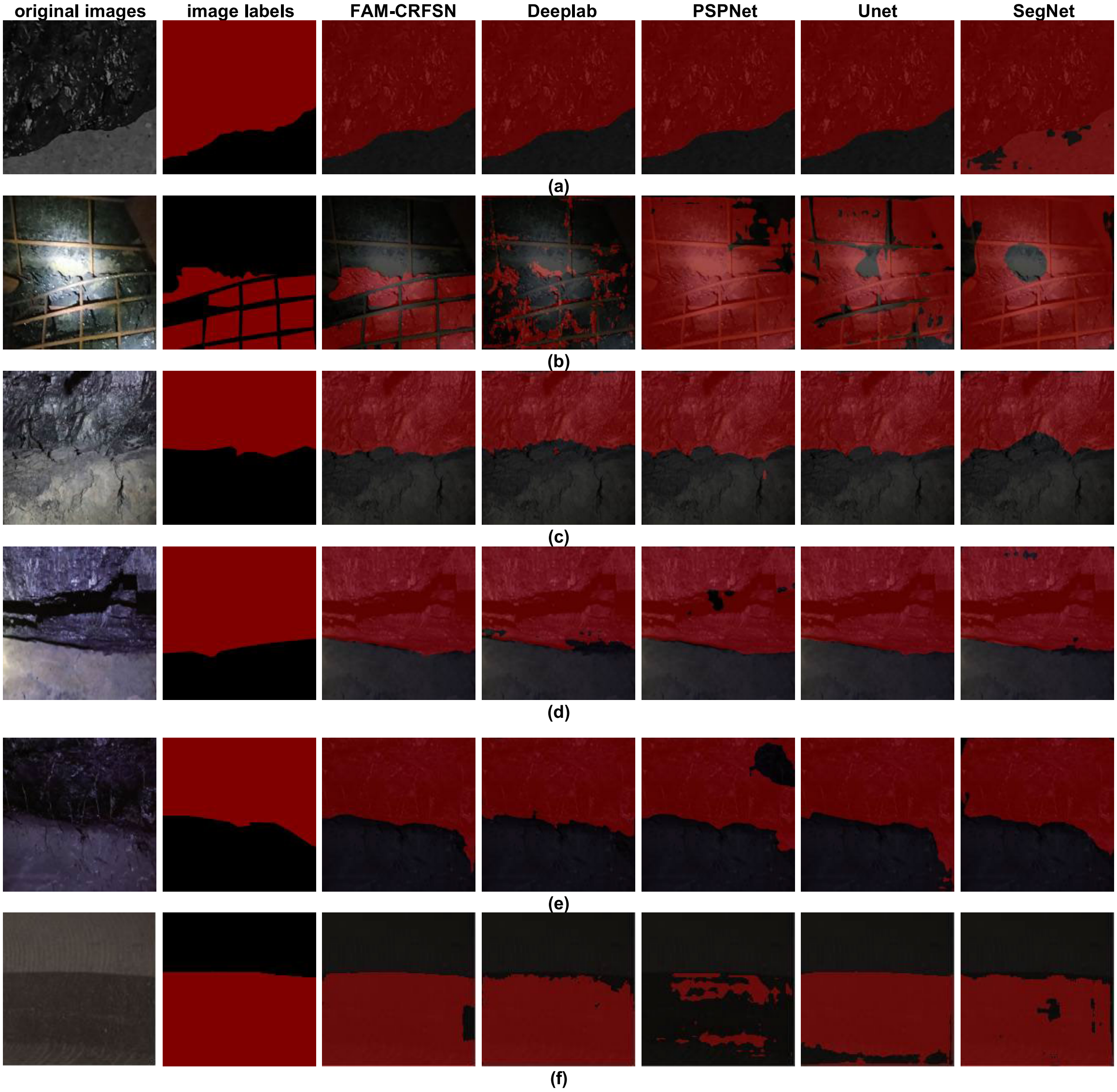


FIGURE 13. Comparison of the prediction effect of different network models.

TABLE 1. Ablation experimental results.

experimental group	1	2	3	4	5
Unet	◎	◎	◎	◎	◎
Full-scale connection				◎	◎
Res2Block		◎	◎		◎
attention module			◎		◎
MIOU	95.78	96.42	97.01	96.55	97.99
MPA	97.75	97.95	98.47	98.20	99.28
parameters/10^6	24.89	25.83	23.87	25.36	24.57

integrates the dilated convolution at the connection between the encoder and the decoder. As a result, the segmentation accuracy of the network is improved under the premise of decreasing the number of overall parameters. Compared with the experimental group 1, the skip structure in the Unet network was replaced by a full-scale connection structure in the experimental group 4, which improved the segmentation accuracy of the network while slightly increasing the number of parameters of the network model. The experimental group 5 is combined with all the improved network models mentioned above, that is, the FAM-CRFSN network model proposed in this paper. This group has achieved the optimal value in each evaluation index result, and has not significantly increased the number of parameters. There is no significant increase in network operation speed, and the task of coal rock image recognition can be well completed in a certain noise signal environment. In summary, the improved scheme added in this paper effectively increases the segmentation accuracy of coal-rock images under the premise of slightly reducing the overall parameters of the network model.

D. EXPERIMENTAL RESULTS AND ANALYSIS

1) SUBJECTIVE ANALYSIS

900 low-illuminance coal rock images of underground coal mines were used for testing in this study. The proposed FAM-CRFSN model, Deeplab, PSPNet, Unet, and SegNet were trained on the same training set, and then they were tested on the testing set for the semantic segmentation task. Due to the limitation of space, six representative images are selected for presentation. As shown in Fig. 13, from the left to right are original images, the recognition result by FAM-CRFSN, Deeplab, PSPNet, Unet, and SegNet, respectively. For the convenience of the display, the coal seam area in the image is marked in red, and the rock and background areas are marked in black.

As shown in Fig. 13, the proposed FAM-CRFSN network model obtains excellent results for the segmenting coal rock images of underground coal mines with low-quality features such as low illumination, weak edges, inhomogeneity, and severe noise interference. Also, it accurately fits the interface between the coal seam and rock. Other network models are not robust in the practical environment of underground coal mines due to their failure, mis-segmentation, over-segmentation, and large segmentation boundary errors for low-quality coal rock images of underground coal mines.

2) OBJECTIVE ANALYSIS

In this study, mean intersection over union (MIOU) and category mean pixel accuracy (MPA) are selected to quantitatively analyze the performance of different models.

IOU is the ratio of the intersection and union of the real value and the predicted value, which represents the overlap between the predicted value and the real value. The MIOU is the average value of IOU of each classification in the whole situation, which represents the classification accuracy under all classifications more effectively:

$$MIOU = \frac{1}{N+1}\sum_{i=1}^{N}\frac{p_{ii}}{\sum_{j=0}^{N}p_{ij}+\sum_{j=0}^{N}p_{ji}-p_{ii}} \tag{14}$$

where N is the number of categories; p_{ii} represents the correctly classified pixel; p_{ij} represents the pixel that belongs to class i but is classified in class j; p_{ji} represents the pixel that belongs to class j but is classified in class i.

PA indicates the number of correctly classified pixels in each category as a percentage of the overall number of pixels, and MPA indicates the global average of the classified PAs:

$$MPA = \frac{1}{N+1}\frac{\sum_{i=0}^{N}p_{ii}}{\sum_{i=0}^{N}p_{i}} \tag{15}$$

where p_i represents all pixels of the corresponding category.

To better verify that the FAM-CRFSN network model has a strong ability to recognize the low illumination coal rock images in the complex environment of coal mines, the coal rock images from the real underground environment are added with different proportions of Gaussian noise and salt-and-pepper noise to simulate the effects of unfavorable factors in underground coal mines. Then, the performance of FAM-CRFSN, Deeplab, PSPNet, Unet, and SegNet network models were evaluated on the test sets with different noise intensities, and the MIOU and MPA results are presented in Table 2 and Table 3. The parameters are shown in Table 4.

TABLE 2. MIOU.

Noise ratio	MIOU					
	FAM-CRFSN	Deeplab	Pspnet	Unet	Segnet	hrnet
0	**98.95**	98.90	98.71	98.77	93.39	91.72
0.01	**98.53**	98.02	97.00	97.69	85.09	90.51
0.02	**97.99**	96.14	94.23	95.78	67.65	89.95
0.03	**97.28**	93.40	89.61	92.82	50.44	89.09
0.04	**96.18**	90.42	83.68	89.04	39.75	87.45
0.05	**94.74**	87.62	76.46	84.71	34.93	85.21
0.06	**93.13**	85.10	68.69	80.25	32.15	82.56
0.07	**91.31**	82.49	60.80	75.85	30.56	79.76
0.08	**88.78**	80.02	54.86	71.02	29.59	77.10
0.09	**85.77**	77.08	50.96	65.99	28.82	73.56

TABLE 3. MPA.

Noise ratio	MPA					
	FAM-CRFSN	Deeplab	Pspnet	Unet	Segnet	hrnet
0	**99.48**	99.45	99.35	99.38	96.68	96.03
0.01	**99.28**	99.04	98.43	98.81	91.44	95.42
0.02	**99.01**	98.12	96.85	97.75	80.09	95.13
0.03	**98.65**	96.68	94.15	96.06	67.8	94.68
0.04	**98.07**	95.02	90.55	93.84	59.7	93.79
0.05	**97.31**	93.39	85.96	91.22	55.94	92.55
0.06	**96.43**	91.86	80.80	88.43	53.75	91.00
0.07	**95.40**	90.23	75.33	85.59	52.79	89.28
0.08	**93.92**	88.62	71.07	82.39	51.72	87.55
0.09	**92.12**	86.68	68.21	78.95	51.10	85.15

TABLE 4. Parameters.

Model	FAM-CRFSN	Deeplab	Pspnet	Unet	Segnet	hrnet
parameters $/10^6$	24.57	2.75	2.44	24.89	5.54	9.67

As shown in Table 2 and Table 3, at a low noise intensity, the FAM-CRFSN model achieves higher segmentation accuracy than other networks; as the noise signal increases, the accuracy of the FAM-CRFSN model is much better than that of other networks; when the noise intensity reaches 0.09, the FAM-CRFSN model still achieves an MIOU of 85.77% and an MPA of 92.12%. Thus, the proposed FAM-CRFSN network model has strong feature extraction and recognition capability for low-illuminance coal rock images in the complex environment of underground coal mines.

IV. CONCLUSION

Aiming at the common characteristics of coal rock images such as low illumination, low contrast, and serious noise interference caused by the complex environment of underground coal mines, this paper proposes a low illumination coal rock image segmentation network model called FAM-CRFSN. The deep extraction of coal rock semantic features is achieved by a multi-channel residual attention mechanism and a full-scale connection structure. Meanwhile, the balance between "deep" stacking and error back-propagation is achieved by structures such as dilated convolution and Res2Block. Besides, a multi-dimensional loss function consists of the cross-entropy loss, IOU loss, and MS-SSIM loss with pixel-level, area-level, and image-level expressions is established, and the network model is trained by deep supervision to achieve automatic learning of deep semantic feature representations. In this approach, the pixel-level segmentation of low-illuminance coal rock images of underground coal mines is achieved.

The performance of the FAM-CRFSN network is tested by adding different proportions of noise signals to practical images to simulate the effects of unfavorable factors in underground coal mines. The experimental results indicate that FAM-CRFSN can accurately fit the interface between the coal seam and rock; at a noise intensity of 0.09, it achieves an MIOU of 85.77% and an MPA of 92.12%. Also, FAM-CRFSN has much better accuracy and generalization performance than mainstream semantic segmentation network models. The results show that the proposed FAM-CRFSN model can effectively segmentize coal rock images with low-quality characteristics such as low illuminance and weak edges caused by the complex environment of underground coal mines.

However, the network model proposed in this paper is completely based on CNN, which leads to a large number of parameters and operation time. In the future, a certain pruning scheme can be considered in practical application and the Transformer structure can be used instead of CNN structure to increase the operation efficiency.

ACKNOWLEDGMENT

The authors are grateful for the support of these projects. They would also like to thank all anonymous reviewers for their constructive comments.

REFERENCES

[1] G. Wang, F. Liu, Y. Pang, H. Ren, and Y. Ma, "Coal mine intellectualization: The core technology of high quality development," *J. China Coal Soc.*, vol. 44, no. 2, pp. 349–357, 2019.

[2] Q. Zhang, H. J. Wang, Z. Wang, and X. Z. Wen, "Analysis of coal—Rock's cutting characteristics and flash temperature of peak based on infrared thermal image testing," *J. Chin. J. Sens. Actuators*, vol. 29, no. 5, pp. 686–692, 2016.

[3] J. Liu, L. Wang, and X. Kong, "Research on coal-rock interface recognition and location measurement," *J. Comput. Eng. Appl.*, vol. 53, no. 8, pp. 246–249, 2017.

[4] Y. Wu and Y. Tian, "Method of coal-rock image feature extraction and recognition based on dictionary learning," *J. China Coal Soc.*, vol. 41, no. 12, pp. 3190–3196, Dec. 2016.

[5] J. Sun and J. She, "Coal-rock image feature extraction and recognition based on support vector machine," *J. China Coal Soc.*, vol. 38, no. S2, pp. 508–512, 2013.

[6] J. Sun and B. Su, "Coal-rock interface detection using cluster prominence based on gray level co-occurrence matrices," *J. Adv. Inf. Sci. Service Sci.*, vol. 4, no. 8, pp. 353–360, 2012.

[7] J. Sun and B. Chen, "Coal-rock recognition approach based on CLBP and support vector guided dictionary learning," *J. China Coal Soc.*, vol. 42, no. 12, pp. 3338–3348, 2017.

[8] J. Sun and B. Chen, "An approach to coal-rock recognition via statistical modeling in dual-tree complex wavelet domain," *J. China Coal Soc.*, vol. 41, no. 7, pp. 1847–1858, 2016.

[9] S. G. Cao, C. M. Sun, P. Guo, F. Luo, and Y. B. Liu, "Image processing and its applications of meso-crack of coal based on modified C-V model," *J. Chin. J. Rock Mech. Eng.*, vol. 34, no. S1, pp. 3074–3081, 2015.

[10] C. Sun, S. Cao, and Y. Li, "Investigation on meso-structure of coal and rock based on the modified LBF model," *J. China Coal Soc.*, vol. 40, no. 2, pp. 331–341, 2015.

[11] C. Sun, *Study on Digital Image Processing and Mesoscopic Damage Constitutive Model of Coal*. Beijing, China: China Atomic Energy Press, 2018, pp. 19–21.

[12] C. Sun, S. Cao, and Y. Li, "Mesomechanics coal experiment and an elastic-brittle damage model based on texture features," *Int. J. Mining Sci. Technol.*, vol. 28, no. 4, pp. 639–647, Jul. 2018.

[13] H. Zhang, J. Wang, X. P. Huang, J. Jin, and J. J. Shan, "Coal rock recognition based on convolutional neural network," *J. Suihua Univ.*, vol. 38, no. 12, pp. 151–153, 2018.

[14] T. Hua, C. E. Xing, and L. Zhao, "Recognition of coal rock and positioning measurement of coal seam based on faster R-CNN," *J. Mining Process. Equip.*, vol. 47, no. 8, pp. 4–9, 2019.

[15] B. Zhang, X. G. Su, Z. X. Duan, L. Z. Chang, and F. Z. Wang, "Application of YOLOv2 in intelligent recognition and location of coal and rock," *J. Mining Strata Control Eng.*, vol. 2, no. 2, pp. 94–101, 2020.

[16] L. Si, Z. B. Wang, X. X. Xiong, and C. Tan, "Coal and rock identification method of fully mechanized mining face based on improved U-Net network model," *J. China Coal Soc.*, vol. 46, no. S1, pp. 578–589, 2021.

[17] L. Si, X. X. Xiong, Z. B. Wang, and C. Tan, "A deep convolutional neural network model for intelligent discrimination between coal and rocks in coal mining face," *Math. Problems Eng.*, vol. 2020, pp. 1–12, Mar. 2020.

[18] X. X. Xiong, "Research on coal rock identification method of fully mechanized mining face based on deep learning," *J. China Univ. Mining Technol.*, pp. 42–55, May 2020.

[19] X. Wang, F. Gao, J. Chen, P. C. Hao, and Z. J. Jing, "Sample generation method of coal and rock image based on GAN network," *J. China Coal Soc.*, vol. 46, no. 9, pp. 3066–3078, 2021.

[20] Y. S. Gao, B. Q. Zhang, and L. Y. Lang, "Technology and implementation of coal gangue recognition based on deep learning," *J. Coal Sci. Technol.*, vol. 49, no. 12, pp. 202–208, 2021.

[21] F. Gao, X. Yin, Q. Liu, X. Huang, Y. Bo, Q. Zhang, and X. Wang, "Coal-rock image recognition method for mining and heading face based on spatial pyramid pooling structure," *J. China Coal Soc.*, vol. 46, no. 12, pp. 4088–4102, 2021.

[22] O. Rukundo, "Effects of image size on deep learning," *Electronics*, vol. 12, no. 4, p. 985, 2023.

[23] O. Rukundo, "Evaluation of extra pixel interpolation with mask processing for medical image segmentation with deep learning," *arXiv*, vol. abs/2302.11522, pp. 1–4, Feb. 2023.

[24] C. M. Sun, Y. P. Wang, C. Wang, R. J. Xu, and X. E. Li, "Coal rock interface recognition method based on improved YOLOv3 and cubic spline interpolation," *J. Mining Strata Control Eng.*, vol. 4, no. 1, pp. 81–90, 2022.

[25] C. M. Sun, R. J. Xu, C. Wang, T. H. Ma, and J. X. Chen, "Coal rock image recognition method based on improved CLBP and receptive field theory," *Deep Underground Sci and Eng.*, vol. 1, no. 2, pp. 165–173, 2022.

[26] H. M. Huang, L. F. Lin, R. F. Tong, H. J. Hu, Q. W. Zhang, Y. Iwamoto, X. H. Han, Y. W. Chen, and J. Wu, "UNet 3+: A full-scale connected unet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, 2020, pp. 1055–1059.

[27] X. P. Qiu, *Neural Networks and Deep Learning*. Beijing, China: Machinery Industry Press, 2020.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, in Lecture Notes in Computer Science, vol. 9351. 2015, pp. 234–241.

[29] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[30] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[31] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[32] H. S. Zhao, J. P. Shi, X. J. Qi, X. G. Wang, and J. Y. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2881–2890.

[33] B. Vijay, H. Ankur, and C. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," *arXiv*, vol. abs/1505.07293, pp. 1–10, May 2015.
[34] D. Fu and W. Wu, "High-resolution representation learning for human pose estimation based on transformer," *J. Phys., Conf. Ser.*, vol. 2189, no. 1, 2022, Art. no. 012023.
[35] X. Li, S. Fang, L. Zeng, Y. Chai, Q. Han, L. Ye, and D. Chen, "An intelligent vehicle-oriented EMC fault dataset augmentation and validity verification method," *Proc. SPIE*, vol. 12451, Oct. 2022, Art. no. 124514W.
[36] A. Bożko and L. Ambroziak, "Influence of insufficient dataset augmentation on IoU and detection threshold in CNN training for object detection on aerial images," *Sensors*, vol. 22, no. 23, p. 9080, Nov. 2022.
[37] C. Hüter, X. Yin, T. Vo, and S. Braun, "A pragmatic dataset augmentation approach for transformation temperature prediction in steels," *Comput. Mater. Sci.*, vol. 176, Apr. 2020, Art. no. 109488.
[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Comput. Sci.*, pp. 1–15, Dec. 2014. [Online]. Available: https://arxiv.org/abs/1412.6980
[39] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *Proc. 18th Int. Conf. Artif. Intell. Statist.*, 2014, pp. 562–570.

SUN CHUANMENG received the B.Sc. and Ph.D. degrees from Chongqing University, in 2010 and 2015, respectively. He is currently a Lecturer with the North University of China. His main research interests include deep learning, computer vision, and information acquisition of various transient processes in harsh environments, such as high voltage, high shock, high temperature, high speed, and strong electromagnetic interference.

LI XINYU received the B.Sc. degree from Shijiazhuang Tiedao University, in 2019. He is currently pursuing the master's degree with the North University of China. His main research interests include deep learning and semantic segmentation.

CHEN JIAXIN received the bachelor's degree from Jinzhong University, in 2020. He is currently pursuing the master's degree with the School of Electrical and Control Engineering, North University of China. His main research interests include deep learning and image processing.

WU ZHIBO received the B.Sc. and Ph.D. degrees from the North University of China, in 2011 and 2020, respectively. He is currently a Lecturer with the North University of China. His main research interests include dynamic test and intelligent instrument.

LI YONG received the B.Sc. degree from Chongqing University, in 2007, and the Ph.D. degree from the University of Bologna, in 2013. He is currently an Associate Professor with Chongqing University. His main research interests include structural engineering and hydraulics.