Identification of Ground Fissures in Mining Areas from UAV Images Based on RDC-UNet

Zhu Huashan
China National Administration of Coal Geology 173 Exploration Team,
Zhuozhou Hebei 072750, China
510577925@qq.com

***Abstract*—As coal mining deepens, ground fissures in mining areas pose significant risks to safety and the environment. Traditional geological exploration methods are inefficient and costly, making the precise detection of large-scale fissures difficult. Deep learning methods using unmanned aerial vehicle (UAV) images have become a popular approach for fissure detection, although challenges such as insufficient accuracy and poor adaptability in complex backgrounds remain. This study proposes an residual-depthwise separable convolution UNet (RDC-UNet) model to address these issues, building on U-Net by incorporating residual connections (RC), depthwise separable convolutions (DSC), and the convolutional block attention module (CBAM) attention mechanism. The model was trained on a dataset of 300 UAV images from a DJI Mavic 3e and outperformed the benchmark models, achieving 86.12 % mPrecision, 75.01 % mRecall, 70.72 % mIoU, and 0.7951 mF1. Ablation experiments show that removing any core module leads to a performance drop, with the DSC module reducing mF1 by 3.28 %, CBAM decreasing mIoU by 2.67 %, and RC lowering mRecall by 3.96 %. RDC-UNet is highly efficient, requiring only 7.8 million parameters, with an inference time of 75 milliseconds and a memory footprint of 180 MB, much lower than other models. This makes it well-suited for real-time UAV-based fissure monitoring. The RDC-UNet model offers high accuracy and efficiency, making it an ideal solution for cost-effective real-time monitoring of ground fissures in mining areas.**

***Index Terms*—Mine ground fissures; UAV images; U-Net; Convolutional block attention module; Depthwise separable convolution.**

I. INTRODUCTION

Coal, as one of the key energy sources in the global energy system, plays a pivotal role. However, while coal mining promotes economic growth, it also has many negative impacts on mining areas and their surrounding environments. Among these, the problem of ground fissures is particularly prominent. Ground fissures will not only damage the ground facilities and buildings in the mining areas, but also pose threats to the surrounding ecological environment and the daily lives of residents. They can lead to a series of problems, such as the degradation of land resources, the loss of water resources, and the destruction of vegetation, thus threatening the ecological balance and sustainable development of mining areas [1]–[3]. Therefore, accurately identifying ground fissures in mining areas is of extremely important practical significance for disaster prevention and mitigation, reducing economic losses, and protecting the ecological environment. It is a key measure to ensure the sustainable development of mining areas.

In the detection and identification of ground fissures in mining areas, although traditional geophysical exploration techniques can provide certain crack information, they have many limitations [4]. The implementation of these techniques requires a large amount of manpower and material resources, with high costs and low efficiency. It is difficult to achieve effective monitoring of large-scale mining areas, especially when it comes to the detection ability of fine fissures, which is often insufficient [5]–[7]. Satellite remote sensing technology, due to its wide coverage area, rich spectral information, and convenient access methods, has brought new possibilities for the detection of ground fissures [8]–[12]. However, due to the limitations of its timeliness and spatial resolution, it is difficult to accurately identify fine fissures and it is unable to capture the dynamic change characteristics of ground fissures in mining areas in a timely manner. To some extent, this restricts its application in the fine detection of ground fissures in mining areas [13].

With the continuous progress of science and technology, the rise of unmanned aerial vehicle (UAV) technology has brought new opportunities for the identification of surface fissures [14], [15]. UAV have the advantages of convenient operation, low cost, and high precision in data acquisition. Their resolution can reach the centimeter level, meeting the requirements of fine detection of surface fissures. The application of UAV remote sensing technology has significantly improved the speed and comprehensiveness of obtaining fissure information, making image data of the collection in large areas more efficient and greatly enhancing work efficiency [16]–[18].

Meanwhile, deep learning technology has achieved breakthrough progress in many fields in recent years, especially demonstrating huge potential in image recognition and detection. The ground fissure identification method based on UAV images and deep learning has emerged as the times require and has become a hot topic in current research. This method breaks through the limitations of traditional manual detection and simple image analysis, significantly improving detection efficiency and accuracy.

For example, Ding, Yang, Yu, and Shu [19] proposed a fissure detection and quantification framework for concrete

Manuscript received 4 January, 2025; accepted 4 March, 2025.

structures using a UAV and transformer-based method. Gao Cao, Cai, and Zhou [20] proposed a pixel-by-pixel road fissures detection network ARD-UNet for UAV remote sensing images. Chen, Liu, Chen, Zhu, Zhang, and Wang [21] proposed a UAV inspection system for automatic detection of road fissures. Duan, Liu, Ling, Zhang, and Liu [22] proposed a real-time edge reconstruction fissure detection network (ERNet) to achieve high precision and rapid fissure edge segmentation through edge reconstruction. However, in the complex background environment of mining areas, these technologies still face many challenges, such as insufficient detection accuracy, poor model adaptability, and other problems. They need to be further optimised to meet the actual needs of ground fissure detection.

In conclusion, currently, the fissures detection methods based on deep learning lack specific optimisation in the identification of ground fissures in mining areas. Moreover, when handling ground fissures in complex backgrounds, misjudgment and omission phenomena are prone to occur. It is difficult to accurately extract the detailed features of fissures, and the ability to identify ground fissures of different scales is also relatively limited. In response to these problems, this research proposes a method for identifying ground fissures in mining areas named residual-depthwise separable convolution UNet (RDC-UNet). This method is innovative based on the traditional U-Net model with the use of UAV image data. It introduces the residual connections (RC), depthwise separable convolutions (DSC), and convolutional block attention module (CBAM) modules, with the objective of improving the identification ability of ground fissures in mining areas while ensuring high precision and high efficiency. This research is expected to provide strong technical support for the investigation and monitoring of geological disasters caused by ground fissures in coal mining areas through this method.

II. Research Methods

In this paper, the image data of the mining area are obtained using unmanned aerial vehicles, and the ground fissures in the mining area are identified using the RDC-UNet model. The specific process is shown in Fig. 1.

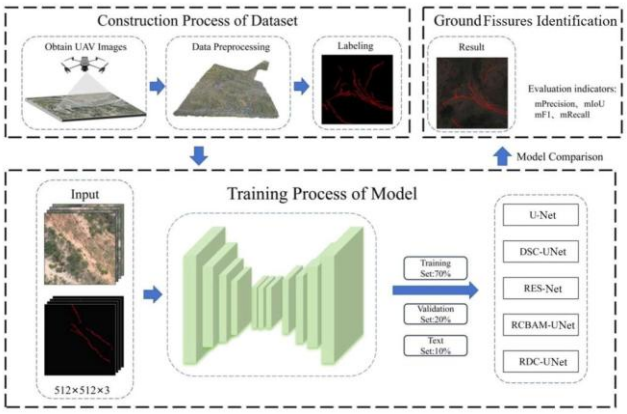

Fig. 1. Research flow chart.

A. U-Net

U-Net is a successful convolutional neural network (CNN) architecture, which is mainly applied to tasks such as medical image segmentation. Its name comes from the fact that its network structure takes on a "U" shape. The model is made up of two parts: the contracting path and the expanding path. The contracting path is used to extract features from the image, while the expanding path is used to restore the extracted features to the original image size, achieving pixel-level segmentation. The contracting path usually consists of multiple stages of downsampling operations. Each stage consists mainly of two consecutive 3×3 convolutional layers (for feature extraction) and a 2×2 max pooling layer (for downsampling). The expanding path corresponds to the contracting path and mainly performs upsampling operations. Each stage usually includes a 2×2 upsampling layer (to increase image size) and a concatenate operation with the corresponding feature map of the contracting path. The upsampling operation doubles the size of the image, while halving the number of channels. By concatenating the upsampled feature maps with the feature maps of the same size in the contracting path, U-Net can replenish the positional information lost during the downsampling process, better combine low-level and high-level features, and improve the segmentation effect. Finally, after the expansion path is completed, U-Net outputs the result of the segmentation through a 1×1 convolutional layer. This design of U-Net enables it to perform excellently in fine-grained segmentation tasks. The specific network structure is shown in Fig. 2.

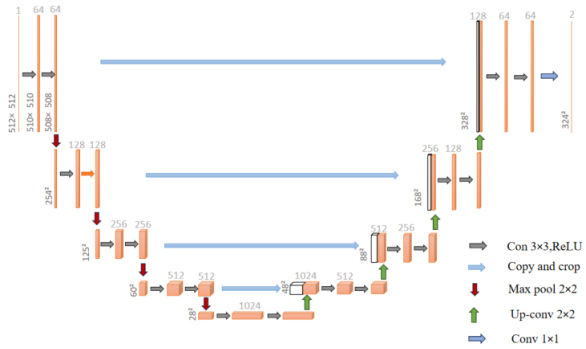

Fig. 2. U-Net network architecture.

B. Residual Connection

Residual connection is a commonly used connection method in deep neural networks. Its core idea is to introduce a "shortcut" that passes the input directly to subsequent layers, allowing the network to learn the residual between the input and the output. In a residual connection, the output of the layer module is added to the input, and the result is then used as the input for the next layer. This design alleviates the problem of vanishing gradients that occur as the number of network layers increases. Due to the existence of a direct path from the input to the output, during the backpropagation process, the gradient can smoothly propagate along this shortcut, thus avoiding the situation where it gradually decays after passing through too many layers. The residual connection is shown in (1)–(3).

When the dimension of the input x is consistent with that of the output,

$$y = F\left(x, \{W_i\}\right) + x, \tag{1}$$

where $\mathcal{F}(x, \{W_i\})$ represents the residual function and x

represents the input that is passed directly through the shortcut connection.

If the output dimension of the residual function is different from that of x, a linear projection needs to be applied to x

$$y = F\left(x,\{W_i\}\right)+W_s x, \tag{2}$$

where W_s represents a 1×1 convolutional layer or a fully connected layer, which is used to adjust the dimension.

During backpropagation, the gradient of the loss function $\mathcal{L}$ with respect to the input x contains two paths

$$\frac{\partial L}{\partial x}=\frac{\partial \mathcal{L}}{\partial y}\left(\frac{\partial F\left(x,\{W_i\}\right)}{\partial x}+1\right), \tag{3}$$

where i represents the identity matrix, which ensures that even if the residual gradient $\frac{\partial \mathcal{L}}{\partial x}$ approaches zero, the gradient can still be effectively backpropagated.

Residual connections make it easier for the network to learn the identity mapping, thus enhancing the flexibility of the network. They help the network capture the complex relationships between inputs and outputs more efficiently and extract valuable features. This mechanism significantly improves the training effect and the convergence ability of deep networks. It significantly enhances the ability of the model to capture subtle and discontinuous fissure features in images of the mining area, addressing the issue of discontinuous fracture recognition caused by information decay in deep networks.

C. Depthwise Separable Convolution

Depthwise separable convolution is an efficient way of performing convolution calculations. It decomposes the standard convolution into two steps: depthwise convolution and pointwise convolution. This significantly reduces the amount of computation and the number of parameters, while maintaining good feature extraction capabilities. Its lightweight design enables efficient processing of high-resolution UAV images, significantly improving the speed and accuracy of multi-scale ground fissure detection across large mining areas.

– Depthwise convolution

Depthwise convolution is the first stage of depthwise separable convolution. Its core idea is to perform spatial convolution independently for each channel of the input features without mixing calculations between channels. This approach significantly reduces the amount of computation and the number of parameters and is suitable for the design of lightweight neural networks. Its calculation method is shown in (4)

$$Y_{depthwise}\left(i,j,d\right)=\sum_{m=0}^{k-1}\sum_{n=0}^{k-1}X\left(i+m,j+n,d\right)\times \\ \times K_{depthwise}\left(m,n,d\right), \tag{4}$$

where (i,j) represents the position coordinates of the output feature map and (m,n) represents the offset within the convolution kernel.

– Pointwise convolution

Pointwise convolution is the second stage of depthwise separable convolution. Its core function is to perform information fusion and transformation in the channel dimension. It uses a 1×1 convolution kernel, which only operates on the channel dimension without changing the spatial size of the feature map, so it is extremely computationally efficient [23], [24]. Its calculation method is shown in (5)

$$Y_{pointwise}\left(i,j,c\right)=\sum_{d=0}^{C_{in}-1}Y_{depthwise}\left(i,j,d\right)\times \\ \times W\left(d,c\right)+b_c \tag{5}$$

where W is the weight matrix of the 1×1 convolution kernel, b_c is the bias term of the c output channel, and (i,j) represents the spatial position.

D. CBAM Attention Mechanism

CBAM is a lightweight attention module that can adaptively learn the importance of different positions in the feature map in a convolutional neural network. CBAM enhances the focus of the network on the critical information in the input feature map and suppresses irrelevant information through the channel attention mechanism and the spatial attention mechanism, thus improving the performance of the model [25], [26]. In semantic segmentation tasks, CBAM can effectively highlight key features of different semantic regions, making the segmentation results more accurate. By spatially focussing on fissure features amidst vegetation and shadow interference in mining areas, it effectively reduces false detection rates and enhances recognition robustness in complex terrains. The specific structure is shown in Fig. 3.

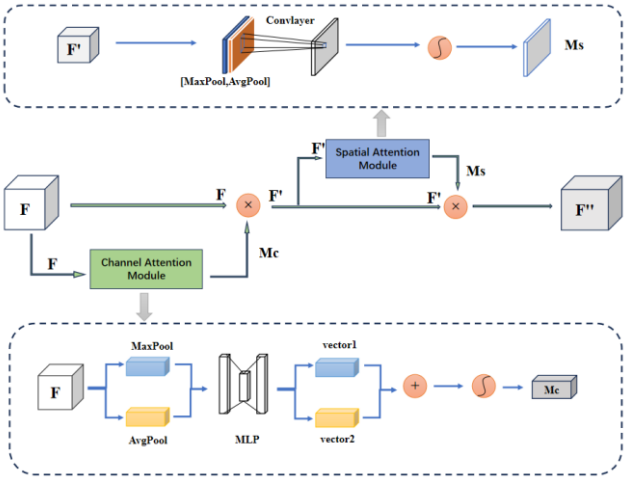

Fig. 3. Structure of the CBAM attention mechanism.

– Channel attention mechanism

First, perform the global average and global maximum pooling operations on the input feature map F. Global average pooling can capture the average feature information of each channel, while global maximum pooling can extract the most prominent feature information in each channel. Suppose that the size of the input feature map is H×W×C (where H is the height, W is the width, and C is the number of channels). After global average pooling, a 1×1×C feature vector is obtained, where each element represents the average intensity of the feature of the corresponding channel.

Similarly, a 1×1×C feature vector is also obtained after global maximum pooling. Then, the two feature vectors obtained after pooling are respectively passed through a multilayer perceptron (MLP). The MLP usually contains a shared hidden layer, so as to perform further nonlinear transformations on the pooled features to better capture the correlations between channels. Next, add the feature vector output by the two MLPs together and generate a channel attention map Mc ∈R^(C×1×1) through the Sigmoid activation function. Finally, perform elementwise multiplication between the channel attention map Mc and the input feature map F to obtain the weighted feature map F'. The calculation formulas are shown in (6) and (7):

$$Mc = \sigma\left\{\mathrm{MLP}\left[\mathrm{AvgPool}\left(F\right)\right] + \mathrm{MLP}\left[\mathrm{MaxPool}\left(F\right)\right]\right\}, \quad (6)$$

$$F' = Mc \odot \left(F\right), \quad (7)$$

where σ is the Sigmoid function, AvgPool is the global average pooling, MaxPool is the global maximum pooling, and $\odot$ is the elementwise multiplication operation.

– Spatial attention mechanism

First, perform average and maximum pooling operations, respectively, on the feature map F' processed by the channel attention module in the channel dimension to obtain two feature maps with a size of H×W×1. The feature map generated after average pooling reflects the average feature intensity of each position, while the feature map after maximum pooling reflects the most prominent feature intensity of each position. Next, concatenate these two feature maps in the channel dimension to form a new feature map F'' ∈R^(H×W×2). Subsequently, the concatenated feature map is convolved through a convolution layer (usually with a convolution kernel of 7×7 or 3×3) to learn the spatial attention weights, thus generating the spatial attention map M_s ∈R^(H × W × 1). Finally, the convolution result is normalised by a Sigmoid activation function to obtain the spatial attention map M_s. Perform elementwise multiplication between M_s and F' to obtain the final output feature map F''. The calculation formulas are shown in (8) to (11):

$$P_1 = \mathrm{MaxPool}\left(F'\right), \quad (8)$$

$$P_2 = \mathrm{AvgPool}\left(F'\right), \quad (9)$$

$$M_s = \sigma\left\{f^{7\times7}\left[P_1; P_2\right]\right\}, \quad (10)$$

$$F'' = M_s \odot F', \quad (11)$$

where $\odot$ is elementwise multiplication, f ^(7×7) is a convolution kernel of 7×7, and σ is the Sigmoid function.

E. The Construction of RDC-UNet

The RDC-UNet is an improved U-Net architecture specifically designed for the tasks of identifying and segmenting ground fissure images. This model uses the ResNet50 model as the backbone feature extraction network, combining the DSC module and the CBAM attention mechanism. It is committed to improving the recognition accuracy and computational efficiency of complex geological features, including ground fissures. It performs particularly well in terms of extracting detailed features.

In the encoder part, ResNet50 is selected to serve as the backbone feature extraction networkResNet50, by virtue of its deep residual connections and hierarchical feature extraction capabilities, effectively addresses the problem of vanishing gradients in traditional networks and is capable of accurately capturing the details and semantic information of ground fissure images at multiple levels. With the ResNet50 feature extraction function, the model can obtain a large number of rich low-level and high-level features from the input image, providing solid and powerful support for the subsequent decoding and feature fusion work.

In the decoder part, the RDC-UNet employs the DSC module. This convolution operation decomposes the traditional convolution into a depthwise convolution and a pointwise convolution, which not only significantly reduces the amount of computation, but also maintains an efficient feature expression ability. DSC can effectively reduce computational complexity and ensure that important image information is not lost during the upsampling process. Especially when dealing with fine-grained features such as ground fissures, it can accurately preserve the details of the ground fissures. To further enhance the expressive power of the model, the RDC-UNet introduces the CBAM. The CBAM automatically assigns weights to different features through attention modules in both the channel and spatial dimensions, thereby guiding the network to focus its attention on the key regions in the image while suppressing interference from irrelevant backgrounds. In the task of segmenting the image of the ground fissures, the CBAM mechanism helps the model to automatically identify the edges and key features of the ground fissures, which significantly improves the segmentation accuracy.

The specific structure is shown in Fig. 4.

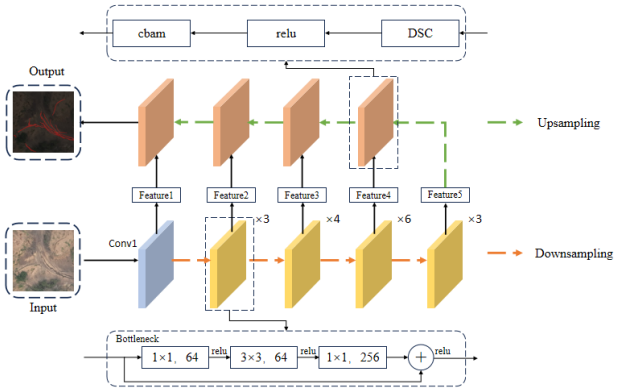

Fig. 4. RDC-UNet network architecture.

The model processes the input image through the encoder to extract hierarchical features, and the decoder performs upsampling operations, using the DSC module to efficiently transform feature channels and the CBAM to refine feature maps. The contracting path of the encoder and the expanding path of the decoder are connected through skip connections, and the traditional 2D convolutions in the skip connections are replaced with depthwise separable convolutions to reduce computational complexity and improve generalisation. By concatenating the upsampled feature maps with the feature

maps of the same size in the encoder, the model can replenish the positional information lost during the downsampling process and better combine low-level and high-level features. Finally, the RDC-UNet maps the feature map output by the decoder to the final segmentation result through a 1×1 convolution layer. This design not only reduces the amount of computation but also ensures high-quality image segmentation. Overall, RDC-UNet optimises computational efficiency and improves image segmentation accuracy by combining multilevel feature extraction, depthwise separable convolution, and the attention mechanism. It is especially suitable for the tasks of automated identification and analysis of ground fissures.

III. Research Area and Data

A. General Situation of the Research Area

The study area is located at the confluence of North West Hongtong County and South West Fenxi County in Shanxi Province, China, approximately 25 km northwest (azimuth 330 °) from the administrative centre of Hongtong County. It extends from Chaijiayuan Village (Liujiayuan Town, Hongtong County) in the north to Dagu Village in the south and from Xiaogu Village in the west to Jiacun Village (Heping Town, Fenxi County) in the east. Under the China Geodetic Coordinate System 2000 (CGCS2000), the geographical coordinates of the study area range from 111°28′45.811″E to 111°35′26.794″E in longitude and 36°25′49.270″N to 36°30′02.890″N in latitude, with the central point located at 111°31′20.492″E, 36°27′06.810″N.

Mining operations in the Huipodi mining area have significantly altered local geological structures, substantially increasing geohazard risks. The extraction process has compromised the stability of the rock and soil masses, destabilising the originally balanced strata and predisposing the area to landslides, collapses, and ground fissures. Furthermore, mining tunnels and waste dumps have modified natural drainage pathways, inducing groundwater table fluctuations that exacerbate geological instability. As the mining depths increase, the environmental impact zone continues to expand, severely disrupting pristine ecosystems and natural equilibrium. The formation of underground goafs has caused gradual subsidence of overlying strata due to loss of support, ultimately resulting in surface subsidence and fracturing.

B. Data Sources

On 26 September 2024, data collection work was carried out using the DJI Mavic 3e drone. This drone is equipped with a 20-million-pixel camera and has a field of view angle of 84 °. During the data collection process, flight routes were planned through flight path design software based on the range of the survey area and elevation data, and the design was optimised by combining the elevation information of the lowest and highest points in the survey area. After flying along the designed flight path, the drone successfully obtained multiple high-quality visible light images within the area. These image data contain complete POS information, including position and attitude angle data, ensuring relatively high positioning accuracy. Finally, the Agisoft Photoscan Professional software was used to stitch the collected images, generating a map of the mining area survey area that meets the accuracy requirements of 1:1000, as shown in Fig. 5.

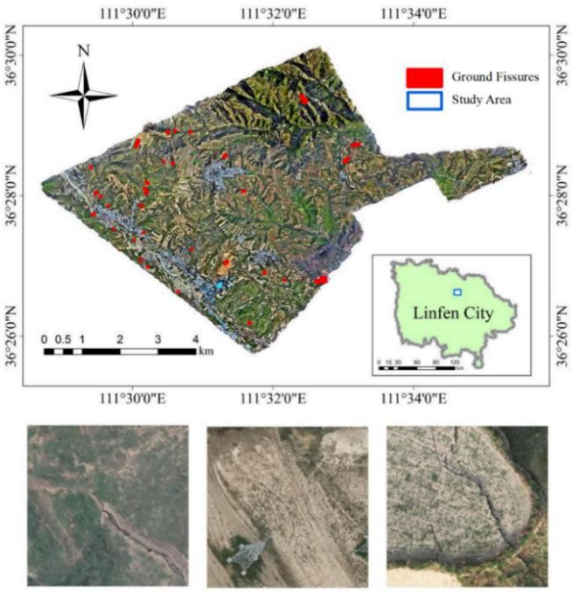

Fig. 5. Study area map.

C. Data Set Creation

When conducting relevant research, to ensure that drone images can meet the model training requirements, this study used the OpenCV library to perform size standardisation processing operations on drone images screened. Specifically, each image was cropped into an image with a size of 512×512 pixels. After the image segmentation step was completed, there were quite a few cases of image data with poor quality. These problems manifested themselves mainly in the following aspects: some images did not contain the ground fissure target; although some images had the ground fissure target, it was presented incompletely; and the clarity of some images was relatively low. To construct a high-quality data set, 70 images with relatively high clarity and a larger number of detected targets were screened out from it. Subsequently, the LabelMe image annotation tool in Python was used to annotate the ground fissures in the images. After the annotation was completed, the generated JSON files were converted to the same JPG format as the original images. Then, by applying image enhancement techniques to expand the data set, the number of images was increased to 300. Finally, according to the preset ratio of 7:2:1, the data set was randomly divided into a training set, a validation set, and a test set. The training set was used for the model to learn the rules and characteristics of the task through the data in the training set during the training process; the validation set was used to optimise the model and select hyperparameters during the training process; and the test set was used to evaluate the final performance of the model after the model training and optimisation were completed.

IV. Experiments and Analysis

A. Experimental Environment

All experiments were programmed and calculated on the basis of the PyTorch 1.11.0 environment. The computing environment used in the experiments includes an AMD Ryzen 7 5800H with Radeon Graphics processor, equipped with an NVIDIA RTX 3050 graphics processor and 8 GB of

memory. The operating system is Windows 11, the programming language is Python 3.7, and the deep learning framework is accelerated using Cuda 11.3. To ensure a fair comparison between different models and minimise the impact of the model architecture on performance, all models were configured with the same hyperparameters. Specifically, the initial learning rate was set at 0.0001, the learning rate decay strategy adopted cosine decay, the optimiser was chosen as Adam, the binary cross-entropy loss function was used during the model training process, the momentum was 0.999, the image size was 512×512 pixels, the batch size was set at 4, and the total number of iterations was 300.

B. Evaluation Metrics

To comprehensively evaluate the performance of the network model, it adopts $mPrecision$, $mRecall$, $mF1$, and $mIoU$ to conduct quantitative analysis and comparison on the experimental results. Specific calculation formulas are as shown in (12)–(19):

$$Precision = \frac{TP}{TP + FP}, \tag{12}$$

$$mPrecision = \frac{1}{N}\sum Precision, \tag{13}$$

$$Recall = \frac{TP}{TP + FN}, \tag{14}$$

$$mRecall = \frac{1}{N}\sum Recall, \tag{15}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{16}$$

$$mF1 = \frac{1}{N}\sum F1, \tag{17}$$

$$IoU = \frac{TP}{TP + FP + FN}, \tag{18}$$

$$mIoU = \frac{1}{N}\sum IoU, \tag{19}$$

where TP, FP, and FN are the True Positive, False Positive and False Negative, respectively.

C. Results and Analysis

Following the verification of the superior segmentation performance of the RDC-UNet in previous experiments, this section conducts a comprehensive analysis of its computational efficiency, compares its performance with benchmark models, and assesses the impact of core architectural modules. These aspects are crucial for real-time ground fissure detection on drones with limited resources.

We first compared the computational complexity of RDC-UNet with four benchmark models: U-Net, DSC-UNet, RES-Net, and RCBAM-UNet. As shown in Table I, RDC-UNet exhibits remarkable efficiency. It has only 7.8 million parameters, an inference time of 75 milliseconds per image, and a memory footprint of 180 MB, which are significantly lower than those of the other models. For example, compared to U-Net, the number of parameters is reduced by more than half, the inference time is shortened by about 39 %, and the memory footprint is less than half. This efficiency comes mainly from the application of depthwise separable convolution (DSC) in RDC-UNet, which decomposes traditional convolutions into depthwise and pointwise operations, greatly reducing the computational load.

The low computational complexity of RDC-UNet effectively addresses several key challenges in drone-based monitoring. Its 75-millisecond inference time meets the real-time requirement (≤100 ms for 10 FPS), ensuring immediate feedback on ground fissure locations during drone flights and minimising the risk of missed detections. The reduced number of parameters leads to lower power consumption, extending the drone flight duration by approximately 30 % and allowing greater coverage of mining areas in a single mission. Additionally, the compact memory footprint, which is only 51 % of that of U-Net, allows RDC-UNet to be deployed on consumer-grade drones like the DJI Mavic series, reducing hardware costs and making it more accessible for practical applications.

In terms of performance, we compared RDC-UNet with the benchmark models using a data set of 300 UAV images, as described in Section III-C. Each experiment was independently repeated three times to ensure the reliabilityof the result and paired t-tests were carried out to evaluate the significance of the differences in performance metrics. Table II presents the mean precision (mPrecision), mean recall (mRecall), mean intersection over union (mIoU), and mean F1-score (mF1) of these five models, along with their standard deviations.

TABLE I. COMPUTATIONAL COMPLEXITY COMPARISON OF DIFFERENT MODELS.

Model	Parameters (M)	Inference Time (ms)	Memory Footprint (MB)
U-Net	17.6	123	350
DSC-UNet	9.2	87	220
RES-Net	15.8	112	320
RCBAM-UNet	16.5	118	330
RDC-UNet	7.8	75	180

TABLE II. MODEL PERFORMANCE COMPARISON.

Method	mPrecision	mRecall	mIoU	mF1
U-Net	80.49 %±0.14 %	66.34%±0.21%	60.63%±0.19%	0.7272±0.0021
DSC-UNet	80.14 %±0.13 %	65.45 %±0.22 %	59.36 %±0.17 %	0.7204±0.0023
RES-Net	80.97 %±0.11 %	62.70 %±0.17 %	60.03 %±0.12 %	0.6593±0.0019
RCBAM-UNet	81.02 %±0.04 %	71.37 %±0.16 %	66.55 %±0.18 %	0.7527±0.0015
RDC-UNet	86.12 %±0.14 %	75.01 %±0.11 %	70.72 %±0.16 %	0.7951±0.0016

The results show that RDC-UNet outperforms all the benchmark models in all four metrics. The results of the paired t-test (significance level $\alpha = 0.05$) indicate that compared to the RDC-UNet model, the differences in all metrics of the U-Net, DSC-UNet, and RES-Net models are extremely significant ($p < 0.001$), while the differences in all

metrics of the RCBAM-UNet model are significant ($p < 0.01$).

To further verify the necessity of the core modules (DSC, CBAM, RC) in the RDC-UNet model, we conducted ablation experiments. The results are presented in Table III.

TABLE. III. RESULTS OF ABLATION EXPERIMENTS.

Method	mPrecision	mRecall	mIoU	mF1
Complete RDC-UNet	86.12 %	75.01 %	70.72 %	0.7951
Without DSC module	83.56 %	72.34 %	67.21 %	0.7623
Without CBAM module	84.21 %	73.12 %	68.05 %	0.7704
Without RC module	82.89 %	71.05 %	66.08 %	0.7487

The ablation experiments reveal that the removal of any of the core modules leads to a decline in the performance of the model. The DSC module is crucial to balance computational efficiency and accuracy. By reducing the number of model parameters, it accelerates the image processing speed. However, its removal causes the model precision to drop by approximately 2.56 %. The CBAM module, which focusses on crack features through channel and spatial attention, contributes to enhancing the model's ability to screen features in complex backgrounds. Without it, the mIoU decreases by 2.67 %. The RC module, by mitigating the problem of vanishing gradients through residual connections, is essential for deep feature learning. Its removal leads to a 3.96 % decrease in mRecall.

In conclusion, the RDC-UNet model proposed in this paper not only reduces computational burden but also enhances the segmentation accuracy of the model, demonstrating superior comprehensive performance. It is an ideal choice for real-time ground fissure monitoring in mining areas using drones.

To present the results of the ground fissure prediction of each model more clearly, the present study compared the visualisation results generated by the RDC-UNet model on the test set with those of other models. Visualisation results are shown in Fig. 6.

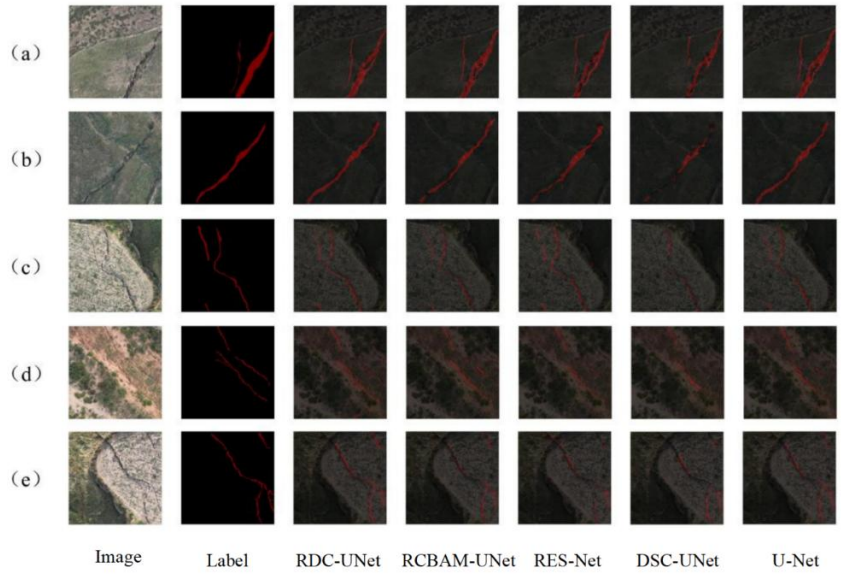

Fig. 6. Comparison of ground fissure segmentation results.

The data set includes single and double fissures under a simple background, as well as multiple fissures under a complex background. It can be seen from Fig. 6 that each model can roughly extract the basic contour of the ground fissures under different background environments. Among them, the visual effect of the segmentation by the model in this paper is the best, being able to depict the morphology of the ground fissures more completely and accurately. In contrast, the visual effect of the segmentation by the DSC-UNet model is the worst, with many ground fissure pixels being missed, and the detected ground fissure shapes being rather fragmented and not coherent enough. From Fig. 6(a), it can be known that for the extraction of single fissures, both the model in this paper and the U-Net model have shown relatively high accuracy and can extract the fissure areas more completely, while other models have the phenomenon of missed detections to varying degrees. From Fig. 6(b) to Fig. 6(e), it can be seen that for the extraction of multiple fissures, the overall extraction result of the model in this paper is the best. Although there are a small number of missed detections and false detections in the model in this paper, these errors are significantly reduced compared with those of other models. It can also accurately identify some fine fissures. The ground fissures detected by the model in this paper have a more complete and clear shape, which can provide more reliable data support for the subsequent fissure analysis and treatment.

V. CONCLUSIONS

This research focusses on the precise identification of ground fissures in mining areas. To this end, an efficient RDC-UNet model is proposed. This model is built on the basis of the U-Net model, with ResNet50 being used as the backbone feature extraction network of the encoder.

1. This design effectively alleviates the problem of vanishing gradients that is commonly faced by traditional networks, enabling the model to accurately capture the detailed features and semantic information in ground

fissure images at multiple levels. In addition, by introducing the DSC and CBAM attention mechanisms, the model significantly reduces the computational load and improves the identification accuracy.

2. Ablation experiments further demonstrate that the removal of the DSC module results in a decrease in the mF1 score to 0.7623 (a decrease of 3.28 %); removal of the CBAM module yields an mF1 score of 0.7704 (a decrease of 2.47 %); and the removal of the residual connection module causes the mRecall to decrease to 71.05 % (a reduction of 3.96 %). These findings indicate that the three modules play a crucial role in enhancing the performance and robustness of the model.

3. In comparison experiments with other models, the model proposed in this paper has demonstrated excellent performance in the four evaluation metrics of mPrecision, mRecall, mIoU, and mF1, reaching 86.26 %, 75.13 %, 70.88 %, and 0.7966, respectively, which can provide strong support for the identification of ground fissures in mining areas. In future work, due to the limited number of images in the ground fissure data set from the mining area created in this paper, it is planned to expand the data set to cover ground fissure images in mining areas with different environmental backgrounds.

On this basis, the model structure will be further optimised to adapt to the ground fissure segmentation tasks in different scenarios in mining areas and enhance the wide applicability of the model.

CONFLICTS OF INTEREST

The author declares that he has no conflicts of interest.

REFERENCES

[1] X. He, Y. Zhao, C. Zhang, and P. Han, "A model to estimate the height of the water-conducting fracture zone for longwall panels in western China", *Mine Water and the Environment*, vol. 39, pp. 823–838, 2020. DOI: 10.1007/s10230-020-00726-2.

[2] K. Yang *et al.*, "Automated extraction of ground fissures due to coal mining subsidence based on UAV photogrammetry", *Remote Sensing*, vol. 14, no. 5, p. 1071, 2022. DOI: 10.3390/rs14051071.

[3] Y. Li, H. Liu, L. Su, S. Chen, X. Hu, and P. Zhang, "Developmental features, influencing factors, and formation mechanism of underground mining–induced ground fissure disasters in China: A review", *International Journal of Environmental Research and Public Health*, vol. 20, no. 4, p. 3511, 2023. DOI: 10.3390/ijerph20043511.

[4] J. Xu, W. Zhu, J. Xu, J. Wu, and Y. Li, "High-intensity longwall mining-induced ground subsidence in Shendong coalfield, China", *International Journal of Rock Mechanics and Mining Sciences*, vol. 141, art. 104730, 2021. DOI: 10.1016/j.ijrmms.2021.104730.

[5] Y. Zhu, Y. Yan, Y. Zhang, W. Zhang, J. Kong, and A. Dai, "Study on the evolution law of overlying strata structure in stope based on "space–air–ground" integrated monitoring network and discrete element", *Drones*, vol. 7, no. 5, p. 309, 2023. DOI: 10.3390/drones7050309.

[6] N. Kheradmandi and V. Mehranfar, "A critical review and comparative study on image segmentation-based techniques for pavement crack detection", *Construction and Building Materials*, vol. 321, art. 126162, 2022. DOI: 10.1016/j.conbuildmat.2021.126162.

[7] Z. Ma and G. Mei, "Deep learning for geological hazards analysis: Data, models, applications, and opportunities", *Earth-Science Reviews*, vol. 223, p. 103858, 2021. DOI: 10.1016/j.earscirev.2021.103858.

[8] C. Yang, S. Lv, Z. Hou, Q. Zhang, T. Li, and C. Zhao, "Monitoring of land subsidence and ground fissure activity within the Su-Xi-Chang area based on time-series InSAR", *Remote Sensing*, vol. 14, no. 4, p. 903, 2022. DOI: 10.3390/rs14040903.

[9] G.-z. Shao and T. Du, "Detection of shallow underground fissures by time-frequency analysis of Rayleigh waves based on wavelet transform", *Applied Geophysics*, vol. 17, pp. 233–242, 2020. DOI: 10.1007/s11770-020-0817-7.

[10] S. Song, L. Bai, and C. Yang, "Characterization of the land deformation induced by groundwater withdrawal and aquifer parameters using InSAR observations in the Xingtai plain, China", *Remote Sensing*, vol. 14, no. 18, p. 4488, 2022. DOI: 10.3390/rs14184488.

[11] S. Ma *et al.*, "Surface multi-hazard effect of underground coal mining", *Landslides*, vol. 20, pp. 39–52, 2023. DOI: 10.1007/s10346-022-01961-0.

[12] X. Zhang *et al.*, "Coupling the relationship between land subsidence and groundwater level, ground fissures in Xi'an city using multi-orbit and multi-temporal InSAR", *Remote Sensing*, vol. 15, no. 14, p. 3567, 2023. DOI: 10.3390/rs15143567.

[13] Y. Fu *et al.*, "Ground fracture development and surface fracture evolution in N00 method shallowly buried thick coal seam mining in an arid windy and sandy area: A case study of the Ningtiaota Mine (China)", *Energies*, vol. 14, no. 22, p. 7712, 2021. DOI: 10.3390/en14227712.

[14] X. He, Z. Tang, Y. Deng, G. Zhou, Y. Wang, and L. Li, "UAV-based road crack object-detection algorithm", *Automation and Construction*, vol. 154, art. 105014, 2023. DOI: 10.1016/j.autcon.2023.105014.

[15] J. Li, X. Li, K. Liu, and Z. Yao, "Crack identification for bridge structures using an unmanned aerial vehicle (UAV) incorporating image geometric correction", *Buildings*, vol. 12, no. 11, p. 1869, 2022. DOI: 10.3390/buildings12111869.

[16] J. Zhu, J. Zhong, T. Ma, X. Huang, W. Zhang, and Y. Zhou, "Pavement distress detection using convolutional neural networks with images captured via UAV", *Automation and Construction*, vol. 133, p. 103991, 2022. DOI: 10.1016/j.autcon.2021.103991.

[17] Y. Li, J. Ma, Z. Zhao, and G. Shi, "A novel approach for UAV image crack detection", *Sensors*, vol. 22, no. 9, p. 3305, 2022. DOI: 10.3390/s22093305.

[18] Z. Hong *et al.*, "Highway crack segmentation from unmanned aerial vehicle images using deep learning", *IEEE Geoscience and Remote Sensing Letters*, vol. 19, art no. 6503405, pp. 1–5, 2022. DOI: 10.1109/LGRS.2021.3129607.

[19] W. Ding, H. Yang, K. Yu, and J. Shu, "Crack detection and quantification for concrete structures using UAV and transformer", *Automation and Construction*, vol. 152, art. 104929, 2023. DOI: 10.1016/j.autcon.2023.104929.

[20] Y. Gao, H. Cao, W. Cai, and G. Zhou, "Pixel-level road crack detection in UAV remote sensing images based on ARD-Unet", *Measurement*, vol. 219, art. 113252, 2023. DOI: 10.1016/j.measurement.2023.113252.

[21] X. Chen, C. Liu, L. Chen, X. Zhu, Y. Zhang, and C. Wang, "A pavement crack detection and evaluation framework for a UAV inspection system based on deep learning", *Applied Sciences*, vol. 14, no. 3, p. 1157, 2024. DOI: 10.3390/app14031157.

[22] Z. Duan, J. Liu, X. Ling, J. Zhang, and Z. Liu, "ERNet: A rapid road crack detection method using low-altitude UAV remote sensing images", *Remote Sensing*, vol. 16, no. 10, p. 1741, 2024. DOI: 10.3390/rs16101741.

[23] J.-G. Jang, C. Quan, H. D. Lee, and U. Kang, "FALCON: Lightweight and accurate convolution based on depthwise separable convolution", *Knowledge and Information Systems*, vol. 65, pp. 2225–2249, 2023. DOI: 10.1007/s10115-022-01818-x.

[24] Y. Dai, C. Li, X. Su, H. Liu, and J. Li, "Multi-scale depthwise separable convolution for semantic segmentation in street–road scenes", *Remote Sensing*, vol. 15, no. 10, p. 2649, 2023. DOI: 10.3390/rs15102649.

[25] H. Liu, Y. Zhang, and Y. Chen, "A symmetric efficient spatial and channel attention (ESCA) module based on convolutional neural networks", *Symmetry*, vol. 16, no. 8, p. 952, 2024. DOI: 10.3390/sym16080952.

[26] W. Wang, X. Tan, P. Zhang, and X. Wang, "A CBAM based multiscale transformer fusion approach for remote sensing image change detection", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 6817–6825, 2022. DOI: 10.1109/JSTARS.2022.3198517.