FS_YOLOv8: A Deep Learning Network for Ground Fissures Instance Segmentation in UAV Images of the Coal Mining Area

Zhihua Xu[a], Yunhao Lin[a], Zhenxin Zhang[b]
z.xu@cumtb.edu.cn; 1056267419@qq.com; zhangzhx@cnu.edu.cn
a College of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China
b College of Resource Environment and Tourism, Capital Normal University, Beijing 100048,

Keywords: Ground fissures identification, UAV image processing, Instance segmentation, Improved YOLOv8, Coal mining area.

Abstract

The ground fissures caused by coal mining have seriously affected the ecological environment of the land. Timely and accurate identification and landfill treatment of ground fissures can avoid secondary geological disasters in coal mine areas. At present, the fissure identification methods based on deep learning show excellent performance on roads and walls, etc. Nevertheless, the automatic and reliable segmentation of ground fissures in remote sensing images poses a challenge for deep learning networks, due to the diverse and complex texture information included in the minning ground fissures and background.
To overcome these challenges, we propose an improved YOLOv8 instance segmentation network to automatically and efficiently segment the ground fissures in coal mining areas. In detail, a model called FS_YOLOv8 is proposed. The DSPP (Dynamic Snake convolutional Pyramid Pooling) module is incorporated into the FS_YOLOv8 model to establish a multi-scale dynamic snake convolution feature aggregation structure. This module replaces the conventional convolution found in the SPPF module of YOLOv8 and aims to enhance the model's ability to extract features related to fissures with tubular structures. Furthermore, the D-LKA (Deformable Large Kernel Attention) module is employed to autonomously collect fissure context information. To enhance the detection capability of challenging samples in remote sensing images with intricate background and fissure texture, we employ a Slide Loss function. Ultimately, the ground fissure dataset of unmanned aerial vehicle (UAV) images in coal mine areas is subjected to experimental analysis. The experimental findings demonstrate that FS_YOLOv8 exhibits exceptional proficiency in segmenting ground fissures within intricate and expansive mining areas.

1. Introduction

Prolonged and intensive mining has led to the eventual exhaustion of coal reserves, resulting in significant harm and deterioration of the surface environment. Surface movement and deformation caused by high-intensity underground coal mining activities lead to the occurrence of rock subsidence and fissures. Ultimately, this process results in the collapse of the surface, resulting in the formation of ground fissures. The presence of ground fissures significantly affects various aspects of the environment, such as inducing mechanical harm to plant roots, diminishing vegetation, degrading soil quality, exacerbating soil water loss, and presenting other ecological and environmental challenges. Therefore, it is imperative to investigate and map ground fissures promptly, efficiently, and comprehensively.

During the twentieth century, the investigation of the origins and patterns of ground fissures was initiated by scholars in several nations due to the heightened occurrence of geological disasters. The researchers developed a set of classification criteria and constructed a prediction model for ground fractures, utilizing fracture development mechanisms, with the aim of offering a theoretical foundation and technological assistance for future research endeavors. However, the aforementioned study is not without flaws. For example, the study of the mechanism of fissure formation is unable to determine the specific position and distribution of fissures.

The current state of research on ground fissure localization demonstrates that field survey and measurement approaches exhibit a high level of accuracy. However, these methods are constrained by several environmental constraints, including terrain and geographical range (Zhao et al., 2021).

Consequently, conducting field surveys and measurements necessitates a significant investment of time and energy. The conventional techniques employed for satellite data collecting exhibit notable constraints, including the inability to detect fissures with narrow widths and the dependence on satellite revisit intervals. Hence, their utilization in the monitoring of ground fracture dynamics poses a significant challenge.

In recent years, the surveying and mapping business has been greatly influenced by the rapid advancement of the UAV Oblique Photogrammetry technology (Wu, 2022; He, et al., 2019). the UAV Oblique Photogrammetry technology enables the quick acquisition of high-resolution digital images of subsidence areas in a short amount of time. Furthermore, the integration of this technology with image processing techniques facilitates the precise retrieval of ground fissures. The foregoing techniques have several significant advantages, including cost effectiveness, lightweight design, flexibility, and high spatio-temporal resolution, which is suitable for ground fissures detection.

Currently, ground fissures detection are performed through manual interpretation, object-oriented approaches, and deep learning-based methods (Kheradmandi and Mehranfar, 2022). Manual interpretation is labor-intensive, inefficient, and error-prone. The use of object-oriented strategies is common in the field of ground fissure extraction. Wei Changjing et al. (2012) created a knowledge model for extracting fissures from the Majiliang mining area in Shanxi Province. This model mixes the UAV images with satellite remote sensing data. Wei Bowen et al. (2018) used ground fissures in the Yaojie mining area as objects and proposed an improved first-order Gaussian difference matching filter (MF-FDOG) method for extracting ground fissure information in the loess area. This strategy was

used with the UAV images. Zhang Xinghang et al. (2019) used Ge-oEye imaging to collect ground fissure data for their investigation. They then proposed an object-based approach for determining the distribution of ground fissures. Tang Fuquan et al. (2023) used the visible light band difference index (VDVI) to reduce the impact of subsidence data on vegetation cover area, improving accuracy. In their study, they used the UAV to acquire coal surface information, and the results showed the effectiveness of this method in coal mine subsidence area detection. Zhao Yixin et al. (2021). used the UAV infrared remote sensing and edge detection technology to discover ground fissures in infrared images. The researchers evaluated the efficacy of several edge detection approaches in detecting fissures and determined the best time frame for UAV infrared remote sensing to perform fissure identification. Nonetheless, because of the abundance of vegetation on the surface of mining areas and the similarity in spectral color characteristics between ground fissures and ground dry vegetation, determining a universally applicable threshold value for extracting ground fissures using threshold segmentation and edge detection techniques is difficult.

Machine learning based on feature engineering extracts fissure features manually.These features are then fed into regression models such as support vector Machine (SVM) (Prasanna et al., 2014), random forest (Shi et al., 2016), and Neural network (Nn) (Chen et al., 2017) to obtain ground fissure extraction results. Nonetheless, these models mostly rely on manually derived features (Fan et al., 2018). It is very difficult to create fracture features that are universally applicable to all coal mine areas manually.

Deep learning-based methods can autonomously generate all the prominent characteristics of fissures and detect ground fissures in large quantities. The present state of deep learning-based methods, including semantic segmentation and object detection, mostly centers on the examination of fissures in artificial structures, such as sidewalks, bridges, buildings, walls, and other civil infrastructures. In contrast to fissures observed in smooth and homogeneous structures, the real large-scale remote sensing scene is characterized by the presence of background noise, light, and shadow interference resulting from diverse types of grass, fallen leaves, and gravel. Additionally, the target of ground fissures exhibits a range of morphologies. As a result, the extraction of ground fissures is more challenging than that of artificial fissures.

Deep learning-based methods for extracting ground fissures from high-resolution remote sensing images have received little attention in the scholarly literature.Yu et al. (2022) created an automated framework for detecting fissures following an earthquake using remote sensing technology. They used the UAV to capture high-resolution aerial images of earthquake-affected areas, which were then processed using photogrammetry software to generate a digital orthophoto map. Then, crack-CADNET, a new neural network for ground fissure detection in terrain, is developed. This paper is the first in-depth study of earthquack fissure detection using the UAV and deep learning-based processes. It differs from previous methods, which were primarily used to detect fissures on the surface of man-made objects (such as flat roads), by studying the spatial properties of curved linear fissures and introducing adaptive the Deformable Convolution with context channel space enhancement mechanism to address this problem. Xiao et al. (2022) introduced the MFPANet deep learning model, which used an encoder-decoder framework to automatically extract ground fissures in mining areas from UAV images. MFPANet introduced a set of modules that aggregate context information, such as DRN, DAM, ASPP, and MFPN, to assist the model in extracting fissures more precisely. To address the issue of complex background noise in ground fissures, Cheng Jian et al. (2020) proposed a mixed domain attentional deformation convolutional network to increase the contribution degree of specific channels and spatial locations in feature maps to ground fissure recognition. Studies have demonstrated that this strategy can considerably enhance the accuracy of fissure detection.

In conclusion, the preceding investigation and examination of ground fissures and artificial structure fissures have revealed that the extraction of multi-scale high-rise features and the integration of multi-level pyramid features may effectively detect finer fissure pixels. Furthermore, the Deformable Convolution (Dai et al., 2017), various attention methods, and other modules have recently been included into the fissure extraction model to focus on the morphological aspects of fissures, allowing for more exact fissure extraction.

Drawing upon previous studies on the extraction of ground fissures, this study aggregates the ground fissure features in the multi-scale pyramid feature fusion module, furthermore, used the attention mechanism and the deformable convolution to focus on the morphological properties of fissures. This study presents a novel multi-scale pyramid feature fusion module called the DSPP module. This module is designed to leverage the morphological qualities of fissures and utilizes a modified version of the deformable convolution known as the Dynamic Snake Convolution (DSConv) (Yaolei Qi et al., 2023). conventional convolution and the dilation convolution (Chen et al., 2017) are incapable of capturing the intricate geometric characteristics of the target. Furthermore, deformable convolution allows for free learning of deformation shifts,which cause the perceived field to deviate from the target, especially for thin tubular structures. The DSConv algorithm directs the convolution kernel towards the extraction of features from tubular targets, specifically fissures, by imposing a tubular constraint on the translation of deformations. Furthermore, the proposed approach incorporates a D-LKA module that facilitates the extraction of comprehensive contextual information from ground fissures.

Many scholarly works pertaining to ground fissure extraction utilize semantic segmentation methods, which lack the capability to directly extract information from the ground fissure instance. In the domain of artificial structural fissure investigation, numerous methodologies employ target identification methods to ascertain the exact spatial coordinates for each fissure. Nevertheless, these methodologies are unable to accurately delineate the intricate morphological characteristics of fissures. In order to provide a thorough comprehension of ground fissures, this study employs an instance segmentation methodology. The most often employed instance segmentation models currently are the R-CNN series network and the YOLO series network. The YOLO series network has superior performance in terms of detection time and model enhancement when compared to the R-CNN series network. In addition, the YOLO algorithm has undergone multiple updates, which evolving into YOLOv9 (Wang et al., 2024). In relation to the accuracy of detection, YOLO has progressively reached a level of parity with the R-CNN framework.

This research uses the widely used YOLOv8 version as the established paradigm for instance segmentation of ground fissures. The DSPP module and D-LKA (Azad et al., 2024) module developed in this study are merged into YOLOv8 to form the FS_YOLOv8 model. This model is intended to reliably detect and segment ground fissures in mining area images. In conclusion, the primary contributions of this study are summarized as follows:

(1) Investigates the viability of using the instance segmentation network FS_YOLOv8 to locate ground fissures in coal mining areas, laying the groundwork for further research and implementation of deep learning technologies in coal mining surface monitoring.

(2) A Dynamic Snake convolutional Pyramid Pooling (DSPP) module is proposed for extracting multi-scale tubular features of ground fissures. By this means, FS_YOLOv8 demonstrates strong recognition capability for ground fissures of varying scales, widths, and shapes in remote sensing images.

2. Method

2.1 YOLOv8 Network Architecture

The YOLOv8 algorithm is a fast single-stage network for target detection. Figure 1 shows the network architecture of YOLOv8, which comprises of a backbone stage and a segmentation head stage. YOLOv8 incorporates two notable modules, namely C2f and SPPF. The C2f module leverages the advantages of the ELAN structure found in YOLOv7 (Wang et al., 2023) and effectively employs the bottleneck module to record supplementary gradient flow information. The SPPF module integrates the feature maps through the process of pooling kernels of varying sizes.

2.2 FS_YOLOv8 Model for Ground Fissures Segmention

Based on YOLOv8 structure, we produce FS_YOLOv8 that replaces the SPPF module with the DSPP module for tubular targets detection. Here, the forcus is on ground fissures detection in UAV high resolution images. This modification enables the extraction of features from multi-scale tubular targets, resulting in a more comprehensive integration of tubular features from multi-layer receptor fields. Furthermore, a D-KLA module is added as an output layer within the backbone architecture to improve the target's overall contextual features. The structure of the FS_YOLOv8 network is shown in Figure 2, where the modifications are marked in Yellow.
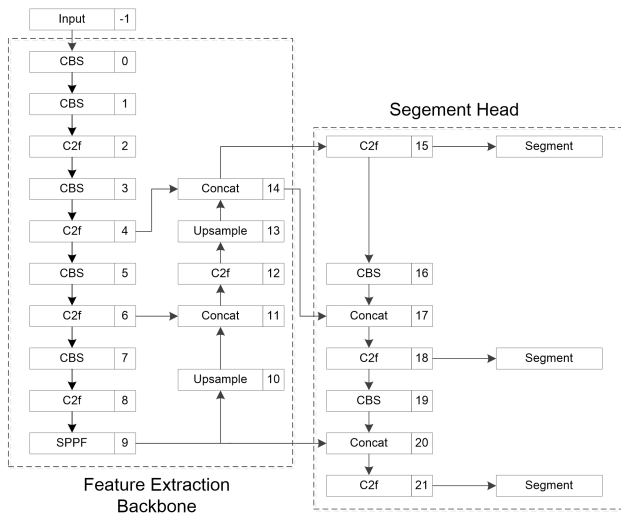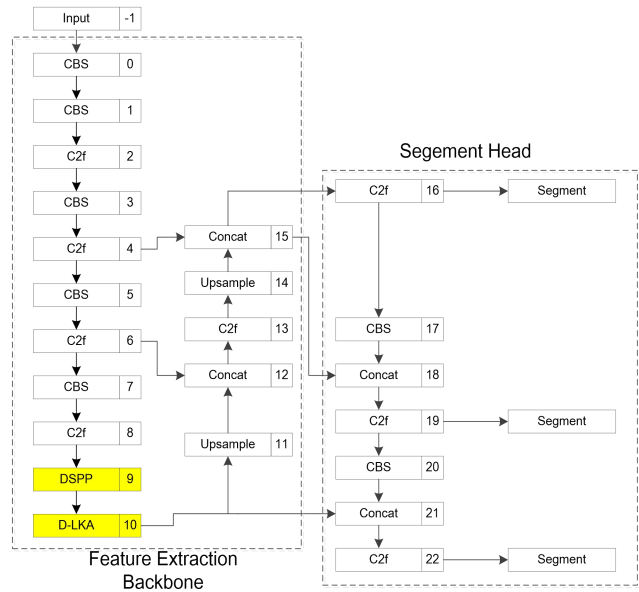

Figure 1. The structure of YOLOv8.

Figure 2. The structure of FS_YOLOv8.

2.2.1 Dynamic Snake Convolution

The Dynamic Snake Convolution demonstrates a favorable performance in extracting the distinctive characteristics of slender, elongated tubular structures, such as roadways and blood arteries. This research applies the method to the ground fissures that possess comparable structures.

Figure 3 depicts the process an iterative strategy of the DSConv. In order to maintain the topological continuity of the convolution kernel deformation, the algorithm selects the observed offset of the subsequent location of each target to be processed sequentially. The offset is limited to the range of [-1,1]. There are two templates available for the DSConv, one positioned along the x-axis and the other along the y-axis. The standard convolution kernel has dimensions of 3x3, resulting in a horizontal grid of 9 convolution nuclei.For the DSConv of the x template, consider the center grid of the convolution kernel as (0,0). The offset in the y direction, which is learnt, represents the vertical distances from other convolution kernels to the center grid. The procedure of selecting each grid point in the convolution kernel commences with the accumulation process of the center grid. In the event that the X-axis distance is increased by one unit, it is necessary to provide the corresponding Y-axis offset. Equation (1) can be utilized to explain the spatial arrangement of each convolution kernel.

$$K_{i\pm c} = \begin{cases} (x_{i+c}, y_{i+c}) = (x_i + c, y_i + \sum_i^{i+c} \Delta y), \\ (x_{i-c}, y_{i-c}) = (x_i - c, y_i + \sum_{i-c}^{i} \Delta y), \end{cases} \quad (1)$$

The offset is a decimal, and the corresponding position cannot be identified in the feature map, so the bilinear interpolation method is used to find the corresponding coordinates on the feature map once the convolution kernel is offset. Equation (2) illustrates the principle of bilinear interpolation.

$$K = \sum_{K'} B(K', K) \cdot K' \quad (2)$$

Figure 4 demonstrates the merging of the DSConvs in both directions with a conventional convolution to form the DSFA module. The DSFA module employs three distinct convolution branches, namely the conventional convolution, the DSConv deformed in the x direction, and the DSConv deformed in the y direction, to compute the input features. The feature aggregation module of DSConv combines the features

produced by the three branches and modifies the output channel number to align with the input channel number using a 1x1 conventional convolution.
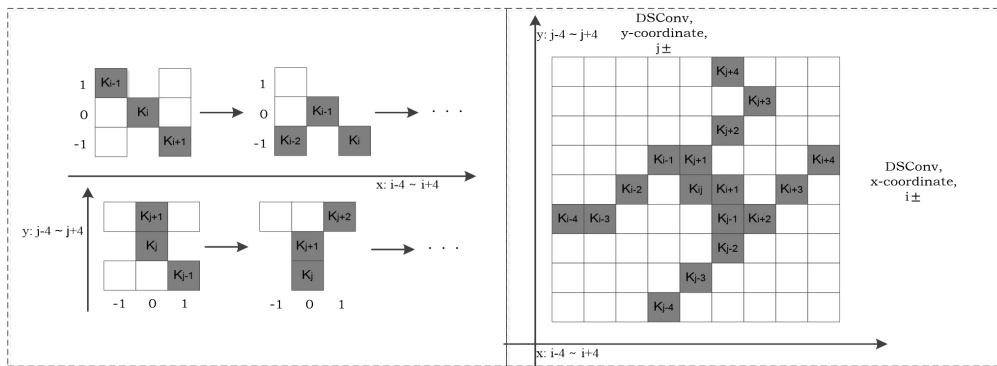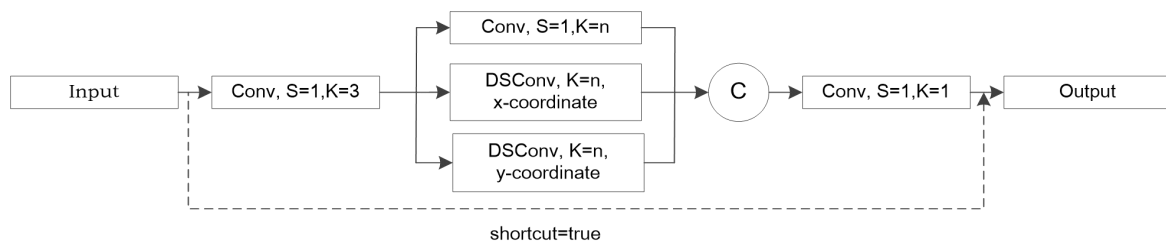

Figure 3. Principle of the DSConv.

Figure 4. The structure of DSFA module.

2.2.2 DSPP Module

The structure of the DSPP module is illustrated in Figure 5. The module for the DSPP structure comprises three main components: 1×1 convolution, SPP structure, and adaptive average pooling. Following the processing of the input features, the three sub-modules employ an alignment connection to splice the resultant three feature maps, so generating the output features. The adaptive average pooling module is designed to capture the global feature information of the input feature map. On the other hand, the pooling pyramid structure is composed of three DSFA modules that utilize different convolution kernels. This enables the model to capture the tubular characteristics of multi-scale receptive fields.

The DSPP module configures the convolution kernels of the three DSFA modules to {5,9,13}. In order to get multi-scale receptive field features, the DSPP module combines the receptive field features from various DSFA modules with different convolution kernels. In contrast, the DSConv module employs a fixed convolution kernel of 9. The YOLOv8 backbone network architecture incorporates the DSPP module as a substitute for the SPPF module, enabling the extraction of multi-scale receptive field features and the integration of high-level semantic information. Both the DSPP module and the SPPF module have a common receptive field. However, the DSPP module utilizes the DSFA module instead of the conventional convolution or pooling layer. The DSFA module is more suitable for segmenting ground fissure targets that possess tubular form features. Simultaneously, the DSPP module has the capability to attain variable scale receptive field characteristics by the manipulation of the convolution kernel, hence enhancing its versatility.
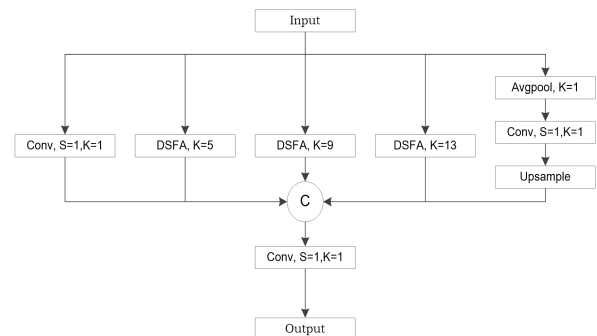

Figure 5. The structure of DSPP module.

2.2.3 D-LKA Module

The fundamental concept underlying Deformable Large Kernel Attention (D-LKA) involves the integration of the attention mechanism of a large convolution kernel and the deformable donvolution. By utilizing a large convolution kernel, D-LKA aims to replicate the receptive field akin to self-attention, while circumventing the computational burden associated with conventional self-attention mechanisms. In addition, D-LKA employs the deformable convolution to dynamically adjust the evaluation grid, hence enhancing the model's capacity to accommodate diverse input patterns.

The model structure of D-LKA is depicted in Figure 5. The integration of depth-separable convolution (DW Conv) kernels and depth-separable convolution with extension (DW-D Conv) in the underlying architecture enables the generation of large convolution nuclei, thereby replicating the receptive field of the self-attention mechanism. This approach facilitates the acquisition of features by the network across a wide range of receptive fields, while simultaneously reducing the parameter count through the utilization of separable convolution. The DW Conv and DW-D Conv of the D-KLA module are created by substituting all conventional convolutions with deformable

conv. By combining a large convolution kernel with deformation-DW-D Conv, the D-LKA module enhances the model's ability to handle complex visual patterns.
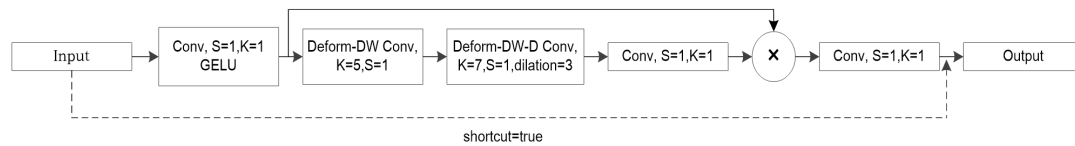

Figure 6. The structure of D-LKA module.

2.2.4 Slide Loss Function

Slide Loss, as proposed by Yu et al. (2022), is employed to direct the model's attention towards difficult samples within the dataset. The remote sensing data comprises a diverse range of ground fissure samples, a significant portion of which pose challenges in terms of identification. The concept of slide loss posits that the model will acquire the ability to optimize these samples and utilize them more efficiently in model training. The distinction between a simple and a difficult sample is established based on the Intersection over Union (IoU) size of the prediction and ground truth boxes. A difficult sample is characterized by an IoU value that is smaller than the average of the IoU values of all bounding boxes (μ). The calculation of Slide Loss is determined by employing the formula presented in equation 3.

$$f(x) = \begin{cases} 1 & x \leq \mu - 0.1 \\ e^{1-\mu} & \mu < x < \mu - 0.1 \\ e^{1-x} & x \geq \mu \end{cases} \quad (3)$$

The slide loss function assigns a greater weight to difficult samples falling within the range of $[\mu\text{-}0.1, \mu]$, while assigning a lower weight to simple samples outside the average IoU as the prediction box IoU increases. This suggests that the model diminishes its focus on easily identifiable samples.

In FS_YOLOv8, the original model's loss functions are utilized, but the Slide loss function is incorporated into the classification loss function calculation. This modification of the weights obtained through CIoU, used for the cross-entropy loss value, enabling the model to prioritize the identification of difficult samples.

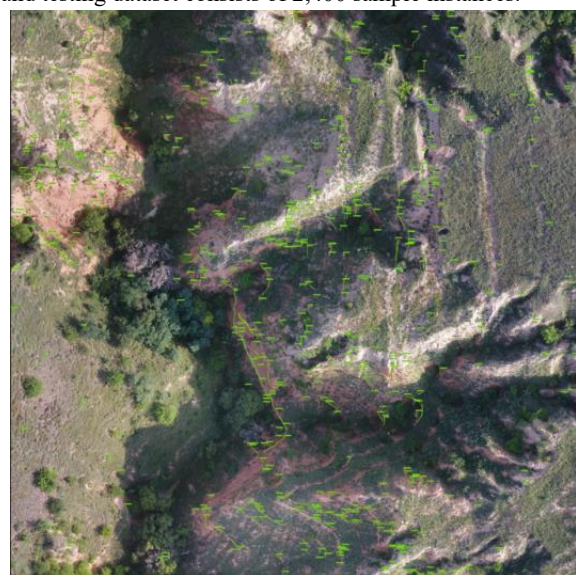## 3. Study Area and Evaluation Evaluation Metrics

3.1 Study Area and Dataset

The study obtained the UAV imagery of areas impacted by subsurface mining operations. Figure 7 shows the example UAV imagery over two mining areas in China, namely the Huangyuchuan and Sandaogou coal mines. The Huangyuchuan mining area is situated in a hilly region with a reasonably gentle slope. The area is characterized by a high vegetation coverage, with vegetation and bare land spread alternately. The Sandaogou mining area is situated in a rugged mountainous region, characterized by steep terrain, abundant vegetation, intricate geological features, and a wide variety of fissures. Figure 8 displays examples of ground fissures within the study area.

The study area yielded a ground fissure datset with an image resolution ranging from 1 to 2cm. This dataset was segmented into 640 640 pixel samples using a clipping approach with a 50% overlap rate, resulting in a total of 3300 images. The images have been categorized into three distinct sets, namely train, validation, and test, with a ratio of 8:1:1. Every image has multiple occurrences of ground fissures. The training dataset comprises over 8,000 cases of fissures, whereas the verification and testing dataset consists of 2,400 sample instances.
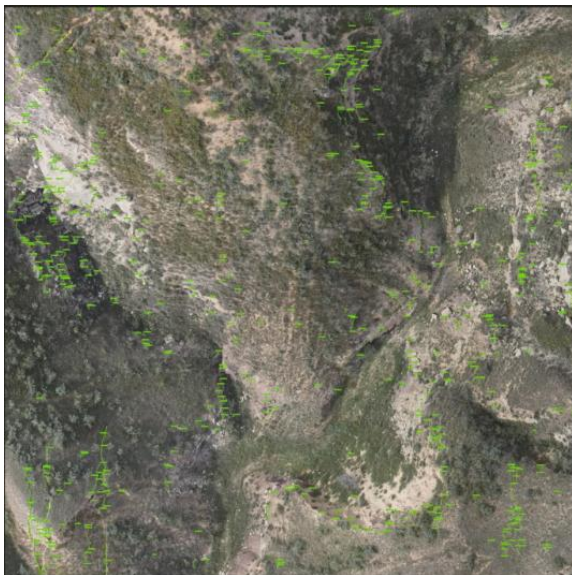
(a) Huangyuchuan mining area   (b) Sandaogou mining area
Figure 7. UAV imagery of the two study areas. The green lines are the fissures annotations.

Figure 8. Examples of ground fissures in the study area.

3.2 Evaluation Metrics

The YOLOv8 evaluation measures were employed to provide an unbiased assessment of the ground fissure instance segmentation model's performance. The metrics used encompassed precision, recall, mAP@0.5, and mAP@0.95 for both the regression boxs and segmentation pixels. The area under the Precision-Recall curve, which interpolates the Precision at different recall rates and calculates the area under the interpolated curve, is referred to as AP. The subscript "@" denotes the iou threshold used to determine the positive or negative nature of a sample. mAP@0.5 represents the average mAP with a threshold exceeding 0.5, whereas mAP@0.95 signifies the average mAP with different IoU thresholds ranging from 0.5 to 0.95, with a step size of 0.05. The fundamental metric for evaluating each regression box is the primary metric. This is because in the instance segmentation task, the regression box is initially positioned, and then the ground fissure pixels are segmented within the positioning box.

4. Experiment Results

4.1 Experimental Setup

The training of all models was conducted using the "start from scratch" methodology, wherein consistent hyperparameters were employed throughout all trials. The hyperparameters utilized during the training process are presented in Table 1. The purpose of setting the training epochs to 1000 is to achieve convergence of the training evaluation metrics for each model on the dataset. This convergence enables the comparison and study of the performance of each model.

Hyperparameters	Value
Learning Rate	0.01
Image Size	640×640
Monetum	0.937
Optimizer	SGD
Batch Size	4
Epoch	1000
Weight Decay	0.0005

Table 1. Hyperparametric configuration.

4.2 Experimental Results and Ablation Study

This section compares the baseline model YOLOv8 and the improved model FS_YOLOv8 using the ground fissure dataset from the mining area. The feasibility of the upgraded technique proposed in this work was confirmed through ablation studies conducted on the baseline model YOLOv8 and the FS_YOLOv8 model. It is noteworthy to emphasize that the YOLOv8 model described in this study represents the default version of YOLOv8n, while the evaluation of different versions remains to be conducted.

4.2.1 Experimental Analysis of Module Ablation

Table 2 presents the evaluation metrics for the YOLOv8 model and its enhanced counterpart, FS_YOLOv8, on the mining ground fissure dataset. The ablation process of module enhancement from YOLOv8 to FS_YOLOv8 is denoted by the symbols "+" and "replaces".

According to the findings presented in Table 2, the replacement of the SPPF module with the DSPP module yields notable enhancements in all evaluation metrics in comparison to the original YOLOv8 model. Notably, the mAP@0.5^{box} and mAP@0.95^{box} metrics experience a substantial increase of 7.9% and 12.6% respectively. The successful application of the DSPP module showcases the efficacy of employing multi-scale dynamic snake convolution for the extraction of ground fissure targets exhibiting diverse tubular, size, and morphological characteristics.

The experimental findings presented in the second row of the table indicate that the YOLOv8 improved model's output layer only included the D-LKA module. In comparison to the original model, the precision of the fissure regression box and segmentation outcomes exhibited improvement, while the recall metric experienced a decline. In FS_YOLOv8, the SPPF module is substituted with the DSPP module, and subsequently, the embedded D-LKA module is incorporated, resulting in a slight enhancement of each metric. One plausible hypothesis is that the SPPF module only incorporates conventional convolution, that is susceptible to interference from the background, leading to the generation of noisy feature graphs. Nevertheless, the deformable convolution in the D-LKA module exhibits excessive flexibility in learning the offset, leading to the acquisition of offset as noise. Consequently, the extraction of high-level characteristics by the D-LKA module yields inaccurate contextual information. The morphological aspects of ground fissures are modeled using the Dynamic Snake Convolution in the DSPP module, resulting in a feature map that is free from noise. The Deformable Convolution in the subsequent D-LKA module learns an offset that is better suited for fissures, resulting in more accurate extraction of fissure context information.

The visual contrast between YOLOv8 and FS_YOLOv8 is depicted in Figure 9. One of the main benefits of FS_YOLOv8 in comparison to the YOLOv8 model is its superior overall confidence level (IoU) in accurately predicting ground fissures. Additionally, the comparative analysis of the models in columns 2 and 3 of Figure 9 reveals that YOLOv8 exhibits a tendency to erroneously classify background items as ground fissures, including branches and tree shadows. By combining the qualitative and quantitative experimental data, FS_YOLOv8 displays better performance than the original model YOLOv8 in instance segmentation of ground fissures.

Method	Precision^{box}	Recall^{box}	mAP@0.5^{box}	mAP@0.95^{box}	Precision^{mask}	Recall^{mask}	mAP@0.5^{mask}	mAP@0.95^{mask}
YOLOv8	76.5%	71.0%	76.9%	52.3%	75.3%	69.3%	73.8%	30.9%
YOLOv8+D-LKA	79.4%	69.7%	76.5%	53.2%	77.8%	68.5%	73.5%	30.9%
DSPP replaces SPPF	88.3%	77.8%	84.8%	64.9%	88.1%	76.0%	82.2%	40.5%
FS_YOLOv8	**90.4%**	**78.3%**	**85.8%**	**66.9%**	**89.4%**	**77.1%**	**84.1%**	**44.5%**

Table 2. Module ablation experiment of FS YOLOv8. The superscript of the evaluation metrics, "box" represents the evaluation metric belonging to the regression box result, "mask" represents the evaluation metric belonging to the segmentation result.
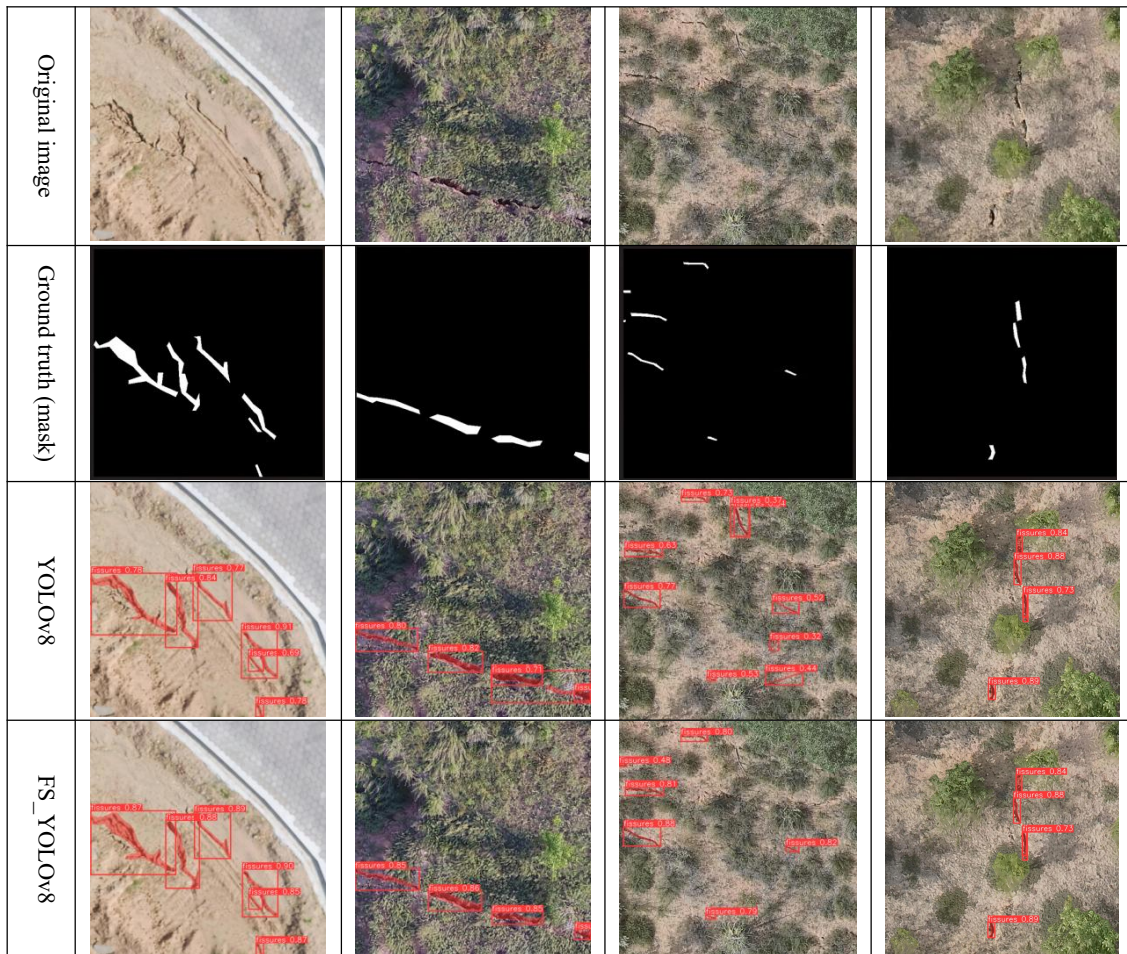

Figure 9. Comparison of prediction results of ground fissures by YOLOv8 and FS_YOLOv8.

IoU Loss	Precision^{box}	Recall^{box}	mAP@0.5^{box}	mAP@0.95^{box}	Precision^{mask}	Recall^{mask}	mAP@0.5^{mask}	mAP@0.95^{mask}
BCE	90.4%	78.3%	85.8%	66.9%	89.4%	77.1%	84.1%	44.5%
BCE+Slide Loss	90.5%	**79.4%**	85.8%	67.0%	90.6%	**78.7%**	84.7%	43.6%

Table 3. Ablation experiment of loss function.

4.2.2 Ablation Study of Loss Function

The impact of incorporating the Slide Loss function into the classification loss of the model, utilizing FS_YOLOv8, is presented in Table 3. According to Table 3, Slide Loss improves the precision of segmenting fissure instances, mostly by increasing the recall of regression boxes and segmentation. This indicates that the Slide Loss function focuses the model's attention on difficult samples. However, because the dataset used in this paper did not purposefully include too many difficult samples, so the Slide Loss function did not significantly improve the metrics, which can only show that Slide Loss function plays a certain role in the model.

5. Discussion

5.1 Parameter Analysis of FS_YOLOv8

Table 4 presents a comprehensive overview of the model complexity parameters observed in the ablation experiment, comparing YOLOv8 to FS_YOLOv8. These parameters include the parameter count, GFLOPs (Gigabit floating point operations per second), which quantifies the network model's execution time in billions of floating point operations per second, and the number of network layers. The utilization of the DSPP module results in a significant increase in both the complexity and parameter count of the model. Because the DSPP module processes the input features through five

branches, with the DSFA module being the primary source of the parameter count. The DSFA module has a more sophisticated network structure than extensible convolution since it includes the Dynamic Snake Convolution and conventional convolution with many big convolution kernels. The process of generating feature graphs and combining features using convolutions leads to a redundancy in the number of feature parameters. In the future, it is necessary to reconstruct the DSPP module.

Method	Para (M)	GFLOPs	Layers
YOLOv8	3.26	12.0	195
YOLOv8+D-LKA	4.89	13.3	203
YOLOv8+DSPP	29.98	30.3	282
FS_YOLOv8	**31.34**	**31.6**	**290**

Table 4. The influence of modules on model complexity in FS YOLOv8.

5.2 Limitations and Future Work

5.2.1 Limitations
This study solely examines the feasibility of YOLOv8 and its improved method on the ground fissure dataset within the coal mining region. Nevertheless, it is important to acknowledge that the research is not without its limitations:
(1) Despite the exploration of multi-scale Dynamic Snake Convolution's potential in ground fissure feature extraction by FS_YOLOv8, the existing structure of the DSPP module was found to be excessively coarse. This resulted in an increased number of parameters within the model, hence requiring the implementation of lightweight processing techniques for the DSPP module.
(2) This research solely evaluated the performance of FS_YOLOv8 on the mining ground fissure dataset, without providing any information about its ability to handle other fissure open datasets. FS_YOLOv8 will be utilized in the future to assess the suitability of each fissure dataset.

5.2.2 Future Work
Regarding the identification of ground fissures at coal mining sites, we are now conducting research on several topics that will be further explored in future studies. These topics include:
(1) Labeling data samples can be a laborious task, particularly when there are numerous additional background targets that resemble ground fissure targets in remote sensing images and require differentiation. Consequently, several weakly supervised and semi-supervised deep learning algorithms are required to automatically generate ground fissure samples from heterogeneous data or crowdsourced data, minimizing models' need on manually labeled samples.
(2) Analyze the spatial arrangement of ground fractures by utilizing remote sensing data from multiple sources. The utilization of thermal infrared imaging enables the visualization of temperature disparities between the fissure region and adjacent background regions at different temporal intervals. InSAR imagery has the capability to identify regions of ground subsidence resulting from mining activities, as well as ground fissures. When combined with multi-source remote sensing data, the model demonstrates enhanced accuracy in detecting ground fissures compared to that solely relies on visible light pictures.
(3) Interpretation of ground fissures using the fissure mechanism model. The general deep learning model is now data-driven, with data determining the model's performance. It is worthwhile to investigate how to spatialize the primary knowledge of mining subsidence, mechanical model, fissure

development law, and so on, in order to build the characteristics and limits of mining areas for the detection of ground fissures, and to incorporate a knowledge-driven model into the model. The knowledge base for ground fissures is added to the model, and the fissure sample data is updated and learned, allowing the model to be applied to a variety of large-scale scenarios in real mining locations.
(4) Creating a deep learning dataset of mining ground fissure is crucial. At now, there is a limited number of individuals that have publicly released ground fissure datasets in mining areas. Due to the intricate nature of remote sensing image scenes, past research has predominantly concentrated on small working surfaces. Researchers tend to favor using simpler scenes as test areas. Consequently, the proposed model method is not applicable to real-world settings. Currently, there is a scarcity of datasets appropriate to diverse scenarios. The following research should look into creating more realistic datasets to explain how deep learning models can be used more effectively in practical engineering applications.

6. Conclusion

This paper introduces a deep learning model called FS_YOLOv8 specifically designed for detecting ground fissures in mining areas. Additionally, we propose a DSPP module within FS_YOLOv8 to extract multi-scale tubular properties of ground fissures. The efficacy of the DSPP module in the detection of ground fissures of diverse sizes and shapes is evaluated through the utilization of the ground fissure dataset.
While our study has made tremendous progress, we continue to face challenges and limits, such as the amount of parameters in the DSPP module and the scenario applicability of the FS_YOLOv8 model. The present study constitutes a first inquiry. Additional areas within the field of ground fissure study warrant additional investigation, including the prospective research directions indicated in this paper, which are anticipated to be progressively incorporated in future studies.

References

Azad R, Niggemeier L, Hüttemann M, et al. 2024: Beyond self-attention: Deformable large kernel attention for medical image segmentation. *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* 1287-1297.

Chen J H, Su M C, Cao R, et al. 2017: A self organizing map optimization based image recognition and processing model for bridge crack inspection. *Automation in Construction* 73: 58-66.

Cheng Jian, Ye Liang, 2020: Guo Yinan et al. Detection method of attention deformation convolutional network in mixed domain of ground fissure in goaf. *Journal of China Coal Society* 45(S2): 993-1002. (in Chinese)

Chen L C, Papandreou G, Kokkinos I, et al. 2017: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence* 40(4): 834-848.

Dai J, Qi H, **ong Y, et al. 2017: Deformable convolutional networks. *In Proceedings of the IEEE international conference on computer vision* 764-773.

Fan Z, Wu Y, Lu J, et al. 2018: Automatic pavement crack detection based on structured prediction with the convolutional neural network. ar**v preprint ar**v: 1802.02208.

Jiang X, Mao S, Li M, et al. 2022: MFPA-Net: An efficient deep learning network for automatic ground fissures extraction in UAV images of the coal mining area. *International Journal of Applied Earth Observation and Geoinformation* 114: 103039.

He Ren,Yanling Zhao,Wu Xiao, et al. 2019: A review of UAV monitoring in mining areas: current status and future perspectives. *International Journal of Coal Science &Technology* 6(3).

Kheradmandi N, Mehranfar V. 2022: A critical review and comparative study on image segmentation-based techniques for pavement crack detection. *Construction and Building Materials* 321: 126162.

Prasanna P, Dana K J, Gucunski N, et al. 2014: Automated crack detection on concrete bridges. *IEEE Transactions on automation science and engineering* 13(2): 591-599.

Qi Y, He Y, Qi X, et al. 2023: Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 6070-6079.

Shi Y, Cui L, Qi Z, et al. 2016: Automatic road crack detection using random structured forests. *IEEE Transactions on Intelligent Transportation Systems* 17(12): 3434-3445.

Tang Fuquan, Sun Wei, Fan Zhigang, et al. 2023: Improvement of surface subsidence information extraction method based on UAV image in western mining area. *Coal Science and Technology* 51(S1): 334-342. (in Chinese)

Wu Xinli. 2022: Application analysis of UAV tilt Photogrammetry technology in mine treatment. *World Nonferrous Metals* (17): 100-102.(in Chinese)

Wei Changjing, Wang Yunjia, Wang Jian, et al. 2012: Research on information technology of UAV image extraction of ground cracks in mining area. *Metal Mine* (10): 90-92, 96.(in Chinese)

Wei Bowen, Liu Guoxiang, Wang Zhiheng. 2018: Extraction of ground fractures in loess area based on improved MF-FDOG algorithm and UAV image. *Geomatics and Mapping* 41(2): 51-56, 61.(in Chinese)

Wang C Y, Yeh I H, Liao H Y M. 2024: YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. ar**v preprint ar**v:2402.13616.

Wang C Y, Yeh I H, Liao H Y M. 2023: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, BC, Canada, 7464–7475.

Yu D, Ji S, Li X, et al. 2022: Earthquake crack detection from aerial images using a deformable convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing* 60: 1-12.

Yu Z, Huang H, Chen W, et al. 2022: Yolo-facev2: A scale and occlusion aware face detector. ar**v preprint ar**v: 2208.02019.

Zhao Yixin, Many, Sun Bo, et al. 2021: Identification of mining ground fractures based on UAV infrared remote sensing and edge detection technology. *Journal of China Coal Society* 46(2): 624-637. (in Chinese)

Zhang Xing-Hang, Zhu Lin, Wang Wei, et al. 2019: Research and application of object-based step extraction method for ground fractures. *Remote Sensing of Land and Resources* 31(1): 87-94. (in Chinese)