electronics

Article

An Automatic Pixel-Level Segmentation Method for Coal-Crack CT Images Based on U^2-Net

Yimin Zhang [1], Chengyi Wu [2], Jinxia Yu [2,*], Guoqiang Wang [1] and Yingying Li [1,3,*]

[1] School of Computer Science, Luoyang Institute of Science and Technology, Luoyang 471023, China; zhangyimin@lit.edu.cn (Y.Z.); wgq@lit.edu.cn (G.W.)
[2] School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China; wuchengyi1997@163.com
[3] State Key Laboratory Cultivation Base for Gas Geology and Gas Control, Henan Polytechnic University, Jiaozuo 454000, China
* Correspondence: jxyu@hpu.edu.cn (J.Y.); liyingying@lit.edu.cn (Y.L.)

Abstract

Automatically segmenting coal cracks in CT images is crucial for 3D reconstruction and the physical properties of mines. This paper proposes an automatic pixel-level deep learning method called Attention Double U^2-Net to enhance the segmentation accuracy of coal cracks in CT images. Due to the lack of public datasets of coal CT images, a pixel-level labeled coal crack dataset is first established through industrial CT scanning experiments and post-processing. Then, the proposed method utilizes a Double Residual U-Block structure (DRSU) based on U^2-Net to improve feature extraction and fusion capabilities. Moreover, an attention mechanism module is proposed, which is called Atrous Asymmetric Fusion Non-Local Block (AAFNB). The AAFNB module is based on the idea of Asymmetric Non-Local, which enables the collection of global information to enhance the segmentation results. Compared with previous state-of-the-art models, the proposed Attention Double U^2-Net method exhibits better performance over the coal crack CT image dataset in various evaluation metrics such as PA, mPA, MIoU, IoU, Precision, Recall, and Dice scores. The crack segmentation results obtained from this method are more accurate and efficient, which provides experimental data and theoretical support to the field of CBM exploration and damage of coal.

Keywords: crack segmentation; coal; CT image; attention mechanism; deep learning

check for updates

Academic Editor: George A. Tsihrintzis

Received: 16 September 2025
Revised: 17 October 2025
Accepted: 24 October 2025
Published: 26 October 2025

Citation: Zhang, Y.; Wu, C.; Yu, J.; Wang, G.; Li, Y. An Automatic Pixel-Level Segmentation Method for Coal-Crack CT Images Based on U^2-Net. *Electronics* **2025**, *14*, 4179. https://doi.org/10.3390/electronics14214179

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the deepening of coal and rock mining, coal and rock dynamic disasters are becoming increasingly severe. The study of surrounding rock stability is essential to ensure the safe construction of coal tunnels, efficient extraction of coalbed methane, and improvement of coal mine economic benefits [1]. This study aims to understand the distribution and characteristics of structural planes such as fractures and fracture zones in coal and rock. With the advancement of X-ray computed tomography (CT), it is now possible to acquire the internal mesostructure of coal, which helps establish a connection between microstructure and properties. This technology also facilitates the creation of accurate 3D reconstruction models. In the realm of numerical calculation, there is a growing trend to investigate the actual composition of coal and rock to establish the most accurate numerical models. Digital image technology can assist researchers in precisely analyzing the physical and mechanical characteristics of rocks. Accurate segmentation of cracks in coal-sample

CT images is crucial for constructing realistic coal structures and numerical models, which can aid in understanding variations in coal rock fracture under load. However, traditional threshold segmentation methods rely on manual threshold settings, making it difficult to avoid noise interference. Therefore, alternative segmentation techniques should be explored to improve accuracy and reduce noise interference. While algorithms exist to determine the threshold, they are not ideal for crack CT images with varying backgrounds and noise levels [2,3]. Therefore, how to effectively extract fractures from small samples of images is a problem to be solved [4]; the accurate and automatic segmentation of cracks in CT images remains a challenging problem for researchers.

Deep learning has become a popular tool across various fields, including computer vision. Convolutional neural networks have proven particularly effective in image classification tasks, with models such as AlexNet [5], VGG [6], ResNet [7], Inception [8,9], and Xception [10] achieving high accuracy. However, the fully connected layers in these networks can hinder their effectiveness in image segmentation tasks. Long et al. [11] proposed Fully Convolutional Neural Networks (FCNs), which involved replacing fully connected layers with deconvolution. This approach enabled pixel-level classification and paved the way for semantic segmentation in image processing. Several subsequent networks have incorporated the concept of FCN and shown promising results, including SegNet [12], U-Net [13], and the DeepLab series [14–17]. In their study, Ali et al. [18] incorporated residual connections into the design of a convolutional neural network for crack segmentation in natural conditions. They also introduced a novel loss function to address the imbalance between positive and negative samples in crack images. Zhang et al. [19], meanwhile, developed a hybrid model by combining object detection and semantic segmentation models. This model accurately detects and segments cracks on bridge surfaces, providing a useful tool for maintenance and repair efforts. While this method achieves effective real-time monitoring, the use of a lightweight model may result in reduced segmentation accuracy. Segmenting coal cracks in CT images requires preserving the original topological structures of cracks, which is challenging due to differences between natural and CT conditions. Furthermore, achieving high segmentation accuracy is crucial for this process.

U-Net is a high-performance tool extensively used for medical segmentation and industrial CT image analysis [20–22]. Its U-shaped structure has proven effective for CT image segmentation. Additionally, various methods have emerged to enhance its feature extraction and fusion capabilities. Residual U-Net (ResU-Net) [23] replaces U-Net's original encoder with ResNet for road extraction in images. Alom et al. [24] also utilized the residual concept and combined it with RCNN [25] to propose the Recurrent Residual U-Net (R2U-Net) for medical image segmentation. Li et al. [26] proposed Dense U-Net, which applied the dense connection idea from DenseNet [27] to segment organ CT images. Meanwhile, Attention U-Net [28] introduced soft attention and proposed an Attention Gate (AG) structure to enhance signals in the region of interest. U-Net++ [26] introduces a nested U-structure by adding sub-networks of varying depths to the original U-Net, resulting in improved performance. Similarly, U2-Net [29], designed for salient object detection (SOD), also utilizes a nested structure. U2-Net achieved excellent performance by applying Residual U-Blocks (RSU) in each block to collect more abundant contextual information. The nested design concept in this approach is noteworthy and worth further exploration. To accurately segment coal cracks in CT images, a deep learning model with strong feature extraction capabilities and multi-level, multi-scale feature fusion is crucial. The key challenge lies in leveraging the feature extraction capabilities of convolution operations to construct a neural network structure that achieves optimal performance.

Unlike naturally occurring cracks, cracks in CT slices generated during coal sample fracturing experiments are often distributed throughout the entire image. Ref. [30]

conducted research on a multi-scale distribution feature recognition method for coal and rock based on CT digital-core deep learning. Convolutional neural networks have a limited receptive field and primarily extract local features, which can negatively impact the accuracy of segmenting large objects. Therefore, utilizing self-attention [31] is crucial for accurate segmentation of coal-crack CT images. This technique, previously applied in natural language processing, enables differentiation of content weights during information processing. In computer vision, a typical characteristic is that feature extraction for each pixel is associated with every pixel in the entire image. This algorithm adds rich global information to feature maps. The non-local module [32] was proposed to introduce a self-attention mechanism that can be integrated into convolutional neural networks. As image data typically contain more information than natural language processing data, Non-Local is applied to feature maps of the neural network's intermediate layers to avoid excessive computation. Despite its benefits, Non-Local still suffers from slow operation and redundant information. In response, CCNet [33] introduced a Criss-Cross module, which can be regarded as a more efficient version of Non-Local. Connecting two such modules enables collection of complete global information, resulting in better performance than the original Non-Local approach. In their study, ANN [34] utilized Non-Local's matrix keys and values to reduce computation and eliminate unnecessary data. They also introduced a fusion module embodying the concept of mutual attention. Res-VGG-U-Net [35] uses the Residual Network–Visual Geometry Group–U-Net deep learning framework in conjunction with two threshold segmentation methods to capture components like fractures, the drilled hole, and minerals. VM-U-Net++ [36] strategically integrates the strengths of the Mamba architecture and U-Net to significantly improve the accuracy of crack segmentation. These methods aim to enhance feature maps with comprehensive global information while minimizing redundant operations and eliminating superfluous data to improve segmentation accuracy. Recently, a novel U-Net-based MCSN [37] model has leveraged transfer learning and multi-scale dilated convolutions to segment coal fractures from Micro-CT images, enhancing gas drainage efficiency. A Trans-U-Net [38] framework that seamlessly integrates transformer-based global context with U-Net's local refinement was proposed to automatically delineate micro-fractures in μCT coal images, outperforming existing models in accuracy and robustness. Its accuracy could reach 90%.

This paper proposes Attention Double U^2-Net, an end-to-end deep learning model for coal-crack segmentation in CT images. Given the scarcity of publicly available data on coal mine CT fractures [39], the authors built a dedicated dataset specifically for coal-crack CT image segmentation. Herein, we propose a Double Residual U-Block (DRSU) based on U^2-Net. This module aims to enhance the feature extraction and fusion capabilities of the original module. The DRSU comprises two serially connected RSU modules with skip connections added between them. The SOD model was modified into a segmentation model to generate binary predictions. To incorporate global information into feature maps, a new block termed Atrous Asymmetric Fusion Non-Local Block (AAFNB) is introduced. This block enhances the sampling method of ANN via atrous convolution and introduces a novel multi-scale fusion attention structure. The AAFNB block is placed between the encoder and decoder, where it fuses the global information of high-level feature maps.

In summary, the proposed method for segmentation of coal-crack CT images has the following three-fold contributions:

1. In this study, we introduce an end-to-end deep learning model for the automatic segmentation of coal cracks in CT images. Our proposed model utilizes a Double RSU structure that efficiently extracts and fuses high-quality features in each block of the encoder and decoder. Our experiments demonstrate that this structure significantly improves the accuracy of segmentation.

2. In this paper, we propose the AAFNB attention module to enhance the collection of global feature information. To evaluate the effectiveness of this module, we conducted ablation experiments and found that it outperformed some typical attention modules.
3. To support our experiment and for performance evaluation, we established a coal-crack CT image dataset. To ensure the segmentation results are directly applicable to subsequent studies, we retained the complete complex crack structure of the original image instead of cutting it into blocks.

2. Proposed Method

The process of coal-crack segmentation from CT images can be divided into three parts, as illustrated in Figure 1. The first part involves data acquisition and processing, which results in obtaining the original images from the dataset. The second part involves building the segmentation network, which uses the Attention Double U^2-Net The final part involves training the network and using it to evaluate the model, resulting in the segmentation results.

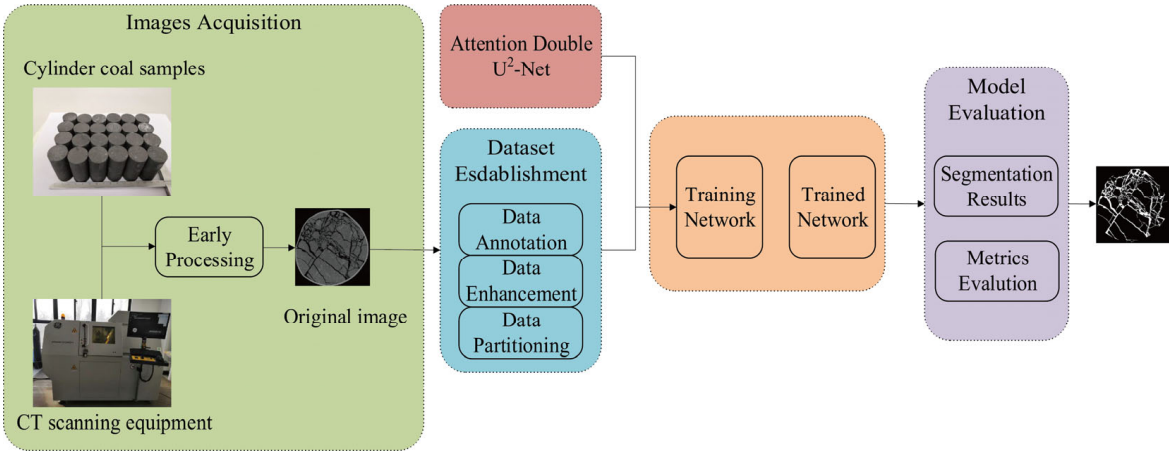


Figure 1. The schematic diagram of coal-crack segmentation of CT images.

2.1. Dataset

Since there are scarce public datasets on the segmentation of coal cracks in CT images, we built a dedicated dataset. Some CT slices with complex structures in the process of coal fracturing experiments are selected as the original data for the dataset. The CT scanning equipment is a phoenix v | tome | x m high-resolution industrial X-ray μCT scanner. It is from the Nation Key of Natural Gas Geology and Natural Gas Control of Henan University Polytechnic laboratory. The equipment diagram [34] is shown in Figure 1. The coal sample used for slice scanning was 25 mm by 50 mm cylindrical raw coal. In order to allow our method to directly segment the original image, all labeling work was without any cutting or other preprocessing.

Most of the images have a resolution of approximately 20 μm, and a total of 742 images were selected. To generate segmentation labels for complex cracks, a hybrid method combining manual annotation and threshold segmentation was adopted (see Figure 2 for the process schematic). The core workflow consists of three steps.

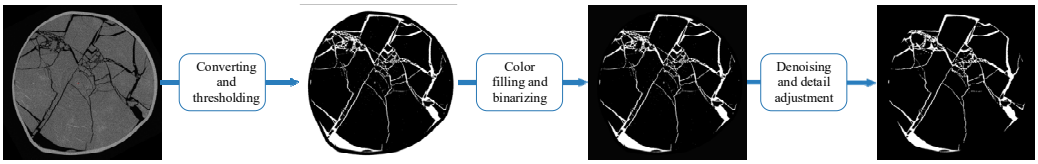


Figure 2. Schematic diagram of dataset production process.

The first step is the conversion and thresholding phase. First, the original three-channel images are converted to an 8-bit single-channel format to support subsequent binarization operations. Then, threshold segmentation is performed by manually calibrating the threshold to optimize the segmentation results.

The second step is the color filling and binarization phase. For the white edges caused by grayscale overlap between the edges of the original images and crack pixels, black filling is applied to optimize the binarized output.

The third step is the denoising and detail adjustment phase. Residual noise—caused by consistent grayscale values between coal matrix pixels and crack pixels—is removed via color filling. Subsequently, the processed images are compared, verified, and adjusted repeatedly with the details of the original images by multiple personnel to minimize deviations caused by human factors. Finally, the ground truth labels are generated.

Data augmentation can generate more equivalent data from limited data, and is valuable for deep learning work. Since we wanted to ensure the integrity of the coal sample slices in images, only rotation and reduction were used to augment the dataset. Specifically, the rotation angle is limited to the range of -30 to 30 degrees, and the scaling ratio is limited to the range of 80% to 120%. A total of 742 original images and the labels are augmented to 6000. An amount of 4200 of them are used as the training set, and the remaining 1800 images are used as the test set.

2.2. Attention Double U^2-Net Structure

Figure 3 shows the structure of our proposed Attention Double U^2-Net. The overall model adopts the nested U-structure and skip connections of U^2-Net, comprising six downsampling stages and five upsampling stages.The last two blocks of the encoder and the first block of the decoder applied RSU4F, the same as U^2-Net, which is the dilated version of RSU and does not include upsampling and downsampling. Because the size of this layer of feature maps is already very small, downsampling will result in a significant loss of pixels. All other blocks applied the proposed Double RSU structure. The purpose of jump connection is to integrate feature maps with different degrees of feature extraction. Although the scales of the two feature maps are the same, the information contained in the feature map obtained through downsampling and the feature map after feature extraction and pixel restoration is different. By integrating this information, it can effectively reduce the semantic gap between the encoder and the decoder, thereby improving the feature extraction ability of the model.

In addition, to adapt the salient object detection model to the segmentation task, this paper adjusts the saliency map fusion part of the original model. The output of the original model has pixel values ranging from 0 to 255, which is not the binary map required for image segmentation. Therefore, this paper converts the grayscale image into a binary image through one-hot encoding before generating the result image. Meanwhile, each feature map output by the DRSU in the decoder is upsampled to the output image size, concatenated, and then fed into the classifier to obtain the final segmentation result, which forms the feature fusion part of the model. The purpose of feature fusion is to integrate multi-scale information: high-level feature maps contain long-range information with large receptive fields, while low-level feature maps contain more detailed information. The fusion of these two types of information helps improve the final segmentation performance, especially for the segmentation of small targets such as fissures. The attention mechanism module AAFNB proposed in this paper is located at the bridge, and its input is the output results of the last two DRSUs in the encoder, which are used for attention computation.

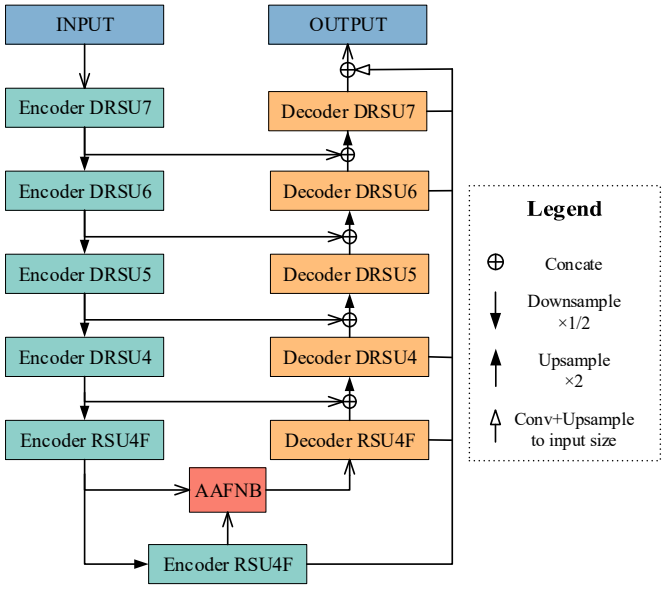


Figure 3. The overall structure of Attention Double U^2-Net.

2.3. Double Residual U-Block Structure

The RSU structure in U^2-Net is an efficient and lightweight U-structure feature extractor. Double RSU structure is proposed to improve the feature extraction and fusion capability of the original structure. The skip connections are added between blocks belonging to two RSUs to collect multi-scale information. The DRSU5 structure is shown in Figure 4. The encoder part of each RSU includes 5 convolution layers, as does the decoder part. Similarly, each part in RSU7 has 7 such convolutional layers.

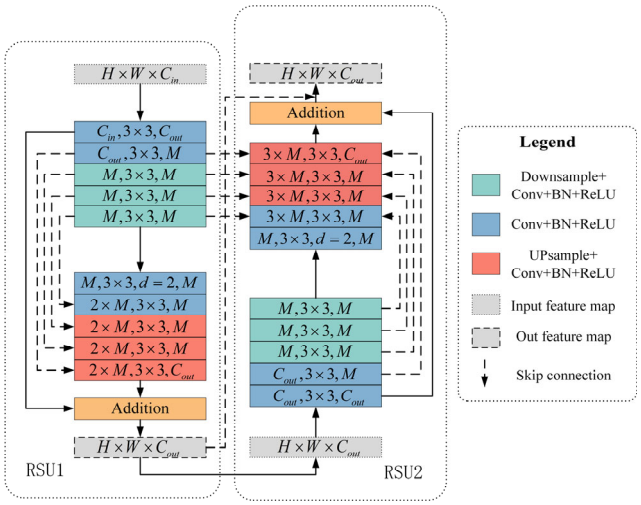


Figure 4. Structure of Double Residual U-Block 5.

The design of DRSU is tailored to address the unique challenges of coal-crack CT image feature extraction. Coal cracks in CT images are often fine (width < 5 pixels at 20 μm resolution), discontinuous, and mixed with a noisy coal matrix (grayscale overlap between matrix and cracks), while also requiring the retention of global topology. The two serially connected RSU modules in DRSU undertake hierarchical tasks:

The first RSU (RSU1) focuses on extracting low-level edge details via shallow convolutions to avoid losing thin crack features.

The second RSU (RSU2) aggregates these low-level features to extract high-level semantic information (e.g., global crack distribution).

The added skip connections between RSU1's encoder and RSU2's decoder directly concatenate low-level edges with high-level semantics, which mitigates the "semantic gap" (a common issue in single RSU modules) and ensures that fine crack details are not distorted during upsampling—effectively balancing local detail retention and global context integration for coal-crack segmentation.

In addition to connecting two RSU in series, feature maps of two encoder parts are concatenated to the last decoder as an input map. This skip connection method causes the last decoder to process maps containing three slightly different pieces of information, which fuses different degrees of feature extraction and different scales. These kinds of information can complement each other, making the output results closer to the ground truth. Moreover, two output maps of each RSU are concatenated as the final output map.

2.4. Atrous Asymmetric Fusion Non-Local Block

In our dataset, cracks in images tend to have large scales close to the image size. Even if after multiple-times image downsampling, the receptive field of the convolution operation is limited. The classification of each pixel cannot connect all pixels of the same object when performing the convolution operation. To solve this problem, we proposed an Atrous Asymmetric Fusion Non-Local Block and set it between the encoder and decoder of Double U^2-Net. In the conventional self-attention mechanism, both *key* and *value* represent dense global information and have a lot of redundant information. So, the AAFNB block contains a sampling structure to sample the *key* and *value* matrices. The sampling structure is shown in Figure 5.

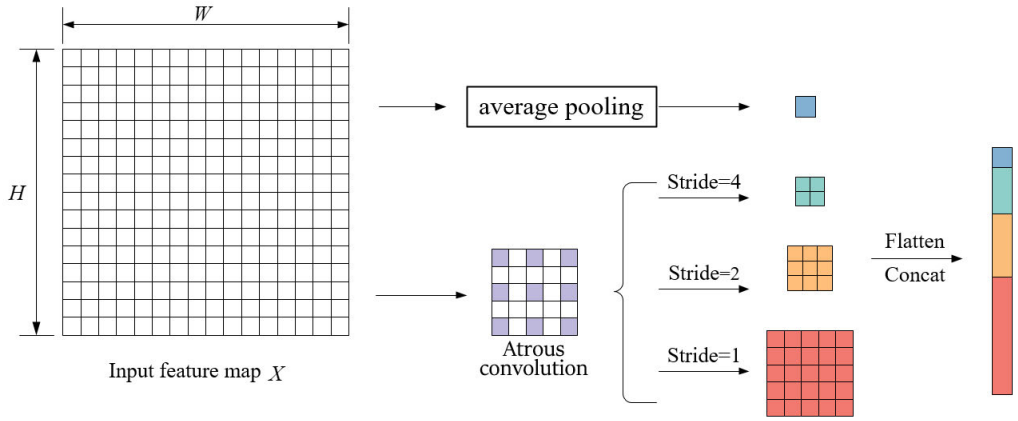


Figure 5. Feature extraction and sampling process.

In the sampling structure, both atrous convolution and global average pooling are applied. Three 3×3 and dilation = 2 atrous convolutions with different strides are used to short-range feature extract and global sample in *key* and *value*. In addition, a global average pooling is applied for high-level global information simultaneously. Strides of atrous convolution are set to 1, 2, and 4—this selection is grounded in the statistical size distribution of coal cracks in the established dataset, where stride = 1 targets fine cracks, stride = 2 captures medium cracks, and stride = 4 covers large continuous cracks—and the output size of average pooling is 1/8 of the input map. It is equivalent to extracting useful information from the original matrix and sampling, which can remove the original redundant information. This structure has an input feature map of $X \in \mathbb{R}^{C \times H \times W}$, where C, H, and W express channel numbers, height, and width of the input map, respectively. The sampling process SMP can be described in the following formula:

$$SMP = F(A_1, A_2, A_4, P) \tag{1}$$

where F denotes flattened concatenation, A_i expresses atrous convolution of $stride = i$, and P expresses average pooling. The numbers of spatial locations are $N = H \times C$. After the sampling process, the output feature map can be described as follows:

$$SMP(X \in \mathbb{R}^{C \times H \times W}) = \hat{X} \in \mathbb{R}^{C \times S} \tag{2}$$

where S is the number of points extracted and sampled from N.

The structure of AAFNB is shown in Figure 6. There are two parts of it: a fusion part and a self-attention part. The fusion module aims to fuse different levels of global information, taking low-level feature maps as *key* and *value*, and high-level feature maps as *query*. The *key*, *value*, and *query* of the self-attention part all come from the low-level feature map. After the two attention parts obtain the results, the two output feature maps will be concatenated and fused again. AAFNB has two feature maps as an input: a high-level map $X_h \in \mathbb{R}^{C_h \times H_h \times W_h}$ and a low-level map $X_l \in \mathbb{R}^{C_l \times H_l \times W_l}$ from the sixth and fifth blocks in the encoder, respectively. In the fusion part, firstly, $X_h \in \mathbb{R}^{C_h \times H_h \times W_h}$ and $X_l \in \mathbb{R}^{C_l \times H_l \times W_l}$ are transformed into $Q_h \in \mathbb{R}^{\hat{C} \times H_h \times W_h}$, $K_l \in \mathbb{R}^{\hat{C} \times H_l \times W_l}$, and $V_l \in \mathbb{R}^{\hat{C} \times H_l \times W_l}$ by 1×1 convolution of C_Q^h, C_K^l, and C_V^l as follows:

$$Q_h = C_Q^h(X_h),\ K_l = C_K^l(X_l),\ V_l = C_V^l(X_l). \tag{3}$$

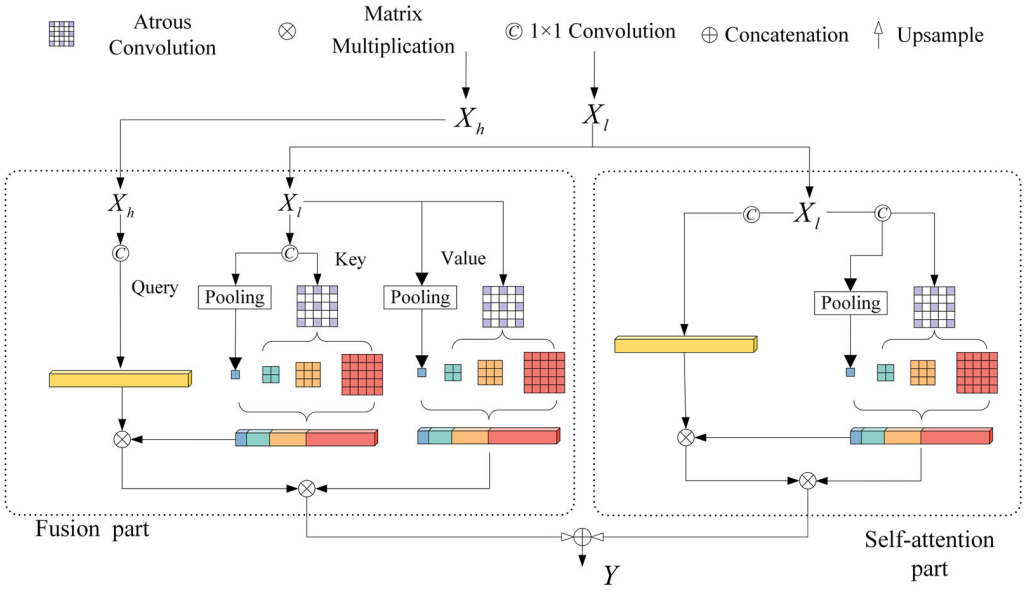


Figure 6. Structure of Atrous Asymmetric Fusion Non-Local Block (AAFNB).

Then, K_l and V_l are sampled by SMP to $\hat{K}_l \in \mathbb{R}^{\hat{C} \times S}$ and $\hat{V}_l \in \mathbb{R}^{\hat{C} \times S}$, and Q_h is flattened to $\hat{Q}_h \in \mathbb{R}^{\hat{C} \times N}$ as follows:

$$\hat{K}_l = SMP(K_l),\ \hat{V}_l = SMP(V_l),\ \hat{Q}_h = F(Q_h) \tag{4}$$

Then the similarity matrix $M_F \in \mathbb{R}^{N_h \times S}$ between $\hat{Q}_h$ and $\hat{K}_l$ is computed by a matrix multiplication:

$$M_F = \hat{Q}_h^T \times \hat{K}_l \tag{5}$$

Then, the $O_F \in \mathbb{R}^{N_h \times \hat{C}}$ is inferred by a matrix multiplication:

$$O_F = M_F \times \hat{V}_l^T \tag{6}$$

By the 1×1 convolution C_O, channel number of O_F is complemented to C_h and concatenated with X_h. Output of the fusion part is obtained by all these steps as follows:

$$Y_F = cat(C_O(O_F), X_h) \tag{7}$$

The output of self-attention part Y_S can infer in a similar way. The next step is to upsample Y_F to the size of Y_S, same as X_l, and concatenate. Finally, we have the output of AAFNB as follows:

$$Y = cat(upsample(Y_F), Y_S) \tag{8}$$

Y is filled with a wide and short range of clues from different levels of features. This kind of feature map is valuable for crack segmentation.

3. Experiments and Results

All evaluations in this paper are under our established coal-crack dataset of CT images. Models were trained and tested in the following environment: Ubuntu 20.04, AMD 3700X CPU, and NVIDIA 2070S GPU. All code applies the deep learning framework Pytorch 1.13.1.

For all models, including our AAFNB and other models used for comparison, the same set of training hyperparameters was adopted. Specifically, CUDA acceleration was employed, the input image size was set to 512×512, the number of training epochs was 100, the batch size was 1, and the learning rate was 1×10^{-4}. These settings fully ensure the fairness and referability of the comparison among different models.

In the next section, some authoritative evaluation metrics will be used to analyze the performance of the proposed method. Moreover, ablation experiments were designed to verify the validity of the proposed structure DRSU and the module AAFNB. Finally, some high-performance methods will be compared with the proposed method.

3.1. Metrics

To verify the effectiveness of our method, we chose several authoritative evaluation metrics to evaluate the model. The following are for semantic segmentation: Pixel Accuracy, Mean Pixel Accuracy, and Mean Intersection over Union. When TP, TN, FP, and FN represent True Positive, True Negative, False Positive, and False Negative, respectively, these metrics can be expressed as follows:

1. Pixel Accuracy (PA): $PA = \frac{(TP+TN)}{TP+TN+FP+FN}$.
2. Mean Pixel Accuracy (mPA): $mPA = \frac{\sum_{i=1}^{N} PA_i}{N}$, where N means the class number, and PA_i means the Pixel Accuracy of each class.
3. Mean Intersection over Union (MIoU): $MIoU = \frac{1}{2}(\frac{TP}{TP+FP+FN} + \frac{TN}{TN+FN+FP})$.

In order to make the model evaluation more convincing, three typical metrics in the segmentation of CT images are used as well, which can be represented as follows:

1. Precision (P): $P = \frac{TP}{TP+FP}$
2. Recall (R): $R = \frac{TP}{TP+FN}$
3. Dice Score: $Dice = \frac{2TP}{2TP+FP+FN}$

Moreover, we introduce the PR curve to evaluate models more intuitively.

3.2. Ablation Experiments

Firstly, in order to verify the effectiveness of our proposed DRSU structure, the original U^2-Net is compared with the Double U^2-Net without AAFNB. The results are shown in Tables 1 and 2. As cracks are the main target of images, the IoU and Dice Score of cracks

are listed separately, and the P and R listed are of cracks. The Double U^2-Net structure is abbreviated as D-U^2-Net in following tables.

Table 1. Comparison of U^2-Net and Double U^2-Net under semantic segmentation evaluation metrics.

Methods	PA (%)	mPA (%)	MIoU	IoU (Crack)
U^2-Net	98.87	98.22	94.26	89.76
D-U^2-Net	99.02	98.45	94.80	90.68

Table 2. Comparison of U^2-Net and Double U^2-Net under crack segmentation evaluation metrics.

Methods	P (%)	R (%)	Dice	Dice (Crack)
U^2-Net	95.32	93.89	96.99	94.60
D-U^2-Net	96.11	94.13	97.29	95.11

Compared to the original U^2-Net, the proposed Double U^2-Net improves the PA, mPA, MioU, and crack IoU by 0.15%, 0.23%, 0.54%, and 0.92%. The P, R, Dice Score, and crack Dice Score of Double U^2-Net are improved by 0.79%, 0.24%, 0.30%, and 0.51%.

Moreover, in order to verify the effectiveness of the proposed AAFNB, several current classic or representative attention mechanism modules are selected to make a comparison: Non-Local Module (NM), Criss-Cross Attention Module (CCAM), and Asymmetric Non-Local Module (ANM). All these modules were plugged at the bridge position of the Double U^2-Net structure. The results are shown in Tables 3 and 4. Compared with the second-best-performing ANM among all these methods, the proposed AAFNB has a 0.04%, 0.08%, 0.32%, and 0.60% higher PA, mPA, MioU, and crack MIoU, respectively. Moreover, the proposed module has a 0.65%, 0.02%, 0.18%, and 0.33% higher P, R, Dice Score, and crack Dice Score than ANM, respectively. Although the P of AAFNB is 0.04% lower than that of CCAM, the other three metrics have substantially better performance.

Table 3. Results of ablation experiments under semantic segmentation evaluation metrics.

Methods	PA	mPA	MIoU	IoU (Crack)
D-U^2-Net	99.02	98.45	94.80	90.68
D-U^2-Net + NM	99.03	97.88	94.89	90.89
D-U^2-Net + CCAM	99.05	97.89	94.95	90.94
D-U^2-Net + ANM	99.05	98.24	95.00	91.04
D-U^2-Net + AAFNB (ours)	99.09	98.75	95.32	91.64

Table 4. Results of ablation experiments under crack segmentation evaluation metrics.

Methods	P (Crack)	R (Crack)	Dice	Dice (Crack)
D-U^2-Net	96.11	94.13	97.29	95.11
D-U^2-Net + NM	96.14	94.34	97.37	95.23
D-U^2-Net + CCAM	96.35	94.18	97.38	95.26
D-U^2-Net + ANM	95.66	94.96	97.39	95.31
D-U^2-Net + AAFNB (ours)	96.31	94.98	97.57	95.64

3.3. Comparing with State-of-the-Art Methods

Several high-performance U-Net series models are compared with the proposed method: U-Net, Attention U-Net, Recurrent Residual U-Net (R2U-Net), U-Net++, and Residual Attention U-Net (RAU-Net). Tables 5 and 6 show the comparison results for our dataset. It can be seen that except for individual metrics, overall, the proposed method outperforms the second-best-performing method RAU-Net. The proposed method has a

0.20%, 0.83%, 1.04%, and 1.85% higher PA, mPA, MioU, and crack IoU than the RAU-Net, respectively. It has a 0.76%, 1.27%, 0.57% and 1.02% higher P, R, Dice Score, and crack Dice Score, respectively.

Table 5. Comparison with state-of-the-art under semantic segmentation evaluation metrics.

Methods	PA (%)	mPA (%)	MIoU	IoU (Crack)
U-Net	98.35	98.45	92.00	85.83
Attention U-Net	97.72	98.71	93.46	88.35
R2U-Net	97.74	96.38	93.67	88.73
U-Net++	98.85	95.04	93.80	88.87
RAU-Net	98.89	97.92	94.28	89.79
Attention D-U^2-Net (ours)	99.09	98.75	95.32	91.64

Table 6. Comparison with state-of-the-art under crack segmentation evaluation metrics.

Methods	P (Crack)	R (Crack)	Dice	Dice (Crack)
U-Net	90.26	94.58	95.91	92.37
Attention U-Net	95.18	92.48	96.55	93.81
R2U-Net	94.12	93.94	96.67	94.03
U-Net++	96.30	92.02	96.74	94.11
RAU-Net	95.55	93.71	97.00	94.62
Attention D-U^2-Net (ours)	96.31	94.98	97.57	95.64

The PR curves for the segmentation results in our dataset are shown in Figure 7. The performance of different models depends on the area under the curves. Obviously, the PR curve generated by the proposed method can wrap all other curves. This shows that the proposed method achieves the best performance on the coal dataset.

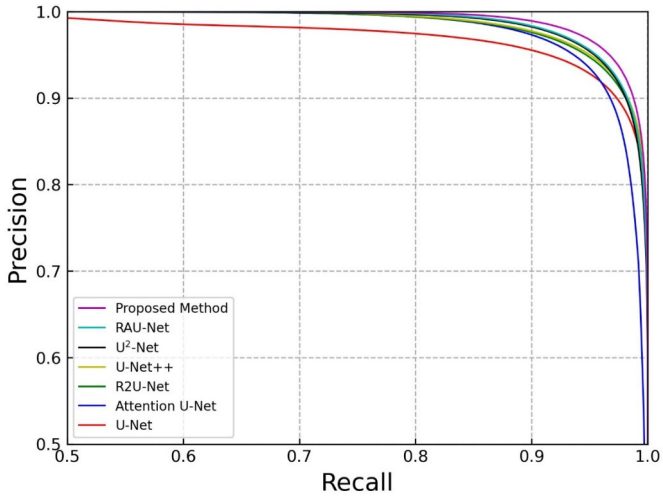


Figure 7. PR curves for crack segmentation in our dataset.

Figure 8 shows several segmentation samples predicted by all these methods. The red circles highlight differences between the proposed method and the others. Firstly, the proposed method effectively avoids some noises caused by misidentified coal matrices. This phenomenon regularly appears in the result maps of complex structures, such as the results of U^2-Net, U-Net, R2U-Net, and U-Net++. Moreover, some small intermittent cracks are ignored in the result maps of Attention U-Net. In addition, U-Net and RAU-Net mistakenly identify the black background as cracks in some result maps. In a word, the proposed method effectively avoids the above problems, and obtains the closest results to the ground truth.

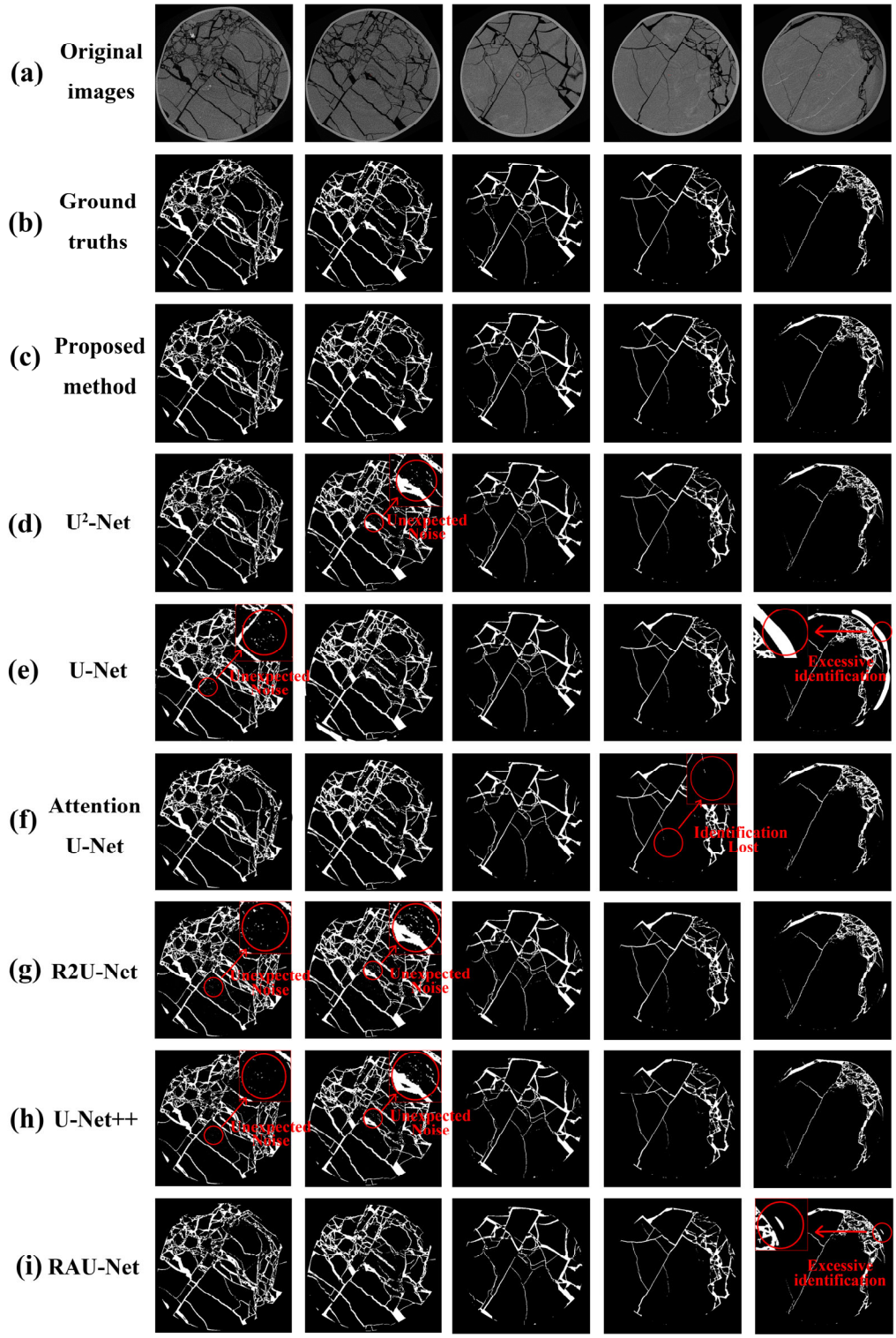


Figure 8. Results predicted by proposed method and others. (**a**) Original images; (**b**) ground truths; (**c**) proposed method; (**d**) U^2-Net; (**e**) U-Net; (**f**) Attention U-Net; (**g**) R2U-Net; (**h**) U-Net++; (**i**) RAU-Net.

4. Conclusions

Automatic coal-crack segmentation of CT images is critical for mine safety, requiring high segmentation precision. In this paper, an end-to-end pixel-level U-structured method based on deep learning was proposed. Furthermore, a CT image dataset of coal cracks

was constructed, and the proposed method achieves superior performance compared with existing approaches.

Based on the U-structured model, we integrated the DRSU structure and skip connections into U^2-Net. We also proposed AAFNB, an attention block designed to incorporate global information into feature maps. The experiments show that the proposed method is effective and suitable for crack segmentation in CT image contexts.

In terms of practical application, this method offers significant value for coal mine safety management and coalbed methane (CBM) exploration. During underground mining operations, the high-precision segmentation results allow real-time quantification of fracture size, orientation, and density, providing data support for roof stability assessment and early warning of geological hazards like collapses. For CBM exploration, the segmented fracture network can further facilitate 3D pore structure reconstruction and the calculation of reservoir permeability and connectivity, thereby guiding fracturing design and well placement to enhance gas extraction efficiency.

In the future, we will expand the dataset scale and increase its diversity and complexity. Additionally, the proposed AAFNB is constrained by high computational costs, which limits its application to low-level feature maps. To address this issue and improve the model's temporal efficiency (a key requirement for industrial on-site deployment), we plan to explore lightweight optimization strategies—including pruning non-critical parameters, adaptive compression of feature map dimensions, and simplification of the AAFNB module's computational logic—while maintaining core segmentation precision.

Author Contributions: Conceptualization, J.Y. and Y.L.; data curation, Y.Z. and C.W.; funding acquisition, Y.L.; methodology, Y.L.; resources, Y.L.; validation, Y.Z., C.W. and G.W.; writing—original draft, Y.Z. and C.W.; writing—review and editing, J.Y. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 62203203, and the State Key Laboratory Cultivation Base for Gas Geology and Gas Control (Henan Polytechnic University), grant number WS2020B13.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. Yuan, L.; Wang, E.; Ma, Y.; Liu, Y.; Li, X. Research progress of coal and rock dynamic disasters and scientific and technological problems in China. *J. China Coal Soc.* **2023**, *48*, 1825–1845.
2. Akagic, A.; Buza, E.; Omanovic, S.; Karabegovic, A. Pavement crack detection using Otsu thresholding for image segmentation. In Proceedings of the 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 21–25 May 2018; pp. 1092–1097.
3. Sha, C.; Hou, J.; Cui, H. A robust 2D Otsu's thresholding method in image segmentation. *J. Vis. Commun. Image Represent.* **2016**, *41*, 339–351. [CrossRef]
4. Wang, W.; Zhou, L. Fracture Extraction from Logging Image Using a Dual Encoder-Decoder Architecture with Swin Transformer. *Petrophysics* **2023**, *64*, 38–49. [CrossRef]
5. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*; Morgan Kaufmann Publishers, Inc.: San Francisco, CA, USA, 2012; p. 25.
6. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
7. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [CrossRef]

8. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A.; Liu, W.; et al. Going Deeper with Convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
9. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
10. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
11. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 640–651. [CrossRef]
12. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]
13. Ronneberger, O.; Fischer, P.; Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*; Springer International Publishing: Cham, Switzerland, 2017. [CrossRef]
14. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *arXiv* **2014**, arXiv:1412.7062.
15. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef] [PubMed]
16. Yu, J.; Wu, C.; Li, Y.; Zhang, Y. Intelligent identification of coal crack in CT images based on deep learning. *Comput. Intell. Neurosci.* **2022**, *2022*, 7092436. [CrossRef] [PubMed]
17. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*; Springer: Cham, Switzerland, 2021.
18. Ali, R.; Chuah, J.H.; Abu Talip, M.S.; Mokhtar, N.; Shoaib, M.A. Automatic pixel-level crack segmentation in images using fully convolutional neural network based on residual blocks and pixel local weights. *Eng. Appl. Artif. Intell.* **2021**, *104*, 104391. [CrossRef]
19. Zhang, J.; Qian, S.; Tan, C. Automated bridge surface crack detection and segmentation using computer vision-based deep learning model. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105225. [CrossRef]
20. Su, Z.; Li, W.; Ma, Z.; Gao, R. An improved U-Net method for the semantic segmentation of remote sensing images. *Appl. Intell.* **2022**, *52*, 3276–3288. [CrossRef]
21. Wang, Y.; Ye, H.; Cao, F. A novel multi-discriminator deep network for image segmentation. *Appl. Intell.* **2022**, *52*, 1092–1109. [CrossRef]
22. Zhao, J.; Dang, M.; Chen, Z.; Wan, L. DSU-Net: Distraction-Sensitive U-Net for 3D lung tumor segmentation. *Eng. Appl. Artif. Intell.* **2022**, *109*, 104649. [CrossRef]
23. Zhang, Z.; Liu, Q.; Wang, Y. Road Extraction by Deep Residual U-Net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1–5. [CrossRef]
24. Alom, Z.; Yakopcic, C.; Hasan, M.; Taha, T.M.; Asari, V.K. Recurrent residual U-Net for medical image segmentation. *J. Med. Imaging* **2019**, *6*, 014006. [CrossRef] [PubMed]
25. Ming, L.; Hu, X. Recurrent convolutional neural network for object recognition. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3367–3375. [CrossRef]
26. Li, X.; Chen, H.; Qi, X.; Dou, Q.; Fu, C.-W.; Heng, P.-A. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes. *IEEE Trans. Med. Imaging* **2018**, *37*, 2663–2674. [CrossRef] [PubMed]
27. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [CrossRef]
28. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**, arXiv:1804.03999. [CrossRef]
29. Qin, X.; Zhang, Z.; Huang, C.; Dehghan, M.; Zaiane, O.R.; Jagersand, M. U^2-Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognit.* **2020**, *106*, 107404. [CrossRef]
30. Feng, X.; Shen, Y.; Zhou, D.; Wang, J.; Wang, M. Multi-scale distribution of coal fractures based on CT digital core deep learning. *Coal Sci. Technol.* **2023**, *51*, 97–104.
31. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 5998–6008.
32. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-Local neural networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; Volume 31, pp. 7794–7803.
33. Huang, Z.; Wang, X.; Huang, L.; Huang, C.; Wei, Y.; Liu, W. Ccnet: Criss-cross attention for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27–28 October 2019; pp. 603–612.

34. Zhu, Z.; Xu, M.; Bai, S.; Huang, T.; Bai, X. Asymmetric nonlocal neural networks for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 593–602.
35. He, C.; Sadeghpour, H.; Shi, Y.; Mishra, B.; Roshankhah, S. Mapping distribution of fractures and minerals in rock samples using Res-VGG-UNet and threshold segmentation methods. *Comput. Geotech.* **2024**, *175*, 106675. [CrossRef]
36. Tang, W.; Wu, Z.; Wang, W.; Pan, Y.; Gan, W. VM-UNet++ research on crack image segmentation based on improved VM-UNet. *Sci. Rep.* **2025**, *15*, 8938. [CrossRef] [PubMed]
37. Wang, D.; Fang, Y.; Wei, J.; Zhang, H.; Zhao, L.; Wang, L.; Xia, Y.; Li, L.; Wang, S.; Zhang, Q.; et al. Intelligent extraction of Micro-CT fissures in coal based on deep learning and its application. *Meitan Xuebao J. China Coal Soc.* **2024**, *49*, 3439–3452. [CrossRef]
38. Hu, Z.; Zhang, X.; Jin, Y.; Gong, L.; Huang, W.; Ren, J.; Klitzsch, N. A method for intelligent information extraction of coal fractures based on CT and deep learning. *Coal Geol. Explor.* **2025**, *53*, 55–66. [CrossRef]
39. Lv, Z.; Fan, Y.; Sha, T.; Cui, Y.; Wu, Y.; Lv, H.; Sun, M.; Tu, Y.; Xu, Z.; Wang, W. A large-scale open image dataset for deep learning-enabled intelligent sorting and analyzing of raw coal. *Sci. Data* **2025**, *12*, 403. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.