China Geology 8 (2025) 643−660

China Geology

Journal homepage: http://chinageology.cgs.cn
https://www.sciencedirect.com/journal/china-geology

# Improved SE-UNet network-based semantic segmentation and extraction of hidden geological significance in geological maps

Kai Ma[a, b], Jun-jie Liu[a, b], Si-qi Lu[a, b], Ze-hua Huang[a, b], Miao Tian[d], Jun-yuan Deng[a, b], Zhong Xie[c, d], Qin-jun Qiu[c, d,*]

[a] *Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering, China Three Gorges University, Yichang* 443002, *China*
[b] *College of Computer and Information Technology, China Three Gorges University, Yichang* 443002, *China*
[c] *School of Computer Science, China University of Geosciences, Wuhan* 430074, *China*
[d] *Key Laboratory of Geological Survey and Evaluation of Ministry of Education, China University of Geosciences, Wuhan* 430074, *China*

ARTICLE INFO

*Article history:*
Received 8 December 2023
Received in revised form 29 June 2024
Accepted 16 July 2024
Available online 27 February 2025

*Keywords:*
Geological map
UNet model
Image segmentation
Semantic segmentation
Pixel pre-segmentation
Clustering algorithm
Attention mechanism
Deep learning
Artificial intelligence
Geological survey engineering

ABSTRACT

Automatic segmentation and recognition of content and element information in color geological map are of great significance for researchers to analyze the distribution of mineral resources and predict disaster information. This article focuses on color planar raster geological map (geological maps include planar geological maps, columnar maps, and profiles). While existing deep learning approaches are often used to segment general images, their performance is limited due to complex elements, diverse regional features, and complicated backgrounds for color geological map in the domain of geoscience. To address the issue, a color geological map segmentation model is proposed that combines the Felz clustering algorithm and an improved SE-UNet deep learning network (named GeoMSeg). Firstly, a symmetrical encoder-decoder structure backbone network based on UNet is constructed, and the channel attention mechanism SENet has been incorporated to augment the network's capacity for feature representation, enabling the model to purposefully extract map information. The SE-UNet network is employed for feature extraction from the geological map and obtain coarse segmentation results. Secondly, the Felz clustering algorithm is used for super pixel pre-segmentation of geological maps. The coarse segmentation results are refined and modified based on the super pixel pre-segmentation results to obtain the final segmentation results. This study applies GeoMSeg to the constructed dataset, and the experimental results show that the algorithm proposed in this paper has superior performance compared to other mainstream map segmentation models, with an accuracy of 91.89% and a MIoU of 71.91%.



## 1. Introduction

A large number of geological maps have accumulated in the geological industry, and play a pivotal role across various domains including the preservation geological environments (Do Valle Júnior RF et al., 2019), mineral exploration (Rahimi H et al., 2021; Qiu QJ et al., 2023b) and geological disaster detection (Ma ZJ and Mei G, 2021). However, these reported geological maps are often stored in raster format (semistructured data) after scanning, which is not structured data and thus cannot be directly used for data analysis and knowledge discover. The traditional method is to vectorize them manually, which requires considerable manpower and material resources, as well as technicians with a certain reserve of expertise and experience. However, the data results are inevitably incomplete and inconsistent, and can even have errors. Image recognition (Liu Y and Wu LZ, 2018; Yu HG et al., 2022) and information extraction (Rauch A et al., 2019; Qiu QJ et al., 2023a; Tian M et al., 2023) provide solutions for the understanding of geological map information, and can increase the information base and sources of geoscience research and thus are of great significance to obtain more data and information sources and further develop geoscience analysis and research.

Segmentation of geological maps stands as a pivotal technology crucial for the intelligent analysis and

First author: *E-mail address*: makai@ctgu.edu.cn (Kai Ma).

* Corresponding author: *E-mail address*: qiuqinjun@cug.edu.cn (Qin-jun Qiu).

Literary editor: Li-qiong Jia
doi:10.31035/cg2023146
2096-5192/© 2025 China Geology Editorial Office.

understanding of geological maps, and provides important auxiliary information for researchers in the domain of geoscience. The regional mineral resources, distribution patterns and other information can be obtained by analyzing the information of various objects and related relation in geological maps. Although large breakthroughs and advances have been made in general-purpose domain image segmentation algorithms (such as deep learning), these algorithms cannot be directly applied to geological map segmentation, and the main problems are as follows: (1) At present, most of the automatic image recognition and information extraction technologies demonstrate commandable efficacy in engineering drawings, but their performance remains constrained when applied to geological maps. The main reason is that the elements (such as geological objects) in the map are complicated and overlap with each other, which increases the difficulty of automatic map recognition and information extraction. (2) There are various distortions in the scanned raster color geological map, for example, dispersion and confusion of geological map colors, and noise, burr and bifurcation of geological map elements, which increase the difficulty of automatic recognition and information extraction from geological maps. (3) A scanned raster color topographic map has only four colors to the human eye, and it cannot achieve the same recognition effect for computers. The color separation (color segmentation) of topographic maps is easier, and the information extraction is more convenient. However, geological maps have many and ever-changing colors, which undoubtedly increases the difficulty of color segmentation. (4) There are many kinds of elements in a geological map, and they overlap in space to a certain extent, such as the interference of line elements such as water systems and kilometer grids on the geological map, and the overlay of characters and various symbols, which creates great difficulties in the recognition and extraction of line elements in the map. Moreover, the intersection of various elements may leave many breakpoints on the single-page map obtained after color segmentation, which increase the duration and difficulty of line tracing.

In response to the aforementioned challenges, this study proposes a color geological image segmentation model that combines the Felz clustering algorithm and an improved SE-UNet network (named GeoMSeg). First, mathematical morphology operations, encompassing techniques like erosion and dilation are employed to pre-process the geological images. The UNet with an encoder-decoder structure is used as the foundational model, and a channel attention mechanism called SENet is added to focus on key image features while suppressing unnecessary features, thereby reducing the influence of complex backgrounds in the image. Second, the Felz clustering algorithm is combined with the pre-segmentation of the geological image, allowing the semantic segmentation results of deep learning to refine and modify the extracted feature information based on the pre-segmentation. The combination of dilation and erosion in mathematical morphology operations can effectively eliminate noise in geological images, and the Felz clustering algorithm can adaptively adjust the threshold based on the local features in the image data to perform pre-segmentation processing on the geological images. Finally, deep learning is used to fit regions with the same semantic information based on the Felz clustering algorithm's pre-segmentation, such that the segmentation results are as consistent as possible with the pre-segmentation results. Experimental results indicate that the algorithm put forth in this study attains notable performance levels on a color geological image dataset.

## 2. Related work

Image segmentation constitutes a crucial component of image analysis, which aims to segment images into different regions based on homogeneous features, including pixel intensities, texture structures, etc. (Ma BJ et al., 2023). Currently, semantic segmentation can be divided into two categories (Zhang WH et al., 2019): Traditional segmentation methods, and the machine learning-based (especially deep learning) methods.

### *2.1. Traditional image segmentation methods*

Traditional segmentation methods employ low-level semantic features like color, shape, and texture to automatically segment the target area, resulting in regions with similar features and differences between regions. This approach categorically encompasses four main classifications: Threshold-based segmentation methods, edge-based segmentation methods, region-based segmentation methods, and segmentation methods combining specific theories.

Threshold-based segmentation method obtains one or more gray thresholds by utilizing the grayscale features of the image, and then compares these thresholds with the gray values of each pixel in the image to classify each pixel based on the comparison result. Common threshold segmentation methods include the maximum inter-class variance method (Otsu) (Otsu N, 1979), the minimum error method (Kittler J and Illingworth J, 1986), and the maximum entropy method (Kapur JN et al., 1985). For example, Levachkine S et al. (2002) used pseudo-color technology to construct a two-dimensional histogram of the image and determined the optimal threshold using a method before and after threshold processing to achieve segmentation of color images. Huang MX et al. (2012) proposed an improved threshold image segmentation algorithm based on Otsu's method, which selects the optimal threshold by narrowing the threshold selection range and searching for the minimum variance ratio. Hosseini-Fard E et al. (2022) proposed a texture attribute extraction strategy based on the concept of direction gradient histogram (HOG) for automatic segmentation and attribute extraction of seismic images. This method can separate diverse shape, distinct seismic patterns, and geological objects exhibiting contrast with the surrounding medium. Threshold-based segmentation methods are simple to implement and

have low computational cost, but they are susceptible to noise and require significant grayscale value differences between the target and surrounding environment, making it difficult to accurately segment complex images.

Edge-based segmentation detects the boundaries between different regions in an image by utilizing the discontinuity of pixel grayscale values between the regions, achieving image segmentation. The key is to perform edge detection on the image. Commonly used edge detection operators include the Robert operator (Rosenfeld A, 1981), Sobel operator (Lang Y and Zheng D, 2016), Prewitt operator (Yang L et al., 2011), LoG operator (Ulupinar F and Medioni G, 1990), and Canny operator (Li ES et al., 2009), among others. For example, Xu JL et al. (2020) proposed an automatic extraction method for linear features based on wavelet edge detection. To reduce tunnel construction accidents, Jiang F et al. (2022) used a threshold-adaptive segmentation algorithm combining the Canny operator to segment and detect tunnel rock fissures from tunnel face images. The experimental results demonstrate that the extraction error rate of this algorithm is less than 2%, and it can accurately extract complete rock fissures.

Region-based segmentation method is a process of continuously merging pixels or sub-regions according to certain rules to form segmentation regions. There are many methods for segmentation based on the region concept (Pham DL et al., 2000; Zhang YJ, 2006), which can be usually divided into three categories: Watershed segmentation (Vincent L and Soille P, 1991), region growing (Adams R and Bischof L, 1994), and region splitting and merging (Tremeau A and Borel N, 1997). Watershed segmentation is usually performed on gradient images and requires the use of a marker image to achieve better segmentation results. For example, Leyk S and Boesch R, (2010) used a constrained seed region growing method based on color similarity and spatial connectivity to segment 19$^{th}$-century archival topographic maps. This method was assessed across map pages exhibiting diverse graphical attributes, and obtained reliable results from accuracy evaluation. Ji XQ et al. (2015) proposed a distance transformation-based watershed algorithm to segment cell adhesion images, which effectively solves problems such as cell proliferation and discontinuous cell boundaries. The algorithm uses Otsu threshold segmentation for coarse segmentation and then uses an optimized watershed algorithm for fine segmentation by optimizing seed points. The region-based segmentation method is less sensitive to noise and has strong noise resistance, but it often suffers from over-segmentation problems. On the other hand, edge-based segmentation methods are characterized by straightforward implementation and rapid processing speeds but are easily affected by noise. Therefore, these two segmentation methods are often combined to segment image content.

The segmentation method based on specific theory can be categorized into two categories: Clustering-based segmentation and graph-based segmentation. Clustering-based segmentation methods cluster pixels with similarities in the image into the same region or image block, and iteratively modify the clustering results to form the segmentation result. Main algorithms include MeanShift algorithm (Cheng YZ, 1995), TurboPixels algorithm (Levinshtein A et al., 2009), and Slic algorithm (Achanta R et al., 2012). For example, Huang CL et al. (2016) proposed a geological segmentation algorithm for different geological features such as rivers, forests, and soil. The experimental results show that this method is effective and reliable in geological segmentation. Graph-based segmentation methods map the entire image into a weighted undirected graph, transforming the image problem into a graph partitioning problem, and the segmentation result of the image is subsequently derived through the optimization of the objective function. For example, Felzenszwalb PF et al. (2004) proposed an algorithm based on the minimum spanning tree of the image, clustering nodes of the graph to achieve segmentation. The generated super pixels are the minimum spanning trees of the pixel set, which can maintain the image boundary well. Van den Bergh M et al. (2015) proposed the energy-driven sampling (SEEDS) algorithm, which uses hill-climbing to optimize the objective function iteratively by searching for the minimum local change to update the solution and continuously refine the boundary to obtain the optimal segmentation effect. Graph-based segmentation methods have strong noise resistance and fast computing power, but their biggest disadvantage is that they can only perform binary segmentation of pixel points in the image, and it is difficult to achieve end-to-end multi-class segmentation of objects in the image.

### *2.2. Machine learning-based image segmentation methods*

There has been a progressive application of deep learning techniques to the domain of geological image segmentation. Compared with traditional image segmentation methods, deep learning-based image segmentation methods (Hinton GE et al., 2006) are less affected by human factors and can obtain abstract image features, mine the internal structures of different features, and have made significant improvements in segmentation performance and efficiency. Deep learning-based image semantic segmentation methods mainly include four types: Segmentation methods based on FCN, segmentation methods based on encoder-decoder, segmentation methods based on dilated convolution, and segmentation methods combined with attention mechanism.

Segmentation methods based on FCN. Fully Convolutional Network (FCN) is the most used network for semantic segmentation and is a pioneering work of deep learning in image semantic segmentation tasks (Long J et al., 2015). Currently, the FCN model has garnered extensive utilization across various computer vision tasks and has shown excellent performance in object detection, classification, image segmentation, and other tasks. For example, Teimouri N et al. (2019) developed and implemented a novel, optimal, and lightweight network

structure based on FCN and ConvLSTM network, using multi-temporal radar images to recognize various crop types. Ji J et al. (2020) proposed a parallel FCN that mainly integrates the Holistically-Nested Edge Detection (HED) network to capture image edge information and improve the segmentation performance of semantic segmentation. Zheng SX et al. (2021) proposed the SETR model based on the FCN framework, mainly treating semantic segmentation as a sequence-to-sequence prediction task.

Segmentation methods based on encoder-decoder. The encoder-decoder structure consists of two fundamental components: Encoder and decoder. The encoding component predominantly comprises a series of stacked convolutional layers, pooling layers, etc., to acquire higher-level semantic information. The decoding part includes convolutional layers and upsampling operations to gradually restore spatial details until the segmentation process is completed. Some representative models include UNet (Ronneberger O et al., 2015), ENet (Paszke A et al., 2016), SegNet (Badrinarayanan V et al., 2017), etc. For example, (Zhang Y et al., 2022) proposed an intelligent image segmentation model RUNet based on residual network ResNet and UNet for rock fracture detection, providing an automated and reliable method for rock quality evaluation. Liu HQ et al. (2023) proposed a U-shaped transformer framework RockFormer for Martian rock segmentation. The framework encompasses a layered encoder-decoder architecture and feature refinement modules connected between them.

Segmentation methods based on dilated convolution. Dilated convolution (Yu F and Koltun V, 2015) essentially injects "zeros" into a standard convolution kernel to expand the receptive field with different sampling rates, allowing for the extraction of multiscale feature information over a larger range without increasing the number of convolution kernel parameters, while maintaining the resolution of the feature map. This effectively utilizes contextual information to solve the problem of spatial information loss in semantic segmentation. For example, Wei YC et al. (2018) used dilated convolution to obtain more reliable target areas, effectively improving the performance of weakly supervised and semi-supervised image semantic segmentation. Chen LC et al. (2018) designed a porous spatial pyramid pooling scheme to achieve robust segmentation of multiscale targets, distributing multiple dilated convolutions in parallel to capture multiscale object and context information. Chen WK et al. (2022) proposed a three-dimensional multithreaded dilated convolution (MTDC) network for automatic segmentation, which can better extract and integrate low-dimensional structural features by combining the MTDC strategy.

Segmentation methods combined with attention mechanism. The fundamental concept underlying the attention mechanism is to generate different weight parameters for objects of varying scales within the image, tailored to the requirements of the segmentation task, selectively emphasizing target objects relevant to the task, and suppressing those irrelevant ones. The key is to generate weight maps with appropriate attention weights, which can capture richer contextual information. For example, Lin GS et al. (2017) proposed a multi-path learning network that uses attention modules with long-range residual connections to fuse low-level and high-level features for high-resolution semantic segmentation models. Wang F et al. (2017) stacked attention modules with residual connections to capture the response results of feature maps at different levels. Ding L et al. (2020) proposed a method that combines patch-based attention modules and attention embedding modules to enhance the embedding of contextual information and enrich the semantic information of low-level features. Zhou TX et al. (2022) proposed a multi-modal segmentation network guided by triple attention fusion, which uses dual attention fusion modules and correlation attention modules to form triple attention fusion blocks, and learns the correlation between modalities and the potential features relevant to segmentation to improve segmentation results.

Within the domain of image segmentation, a large amount of research has been devoted to extracting information from unstructured images. Although these efforts have achieved certain achievements, existing image segmentation methods still have limitations in geological image scenes. Firstly, when extracting information in geological image scenes, there is a dearth of image segmentation methodologies capable of diminishing manual intervention and attaining high-performance results. On one hand, most of existing image segmentation methods use rule-based traditional segmentation methods or supervised machine learning methods. Rule-based and supervised machine learning-based image segmentation methods can solve the heterogeneity and complexity of images by learning many representative examples, thereby achieving high-performance image semantic segmentation. However, traditional rule-based segmentation methods, are difficult to adapt to the segmentation requirements of geological image scenes that have overlapping features and are complex and diverse. Specific problems mainly include the following two points: (1) Segmentation methods and improved segmentation algorithms have increased the limitations of the algorithm on usage conditions or segmentation rules while improving segmentation accuracy or efficiency, such as threshold-based segmentation methods requiring different probability distributions for background areas and target areas, and region-based segmentation methods requiring similarity rules, merging rules, and splitting rules to be set. (2) Traditional segmentation methods often require human intervention and mainly use manually annotated feature rules for segmentation, using visual surface features like texture and color within the image and external structural characteristics for processing. The performance and robustness of the features are greatly limited, and the segmentation accuracy is easily affected. Supervised machine learning methods are not suitable for geological image segmentation scenes, as geological images contain element symbols such as text annotations, numeric annotations, point symbols, and line symbols that are intertwined and complex, and the characteristics of each element formed in different locations and times. Therefore, it is impossible to uniformly label all elements in geological images, making supervised

machine learning methods ineffective for geological image segmentation. On the other hand, while image segmentation methods founded on semi-supervised and unsupervised machine learning offer practicable solutions to address labor-intensive challenges, existing semi-supervised and unsupervised machine learning-based image segmentation methods still have certain limitations in extracting information from geological images characterized by high levels of heterogeneity and complexity. Therefore, further research is still needed on how to improve or combine other theories or algorithms to make the segmentation results more accurate and robust for the problem of complex and diverse features in geological image segmentation.

## 3. Methodology

Within this study, a method that integrates the combination of the UNet model and Felz clustering algorithm is presented for the segmentation of geological maps (Fig. 1). To reduce background noise interference within geological maps and enhance the feature extraction proficiency of the segmentation model. Firstly, pre-processing operations such as mathematical morphology are adopted before pre-segmentation of geological maps to reduce the impact of noise and other factors in maps on segmentation performance. Secondly, the channel attention mechanism SENet is added to the UNet network to enhance the attention to target information and suppress unnecessary feature information to improve the feature expression ability of the model. In terms of model segmentation metrics, commonly used indicators such as the pixel accuracy (PA), mean intersection over union (MIoU) (Rahman MA and Wang Y, 2016), and Dice coefficient (Shamir RR et al., 2019) in pixel level segmentation are selected to evaluate the segmentation effect, in order to more objectively and directly reflect the segmentation accuracy of the model.

### *3.1. Geological image dataset construction*

Due to the absence of publicly available colored image datasets in the domain of geoscience, the colored geological image dataset (named Geo_Dataset) was obtained from various websites, such as the U.S. Geological Survey (https://www.usgs.gov/), Geo cloud (https://geocloud.cgs.gov.cn/), National Geological Archives Data Center (http://dc.ngac.org.cn/Home), and the National Geological Data Museum (http://www.ngac.org.cn/), as well as large amount of regional geological reports, this study mainly selected color planar

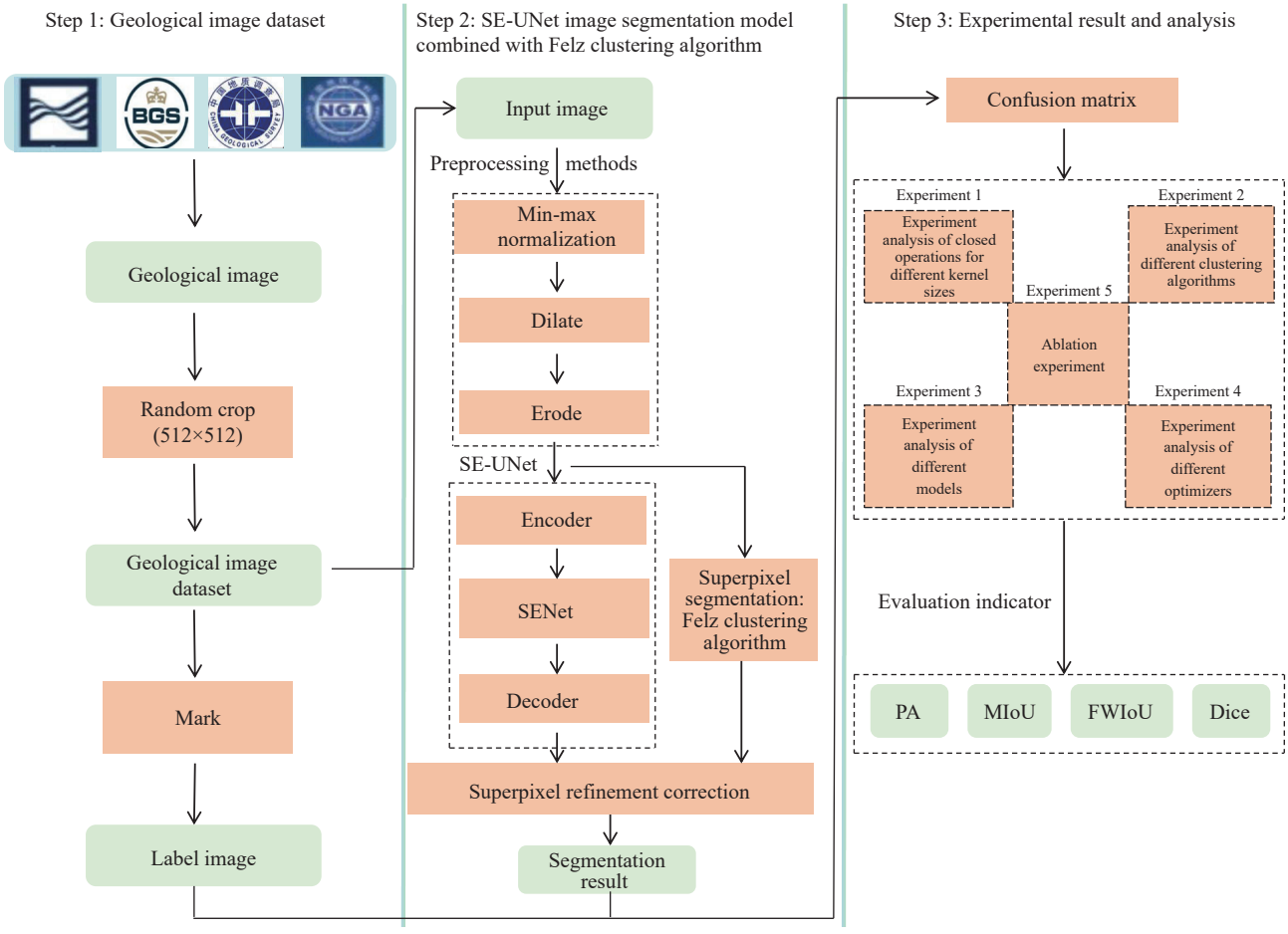


**Fig. 1.** Overall architecture of unsupervised color geological image segmentation model. Unsupervised color geological image segmentation model can be divided into three steps: Geological image data set construction, segmentation model combined with clustering algorithm and improved SE-UNet network, and image segmentation result analysis.

geological image data with scales of 1∶50000, 1∶200000, and 1∶250000, with an image size of 512 × 512, in RGB format. The geological images in this dataset cover the geological features of several countries, including the Kaiping sheet geological map and Xingxing Gorge Geological map of China, and Alaska geological map of the United States, and more. The constructed color geological atlas contains 200 images in total. The partial sample images of the Geo_dataset are shown in Fig. 2. An unsupervised image segmentation method that does not require training data or labeled data is used in this paper, where the training process is also the prediction process, eliminating the need to partition the dataset into training and test sets.

As the collected geological images have different resolutions and sizes, the image size is uniformly adjusted to 512 × 512 through cropping to adapt to the model's input size. To quantitatively evaluate the segmentation results of the model, the collected data is manually labeled using the Labelme annotation software, generating a corresponding JSON file for each original image. Python is then used to process the original image and JSON file, generating a mask image with the same name as the original image.

### *3.2. SE-UNet image segmentation model combined with Felz clustering algorithm*

#### *3.2.1. Image preprocessing*

Image normalization. In this study, the maximum minimum normalization method was used to normalize the input image, and Formula 1 is as follows:

$$norm = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{1}$$

Where $x_i$ denotes the image pixel value, $\min(x)$, $\max(x)$ denotes the minimum and maximum value of the image pixel respectively.

By converting the value range of image data to between [0, 1] through the above normalization formula, effective enhancements in both the convergence speed and accuracy of the model can be achieved, and the problem of gradient explosion or gradient disappearance caused by too large or too small image pixel values can be avoided in the optimization process.

Mathematical morphology. Mathematical Morphology was proposed by Serra and Matheron in France in 1964 (Angulo J and Serra J, 2007) to obtain topological and structural information of objects by exploiting the interaction between objects and structural elements. In this study, closed operations are used to process the input images mathematically morphologically. The closed operation improves the quality and morphology of the geological image by combining dilation (Fig. 3) and erosion (Fig. 4) operations. The computational formulas for dilation and erosion, respectively, and are defined as follows:

$$A \oplus B = \{z | [(B^*)_z \cap A] \subseteq A\} \tag{2}$$

$$A \ominus B = \{z | (B)_z \subseteq A\} \tag{3}$$

Where $A$ represents the set of foreground pixels, $B$ and $B^*$ represent structuring elements, and $z$ denotes the foreground pixel value. $A \oplus B$ denotes the dilation of $B$ with respect to $A$, which is the collection of all displacements $z$ for which the foreground elements of $B^*$ overlap with at least one element of $A$. $A \ominus B$ denotes the erosion of $B$ with respect to $A$, which is the collection of all points $z$ for which the translated $B$ is contained within $A$. By using Formula 2 and Formula 3, the Formula 4 for the closed operation can be derived, defined as follows:

$$A \cdot B = (A \oplus B) \ominus B \tag{4}$$

Closed operation can smooth the boundaries of geological images, enhance their continuity and integrity. Moreover, by

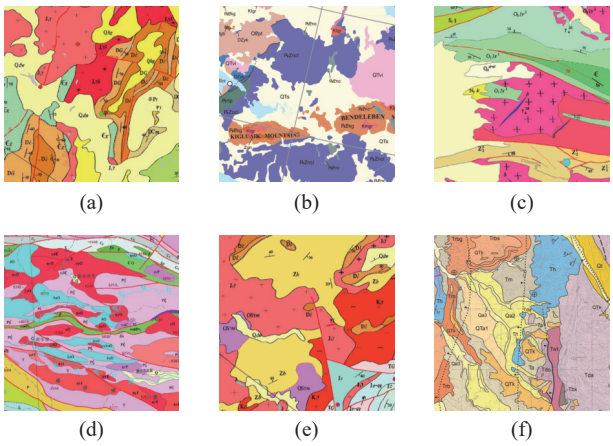


**Fig. 2.** Partial sample graph. a–China, Kaiping sheet; b–United States, Alaska; c–China, Xingxing Gorge sheet; d–China, Hindu Kush-West Kunlun; e–China, Zijin sheet; f–United States, The northern White Hills, Mohave County, Arizona.

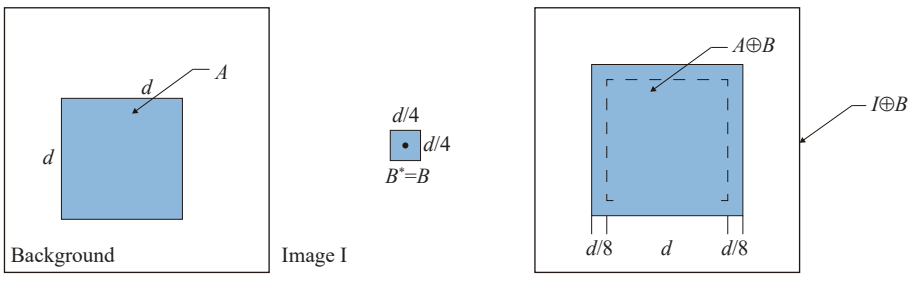


**Fig. 3.** Dilation diagram.

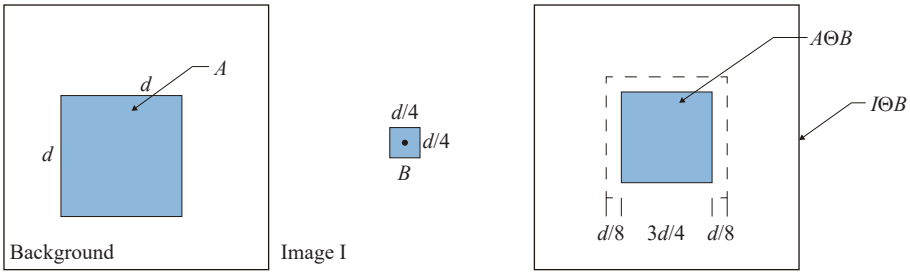


**Fig. 4.** Erosion diagram.

combining erosion operations, it can effectively eliminate the influence of characters in geological images on element segmentation, thereby improving the quality of segmentation results.

*3.2.2. Clustering algorithm: Felz*

Before introducing the basic idea and calculation steps of the Felz (Felzenszwalb PF and Huttenlocher DP, 2004) clustering algorithm, the related definition based on graph representation is as follows:

Definition 1: $G=(V,E)$ denotes an undirected graph with $n$ vertices and $m$ edges, consisting of vertex set $V$ and edge set $E$. $V$ is the set of vertices to be partitioned, which is a single pixel point in the image, where $v_i \in V$ and $(v_i,v_j) \in E$ denotes the edges connected between adjacent vertices $(v_i,v_j)$.

Definition 2: Each connected edge $(v_i,v_j) \in E$ in a graph $G$ has a weight $\omega=(v_i,v_j)$, which indicates a nonnegative measure of dissimilarity between vertices (such as brightness, color, motion, position, or other properties of itself), meaning that the larger the weight, the more dissimilar the corresponding vertex.

Definition 3: Segment the graph $G$ to obtain $S=(C_1,\cdots,C_r)$, where $C_i$represents the mutually disjoint regions after segmentation, where $C_i \subseteq V(1 \leqslant i \leqslant r)$.

The Felz clustering algorithm is a graph-based greedy clustering algorithm, as shown in Fig. 5. The primary objective of image segmentation is to segment the image into several specific regions characterized by unique properties, and then extract the target of interest from these areas. The delineation of boundaries between image regions constitutes the key of the image segmentation algorithm. The algorithm provides the judgment standard for the boundary definition between image regions based on graph-based representation. The basic idea is to merge regions by the judgment standard of the distance difference between regions and the distance difference within regions to adaptively adjust the threshold according to the local features of image data and use the greedy decision algorithm to segment images.

The judgment criteria for the definition of the boundary between image regions of the Felz clustering algorithm are the interregion spacing. The intraregion spacing is the maximum weight in the minimum spanning tree (MST) in the segmented region$C_i(1 \leqslant i \leqslant r)$, which corresponds the weight of an edge with the largest dissimilarity in the corresponding region, expressed as$Int(C)$. For any$C_i \subseteq V(1 \leqslant i \leqslant r)$, Formula 5 of $Int(C)$ is denoted as:

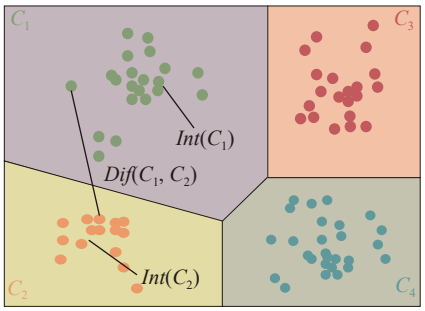


**Fig. 5.** Felz clustering algorithm diagram.

$$Int(C)=\max_{e\in MST(C,E)}\omega(e) \tag{5}$$

That is, for a given region $C$ will remain connected only when the weight is at least $Int(C)$.

The interregion spacing is the minimum weight value among the pairs of points belonging to two regions and connected by edges to each other, that is, the weight of the edge with the minimum dissimilarity between two regions, denoted by $Dif(C_1,C_2)$, that is, for any $v_i,v_j \in V(1 \leqslant i,j \leqslant r, i \neq j)$, Formula 6 of $Dif(C_1,C_2)$ is denoted as:

$$Dif(C_1,C_2)=\min_{v_i\in C_1,v_j\in C_2,(v_i,v_j)\in E}\omega(v_i,v_j) \tag{6}$$

When there is no connected edge between regions $C_1$ and $C_2$, $Dif(C_1,C_2)=+\infty$.

By comparing the differences between intra-area spacing and inter-area spacing, it is possible to evaluate whether there is a boundary between different areas. The boundary judgment function is represented by$D(C_1,C_2)$, and defined as Formula 7:

$$D(C_1,C_2)=\begin{cases}True, Dif(C_1,C_2)>MInt(C_1,C_2)\\ False, Dif(C_1,C_2)\leqslant MInt(C_1,C_2)\end{cases} \tag{7}$$

where the minimum internal variance $MInt(C_1,C_2)$ formula is defined as:

$$MInt(C_1,C_2)=\min(Int(C_1)+\tau(C_1),Int(C_2)+\tau(C_2)) \tag{8}$$

In Formula 8, $\tau$ serves as a threshold function to control the extent to which the difference between two regions must be greater than the intraregional disparity to prove the existence of a boundary between them, that is, the boundary function $D(C_1,C_2)$ is true. For small regions, $Int(C)$does not reflect the intraregional spacing well, because in the extreme case, when the region $C$ contains only one vertex, that is, the

region is an isolated pixel point, $Int(C) = 0$, which will lead to oversegmentation in the absence of restriction, so a threshold function based on the region size is added, defined as Formula 9:

$$\tau(C) = \frac{k}{|C|} \tag{9}$$

where $|C|$ denotes the count of pixel points it encompasses and $k$ is a fixed parameter that sets the observation range and is used to control the size of the formed region. When $k = 0$, each pixel point in the graph is an independent region, while when $k = +\infty$, the whole graph will become a region, so the image can be partitioned into regions of a specific shape by adjusting the threshold function $\tau$. The Felz clustering algorithm is as follows:

In terms of the time efficiency, the algorithm is basically linear correlation with the quantity of edges corresponding to the graph representation of the image, and the edges of the graph representation are proportional to the pixel points, which means that the time efficiency of image segmentation is linear with the number of pixel points in the image, and the actual operation is fast. Another advantage is that it can keep the details of low-variation regions while ignoring the details of high-variation regions, which is conducive for finding visually consistent regions and has a good segmentation effect on images.

**Algorithm 1:** Felz clustering algorithm

**Input:** a graph $G = (V,E)$ with $n$ vertices and $m$ edges

**Output:** a segmentation of $V$ into components $S = (C_1,\cdots,C_r)$

**1: //clustering phase:**

**2: for** each edge $o \in E$ **do**

**3:** $\varphi = (e_1,\cdots,e_m)$←sort by weight in descending order of dissimilarity of each edge

**4: end for**

**5:** $S_0 \leftarrow v_i \in V(1 \leqslant i \leqslant r)$

**6:** $v_i, v_j$ ←represents the vertices connected by the qth edge

**7: for** $q = 1$ to $m$ **do**

**8:** $O_q = (v_i, v_j)\ \ v_i \in C_i^{q-1}\ \ v_j \in C_j^{q-1}$

**9: if** $C_i^{q-1} \neq C_j^{q-1}$ && $D(O_q) \leqslant MInt(C_i^{q-1}, C_j^{q-1})$

**10:** $S_q$ ←merge $C_i^{q-1}$ and $C_j^{q-1}$

**11: else** $S_q = S_{q-1}$

**12: end for**

**13: return** $S = S_m$

### 3.2.3. Attention modules: SENet

SENet (Hu J et al., 2018) model structure consists of three parts: Squeeze, excitation and reweight, and the core idea is that the network can learn the importance of feature channels autonomously according to the loss, and increase the weight of more effective feature information and decrease the weight of less effective feature information according to the importance of feature channels, so that the model can achieve better results. The structure of SENet is shown in Fig. 6.

Squeeze operation. A feature map of size C × W × H is obtained by compressing and encoding the spatial information on each channel into global features using global average pooling to acquire a feature map of size 1 × 1 × C. The mapping relationship is shown in Formula 10:

$$z_c = F_{sq}(u_c) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j) \tag{10}$$

where $C$, $H$, $W$ are feature map dimension dimensions, $u_c$ is the feature channel, $F_{sq}$ is the Squeeze operation definition, $i$, $j$ are pixel location variables.

Excitation operation. The global features obtained from the Squeeze operation are passed into the two fully connected layers, and the weights of the different feature channels and channel correlations are represented by the parameter $W$. The mapping relationship is shown in Formula 11:

$$s = F_{ex}(z,W) = \sigma[g(z,W)] = \sigma[W_2\delta(W_1,z)] \tag{11}$$

where $F_{ex}$ is the Excitation operation definition, $z$ is the Squeeze operation output; $W_1$ and $W_2$ are the channel weights, $\delta$ is the activation function, $\sigma$ is the normalization function.

Reweight operation. The weight s obtained from the Excitation operation is weighted onto the original input features to acquire the output and used as input for the next level. The mapping relationship is shown in Formula 12:

$$x_c = F_{scale}(u_c, s) = s_c u_c \tag{12}$$

where $F_{scale}$ is the definition of Reweight operation, $s_c$ is the output matrix channel of Excitation operation.

### 3.2.4. UNet network

UNet is a deep learning structure for image segmentation tasks, consisting of a symmetric encoder and decoder, and the network structure is shown in Fig. 7. The primary function of the encoder lies in the extraction of feature information from the input image and abstracting it for fine segmentation in the decoder. The encoder part has four down sampling modules, each of which consists of a convolutional layer, a ReLU (Glorot X et al., 2011) correction linear unit and a pooling layer. The convolution layer serves as the primary instrument for extracting features from the input image, and the convolution kernel applies sliding windows to different regions of the image to extract the local feature matrix of the image, defined as Formula 13:

$$z_{i,j} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \omega_{m,n} x_{(i+m),(j+n)} + b \tag{13}$$

where $x$ denotes the input image, $\omega$ denotes the convolution kernel, $b$ denotes the bias term, $M$, $N$ represent the size of the convolution kernel, $z$ represents the output feature map, and $(i, j)$ denotes the location index of the output tensor. The correction unit ReLU is an activation function that performs a nonlinear mapping of the features extracted from the convolutional layer to make the special features easier to distinguish, and its calculation is shown in Formula 14:

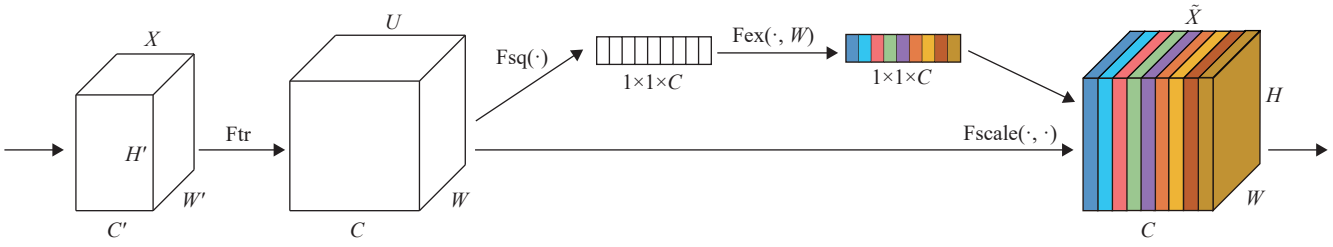


**Fig. 6.** Felz clustering algorithm diagram.

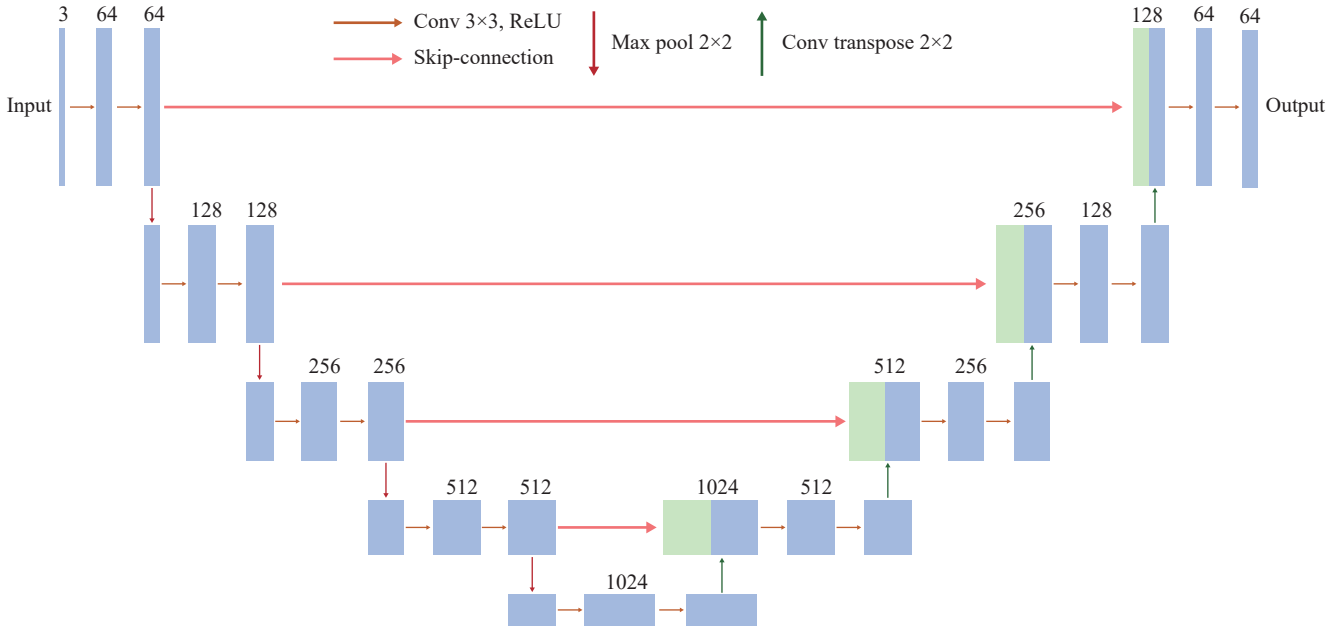


**Fig. 7.** UNet network structure.

$$y = \max(0, x) \tag{14}$$

where $x$ denotes the output feature map of the convolutional layer. the ReLU function sets the input less than 0 to 0 and the input greater than or equal to 0 remains unchanged. this nonlinear transformation contributes to augmenting the network's capability for nonlinear expression, solve the gradient disappearance problem in model training, and accelerate the convergence speed. The maximum pooling layer compresses the size of the feature map's size through down-sampling, chunking the input feature map according to the specified pooling size and retaining only the strongest features in the local area, which can effectively reduce the dimensionality and size of the feature map, reduce the computational effort and prevent overfitting, as shown in Formula 15:

$$c_{i,j} = \max_{m=0}^{M-1} \max_{n=0}^{N-1} x_{(i+m),(j+n)} \tag{15}$$

where $x$ represents the input feature map, $c$ represents the output tensor, and $(i, j)$represents the positional index within the output tensor.

The decoder uses a skip connection to connect each layer in the encoder to the corresponding layer in the encoder to pass the multi-scale feature information from the encoder to the decoder, and then restores the spatial resolution of the feature map to the original size through the upsampling layer to output a segmentation result of the same size as the input. The network structure of the decoder part is similar to the symmetric structure of the encoder, which is also divided into four layers, and each upsampling module is composed of a convolutional layer, a ReLU-corrected linear unit and a transposed convolution. The transposed convolution recovers the position information of the image through the upsampled feature map, while combining it with the corresponding low-level features of the encoder, aiming to leverage multi-scale feature information and enhance segmentation accuracy. The calculation method is shown in Formula 16:

$$\mu(x,y) = \sum_{i=0}^{k-1} \sum_{j=0}^{l-1} \upsilon(x+i, y+j) \cdot \omega(i,j) + p \tag{16}$$

where $\omega$ is the transposed convolution kernel, $\upsilon$ denotes the input feature map, $p$ represents the bias term, $k$, $l$ represents the size of the transposed convolution kernel in the horizontal and vertical directions, and $x$, $y$ represents the position index of the output feature map.

Skip connection is the key to the structure of UNet network. The skip connection splices the feature maps in the encoder with those in the corresponding decoder by channel to build a richer feature representation, and then downscales them by convolution operation to obtain a more compact feature representation as the input to the decoder. The mutual

amalgamation of low-level and high-level features allows the model to access rich multi-scale information, thus restoring the feature information of the original image more accurately.

*3.2.5. SE-UNet model for color geological image segmentation*

The overall structure of the SE-UNet model proposed in this paper is shown in Fig. 8. The model is composed of a U-shaped convolutional network and an attention mechanism module SENet. The U-shaped convolutional network is primarily employed for feature extraction, and it contains an encoder and a decoder. The encoder comprises two modules consisting of 3×3 convolutions repeatedly, each convolutional layer is followed by a batch normalization and ReLU operation. Each module is followed by a 2×2 max-pooling layer for downsampling, where the size of the feature map is doubled for each downsampling and the number of channels is doubled. The decoder follows each 3×3 convolution module with a 2×2 deconvolution layer for upsampling, doubling the size of each upsampled feature map and doubling the number of channels. The decoder of U-shaped convolutional network is structured upon the architecture of VGGNet. In comparison to the original UNet network, the SE-UNet, as proposed, removes the last downsampling layer and the upsampling layer, because the multiple downsampling is easy to the gradient disappearance will occur if there are too many convolutional layers.

The attention mechanism module SENet is mainly used to process the attention mechanism for the features of the high and low spatial network layers. Since the semantic information extracted from different layers of the U-shaped convolutional network varies greatly, the feature maps obtained by fusion of jump connections are not conducive to the learning of the network. Consequently, the integration of the attention module to the skip connection can highlight the salient features through the jump connection, allowing the network to learn the task-relevant feature information and disregard the interference of irrelevant feature information. The attention mechanism enables the network to acquire more useful information, heightens the sensitivity of the network to the target, and obtain better segmentation accuracy.

## 4. Experimental results and analysis

### *4.1. Experimental parameter settings*

To evaluate the proposed algorithm objectively and fairly, all experiments were conducted under the same experimental environment. The experiments were performed on a Windows operating system with an NVIDIA GTX 3060 GPU and CUDA version 12.0.134. PyTorch deep learning framework was used, and the programming language was Python.

The input size of the network images was set to 512 × 512. The optimizer was chosen as Stochastic Gradient Descent (SGD) (Tian Z et al., 2016), with 64 epochs of model iterations, a batch size of 1, an initial learning rate of 0.05, and momentum set to 0.9, which effectively reduced oscillation caused by gradient descent direction changes. Cross-entropy (Chockler H et al., 2007) was used as the loss function.

### *4.2. Evaluation metrics*

The predicted images and ground-truth images were used to calculate the confusion matrix, which is also known as the error matrix and is a standard format for accuracy evaluation. The confusion matrix is represented in a matrix form with *n* rows and *n* columns.

When verifying the effectiveness of image segmentation results, existing studies often used the pixel count of the target region to evaluate the segmentation effect (Sezgin M and Sankur B, 2004). In this experiment, the confusion matrix was used to reflect the segmentation of each area in the geological image by the algorithm; to better reflect the performance of the algorithm and compare the geological images based on different models or clustering algorithms more comprehensively and intuitively, PA, MIoU, frequency-weighted intersection over union (FWIoU) and Dice

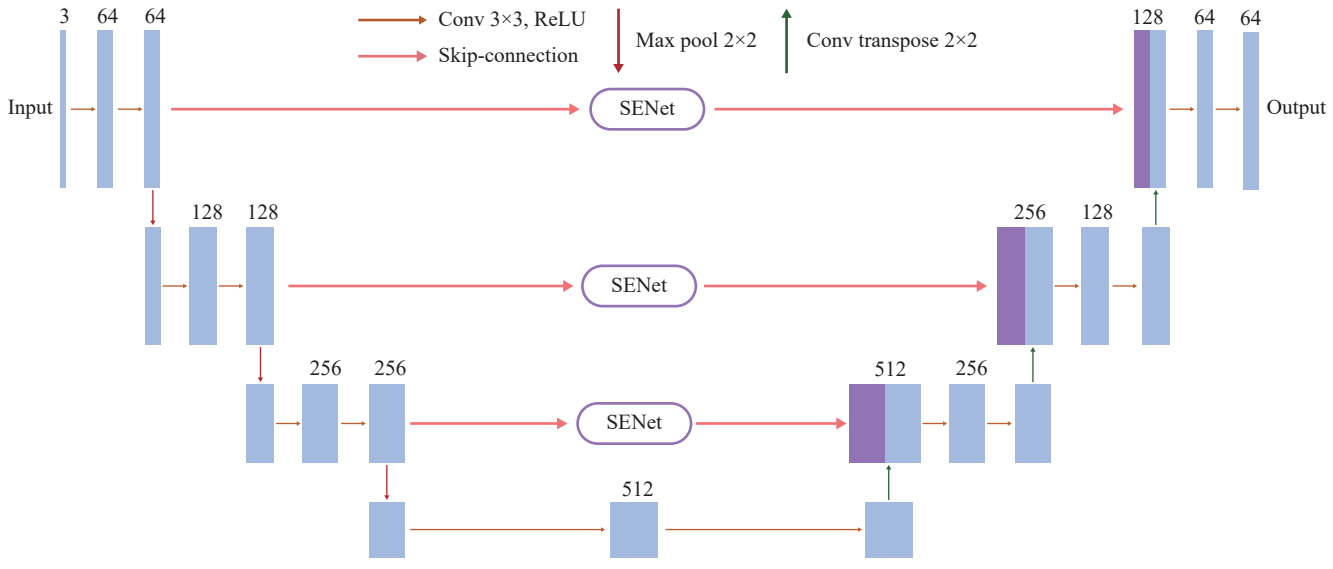


**Fig. 8.** SE-UNet network structure.

coefficient were used to evaluate the performance difference of the color segmentation algorithm, where PA signifies the ratio of pixel correctly identified to the total pixels, the larger the PA, the better the recognition effect of various geological elements in the image. MIoU is the standard metric of semantic segmentation, which computes the ratio of the intersection of the predicted results of the categories and the true value of the segmentation algorithm to the merged set and then to the total number of its categories. A higher MIoU value indicates a higher degree of overlap between the predicted and ground-truth segmentation maps. When the MIoU is greater than 0.5, the segmentation model is considered to have good recognition ability. FWIoU is the lifting formula of the MIoU, which can be calculated by setting the corresponding weight according to the frequency of each category; the Dice coefficient is used to calculate the similarity of two samples and takes the value of [0,1], and can be used to calculate the similarity between the ground-truth and segmentation maps, the closer the Dice value is to 0, the less similar the two contour areas are, and the closer it is to 1, the more similar the two contour areas are. The calculation formula for the four evaluation indicators is as follows (Formula 17–20):

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

$$MIoU = \frac{1}{k+1} \cdot \frac{TP}{TP + FP + FN} \tag{18}$$

$$FWIoU = \frac{1}{\sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij}} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{19}$$

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{20}$$

In the above formula, it is assumed that there are $k+1$ classes, $i$ denotes the true value, $j$ denotes the predicted value, and $p_{ij}$ denotes the number of pixels that predict class $i$ as class $j$. $TP$ denotes positive samples whose model predicts the positive class; $FP$ denotes negative samples whose model predicts the positive class; $FN$ denotes positive samples whose model predicts the negative class; $TN$ denotes negative samples whose model predicts negative class. $TP$ and $TN$ denote the total number of pixels that correctly segment the geological image area of the corresponding category, while $FP$ and $FN$ denote the total number of pixels that do not correctly segment the geological image area of the corresponding category. Therefore, $p_{ii}$indicates that the number of pixels in the correctly segmented category of the model is equivalent to $TP$, $p_{ij}$ is equivalent to $FN$, $p_{ji}$ is equivalent to $FP$, and $p_{jj}$ is equivalent to $TN$.

### 4.3. Experimental results

#### 4.3.1. Experimental analysis of closed operations for different kernel sizes

In unsupervised segmentation of color geological maps, due to the complex structure of geological maps and the inhomogeneity of color distribution, direct segmentation is prone to incorrect segmentation, so this study introduces the closed operation in mathematical morphology in the preprocessing stage of geological maps, that is, the maps is first subjected to the expansion operation and then the erosion operation. Geomorphology is used as a tool to extract useful image components from the image that express and depict the shape of the region, such as boundaries, skeletons, etc. (Gonzalez RC and Woods RE, 2002). In layman's terms, the expansion and erosion operations are to convolve the image with the corresponding kernel. By using the erosion and expansion operations in mathematical morphology, the fine noise in the image can be removed, while the connectivity of the region can be preserved, so that the structural and edge information of the image can be better preserved, providing a more accurate information basis for subsequent unsupervised segmentation. In addition, the processing process is simple and the computational effort is small, which is feasible for the image segmentation task in practical applications and assumes a pivotal role in enhancing the segmentation accuracy and preserving the image structure information.

In order to verify the effect of the closed operation with different convolutional kernel sizes on the segmentation performance, the geological image segmentation was performed using convolutional kernel sizes of 2, 3, 4 and 5, respectively, and the experimental results were compared as shown in Table 1. Segmentation performance for closed operation with different sizes of convolutional kernels. The experimental results show that the best segmentation results were obtained with the convolutional kernel size of 5, with PA values up to 91.89%, MIoU values up to 71.91%, and Dice values up to 77.91%. Compared with other sizes of convolutional kernels, the PA improved by 0.63%–4.1%, MIoU improved by 0.39%–14.92%, and Dice values improved by 0.04%–14.71%. The segmentation performance is worst for convolutional kernel size of 2. The visualization results of the closed operation with a convolution kernel size of 5 are shown in Fig. 9. It is obvious that the closed operation with a convolution kernel size of 5 can effectively remove the characters and other factors affecting the geological image segmentation, and better maintain the shape and boundary of the geological entities, while too small convolution kernels cannot completely cover the target area of the geological image, which leads to the degradation of the segmentation performance. The segmentation gap is not obvious when the convolutional kernel size is 3, 4, 5, which means that the constructed geological image dataset does not

**Table 1. Segmentation performance for closed operation with different sizes of convolutional kernels.**

| Kernel size | PA | MIoU | FWIoU | Dice |
|---|---|---|---|---|
| (2, 2) | 0.8779 | 0.5699 | 0.7971 | 0.6320 |
| (3, 3) | 0.9174 | 0.7152 | 0.8522 | 0.7787 |
| (4, 4) | 0.9126 | 0.7111 | 0.8450 | 0.7751 |
| (5, 5) | 0.9189 | 0.7191 | 0.8563 | 0.7791 |

have a significant impact on the segmentation results for the closed operation with the convolutional kernel size between 3–5.

*4.3.2. Experimental analysis of different clustering algorithms*

The unsupervised geological image segmentation algorithm proposed in this study obtains the superpixel segmentation map by Felz clustering algorithm, then uses different segmentation models to obtain the feature map of the geological image, and then refines and corrects the feature map by the superpixel segmentation map to finally arrive at the segmentation map. The Felz clustering algorithm uses the edge weight of the difference between the gray value of each pixel and the gray value of the neighboring pixels to combine into connected regions by the similarity between pixels of the geological image is grouped to form different regions, so the performance of the clustering algorithm directly affects the quality of the segmentation results.

To evaluate the presegmentation effect of the Felz clustering algorithm more comprehensively on the geological maps, this study also adds Slic (Achanta R et al., 2012) and K-means (Likas A et al., 2003) clustering algorithms for comparison. Compared with the Slic clustering algorithm, the Felz clustering algorithm can better capture the boundaries of regional elements for geological images, which are rich in texture and have different target sizes, because the segmentation boundaries are calculated based on the degree of difference between pixels in the region; the K-means clustering algorithm is a prototype-based objective function clustering method, and it is also a basic cluster partitioning method (Recky M and Leberl F, 2010; Hu JL et al., 2010). Compared with the Felz clustering algorithm, the K-means clustering algorithm relies heavily on the number of clusters *K*. In geological images, the number and density of features in different regions are different, and it is difficult to determine the appropriate number of clusters, so for geological images, which have different numbers of image features, it requires several attempts to determine the number of cluster centers, and the K-Means clustering algorithm is vulnerable to noise and outliers Therefore, this study mainly compare Felz clustering algorithm and Slic clustering algorithm, and the images processed by Felz clustering algorithm and Slic clustering algorithm are shown in Fig. 10. The parameter settings of Felz clustering algorithm and Slic clustering algorithm are shown in Table 2 and Table 3.

In this study, two clustering algorithms, Slic and Felz, were used to compare the unsupervised segmentation experiments of geological images. To test the performance impact of different clustering algorithms on the unsupervised geological image segmentation, the Slic clustering algorithm and Felz clustering algorithm were used in this paper for a comparative study, and the experimental results are shown in Table 4. The experimental results show that the Felz clustering algorithm outperforms the Slic algorithm in the segmentation task of geological images. Compared with the Slic algorithm, the segmentation performance of the Felz

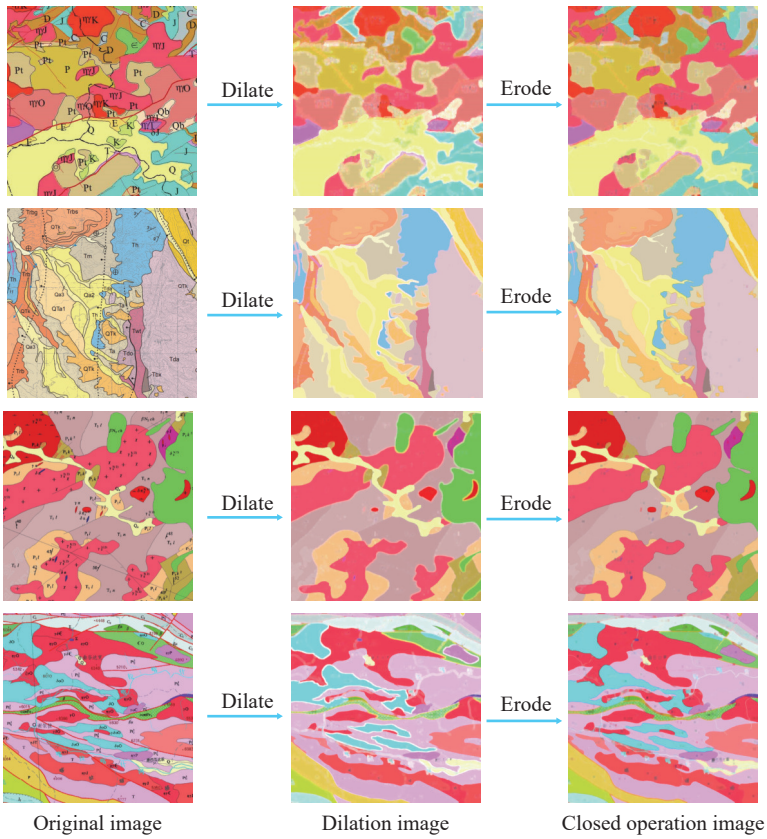


**Fig. 9.** Schematic diagram of mathematical morphology preprocessing. By performing a closed operation on the image using a kernel size of 5 × 5.

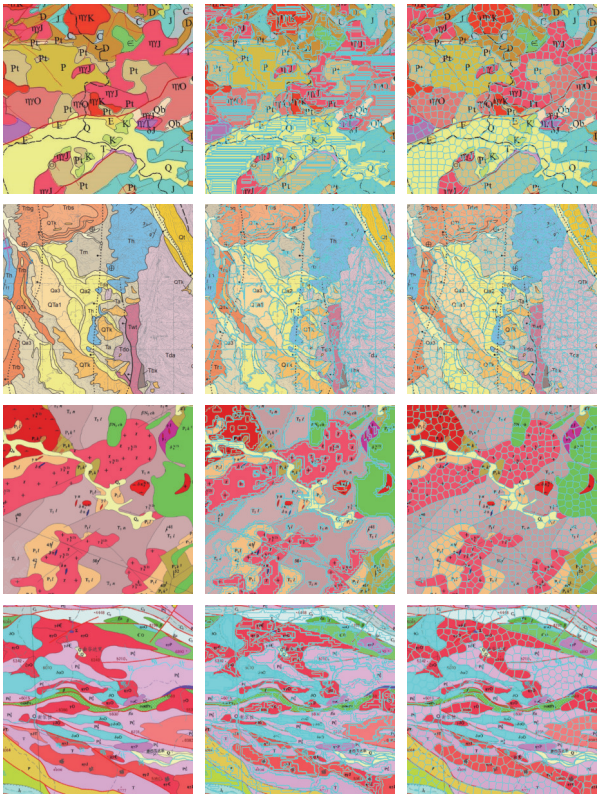


**Fig. 10.** Clustering algorithm processing schematic.

clustering algorithm is significantly improved, for example, the PA value is increased by 3.04%, the MIoU value is increased by 7.45%, the FWIoU value is increased by 5.62%, and the Dice value is increased by 6.81%. In order to reflect the clustering effect of Felz clustering algorithm more visually and Slic clustering algorithm, the visualization results are given in Fig. 10, from which Felz clustering algorithm is more accurate in segmenting the edge stripes and edges of the image, and finer in segmenting the characters and small regional elements in the image. This is because it can perform adaptive segmentation based on the color and spatial distance between pixels, thus accurately capturing the shape and boundaries of geological entities. In contrast, the Slic algorithm is rougher in the processing of small regions as well as boundaries, resulting in less refined segmentation of geological entities, and the Felz algorithm has better computational efficiency. Since geological images usually have high resolution and complex texture features, the computational effort is larger. In contrast, Felz algorithm uses a graph-based segmentation strategy, which can complete the segmentation task in a shorter time and improve the processing efficiency. In summary, this study found that Felz algorithm performs better than Slic algorithm in the unsupervised segmentation task of geological images. It can accurately capture the shape and boundary of geological entities with better robustness and computational efficiency, and is suitable for processing geological images with varying number of features.

*4.3.3. Experimental analysis of different models*

To assess the efficacy of the proposed method in this study, comparison experiments were conducted between the method in this paper and FCN, SegNet, CNN and UNet segmentation algorithms on the color geological image dataset, and PA, MIoU, FWIoU and Dice coefficients were used as evaluation indexes, and the quantitative evaluation results of each image segmentation model are shown in Table 5. The segmentation visualization results are shown in Fig. 11.

As depicted in Fig. 11, compared with other segmentation networks such as UNet and SegNet, the SENet-optimized UNet network is better than other networks in terms of detail processing of region edges and the accuracy of recognizing the location of regions with similar colors, reflecting the advantages of SE-UNet in edge details and combining multi-scale information. From the first geological image, SE-UNet is slightly better than other networks in recognizing multiple regions and region boundary accuracy in the processing of multiple geological regions, as well as in recognizing regions with similar colors accurately compared with other networks. In the geological map with rich colors and many similar colors, SE-UNet can recognize the similar color areas with more accurate edge details than other networks, while other networks can recognize different areas as the same area and miss the recognition. In general, the SENet-optimized network can improve the segmentation ability of color geological image elements more obviously and the overall recognition is more accurate. It is conducive to the automatic recognition of color geological images, and achieves the effect of real-time intelligent and fast recognition. It can effectively provide a reasonable solution for researchers to analyze regional mineral distribution and predict disaster information.

**Table 2. Felz clustering algorithm parameter setting.**

| Clustering algorithm | Scale | Sigma | Min_size | Others |
|---|---|---|---|---|
| Felz | 40 | 0.4 | 60 | Default |

**Table 3. Slic clustering algorithm parameter setting.**

| Clustering algorithm | Compactness | N_segments | Max_iter | Others |
|---|---|---|---|---|
| Slic | 10 | 1000 | 10 | Default |

**Table 4. Segmentation performance for different clustering algorithms.**

| Clustering algorithm | PA | MIoU | FWIoU | Dice |
|---|---|---|---|---|
| Felz | 0.9189 | 0.7191 | 0.8563 | 0.7791 |
| Slic | 0.8885 | 0.6446 | 0.8001 | 0.7110 |

**Table 5. Evaluation of segmentation performance for different models.**

| Model | PA | MIoU | FWIoU | Dice | Inference Time |
|---|---|---|---|---|---|
| SegNet | 0.7777 | 0.5317 | 0.6426 | 0.5871 | 8.70s |
| CNN | 0.8908 | 0.6162 | 0.8077 | 0.6641 | 8.14s |
| FCN-8s | 0.7836 | 0.4896 | 0.6571 | 0.5535 | 11.66s |
| UNet | 0.9035 | 0.6479 | 0.8355 | 0.7066 | 15.62s |
| SE-UNet | 0.9189 | 0.7191 | 0.8563 | 0.7791 | 14.72s |

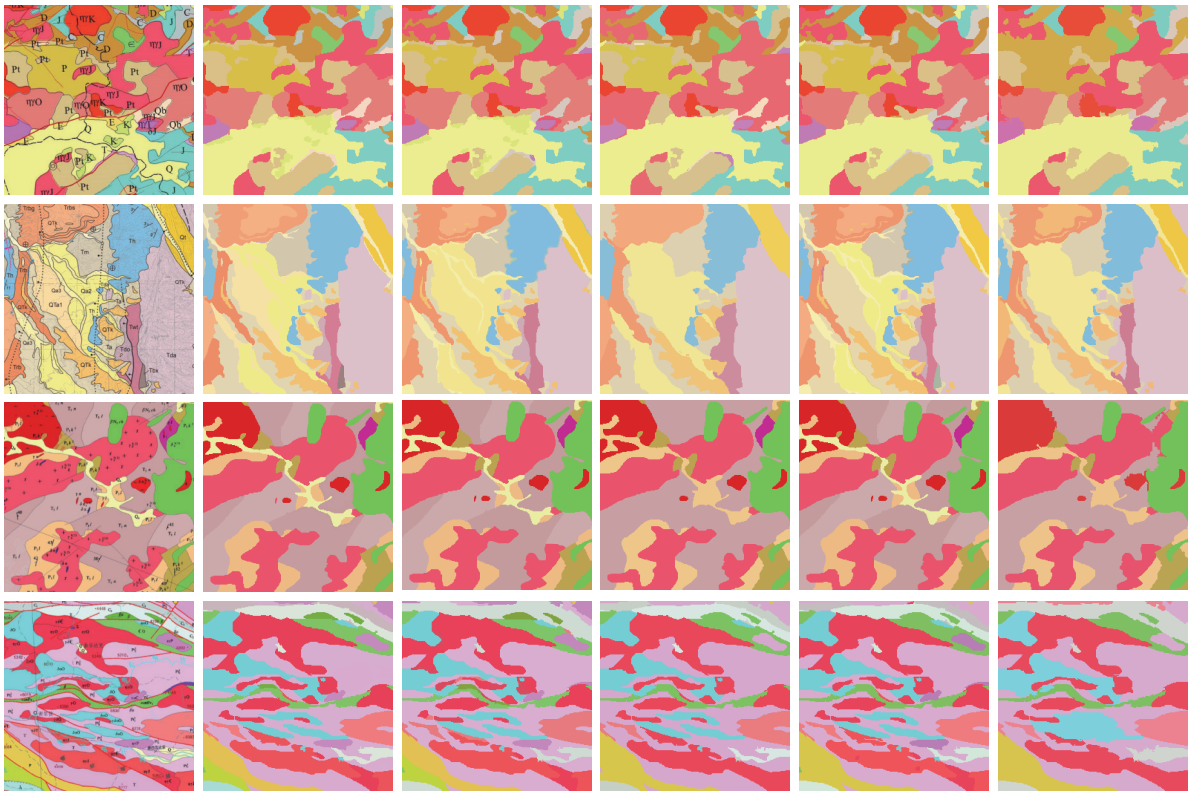


**Fig. 11.** Effect comparison of different models on color geological image segmentation.

To test the effects of diverse models on the efficacy of unsupervised geological image segmentation, this paper tested several commonly used segmentation models in the domain of image segmentation, including UNet, SegNet, FCN, etc., and evaluated their performance in terms of segmentation performance, and the experimental results are shown in Table 5. The experimental results show that SE-UNet has the best performance in the unsupervised geological image segmentation task, and compared with the original UNet network, the PA value increases from 90.35% to 91.89%, MIoU value increases from 64.79% to 71.91%, FWIoU value increases from 83.55% to 85.63%, and Dice value increases from 70.66% to 77.91%. And all of them are also improved compared to SegNet, FCN and CNN networks. Although the inference time of SE-UNet is slightly longer compared to some models, its performance advantage can compensate for the difference in inference time. Moreover, the inference time of SE-UNet remains within an acceptable range and will not have a significant impact on practical applications. Additionally, to some extent, it is constrained by the computational power of the experimental equipment, and subsequent testing can be conducted using devices with higher computational capabilities. In summary, the SE-UNet network proposed in this study is more suitable for the segmentation task of geological images.

*4.3.4. Experimental analysis of different optimizers*

In the domain of image segmentation, the commonly used optimizers include SGD (Song S et al., 2013), Adam (Kingma DP and Ba J, 2014), and RMSprop (Tieleman T, 2012). SGD optimizer is able to achieve a global optimal solution compared to the Adam optimizer, and its optimal accuracy during training is also higher than that of other optimization algorithms, as it can use the limited information more effectively in the case of more redundant information. Its simple structure and faster calculation speed can effectively save time; while Adam tends to oscillate at the local minimum, which may cause a sudden increase in the learning rate in some datasets, thus resulting in a nonconvergence situation.

In this study, the performance impact of SGD, Adam, and AdamW optimizers on unsupervised geological image segmentation was compared through experiments, and the quality of segmentation results was evaluated through indicators such as PA, MIoU, FWIoU, and Dice values. The experimental results are shown in Table 6. The SGD optimizer showed the best test results with the following metrics: 91.89% for PA, 71.91% for MIoU, 85.63% for FWIoU, and 77.91% for Dice value. The PA values of all three optimizers are above 85%, and most of the pixels in the geological images are better classified, but the optimizers of SGD improve 10.82% and 8.87% in MIoU values compared with Adam and AdamW, respectively, indicating that the learning rate adjustment and parameter update strategies of Adam and AdamW in the optimization process do not fully match the characteristics of geological image segmentation match, which leads to a certain degree of performance degradation.

*4.3.5. Ablation experiment*

To further confirm that each module added in this paper really plays a role in improving the segmentation accuracy, ablation experiment is carried out on the constructed color geological image data set. Ablation experiment refers to the comparison of the segmentation performance of geological images after different modules are added based on the basic network used in this paper. The experimental results are shown in Table 7. In order to facilitate statistics, C and S are used to represent the removed convolutional layer and SENet. In addition, "T" represents the modules left in the UNet network, and "F" represents the modules eliminated.

Table 7 shows that based on the original UNet network, adding the SENet attention mechanism alone, PA, MIoU, FWIoU, and Dice are improved by 1.28%, 4.79%, 1.30%, and 4.81%, respectively, and from the above data, it can be seen that the UNet with SENet attention module shows a certain improvement compared to the original UNet, because SENet can effectively capture important features such as edges and element details in the geological images, and add more attention to the key regions that help the segmentation task, which makes the network have stronger perceptual ability and self-adaptability, and can better adapt to the characteristics of different geological images, thus improving the quality of the segmentation results. The UNet with the convolutional layer removed alone has a slightly better performance compared to the original. This indicates that reducing the number of down-sampling and up-sampling layers can reduce the information loss and blurring to a certain extent in the task of geological image segmentation, but the magnitude of this improvement is relatively small; fusing the above two modules to obtain SE-UNet, its PA value is 91.89%, MIoU value is 71.91%, FWIoU value is 85.63% and Dice value of 77.91%, which are better than the first three models in terms of performance, so it can be seen that incorporating the two modules into the UNet network gives the best results, and the addition of each module has a certain improvement on geological image segmentation.

**Table 6. Evaluation of segmentation performance for different optimizers.**

| Optimizer | PA | MIoU | FWIoU | Dice |
|---|---|---|---|---|
| Adam | 0.8896 | 0.6109 | 0.8069 | 0.6612 |
| AdamW | 0.8870 | 0.6304 | 0.8181 | 0.6928 |
| SGD | 0.9189 | 0.7191 | 0.8563 | 0.7791 |

**Table 7. Ablation study of individual modules on overall model performance.**

| Model | UNet | S | C | SE-UNet |
|---|---|---|---|---|
| C | F | F | T | T |
| S | F | T | F | T |
| PA | 0.9035 | 0.9163 | 0.9041 | 0.9189 |
| MIoU | 0.6479 | 0.6958 | 0.6641 | 0.7191 |
| FWIoU | 0.8355 | 0.8485 | 0.8396 | 0.8563 |
| Dice | 0.7066 | 0.7547 | 0.7263 | 0.7791 |

## 5. Discussion

Geological images typically encompass a diverse array of geological entities, such as rocks, minerals, and sediments, each characterized by distinct features like color, shape, size, and texture. Given the irregularity in the number and spatial distribution of these entities, achieving comprehensive coverage of all possible entities within geological images, including those not present in training data, poses a significant challenge in supervised image segmentation methods. To address this issue, this study introduces an enhanced unsupervised segmentation approach that combines the SE-UNet network with the Felz clustering algorithm specifically tailored for color geological images. The research evaluates and compares five deep learning methods, conducting ablation experiments to validate the efficacy of the enhanced components and assess their impact on segmentation performance. Evaluation metrics including PA, MIoU, FWIoU, and Dice are employed to confirm the accuracy and reliability of the models, showcasing their adaptability in handling scenarios involving color planar geological images. The proposed method leverages the intrinsic color feature information of geological images for clustering, enabling superpixel segmentation without external label data dependency. By effectively capturing key geological image features, the method achieves precise geological element segmentation, particularly suited for intricate and dynamic geological settings. The application of the model to the geological map of the Guangdong coastal zone demonstrates its proficiency, with the SE-UNet model successfully extracting 13 out of 15 geological elements, outperforming the UNet and SegNet models extracting 12 and 8 geological elements, respectively. Despite this success, two limitations warrant consideration concerning the proposed unsupervised color geological image semantic segmentation method and the associated experimental procedures.

The proposed segmentation approach encounters challenges in fully accommodating various color geological images within the geological domain. Instances arise where regions with analogous colors in geological images result in mis-segmentation issues. This dilemma stems from the reliance on pixel feature similarity and the associated adaptive threshold $\tau$ in the Felz clustering algorithm for region determination. Notably, the adaptive threshold's sensitivity to the parameter $k$, which regulates segmentation fineness as a relative scale, plays a crucial role in this context. The selection of an appropriate $k$ value necessitates consideration of factors such as image resolution, segmentation granularity, and computational capacity, as adjusting $k$ impacts the trade-off between detail richness and refinement level in superpixel generation. Furthermore, the method exhibits limitations in accurately segmenting diminutive targets within images due to their limited discernible features. The inherent challenge lies in the neural network's ability to effectively extract and represent intricate details of small targets, compounded by potential visibility issues hindering feature capture. Although

efforts were made in this study to fine-tune the network depth to mitigate detail loss and demonstrate efficacy through ablation experiments, opportunities for enhancement persist. Enhancing the network's feature extraction capability through further refinements stands as a viable avenue for improvement. Moreover, the parameter min in the clustering algorithm governs the minimum size of segmented superpixels, dictating the minimum pixel count per superpixel and guiding the merging of excessively small superpixels into neighboring entities. Tailoring the min parameter to specific image attributes and application contexts is crucial for optimal performance. Typically, a min value ranging between 10–100 is deemed reasonable for effective segmentation outcomes. In essence, while the method does not strive for perfection in color geological image segmentation, acknowledging the impossibility of a universal solution, it aims for practical effectiveness. Experimental validations underscore the method's utility, emphasizing its pragmatic approach to addressing segmentation challenges within the geological domain.

While the unsupervised color geological image segmentation model introduced in this study has demonstrated a certain level of success in the conducted experiments, there exist prospective limitations and avenues for refinement. To enhance the model's performance and broaden its applicability, the authors outline three key research directions for future exploration. (1) The authors propose incorporating additional feature information, such as texture, gradient, and structural details, into the clustering algorithm to refine region differentiation based on color similarities but distinct features. This augmentation aims to mitigate missegmentation errors, thereby augmenting the accuracy and robustness of the segmentation methodology. (2) The focus is on optimizing the encoding-decoding module of the neural network by integrating more advanced attention mechanisms, multi-scale feature fusion strategies, and leveraging pre-trained backbone networks. These enhancements are intended to bolster the model's feature extraction capabilities, particularly in detecting and segmenting small targets within images effectively. (3) To bolster the model's adaptability, future investigations may entail incorporating a broader spectrum of geological image data encompassing diverse geological settings and conditions. This diversified dataset approach seeks to facilitate enhanced learning and generalization within the model, fostering improved performance across varied scenarios. Future endeavors are poised to refine the unsupervised color geological image segmentation method by iteratively refining the algorithm, enhancing network architectures, and fortifying model adaptability. These enhancements aim to elevate the precision of information extraction from geological images, furnishing a dependable and efficacious tool for geological image analysis and interpretation.

## 6. Conclusions

This paper proposes an unsupervised geological image segmentation model combining Felz clustering algorithm and improved SE-UNet network, which is based on the UNet network with encoder-decoder structure, and incorporates the channel attention mechanism SENet, focusing on the key features of the image to enhance the feature expression ability of the network, and using Felz clustering algorithm to guide the neural network to segment the image.

(i) The algorithm proposed in this paper, compared with other mainstream segmentation models, improves the performance on PA, MIoU, FWIoU, and Dice on the constructed colour geological image dataset, and the PA is up to 91.89%, and MIoU is 71.91%, which verifies the effectiveness of the model.

(ii) The image semantic segmentation model proposed in this paper is an unsupervised segmentation scheme without training data, and its training process is the inference process, and the segmentation results can be obtained when training the image, which means that the model can be used without labelled data, and effectively solves the problem of the elements in the coloured geologic image not being able to be effectively labelled (due to the variable nature of the elements in the geologic maps) and the inability to perform segmentation in a supervised manner.

(iii) The model can help geologists quickly and accurately extract the geological elements in the image, help geologists identify and understand the geological elements, and provide reliable technical support for geological image analysis and interpretation. Moreover, geological exploration and resource evaluation often need to describe and analyze the spatial distribution and characteristics of geological elements to guide mineral exploration and resource development, so the model can provide support for the above research.

## CRediT authorship contribution statement

Experiment proposal: Kai Ma and Qin-jun Qiu; funding acquisition: Zhong Xie, Ma Kai and Qin-jun Qiu; preliminary research: Jun-jie Liu and Si-qi Lu; data collection: Jun-jie Liu, Miao Tian, Jun-yuan Deng and Ze-hua Huang; experimental design and analysis: Jun-jie Liu and Qin-jun Qiu; writing the original manuscript: Kai Ma, Jun-jie Liu, Qin-jun Qiu; writing-review and editing: Jun-jie Liu, Kai Ma, Qin-jun Qiu. All authors have read and agreed to the published version of the manuscript.

## Declaration of competing interest

The authors declare no conflicts of interest.

## Acknowledgments

The authors are very grateful to National Geological Archives for providing data at no cost. This study was financially supported by the Natural Science Foundation of China (42301492), the Open Fund of Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering (2022SDSJ04, 2024SDSJ03), and the Opening

Fund of Key Laboratory of Geological Survey and Evaluation of Ministry of Education (GLAB 2023ZR01, GLAB2024ZR08) and the Fundamental Research Funds for the Central Universities.

## References


Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. 2012. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11), 2274–2282. doi: 10.1109/TPAMI.2012.120.

Adams R, Bischof L. 1994. Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(6), 641–647. doi: 10.1109/34.295913.

Angulo J, Serra J. 2007. Modelling and segmentation of colour images in polar representations. Image and Vision Computing, 25(4), 475–495. doi: 10.1016/j.imavis.2006.07.018.

Badrinarayanan V, Kendall A, Cipolla R. 2017. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495. doi: 10.1109/TPAMI.2016.2644615.

Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. 2018. 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. doi: 10.1109/TPAMI.2017.2699184.

Chen WK, Zhou WF, Zhu L, Cao Y, Gu HM, Yu B. 2022. MTDCNet: A 3D multi-threading dilated convolutional network for brain tumor automatic segmentation. Journal of Biomedical Informatics, 133, 104173. doi: 10.1016/j.jbi.2022.104173.

Cheng YZ. 1995. Mean shift, mode seeking, and clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(8), 790–799. doi: 10.1109/34.400568.

Chockler H, Farchi E, Godlin B, Novikov S. 2007. Cross-entropy based testing. Formal Methods in Computer Aided Design (FMCAD'07). November 11–14, 2007, Austin, TX, USA. IEEE, 101–108. doi: 10.1109/FAMCAD.2007.19.

Ding L, Tang H, Bruzzone L. 2021. LANet: Local attention embedding to improve the semantic segmentation of remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 59(1), 426–435. doi: 10.1109/TGRS.2020.2994150.

do Valle RF Jr, Siqueira HE, Valera CA, Oliveira CF, Sanches Fernandes LF, Moura JP, Pacheco FAL. 2019. Diagnosis of degraded pastures using an improved NDVI-based remote sensing approach: An application to the Environmental Protection Area of Uberaba River Basin (Minas Gerais, Brazil). Remote Sensing Applications: Society and Environment, 14, 20–33. doi: 10.1016/j.rsase.2019.02.001.

Felzenszwalb PF, Huttenlocher DP. 2004. Efficient graph-based image segmentation. International Journal of Computer Vision, 59(2), 167–181. doi: 10.1023/B:VISI.0000022288.19776.77.

Glorot X, Bordes A, Bengio Y. 2011. Deep sparse rectifier neural networks. In Proceedings of the fourteenth international conference on artificial intelligence and statistics, 315–323.

Gonzalez RC, Woods RE. 2002. Digital Image Processing (Second Edition). Beijing, Publishing House of Electronics Industry, 455.

Hinton GE, Osindero S, Teh YW. 2006. A fast learning algorithm for deep belief nets. Neural Computation, 18(7), 1527–1554. doi: 10.1162/neco.2006.18.7.1527.

Hosseini-Fard E, Roshandel-Kahoo A, Soleimani-Monfared M, Khayer K, Ahmadi-Fard AR. 2022. Automatic seismic image segmentation by introducing a novel strategy in histogram of oriented gradients. Journal of Petroleum Science and Engineering, 209, 109971. doi: 10.1016/j.petrol.2021.109971.

Hu J, Shen L, Sun G. 2018. Squeeze-and-excitation networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18–23, 2018, Salt Lake City, UT, USA. IEEE, 7132–7141. doi: 10.1109/CVPR.2018.00745.

Hu JL, Deng JB, Zou SS. 2010. A novel algorithm for color space conversion model from CMYK to LAB. Journal of Multimedia, 5(2), 159. doi: 10.4304/jmm.5.2.159-166.

Huang CL, Chen JJ, Chen CJ, Wu YG. 2016. Geological segmentation on UAV aerial image using shape-based LSM with dominant color. 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA). March 23–25, 2016, Crans-Montana, Switzerland. IEEE, 928–933. doi: 10.1109/WAINA.2016.82.

Huang MX, Yu WJ, Zhu DH. 2012. An improved image segmentation algorithm based on the otsu method. 2012 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. August 8–10, 2012, Kyoto, Japan. IEEE, 135–139. doi: 10.1109/SNPD.2012.26.

Ji J, Lu XC, Luo M, Yin MH, Miao QG, Liu XZ. 2020. Parallel fully convolutional network for semantic segmentation. IEEE Access, 9, 673–682. doi: 10.1109/ACCESS.2020.3042254.

Ji XQ, Li Y, Cheng JZ, Yu YH, Wang MJ. 2015. Cell image segmentation based on an improved watershed algorithm. 2015 8th International Congress on Image and Signal Processing (CISP). October 14–16, 2015, Shenyang, China. IEEE, 433–437. doi: 10.1109/CISP.2015.7407919.

Jiang F, Wang G, He P, Zheng CC, Xiao ZY, Wu Y. 2022. Application of canny operator threshold adaptive segmentation algorithm combined with digital image processing in tunnel face crevice extraction. The Journal of Supercomputing, 78(9), 11601–11620. doi: 10.1007/s11227-022-04330-9.

Kapur JN, Sahoo PK, Wong AKC. 1985. A new method for gray-level picture thresholding using the entropy of the histogram. Computer Vision, Graphics, and Image Processing, 29(3), 273–285. doi: 10.1016/0734-189X(85)90125-2.

Kingma DP, Ba J, Hammad MM. 2014. Adam: A method for stochastic optimization: 1412.6980. https://arxiv.org/abs/1412.6980v9.

Kittler J, Illingworth J. 1986. Minimum error thresholding. Pattern Recognition, 19(1), 41–47. doi: 10.1016/0031-3203(86)90030-0.

Lang Y, Zheng D. 2016. An Improved Sobel Edge Detection OperatorAdvances in Intelligent Systems Research", "Proceedings of the 2016 6th International Conference on Mechatronics, Computer and Education Informationization (MCEI 2016). November 11-13, 2016. Shenyang, China. Atlantis Press, 590–593. doi: 10.2991/mcei-16.2016.123.

Levachkine S, Velàzquez A, Alexandrov V, Kharinov M. 2002. Semantic analysis and recognition of raster-scanned color cartographic images. Graphics Recognition Algorithms and Applications. Berlin, Heidelberg: Springer Berlin Heidelberg, 178–189. doi: 10.1007/3-540-45868-9_15.

Levinshtein A, Stere A, Kutulakos KN, Fleet DJ, Dickinson SJ, Siddiqi K. 2009. TurboPixels: Fast superpixels using geometric flows. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(12), 2290–2297. doi: 10.1109/TPAMI.2009.96.

Leyk S, Boesch R. 2010. Colors of the past: Color image segmentation in historical topographic maps based on homogeneity. GeoInformatica, 14(1), 1–21. doi: 10.1007/s10707-008-0074-z.

Li ES, Zhu SL, Zhu BS, Zhao Y, Xia CG, Song LH. 2009. An adaptive edge-detection method based on the canny operator. 2009 International Conference on Environmental Science and Information Application Technology. July 4–5, 2009, Wuhan, China. IEEE, 465–469. doi: 10.1109/ESIAT.2009.49.

Lin GS, Milan A, Shen CH, Reid I. 2017. RefineNet: Multi-path refinement networks for high-resolution semantic segmentation. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21–26, 2017, Honolulu, HI, USA. IEEE, 5168–5177. doi: 10.1109/CVPR.2017.549.

Liu HQ, Yao MB, Xiao XM, Xiong YG. 2023. RockFormer: A U-shaped transformer network for Martian rock segmentation. IEEE Transactions on Geoscience and Remote Sensing, 61, 4600116. doi: 10.1109/TGRS.2023.3235525.

Likas A, Vlassis N, Verbeek JJ. 2003. The global k-means clustering algorithm. Pattern Recognition, 36(2), 451–461. doi: 10.1016/S0031-3203(02)00060-2.

Liu Y, Wu LZ. 2018. High performance geological disaster recognition using deep learning. Procedia Computer Science, 139, 529–536. doi: 10.1016/j.procs.2018.10.237.

Long J, Shelhamer E, Darrell T. 2015. Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 7–12, 2015, Boston, MA, USA. IEEE, 3431–3440. doi: 10.1109/CVPR.2015.7298965.

Ma BJ, Pereira JLJ, Oliva D, Liu S, Kuo YH. 2023. Manta ray foraging optimizer-based image segmentation with a two-strategy enhancement. Knowledge-Based Systems, 262, 110247. doi: 10.1016/j.knosys.2022.110247.

Ma ZJ, Mei G. 2021. Deep learning for geological hazards analysis: Data, models, applications, and opportunities. Earth-Science Reviews, 223, 103858. doi: 10.1016/j.earscirev.2021.103858.

Otsu N. 1979. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62–66. doi: 10.1109/TSMC.1979.4310076.

Paszke A, Chaurasia A, Kim S, Culurciello E. 2016. ENet: A deep neural network architecture for real-time semantic segmentation, 1606.02147. https://arxiv.org/abs/1606.02147v1.

Pham DL, Xu CY, Prince JL. 2000. Current methods in medical image segmentation. Annual Review of Biomedical Engineering, 2, 315–337. doi: 10.1146/annurev.bioeng.2.1.315.

Qiu QJ, Ma K, Lv HR, Tao LF, Xie Z. 2023. Construction and application of a knowledge graph for iron deposits using text mining analytics and a deep learning algorithm. Mathematical Geosciences, 55(3), 423–456. doi: 10.1007/s11004-023-10050-4.

Qiu QJ, Tan YJ, Ma K, Tian M, Xie Z, Tao LF. 2023. Geological symbol recognition on geological map using convolutional recurrent neural network with augmented data. Ore Geology Reviews, 153, 105262. doi: 10.1016/j.oregeorev.2022.105262.

Rahimi H, Abedi M, Yousefi M, Bahroudi A, Elyasi GR. 2021. Supervised mineral exploration targeting and the challenges with the selection of deposit and non-deposit sites thereof. Applied Geochemistry, 128, 104940. doi: 10.1016/j.apgeochem.2021.104940.

Rahman MA, Wang Y. 2016. Optimizing intersection-over-union in deep neural networks for image segmentation. Advances in Visual Computing. Cham, Springer International Publishing, 234–244. doi: 10.1007/978-3-319-50835-1_22.

Rauch A, Sartori M, Rossi E, Baland P, Castelltort S. 2019. Trace information extraction (TIE): A new approach to extract structural information from traces in geological maps. Journal of Structural Geology, 126, 286–300. doi: 10.1016/j.jsg.2019.06.007.

Recky M, Leberl F. 2010. Windows detection using K-means in CIE-lab color space. 2010 20th International Conference on Pattern Recognition. August 23–26, 2010, Istanbul, Turkey. IEEE, 356–359. doi: 10.1109/ICPR.2010.96.

Ronneberger O, Fischer P, Brox T. 2015. U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham, Switzerland: Springer International Publishing, 234–241. doi: 10.1007/978-3-319-24574-4_28.

Rosenfeld A. 1981. The max Roberts operator is a hueckel-type edge detector. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-3(1), 101–103. doi: 10.1109/TPAMI.1981.4767056.

Sezgin M, Sankur B. 2004. Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic Imaging, 13, 146–165 doi: 10.1117/1.1631315.

Shamir RR, Duchin Y, Kim J, Sapiro G, Harel N. 2019. Continuous dice coefficient: A method for evaluating probabilistic segmentations, 1906.11031. https://arxiv.org/abs/1906.11031v1.

Song S, Chaudhuri K, Sarwate AD. 2013. Stochastic gradient descent with differentially private updates. 2013 IEEE Global Conference on Signal and Information Processing. December 3-5, 2013, Austin, TX, USA. IEEE, 245–248. doi: 10.1109/GlobalSIP.2013.6736861.

Teimouri N, Dyrmann M, Jørgensen RN. 2019. A novel spatio-temporal FCN-LSTM network for recognizing various crop types using multi-temporal radar images. Remote Sensing, 11(8), 990. doi: 10.3390/rs11080990.

Tian M, Ma K, Liu ZH, Qiu QJ, Tan YJ, Xie Z. 2023. Recognition of geological legends on a geological profile *via* an improved deep learning method with augmented data using transfer learning strategies. Ore Geology Reviews, 153, 105270. doi: 10.1016/j.oregeorev.2022.105270.

Tian Z, Huang WL, He T, He P, Qiao Y. 2016. Detecting text in natural image with connectionist text proposal network. Computer Vision – ECCV 2016. Cham, Switzerland: Springer International Publishing, 56–72. doi: 10.1007/978-3-319-46484-8_4.

Tieleman T. 2012. Lecture 6. 5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural networks for machine learning, 4(2), 26.

Tremeau A, Borel N. 1997. A region growing and merging algorithm to color segmentation. Pattern Recognition, 30(7), 1191–1203. doi: 10.1016/S0031-3203(96)00147-1.

Ulupinar F, Medioni G. 1990. Refining edges detected by a LoG operator. Computer Vision, Graphics, and Image Processing, 51(3), 275–298. doi: 10.1016/0734-189X(90)90004-F.

Van den Bergh M, Boix X, Roig G, Van Gool L. 2015. SEEDS: Superpixels extracted *via* energy-driven sampling. International Journal of Computer Vision, 111(3), 298–314. doi: 10.1007/s11263-014-0744-2.

Vincent L, Soille P. 1991. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(6), 583–598. doi: 10.1109/34.87344.

Wang F, Jiang MQ, Qian C, Yang S, Li C, Zhang HG, Wang XG, Tang XO. 2017. Residual attention network for image classification. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21–26, 2017, Honolulu, HI, USA. IEEE, 6450–6458. doi: 10.1109/CVPR.2017.683.

Wei YC, Xiao HX, Shi HH, Jie ZQ, Feng JS, Huang TS. 2018. Revisiting dilated convolution: A simple approach for weakly- and semi-supervised semantic segmentation. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18–23, 2018, Salt Lake City, UT, USA. IEEE, 7268–7277. doi: 10.1109/CVPR.2018.00759.

Xu JL, Wen XP, Zhang HN, Luo DY, Li JB, Xu LL, Yu M. 2020. Automatic extraction of lineaments based on wavelet edge detection and aided tracking by hillshade. Advances in Space Research, 65(1), 506–517. doi: 10.1016/j.asr.2019.09.045.

Yang L, Wu XY, Zhao DW, Li H, Zhai J. 2011. An improved Prewitt algorithm for edge detection based on noised image. 2011 4th International Congress on Image and Signal Processing. October 15–17, 2011, Shanghai, China. IEEE, 1197–1200. doi: 10.1109/CISP.2011.6100495.

Yu F, Koltun V. 2015. Multi-scale context aggregation by dilated convolutions. 1511.07122. https://arxiv.org/abs/1511.07122v3.

Yu HG, Tao JF, Qin CJ, Liu MY, Xiao DY, Sun H, Liu CL. 2022. A novel constrained dense convolutional autoencoder and DNN-based semi-supervised method for shield machine tunnel geological formation recognition. Mechanical Systems and Signal Processing, 165, 108353. doi: 10.1016/j.ymssp.2021.108353.

Zhang WH, Wang X, You W, Chen JF, Dai P, Zhang PB. 2019. RESLS: Region and edge synergetic level set framework for image segmentation. IEEE Transactions on Image Processing, 29, 57–71. doi: 10.1109/TIP.2019.2928134.

Zhang Y, Chen JQ, Li YL. 2022. Segmentation and quantitative analysis of geological fracture: A deep transfer learning approach based on borehole televiewer image. Arabian Journal of Geosciences, 15(3), 300. doi: 10.1007/s12517-022-09536-y.

Zhang YJ. 2006. An overview of image and video segmentation in the last 40 years. Advances in Image and Video Segmentation, IGI Global, 1–16. doi: 10.4018/978-1-59140-753-9.ch001.

Zheng SX, Lu JC, Zhao HS, Zhu XT, Luo ZK, Wang YB, Fu YW, Feng JF, Xiang T, Torr PHS, Zhang L. 2021. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 20–25, 2021, Nashville, TN, USA. IEEE, 6877–6886. doi: 10.1109/CVPR46437.2021.00681.

Zhou TX, Ruan S, Vera P, Canu S. 2022. A Tri-Attention fusion guided multi-modal segmentation network. Pattern Recognition, 124, 108417. doi: 10.1016/j.patcog.2021.108417.