minerals

MDPI

Article

High-Accuracy Image Segmentation Based on Hybrid Attention Mechanism for Sandstone Analysis

Lanfang Dong [1,2,3,4,*], Hao Gui [3], Xiaolu Yu [2], Xinming Zhang [4] and Mingyang Xu [5]

[1] State Key Laboratory of Shale Oil and Gas Enrichment Mechanisms and Effective Development, Petroleum Exploration and Production Research Institute, China Petrochemical Corporation, Beijing 102206, China
[2] Sinopec Key Laboratory of Petroleum Accumulation Mechanisms, Petroleum Exploration and Production Research Institute, China Petrochemical Corporation, Wuxi 214126, China; yuxl.syky@sinopec.com
[3] School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China; guihao@mail.ustc.edu.cn
[4] Institute of Advanced Technology, University of Science and Technology of China, Hefei 230031, China; xm_zhang@mail.ustc.edu.cn
[5] Anhui Rank Artificial Intelligent Technology Co., Ltd., Hefei 230088, China; xumingyang@ranktek.cn
* Correspondence: lfdong@ustc.edu.cn

Abstract: Mineral image segmentation based on computer vision is vital to realize automatic mineral analysis. However, current image segmentation methods still cannot effectively solve the problem of sandstone grains that are adjoined and concealed by leaching processes, and the segmentation performance of small and irregular grains still needs to be improved. This investigation explores and designs a Mask R-CNN-based sandstone image segmentation model, including a hybrid attention mechanism, loss function construction, and receptive field enlargement. Simultaneously, we propose a high-quality sandstone dataset with abundant labels named SMISD to facilitate comprehensive training of the model. The experimental results show that the proposed segmentation model has excellent segmentation performance, effectively solving adhesion and overlap between adjacent grains without affecting the classification accuracy. The model has comparable performance to other models on the COCO dataset, and performs better on SMISD than others.

Keywords: mineral image analysis; sandstone grain segmentation; deep learning; Mask R-CNN

Citation: Dong, L.; Gui, H.; Yu, X.; Zhang, X.; Xu, M. High-Accuracy Image Segmentation Based on Hybrid Attention Mechanism for Sandstone Analysis. *Minerals* **2024**, *14*, 544. https://doi.org/10.3390/min14060544

Academic Editors: Gene Hall and Begoña González

Received: 18 March 2024
Revised: 29 April 2024
Accepted: 22 May 2024
Published: 25 May 2024

1. Introduction

Most mineral identification and textural characterization in coarse-grained clastic sediments (such as sandstones) necessitate the use of microscopy assessment, which involves interpretation and the gathering of counting statistics using a light microscope [1–6]. The analysis of sandstone images is a vital part of geological exploration research, as rocks are rich in hydrocarbon resources [7]. Sedimentary petrology is a specialized field that demands significant expertise to interpret and classify porosity, minerals, matrix materials, and mineral cements, along with their associated subclasses. A range of objective lenses, optical techniques, and light sources are employed to identify and quantify components. Certain components in sandstone can be small, dark, or opaque, posing challenges for interpretation even to an experienced petrologist. In such cases, methods beyond light microscopy, such as scanning electron microscopy or bulk measurement techniques like X-ray diffraction, are often utilized to assist in mineral identification [8]. Traditionally, the analysis of sandstone thin sections is primarily conducted by professionals, which is labor-intensive, expensive, and subjective.

At present, classic mineral image segmentation technology primarily relies on the low-level visual information of image pixels and is mainly categorized into three types: (1) the threshold-based mineral image segmentation algorithm, which compares each pixel in the input images with a preset threshold value to segment the target areas [9–12]; (2) the

region-based mineral image segmentation algorithm that divides the original image into different pixel regions, separating the target areas from the background [13–15]; and (3) the specific theory-based mineral image segmentation algorithm that employs more targeted computational methods such as cluster analysis to separate mineral grain images [16]. Although these classic mineral image segmentation methods have improved the efficiency of sandstone analysis, they cannot effectively address the issue of adhesion and overlap between adjacent grains, and their segmentation performance for small and irregular grains is relatively poor. Additionally, the classic image segmentation algorithm requires manual tuning for different types of mineral grains, which lowers the efficiency of mineral image segmentation and increases the time consumption.

Owing to the booming development of deep learning and the exceptional feature extraction capabilities of CNNs (convolutional neural networks), deep learning-based approaches have been increasingly employed in mineral image segmentation. These methods, diverging from traditional ones, adhere to an end-to-end paradigm and significantly surpass conventional methods in accuracy and efficiency. Furthermore, they are data-driven methods (i.e., the larger the dataset or the more refined the data analysis, the higher the achievable accuracy), allowing for enhanced performance with continuous data expansion. Some studies have explored and applied it in mineral image segmentation tasks. An RDU (R: residual connection; DU: DUNet) ore image segmentation model was proposed to estimate the grain size of ore fragments in conveyor belts, which can adjust the receptive field adaptively according to the size and shape of different ore fragments and achieve accurate segmentation [17]. Deep learning-based methods for mineral image segmentation also excel in segmenting adherent, overlapping, and multi-scale mineral grains, effectively addressing the typical challenges encountered in previous approaches [18]. However, existing image segmentation methods mainly focus on semantic segmentation, which cannot meet the requirements to compute specified morphological data in the mineral analysis, such as circularity, particle size, and contact relationship. Furthermore, these models are mostly designed for scenarios with clean image backgrounds, whereas in sandstone analysis scenarios, the image backgrounds are complex and filled with indistinguishable fillers, which makes it a challenge to segment grains out [19].

In order to solve the above problems and further improve the application prospects of automatic sandstone analysis, we propose a high-accuracy instance segmentation model based on Mask R-CNN [20] by introducing a hybrid attention mechanism to better adapt to the complex shapes of sandstone particles. Secondly, we design a shape-aware training loss function for the improved model and conduct an ablation experiment to confirm its advantages. Additionally, original convolution is replaced with dilated convolution for the purpose of obtaining more global information. Finally, this experiment establishes a dataset of sandstone instance segmentation with 40,122 grain labels.

2. Methodology

Section 2 mainly introduces various methods and ideas involved in the improved model of the sandstone image segmentation system. In the model-building stage, backbone selection, module setting, and loss function design are the main considered aspects to improve the segmentation model performance and solve edge blurring and incomplete segmentation in the segmentation process.

2.1. Mask R-CNN

The deep learning-based Mask R-CNN network segmentation method has superior performance and can achieve good results in the sandstone microscopic image segmentation task [21]. We customize and optimize the Mask R-CNN network to form a task specified model, based on the distribution characteristics of sandstone grains in sandstone microscopic images.

The Mask R-CNN network is divided into three steps: object localization, object category calculation, and segmentation mask prediction. The process of the Mask R-CNN

network is as follows: the image passes through the ResNet backbone network, and different levels of feature maps are obtained using the feature pyramid. This facilitates the model in extracting features at different levels. It then enters the region proposal network to generate candidate areas where grains might be present. On the one hand, a classifier determines whether pixels belong to sandstone grains or the background. On the other hand, a box regressor corrects the boundaries of the sandstone grains. The combination of these two steps forms the candidate target areas. Finally, through a fully convolutional neural network, accurate segmentation results are achieved, completely extracting the sandstone grains. The processing procedure is shown in Figure 1.

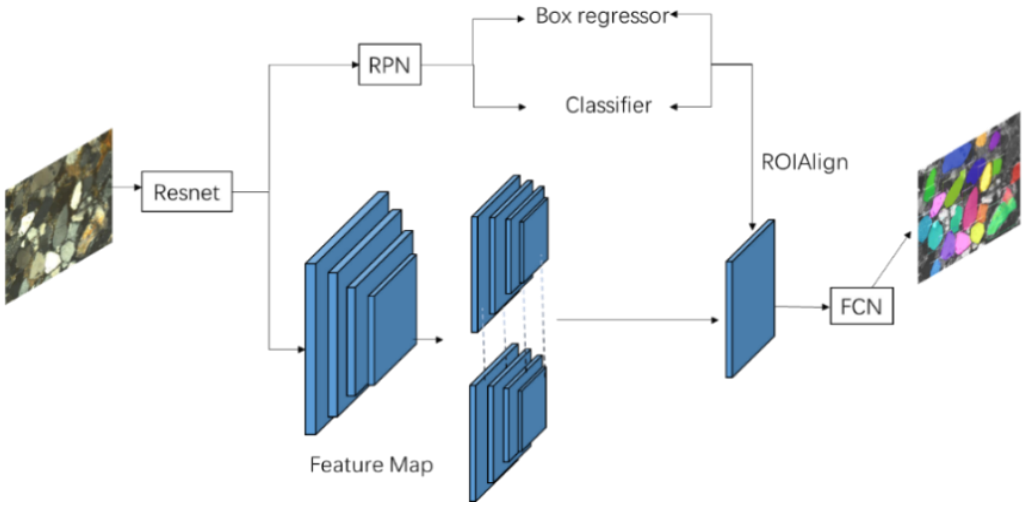


Figure 1. Original Mask R-CNN sandstone image segmentation processing flowchart.

2.2. SE-Net

Squeeze-and-Excitation Network (SE-Net [22]) introduces the concept of channel attention. By modeling and assigning weights to feature channels, it forms the parameters that can be learned and updated and continuously increases the weights of useful feature channels to optimize the generalization capability of the model. In this paper, SE-Net is added to the backbone recognition network to improve the spatial information processing capability of the model and therefore achieve better results for sandstone grains at different scales. The specific structure of SE-Net is shown in Figure 2.

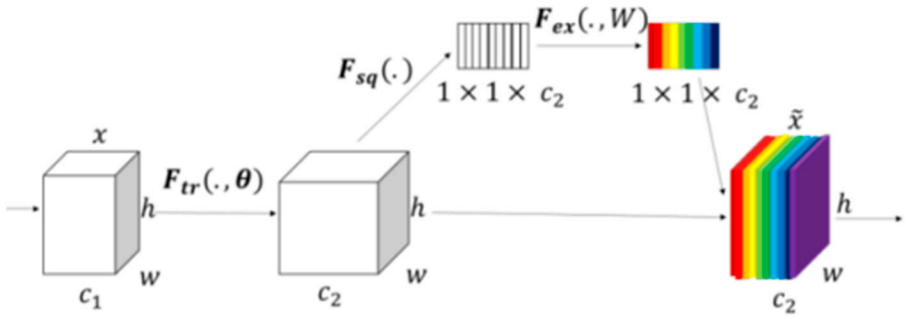


Figure 2. SE-Net structure diagram.

The first operation is F_{tr} conversion, which converts the input x with a feature channel number C_1 by a series of general transformations such as convolution to obtain a feature with a feature channel number C_2. The F_{tr} operation is shown in Equation (1).

$$F_{tr} : X \rightarrow U, \ X \in R^{H \times W \times c_1}, \ U \in R^{H \times W \times c_2} \quad (1)$$

The Squeeze operation is shown in Equation (2), where the two-dimensional feature channels are mapped by compression to obtain a real number, which is connected to form a one-dimensional vector to obtain the global distribution and construct the global receptive field of the model. The dimension number of the vector equals the number of input feature channels.

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \tag{2}$$

Next is the Excitation operation. The parameter w is used to generate the weights for each feature channel. Multiply W_1 (which dimension is $C/r \times C$) and z to reduce the computational complexity by scale operation. The dimension of the feature map of $W_1 z$ is still $1 \times 1 \times C/r$; it is then passed through the ReLU layer and multiplied with W_2 (which dimension is $C \times C/r$) to obtain the feature map dimension $1 \times 1 \times C$. Finally, the Sigmoid is derived to generate the weights (s) of the feature maps (at a total of C). The parameter s incorporates the feature map information of each feature channel and is part of the neural network, which can be learned and optimized.

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \tag{3}$$

Finally, the weights calculated by the model are weighted onto the original feature map by the previous channels through the scale operation, as shown:

$$\widetilde{x_c} = F_{scale}(u_c, s_c) = s_c u_c \tag{4}$$

2.3. Coordinate Attention and Spatial Attention

The channel attention mechanism models global information through channels, allowing the model to effectively extract sandstone grains. When extracting, the location information of the sandstone is directly related to the accuracy of the boundary fit. In this paper, a CA + SP hybrid attention mechanism is proposed to optimize the model by combining the coordinate attention mechanism [23] and the spatial attention mechanism [24].

Coordinate attention takes a similar operation to channel attention in both horizontal and vertical directions to obtain relatively independent feature maps in both directions, effectively preserving one-dimensional location information and establishing spatial long-range dependence in one dimension. This mechanism is very sensitive to coordinate information and can effectively pinpoint spatial coordinate information. Similar to the channel attention mechanism, the coordinate attention mechanism first performs coordinate position encoding and then generates coordinate attention. A coordinate attention module can be seen as a computational unit used to augment feature representation capabilities. It can take any intermediate tensor $X = [x_1, x_2, \ldots, x_C] \in R^{C \times H \times W}$ as input and produces an output $Y = [y_1, y_2, \ldots, y_C]$ of the same size with enhanced representational power. The horizontal and vertical directions are treated separately and computed using pooling kernels of dimensions (H, 1) and (1, W), giving the following outputs for the vertical and horizontal channels.

$$z_c^h(h) = \frac{1}{W} \sum_{0 \le i < W} x_c(h, i), z_c^w(w) = \frac{1}{W} \sum_{0 \le j < W} x_c(j, w) \tag{5}$$

The two feature maps generated by the previous module are cascaded and then transformed using a shared 1×1 convolution to perform the transformation F_1, expressed as in Equation (5); the generated $f \in R^{C/r \times (H+W)}$ is the intermediate feature map during the computation, where r denotes the down-sampled ratio, which is used to control the size of the module, like the SE module.

$$f = \delta\left(F_1\left(\left[z^h, z^w\right]\right)\right) \tag{6}$$

Next, f is cut into two direction-independent tensors, $f^h \in R^{\frac{C}{r} \times H}$, $f^w \in R^{\frac{C}{r} \times W}$, and then using the two 1 × 1 convolutions F_h and F_w, we transform two tensors to the same number of channels as the input and output, as in the following equation:

$$g^h = \sigma\left(F_h\left(f^h\right)\right), g^w = \sigma\left(F_h\left(f^h\right)\right) \tag{7}$$

The two are then expanded as attention weights, and final output of the CA module is

$$\tilde{y_c}(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{8}$$

The spatial attention mechanism can simulate the function of the human eye and extract the parts of interest to the model, which generates a mask for the space, and draws out one way, which undergoes similar operations as described above to form the spatial attention mechanism, effectively extracting its relative spatial information.

2.4. Dilated Convolution

The up-sampling process of bilinear interpolation has large errors, leading to problems such as distortion when generating sandstone grain contours, and this paper achieves refinement of sandstone contours by introducing dilated convolution. The dilated convolution [25] has the following main functions:

1. Expanding the Mask R-CNN network receptive fields more efficiently while taking into account image resolution;
2. Changing the size of the convolution kernel and the perceptual field of the model by adjusting the expansion rate (r) to obtain multi-scale global semantic information.

The dilated convolution is shown in Figure 3. Only non-zero elements play a role in the calculation, and the dilated convolution fills the ordinary convolution with zeros to increase the size of the receptive field, as shown in Equation (9).

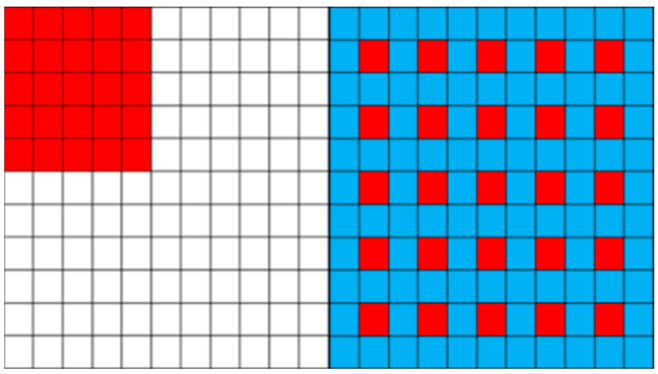


Figure 3. Dilated convolution diagram.

$$K = k + (k-1)(r-1) \tag{9}$$

K is the size of the expanded convolution kernel, k is the size of the original convolution kernel, and r represents the expansion rate. As shown in the figure, when k = 5 and r = 2, compared with the ordinary 5 × 5 convolution on the left, the dilated convolution expands the convolution kernel to 11 × 11, and the range of the sensory field is greatly improved.

3. Methods and Materials

3.1. Experimental Methods

3.1.1. Hybrid Attention Mechanism

In this paper, the SE module is embedded in the backbone network ResNet and its processing flow is shown in Figure 4. In contrast to the SE-Net structure mentioned earlier, Squeeze takes global average pooling and Excitation takes two full connection layers for

computation. To maintain the ResNet structure, it is also necessary to ensure that the input and output dimensions are the same: the feature dimension is first reduced to 1/4 of the input, and then the dimension is recovered after ReLU activation. This operation not only complies more with the specification of the Excitation operation but also reduces the computational costs by reducing the dimensionality. The neural network architecture with two fully connected layers allows for better modeling of correlations between channels. Finally, analogous to the SE-Net model, the scale operation is set to weight the weights by channel onto the corresponding original feature maps to form the SE-ResNet model.

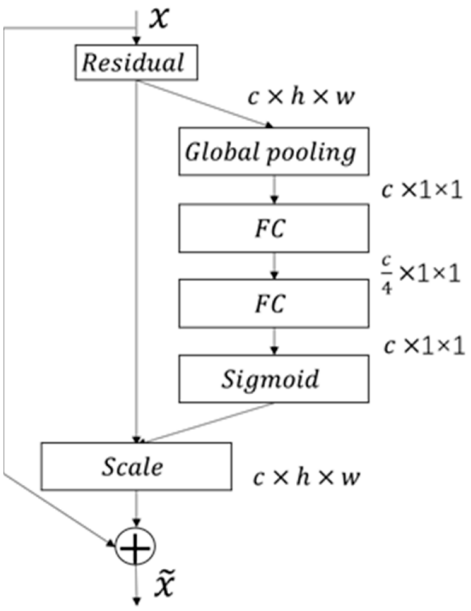


Figure 4. SE-ResNet module architecture diagram.

Additionally, a CA + SP-ResNet backbone network is proposed to optimize the model by combining the coordinate attention mechanism and the spatial attention mechanism in a re-weighted way, which is shown in Figure 5.

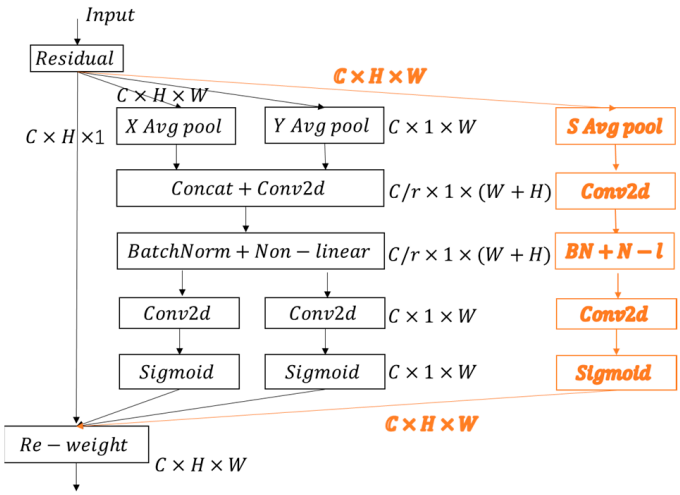


Figure 5. Coordinate attention + spatial attention ResNet module architecture diagram.

After the spatial attention mechanism, the output section of the module should also multiply the coordinate attention output above by the effect of spatial attention and modify the final output to Equation (10).

$$y(i,j) = \widetilde{y}(i,j) \times g_c^s(i,j) \tag{10}$$

Finally, the output of SE-ResNet and CA + SP-ResNet also uses the similar re-weighted method to obtain the final result.

3.1.2. Loss Function

Sandstone grains are often dense and complex in shape. To reduce segmentation error, an accurate sandstone grain boundary is of great importance. In fact, it is necessary to compare the real grain area with the predictions to ensure that their boundaries and contours fit perfectly. Thus, we use DiceLoss [26] to calculate the segmentation error, which is used to calculate the Dice (the graphical similarity between the true and predicted segmentation results of the sandstone grains). Let the true result of segmentation be $|T|$ and the predicted result be $|P|$, and then $|T \cap P|$ be the dot product of the binary plot of the predicted segmentation result and the true segmentation result. The elements in $|T|$ and $|P|$ directly take the summation. The similarity coefficient Dice and segmentation error DiceLoss are as follows:

$$\text{Dice}(P,T) = \frac{|P \cap T|}{(|P+T|)/2} = \frac{2 * \text{gtmask} \cap \text{predmask}}{\text{gtmask} + \text{predmask}} \tag{11}$$

$$\text{DiceLoss}(P,T) = 1 - 2\frac{|P \cap T|}{|P|+|T|} \tag{12}$$

To avoid the denominator of the equation being zero and to reduce over-fitting, smoothing is added to the numerator and denominator at the same time:

$$\text{DiceLoss}(P,T) = 1 - 2\frac{|P \cap T| + \text{smooth}}{|P| + |T| + \text{smooth}} \tag{13}$$

Since cross-entropy is only a form of proxy, we directly adopt DiceLoss as the loss function. The real target of the segmentation is to maximize the overlap between the predicted results and the true segmentation results, i.e., the Dice coefficient, thus minimizing DiceLoss. Also, the prediction of segmentation network accuracy performance often needs to be tested with IoU, which is very similar to DiceLoss; this operation allows the performance metric to be trained and optimized directly as a component of the loss function. Due to the non-convexity of the DiceLoss function, there can be problems such as gradient explosion, which typically leads to problems such as more unstable training and difficulty in converging the training. Through several experiments, this paper uses Log-Cosh DiceLoss, based on the Log-Cosh function, to smooth the DiceLoss. The Cosh function as well as the Log-Cosh DiceLoss equation, i.e., $L_{\text{lc-dce}}$, are shown in Equations (14) and (15).

$$\text{Cosh } x = \frac{e^x + e^{-x}}{2} \tag{14}$$

$$L_{lc-dce} = \log(\cosh(\text{DiceLoss})) \tag{15}$$

By changing the segmentation loss from L_{mask} to $L_{\text{lc-dce}}$, we obtain a more refined simulation of sandstone microscopic image boundaries, effectively improving the segmentation effect and contour refinement for irregular sandstone grains.

3.2. Material Preparation

Physical thin sections were acquired in a standard manner, in which blue epoxy (variable hue) was impregnated into the sandstone (a strictly coarse-grained clastic sediment), an approximately 30 μm thick thin section on a glass slide was prepared, and a mechanical polish was applied to the surface. Immediately prior to imaging, mineral oil and a cover slip were applied to the thin section. Generally, few label sets for petrology disciplines exist and their generation is extremely time-consuming. The images, used for labeling,

were acquired on several different optical microscopes using objective lenses with 5× or 10× magnification.

The dataset SMISD has 288 typical labeled, single-polarized sandstone grain images with a cumulative total of 40,122 sandstone grains. An approach was taken to expand the dataset by rotating the images counterclockwise by different degrees, as shown in Figure 6, to a total of 1121 images with a total of 153,466 grains.

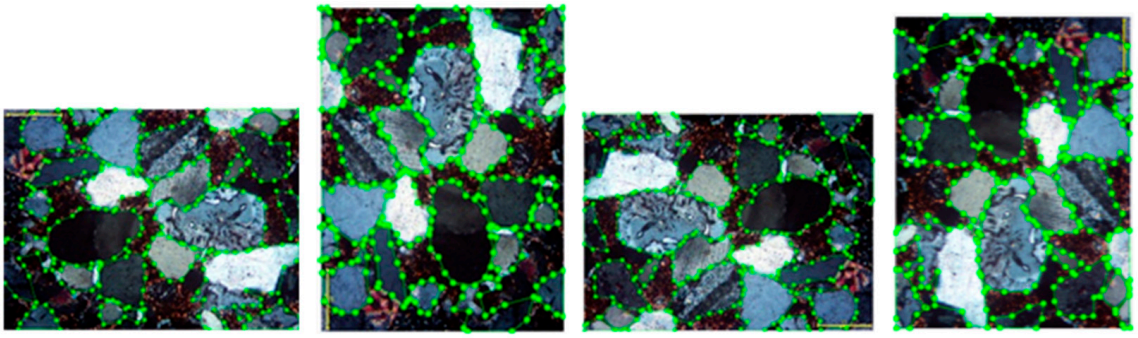


Figure 6. Image enhancement methods for sandstone microscopic image segmentation dataset.

In order to obtain better generalization ability, in this paper, 85% of them are divided into a training set and the remaining 15% are divided into a test set.

4. Results

4.1. Experiments and Analysis

To test the performance of the improved Mask R-CNN model based on the above theory for sandstone microscopic image segmentation, several experiments were conducted based on the traditional Mask R-CNN network and the improved Mask R-CNN network, and the experimental environment is shown in Table 1.

Table 1. Table of experimental environment.

Software/Hardware	Configuration
Operating system	Ubuntu 20.04
Memory	32 GB
CPU	Intel(R) Core(TM) i9-10920X CPU @ 3.50 GHz
GPU	NVIDIA GeForce RTX 3090
Related software	Python3.8/Torch1.8.0/cuda11.1

The basic parameters of the Mask R-CNN network were adjusted to obtain the best experimental results, and the hyper-parameters were set as follows: the model threshold was set to 0.5, the MINI_MASK was set to (28,28) (to facilitate the detection of fine grains), ResNet50 was used for the skeleton network (to increase the model processing speed), and the maximum number of grains present in a single image was set to 1000 (taking into account the dense distribution of grains in some sandstone micro-graphs).

4.2. Performance Metric

For the sandstone microscopic image segmentation task, the IoU is fixed and the average precision rate and average recall rate are taken as evaluation metrics.

4.2.1. Accuracy and Recall Rate, AP

To introduce the performance metric of sandstone microscopic image segmentation models (i.e., the generalization ability), the following four concepts are introduced: TP and FP, respectively, refer to the number of correct and incorrect determinations for all samples judged to be sandstone grains. TN and FN, respectively, refer to the number of correct and

incorrect determinations for samples judged to be non-sandstone grains. Accuracy is the probability that all samples predicted to be sandstone grains are indeed sandstone grains, and it addresses the accuracy of the prediction results for a specific sample and sets the accuracy rate as P, given by the following equation:

$$P = \frac{TP}{TP + FP} \tag{16}$$

The recall rate is the probability that the model correctly predicts a sample that is indeed sandstone grains and it focuses on the prediction of the specific sample. If we set the recall as R, the equation is as follows:

$$R = \frac{TP}{TP + FN} \tag{17}$$

Accuracy and recall rate are interactive, with both being high when the model generalization is strong. The two are usually negatively correlated. If the X-axis is the recall rate and the Y-axis is the precision rate, the PR curve is obtained by tracing the points. The average accuracy (AP) is the PR curve integrated from 0 to 1. The calculation formula is as follows:

$$AP = \int_0^1 P(r)dr \tag{18}$$

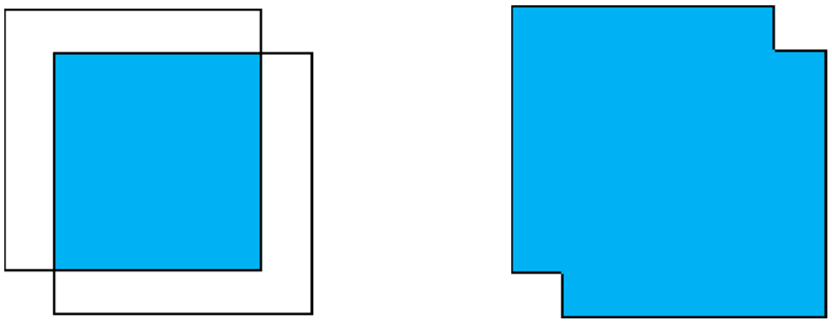
The calculation is simplified by smoothing, which is calculated as follows:

$$AP = \frac{1}{11} \sum_{i=0,0.1,\ldots,1.0} \text{smooth}(i) \tag{19}$$

4.2.2. IoU

IoU [27] represents the overlap ratio between the generated and real image segmentation regions. The calculation of IoU is shown in Equation (20), where C is the real image segmentation area and G is the segmentation area generated by the model. A diagram of IoU is shown in Figure 7. The left side of the image is the intersection of the two and the right side is the merging of the two.

$$IoU = \frac{Area(C) \cap Area(G)}{Area(c) \cup Area(G)} \tag{20}$$



(**a**) Area(C) ∩ Area(G) (**b**) Area(c) ∪ Area(G)

Figure 7. Schematic diagram of the IoU.

The following six metrics for characterizing the performance of image segmentation can be obtained based on the above concepts, separately shown in Table 2.

Table 2. Image segmentation performance indicators.

Average Precision (AP)	
AP	AP When IoU = 0.50:0.05:0.95
AP_{50}	AP When IoU = 0.50
AP_{75}	AP When IoU = 0.75
AP Across Scales	
APs	AP When the grain is small: pixel area < 32^2
AP_M	AP When the grain is medium: 32^2 < pixel area < 96^2
AP_L	AP When the grain is large: pixel area > 96^2

AP_{50} means that the model is judged to be correct when the IoU of the predicted and real grain area is greater than 0.5; then, the classification AP of the model is calculated. AP_{75} needs a higher IoU threshold; thus, the segmentation needs to be more refined.

AP, i.e., AP:IoU = 0.50:0.05:0.95:, means, respectively, setting the correct determination IoU to 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, or 0.95 and then averaging it after calculating the AP value to obtain the final AP.

At the same time, the test results are classified according to the area occupied by the pixels. AP Across Scales means that the detection targets are classified according to area into three categories—small, medium, and large—and then the predictive ability of the model is judged.

The accuracy rate of the model was calculated separately by several indicators such as IoU accuracy and target size to obtain the final image segmentation performance metrics, as shown in Table 2.

Compared to the conventional model, the improved model identified significantly more grains, segmented grain contours were significantly finer, and AP values were improved by about 5%.

The effectiveness of the hybrid attention mechanism, its shape awareness, and its dilated convolution are shown in Tables 3–5, respectively.

Table 3. The effectiveness of the attention mechanism.

Backbone	**SMISD (%)**		
	AP_S	**AP_M**	**AP_L**
ResNet	15.4	37.4	41.2
+SE	18.9	38.1	42.3
+X attention	19.3	38.9	41.9
+Y attention	19.2	38.7	42.0
+CA	20.3	39.5	43.1
+CA + SP	20.8	39.7	43.2

Table 4. The effectiveness of the shape-aware loss function.

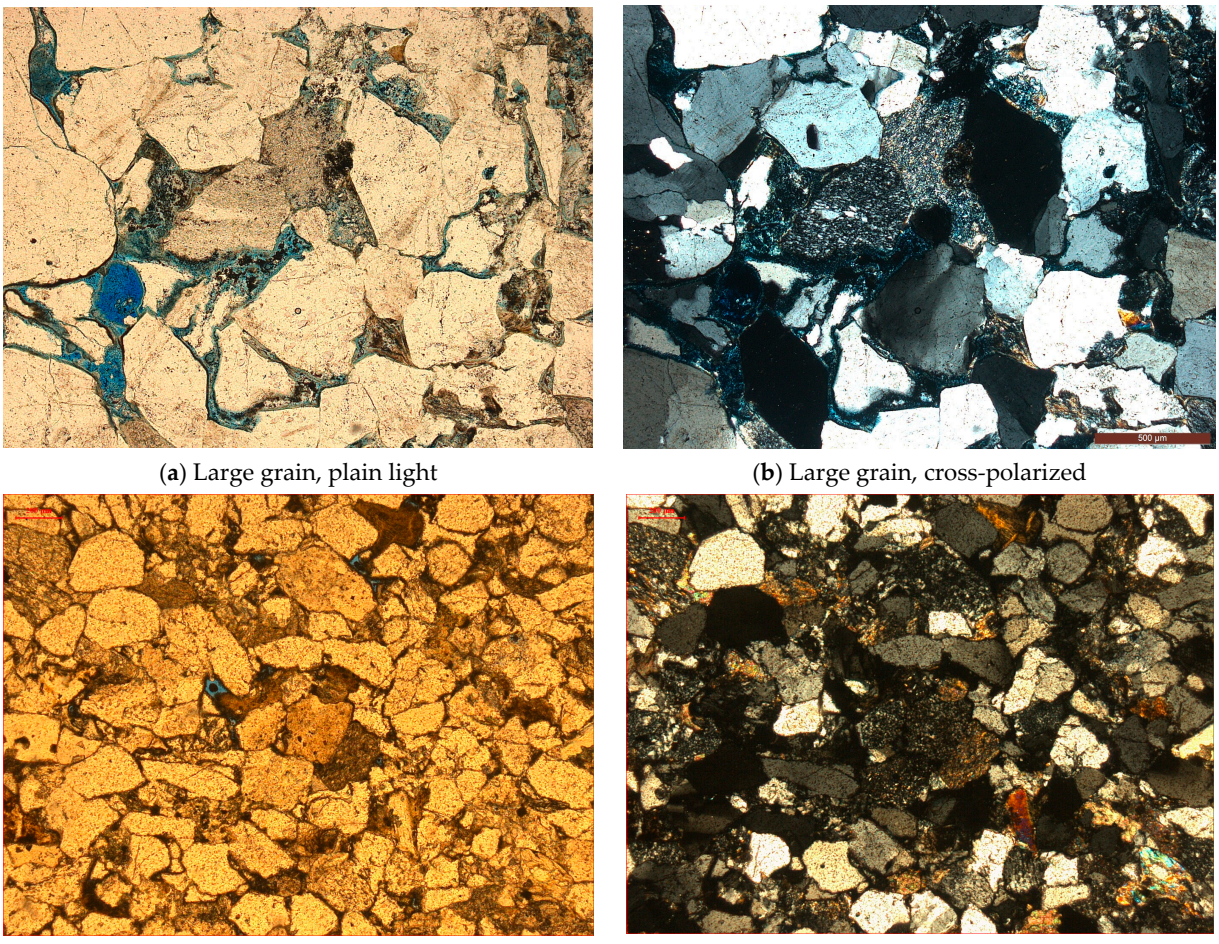
Loss Function	**SMISD (%)**		
	AP	**AP_{50}**	**AP_{75}**
L_{mask}	33.2	38.6	29.1
L_{dice}	36.3	41.3	37.3

Table 5. The effectiveness of dilated convolution.

Type	**SMISD (%)**		
	AP	**AP_{50}**	**AP_{75}**
Original convolution	33.2	38.6	29.1
Dilated convolution	37.9	43.2	34.2

4.2.3. Results and Visualization

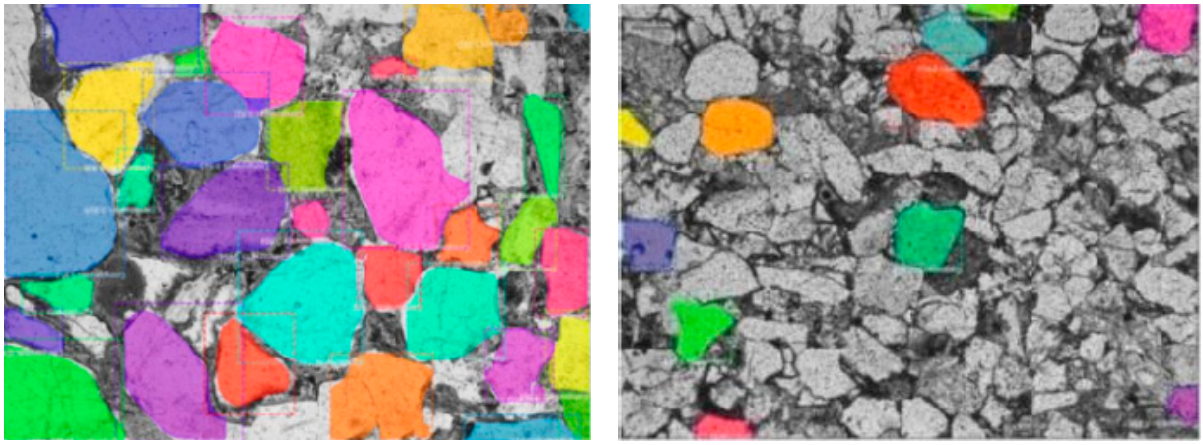
The sandstone micrographic images used to visualize the effectiveness our model are shown in Figure 8. The results of segmentation when directly using the original Mask R-CNN network are shown in Figure 9; the contour fitting was poor for sandstone grains of varying sizes and shapes. The network can only fit about 10% when grains are small.



(**a**) Large grain, plain light (**b**) Large grain, cross-polarized

(**c**) Small grain, plain light (**d**) Small grain, cross-polarized

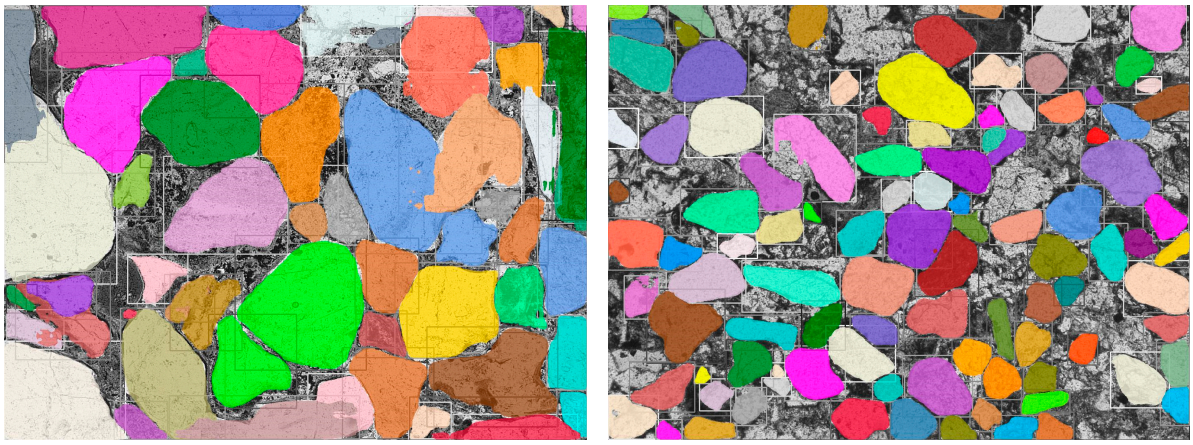
Figure 8. Images used to visualize the effectiveness.



(**a**) Large grains (**b**) Small grains

Figure 9. Processing results of original Mask R-CNN network.

After the above modification, new segmented images are shown in Figure 10. Obviously, the model can identify about 80% of the grains regardless of their size, and the contours fit more closely. The mis-segmentation and under-segmentation are greatly reduced, and the generalization ability of the model is greatly improved.



(**a**) Large grains (**b**) Small grains

Figure 10. Processing results of the improved Mask R-CNN network in this paper.

Comparison of the Results of the Improved Mask R-CNN Network in Fitting Irregular Sandstone Grain Images

For the irregular, narrow sandstone grains, the new model is far superior to the original model in segmentation integrity as well as fit, which is shown in Figure 11.

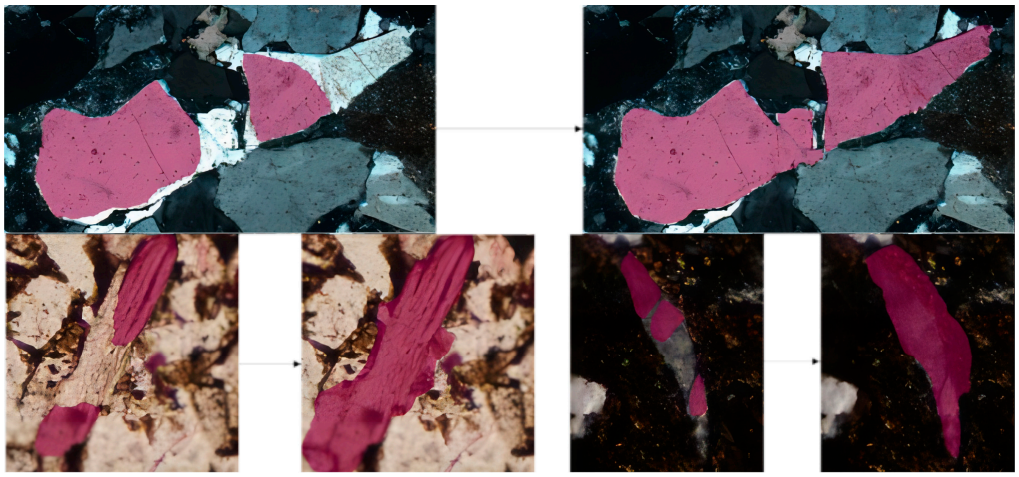


Figure 11. Comparison of the results of the original and improved Mask R-CNN in fitting irregular sandstone grain images.

Experiments on the Performance of the Improved Segmentation Network

The performance of the proposed image segmentation network is predicted by testing 238 sandstone images in the test set. Experiments were conducted using Mask R-CNN, HTC, PointRend, the latest Refine Mask, and the improved image segmentation algorithm (labeled Hybrid + Loss in the table), based on the publicly available COCO image segmentation dataset [28] and the sandstone microscopic dataset in this paper. For the best results, the recognition backbone networks all adopted ResNet101.

As shown in Table 6, our algorithm achieved the best results out of all the algorithms in the sandstone microscopic image segmentation task, based on SMISD. Compared to the conventional Mask R-CNN network, the improvement is 4.3%, 5.2%, and 6.9% for

each of the three grain types (large, medium, and small), which is slightly better than the state-of-the-art Refine Mask network. Hybrid + Loss gained significant improvement in AP values for all sizes of grains, which indicates that the image segmentation algorithm tailored to the sandstone microscopic image segmentation task can effectively identify and extract sandstone grains, and can better fit the contours to achieve fine segmentation of sandstone grains.

Table 6. Performance comparison of segmentation algorithms.

Dataset / Algorithm	COCO (%)				SMISD (%)			
	AP	**AP_S**	**AP_M**	**AP_L**	**AP**	**AP_S**	**AP_M**	**AP_L**
Mask R-CNN	39.6	27.2	49.0	57.7	32.3	15.4	37.4	41.2
HTC	41.2	27.2	51.9	61.5	33.9	15.6	38.9	44.0
PointRend	41.1	27.8	52.0	62.0	35.1	16.3	39.9	45.7
RefineMask	41.8	28.6	53.1	62.8	36.7	18.0	41.1	47.3
Hybrid + Loss	41.7	28.9	52.7	62.5	37.9	19.7	42.6	48.1

5. Discussion

Channel attention can enhance the network's ability to extract image information. Coordinate attention can improve the model's ability to locate boundaries. Spatial attention can enhance the model's receptive capability and optimize the model's generalization ability, that is, its performance on images that have not been used for training. By introducing the hybrid attention mechanism, the model proposed in this paper surpasses other models on SMISD. With the goal of the task—to make the boundaries of grain segmentation results more precise—as the performance evaluation index, a shape-aware loss function can improve the model's segmentation effect on grain contours, especially irregular grains. Therefore, the model adopts Log-Cosh DiceLoss as the model's loss function. Finally, the dilated convolution can expand the receptive field and segment by combining more regional information surrounding the grains, which makes model perform better. Additionally, the dataset with rich labeled grains helps the model fully learn the segment rules, which is vital for data-driven methods.

6. Conclusions

Intelligent analysis of images of thin sandstone sections is of high research and application value, but the search for automatic segmentation and recognition is challenging because of the variety and complexity of sandstone images. This paper designed a high-accuracy sandstone segmentation model by mixing several attention mechanisms and applying appropriate loss functions and dilated convolution. The segmented particle images can be classified using other methods to obtain their mineralogical categories. The research on the automatic analysis of thin sandstone sections has made some progress in this paper, but it is difficult to identify filler, especially heterogeneous groups and cementation, and there is still much room for improvement in recognition accuracy.

Author Contributions: Methodology, H.G.; software, M.X.; validation, X.Z.; investigation, H.G. and L.D.; resources, X.Y.; writing—original draft preparation, H.G.; writing—review and editing, X.Z.; visualization, H.G.; supervision, L.D.; project administration, L.D.; funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the SINOPEC Key Laboratory of Petroleum Accumulation Mechanisms's Microscopic Panoramic Segmentation of Dense Sandstone Open Fund, funding number 33550007-22-ZC0613-0038, and SINOPEC Excellent Youth Technology Innovation Fund, funding number P19028.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from the SINOPEC Key Laboratory of Petroleum Accumulation Mechanisms and are avail-

able from the author Xiaolu Yu with the permission of SINOPEC Key Laboratory of Petroleum Accumulation Mechanisms.

Acknowledgments: Thanks to the SINOPEC Key Laboratory of Petroleum Accumulation Mechanisms for the given sandstone microscopic images.

Conflicts of Interest: Xiaolu Yu is employee of Sinopec. The paper reflects the views of the scientists and not the company. Mingyang Xu are employee of Anhui Rank Artificial Intelligent Technology Co., Ltd. The paper reflects the views of the scientists and not the company.

References

1. Taylor, T.R.; Giles, M.R.; Hathon, L.A.; Diggs, T.N.; Braunsdorf, N.R.; Birbiglia, G.V.; Kittridge, M.G.; Macaulay, C.I.; Espejo, I.S. Sandstone diagenesis and reservoir quality prediction: Models, myths, and reality. *AAPG Bull.* **2010**, *94*, 1093–1132. [CrossRef]
2. Milliken, K. *Late Diagenesis and Mass Transfer in Sandstone Shale Sequences*; Elsevier: Amsterdam, The Netherlands, 2003; Volume 7.
3. Makowitz, A.; Milliken, K.L. Quantification of brittle deformation in burial compaction, Frio and Mount Simon Formation sandstones. *J. Sediment. Res.* **2003**, *73*, 1007–1021. [CrossRef]
4. Makowitz, A.; Lander, R.; Milliken, K. Diagenetic modeling to assess the relative timing of quartz cementation and brittle grain processes during compaction. *AAPG Bull.* **2006**, *90*, 873–885. [CrossRef]
5. Dutton, S.P.; Loucks, R.G.; Day-Stirrat, R.J. Impact of regional variation in detrital mineral composition on reservoir quality in deep to ultradeep lower Miocene sandstones, western Gulf of Mexico. *Mar. Pet. Geol.* **2012**, *35*, 139–153. [CrossRef]
6. Dutton, S.P.; Loucks, R.G. Diagenetic controls on evolution of porosity and permeability in lower Tertiary Wilcox sandstones from shallow to ultradeep (200–6700 m) burial, Gulf of Mexico Basin, USA. *Mar. Pet. Geol.* **2010**, *27*, 69–81. [CrossRef]
7. McRae, L.; Holtz, M.; Hentz, T. *Strategies for Reservoir Characterization and Identification of Incremental Recovery Opportunities in Mature Reservoirs in Frio Fluvial-Deltaic Sandstones, South Texas: An Example from Rincon Field, Starr County*; Topical report; U.S. Department of Energy: Washington, DC, USA, 1995. Available online: https://www.osti.gov/servlets/purl/123238 (accessed on 17 March 2024).
8. Saxena, N.; Day-Stirrat, R.J.; Hows, A.; Hofmann, R. Application of deep learning for semantic segmentation of sandstone thin sections. *Comput. Geosci.* **2021**, *152*, 104778. [CrossRef]
9. Perez, C.A.; Estévez, P.A.; Vera, P.A.; Castillo, L.E.; Aravena, C.M.; Schulz, D.A.; Medina, L.E. Ore grade estimation by feature selection and voting using boundary detection in digital image analysis. *Int. J. Miner. Process.* **2011**, *101*, 28–36. [CrossRef]
10. Patel, A.K.; Chatterjee, S.; Gorai, A.K. Development of a machine vision system using the support vector machine regression (SVR) algorithm for the online prediction of iron ore grades. *Earth Sci. Inform.* **2019**, *12*, 197–210. [CrossRef]
11. Patel, A.K.; Chatterjee, S.; Gorai, A.K. Development of machine vision-based ore classification model using support vector machine (SVM) algorithm. *Arab. J. Geosci.* **2017**, *10*, 1–16. [CrossRef]
12. Ma, X.-M. A Revised Edge Detection Algorithm Based on Wavelet Transform for Coal Gangue Image. In Proceedings of the 2007 International Conference on Machine Learning and Cybernetics, Hong Kong, China, 19–22 August 2007; pp. 1639–1642.
13. Xu, D.; Chen, X.; Xie, Y.; Yang, C.; Gui, W. Complex networks-based texture extraction and classification method for mineral flotation froth images. *Miner. Eng.* **2015**, *83*, 105–116. [CrossRef]
14. Chatterjee, S.; Bandopadhyay, S.; Machuca, D. Ore grade prediction using a genetic algorithm and clustering based ensemble neural network model. *Math. Geosci.* **2010**, *42*, 309–326. [CrossRef]
15. Andersson, T.; Thurley, M.J.; Carlson, J.E. A machine vision system for estimation of size distributions by weight of limestone particles. *Miner. Eng.* **2012**, *25*, 38–46. [CrossRef]
16. Delbem, I.; Galéry, R.; Brandão, P.; Peres, A. Semi-automated iron ore characterisation based on optical microscope analysis: Quartz/resin classification. *Miner. Eng.* **2015**, *82*, 2–13. [CrossRef]
17. Xiao, D.; Liu, X.; Le, B.T.; Ji, Z.; Sun, X. An ore image segmentation method based on RDU-Net model. *Sensors* **2020**, *20*, 4979. [CrossRef] [PubMed]
18. Liu, Y.; Zhang, Z.; Liu, X.; Wang, L.; Xia, X. Efficient image segmentation based on deep learning for mineral image classification. *Adv. Powder Technol.* **2021**, *32*, 3885–3903. [CrossRef]
19. Baraian, A.; Kellokumpu, V.; Paaso, J.; Koresaar, L.; Kaartinen, J. Computing Particle Size Distribution of Mineral Rocks Using Deep Learning-Based Instance Segmentation. In Proceedings of the 2022 10th European Workshop on Visual Information Processing (EUVIP), Hong Kong, China, 19–22 August 2022; pp. 1–6.
20. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
21. Wu, S.; Wang, Q.; Zeng, Q.; Zhang, Y.; Shao, Y.; Deng, F.; Liu, Y.; Wei, W. Automatic extraction of outcrop cavity based on a multiscale regional convolution neural network. *Comput. Geosci.* **2022**, *160*, 105038. [CrossRef]
22. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
23. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 19–25 June 2021; pp. 13713–13722.

24. Jaderberg, M.; Simonyan, K.; Zisserman, A. Spatial transformer networks. In *Advances in Neural Information Processing Systems, Proceedings of the NIPS 2015, Montreal, Canada, 7–12 December 2015*; Neural Information Processing Systems Foundation, Inc. (NeurIPS): La Jolla, CA, USA, 2015; pp. 1–9.
25. Yu, F.; Koltun, V.; Funkhouser, T. Dilated Residual Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Venice, Italy, 22–29 October 2017; pp. 472–480.
26. Li, X.; Sun, X.; Meng, Y.; Liang, J.; Wu, F.; Li, J. Dice loss for data-imbalanced NLP tasks. *arXiv* **2019**, arXiv:1911.02855.
27. Wu, S.; Li, X.; Wang, X. IoU-aware single-stage object detector for accurate localization. *Image Vis. Comput.* **2020**, *97*, 103911. [CrossRef]
28. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft Coco: Common Objects in Context. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.