



TECHNICAL ARTICLE

An Improved Mineral Image Recognition Method Based on Deep Learning

HUAMING TANG,¹ HONGMING WANG,² LING WANG,^{1,3}
CHONG CAO,¹ YIMIAO NIE,¹ and SHUXIAN LIU¹

1.—School of Mining Engineering, North China University of Science and Technology, Tangshan 063210, China. 2.—School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China. 3.—e-mail: wanglings_@163.com

Identification of rock minerals is one of the fundamental procedures of geology and mineralogy. Computer vision technologies and the theory of deep learning (DL) make intelligent rock mineral identification possible. The polarizing microscope pictures of iron ore as the data source, a composite dataset consisting of transmitted light images and reflected light images was developed in this investigation. Using the Deeplabv3+ network, a targeted mineral identification network model was created based on the DL theory. This model can effectively and automatically extract the deep feature information of ore mineral images under a polarizing microscope, as well as achieve intelligent identification and classification of transparent minerals and non-transparent minerals. The model was then improved by freezing training, enlarging the receptive area, and utilizing FC-CRF. The outcome demonstrated the outstanding performance. The total mineral recognition accuracy reached 97.56%, and the recognition accuracy of certain minerals was up to 99%. The identification result obtained by the improved mineral identification model accurately depicts the mineral species information of the microscope photographs, providing a convenient and trustworthy data source for the development of intelligent mineralogy.

INTRODUCTION

Identification of rocks and minerals is crucial in many disciplines, such as geology, mineral processing, etc. However, traditional mineral identification relies heavily on the experience of the identification staff, and requires a high degree of professional knowledge.¹ It often has high costs and labor, and the recognition results are likely to be different because of subjective elements, like the observer's operating experience and professional level. Although the techniques and tools for mineral identification, such as Raman spectroscopy, x-ray fluorescence spectroscopy, x-ray diffraction, scanning electron microscopy, and more sophisticated tools like MLA mineral analyzers and automated mineralogy systems,² have gradually improved with

the advancement of science and technology, optical microscopy observation also serves as the foundation for mineral characterization and quantification, and continues to hold a prominent position due to its simplicity of method, accessibility of equipment, and the ease of operation.

Major technological breakthroughs have recently been made in the field of artificial intelligence (AI), and the integration of applications in other fields has become a popular research. A deep-learning (DL)-based convolutional neural network image identification system can effectively recognize minerals under a microscope, which can also make up for the limitations of larger errors and the higher thresholds of traditional mineral identification.³ Therefore, the development of a DL-based intelligent mineral identification system, analyzing the mineral structure and composition according to mineral morphological traits, may significantly improve the productivity of researchers by providing a prompt assisting identification. Zhang et al.⁴

(Received December 16, 2022; accepted March 9, 2023;
published online April 6, 2023)

used the Inception-v3 network to extract four transparent mineral image features, and, based on the obtained feature data, adopted a logistic regression (LR), a support vector machine (SVM), a multilayer perceptual neural network (MLP), and other machine-learning methods as the base models to identify the minerals. They demonstrated that multiple model stacking can effectively improve the model performance using comparative experiments. Aligholi et al.⁵ were able to identify 15 non-metallic minerals, including garnet, quartz, and black mica, by recording the color shift of minerals in CIELab color space under single polarization and orthogonal polarization, and calculating their Hausdorff distance with known minerals. Ramil et al.⁶ optimized a back propagation artificial neural network to recognize mineral particles of quartz, potassium feldspar, plagioclase, and black mica in various granite samples. A DL-based image segmentation method was proposed by Liu et al.⁷ to address the adhesion and overlap issues of existing mineral image segmentation methods. Compared with the segmentation performances of various network models, it was found that the model using MobileNet⁸ as the decoder can segment a more complete image of coal particles. Zeng et al.⁹ proposed an EfficientNet-b4 deep neural network-based mineral recognition model, and built a dataset combining optical images of minerals with their Mohs hardness, through which 73 mineral samples were recognized with the highest recognition accuracy of 99.6%.

Most current research on the recognition problem of various mineral images is solely to obtain one type of data source images, i.e., mineral images under reflected light or mineral images under transmitted light. However, the existing recognition models are built with too much emphasis on the performance of the model, and ignore its application to practical situations. In mineral and rock identification, the mineral identification of ore samples cannot be accomplished by observing only under transmitted light or only under reflected light. In particular, there are both metallic and non-metallic minerals in metallic ore, and most of the time we need to study the mineral characteristics under both transmitted light and reflected light.

In this research, we present a comprehensive species mineral identification model both for transparent and for opaque minerals of metallic ore, based on the Deeplabv3+ model, to solve the limitations of the previous research. By using freeze training to reduce the training time, the model has been efficiently constructed. Additionally, by optimization of the astrous spatial pyramid pooling (ASPP) perceptual field, the range of picture feature extraction is widened. Therefore, the identification accuracy of the model is greatly improved. In order to better detail the recognition outcomes, a full-connected conditional random field has also been added in the model.

DEEP LEARNING

Hinton et al.¹⁰ initially suggested a class of machine-learning algorithms that may be grouped together as DL.¹¹ It is an algorithm that seeks to accomplish high-level abstraction of data utilizing several processing layers, including complicated structures or consisting of numerous nonlinear transformations. It is frequently used to process multi-layer nonlinear transformations of very complex data,¹² when dealing with complex nonlinear separable problems in real-world applications, which are challenging to analyze both qualitatively and quantitatively through mathematical modeling, optimization, etc. It is said to possess nonlinearity. These problems are typically approximated by using complex fitting functions. It is clear that no mathematical formula can be used to create the link between mineral properties and mineral species, which is a typical nonlinear problem.

As shown in Fig. 1, a deep neural network consists of multiple layers of interconnected nodes, which contain black nodes and white nodes. The black nodes, called the visible layers, form the input and output layers. The input layer is the section of the DL model that receives data, such as the images and videos of the mineral processing, or data originating from other production equipment. The output layer is the section of the DL model that obtains the results of ultimate identification, categorization, or description. The white nodes, located between the input and output layers and called the hidden layers, are used to discern and extract the patterns and features of the data. The hidden layer is the section of the network with the most layers, of which each node employs ever more sophisticated DL algorithms to discern and extract the patterns and features of the data. Compared with machine-learning algorithm models of only one-layer, DL with more than three hidden layers can offer a better fit to the complicated models.¹³

DL architectures have been applied in the areas of computer vision, machine vision, medical image analysis, material testing, and board game programs, etc.¹⁴ In some circumstances, the achievements of DL are even superior to that of humans. DL has also been studied primarily in the area of mineral processing, and some good research results have been achieved in mineral identification,¹⁵ flotation,¹⁶ grinding,¹⁷ picking, and so on.¹⁸

DATASETS

Data Collection

The training of deep neural networks relies on a large amount of training data. The quality and quantity of images in the training set have a great effect on the performance and generalization ability of the neural network.¹⁹ Magnetite ore from the Shui Chang Iron Mine of Tangshan City, Hebei Province, China, was used to gather the training

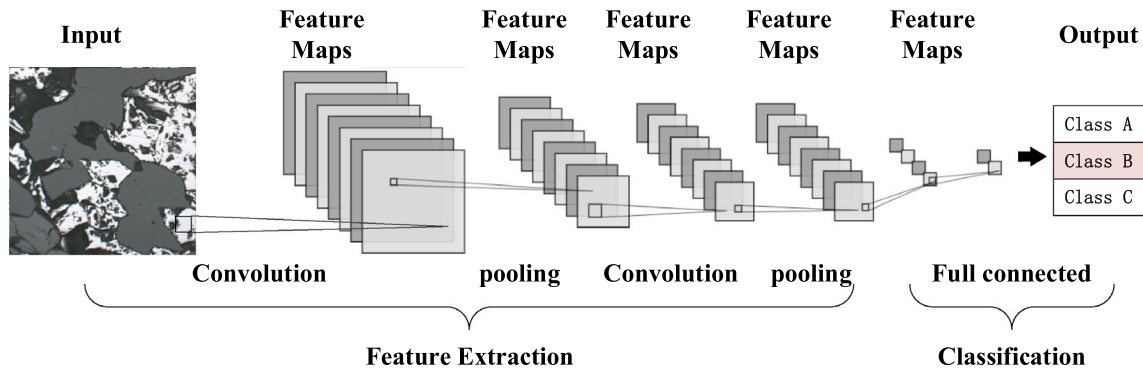


Fig. 1. Deep neural network architecture.

data for the experiment. Magnetite and pyrite are the predominant metallic minerals in the ore, while quartz, pyroxene, amphibole, biotite, zircon, and carbonate minerals are the predominant non-metallic minerals. With a polarizing microscope, a polished thin section and a polished section of magnetite ore were utilized to capture images under transmitted light and reflected light for the experiment, respectively.

To prepare the polished thin section, the selected ore block must be cut to less than 0.3 mm, then glued to the slide and polished.¹⁹ The intricacy of the polished thin section images of ore minerals is considerably different from that of images of other things, like landscapes, people, etc. In addition to the properties of the mineral itself, the polished thin section image also comprises a number of sophisticated internal structural features, such as pore size, pore connectivity, pore type, particle distribution and arrangement form, particle size, debris, permeability, and so on. The polished thin sections are mainly used to observe the transparent minerals identified by the distinctive color, crystal shape, cleavage, angles, etc.²⁰

The polished section was polished on one side of the ore block. The polished surface should be as smooth as a mirror, ideally devoid of small pits, fine fractures, or scratches. Frequently, the polished sections are made in order to observe opaque minerals for their structure, mineral composition, and grain size.²¹ Compared with the polished thin section, the polished section is directly observed under the optical microscope, so there is no interference with the non-mineral substances such as air bubbles and resins in the polished section. In view of the influence of interference color and illumination angle, the images were collected under different angles with the rotation of the worktable.

Collection of the data follows these principle:

- (i) 10 times eyepiece and objective lens are used to observation.
- (ii) Adjust the sample position so that as many minerals as possible are visible in the field of view.

- (iii) Under the crossed polarizer, take photographs of the field of view using transmitted or reflected light, and collect the image data.
- (iv) When taking the images of the polished thin sections under transmitted light, the photographs are collected every 45°, and a total of 5 photographs are collected with 180° rotation of the worktable in the field of view.

Figure 2 shows an example of the gathered minerals. For the polished section, 1 photograph was captured each field of view under the reflected light (Fig. 2A), and a total of 100 photographs were captured. For the polished thin section, 5 photographs are captured with different angles in each field of view under the transmitted light (Fig. 2C, D, E, and F), and also 1 photograph under reflected light (Fig. 2B), and a total of 300 photographs were captured.

The photographs were manually annotated, as shown in Fig. 3. For the transmitted light photographs, all the mineral images have been annotated; while only the opaque mineral images (magnetite and pyrite) have been labeled for the reflected light photographs. The matching information for each color label is presented in Table 1.

Data Augmentation

It is not sufficient to train a deep neural network by 400 data photographs, as gathered in section 2.1, so data augmentation methods are employed to gain additional training data. For a normal dataset, there are various data augmentation techniques,²² such as random cropping, random inversion, random contrast enhancement, color modification, etc., or a combination of them. Adjusting the brightness, contrast, saturation, and hue can help lower the model's color sensitivity,²³ but the color and reflectivity are the important characteristics in mineral identification, so we have just adopted random cropping and mirror flipping to augment the original dataset in the experiment.

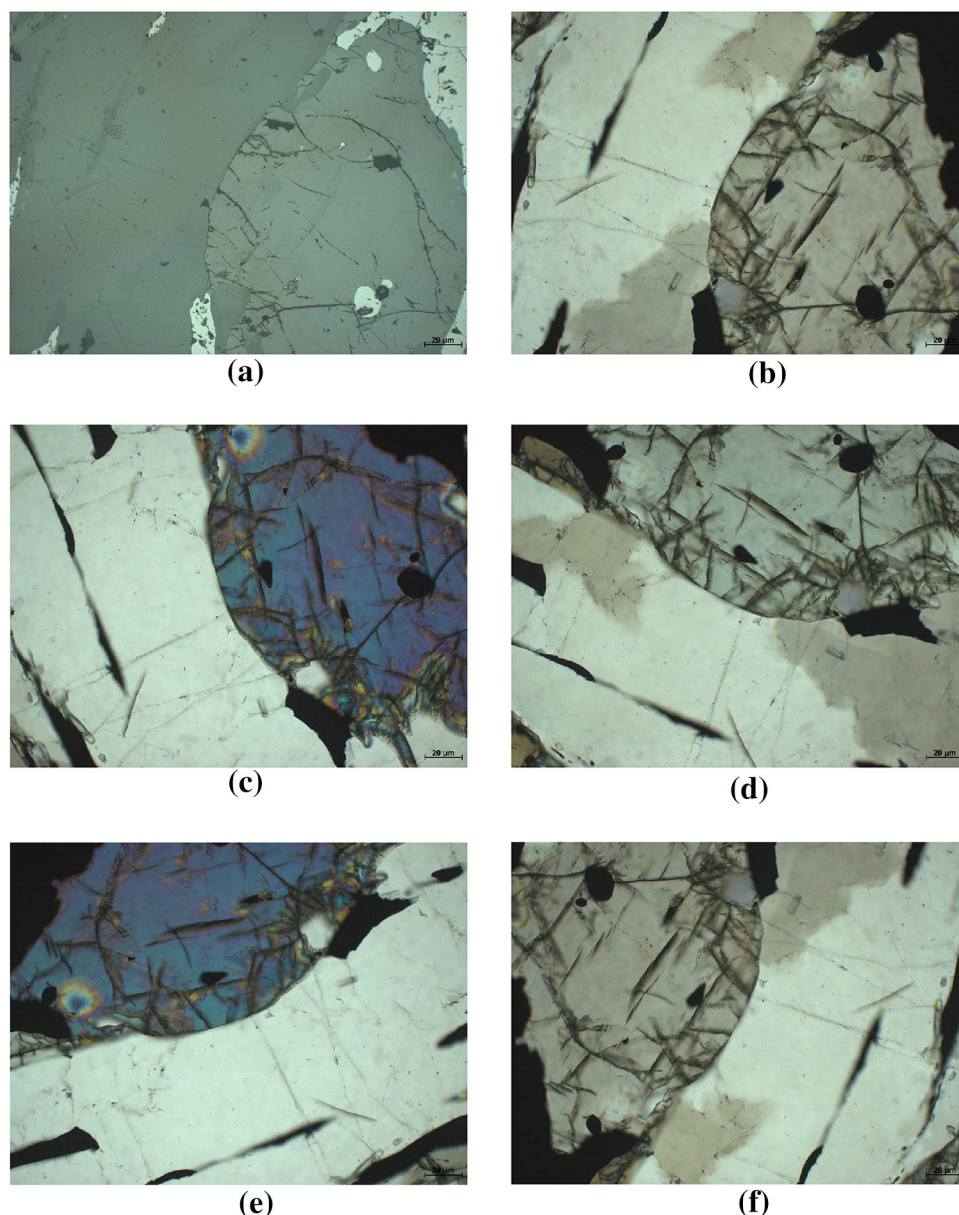


Fig. 2. Original photomosaic of the sample surface: (A) reflected light image of a thin sheet, (B–F) transmitted light images of the same field of view.

Random Cropping

The resolution of the original photograph is 4164×3120 pixels. By randomly setting the initial cropping point a random cropping image with a fixed size of 512×512 pixels is obtained, which makes the minerals appear in various positions. Random cropping can lower the model's sensitivity to the target position.²⁴ If the cropping result exceeds the edge of the original photograph, the extra area is filled with black and set as the background image, which has no effect on the mineral image.

Mirror Rotation

According to crystal optics, images obtained from the rotation by left and right, or up and down, are still reasonable, which is equivalent to resetting the position of the polished section or the thin section on the worktable. Therefore, the random cropping images were also rotated horizontally, vertically, and horizontally and vertically to augment the experimental data.

Through the foregoing data augmentation, the initial 400 photographs were enlarged to 9000, which is sufficient for training a deep neural

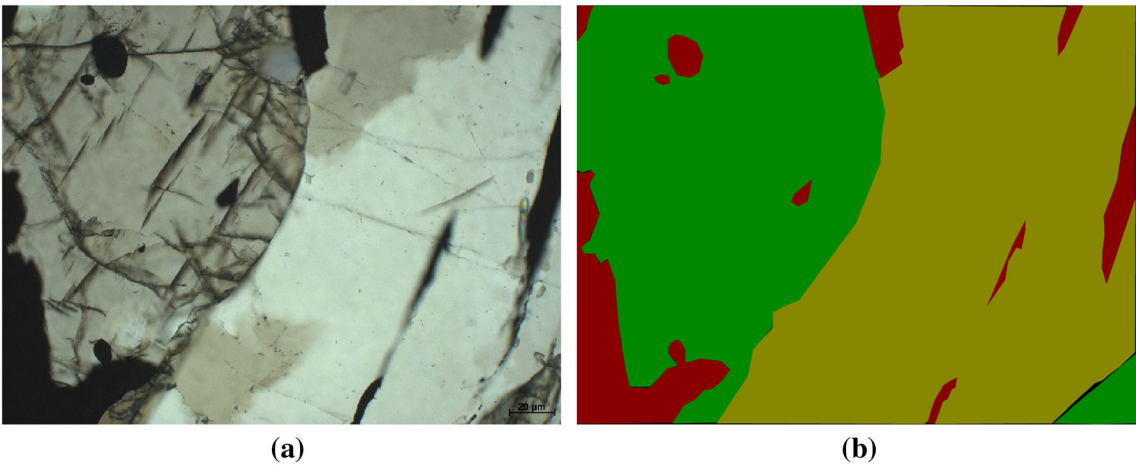









Fig. 3. (a) Original image, (b) labeled image (Color figure online).

Table I. Correspondence of mineral and color (Color table online)		
Mineral species	Corresponding color code	Annotated color
Magnetite	#800080	
Pyrite	#008080	
Quartz	#808000	
Pyroxene	#008000	
Amphibole	#000080	
Opaque minerals (transmitted light image)	#800,000	
Background	#000000	

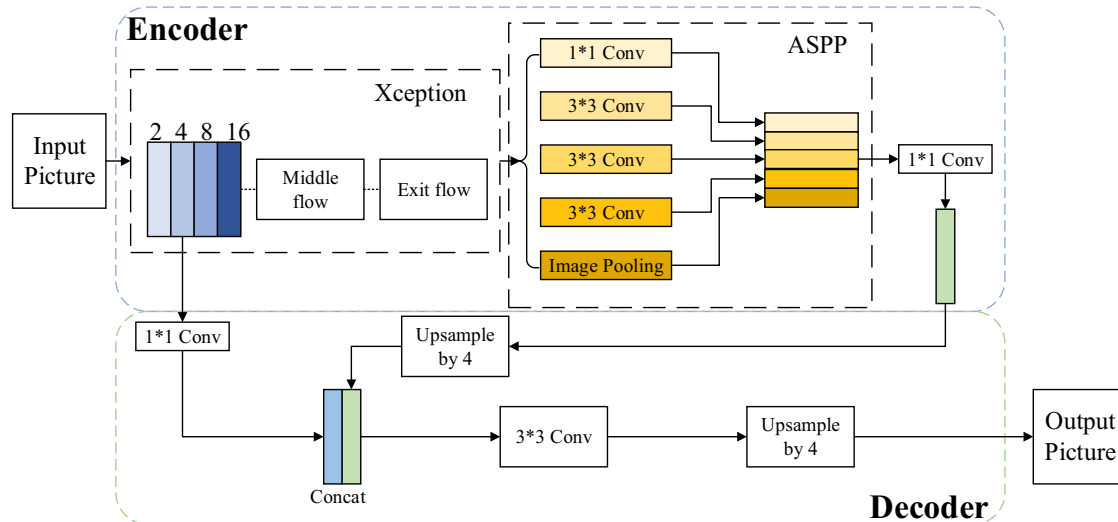


Fig. 4. Deeplabv3+ network structure.

network. During the experiment, 80% of the data (7200 images) were set as training data and 20% of the data (1800 images) were set as validation data.

The Deeplabv3+ Model

DeepLabv3+ is a type of semantic segmentation algorithm that employs a spatial pyramid module and an encoder-decoder architecture. It can augment contextual information by combining operations at various resolutions and gradually recover image boundary information via an encoder-decoder architecture. The network layout of DeepLabv3+ is depicted in Fig. 4. The encoder module uses ResNet²⁵ as the backbone network for feature extraction, subsequently connecting with the (ASPP) module which extracts multi-scale information from the fused pictures.²⁶ The decoder module combines the feature maps of the encoder module with the intermediate information of ResNet. The feature map of the encoder module offers semantic information, down-sampling in the middle of ResNet provides detailed information, and up-sampling yields semantic segmentation.

MODIFIED DEEPLABV3+

On the basis of DeeplabV3+, the model was improved by the following three aspects. First, the frozen training module was installed to reduce the training resource consumption of the mode and to increase the training efficiency. Second, the structure of the ASPP was improved by fusing the receptive fields to enhance the information utilization. Then, a fully connected conditional random field (FC-CRF) module was introduced at the end of the model to reinforce the correlation among the pixels and promote the effect of edge processing. Figure 5 depicts the network structure of the improved DeeplabV3+ model.

The loss function for stochastic gradient descent²⁷ model training is cross-entropy. The number of loops has been set to 100, the starting learning rate has been set to 0.0005, and the early stop mechanism has been configured, by setting the batch size to eight based on the current device performance. All models were trained and evaluated using GPU computing resources (NVIDIA GeForce RTX 2060).

Freeze Training

For the Deeplabv3+ model, the characteristics extracted by the backbone network are universal, so the weights of the backbone network can be frozen, which is conducive to enhancing the training speed of the model and ensuring the consistency of the weights.

The backbone network is divided into the freeze phase and the thaw phase in freeze training. During the freeze phase, the system consumes fewer resources because the network for feature extraction does not change. During the thaw phase, the network for feature extraction is unfrozen, and its parameters will vary, based on the number of learning iterations, which will consume a significant amount of memory resources. Consequently, freeze training minimizes the experimental environment needed for the model training while simultaneously decreasing the model training.

For the improved Deeplabv3+ model, the first 50 learning training sessions were carried out by the freeze phase using the original general feature network to limit the consumption of the resources available. The next 50 learning training sessions were carried out by the thaw phase which increases the input of the training resources so that the full mineral picture characteristics was extracted, completing the total learning and training of the mineral identification network model.

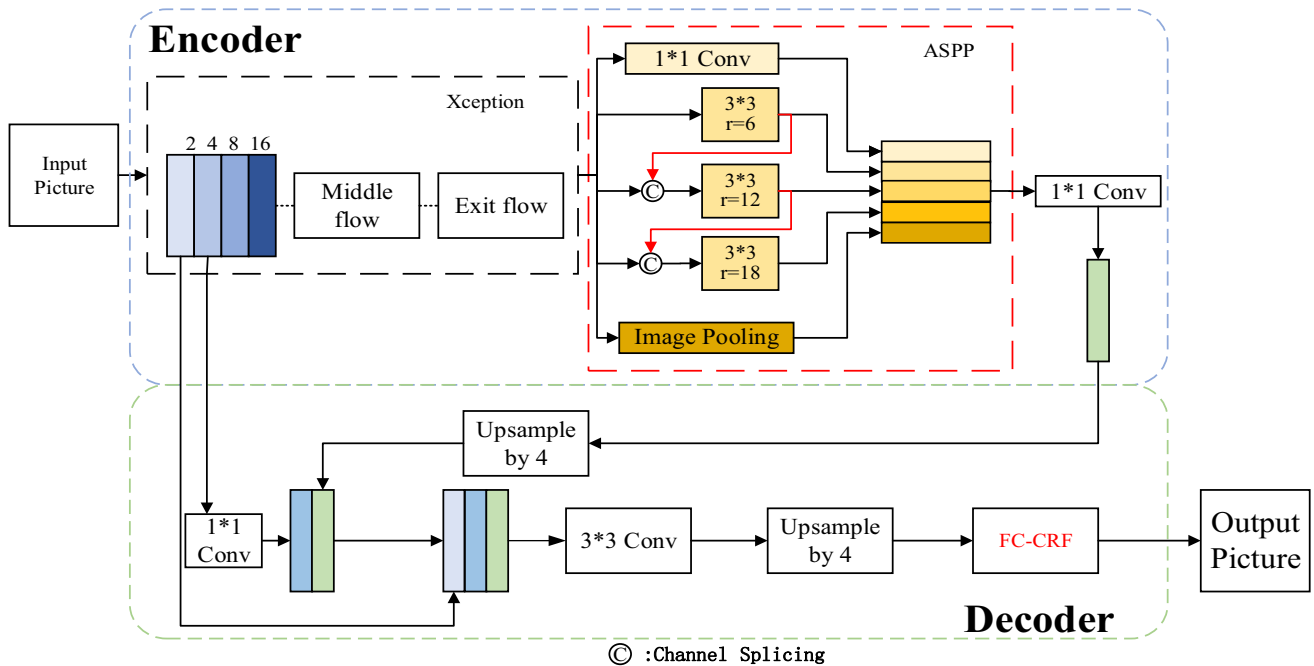


Fig. 5. Improved Deeplabv3+ network structure.

Widen the Scope of Receptivity

The receptive field of the ASPP structure of Deeplabv3+ model was widened via dilated convolution, which originated by adding the null value filling on the foundation of ordinary convolution. However, when the dilation rate grows too quickly, the number of faulty sample points in the dilated convolution will rapidly increase, which will affect the information gathering of the picture too much to extract the correct feature information and so obtain an expected network model. Therefore, the structure of ASPP was optimized. The original feature map and the feature map processed by the convolutional layer were spliced and first given to the subsequent convolutional layer. Then, the sampling point distribution of the original feature map of the convolution layer of the picture given was adjusted, and the receptive field size and the number of elements involved in the valid operations were set up. Finally, the subsequent convolution layer' sampling point distribution of the feature map of the previous processed convolution layer was adjusted.

By increasing the association between the diverse receptive field information, this method distinguishes the category attribute of a single pixel in a larger area, and substantially reduces the loss of the information. Compared with the original ASPP structure, the parameters of the optimized ASPP structure are effectively increased, the receptive field is expanded in the simpler network complexity and the information utilization rate is enhanced.

Fully Connected Conditional Random Field

To improve the accuracy of the recognition results, the FC-CRF was introduced in the model. FC-CRF techniques perform well on semantic segmentation issues.²⁸ Zheng et al.²⁹ defined FC-CRF as the structure of a recurrent neural network and coupled it with a fully convolutional neural network (FCN) to simultaneously train FCN and CRF, which greatly improved the segmentation accuracy. The CRF energy function includes a data term and a smoothing term. The data term is determined by the likelihood that each pixel belonging to each class, whereas the smooth term is determined by the difference in gray values and the spatial distance between the pixels. The smoothing term of conventional CRF only analyzes the correlation among the neighboring pixels, but FC-CRF considers the correlation between any two pixels in the image, and thus it can differentiate the borders between distinct categories.³⁰ Conditional random fields with Gaussian pairwise potentials can successfully erase erroneously detected isolated pixels in mineral grain images. For the smooth term, adjacent pixels with identical features may be derived from the same mineral.

Model Evaluations

To better comprehend and explain the recognition impact, the model's outcomes were evaluated using the confusion matrix model provided in Table II. Each row of the confusion matrix represents the labeled real mineral species, whereas each column represents the pixel class determined by the

Table II. Confusion matrix model

Predicted class	Actual class	
	Positive	Negative
Positive	TN	FN
Negative	FP	TP

model.³¹ Using magnetite as an example, True Negative (TN) indicates the number of correctly detected pixels as magnetite, while False Positive (FP) denotes the identification of a mineral as magnetite when, in fact, it is composed of a greater number of pixels of other minerals. False Negative (FN) reflects the number of pixels that are incorrectly recognized as other minerals when they are in fact magnetite. True Positive (TP) is the number of pixels properly identified as other minerals. These statistics show the number of pixels in the image that were recognized using the validation set in the dataset. There are 1800 photos with a resolution of 512×512 pixels.

The loss function and the F1-score value can properly evaluate the semantic segmentation model's performance. The loss function is utilized to evaluate the difference between the model's recognized value and the actual value. The smaller the value of the loss function, the closer the recognition result is to the genuine result. The F1-score³¹ is a statistic used in classification models to test the model's accuracy, allowing for the most effective feature selection. It consists of two metrics: Precision and Recall. Precision is defined as the ratio of successfully identified positive samples to the total number of positive samples, divided by the classifier:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

Recall is the proportion of accurately detected true positive samples (without any negative samples):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

The formula for F1-score calculation is:

$$F_1 - \text{score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

Mean intersection over union (MIoU)³² and mean pixel accuracy (MPA) are useful metrics for evaluating the model's segmentation accuracy. MIoU is computed by:

$$\text{MIoU} = \frac{1}{2} \left[\frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} + \frac{\text{TN}}{\text{TN} + \text{FN} + \text{FP}} \right] \quad (4)$$

MPA can be calculated by:

$$\text{MPA} = \text{sum}(P_i) / \text{Number of Categories} \quad (5)$$

where P_i is the diagonal value divided by the total number of pixels in the column in question.

RESULTS AND ANALYSIS

The normalized confusion matrix of the recognition results is given in Fig. 6. The color of the matrix is related to the accuracy of the mineral recognition; that is, the deeper the color, the higher the accuracy. The "background" in the matrix is the segmentation of the black area during the mineral identification of the transmitted light image, which represents the opaque mineral portion. The number of the diagonal from upper left to lower right denotes the fraction of correct identification, and the other diagonals are the fraction of misrecognition. As can be seen, although there are more incorrectly recognized pixels, less than 90% of the accuracy, of the pyrite identification section, the accuracy of magnetite, pyroxene, quartz, amphibole and background are all higher to 90%. Due to the substantial contrast between the opaque mineral data photos and the transparent mineral data images, the error ratio of mineral identification in the two categories is near to 0. This demonstrates that multi-source datasets may minimize the complexity of model classification and make mineral identification be more feasible.

On the basis of the confusion matrix, variations of the loss function and the F1-score value of the experimental model on the training set are shown in Figs. 7 and 8, respectively. The steady stability of the F1-score and the loss values of the two models with increasing training times demonstrate that there is no overfitting in the training of the model. It can also be observed that, as training times grow, the value of the model F1-score increases, while the loss function decreases. The loss function is essentially stable after 45 iterations of the original Deeplabv3+, while the improved model is essentially stable after about 83 but is lower. The F1-score of the original Deeplabv3+ model is more unstable and lower for the F1-score of the improved Deeplabv3+ model. In terms of convergence speed and stability, the improved Deeplabv3+ model has clear benefits over the original Deeplabv3+.

Although the loss function curve demonstrates that the addition of frozen training caused severe fluctuation in the 50th training process, which is the result of the growth in data volume and the modification of parameters, the model is able to lower the rate of information loss and improve the learning effect in the subsequent learning process.

The trained model was tested using the validation data to assess the recognition accuracy of each mineral. The validation data independent of the training data is generated separately and can obtain

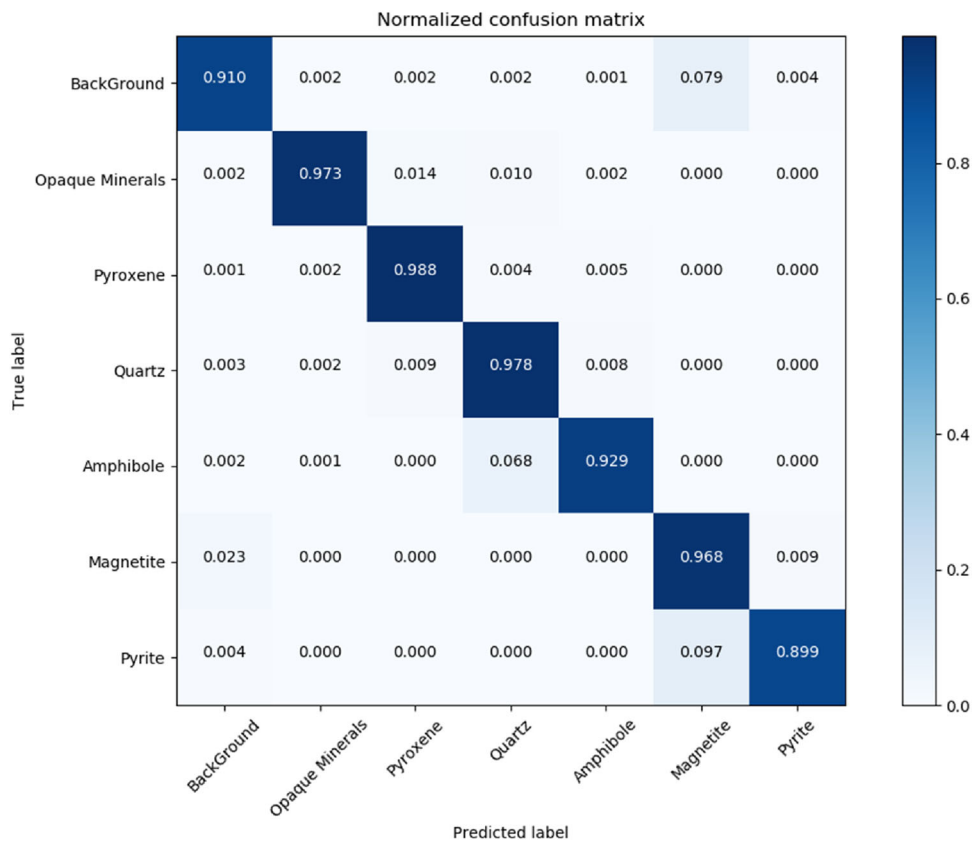


Fig. 6. Normalized confusion matrix.

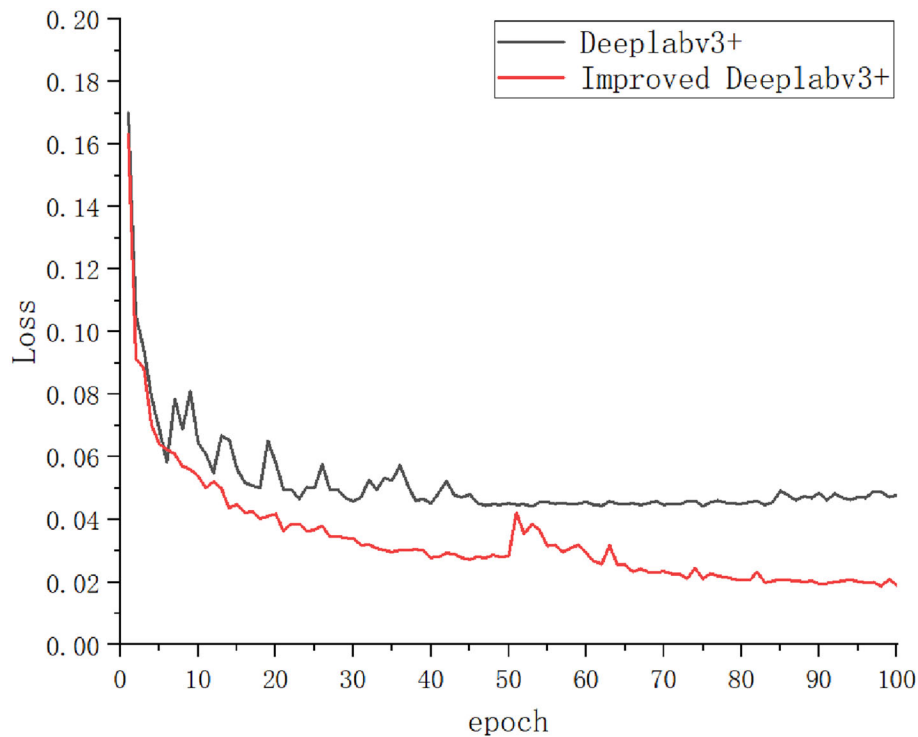


Fig. 7. Loss evolution as a function of the number of training epoch.

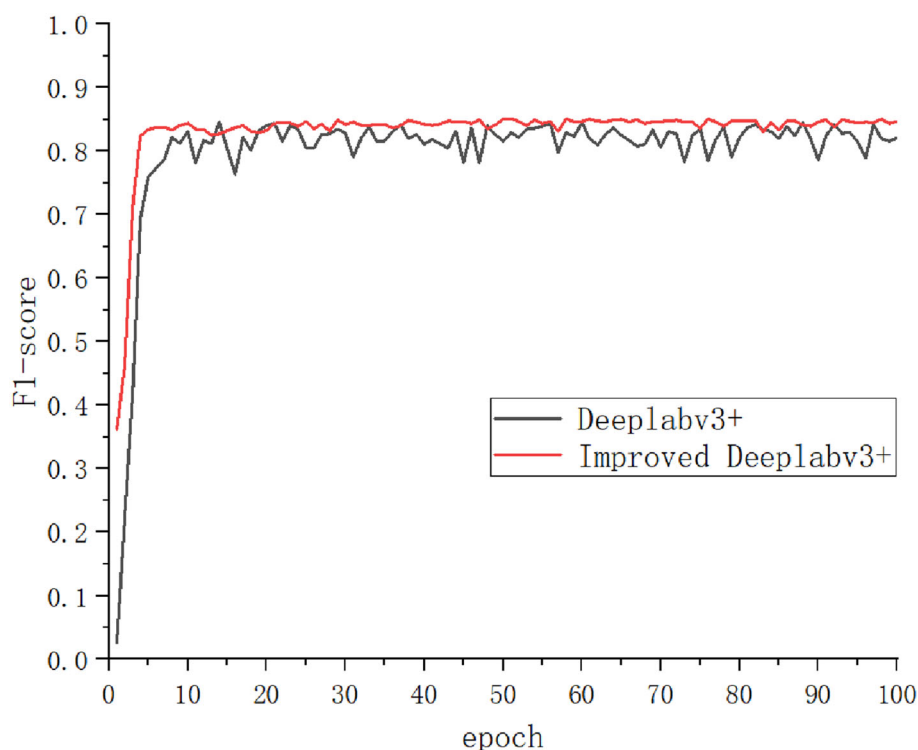


Fig. 8. F1-score evolution as a function of the number of training epoch.

Table III. Segmentation results of Deeplabv3+ model and improved Deeplabv3+ model

Minerals	Deeplabv3+		Improved Deeplabv3+	
	MIoU (%)	MPA (%)	MIoU (%)	MPA (%)
Magnetite	81.74	88.78	94.93	97.33
Pyrite	87.85	92.75	95.41	96.85
Quartz	95.41	96.85	96.91	99.13
Pyroxene	62.58	73.85	96.36	97.61
Amphibole	96.22	98.76	96.22	98.76
Opaque minerals (transmitted light image)	88.57	89.76	94.99	97.04
Background	85.54	90.23	93.2	96.45
Total	85.42	90.14	95.43	97.56
Training time (s)	24,493		21,140	

more precise validation results. The MIoU and accuracy of each mineral identification in the validation data are displayed in Table III, from which it can be seen that the improved model can effectively increase the model's recognition performance in semantic segmentation, in which both the specific details and the overall recognition have the better recognition effect. The total MIoU and MPA increased by 10.01% and 7.42%, respectively. In terms of specific mineral recognition, the MIoU and MPA of pyroxene showed the largest improvement by 33.78% and 23.76%, respectively, while magnetite which increased by 13.19% and 8.55% in MIoU and MPA, respectively. The recognition effects of the other minerals were also enhanced by varying degrees. In addition, it was also showed

that the incorporation of frozen training considerably accelerates network model training, resulting in time savings of 15.86%.

To further illustrate the identification results of the model, the comparison of the original mineral images under optical microscopy and their recognition outcomes using the model are shown in Fig. 9. This shows that the model performs well in the vast majority of instances, and the identification color of the model recognition result was well matched with the color of the manual annotation (Table I). In Fig. 9A, there are partial recognition flaws on the image of pyrite (Fig. 9A and B), indicated by the yellow arrow). Additionally, due to the local fragmentation and various color changes, the partial amphibole was incorrectly identified (Fig. 9C and

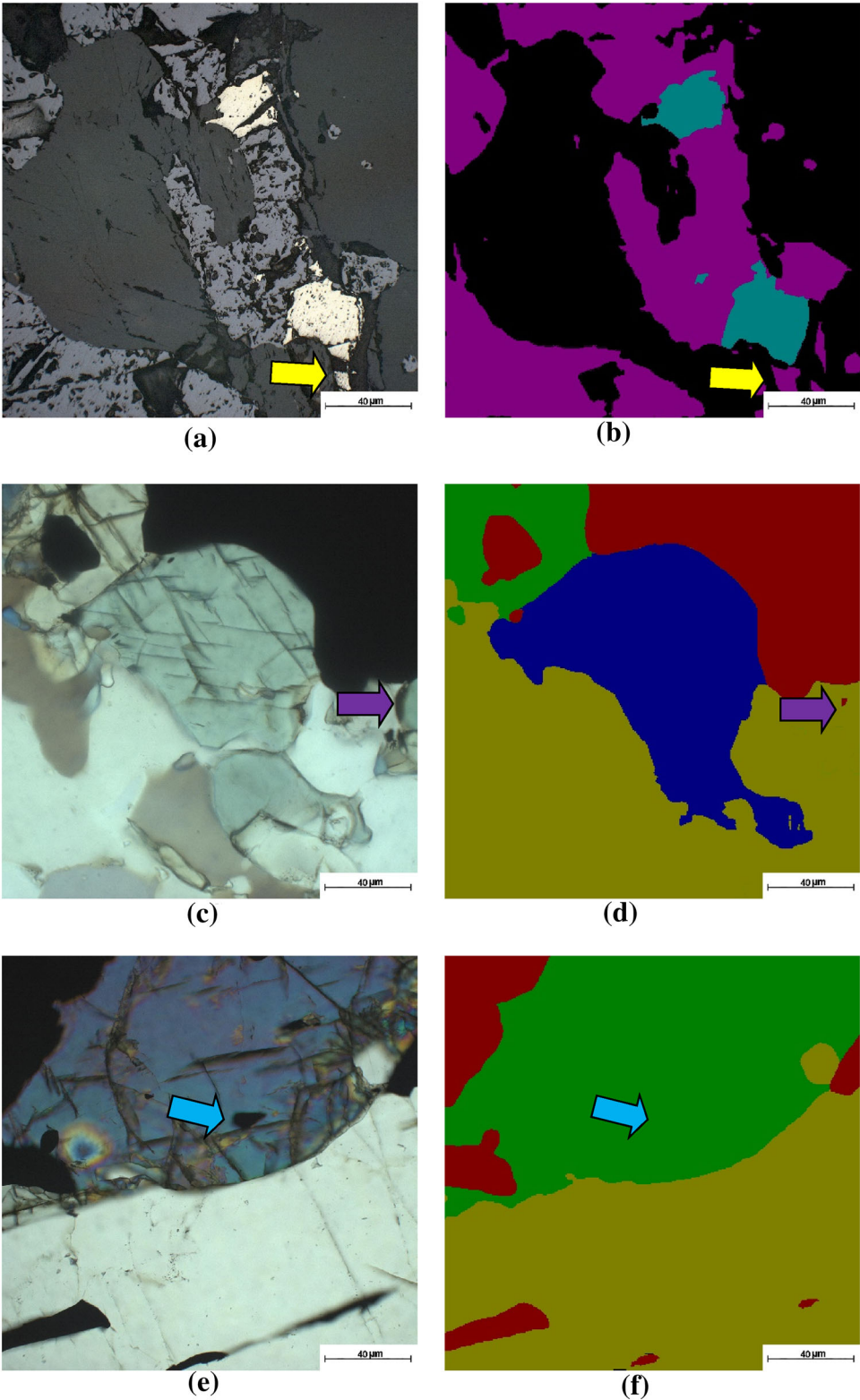


Fig. 9. Recognition results. (A), (C), and (E) are the original images, (B), (D), and (F) are the recognition results. Problematic identification results are marked with 0 (Color figure online).

D), indicated by purple arrows). Evidently, color information continues to play a significant role in mineral identification. In addition, quartz identification performance is satisfied (Fig. 9E and F), and only small opaque mineral pictures embedded in the ore were recognized incorrectly (blue arrows).

CONCLUSION AND PROSPECTS

There has been little comprehensive identification of transmitted light images and reflected light images of minerals in the research of intelligent mineral identification; in practice, however, it is frequently encountered. A composite dataset of photographs of transmitted light and reflected light under a polarizing microscope were built in the investigation. It was found that the Deeplabv3+ model can be used to comprehensively detect both transmitted light and reflected light photographs based on deep learning. An improved Deeplabv3+ model was proposed by stopping training, enlarging the receptive field, and implementing a fully connected conditional random field, by which the training time was decreased by 15.86% and the total accuracy was enhanced by 7.42%. Maximum accuracy in classification and recognition can reach 99.13%. In the identification findings, the same mineral is denoted by the same solid color block, and the computation of mineral particles and content can be completed using straightforward image-processing technology, which provides an easy and trustworthy data source for the subsequent process of mineralogy development. The application of this paradigm in various disciplines has yielded more than 1000 labels, and adequate space has been allocated for the future expansion of multi-source datasets.

The development of a training dataset is essential for the intelligent recognition of mineral images using neural network. Numerous elements, including geology, mines, history, etc., influence the appearance of minerals. The fact that the same mineral can take on a variety of shapes makes it extremely challenging to compile a consistent database of mineral photos. Massive face photos currently available constitute the foundation for the rapid growth of face recognition. For the advancement of intelligent mineral identification, the construction of picture databases of minerals under transmitted and reflected light is crucial. Existing digital rock gateways have already begun this process, but there are many more non-metallic rock varieties to be added. Using large data technologies and other relevant research methods, it is possible to build standard dataset standards and manufacturing techniques to add more mineral species to the database.

ACKNOWLEDGEMENTS

This work was supported by the Natural Science Foundation of Hebei Province (E2022209119 &

D2020209017), the National Natural Science Foundation of China (42002098 & 52004091) and the Central Government Guides Local Science and Technology Development Fund Project (226Z4103G).

CONFLICT OF INTEREST

On behalf of all authors, the corresponding author states that there is no conflict of interest.

REFERENCES

1. K. Misra, *Understanding Mineral Deposits* (Springer Science & Business Media, Amsterdam, 2000), pp3–4.
2. W. Lou, D. Zhang, and R.C. Bayless, *Appl. Geochem.* 122, 104727 (2020).
3. Y.-C. Wu and J.-W. Feng, *Wireless Pers. Commun.* 102, 1645–1656 (2018).
4. Y. Zhang, M. Li, S. Han, Q. Ren, and J. Shi, *Sensors* 19, 3914 (2019).
5. S. Aligholi, G.R. Lashkaripour, R. Khajavi, and M. Razmara, *Pattern Recogn.* 65, 164–174 (2017).
6. A. Ramil, A. López, J. Pozo-Antonio, and T. Rivas, *Measurement* 117, 90–95 (2018).
7. Y. Liu, Z. Zhang, X. Liu, L. Wang, and X. Xia, *Adv. Powder Technol.* 32, 3885–3903 (2021).
8. A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto and H. Adam, arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861) (2017).
9. X. Zeng, Y. Xiao, X. Ji, and G. Wang, *Minerals* 11, 506 (2021).
10. G.E. Hinton and R.R. Salakhutdinov, *Science* 313, 504–507 (2006).
11. Y. LeCun, Y. Bengio, and G. Hinton, *Nature* 521, 436–444 (2015).
12. C.M. Bastuscheck, In *CVPR*, (1989), pp 262–268.
13. G. Montavon, W. Samek, and K.-R. Müller, *Digital Signal Process.* 73, 1–15 (2018).
14. D. Ciregan, U. Meier, and J. Schmidhuber, In *2012 IEEE conference on computer vision and pattern recognition* (IEEE: 2012), pp 3642–3649.
15. A. Khvostikov, D. Korshunov, A. Krylov, and M. Boguslavskiy, *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 44, 113–118 (2021).
16. H. Zhang, Z. Tang, Y. Xie, X. Gao, Q. Chen, and W. Gui, *Miner. Eng.* 160, 106677 (2021).
17. A. Zakamaldin and A.A. Shilin, In *IOP Conference Series: Materials Science and Engineering* (IOP Publishing: 2020), p. 012010.
18. A.A.C. Pereira, C.A.C. Olivera, A.G. Merma, R.R. Hacha, B.F. dos Santos, and M.L. Torem, *Miner. Eng.* 169, 106983 (2021).
19. M.P. Filippo, O.F.M. Gomes, G.A.O.P. da Costa, and G.L.A. Mota, *Miner. Eng.* 170, 107007 (2021).
20. D.G. Tang, K.L. Milliken, and K.T. Spikes, *Mar. Pet. Geol.* 120, 104518 (2020).
21. G. Latif, K. Bouchard, J. Maitre, A. Back, and L.P. Bédard, *Minerals* 12, 455 (2022).
22. P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, *Procedia Comput. Sci.* 199, 1066–1073 (2022).
23. G. Xu, Y. Zhang, Q. Zhang, G. Lin, Z. Wang, Y. Jia, and J. Wang, *Fire Saf. J.* 105, 277–285 (2019).
24. B.I.M. Shakkak and S.A.K. Al Mazruii, *Appl. Comput. J.* 2, 193–212 <https://doi.org/10.52098/acj.202247> (2022).
25. K. He, X. Zhang, S. Ren and J. Sun, In *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), pp 770–778.
26. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff and H. Adam, In *Proceedings of the European conference on computer vision (ECCV)* (2018), pp 801–818.

27. G. Andrew and J. Gao, In *Proceedings of the 24th international conference on Machine learning* (2007), pp 33–40.
28. A. Arnab, S. Zheng, S. Jayasumana, B. Romera-Paredes, M. Larsson, A. Kirillov, B. Savchynskyy, C. Rother, F. Kahl, and P.H. Torr, *IEEE Signal Process. Mag.* 35, 37–52 (2018).
29. S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P.H. Torr, In *Proceedings of the IEEE international conference on computer vision* (2015), pp 7–13.
30. M. Ovsjanikov, J. Sun, and L. Guibas, In *Computer graphics forum* (Wiley Online Library: 2008), pp 1341–1348.
31. Y.-W. Chen and C.-J. Lin, *Feature Extr. Found. Appl.* 207, 315–324 (2006).
32. H. Torbati-Sarraf, S. Niverty, R. Singh, D. Barboza, V. De Andrade, P. Turaga, and N. Chawla, *JOM* 73, 2173–2184 (2021).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.