Rock Mechanics and Rock Engineering
https://doi.org/10.1007/s00603-022-02907-7

ORIGINAL PAPER

Influence of Neural Network Structure on Rock Intelligent Classification Based on Structural and Tectonic Features of Rocks

Mingyang Wang[1] · Enzhi Wang[1] · Xiaoli Liu[1] · Zheng Wang[1] · Congcong Wang[1]

Received: 22 September 2021 / Accepted: 24 April 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract

Fast and accurate intelligent classification of rocks is the most important prerequisite to ensure the unmanned construction of geological projects. In this paper, a dataset containing 660 images is established, classified based on the structure and the fabric characteristics of rocks, and investigate the recognition accuracy of different structural and tectonic features. Moreover, the accuracy, training time, and robustness of different neural network architectures on a multi-scale image set are compared by transfer learning. The results show that InceptionResnetV2 has the best robustness and high accuracy, which can reach 74.24%, 74.24%, and 72.73% on the three image sets respectively; and it is found that the Inception module significantly improves the recognition accuracy and the Fire module reduces the learning time by a factor of 5–10. Finally, MobileNetV2 was tested and it obtained the highest accuracy of 81.82%. In addition, MobileNetV2 has a strong ability to recognize typical rock structures such as bedding, columnar joints, and leaf veins, and can achieve a recognition accuracy of 95.83%.

Highlights

- A data set based on the rock structure and fabric characteristics is built as the basis for lithology identification.
- Based on transfer learning, structural and tectonic features of rocks can be recognized with an accuracy of 95.83%.
- InceptionResnetV2 achieved the highest accuracy for recognition on a mixed-scale dataset, at 74.24%.
- MobileNetV2 has powerful capabilities in identifying lithological structural features.

Keywords Transfer learning · Lithology recognition · MobileNetV2 · Neural network structure · Structural and tectonic features

List of Symbols

e	Napier constant
I	Input layer matrix
W	Neural network weight matrix
o	Size of the image after the pooling layer
x_i	Input value corresponding to the i-th neural network neurons
K	Pooling kernel size
S	String of the pooling operation
P	Padding of the pooling operation
C	Total number of input vectors entering SoftMax layer

✉ Xiaoli Liu
xiaoli.liu@tsinghua.edu.cn

[1] State Key Laboratory of Hydro-Science and Engineering, Tsinghua University, Beijing 100084, China

1 Introduction

Lithologic recognition and classification are very important tasks in the field of geology. Whether it is laboratory rock sample analysis or field geological exploration (Feng et al. 2019; Li et al. 2020a; Liu et al. 2016), both need highly accurate rock lithology classification. The traditional laboratory lithologic identification and classification work mainly relies on the human subjective judgment of hand scale rock samples; then, the rock samples are classified and

Published online: 21 June 2022

named according to their mineral composition and proportion (Li et al. 2020b). Field geological exploration methods mainly include: exploration methods based on field outcrops (Amosu and Sun 2017; Yan et al. 2020), geophysical exploration methods (seismic wave method, electromagnetic method, high-density electrical method, magnetic resonance, etc.) (Han et al. 2020; Tchokpon et al. 2020), and using remote sensing data for regional geological explorations (Kereszturi 2011; Tchokpon et al. 2020). Traditional lithologic recognition methods are time-consuming, require substantial expertise, and are prone to errors, which makes them dangerous and expensive for engineering construction.

With the rapid development of artificial intelligence, methods such as intelligent image recognition, classification, and regression prediction have become important elements in the integration of statistics and other disciplines (Deepnarain et al. 2020). Aiming at the problem of intelligent image recognition and classification, the convolution neural network (CNN) has been proposed. It has been tremendously successful in image processing, pattern recognition, and object detection systems; it also greatly improves the accuracy of image recognition (Kim et al. 2021). As a result, CNNs have become the dominant machine learning approach for visual object recognition (LeCun et al. 1989).

At present, a large number of studies have been carried out on intelligent classification, recognition, and information extraction of geological body and rock features(Bahmani et al. 2020). Baker et al. (2008) present the intelligent detection of anisotropic features on rock surface based on rock surface roughness. The seismic facies classification is presented in (Dramsch and Lüthje 2018; Duan et al. 2019; Zhao 2018). Puzyrev and Elders (2020) use a deep convolutional autoencoder to propose an unsupervised seismic facies classification method. Dunlop (2006) uses albedo, color, texture, and shape properties to characterize rocks of natural terrain. Feng et al. (2019) establish an AlexNet twin neural network to identify rock lithology based on fresh rock surface image and AlexNet neural network structure. Li et al. (2020a) establish a support vector machine (SVM) regression model of waveform and strength, and combine it with deep CNNs. Rock lithology classification based on coupled rock images and hammering audios has been realized.

The intelligent recognition and classification of rock lithology are based on various aspects: physical parameters and mechanical strength characteristics of rocks (Fan et al. 2012; Guo et al. 2010; Wang et al. 2018b); image optical characteristics of rock surfaces (Lepisto et al. 2003; Thompson et al. 2005a, b; Castafio et al. 2003; Xu et al. 2021), texture characteristics of rocks(Julesz 1962, 1981), geometric features of rock image information (color space combination relations, shape edges, etc.) (Gotlieb and Kreyszig 1990; Leung and Malik 2001; Tuceryan and Jain 1998; Varma and Zisserman 2005), and microstructure and mineral composition characteristics of rock slices (Galdames et al. 2019; Iwaszenko and Smoliński 2021; Purswani et al. 2020; Valentín et al. 2019). Although there is a large number of classification research articles, the intelligent classification of rocks remains an open problem. At present, the accuracy of mineral recognition software applied to the market for fresh mineral identification reaches 80–90%, while the recognition accuracy for rock is only 40–50%. When the rock surface is seriously weathered, the fresh surface is not obvious, the holes are filled, or the scratch marks are complex, it is difficult to accurately extract mineral information from rock images. As a result, the information cannot be accurately extracted for rock lithology recognition and classification.

More importantly, the ability to extract information from hand scale rock images cannot meet the rigorous recognition demands of field scale rock images for engineering projects. In recent years, the application of tunnel building machine and intelligent drilling systems has gradually replaced the traditional manual construction method as the main construction method of tunnel engineering and underground engineering (Jing et al. 2021; Sehsah et al. 2017). The rock image information collected in the process of intelligent excavation is a mix of field scale and hand scale. The scales used correspond to different attribute features; correspondingly, complex neural networks trained by large-scale data sets also show different recognition abilities for different attribute feature data sets. According to such characteristics, Zhang et al. (2018) used machine learning, CNN, and transfer learning methods to verify the robustness of different models for lithologic image sets. The results show that transfer learning based on InceptionV3 can extract features of small geological structure data effectively, and the Inception module has a strong ability to extract and identify the phyllite and breccia porphyry structures. In order to improve the speed of transfer learning and recognition, MobileNet and MobileNetV2 were used to study the features of granite, gneiss, marble, and sandstone(Wang et al. 2018a; Xia et al. 2021). However, the rock image sets mentioned above include multiple scales (micro, hand, and field). Without analyzing the sensitivity of different network structures to different scale features after transfer learning, the establishment of new neural network structures (suitable for more specialized classifications) is impeded.

Most of the existing rock identification methods are based on mineral composition and physical attribute information. For engineering practice, most of the lithologic recognition methods for geological engineers are mainly based on the structure and fabric characteristics of rock samples or rock masses.

This paper builds an image dataset from structural and tectonic classification methods of rock masses. In Sect. 1, the shortcomings of current lithology recognition research

are presented, and in Sect. 2, transfer learning methods and their advantages in engineering lithology recognition are presented. Section 3 shows the results of each network after transfer learning is performed in this study. These results include calculation and analysis of recognition results and feature extraction capabilities for hand-sampled scale, field scale and mixed scale image sets. Section 4 investigates the effect of different network structures on learning time and training accuracy. Section 5 shows the test results after fine-tuning the MobileNetV2 network in comparison with other networks.

2 Principle of Mixed Scale Lithology Recognition

2.1 Deep Convolutional Neural Networks

Intelligent image recognition, information extraction, and classification are developing rapidly. In the past, traditional artificial neural networks (ANNs) were mainly driven by increasing the model capacity, and to improve pixel-scale based image recognition problems, models are moving towards having billions of parameters, resulting in hardware devices that cannot meet the requirements; thus, deep convolutional neural networks (DCNNs) have become the basis of practically all state-of-the-art approaches for image classification, object detection, semantic segmentation, and many other applications.

CNN is a kind of feedforward neural network which can represent the features extracted from different regions of input data. A CNN is composed of several layers with different functions; it provides convolutional computation and back propagation through multiple filters $f^{(n)}$ to update weights in a matrix. The basic working principle is as follows (Fig. 1). Firstly, the input image is vectorized according to the pre-set input size. Then, the convolution calculation, activation function, pooling, regularization processing, and other operations are used to integrate a large number of distributed discrete input signals into an output signal. The weight matrix is updated through the loss function and the continuous backward propagation. After that, the data obtained from the last filter are passed to the fully connected layer to achieve parameter sharing. Finally, a flattening operation is used to classify by SoftMax functions.

The conceptual definition and theoretical analysis of CNN are formulated as follows.

$$\text{Basic principle: } y = \sum_{i=1}^{n} W_i X_i + \text{bias}, \tag{1}$$

where y is the regression dependent variable of the neural network, W is the weight matrix, and X is the vectorization matrix of input image.

$$\text{Input layer}: I = \text{Input} = \text{vectorize}(\text{image}(A, B)), \tag{2}$$

where I is input image vectorization matrix, A and B are the input image sizes.

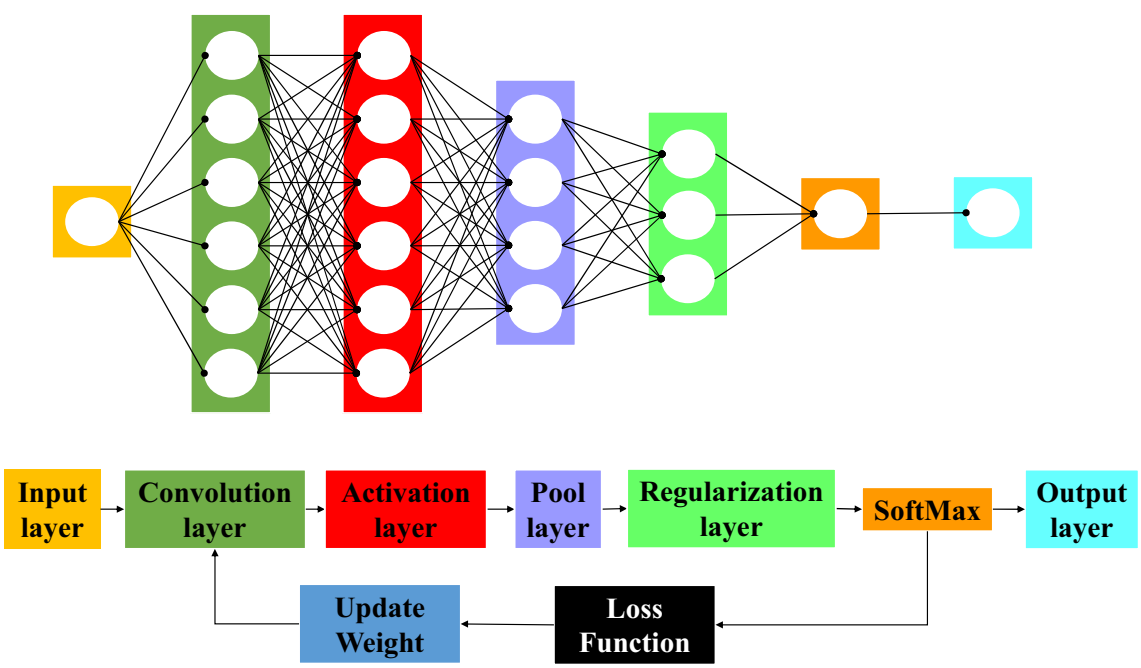


Fig. 1 Schematic diagram of Convolution neural network (CNN). The input data in turn through layers of convolution, activation function, pooling and regularization for the calculation of loss, and the weight matrix is updated iteratively

$$\text{Convolution layer}: \text{Output} = \sigma\left(\sum_{l=0}^{N}\sum_{m=0}^{N} W_{l,m} \cdot I_{j+l,k+m} + \text{bias}\right), \tag{3}$$

where n is the convolution kernel size ($N \times N$); j and k are the iteration times respectively; W is the neural network weight matrix; I is the input layer matrix; and σ is the activation function.

$$\text{Pool layer}: o = \left[(I - K + 2P)/S\right] + 1. \tag{4}$$

To express the pooling function conveniently, assume that the image processed by the pool layer is a square. In Eq. (4), o is the size of the image after the pooling layer ($o \times o$), K is the pooling kernel size, S is the string of the pooling operation, and P is the padding of the pooling operation.

$$\text{ReLU}(x) = \begin{cases} x \text{ if } x > 0 \\ 0 \text{ if } x \le 0 \end{cases}. \tag{5}$$

Due to its simplicity, Eq. (5) is one of the most commonly used activation functions in deep learning.

$$\text{SoftMax: Softmax}\left(x_i\right) = \frac{e^{x_i}}{\sum_{c=1}^{C} e^{x_c}}. \tag{6}$$

In Eq. (6), e is Napier constant, x_i represents the input value corresponding to the i-th neural network neurons, and C is the total number of input vectors entering the SoftMax layer.

2.2 Transfer Learning

In the research on deep learning, a large number of neural networks with different structures have been established for different feature attribute datasets. The purpose of these neural networks is to achieve higher recognition accuracy and better robustness on recognition problems. However, it is difficult to obtain such a large data set from life to accurately train new neural networks for more specialized image data or other types of data recognition and classification problems. It's time-consuming, laborious, and the results are variable.

Transfer learning (TL) inherits the weights and biases of a trained neural network as initial values (we call these "knowledge") and updates the weights and biases on new and smaller datasets to obtain new feature extraction capabilities and is a method to improve recognition accuracy, reduce model capacity and gradient fast descent. Xu et al. (2017) and Zhang et al. (2018) both report that the use of TL based on pre-trained models can reduce time, cost and dependence on big data. Therefore, the wider use of TL can better introduce the concept of intelligence to more domains to solve more specific and complex problems. For these reasons, TL is considered to be a key approach to achieving intelligent solutions.

When using neural networks to classify rock images, the training model has very unstable recognition results due to the quality of the images in the dataset and the setting of the training parameters. This instability is reflected in: (1) the difficulty in accurately recognising specific rock structures and formations; (2) the very high recognition accuracy of the images in the training dataset and the very low accuracy of the test dataset, i.e. the very poor robustness of the neural network; (3) the difference in recognition results between multiple training sessions on the same image. However, the transfer learning approach allows for the training of hidden skeletal and contour features in the image in advance, making it more sensitive to structural features. By retaining the parameters of the training model, this 'knowledge' can be inherited into the rock classification model, resulting in a more targeted lithology recognition model.

2.3 Mix-Scale Rock Automatic Recognition

To promote the society and ensure the safety of geologists, information such as remote sensing data, UAV (Unmanned Aerial Vehicle) photographs, geophysical survey data, drilling data and electron microscope thin sections are used to investigate the geology of the study area. However, these methods of investigation can only provide a reference for the project and can hardly replace the professionals. The mechanical properties and regional geological environment characteristics of igneous rock, sedimentary rock, and metamorphic rock are obviously different. The accurate identification of lithology is beneficial to the evaluation of surrounding rock strength and the exploration of underground water bearing and mineral resources. This is of great significance for engineering construction and resource exploration. There are three key problems in establishing a new neural network to realize the accurate classification of rock images. First, the characteristics of rock images with different scales are obviously different. Second, it is difficult to establish enough data sets with accurate corresponding relationships. Third, metamorphic rocks inherit the characteristics of parent rocks and are easily misidentified. To achieve the automatic recognition of mixed scale rock lithology, the TL method is adopted to inherit the feature extraction ability of pre-trained neural network and update the weight matrix. This is more suitable for the extraction ability of rock hand scale and field-scale structure features (Fig. 2).

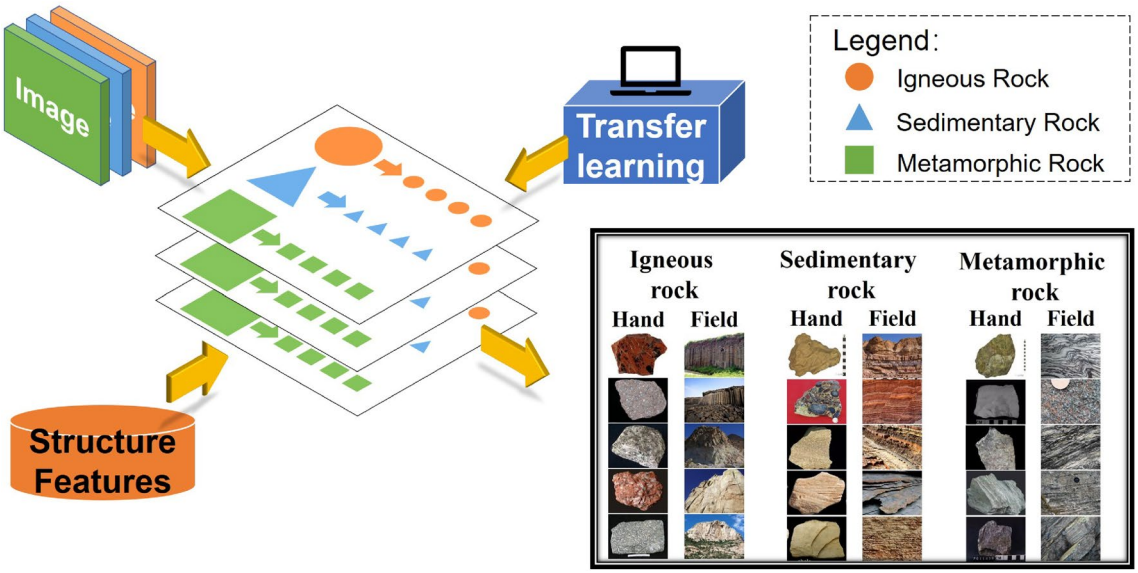


Fig. 2 Schematic diagram of mixed scale lithology classification in rock image. The input images go through different structure pre-trained neural networks to achieve automated classification for rocks

Table 1 Features summary of the dataset on rocks structure and fabric characteristics

Igneous rock	Feature	Porphyritic texture	Massive structure	Air hole and almond construct	Intrusive feature	Others
	110	60	18	17	12	3
	Hand	54.55%	16.36%	15.45%	10.91%	2.73%
	Feature	Columnar joint structure	Massive structure		Intrusive feature	
	110	54	47		9	
	Field	49.09%	42.73%		8.18%	
Sedimentary rock	Feature	Sandstone	Conglomerate	Limestone	Clay rock	Coal and others
	110	40	29	20	11	10
	Hand	36.36%	26.36%	18.18%	10.00%	9.09%
	Feature	Bedding structure	Bedding plane structure	Clastic characteristics	Others	
	110	76	20	7	7	
	Field	69.09%	18.18%	6.36%	6.36%	
Metamorphic rock	Feature	Schistosity structure	Crystalloblastic texture	Palimpsest texture	Metasomatic texture	
	110	60	36	8	6	
	Hand	54.55%	32.73%	7.27%	5.45%	
	Feature	Characteristics of dynamic metamorphism			Characteristics of thermal contact metamorphism	
	110	55			55	
	Field	50%			50%	

Table 2 Table of Initial hyper-parameter setting

Parameters/function	Value
Loss function	RMSprop
Max epochs	80
Mini batch size	4 or 5
Validation frequency	50
Validation patience	Inf
Initial learn rate	1e−5
Learn rate drop period	10
Learn rate drop factor	0.1
Momentum	0.9

3 Recognition Results and Analysis of Different Scales

3.1 Dataset Preparation

Based on the structure and fabric characteristics of rocks, the image data sets of igneous rocks, sedimentary rocks, and metamorphic rocks are established. Part of the data is from the internet, and the other is from the website: http://www.earth.ox.ac.uk/~oesis/, there are 660 images in total. The detailed data characteristics are shown in Table 1.

3.2 Operating Environment and Parameter Setting

The neural networks trained by different data sets have different latent space features extraction abilities. Firstly, seven typical and single-structured neural network models (AlexNet, Vgg16, Vgg19, Densenet201, InceptionV3, InceptionresnetV2, SqueezeNet) are used to calculate and analyze the data sets of hand scale, field scale and mixed scale. To speed up the calculation, the execution environment utilizes a GPU; the hyper parameters used in the TL are shown in Table 2. The system has an Intel®Core™ i9-9900 K CPU @3.60 GHz, 32 GB of RAM, and an × 64 operating system.

Table 3 Classification accuracy and training time results of established rock data sets with different network structures

Neural network	Layers	Hand scale accuracy (%)	Hand scale time (s)	Field scale accuracy (%)	Field scale time (s)	Learn rate
AlexNet	24	33	7 min 48	33	11 min 14	1e−11
VGG16	41	36.36	36 min 53	57.58	42 min 32	1e−11
VGG19	47	51.52	46 min 39	54.55	50 min 20	1e−11
Densenet201	709	68.18	243 min 43	75.76	265 min 50	1e−11
InceptionV3	316	60.61	76 min 3	77.27	75 min 37	1e−11
InceptionResnetV2	825	74.24	249 min 30	74.24	252 min 52	1e−11
Squeeze net	68	46.87	11 min 10	74.24	14 min 3	1e−11

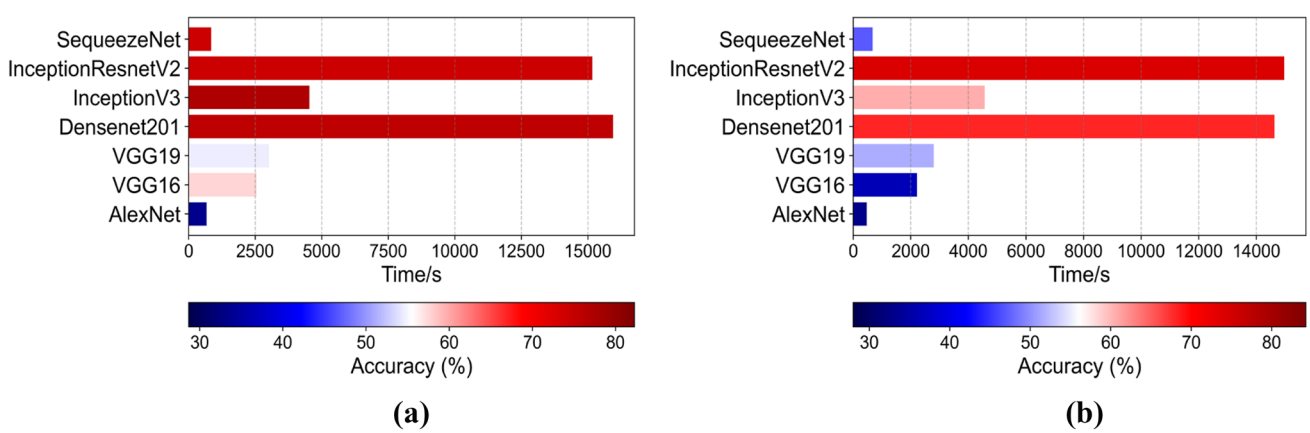


Fig. 3 **a** Classification accuracy and training time of seven neural networks with different structures for field-scale rock images. Squeeze Net obtained the highest accuracy with a very short training time. **b** Classification accuracy and training time of seven neural networks with different structures for hand scale rock images. InceptionResnetV2 showed the best performance on hand scale images. Instead, Squeeze Net showed an inferior result

Table 4 Table of hyper-parameter setting of four deep neural networks for retraining

Parameters/function	Value
Loss function	RMSprop
Verbose frequency	100
Max epochs	200
Mini batch size	5
Learn rate drop period	5
Learn rate drop factor	0.1

Table 5 Summary of the results of the retrained neural network for the rock classification task

Neural network	Hand scale accuracy (%)	Time (s)	Field scale accuracy (%)	Time (s)
InceptionV3	63.64	12,036	69.70	11,663
Densenet201	51.52	67,202	66.67	64,303
InceptionResnetV2	57.58	64,532	75.76	64,565
Squeeze net	59.09	856	80.30	961

3.3 Recognition Results of Each Dataset

In this work, TL is implemented with the Deep Network Designer and Transfer Learning open-source App in MATLAB R2019a. 80% of the 660 photos in the data set are used for training and 20% for validation. To improve the training accuracy, the images are pre-processed with the data augmentation methods of flipping and rotation.

3.3.1 Hand Scale and Field Scale Results

The data sets are divided into 330 hand scale data sets and 330 field-scale data sets with igneous rock, sedimentary rock, and metamorphic rock as the classification criteria. The recognition accuracy and training time are used as evaluation indexes. The calculation results are shown in Table 3 and Fig. 3.

The characteristics of the hand scale data set mainly include porphyritic texture, air hole and almond structure characteristics of igneous rocks, clastic grain characteristics of sedimentary rocks, schistosity, and metamorphic structure characteristics of metamorphic rocks. The above-mentioned structure and fabric characteristics all reflect the geometric shape, color difference, and spatial distribution characteristics of mineral particles in rock hand specimen images. The characteristics of the field-scale data set mainly include columnar joint structure and massive structure of igneous rock, bedding structure and bedding plane structure of sedimentary rock, dynamic metamorphism, and thermal contact metamorphism of metamorphic rock. Compared with the characteristics of mineral particles, these structure and fabric features have larger scales and more obvious macro characteristics; these are similar to the identification and detection of object roughness and geometry shape.

The accuracy and training time of the test are analyzed as indicators. The recognition accuracy of the data sets for the hand scale in the InceptionalResnetV2 and Densenet201 networks are higher, reaching 74.24% and 68.18%, respectively. These results are consistent with the traditional concept of achieving higher accuracy with

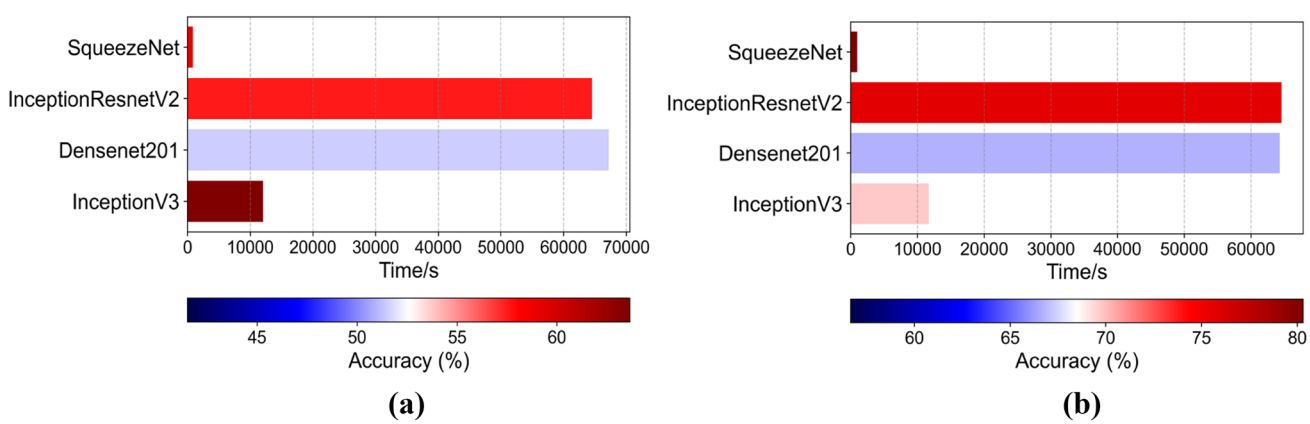


Fig. 4 **a** Classification accuracy and training time of four deep neural networks on field-scale rock images with resetting hyper parameters. Squeeze Net consumed the shortest time but get the highest accuracy. **b** Classification accuracy and training time of four deep neural networks on hand scale rock images with resetting hyper parameters. Surprisingly, InceptionV3 showed a great performance on the classification of hand scale rock images

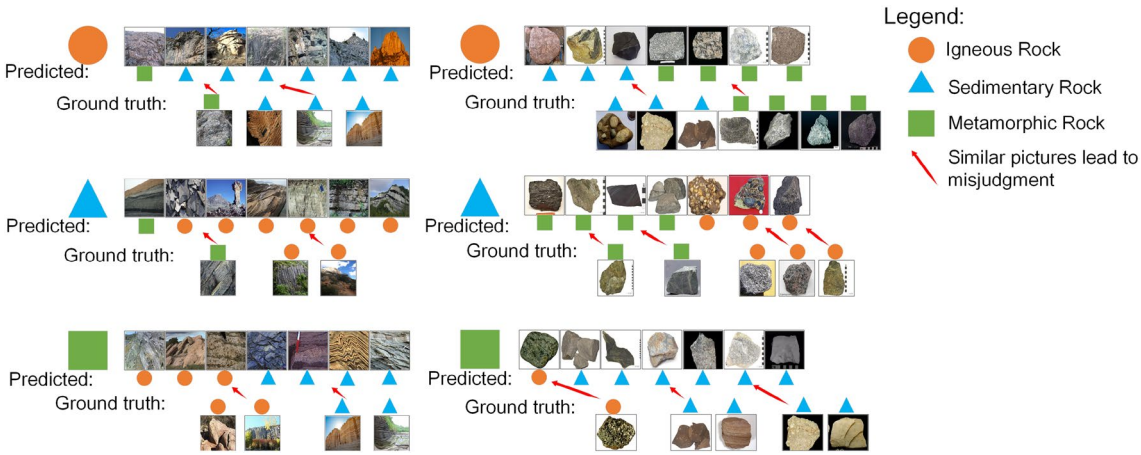


Fig. 5 Hand and Field-scale classification error set analysis chart. The right side of each type of rock (circle, triangle, rectangle) in the figure corresponds to the image with the wrong prediction, the bottom of each image corresponds to the wrong prediction classification, and the image starting from the red arrow which is similar to the wrong prediction set

an increasing number of layers, but the training time has exceeded four hours. For the field scale data set, Dense net201,InceptionResnetV2,InceptionV3 and SqueezeNet all achieve high precision: 75.76%, 77.27%, 74.24% and 74.24%, respectively. When considering the index of training time, the results of InceptionV3 and SqueezeNet are better than other networks; in particular, the lightweight network Squeezenet performs very well: 843 s training time to reach 74.24% accuracy.

According to the training process curves of accuracy and loss, it is found that for the deeper layers' networks InceptionResnetV2 and Densenet201, the best effect may not be achieved due to the parameter settings of epoch and mini-batch. Therefore, the hyperparameters are adjusted. The specific parameters and training results are shown in Tables 4, 5 and Fig. 4.

Based on the re-training results, three important observations can be made. First, blindly increasing the number of iterations and the depth of the neural network layer does not necessarily achieve higher recognition accuracy. Second, the network structure of SqueezeNet has a good effect on the feature extraction of field scale rock image set, and the training time is far lower than other deep networks. Third, residual connection structures have shown a good extraction ability on hand-scale rock images.

3.3.2 Error Sets Analysis

It is not enough to evaluate the ability of the network to extract rock image features only based on accuracy and time. By integrating and analyzing the error set and ground truth set after TL Densenet201, InceptionV3, InceptionResnetV2, and SqueezeNet. Compared with the characteristics of error and ground truth solutions shown in Fig. 5, we found the following for the field scale and the hand scale results:

Table 6 Calculation results of neural networks after mixing different scale rock image sets

Neural network	Mix scale accuracy (%)	Time (s)	Learn rate
AlexNet	33	1401	1e−11
VGG16	31.82	2704	1e−11
VGG19	43.94	2919	1e−11
Densenet201	65.15	33,859	1e−11
InceptionV3	63.64	10,554	1e−11
InceptionResnetV2	72.73	38,890	1e−11
Squeeze net	62.88	1595	1e−11

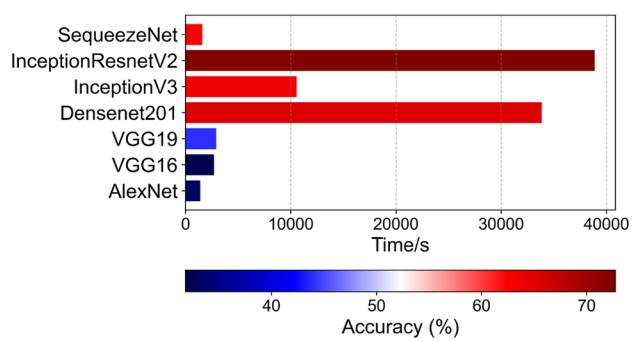


Fig. 6 The classification accuracy and training time consumption of each neural network for different scale data sets after mixing. InceptionResnetV2 obtained the best recognition accuracy but took the longest time to train

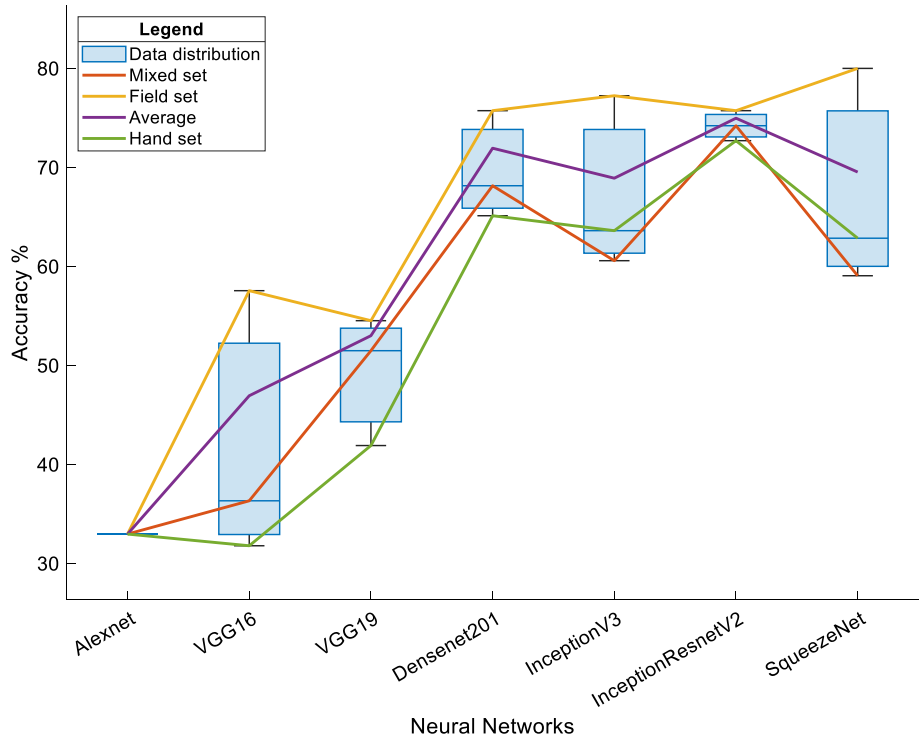


Fig. 7 Accuracy distribution box chart and line chart (Note1: Data shown in average line is the arithmetic mean of the precision of handset and fieldset. Note2: Data distribution represents the accuracy gap between hand scale set and field-scale set)

Field scale Based on the analysis, three important observations are made. First, the identification of igneous rocks is seriously affected by weathering, some columnar joint structures and massive structures after weathering are easy to be predicted as sedimentary rocks with and bedding structure due to their light surface color. Second, sedimentary rocks are affected by special strata and denudation. For example, coal seams and klippe are easily predicted as metamorphic rocks and igneous rocks. Third, some features caused by thermal contact metamorphism are easily confused with porphyritic texture; some weathered metamorphic rocks with schistose structure are identified as sedimentary rocks.

Hand scale Based on the analysis, four important observations are made. First, the accuracy of the data set verification is greatly affected by the structure of the neural network compared with field-scale results. Second, the rock sample with a pillow structure is easy to be identified as a conglomerate; the seriously weathered and light-colored rock is easily identified as sedimentary rock. Third, the Inception module and residual connection have a positive impact on the recognition of particle characteristics of sedimentary rock. Fourth, the recognition of metamorphic rocks is greatly influenced by color and boundary shape; this is impacted by the characteristics of the parent rock.

3.3.3 Mixed Scale Analysis

It is generally considered that the recognition accuracy is expected to fall between the hand scale and the field scale after mixing the two-scale data set. In addition, the deeper the layer is, the better accuracy and robustness are anticipated. However, experimental results are shown in Table 6, Figs. 6, and 7 reveal three important results. First, some networks acquire good recognition accuracy on field-scale images but not on the hand scale images; these include IncceptionV3 and SqueezeNet. When the sets are mixed,

Table 7 Precision and time results on hand scale and field-scale sets of each neural network (The accuracy improvement ratio is the percentage of accuracy improvement compared with the Alexnet network)

Neural network	Hand scale accuracy (%)	Time (s)	Accuracy improvement ratio (%)	Field scale accuracy (%)	Time (s)	Accuracy improvement ratio (%)
ResNet18	46.87	1426	42.03	71.21	1379	115.79
ResNet50	59.09	4579	79.06	74.24	4443	124.97
ResNet101	56.06	5198	69.88	69.79	5459	111.48
InceptionV3	60.61	4563	83.67	77.27	4537	134.15
InceptionResnetV2	74.24	14,970	124.97	74.24	15,172	124.97

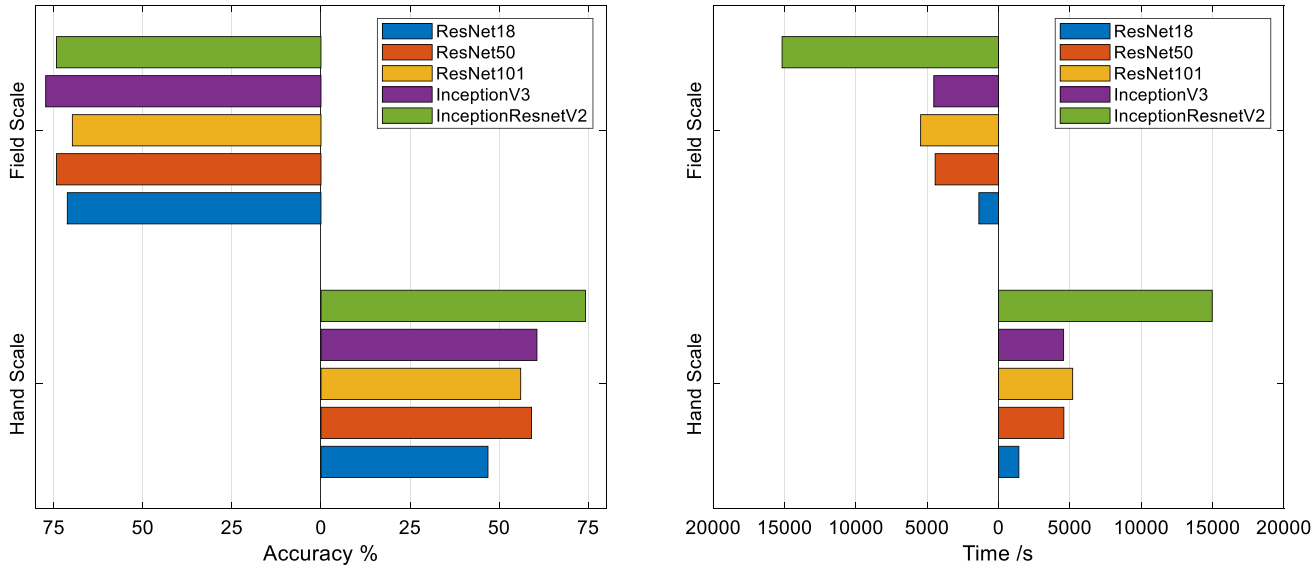


Fig. 8 Accuracy and Time results from bar chart (left: the precision result of each network in hand scale and field-scale sets; right: the training time result of each network in hand scale and field-scale sets, respectively)

the accuracy is greatly reduced, even lower than the hand scale. Second, the accuracy of mixed scale verification is generally lower than the arithmetic average of hand and field-scale verification. Third, the super deep neural network InceptionResnetV2 shows good robustness for different scale data sets; this is followed by the densenet201 network. Although the field scale feature recognition ability of the SqueezeNet network is better, the robustness is poor due to its network structure.

4 Influence of Neural Network Structure

4.1 Influence of Network Structure on Accuracy

The results presented in Sect. 3 show that the neural networks based on different data sets show better effects for new data sets with similar features after TL, and the structure of the neural network affects the updating ability of weight, that is, the robustness of the network. Compared with the calculation results of each network, the recognition accuracy of the network only relying on linear convolution layer stacking after TL is low; however, Densenet201 and SqueezeNet which are based on the neural network structure of parallel convolution layers and a residual connection show a good ability for field-scale rock recognition. For hand scale images, the accuracies are greatly reduced for the mineral particle feature. However, InceptionV3 and InceptionResnetV2 (that use the Inception module) have shown a good recognition ability on the hand scale image set. Therefore, it is speculated that a residual connection structure or an Inception module structure has a better feature extraction ability for the mineral particle characteristics of rock in the established data set. Thus, the recognition effect of ResNet18, ResNet50, ResNet101, Inceptionv3, and InceptionResnetv2 on a field scale and hand scale is verified. The results are shown in Table 7 and Fig. 8.

Compared with the recognition results of the five networks, although the training time of InceptionResnetV2 is long, its structure and fabric stability, robustness, and

Table 8 Transfer learning time comparison table (Time 1 = training time on hand scale set, Time 2 = training time on field-scale set, Time 3 = training time on mixed scale set, Ratio 1 = Time 3/(Time 1 + Time 2), Ratio 2 = (Time 1 of Squeezenet)/(Time 1 of each network), Ratio 3 = (Time 2 of Squeezenet)/(Time 2 of each network), Ratio 4 = (Time 3 of Squeezenet)/(Time 3 of each network))

Neural network	Time 1 (s)	Time 2 (s)	Time 3 (s)	Ratio 1 (%)	Ratio 2 (%)	Ratio 3 (%)	Ratio 4 (%)
Squeezenet	856	961	1595	87.78	–	–	–
InceptionResnetV2	14,970	15,172	38,890	129.02	5.7181	6.334	4.1013
InceptionV3	4563	4537	10,554	115.98	18.7596	21.1814	15.1128
Densenet201	14,623	15,950	33,859	110.75	5.8538	6.0251	4.7107

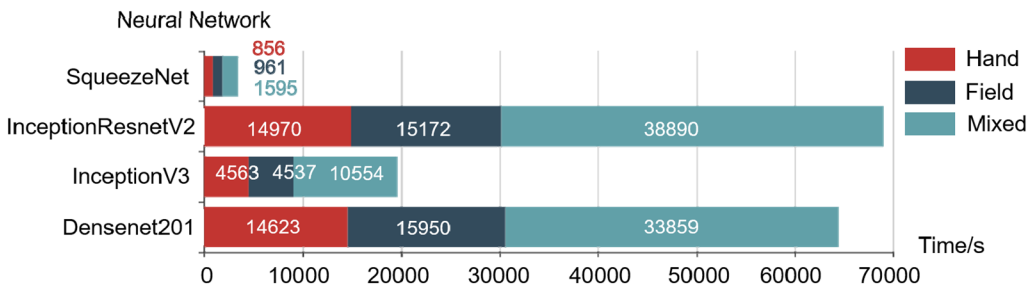


Fig. 9 Time cost stack bar chart of neural networks (Squeeze Net, InceptionResnetV2, InceptionV3, Densenet201)

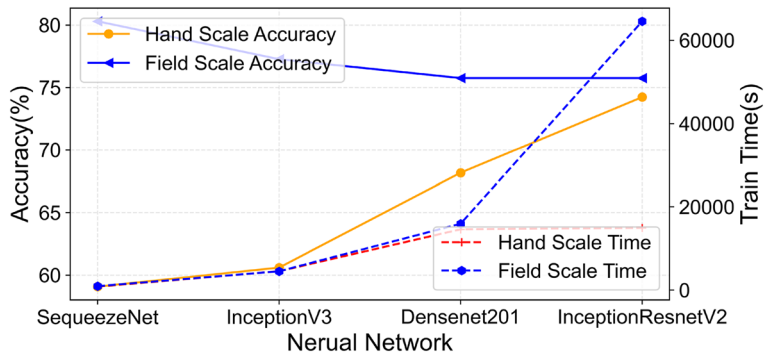


Fig. 10 Comparison line chart of classification accuracy and training time consumption of tested neural networks. (The number of layers of networks increases from left to right (Squeeze Net to Inception-ResnetV2). Solid lines correspond to accuracy-related indicators, dotted lines correspond to time-related indicators)

recognition accuracy of each scale data set are very high. Surprisingly, ResNet50 shows similar recognition ability and computational efficiency as InceptionV3, but Inception-ResnetV2 shows a better effect on fine-grained feature recognition. Moreover, the ResNets have all achieved high recognition accuracy for the field scale set. Compared with the accuracy of Alexnet network, the accuracy can be improved by up to 124.97%. In conclusion, the parallel convolution layer branch structure and residual connection structure have good feature recognition ability for rock data sets after TL.

4.2 Effect of Compression Structure on Time

In addition to the accuracy, the time cost of neural network training is also an important evaluation index. Due to the low recognition accuracy of AlexNet,VGG16, and VGG19, they are not included in the comparison of time cost. The calculation results are shown in Fig. 7. To achieve the same recognition accuracy in a shorter time, the fire module used by SqueezeNet is proposed, through TL training, it is found that this network structure has a shorter training time and similar recognition accuracy compared with other neural networks on the field scale set. From Table 8 and Fig. 9, we see that the training time of the mixed data set of InceptionResnetV2, InceptionV3, and Densenet201 is greater than the sum of the time needed for single scale set training (Ratio1 is greater than 100%). SqueezeNet substantially shortens the training time due to its unique network module (Ratio1 = 87.78%). Moreover, the computation time of Squeezenet is only 4–21% of the learning time of other network structures. Specially, it takes only 4.1% of the learning time on a mixed scale set of InceptionResnetV2.

According to the data shown in Fig. 10, three important observations can be made. Firstly, SqueezeNet has the highest accuracy on the field scale with the shortest training time, but the hand scale feature recognition effect is poor. Secondly, InceptionResnetV2 has good robustness, but it costs much more time than SqueezeNet on training. Thirdly, with the deepening of neural network layers, hand scale recognition accuracy increases, but the training time also increases. However, SqueezeNet has the best recognition effect and spends the least time with a small number of layers.

5 Comparative Analysis of MobileNetV2

5.1 MobileNet Network Structure

To increase the recognition accuracy and reduce the training speed of neural networks for the data sets (two scales and mixed scales), the network structure needs to be improved. Ideally, a solution that inherits the fire module of the SqueezeNet network structure in addition to the Inception module and residual module (Figs. 11 and 12) is needed. Therefore, MobileNetV2 is selected

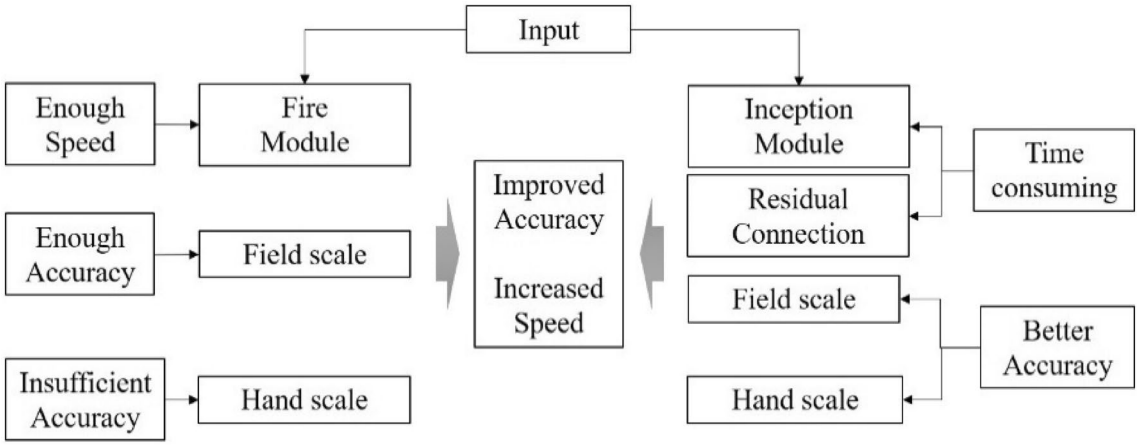


Fig. 11 Requirements diagram of MobileNetV2. Because we need to shorten the training time of the neural network, we introduce the fire module to accelerate. Coincidentally, the Squeeze Net based on fire module has good recognition accuracy for field-scale data sets. Because we need high classification accuracy for hand data set, field data set and mixed data set, we hope that the neural network structure inherits the characteristics of the concept module and residual connection

to test whether the training time can be reduced and the recognition accuracy can be improved at the same time. MobileNets was built for mobile and embedded vision applications. MobileNetV2 added inverted residuals and linear bottlenecks on the basis of MobileNet, and it is the state-of-the-art fabric at this stage.

5.2 Comparative Analysis of MobileNetV2

After fine tuning the neural network fabric, TL is adopted. The parameters are set as follows: Max epochs = 120, mini batch size = 12. The calculation results are shown in Fig. 13.

5.2.1 Analysis of Accuracy and Time

MobileNetV2 obtained 81.82% accuracy on field-scale recognition, which is the first-time results have exceeded 80% in machine recognition. MobileNetV2 shows a good performance on the hand scale set and mixed scale set, and the accuracy is 71.21% and 66.67%, respectively; this is followed by InceptionResnetV2. The results also indicate substantial reductions in training time.

The area of the circle is used as an index, and it is defined as A = time/iteration. The higher the value of A is, the less time it takes for the same iteration times and the lower the time cost. From Fig. 14, it can be seen that compared with other networks with a similar time cost, MobileNetV2 has higher precision and lower time cost than networks with similar precision.

5.2.2 Analysis of Error Set Characteristics

By observing the characteristics of the error sets of MobilNetV2 in the hand scale, field-scale and mixed-scale image sets are shown in Fig. 15, three important observations can be made. Firstly, the recognition of the field scale is obviously affected by weathering. All light-colored rocks are easily identified as sedimentary rocks after weathering, and some column joint structures with broken weathering are also easily identified as sedimentary rock. At the same time, coal seams are generally identified as intrusive bodies. Secondly, it is difficult to extract the fine-grained features of minerals accurately, and ability to distinguish between metamorphic rocks and their parent rocks. Most of the lithology identification results are sedimentary rocks due to weathering. Thirdly, although the mixed scale has achieved high accuracy, the classification characteristics of the data set are not ideal, and it is still difficult to identify coal seams. For columnar and massive sedimentary rocks formed by wind erosion, they are easily identified as igneous rocks. Interestingly, one field scale sedimentary rock image with bedding structure and another hand scale metamorphic rock image with a schistose structure obtain the opposite classification result. Overall, the robustness of MobileNetV2 network is not good enough, but it does provide good recognition abilities for field-scale features with fast training speeds.

6 Discussion

The traditional concept of lithologic recognition is to identify minerals first, and then classify rocks according to their minerals content. This paper starts with the natural

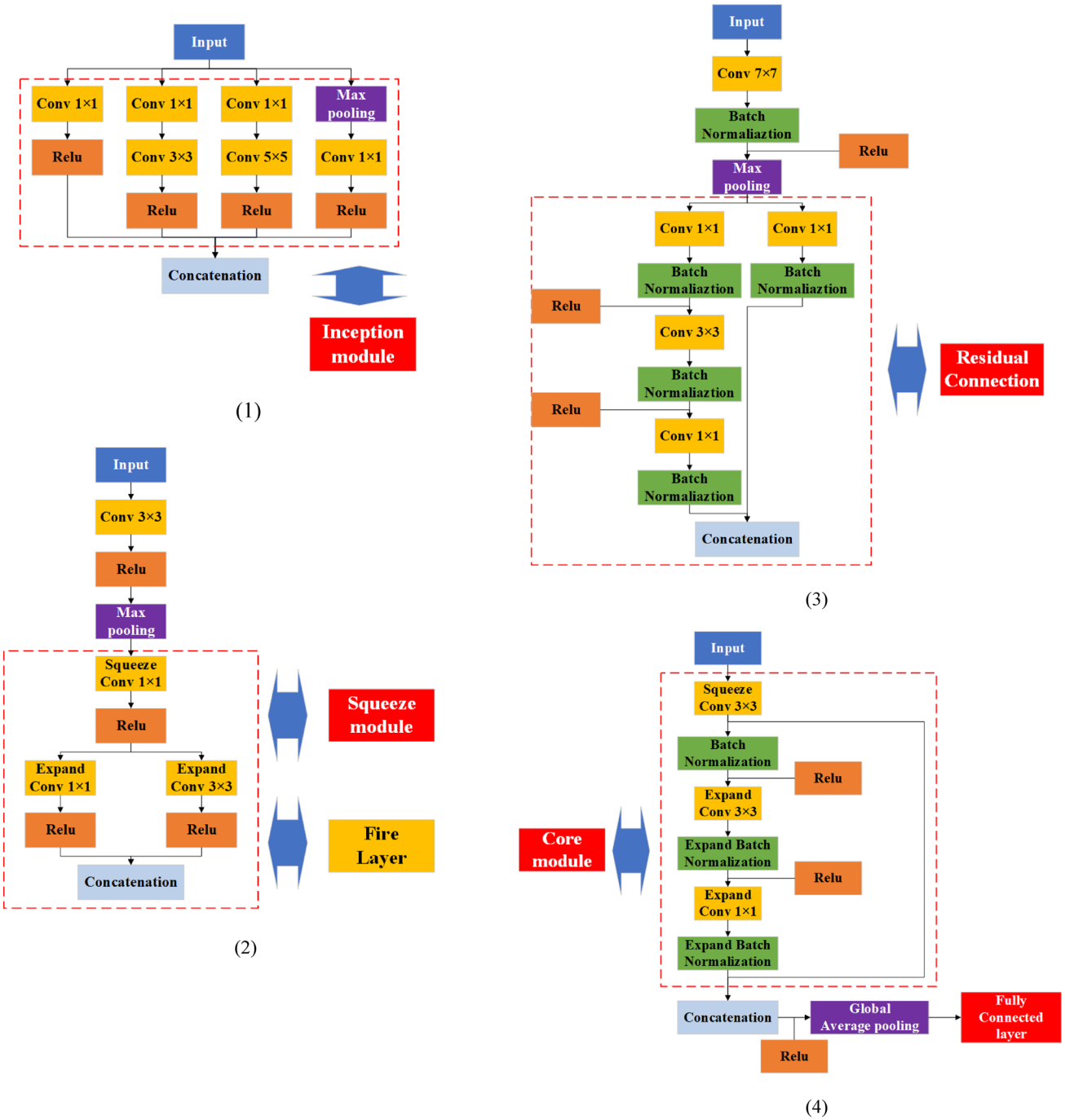
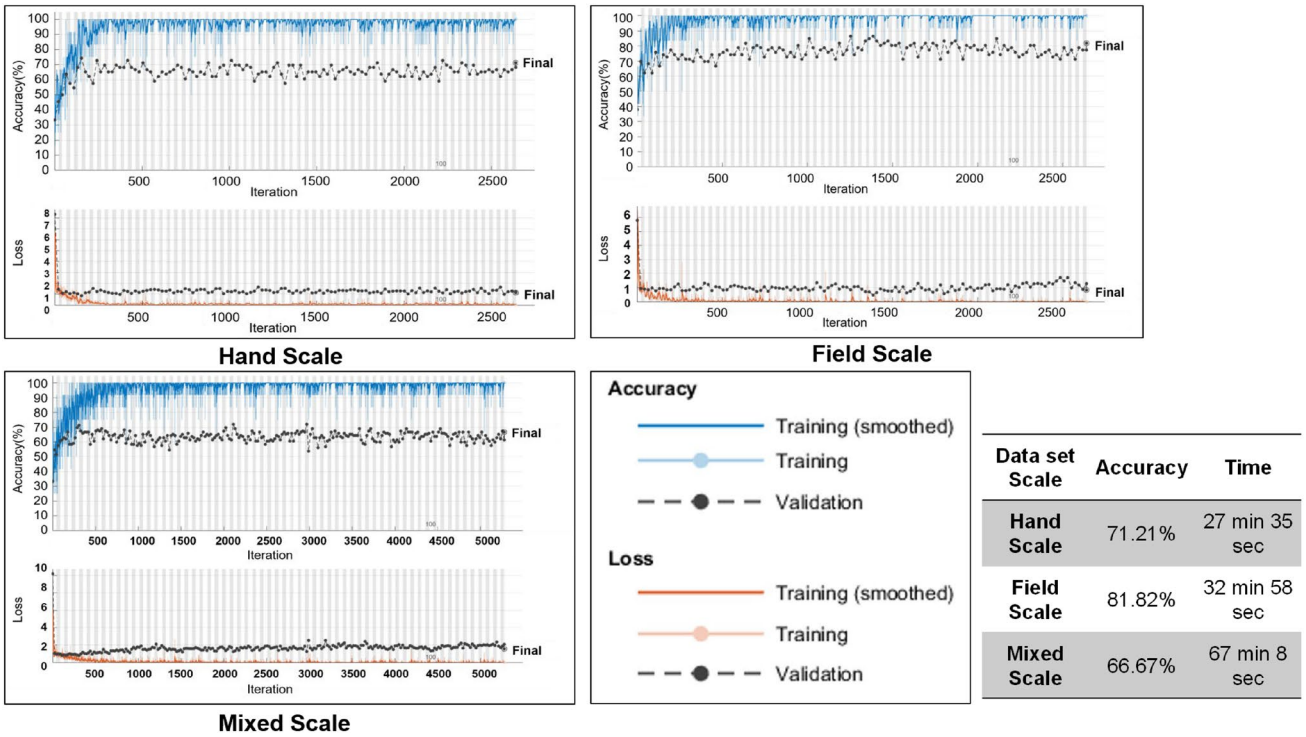


Fig. 12 Flow chart of different neural network architecture core module(Inception module(1), Squeeze module(2), Residual Connection(3), MobileNetV2(4))

characteristics of rocks. The structure and fabric characteristics of rocks displayed in the field scale and the mineral structure features shown in the hand scale are used as the basis for lithology classification. The TL method is used to inherit the original dataset knowledge, and the weight matrix is adjusted through different network structures. The conclusion is that the fire module can greatly reduce the training time; in addition, it has a strong edge-feature extraction ability for the structure and fabric characteristics of rock lithology. The ability to recognize the specialized characteristics of minerals and network robustness are significantly improved by the Inception module and residual connection. It is feasible to carry out the automatic classification of rocks according to the macro characteristics of the rock.



Data set Scale	Accuracy	Time
Hand Scale	71.21%	27 min 35 sec
Field Scale	81.82%	32 min 58 sec
Mixed Scale	66.67%	67 min 8 sec

Fig. 13 MobileNetV2 calculating results (including curves of accuracy and loss in transfer learning, final recognition accuracy and training time results)

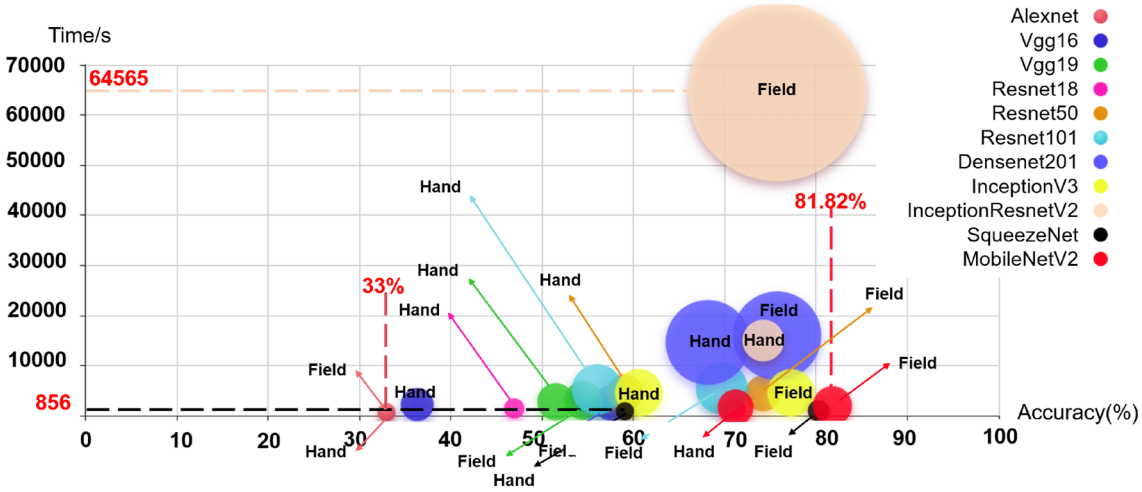


Fig. 14 Scatter comparison chart of each neural network results after transfer learning (A = area of the circle = time/iteration)

To evaluate the capability of each neural network in lithology recognition more comprehensively, an index = Accuracy/A is introduced to comprehensively evaluate the recognition accuracy and time cost. The detailed data are shown in Table 9 and Fig. 16.

By comparing the time, iteration times, and accuracy of each network on a mixed scale, SqueezeNet achieves high accuracy with the shortest time, MobileNetV2 achieves higher accuracy with fewer iterations, and IneptionResnetV2 achieves the highest accuracy with the longest time and the most iterations. In contrast, if only considering the recognition accuracy and the robustness of the neural network, InceptionResnetV2 has the best performance; considering the time cost, MobileNetV2 is better.

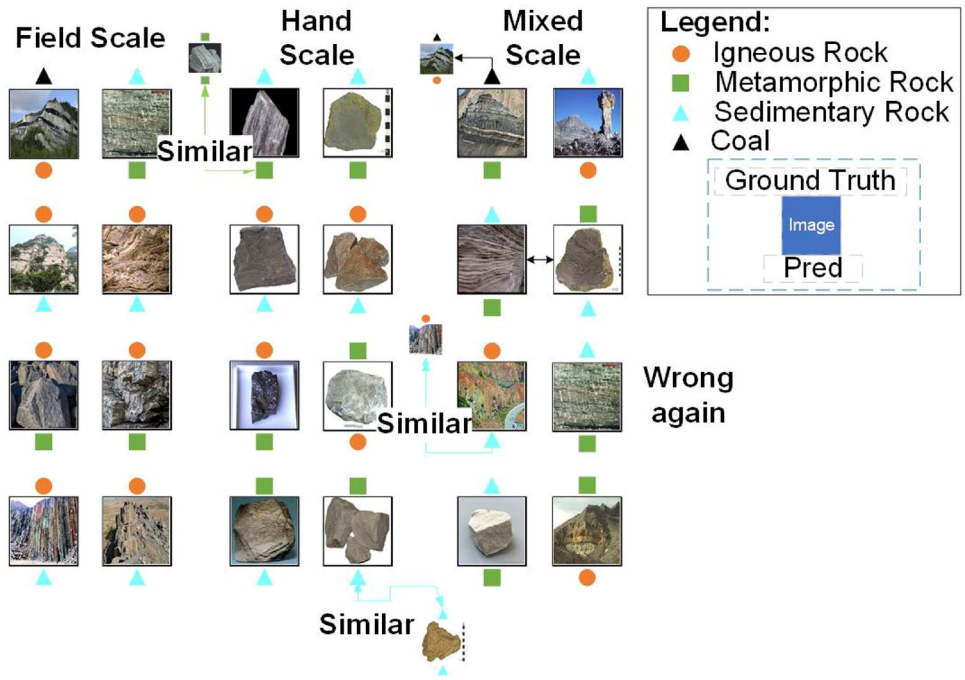


Fig. 15 MobileNetV2 wrong set analysis. (Igneous rocks, sedimentary rocks and metamorphic rocks are represented by circle, triangle and rectangle. Ground truth is shown above the image, and the prediction solution is shown below the image)

Table 9 Comparison table of training time consumption, lithology classification accuracy and iteration times of different neural networks

Neural network	Time (s)	Iteration	Accuracy	Accuracy/A
AlexNet	1401	8400	0.33	1.978587
VGG16	2704	8400	0.3182	0.988491
VGG19	2919	8400	0.4395	1.264748
Resnet101	14,916	10,560	0.5682	0.402265
Densenet201	33,859	8400	0.6515	0.161629
InceptionV3	10,554	8400	0.6364	0.506515
InceptionResnetV2	38,890	10,560	0.7273	0.197487
SqueezeNet	1595	8400	0.6288	3.311549
MobileNetV2	4028	5280	0.6667	0.873927

A = time/iteration

By analyzing the error sets of InceptionResnetV2, SqueezeNet, and MobileNetV2 for hand, field, and mixed scale image sets, it is found that InceptionResnetV2 has the best robustness for lithologic structure and fabric feature recognition, but the accuracy is not the highest. MobileNetV2, due to its special network structure, has an excellent recognition ability. For the large proportion features in field-scale data set such as massive structures and columnar joint structures it achieves 91.82% accuracy, and for bedding structures and bedding plane structures it achieves 87.27% accuracy. It is difficult for it to accurately identify some severely weathered rocks; here, the recognition accuracy reaches 81.82% on the field scale. In the hand scale set, the mineral characteristics (mineral size, color, grinding, compactness and so on) of igneous rock, sedimentary rock, and some metamorphic rock have a high recognition ability after TL of the Inception module and the residual connection network structure. Besides, for the porphyritic structure (54.55%) and the schizositity structure (54.55%) demonstrate excellent recognition ability, the best recognition accuracy of the neural network used in this study is 95.83% (MobileNetV2). However, for the crystalloblastic texture, because it inherits some characteristics of parent rock, it produces numerous similar misinterpretations. The causes of misinterpretation are shown in Fig. 17.

It is very effective to identify the rock on a field scale image set through the structure and fabric characteristics. This is helpful for lithology classification and UAV regional exploration before engineering construction begins, but it needs to find enough accurate and clear data sets on hand scale image sets. In addition, a special network structure to update weights continuously is needed such as the Inception module and the residual connection. These make the neural network more sensitive to subtle changes. Therefore, it is difficult to blindly use different kinds of a neural network to realize more professional lithologic recognition and identification. At this time, the role of the neural network is more like a black box. It is suggested that we can use a combination of various identification methods to determine the rock classification. First, we classify the lithology (igneous rock, sedimentary rock, metamorphic rock) according to the image of the structure and the fabric features. Nest, we distinguish the rock according to the hand scale mineral characteristics. Finally, the accurate identification of the rock according is determined based on the scanning results of the

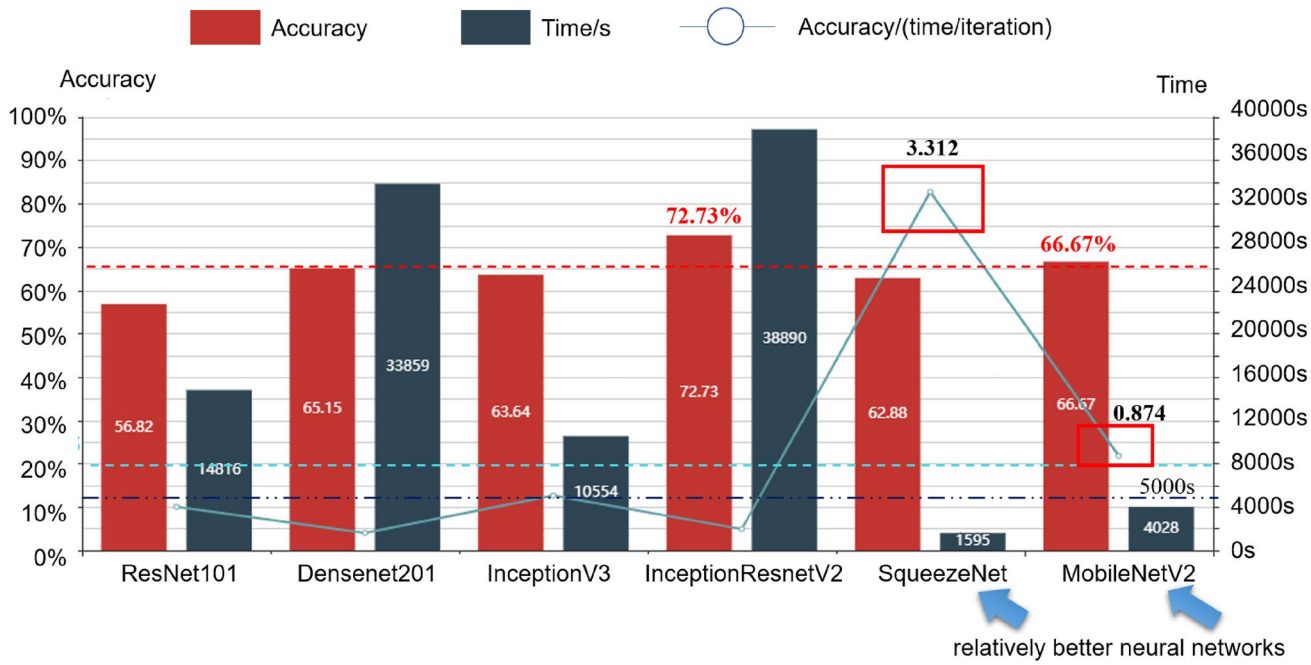


Fig. 16 Bar chart of neural network lithology classification results and line chart of comprehensive performance of neural networks. [Red dotted line represent classification accuracy equals 65.15%, blue dotted line represent index accuracy/(time/iteration) equals 0.85] (colour figure online)

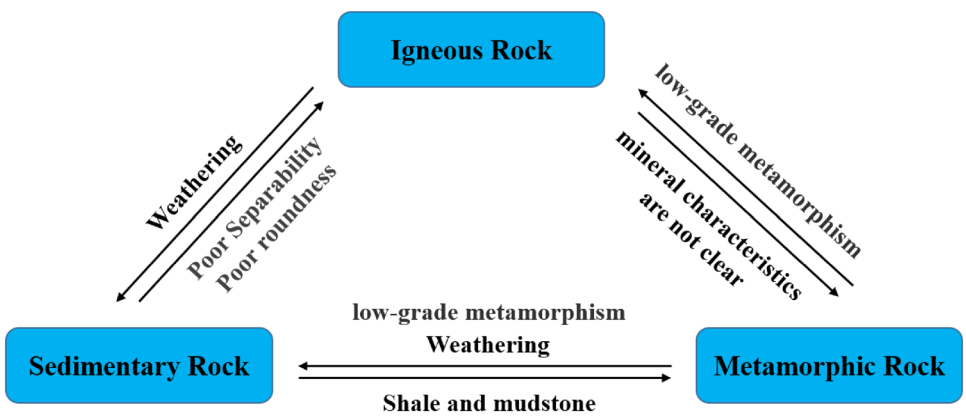


Fig. 17 Analysis chart of misinterpretation. Weathering makes the prediction result of rock tend to sedimentary rock, Low grade meteorology makes the prediction result biased to the parent rock lithology

electron microscope, which could effectively avoid the problem of endlessly expanding the data set.

7 Conclusions

Compared with the results of TL for each neural network, it is found that the robustness and accuracy of the InceptionResnetV2 neural network are the best, and the recognition accuracy in hand scale, field scale, and mixed scale are 74.24%, 74.24%, and 72.73% respectively. However, the training time is too long: the longest time can reach 64,565 s (1076 min 5 s). While MobileNetV2 achieves the highest accuracy of 81.82% in the field scale image set, and takes 1978s (32 min 58 s). In addition, SqueezeNet achieves 80.3% in field-scale accuracy and spends the shortest amount of time, 961 s (16 min 1 s).

The Inception module and the residual connection can significantly improve the recognition accuracy of mineral features in rock specimens, compared with Alexnet, the addition of these two network structures improve the recognition accuracy by up to 124.97%. And the fire module can greatly reduce the training time of the neural network, about 5 to 10 times, specially, the training time of Squeezenet is only 4.1–6.3% of InceptionResnetV2.

By observing the correct and incorrect solutions of MobileNetV2, it is found that the network has a strong recognition ability for bedding, columnar joint, and foliation,

recognition accuracy can up to 95.83%; these are difficult for humans to distinguish. However, it is weak for the feature extraction of the crystalloblastic texture and the palimpsest texture. In contrast, the existence of the Inception module can greatly improve the recognition accuracy of these features. Besides, weathering seriously affects the recognition and classification effect of the neural network on lithology.

Not only can the lithology identification be realized through mineral identification, but qualitative classifications can also be made by the structure and fabric differences of igneous rock, sedimentary rock, and metamorphic rock. Then, more specialized classification and rock naming can be carried out by examining the hand scale mineral characteristics or the scanning electron microscope results. After that, the engineering geological conditions and regional geological structure attributes can be analyzed.

Acknowledgements This work was financially supported by the National Key Research and Development Plan (Grant no. 2018YFC1504902), the National Natural Science Foundation (Grant nos. 52079068, 41941019), and the State Key Laboratory of Hydroscience and Hydraulic Engineering (Grant no. 2021-KY-04).

References

Amosu A, Sun Y (2017) WheelerLab: an interactive program for sequence stratigraphic analysis of seismic sections, outcrops and well sections and the generation of chronostratigraphic sections and dynamic chronostratigraphic sections. SoftwareX 6:19–24

Bahmani AA, Riahi MA, Ramin N (2020) Detection of stratigraphic traps in the Asmari Formation using seismic attributes, petrophysical logs, and geological data in an oil field in the Zagros basin, Iran. J Petrol Sci Eng 194:107517

Baker BR, Gessner K, Holden E-J, Squelch AP (2008) Automatic detection of anisotropic features on rock surfaces. Geosphere 4(2):418–428

Castafio R, Anderson, R, Estlin, T, DeCoste, D, Fisher, F, Gaines, D, Mazzoni, D, Judd, M (Year) Rover traverse science for increased mission science return. In: 2003 IEEE aerospace conference. 8–15 March Big Sky, MT, United States, pp 3629–3636

Deepnarain N, Nasr M, Kumari S, Stenström TA, Reddy P, Pillay K, Bux F (2020) Artificial intelligence and multivariate statistics for comprehensive assessment of filamentous bacteria in wastewater treatment plants experiencing sludge bulking. Environ Technol Innov 19:100853

Dramsch JS, Lüthje M (2018) Deep-learning seismic facies on state-of-the-art CNN architectures. In: Seg technical program expanded abstracts 2018. Society of Exploration Geophysicists, pp 2036–2040

Duan Y, Zheng X, Hu L, Sun L (2019) Seismic facies analysis based on deep convolutional embedded clustering. Geophysics 84(6):IM87–IM97

Dunlop H (2006) Automatic rock detection and classification in natural scenes. Masters Thesis, Carnegie Mellon University.

Fan L, Zhao W, Gong ZN (2012) Correspondence analysis of rock spectra based on continuum removing. J Jilin Univ 42(2):575–582

Feng YX, Gong X, Xu YY, Xie Z, Cai HH, Lv X (2019) Lithology recognition based on fresh rock images and twins convolution neural network. Geogr Geo-Inf Sci 5:89–94 (**(in Chinese)**)

Galdames FJ, Perez CA, Estévez PA, Adams M (2019) Rock lithological classification by hyperspectral, range 3D and color images. Chemom Intell Lab Syst 189:138–148

Gotlieb CC, Kreyszig H (1990) Texture descriptors based on co-occurrence matrices. Comput vis Gr Image Process 51(1):70–86

Guo Q, Zhou Y, Cao S, Qiu Z, Xu Z, Zhang Y (2010) Study on mineralogy of Guangning jade. Acta Sci Natur Univ Sunyatseni 49(3):146–151

Han S, Wang S, Tang Z, Tan K, Duan X, He H, Feng Z, Xie Y (2020) Integrated geophysical exploration of the coupling of a concealed rock body and metallogenic structures—Ag–Pb–Zn mining area case study in Jilinbaolige, Inner Mongolia, China. J Appl Geophys 178:104048

Iwaszenko S, Smoliński A (2021) Texture features for bulk rock material grain boundary segmentation. J King Saud Univ Eng Sci 33(2):95–103

Jing LJ, Li JB, Zhang N, Chen S, Yang C, Cao HB (2021) A TBM advance rate prediction method considering the effects of operating factors. Tunn Undergr Sp Technol 107:103620

Julesz B (1962) Visual pattern discrimination. IRE Trans Inf Theory 8(2):84–92

Julesz B (1981) Textons, the elements of texture perception, and their interactions. Nature 290(5802):91–97

Kereszturi A (2011) Geologic field work on Mars: distance and time issues during surface exploration. Acta Astronaut 68(11):1686–1701

Kim T, Hong K, Byun H (2021) The feature generator of hard negative samples for fine-grained image recognition. Neurocomputing 439:374–382

LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD (1989) Backpropagation applied to handwritten zip code recognition. Neural Comput 1(4):541–551

Lepisto L, Kunttu I, Autio J, Visa A (2003) Classification method for colored natural textures using gabor filtering. In: 12th International conference on image analysis and processing, 2003. Proceedings, pp 397–401

Leung T, Malik J (2001) Representing and recognizing the visual appearance of materials using three-dimensional textons. Int J Comput vis 43(1):29–44

Li MC, Fu JK, Zhang Y, Liu C (2020a) Intelligent recognition and analysis method of rock lithology classification based on coupled rock images and hammering audios. Chin J Rock Mech Eng 39(5):996–1004 (**(in Chinese)**)

Li N, Hao H, Jiang Z, Jiang F, Guo R, Gu Q, Hu X (2020b) A multi-task multi-class learning method for automatic identification of heavy minerals from river sand. Comput Geosci 135:104403

Liu Y, Cheng G, Ma W, Guo C (2016) Rock classification based on features form color space and morphological gradient of rock thin section image. J Cent South Univ (science and Technology) 47:2375–2382

Purswani P, Karpyn ZT, Enab K, Xue Y, Huang X (2020) Evaluation of image segmentation techniques for image-based rock property estimation. J Petrol Sci Eng 195:107890

Puzyrev V, Elders C (2020) Unsupervised seismic facies classification using deep convolutional autoencoder. arXiv preprint, arXiv:.01995. https://arxiv.org/abs/2008.01995. Accessed 1 Oct 2021

Sehsah O, Ghazzawi A, Vie G, Al-Tajar T, Ali A, Al-Mohammed A, Itani M, Ullah S, Escalera H, Balka M (2017) Intelligent drilling system: expanding the envelope of wired drill pipe. In: Abu Dhabi international petroleum exhibition & conference

Tchokpon KG, Kaki C, Kourouma M, Yalo N (2020) Detection of gold-bearing quartz veins in the meta-sedimentary formation in the North-Eastern Guinea using remote sensing and geophysical exploration. J Afr Earth Sci 168:103869

Thompson D, Niekum S, Smith T, Wettergreen D (2005a) Automatic detection and classification of features of geologic interest. In: 2005 IEEE aerospace conference, pp 366–377

Thompson DR, Smith T, Wettergreen D (2005b) Data mining during rover traverse: from images to geologic signatures. In: 8th International symposium on artificial intelligence, robotics and automation in space, pp 31–32

Tuceryan M, Jain AK (1998) Texture analysis. In: Handbook of pattern recognition and computer vision, 2nd edn. World Scientific Publishing, Singapore, pp 235–276. https://doi.org/10.1142/9789814343138_0010

Valentín MB, Bom CR, Coelho JM, Correia MD, Márcio P, Marcelo P, Faria EL (2019) A deep residual convolutional neural network for automatic lithological facies identification in Brazilian pre-salt oilfield wellbore image logs. J Petrol Sci Eng 179:474–503

Varma M, Zisserman A (2005) A statistical approach to texture classification from single images. Int J Comput vis 62(1–2):61–81

Wang C, Li Y, Fan G, Chen F, Wang W (2018a) Quick recognition of rock images for mobile applications. J Eng Sci Technol Rev 11(4):111–117

Wang Z, Liu X, Fu Y, Yuan W, Miao L (2018b) Comprehensive screening method of "ultrasonic-rebound-density" for two kinds of rock specimens. Chin J Rock Mech Eng 37:3575–3583

Xia Y, Li Q, Deng C, Yao J, Long B (2021) Rapid identification method for lithology of tunnel based on lightweight model. J Southw Jiaotong Univ 56(2):420–427

Xu G, Zhu X, Fu D, Dong J, Xiao X (2017) Automatic land cover classification of geo-tagged field photos by deep learning. Environ Model Softw 91:127–134

Xu Z, Shi H, Lin P, Liu T (2021) Integrated lithology identification based on images and elemental data from rocks. J Petrol Sci Eng 205:108853

Yan D, Xu H, Xu Z, Lei Z, Tian M, Cheng L, Ma Y, Wang Z, Ostadhassan M (2020) Sedimentary architecture of hyperpycnal flow deposits: cretaceous Sangyuan outcrop, from the Luanping Basin, North East China. Mar Pet Geol 121:104593

Zhang Y, Wang G, Li M, Han S (2018) Automated classification analysis of geological structures based on images data and deep learning model. Appl Sci 8(12):2493

Zhao T (2018) Seismic facies classification using different deep convolutional neural networks. In: SEG Technical program expanded abstracts 2018. Society of Exploration Geophysicists, pp 2046–2050

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