Full Length Article

Deep learning of rock microscopic images for intelligent lithology identification: Neural network comparison and selection

Zhenhao Xu [a,b,*], Wen Ma [a], Peng Lin [a], Yilei Hua [c]

[a] Geotechnical and Structural Engineering Research Center, Shandong University, Jinan, 250061, China
[b] School of Qilu Transportation, Shandong University, Jinan, 250061, China
[c] Institute of Robotics and Intelligent Systems, Wuhan University of Science and Technology, Wuhan, 430000, China

ARTICLE INFO

ABSTRACT

An intelligent lithology identification method is proposed based on deep learning of the rock microscopic images. Based on the characteristics of rock images in the dataset, we used Xception, MobileNet_v2, Inception_ResNet_v2, Inception_v3, Densenet121, ResNet101_v2, and ResNet-101 to develop microscopic image classification models, and then the network structures of seven different convolutional neural networks (CNNs) were compared. It shows that the multi-layer representation of rock features can be represented through convolution structures, thus better feature robustness can be achieved. For the loss function, cross-entropy is used to back propagate the weight parameters layer by layer, and the accuracy of the network is improved by frequent iterative training. We expanded a self-built dataset by using transfer learning and data augmentation. Next, accuracy (acc) and frames per second (fps) were used as the evaluation indexes to assess the accuracy and speed of model identification. The results show that the Xception-based model has the optimum performance, with an accuracy of 97.66% in the training dataset and 98.65% in the testing dataset. Furthermore, the fps of the model is 50.76, and the model is feasible to deploy under different hardware conditions and meets the requirements of rapid lithology identification. This proposed method is proved to be robust and versatile in generalization performance, and it is suitable for both geologists and engineers to identify lithology quickly.

1. Introduction

Lithology identification is the basis for stratigraphy analysis, resource/reserve estimation, and geological modeling. It can provide specific information about the adverse geological characteristics of the engineering area (Xu et al., 2021a; Lin et al., 2022), and also provide evidence for geohazards prevention and mitigation (Martinez-Martinez et al., 2017). Lithology identification is an important and fundamental indicator in geology, geotechnical investigation, tunneling and underground engineering (Kearsey et al., 2015; Kumar et al., 2019; Xu et al., 2021b). Rapid and accurate identification of lithology has important engineering applications. In engineering practices, visual inspection of hand specimens in field is insufficient, thus it is necessary to conduct accurate lithology identification in the laboratory (Bai et al., 2019). Rock lithology can be identified based on rock density, magnetism, conductivity, and elemental content, using scanning electron microscope (SEM), X-ray diffraction (XRD), and electron probe microanalyzer (EPMA) (Izadi et al., 2017). In general, laboratory lithology identification requires high-precision equipment and a specific working environment, and different equipment facilities may generate different types of data (Vaneghi et al., 2021). Since most equipment is costly and the experiment is time-consuming, the thin section identification still remains the main method for lithology identification in current engineering practices.

Thin section identification is a traditional method that uses images to identify mineral lithology. Rock samples are cut into thin slices, and the crystallization characteristics of minerals are then observed under a polarizing microscope by the geologists. The mineral composition of rocks is determined by measuring their optical properties. Rock type, its genetic characteristics, and lithology can be determined based on the rock structure, rock fabric, and mineral sequence. Compared with the experimental analysis, the entire identification process is time- and cost-effective. However, due to the strong subjectivity of results and high requirements

* Corresponding author. Geotechnical and Structural Engineering Research Center, Shandong University, Jinan, 250061, China.
E-mail address: zhenhao_xu@sdu.edu.cn (Z. Xu).
Peer review under responsibility of Institute of Rock and Soil Mechanics, Chinese Academy of Sciences.

for researchers (Fan et al., 2020), it is often challenging to identify lithology (de Lima et al., 2020). If intelligent lithology identification can be achieved, it can not only reduce the workload of researchers but also enable more practitioners to achieve efficient and objective identification results.

With the rapid development in computer vision technology in recent years (Duan et al., 2021; Isleyen et al., 2021), significant progress has been made in the automatic identification of rock microscopic images. At present, studies on automatic rock image identification are primarily focused on image analysis and feature extraction. According to the characteristics of rock texture, rock fabric and particle distribution, feature extraction based on image processing technology and lithology identification based on machine learning methods have been widely used. For instance, Khorram et al. (2017) proposed a vision-based rock type and classification algorithm based on images of samples collected from a limestone mine. The support vector machine (SVM) and Bayesian techniques were used for classification, enabling the classification of lithology in different stages of mining. Mlynarczuk et al. (2013) used a polarization microscope to obtain digital images from thin sections of nine types of rock samples. Four pattern-identification methods (nearest neighbor, K-nearest neighbor, nearest mode, and optimal spherical neighborhood) were used to automatically identify rock samples. Singh et al. (2010) proposed a texture identification method based on image processing of different basalt thin sections. In their method, the red-green-blue color mode (RGB) or grayscale images of rock samples were used as inputs, and the estimated rock texture categories were outputs, which were provided by the multi-layer perceptron neural networks.

The aforementioned machine learning methods can greatly reduce the subjectivity in lithology identification. However, machine learning based lithology identification still requires manual image processing. To improve the automation of the whole identification process and reduce the difficulty of image processing, deep learning has been gradually applied to intelligent lithology identification based on rock microscopic images. For example, Polat et al. (2021) used a transfer learning model based on Densenet121 and ResNet 50 to extract the features of microscopic images of volcanic rocks to achieve a rapid and intelligent identification of six types of volcanic rocks. Bai et al. (2019) proposed a model of rock microscopic image classification based on visual geometry group (VGG) to identify six common rock types, such as andesite, dolomite, and granite. Because of the diversity and complexity of rock features, many factors need to be considered when selecting a deep learning model.

In previous studies, building a single model is often insufficient to explain the best effect of convolutional neural network (CNN) in image identification. When selecting a CNN, it is necessary to consider not only the identification accuracy but also the identification speed and portability of the proposed model. Considering the diversity and abstraction of rock characteristics, we choose seven kinds of CNNs commonly used in the image classification area. The accuracy of the networks is improved through iterative training. By loading the pre-trained model, the convergence speed is improved, and the training difficulty is reduced. Accuracy (acc) and frames per second (fps) are used as the evaluation indices for assessing model identification accuracy and speed, respectively. The well-trained classification models have good lithology identification ability on the rock microscopic data.

2. Comparison and selection of neural network for lithology identification

We used the CNN to classify rock microscopic images. The main process is to give an input image and use a deep learning model to assign it with a label of a known mixed type. The input is a collection of images under the microscope, and the label of each image is one of the rock types. The training dataset is used to learn different features of each type and then generate a microscopic image classification model. The classifier in the model is used to predict images in the testing dataset, and the labels predicted by the classifier are compared against the ground-truth labels to evaluate the quality of the model. In general, different CNNs have different effects on different datasets. The selection of the CNN is very important for lithology identification. By comparing the structure and design characteristics of commonly used CNNs, the best-performing networks in rock microscopic image datasets are selected to build the classification model.

2.1. Comparison of different neural networks

By designing appropriate network structures, the performance of CNNs can be improved. Increasing the image resolution can allow for more information to be added to the network. Increasing the width and depth of the network can enable the network to learn more parameters. Adding skip connections can increase the complexity of the network and therefore improve the representation of the network. Different network designs will bring different benefits and will also have different effects on model parameters and identification speed.

Prior to the series of inception networks, most popular CNNs simply stack the convolutional layers to obtain better performance by using deeper networks. The model requires to be transplantable, suitable for lithology identification in practical conditions, and can be used in different hardware environments. The inception model is designed to build a network with an excellent local topology structure with fewer parameters. Specifically, multiple convolution and pooling operations are performed on the input image in parallel, and then all output results are spliced into a very deep feature map (Szegedy et al., 2016a). The continuous improvement in the inception series network has led to a variety of network versions. Each version is an iterative evolution of the previous version. Choosing the appropriate version helps optimize the speed and accuracy. In this work, we selected Inception_v3 and Inception ResNet_v2 (Szegedy et al., 2016b) for comparison.

In addition to the inception network, the residual structure in ResNet is also a very important network design in the development of CNNs. Empirically, the impact of network depth on model performance is very important. As the number of network layers increases, the network can extract more complex features. However, experiments show that when the network is too deep, the accuracy will not continue to improve. As a result, residual learning is used to address the problem, whereby the residual unit is added through shortcut connections (He et al., 2016a). Due to the complexity and abstraction of rock characteristics, we selected ResNet-101 and ResNet101_v2 (He et al., 2016b) for comparison. Experiments show that the structure of ResNet101_v2 is superior to that of ResNet-101.

The above two design ideas for a CNN are: (1) deepening the network (such as ResNet), and (2) widening the network (such as the Inception network). Densenet starts with features and achieves better results through the extreme use of features (Huang et al., 2016). The narrower network structure and fewer parameters of Densenet are largely due to the design of dense blocks. Through dense connections, the transmission of features and gradients is more efficient, helping reduce the disappearance of gradients.

When improving the network performance, we also needed to consider the practicability of the network. Under different hardware conditions, there are many limitations in computing performance and storage space, and there are also high requirements for computational speed. Therefore, it is important to achieve

lightweight while ensuring network accuracy. We selected two lightweight networks, i.e. MobileNet v2 and Xception (Chollet, 2017). MobileNet v2 abandons the conventional convolution operations and introduces the depthwise separable convolution as the basic unit. Experiments show that its overall effect is equivalent to a standard convolution, but can greatly reduce the computation and the number of model parameters, which is beneficial to reduce the time and space complexity of convolution (Howard et al., 2017). Xception adopts the depthwise separable convolution similar to MobileNet (the specific structure is described in the following section), which not only makes full use of hardware resources but also maximizes the efficiency and performance of the network.

By comparing the effects of different CNNs, Xception is selected as the backbone network of the microscopic image classification model for intelligent lithology identification.

2.2. Network selection and evaluation

To evaluate the performance of different microscopic image classification models, acc represents the evaluation index for accuracy, fps represents the evaluation index for speed, and the confusion matrix is used as the evaluation index to describe the specific situation of microscopic image identification for different rock types.

Since acc is generally used to evaluate the global accuracy of a model, it is necessary to use the confusion matrix to comprehensively evaluate the model classification performance on individual categories. The x-coordinate in the confusion matrix represents the statistical quantity of categories predicted by the model, and the y-coordinate represents the statistics of the quantity of real labels. Diagonals represent the probability of labels predicted by the model that are consistent with the ground-truth labels. The larger the diagonal value, the better the identification result from the model on this type of rock (which is denoted by the darker color in the visualization results). Off-diagonal values represent the probability of misprediction for other types of rocks, i.e. the lower value indicates better a prediction result. True positive (TP) indicates the number of positive samples correctly identified as positive; true negative (TN) indicates the number of negative samples correctly identified as negative; false positive (FP) indicates the number of negative samples incorrectly identified as positive; and false negative (FN) indicates the number of positive samples incorrectly identified as negative. Then, acc can be expressed as

$$acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

In addition to the detection accuracy, another important performance index for the model evaluation is computational speed. Rapid identification improves the model's efficiency in engineering applications. The comparison of fps needs to be done under the same hardware condition. The larger the fps value, the faster the speed of identification.

3. Xception-based intelligent lithology identification

On the basis of deep learning of rock microscopic images, an intelligent lithology identification method is proposed. Rock microscopic images are divided into training dataset and testing dataset. First, The Xception-based microscopic image classification model is used to directly extract advanced features through convolution operation and pooling operation as the inputs to the full connection layer, and then use these features to classify input images based on the training dataset. Then, in the training, the transfer learning method is used to improve the learning ability of

rock characteristics by loading the pre-trained weights. Finally, this method is verified using the testing dataset to achieve the intelligent lithology identification.

3.1. Microscopic image classification model based on deep learning

When deep learning is used to classify rock microscopic images, rock features are often abstract representations of deep features (Xu et al., 2021c). CNN has become the dominant deep learning method to extract deep features (Men et al., 2017). Image identification using CNN typically consists of four operations, i.e. convolution, nonlinear processing, pooling, and classification. The purpose of convolution is to extract features from the input images. The feature map can be obtained by sliding the filter over the image. For the same input image, different convolution operations generate different feature maps.

As shown in Fig. 1, the lightweight deep learning model based on the Xception architecture contains 36 convolution layers, divided into the entry flow, middle flow, and exit flow. The entry flow contains 8 convolution layers, the middle flow contains 24 convolution layers, and the exit flow contains 4 convolution layers (Chollet, 2017). Xception is combined with deep separable convolutions to learn deep features from a small fraction of data in the image and preserve the spatial relationships between pixels. Different from other lightweight networks, the function of Xception is not to compress the model, but to improve the performance. Because it expands the network with an equivalent number of parameters with Inception_v3. Therefore, the Xception-based microscopic image classification model not only makes full use of hardware resources but also maximizes the network efficiency and performance, thus extracting richer rock features.

The main structure of Xception is a block stack containing a residual network and separable conv Fig. 2 shows the structure of two common blocks in Xception (see Fig. 1), mainly in the entry flow and exit flow. In conventional convolution, the convolution kernel is usually responsible for both channel and spatial relationship mapping. The separable conv module of Xception draws the idea from the depth separable convolution, which separates the two relationship mappings to make the entire convolution process simpler and more efficient. Depth separable convolution includes two operations: depthwise convolution and pointwise convolution. The depthwise convolution performs the first convolution operation on a two-dimensional plane, after which three feature maps are generated. The number of feature maps after the depthwise convolution is the same as the number of channels in the input layer, and thus the feature maps cannot be extended. In addition, the convolution operation of each channel is independent, and the feature information of different channels in the same spatial position is not used effectively. Therefore, the pointwise convolution is used to weight the feature maps of the previous steps in the depth direction and combine them to generate a new feature map. When the inputs are the same and the number of feature maps is the same, the number of depth separable convolution parameters is approximately 1/3 that of conventional convolution. Therefore, the number of CNN layers with depth separable convolution can reach deeper under the premise of the same number of parameters.

As shown in Fig. 2, the residual network is introduced into the Xception block for operation. This depth separable convolution structure with residual connection is easy to define and modify, speeding up training and improving overall performance. The ReLU nonlinear activation function is set after each operation in Xception. Experiments show that using ReLU as an activation function can avoid the disappearance of gradients in back propagation. At the same time, part of the output will be 0 after the ReLU operation. By forming a sparse network, the interdependence of parameters
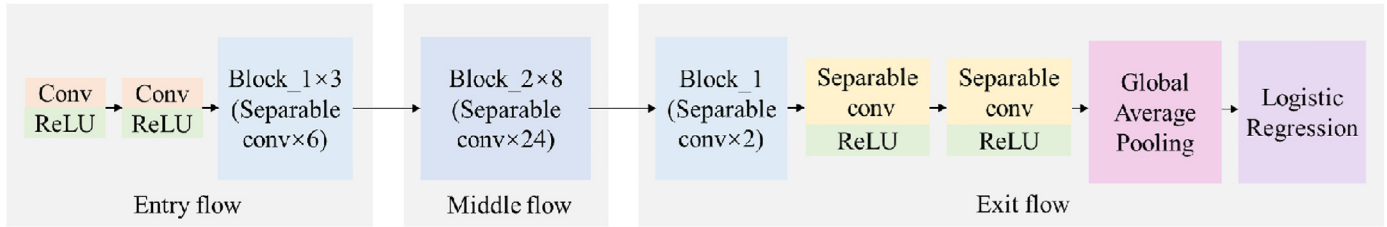
Fig. 1. The network architecture of microscopic image classification model based on Xception.

can be reduced and the overfitting issue can be alleviated. Compared to the sigmoid and tanh activation functions, the derivation of the ReLU function is simpler and computational more efficient (Xu et al., 2021d). Therefore, ReLU is used as the activation function to extract rock features in our classification models, which can be expressed as

$$f(\boldsymbol{x}) = \max(0, \boldsymbol{x}) \tag{2}$$

where $f(\boldsymbol{x})$ represents the activation function, i.e. a function that maximizes the input value $\boldsymbol{x}$ of a neuron.

The output of the CNN is used as the input to the full connection layer and is trained by back propagation. The full connection layer is a traditional multi-layer perceptron. The high-level features of the input image can be classified by the full connection layer. Adding a full connection layer is also an easy way to learn the nonlinear combination of these features. Finally, SoftMax is used as the activation function in the output layer, which converts any input vector greater than 0 into a numeric vector between 0 and 1, with the sum of the output probabilities obtained from the full connection layer being 1. Experiments show that the most advanced features are trained in combination, which is more effective for classification tasks. Taking the output of the i th node as an example, the SoftMax function can be expressed as

$$z_i = \log_{10} \widehat{P}(y = i | \boldsymbol{x}) \tag{3}$$

$$Softmax(z_i) = \frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}} \ (i = 1, 2, \ldots, k) \tag{4}$$

where $\boldsymbol{x}$ is the input vector; z_i is the output probability of the i th node, and k is the number of output nodes, that is, the number of types classified.

3.2. Network training

Compared with vehicle or face identification, there are some special issues with rock microscopic image identification:

(1) Complex data acquisition. Rocks are formed in fundamentally different ways and have clearly different physical and chemical characteristics. There are three main types of rocks: sedimentary, igneous, and metamorphic. Within each of these types, these are many classes of rocks formed by physical changes such as melting, cooling, eroding, compacting, or deforming. Collecting a large number and variety of samples manually is a time- and labor-consuming long-term task.
(2) Complex feature representation. Different rocks have different features because of their minerals, the ways that the rocks were formed, and the geological processes that act on them since they are formed. The thin sections of rocks have complex mineral crystallization and optical properties under the microscope. It is necessary to determine the mineral composition of a rock and study its structure and fabric. Also we should analyze the generation sequence and genetic characteristics of minerals. Therefore, the microscopic features of a rock cannot be described only by simple features such as contours and colors.

Transfer learning and a large amount of source domain information are used to improve the prediction performance of the training model in the target domain. The source domain refers to as the set of annotated instances, and the target domain the set of instances to be annotated. These two domains have different feature spaces. Loading the ImageNet pre-trained model can improve the convergence speed of gradient descent, and obtain a model with a low generalization error. The pre-trained model has a higher performance and faster training speed, which also reduces the gradient disappearance or gradient explosion problem caused by no initialization or improper initialization.

In order to fully use the dataset of rock microscopic images, data augmentation operations such as random scaling, flipping, and rotation are used in training to improve the learning ability of rock features. The best model is verified through the testing dataset.

During the model training, we first initialized all filters, loaded the pre-trained model, and set parameters and weights with random values. Random initialization can break the symmetry of the data so that different hidden units can learn different information. The convolution network receives the training images as the inputs and obtains different types of output probabilities through the forward propagation (including convolution, ReLU, pooling operations, and forward propagation of the full connection
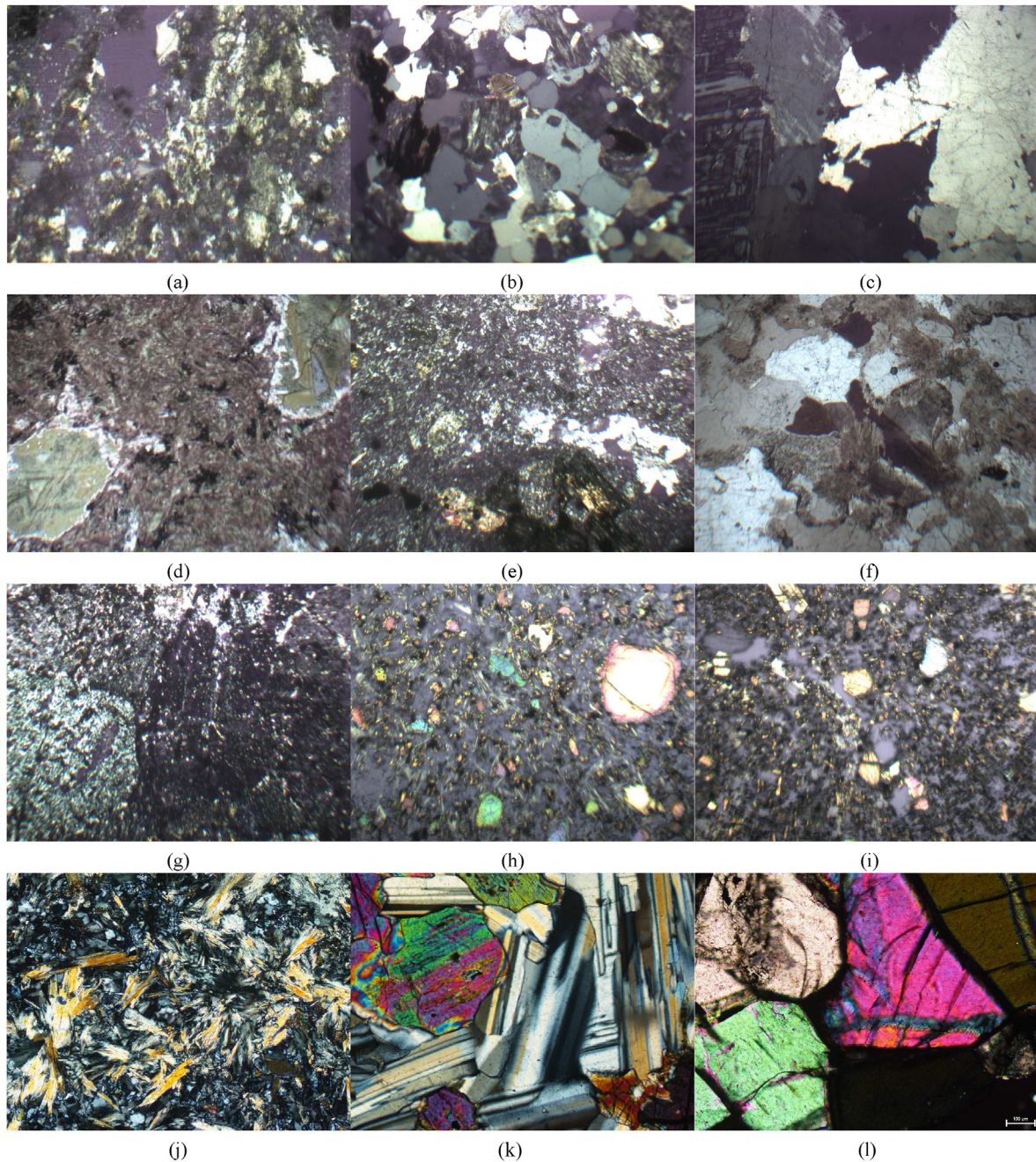

Fig. 2. The Block_1 and Block_2 in Xception: (a) Block_1 and (b) Block_2.

Fig. 3. Examples of typical igneous rock microscopic images in the dataset: (a) Rhyolite; (b) Granite; (c) Granite pegmatite; (d) Andesite; (e) Diorite-porphyry; (f) Syenite; (g) Anorthosite; (h) Stomatal basalt; (i) Amygdaloidal basalt; (j) Diabase; (k) Gabbro and (l) Peridotite.

layer). The total error is calculated at the output layer, and the back propagation is used to calculate the error gradient based on the weight of the network. The gradient descent algorithm updates the values, weights, and parameters of all filters to minimize output errors. We repeated the steps above for all images in the training dataset. The update step is obtained by calculating the adaptive learning rate for each parameter by using the Adam optimizer.

Forward propagation is the process of using a SoftMax classifier to calculate the probability score and obtain the corresponding loss function. After obtaining the loss function, the microscopic image classification model is optimized according to the loss function to lower the loss value. We used cross-entropy as the loss function. Cross-entropy measures the difference between two different probability distributions for the same random variable. In deep learning, it represents the difference between the real probability distribution and the predicted probability distribution. The smaller the cross-entropy, the better the model predicts. Usually, SoftMax is used to process outputs, and the sum of the predicted values of multiple classifications is 1. Cross-entropy is used to calculate the loss, which can be expressed as

$$L = -\sum_{i=1}^{n} p_i \ln p_i' \tag{5}$$

where i is the index of outputs, L is the loss, p_i is the actual probability distribution, and p_i' is the predicted probability distribution.
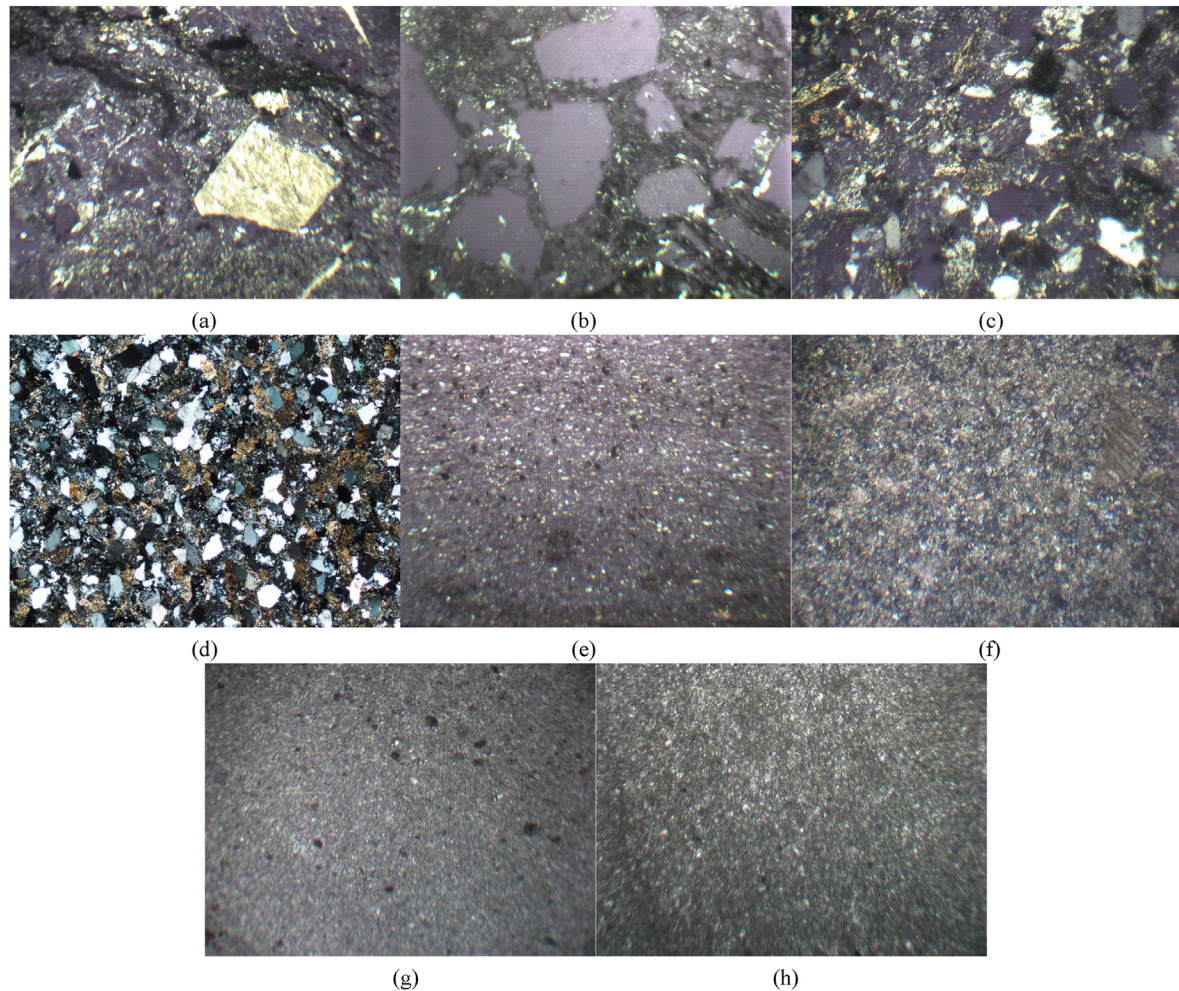
Fig. 4. Examples of typical sedimentary rock microscopic images in the dataset: (a) Volcanic breccia; (b) Tuff; (c) Conglomerate; (d) Siltstone; (e) Shale; (f) Limestone; (g) Wormkalk and (h) Dolomite.

4. Case study and verification

To verify the fidelity of this method and model, 30 types of rock microscopic images are selected for tests, and lithology identification is conducted by using the microscopic image classification model based on Xception. To compare the performance of different CNNs in rock microscopic image datasets, ResNet101_v2, Mobile-Net_v2, Inception_ResNet_v2, Inception_v3, Densenet121, and ResNet-101 are used to establish microscopic image classification models for comparative tests.

4.1. Dataset

We used a total of 30 rock types from three categories as the dataset, which are common and representative. In our dataset, 12 types of igneous rocks are classified as Category A, including rhyolite, granite, granite pegmatite, andesite, diorite porphyry, syenite, anorthosite, stomatal basalt, amygdaloidal basalt, diabase, gabbro, and peridotite. Eight types of sedimentary rocks are classified as Category B, including volcanic breccia, tuff, conglomerate, siltstone, shale, limestone, wormkalk, and dolomite (tuff and volcanic breccia are in the form of sedimentary rocks). Ten types of metamorphic rocks are classified as Category C, including black slate, phyllite, granite schist, granite gneiss, garnet gneiss, quartzite, serpentinized marble, greisen, skarn, and striped migmatite.

Samples are cut into thin slices, and then are photographed by an Olympus DP74 polarizing microscope (orthogonal polarizing, magnification: 10 times). Figs. 3−5 show an example of 30 types of rock microscopic images in the dataset. Collecting thin section images from the same rock in different angles can further improve the model's ability to extract rock features.

A total of 14,950 rock microscopic images are used to create the dataset. The images are randomly selected with a ratio of 9:1 between the training dataset and the testing dataset. Specifically, the training dataset has 13,463 images, and the testing dataset has 1487 images. The number of training and testing datasets for each rock type is listed in Table 1. In order to ensure that the model is insensitive to the missing values of samples in training, the weight relationship between different feature factors and corresponding types is established. The uniformly distributed rock microscopic images dataset is beneficial to improving the identification accuracy and generalization ability of the model. Images in the training set are labeled and trained according to the ground-truth labels.

4.2. Network training

Deep learning models are used to identify rock microscopic images. In this paper, seven network models including Xception, ResNet101_v2, MobileNet_v2, Inception_ResNet_v2, Inception_v3,
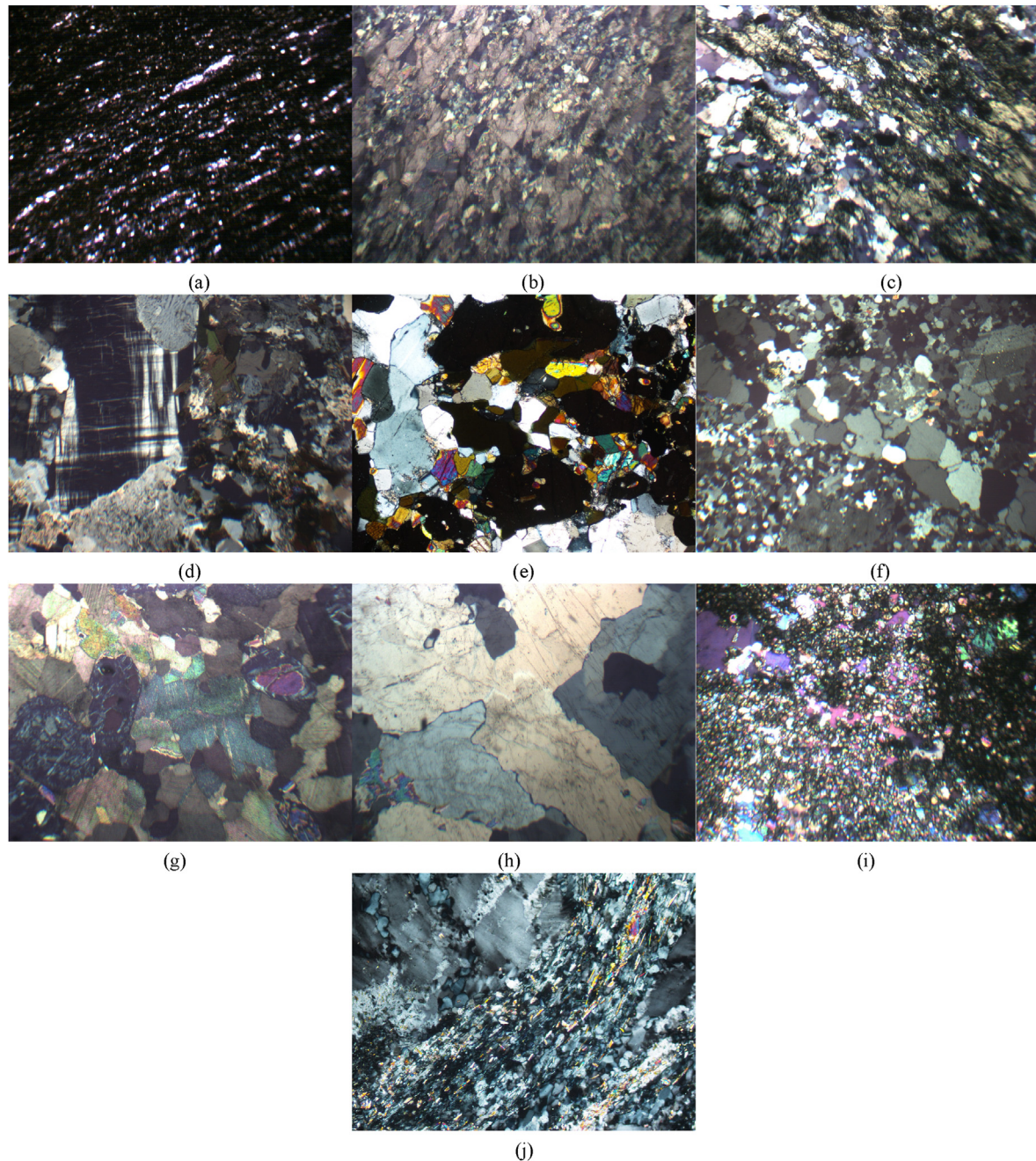
Fig. 5. Examples of typical metamorphic rock microscopic images in the dataset: (a) Black slate; (b) Phyllite; (c) Granite schist; (d) Granite gneiss; (e) Garnet granulite; (f) Quartzite; (g) Serpentinized marble; (h) Greisen; (i) Skarn and (j) Striped migmatite.

Densenet121 and ResNet-101 are used to build classification models, which are trained under the framework of Keras in Python. Cross-entropy is used as the loss function to further optimize the model, and the pre-trained model is loaded to speed up the convergence and improve the accuracy of the model. To better evaluate the seven models, experiments must be performed under the same hardware condition and the model parameters are adjusted to be the same. We used quad core CPU (2.6 GHz), and the NVIDIA geforce GTX 1080 was used as the graphics card. Through many experiments, we set the batch size to be 10, the learning rate to be 0.0001, the learning rate decay to be 0.0001, and the weight decay to be 0.0001. As shown in Fig. 4, after 40 iterations, the losses of all seven models are basically stabilized, indicating that the classification models can extract rock features. In the training dataset, the loss value of Xception,

MobileNet_v2, Inception_ResNnet_v2, Inception_v3, and Densenet121 are stable at about 0.05, which have better convergence compared with ResNet101_v2 and ResNet-101.

As shown in Fig. 6, Xception, Inception_ResNet_v2, Inception_v3, and Densenet121 have relatively modest loss performance in the test set. After 40 iterations, Xception, MobileNet_v2, Inception_ResNet_v2, Inception_v3, and Densenet121 have an acc value of 97.3%–97.6% on the training dataset. This also shows that the model with better convergence yields better results in multi-class image identification. As shown in Figs. 7 and 8 due to the randomness of testing dataset, the performance of the seven models on the test set is relatively volatile. Among them, Xception's acc is relatively stable, after 30–40 rounds of iterations, it remains at a relatively stable value.

Table 1
Datasets for the image classification of rock lithology.

Category	Rock type	Number of images	Training dataset	Testing dataset
A	Rhyolite	440	396	44
	Granite	546	492	54
	Granite pegmatite	527	475	52
	Andesite	477	429	48
	Diorite-porphyrite	493	440	53
	Syenite	480	432	48
	Anorthosite	246	225	21
	Stomatal basalt	541	487	54
	Amygdaloidal basalt	462	416	46
	Diabase	461	415	46
	Gabbro	457	412	45
	Peridotite	504	454	50
B	Volcanic breccia	522	470	52
	Tuff	536	483	53
	Conglomerate	489	441	48
	Siltstone	533	480	53
	Shale	513	462	51
	Limestone	527	475	52
	Wormkalk	474	427	47
	Dolomite	528	476	52
C	Black slate	502	452	50
	Phyllite	523	471	52
	Staurolite schist	530	477	53
	Granitic gneiss	446	402	44
	Garnet granulite	484	436	48
	Quartzite	550	495	55
	Serpentinized marble	531	478	53
	Greisen	527	475	52
	Skarn	541	486	55
	Striped migmatite	560	504	56

4.3. Network evaluation

MobileNet_v2 is the fastest among the seven models, processing 54.76 images per second. Xception ranks second, processing 50.76 images per second. MobileNet_v2 and Xception, as the typical lightweight CNNs, can achieve the purpose of rapid lithology identification. Although MobileNet_v2 is slightly faster than Xception, this is achieved at the expense of model accuracy. Under the premise of ensuring high accuracy, the Xception-based microscopic image classification model can meet the rapid identification of lithology in engineering practices. Although the other five models are relatively slow, in general, deep learning-based rock microscopic image identification has a significant advantage over manual identification.

For model evaluation, the size of the model is also important. When the hardware conditions are not the same, it is necessary to consider the difficulty of deploying on different computing devices. The smaller the model size, the less the calculations and the more scenarios that can be applied. It is more suitable for a wide range of laboratory environments. As shown in Table 2, MobileNet_v2 has the smallest model (only 80.91 MB), followed by Densenet121 (130.36 MB), and Xception (311.56 MB). For models with small sizes, hardware conditions are less demanding, so the portability of the model is stronger. Fig. 10 shows the acc from the Xception-based microscopic image classification model is the highest, which can reach 98.65%. Similarly, acc of ResNet101_v2, Mobile-Net_v2, Inception_ResNet_v2, Inception_v3, Densenet121, and ResNet-101 can reach 94.81%, 96.43%, 97.64%, 97.71%, 97.17%, 91.99%, respectively.

The indicators commonly used in classification problems are used to compare microscopic image classification models. Fig. 9 shows the comparison of the confusion matrices of different
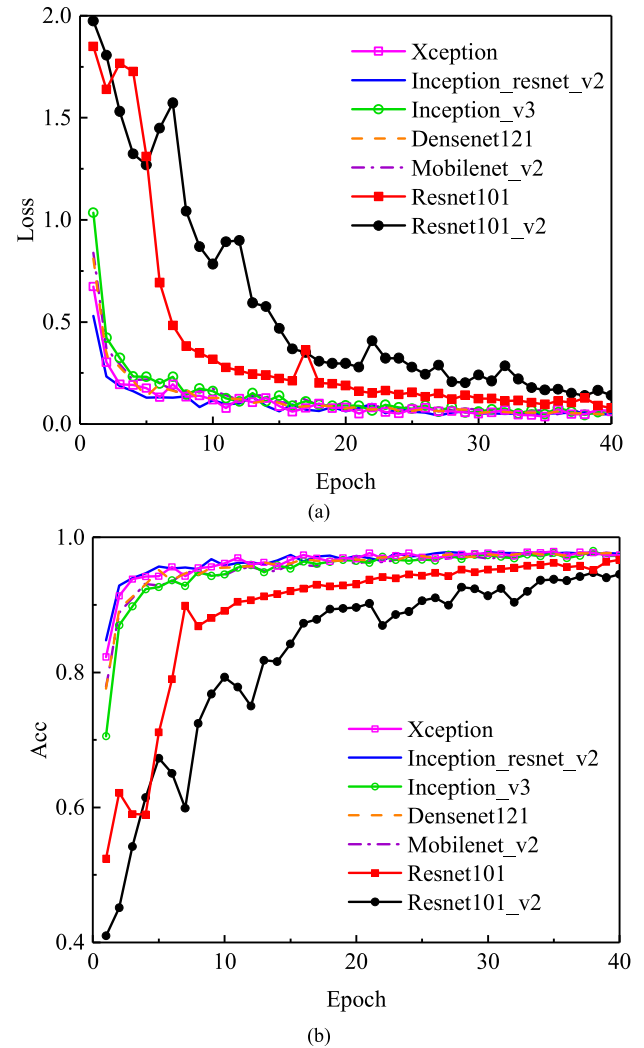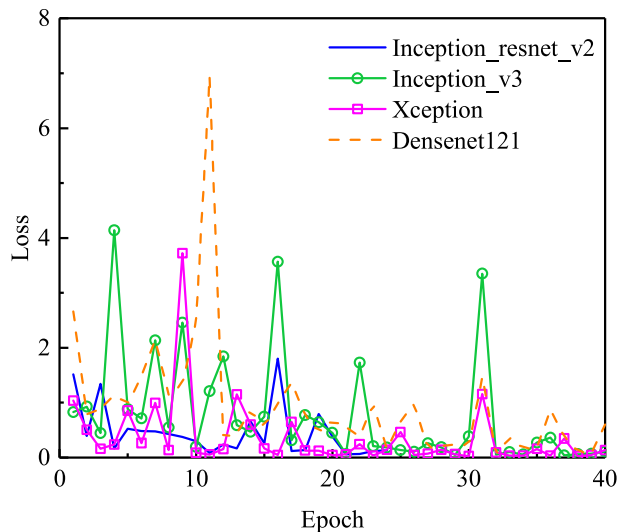

Fig. 6. The convergence loss and accuracy curves in the training dataset: (a) Loss and (b) Accuracy.
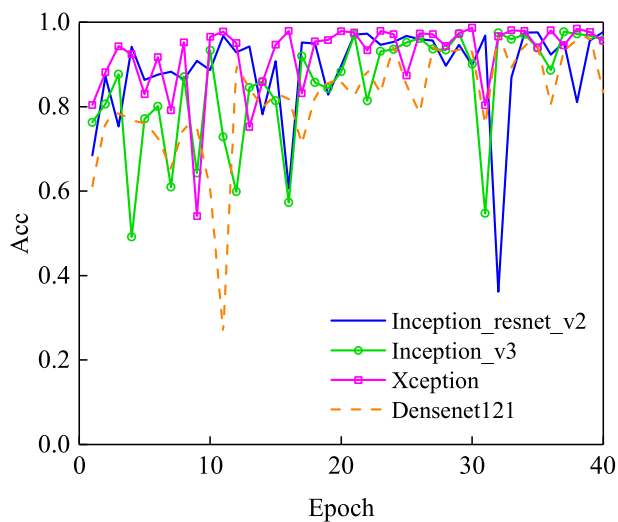
models. As shown in Fig. 9a, The Densenet121-based microscopic image classification model has a 19% probability of identifying diabase as gabbro, and a 36% probability of identifying gabbro as diabase. The probabilities of identifying siltstone as anorthosite, phyllite, and serpentinized marble are 15%, 11%, and 19%, respectively. Similarly, the probabilities of identifying rhyolite as anorthosite and diabase are 9% and 2%, respectively. The identification acc of other rock types is higher than 90%, and 24 rock types out of 30 are identified completely correct.

As shown in Fig. 9b, there is a 27% probability that gabbro will be identified as diabase from the microscopic image classification model based on Inception_ResNet_v2. The probability of identifying gabbro as diabase is 17%, and the probabilities of identifying black slate as volcanic breccia and dolomite are 10% and 2%, respectively. The identification acc of other types of rocks is higher than 90%, and 22 types of rocks are identified completely correct.

Fig. 9c shows that for the microscopic image classification model based on Inception_v3, there is a 25% probability that gabbro is misidentified as diabase. The probability of misidentifying gabbro as diabase is 17%, and the probability of misidentifying conglomerate as siltstone is 12%. The identification acc of other types of rocks is higher than 90%, and 24 types of rocks are identified completely correct.

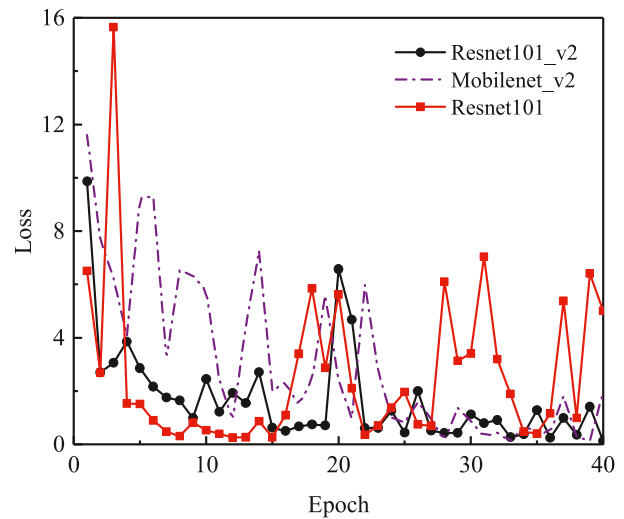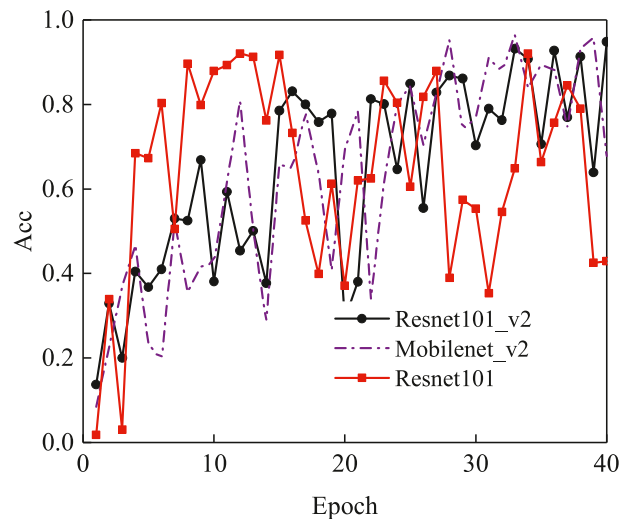**Fig. 7.** The convergence loss and accuracy curves in the testing dataset using Xception, Inception_ResNet_v2, Inception_v3, and Densenet121: (a) Loss and (b) Accuracy.

Fig. 8. The convergence loss and accuracy curves in the testing dataset using Resnet 101_v2, Mobilenet_v2, Resnet101: (a) Loss and (b) Accuracy.

For the microscopic image classification model based on MoblieNet_v2, as shown in Fig. 9d, there is a 40% probability that gabbro is misidentified as diabase. The probabilities of misidentifying black slate as granite pegmatite, andesite, and volcanic breccia are 8%, 2%, and 2%, respectively. The identification acc of other types of rocks is higher than 90%, and 16 types of rocks are identified completely correct.

Fig. 9e shows that, for the ResNet10-based microscopic image classification model, there is a 25% probability that gabbro is misidentified as diabase. The probability of misidentifying gabbro as diabase is 11%. The probabilities of misidentifying granite as syenite, volcanic breccia, and garnet granulite are 2%, 2%, and 13%, respectively. The probabilities of misidentifying granite pegmatites as syenite, volcanic breccia, and garnet breccia are 2%, 2%, and 13%, respectively. The probabilities of misidentifying andesite as amygdaloidal basalt, shale, dolomite, serpentinized marble, and greisen are 6%, 2%, 30%, 2%, and 2%, respectively. The probabilities of misidentifying anorthosite as diorite porphyry, and garnet granulite

Table 2
Test result of the seven classification models.

Method	fps	Size of model (MB)	Input size	Acc (%)
Resnet 101_v2	24.37	561.32	160 × 160	94.81
Mobilenet_v2	54.76	80.91	224 × 224	96.43
Inception_resnet_v2	10.7	685.09	299 × 299	97.64
Inception_v3	27.25	323.08	224 × 224	97.71
Xception	50.76	311.56	299 × 299	98.65
Densenet121	22.48	130.36	224 × 224	97.17
Resnet-101	25.32	561.74	224 × 224	91.99

are 10% and 2%, respectively. The probabilities of misidentifying volcanic breccia as granite, syenite, black slate, garnet granulite, and quartzite are 2%, 6%, 4%, 2%, and 2%, respectively. The probability of misidentifying volcanic tuff as serpentinized marble is 24%. The probability of misidentifying conglomerate as peridotite or garnet granulite is 6%. The probabilities of identifying wormkalk as shale, garnet granulite, and serpentinized marble are 9%, 5%, and

20%, respectively. The probability of misidentifying serpentinized marble as garnet granulite is 19%. The identification acc of other types of rocks is higher than 90%, and eight types of rocks are identified completely correct.

For the ResNet101_v2-based microscopic image classification model, there is a 21% probability that gabbro is misidentified as diabase, as shown in Fig. 9f. The probability of misidentifying gabbro as diabase is 23%. The probabilities of misidentifying volcanic breccia as granite, syenite, shale, dolomite, black slate, and garnet granulite are 19%, 2%, 2%, 4%, 6%, and 6%, respectively. The probabilities of misidentifying black slate as granite and syenite are 6% and 8%, respectively. The probability of misidentifying serpentinized marble as siltstone is 14%. The identification acc of other types of rocks is higher than 90%, and 10 types of rocks are identified completely correct.

Fig. 9g illustrates that, for the Xception-based microscopic image classification model, there is a 37% probability that gabbro is misidentified as diabase. The identification acc of other types of rocks is higher than 90%, and 23 types of rocks are identified completely correct.

Table 3 shows the average accuracy of microscopic image classification models based on Inception_ResNet_v2, Inception_v3, Densenet121, MobileNet v2, ResNet101 v2, ResNet-101, and Xception. The results show that Xception has the highest average acc and the accuracy for the overall model can reach 98.02%. Models based on Inception_ResNet_v2, Inception_v3, and Densenet121 perform well, with an average accuracy of 96.88%, 96.85%, and 94.75%, respectively. Models based on MobileNet_v2, ResNet101_v2, and ResNet-101 perform modest, with an average accuracy of 90.53%, 88.66%, and 89.18%, respectively. Again, these results show that Xception has the optimum identification ability on the rock microscopic image dataset used in this work among the seven models, with the highest accuracy and stability.

5. Discussion

Because thin section identification usually depends on empirical methods, and experience is difficult to be described in mathematical language. It is difficult to judge which traditional image features play a universal role in lithology identification. Therefore, it is
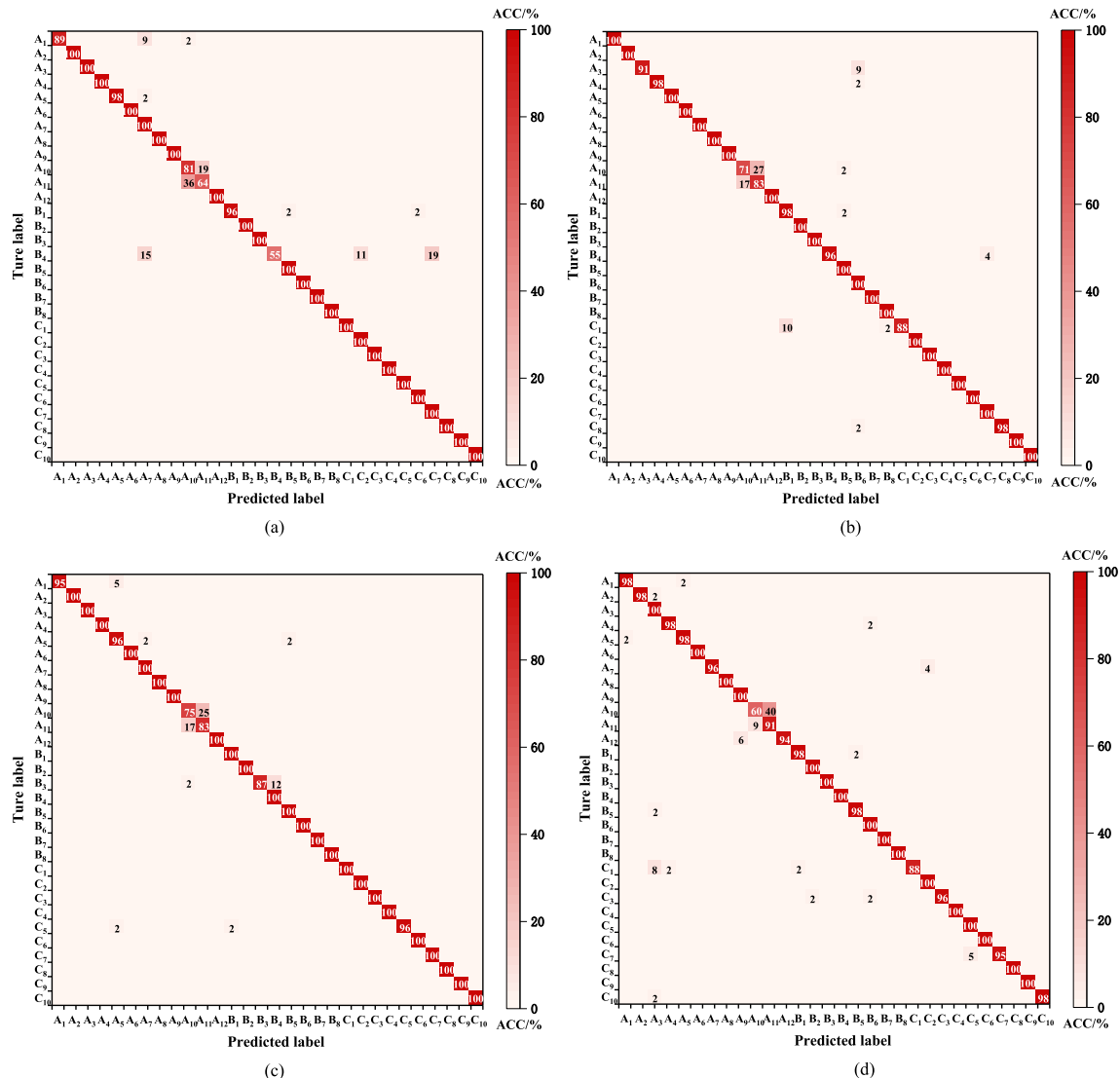

Fig. 9. The confusion matrices from different microscopic image classification models: (a) Densenet121; (b) Inception_ResNet_V2; (c) Inception_v3; (d) MoblieNet_v2; (e) ResNet-101; (f) ResNet101_V2 and (g) Xception.
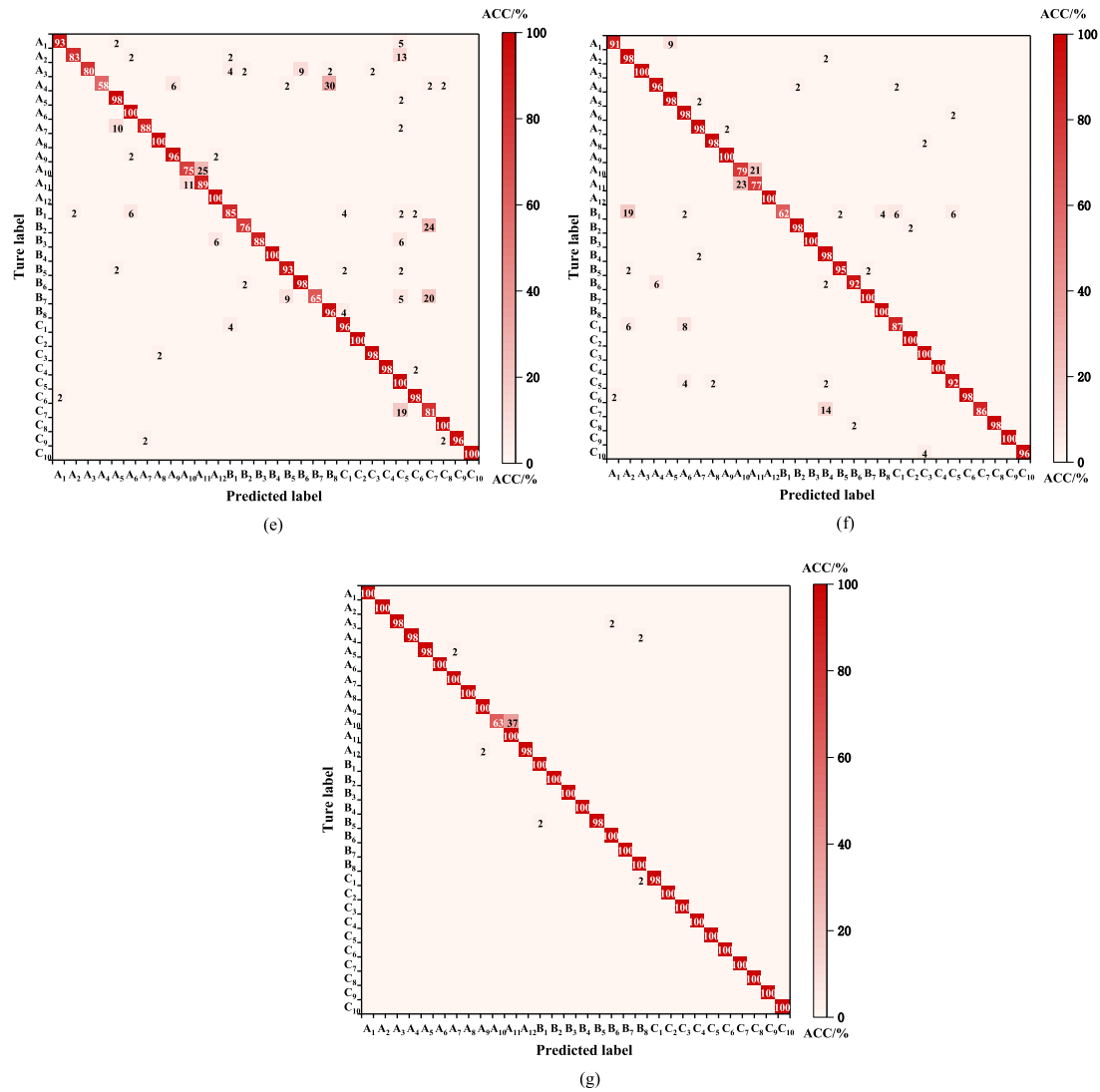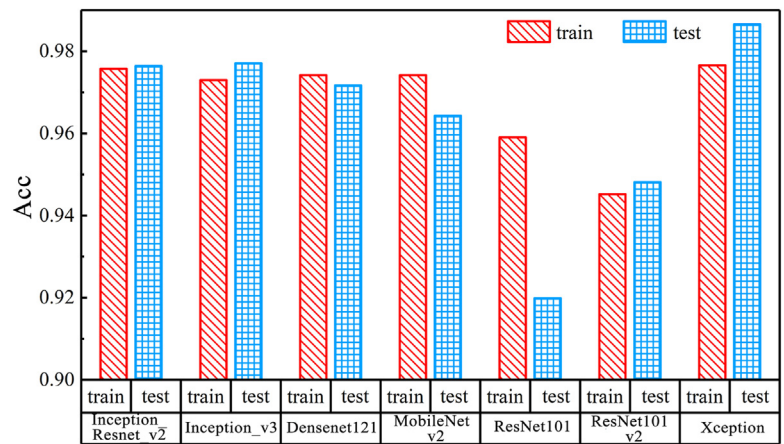
(e)

(f)

(g)

Fig. 9. Continued.

Fig. 10. The maximum accuracy of the lithology identification models.

Table 3
Result of the average accuracy of the seven classification models.

Method	Average of accuracy (%)
Resnet 101_v2	88.66
Mobilenet_v2	90.53
Inception_resnet_v2	96.88
Inception_v3	96.85
Xception	98.02
Densenet121	94.75
Resnet-101	89.18

difficult to use traditional image processing methods to extract and reduce the dimension of microscopic features. Due to the rotation invariance and translation invariance of features, the CNN can adaptively form filters, which are sensitive to various key deep features in the learning process. Moreover, the CNN can fully consider the relationship between the local pixels of the image. Therefore, the CNN can better extract features and carry out downstream classification tasks than traditional methods.

Different CNNs have different performances on different datasets, which requires to consider all aspects simultaneously. In this paper, seven different CNNs are compared under the same hardware condition. The model based on Xception has the highest accuracy, with a fast calculation speed and appropriate scale. Compared with other models, Xception uses depth separable convolution that makes the CNN layer deeper but with the same parameters, thus maximizing the network efficiency and performance. During the training, the transfer learning method is used to speed up the convergence of gradient descent by loading the ImageNet pre-trained model. Finally, the model with low generalization error is obtained.

The identification of constituent minerals is fundamental to rock microscopic image identification. Some rocks contain similar mineral components and thus have similar microscopic manifestations. When using deep learning models for identification, the CNN may be prone to insufficient extraction of features, and the extracted features may not be targeted. These factors can eventually lead to misidentification. For example, diabase and gabbro have a large probability to be misidentified with each other in all seven models. In addition, siltstone, black slate, volcanic breccia, and conglomerate can also be misidentified. The characteristics of rock constituent minerals can be further analyzed in future work. At present, we used common CNNs, and transformer-based deep learning models can be explored in future. In addition, considering the variety of rock types and the different compositions of minerals, the dataset of rock microscopic images can be expanded in future work.

6. Conclusions

We used rock microscopic images as the research object, and combine the image identification technology with lithology identification. The main conclusions can be drawn as follows:

(1) An intelligent lithology identification method is proposed using deep learning of rock microscopic images. The Xception, MobileNet_v2, Inception_ResNet_v2, Inception_v3, Densenet121, ResNet101_v2, and ResNet-101 are used to establish the microscopic image classification models, and rapid intelligent lithology identification can be realized.

(2) In terms of accuracy, the Xception-based model is compared with six other models. The Xception model has the highest accuracy of 98.65% and an average accuracy of 98.02%. In terms of the identification speed and model size, the Xception-based model is of medium size, and the fps of

Xception can reach 50.76, indicating high identification speed.

(3) Transfer learning and data augmentation are used to expand the data set, which helps optimize the training speed. Compared with traditional methods, this method has a better ability to extract features. Without preprocessing images, the lithology identification process can be greatly simplified.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
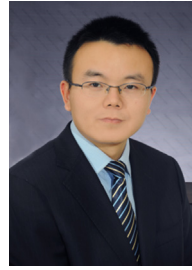
References

Bai, L., Wei, X., Liu, Y., Wu, C.Y., Chen, L.H., 2019. Rock thin section images identification based on VGG model. Geol. Bull. China 38 (12), 2053–2058.
Chollet, F., 2017. Xception: deep learning with depthwise separable convolutions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1800–1807.
de Lima, R.P., Duarte, D., Nicholson, C., Slatt, R., Marfurt, K.J., 2020. Petrographic microfacies classification with deep convolutional neural networks. Comput. Geosci. 142, 11.
Duan, Y., Shen, Y.R., Canbulat, I., Luo, X., Si, G.Y., 2021. Classification of clustered microseismic events in a coal mine using machine learning. J. Rock Mech. Geotech. Eng. 13, 1256–1273.
Fan, G.P., Chen, F.X., Chen, D.Y., Dong, Y.Q., 2020. Recognizing multiple types of rocks quickly and accurately based on lightweight CNNs model. IEEE Access 55269–55278.
He, K., Zhang, X.Y., Ren, S.Q., Sun, J., 2016a. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778.
He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J., 2016b. Identity mappings in deep residual networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016, pp. 630–645.
Howard, A.G., Zhu, M.L., Chen, B., et al., 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, 04861 arXiv: 1704.
Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2016. Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269.
Isleyen, E., Duzgun, S., Carter, R.M., 2021. Interpretable deep learning for roof fall hazard detection in underground mines. J. Rock Mech. Geotech. Eng. 13, 1246–1255.
Izadi, H., Sadri, J., Bayati, M., 2017. An intelligent system for mineral identification in thin sections based on a cascade approach. Comput. Geosci. 99, 37–49.
Kearsey, T., Williams, J., Finlayson, A., et al., 2015. Testing the application and limitation of stochastic simulations to predict the lithology of glacial and fluvial deposits in Central Glasgow, UK. Eng. Geol. 187, 98–112.
Khorram, F., Morshedy, A.H., Memarian, H., Tokhmechi, B., Zadeh, H.S., 2017. Lithological classification and chemical component estimation based on the visual features of crushed rock samples. Arabian J. Geosci. 10 (15), 324.
Kumar, C.V., Vardhan, H., Murthy, C.S.N., Karmakar, N.C., 2019. Estimating rock properties using sound signal dominant frequencies during diamond core drilling operations. J. Rock Mech. Geotech. Eng. 11 (4), 850–859.
Lin, P., Yu, T.F., Xu, Z.H., Shao, R.Q., Wang, W.Y., 2022. Geochemical, mineralogical, and microstructural characteristics of fault rocks and their impact on TBM jamming: a case study. Bull. Eng. Geol. Environ. 81, 64.
Martinez-Martinez, J., Corbi, H., Martin-Rojas, I., Baeza-Carratala, J.F., Giannetti, A., 2017. Stratigraphy, petrophysical characterization and 3D geological modelling of the historical quarry of Nueva Tabarca island (western Mediterranean): implications on heritage conservation. Eng. Geol. 231, 88–99.
Men, K., Dai, J.R., Li, Y.X., 2017. Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. Med. Phys. 44 (12), 6377–6389.
Mlynarczuk, M., Gorszczyk, A., Slipek, B., 2013. The application of pattern recognition in the automatic classification of microscopic rock images. Comput. Geosci. 60, 126–133.
Polat, O., Polat, A., Ekici, T., 2021. Automatic classification of volcanic rocks from thin section images using transfer learning networks. Neural Comput. Appl. 33 (18), 11531–11540.

Singh, N., Singh, T.N., Tiwary, A., Sarkar, K.M., 2010. Textural identification of basaltic rock mass using image processing and neural network. Comput. Geosci. 14 (2), 301–310.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A., 2016a. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning arXiv: 1602.07261.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016b. Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818–2826.

Vaneghi, R.G., Saberhosseini, S.E., Dyskin, A.V., Thoeni, K., Sharifzadeh, M., Sarmadivaleh, M., 2021. Sources of variability in laboratory rock test results. J. Rock Mech. Geotech. Eng. 13 (5), 985–1001.

Xu, Z.H., Liu, F.M., Lin, P., Shao, R.Q., Shi, X.S., 2021a. Non-destructive, in-situ, fast identification of adverse geology in tunnels based on anomalies analysis of element content. Tunn. Undergr. Space Technol. 118, 104146.

Xu, Z.H., Wang, W., Lin, P.L., Nie, C., Li, Z.M., 2021b. Hard-rock TBM jamming subject to adverse geological conditions: influencing factor, hazard mode and a case study of Gaoligongshan Tunnel. Tunn. Undergr. Space Technol. 108, 103683.

Xu, Z.H., Ma, W., Lin, P., Shi, H., Pan, D.D., Liu, T.H., 2021c. Deep learning of rock images for intelligent lithology identification. Comput. Geosci. 154, 104799.

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

Prof. Zhenhao Xu received his BSc degree in Civil Engineering from China University of Geosciences, and a PhD degree in Geological Engineering from Shandong University and the University of Western Australia. After that, he conducted postdoctoral research in Geological Resources and Geological Engineering at Chinese Academy of Geological Sciences, China. He is now a professor at Shandong University, China. His research interests include (1) intelligent identification of lithology and adverse geology, disaster prevention and control in tunneling and underground engineering; (2) numerical simulation of groundwater and grouting process; (3) advanced geological prediction, and water and mud inrush in tunnels; and (4) urban underground space development and utilization.