



Original Paper

Intelligent Lithology Identification Methods for Rock Images Based on Object Detection

Hou Zhenlong ^{1,2,4} Wei Jikang,² Shen Jinrong,² Liu Xinwei,^{2,3} and Zhao Wentian²

Received 6 May 2023; accepted 5 October 2023

Published online: 26 October 2023

Lithology identification is a crucial step in geological research. In recent years, the development of artificial intelligence technologies has provided new insights into solving problems associated with subjectivity and labor intensity of traditional manual identification. However, when rocks are identified in situ, existing algorithms cannot accurately identify them if the image features of different types of rocks are similar or the rock textures are varied. In this regard, the study of lithology identification for the rock images captured from the field was carried out. First, the object detection algorithm of single shot multibox detector was improved by adding residual net and adaptive moment estimation, and a lithology identification model was constructed. Second, based on the above improved algorithm, the technologies of database and geographic information system were combined to develop an integrated identification method. Third, the proposed methods were applied to 12 types of rocks in Xingcheng area, China, for testing their validity, and feasibility in field geological surveys. Finally, the effects of learning rate and batch size on the identification were discussed, as the epoch number was increased. We found that the average accuracies of the improved single shot multibox detector and integrated method were 89.4% and 98.4%, respectively. The maximum accuracy could even reach 100%. The identification results were evaluated based on accuracy, precision, recall, F₁-score, and mean average precision. It was demonstrated that the integrated method has a strong identification ability compared with other neural network methods. Generally, a small learning rate can lead to low loss and high accuracy, whereas a small batch size can lead to high loss and high accuracy. Moreover, the newly proposed methods helped to improve the lithology identification accuracy in the field and support the study of intelligent in situ identification for rock images.

KEY WORDS: Lithology identification, Rock image, SSD, Database, GIS.

INTRODUCTION

Lithology identification and classification always play important roles in regional geological surveys, geotechnical engineering, and resource exploration (Pour et al., 2018; Kumar et al., 2019; Xu et al., 2021). Lithology identification can be implemented through the visual observation of rock specimens and thin or polished sections. Methodologies associated with geophysics (Bosch and McGaughey, 2001; Asfahani et al., 2015) and geo-

¹Key Laboratory of Ministry of Education on Safe Mining of Deep Metal Mines, Northeastern University, Shenyang 110819, Liaoning, China.

²School of Resources and Civil Engineering, Northeastern University, Shenyang 110819, Liaoning, China.

³School of Earth Sciences and Engineering, Sun Yat-sen University, Zhuhai 519080, Guangdong, China.

⁴To whom correspondence should be addressed; e-mail: houzlatjlu@163.com, houzhenlong@mail.neu.edu.cn

chemistry (Han et al., 2018; Gleeson et al., 2020; Cheng et al., 2022) can also be used for lithology identification. However, specialized geological knowledge and experience are indispensable for traditional identification. When the number of rock samples is large, the accuracy is inevitably decreased due to human subjectivity. Hence, the study of automated and intelligent methods is important for effective lithology identification and classification. This study also makes it more convenient for scholars and technicians without petrology background to conduct lithology identification.

Artificial intelligence (AI) has developed rapidly in recent years. Image recognition technologies using artificial neural network (ANN) are among the hotspots in AI and they have been applied widely in various fields, such as monitoring cultivated land changes (Song et al., 2018), biological image classification (Qin et al., 2020), three-dimensional face recognition (Li et al., 2022), and skin attribute detection (Nguyen et al., 2022). In this regard, some studies on intelligent lithology identification have been conducted. For instance, Marmo et al. (2005) used a multi-layer perceptron neural network to develop a textural identification method for carbonate rocks. Singh and Rao (2005) implemented ore sorting and classification using a radial basis neural network. Singh et al. (2010) proposed an approach to identify textures of basaltic rock based on image processing and neural network. Chatterjee (2013) applied a multi-class support vector machine for rock-type classification of limestone, and an accuracy of 96.2% was obtained. Młynarczyk et al. (2013) conducted automatic classification of microscopic images for nine types of rocks with four pattern recognition methods. Izadi et al. (2017) proposed an intelligent system for mineral identification in thin sections based on ANN. Sun et al. (2019) optimized the logging method while drilling using machine learning algorithms for rapid lithology identification. Xie et al. (2021) developed a coarse-to-fine approach with extremely randomized trees for logging lithology identification. Xu et al. (2022) proposed an on-site identification method of rock images and elemental data using deep learning. Among the above studies, some methods successfully implemented lithology identification using rock images; however, methods for in situ lithology identification in the field are few and have yet to be improved. Moreover, as deep learning has great ability in object detection and identification, meth-

ods using deep learning show potential for rapid and accurate lithology identification.

Convolutional neural network (CNN) is an important and widely used part of deep learning. At present, CNN types used commonly in object detection include region-based CNN (R-CNN) (Girshick et al., 2014), fast R-CNN (Girshick, 2015), faster R-CNN (Ren et al., 2015), you only look once (YOLO) (Redmon et al., 2016), and single shot multibox detector (SSD) (Liu et al., 2016). Fast R-CNN and faster R-CNN were developed based on R-CNN, and R-CNN-based methods can make accurate but slow identification. Meanwhile, the methods using YOLO are fast; however, their accuracy needs to be improved. SSD combines the ideas of the grid in YOLO and the anchor in faster R-CNN; moreover, it is faster and more accurate than YOLO and faster R-CNN. SSD has thus been applied for electronic component recognition (Sun et al., 2020), remote sensing (Lu et al., 2021), and automatic driving (Chen et al., 2022). Hence, SSD can help improve the accuracy of lithology identification in the field.

In this study, SSD was improved first and an intelligent lithology identification method based on improved SSD for rock images was proposed. Then, database (DB) and geographic information science (GIS) technologies were introduced to construct an integrated identification method. The proposed methods were applied to the rocks in Xingcheng area, China. Finally, the influencing factors of the identification results were discussed.

METHODS

Aiming at the problem associated with low in situ lithology identification accuracy for rocks in the field caused by complicated image features, two methods are proposed in this section: improved SSD and integrated method based on DB, GIS, and improved SSD (DGS).

Building Datasets

The prerequisite for intelligent lithology identification is building datasets, which is important for training models. During training, large amounts of labeled samples are provided to improve the accuracy and generalization ability of a network. However, limited by the manual capturing of

photographs, the number of rock images from the field cannot meet the training demands. Hence, the image number is increased through data augmentation. The method of building datasets is as follows.

- (1) Data acquisition: All photos of rock outcrops were captured from the field, and this method was different from a study that utilized images of rock samples (Xu et al., 2022); meanwhile, only one lithology was shot in one image to reduce the effect of non-rock objects, such as vegetation, and water.
- (2) Preprocessing: The parts with contents irrelevant to the rock, such as references in the images, were removed.
- (3) Labeling: The images captured from the field were labeled (Fig. 1) according to the features of the rock images, and the files with the suffix “.xml” were saved in the format of PASCAL Visual Object Classes.
- (4) Augmentation: The files with the suffix “.xml” and images were augmented through (horizontal or vertical) flipping, changing the brightness (increasing to 120–150% of the original image), random fuzzification, translation, scaling, and rotation (Fig. 2). More images were obtained through augmentation. Moreover, the robustness of the model and the accuracy of the results can be improved by augmentation.
- (5) Filtering: The augmented images were filtered by removing images that had very few or even no target rocks so that the features could be more accurately captured and extracted.
- (6) Dividing: 80% of the filtered images were used to build a training dataset, and the remaining 20% served as the testing dataset.

Improved SSD

Network Structure

In SSD, visual geometry group network with 16 weight layers (VGG16) (Simonyan and Zisserman, 2014) is used commonly to extract image features. However, the vanishing gradient problem is caused by the increased depth of VGG16. Residual net (ResNet) helps solve this problem using the residual function (He et al., 2016). Thus, ResNet with 50

layers (ResNet50) was applied as a basic network instead of VGG16.

There were three parts in the structure of improved SSD (Fig. 3): ResNet50, extra feature layers, and prediction layer. ResNet50 was used for feature extraction. Each feature map had default boxes in different sizes. Among these default boxes, the highest layer had a scale of 0.85, the lowest layer had a scale of 0.10, and the layers in between had scales of 0.25, 0.40, 0.55, and 0.70. Extra feature layers were added to the end of ResNet50 for predictions at multiple scales, and the sizes of the layers decreased progressively. Prediction layer was used to obtain the location, confidence, and classification of the bounding boxes.

Loss Function

During identification, the loss function *loss* was used to indicate the difference between prediction and ground truth. The higher the value of *loss*, the lower the precision. The equation for *loss* (Liu et al., 2016) is:

$$loss = \frac{1}{N}(L_{conf} + \alpha L_{loc}) \quad (1)$$

where N is the number of positives, L_{conf} is the confidence loss, L_{loc} is the localization loss, and α is a weight factor, which was set to 1. The equation for L_{conf} is:

$$L_{conf} = - \sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^0), \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)} \quad (2)$$

where Pos represents positives, which are samples belonging to the target lithology; Neg represents negatives, which are samples that do not belong to the target; x_{ij}^p is the indicator for matching the i th default box to the j th ground truth box of category p ; and c_i^p is the confidence of lithology p for the i th default box.

The equations for L_{loc} are:

$$L_{loc} = \sum_{i \in Pos} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k \text{smooth}_{L_1}(l_i^m - g_{o,j}^m) \quad (3)$$

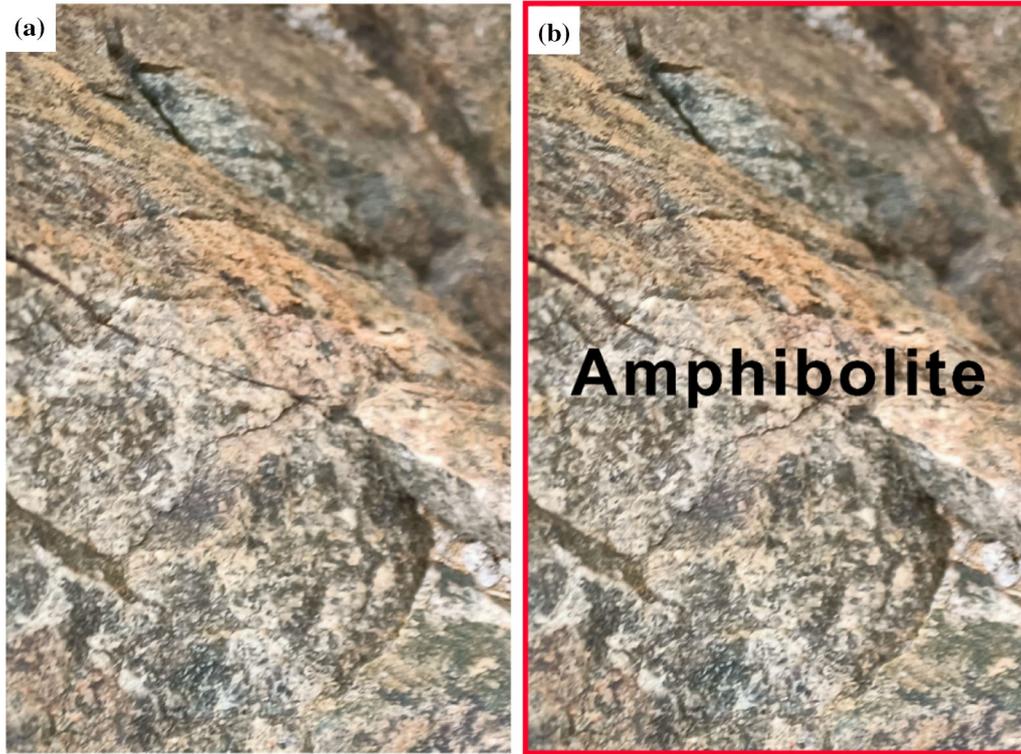


Figure 1. Comparison of unlabeled and labeled images with amphibolite as an example: (a) original image; (b) image after labeling. Red rectangular boxes represent the ground truth boxes.

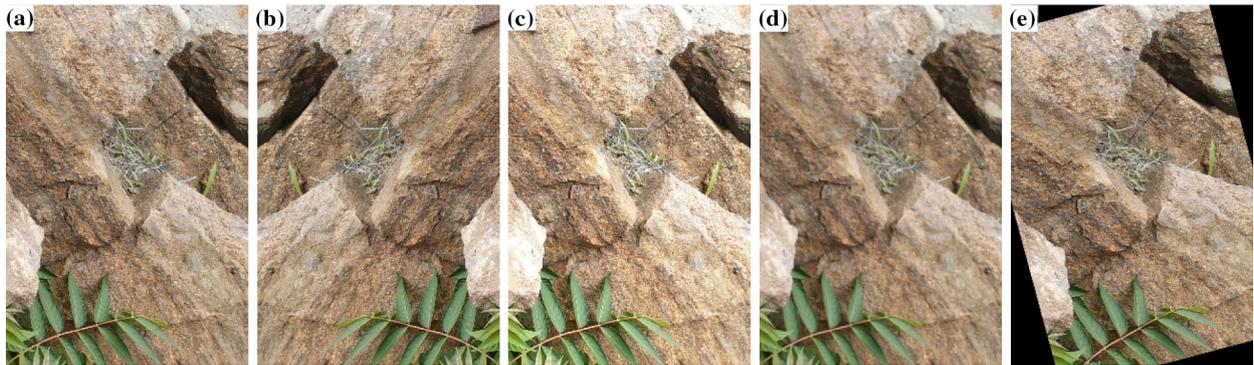


Figure 2. Comparison of images of conglomeratic feldspathic quartz sandstone before and after augmentation: (a) original image; (b) horizontally flipped image; (c) brightened image; (d) fuzzy image; (e) rotated image.

$$\begin{aligned}
 g_{o,j}^{cx} &= (g_j^{cx} - d_i^{cx})/d_i^w, g_{o,j}^{cy} = (g_j^{cy} - d_i^{cy})/d_i^h, g_{o,j}^w \\
 &= \log\left(\frac{g_j^w}{d_i^w}\right), g_{o,j}^h = \log\left(\frac{g_j^h}{d_i^h}\right)
 \end{aligned}
 \tag{4}$$

$$\text{smooth}_{L_1} = \begin{cases} 0.5 \times (x_i - y_i)^2 / \beta, & \text{if } |x_i - y_i| < \beta \\ |x_i - y_i| - 0.5 \times \beta, & \text{otherwise} \end{cases}
 \tag{5}$$

where l is the predicted box, g is the ground truth box; (cx, cy) is the center coordinate of the default

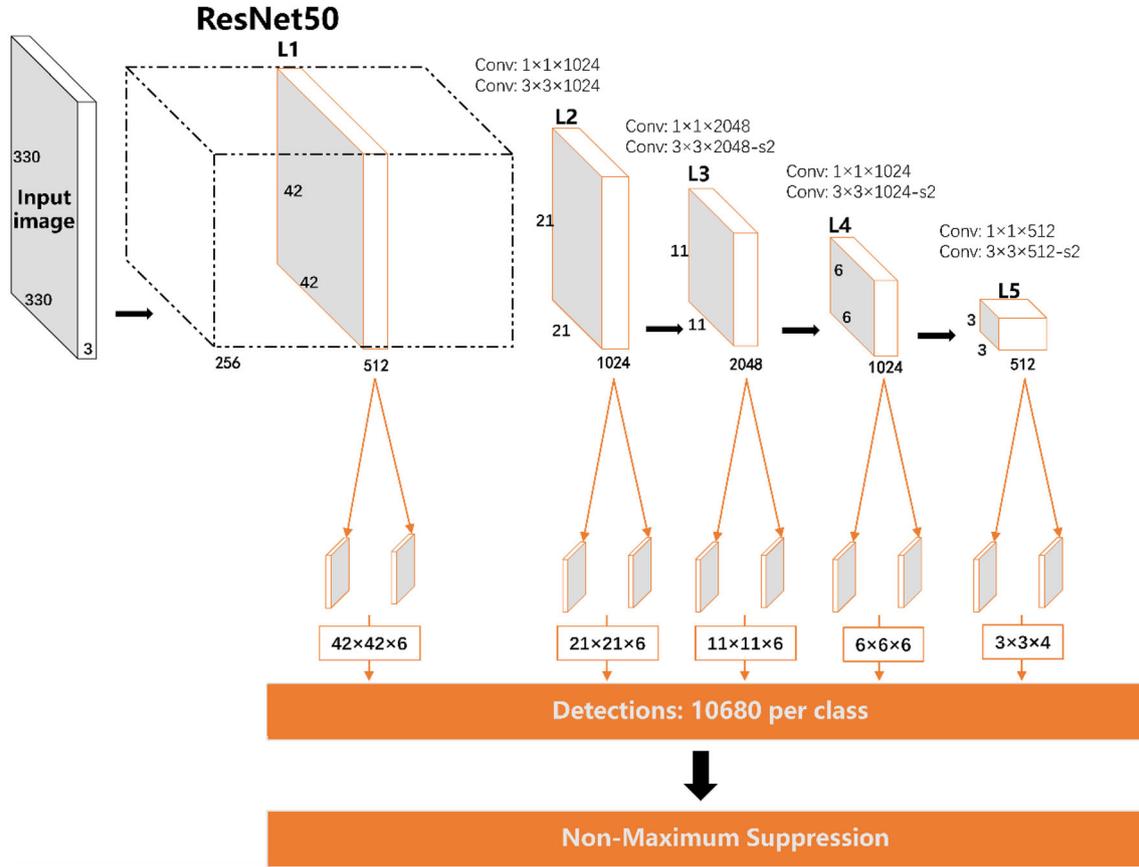


Figure 3. Network structure of improved SSD.

box d ; w and h are the width and height of d , respectively; l_i^m , g_j^m , and d_i^m are the parameters m of the i th predicted box, the j th ground truth box, and the i th default box, respectively; $g_{o,j}^m$ is the offset of g_j^m ; smooth_{L_1} is the robust L_1 loss function (Girshick, 2015); x_i and y_i are the values of the target and output, respectively; and β specifies the threshold at which to change between L_1 and L_2 loss, which is a constant set to 1 in this study.

Optimization Strategy

The adaptive moment estimation (Adam) algorithm (Kingma and Ba, 2014) was adopted instead of stochastic gradient descent as the optimization strategy. Adam combines the advantages of adaptive gradient (Duchi et al., 2011) and root mean square propagation (Tieleman and Hinton, 2012). Hence, it can deal with sparse gradients and

non-stationary objects. The equations of Adam are as follows:

$$g_t = \nabla_{\theta} f_t(\theta_{t-1}), \theta_t = \theta_{t-1} - \frac{\alpha}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t \quad (6)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (7)$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (8)$$

where g_t is the gradient of loss function f with respect to parameter θ at time step t ; m_t is the biased first moment estimate at time step t ; v_t is the biased second raw moment estimate at time step t ; \hat{m}_t and \hat{v}_t are the bias-corrected first moment estimate, and bias-corrected second raw moment estimate, respectively; α is the step size; β_1 and β_2 are the exponential decay rates for the moment estimates;

and ε is a small positive number to avoid generating a singular value.

Integrated Identification Method

DGS, a method based on improved SSD, and using the constraint provided by DB and GIS, is proposed to further improve lithology identification in the field.

Rocks of different lithology can have similar appearances. Therefore, identification using only rock images can lead to misclassifications. In most cases, historical work can provide rock information, such as location and lithology, which can be the basis for identification. Once imported into the database, the known rock information provides constraints for in situ identification for further improved accuracy. Specifically, after the preliminary identification using improved SSD, the lithology with maximum confidence was not output as an identification result. Instead, a set of lithology candidates was output as the preliminary result. According to rock location, the known lithology information in the area was obtained from the database, which was regarded as a constraint to compare with the preliminary identification result. Based on this comparison, the lithology candidate that was not in the set of known lithology was removed from the preliminary result. Then, the lithology with maximum confidence in the rest was considered the final result. If the lithology in the constraints were quite different from the candidates, only improved SSD was used and the constraints did not affect the lithology identification. Information on newly identified rocks could be imported into the database to provide the constraints for future works. In DGS, the DB technique was used to provide the constraints, the GIS technique was used to provide spatial information, and the improved SSD was used for identification. The combination of DB, GIS, and improved SSD contributed to high-accuracy in situ identification.

It should be noted that there is a prerequisite for this integrated method: the historical work has been performed in the study area (i.e., the geological background, including the lithology, is known). If no geological information was collected, only improved SSD was used. It could be inferred from the above that DGS could be adopted in many application scenarios of lithology identification. For example, exposed rocks can be identified in situ using DGS during mine production, so that the ore body can be

distinguished from the surroundings and a database of lithology distribution could be built, supporting high-efficiency mining.

Compared with the improved SSD, DGS apparently helped obtain optimal identification, build the lithology database, and provide new insights into the in situ identification in the field. The steps of DGS are as follows (Fig. 4):

- (1) Build the database for known lithology using MySQL.
- (2) Import rock information, including location and lithology, collected from different ways into the database.
- (3) Capture images of the rocks in the field and obtain the coordinates using a positioning tool, such as, the GPS module in cameras or smartphones.
- (4) Import the rock images into improved SSD and simultaneously search the lithology information from the database according to the coordinates.
- (5) Implement lithology identification using improved SSD to produce candidates and the constraints are provided by the database.
- (6) Obtain the final result by integrating the information of the candidates and constraints and import it into the database.
- (7) Evaluate the results with the metrics after identification (for details, see section of Assessment Methodology).

TESTS AND RESULTS

The proposed methods were applied to rock images obtained from Xingcheng area, China. The tests were implemented to verify the effectiveness and feasibility of the methods.

Geological Setting and Rock Images

Xingcheng area is in the southwestern part of Liaoning Province. In terms of regional geomorphology, it is located in a coastal hilly area on the eastern margin of Heishan Hills in the West Liaoning Mountainous Region. In terms of regional tectonics, it is located north of the North China Craton. In this area, the strata are well-developed, with various types of rocks (Liang et al., 2015).

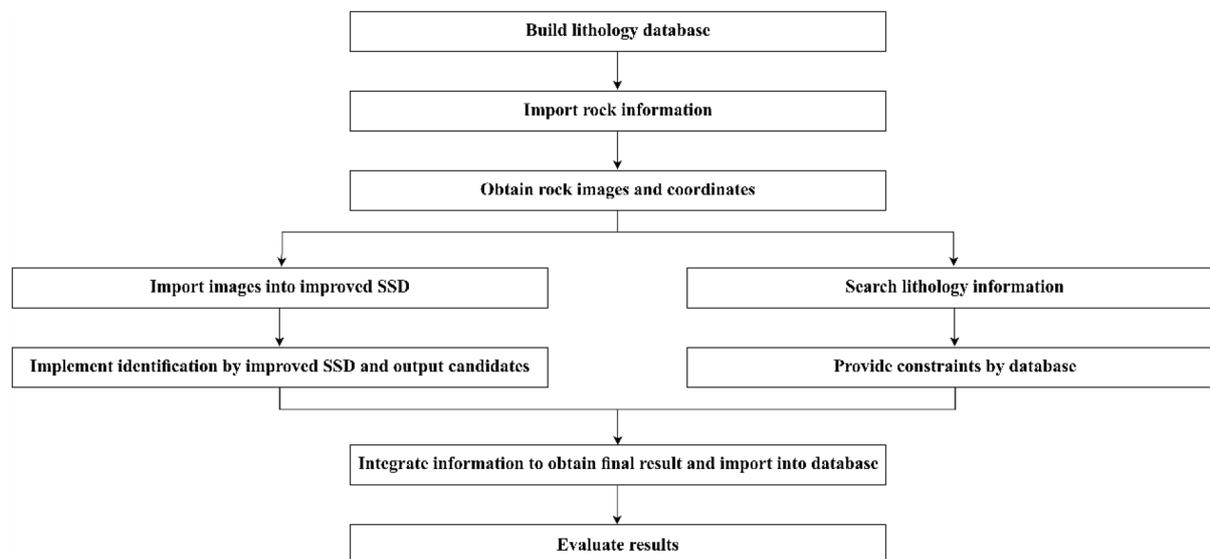


Figure 4. Flow of DGS.

The rock images were captured from Diaoyutai, Jiashan, Longhuitou, Heiyugou, and Taili in Xingcheng area. Well-exposed rocks with distinguishing lithologic features were considered typical rocks and used for identification. Among the rock images (Fig. 5), those of biotite monzonitic granite, monzonitic granite, quartz syenite, amphibolite, diabase, and granite pegmatite were captured from Diaoyutai, those of conglomeratic feldspathic quartz sandstone and quartz conglomerate were captured from Jiashan, those of polymictic conglomerate, quartz sandstone, quartz conglomerate, and diabase were captured from Longhuitou, those of oolitic limestone were captured from Heiyugou, and those of granite pegmatite, diabase, and mylonite were captured from Taili.

In total, 1187 raw images were captured from the field using the smartphone OnePlus 6, and one target lithology was captured in one image. The location information of the rocks was acquired using the smartphone's GPS synchronously when the photographs were captured. According to the section Building Datasets, 11870 images were finally obtained after preprocessing, labeling, augmentation, and filtering. Of these images, 80% were used to build a training dataset, and 20% served as the testing dataset.

Lithology Identification Tests

In the tests, all of the codes were written in Python language. Python tool boxes were used, including LabelImg for labeling the images, imgaug and PIL for data augmentation, and PyTorch for designing the network and calling graphics processing unit (GPU). The tests were implemented on the GPU node of the supercomputer TianHe HPC4. The node had $2 \times$ Intel Xeon Gold 6354 18-core processors with 3.00 GHz, $2 \times$ NVIDIA HGX A100 GPU, and 256 GB memory. The operating system was RedHat Enterprise Linux 8.4.

The built dataset was adopted to test the trained improved SSD, and the constraints were introduced by DB and GIS. The results obtained from improved SSD and DGS were compared. In Figure 6, polymictic conglomerate, quartz conglomerate, and mylonite were taken as examples to show the identification results of DGS. When the epoch number was 200, learning rate was 1×10^{-4} , and batch size was 4. The identification accuracies of the improved SSD and DGS were recorded (Table 1).

As shown in Table 1, the average accuracies of the improved SSD and DGS were 89.4% and 98.4%, respectively. After introducing the constraints, the lithology candidates were removed, that might have had high confidence but did not exist in the area where the target rock was located. Hence, the

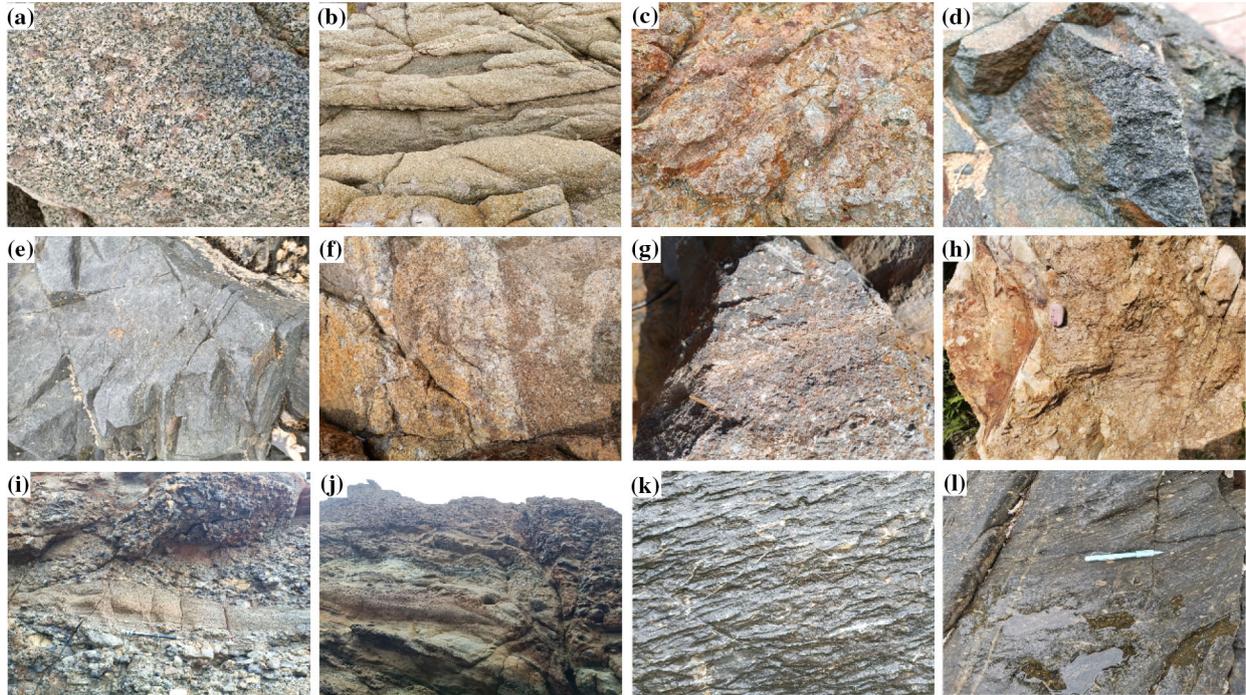


Figure 5. Raw images of typical rocks: (a) biotite monzonitic granite; (b) monzonitic granite; (c) quartz syenite; (d) amphibolite; (e) diabase; (f) granite pegmatite; (g) conglomeratic feldspathic quartz sandstone; (h) quartz conglomerate; (i) polymictic conglomerate; (j) quartz sandstone; (k) oolitic limestone; (l) mylonite.

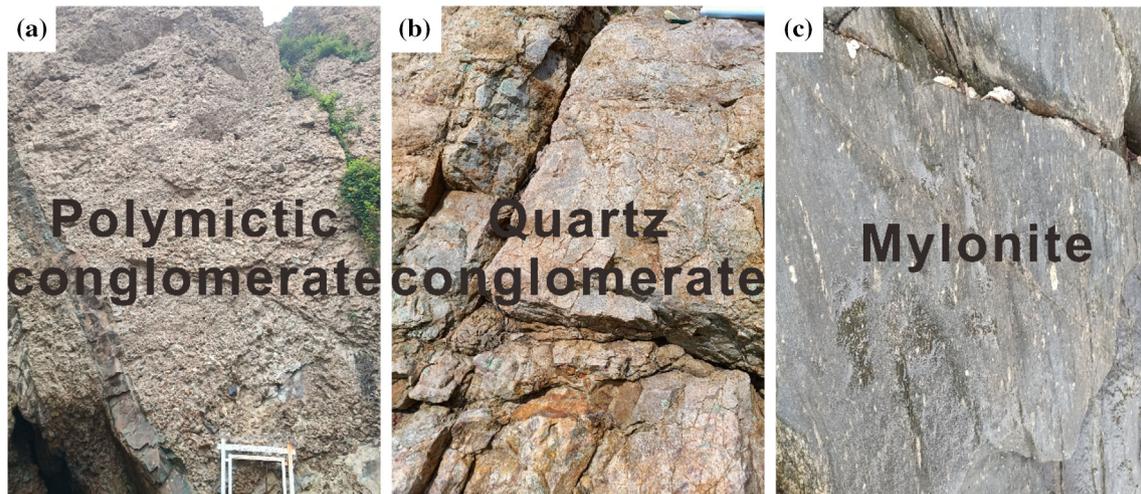


Figure 6. Identification results of typical rocks: (a) polymictic conglomerate; (b) quartz conglomerate; (c) mylonite.

accuracy could be improved and could even reach 100% for some types of rocks. Compared with the improved SSD, DGS resulted in varying degrees of accuracy improvement. However, the accuracies for biotite monzonitic granite, diabase, granite pegmatite, and polymictic conglomerate were not im-

proved by DGS. Moreover, the accuracies for these four rocks were more than 90%. This shows that, in most cases, DGS can improve the accuracy for rocks exposed in only one area; for widely distributed rocks, the accuracy may not be improved considerably by DGS; however, DGS at least does not pro-

Table 1. Identification accuracies of improved SSD and DGS for the studied rocks

Lithology	Improved SSD (%)	DGS (%)
Biotite monzonitic granite	90.0	90.0
Monzonitic granite	96.2	100.0
Quartz syenite	86.1	97.2
Amphibolite	75.0	100.0
Diabase	93.3	93.3
Granite pegmatite	90.9	90.9
Conglomeratic feldspathic quartz Sandstone	93.1	100.0
Quartz conglomerate	93.1	100.0
Polymictic conglomerate	96.9	96.9
Quartz sandstone	92.9	100.0
Oolitic limestone	76.5	100.0
Mylonite	94.4	100.0
Average Accuracy	89.4	98.4

duce negative effects. This indicates that the proposed methods are effective and feasible. The identification ability of the improved SSD was proven to be unaffected by location, but that of DGS was related to the lithology distribution.

DISCUSSIONS

As DB and GIS only provided the constraints, and the basis of DGS was the improved SSD, analyzing the factors that impact on training was important to DGS as well as the improved SSD. The identification results obtained through the training with different parameters are discussed in the following tests. The effects of parameter variations on the identification performance are analyzed.

Assessment Methodology

Assessments of identification performance were implemented during the tests.

The first was the assessment of the predicted boxes when outputting the identification result of a single image. For each image in the dataset, non-maximum suppression (NMS) was adopted for the filtering (Fig. 7). The maximum confidence of all the predicted boxes was found, and the identification result of the corresponding box was output. Hence, confidence was used as the metric to judge which box to output.

The second one was the assessment with *loss* for model training. The values of *loss* during training and testing were used typically as the metric for

evaluating the network. However, as this evaluation was conducted only for training during the test, only the values of *loss* in the training were used.

The third assessment was made for the results of the methods. Based on the section of Building Datasets, only one type of rock was captured in each image when acquiring data. Therefore, for the improved SSD and DGS, the candidate with maximum confidence was considered the prediction result, and accuracy was equal to the ratio of correctly identified image numbers to the total image numbers, which was different from the common way. Accuracy (*Acc*), precision (*Pre*), recall (*Rec*), F₁-score (*F*), and mean average precision (mAP) were used to evaluate the identification. Their equations are, respectively, as follows:

$$Acc = \frac{N_c}{N} \quad (9)$$

$$Pre = \frac{TP}{TP + FP} \quad (10)$$

$$Rec = \frac{TP}{TP + FN} \quad (11)$$

$$F = \frac{2 \times Pre \times Rec}{Pre + Rec} \quad (12)$$

$$mAP = \frac{\sum_{i=1}^K AP_i}{K} \quad (13)$$

where N_c is the number of images identified correctly, N is the total number of images, mAP is the mean average precision, AP_i is the average

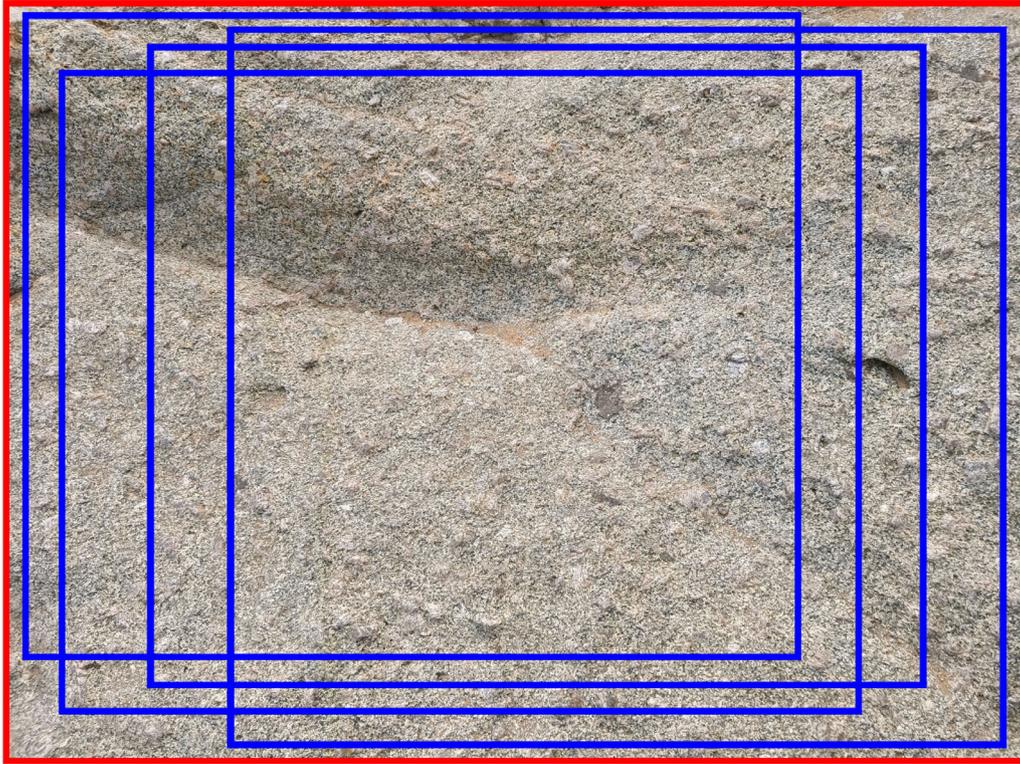


Figure 7. Evaluation with NMS. Blue boxes represent predicted boxes, and the red box represents the ground truth box.

accuracy of lithology i , and K is the total number of lithology.

Performance Analysis

Different learning rates and batch sizes were used for the model training. In the methods, learning rate is the stride in epochs during the training, and batch size is the sample number utilized in one iteration. The effects of parameter variation on identification performance were discussed and analyzed. The epoch number was set to 200.

First, we set batch size to 16 and learning rates to 5×10^{-2} , 1×10^{-2} , 5×10^{-3} , 1×10^{-3} , 5×10^{-4} , 1×10^{-4} , and 5×10^{-5} . The effect of learning rate on identification was discussed. All the other parameters and dataset were the same in the training. The values of *loss* (Fig. 8) and accuracies (Fig. 9) were recorded during the training with different learning rates. When the learning rate was 5×10^{-2} , *loss* was too high, and its curve had a

different shape from those of the others; therefore, it was plotted separately.

In the *loss* figure (Fig. 8a), as the epoch number increased, the values of *loss* decreased quickly and tended to stabilize, except when the learning rates were 1×10^{-2} and 5×10^{-3} . Based on this test, an extremely large learning rate led to the case where *loss* could not decrease to a low level, such as when learning rate was 1×10^{-2} . Similarly, when the learning rate was 5×10^{-3} , *loss* had a faster convergence and smaller final value than *loss* obtained when the learning rate was 1×10^{-2} ; however, this was still unexpected. The final values of *loss* were 1.445 and 0.370 when the learning rates were 1×10^{-2} and 5×10^{-3} , respectively, which were much higher than the values obtained when using other learning rates. When the learning rates were 1×10^{-3} , 5×10^{-4} , 1×10^{-4} , and 5×10^{-5} , the final values of *loss* were 0.023, 0.015, 0.005, and 0.004, respectively. These indicated that the smaller the learning rate, the smaller the values of *loss*. Moreover, in the first 25 epochs, the smaller the learning rate, the faster the convergence rate. However, the

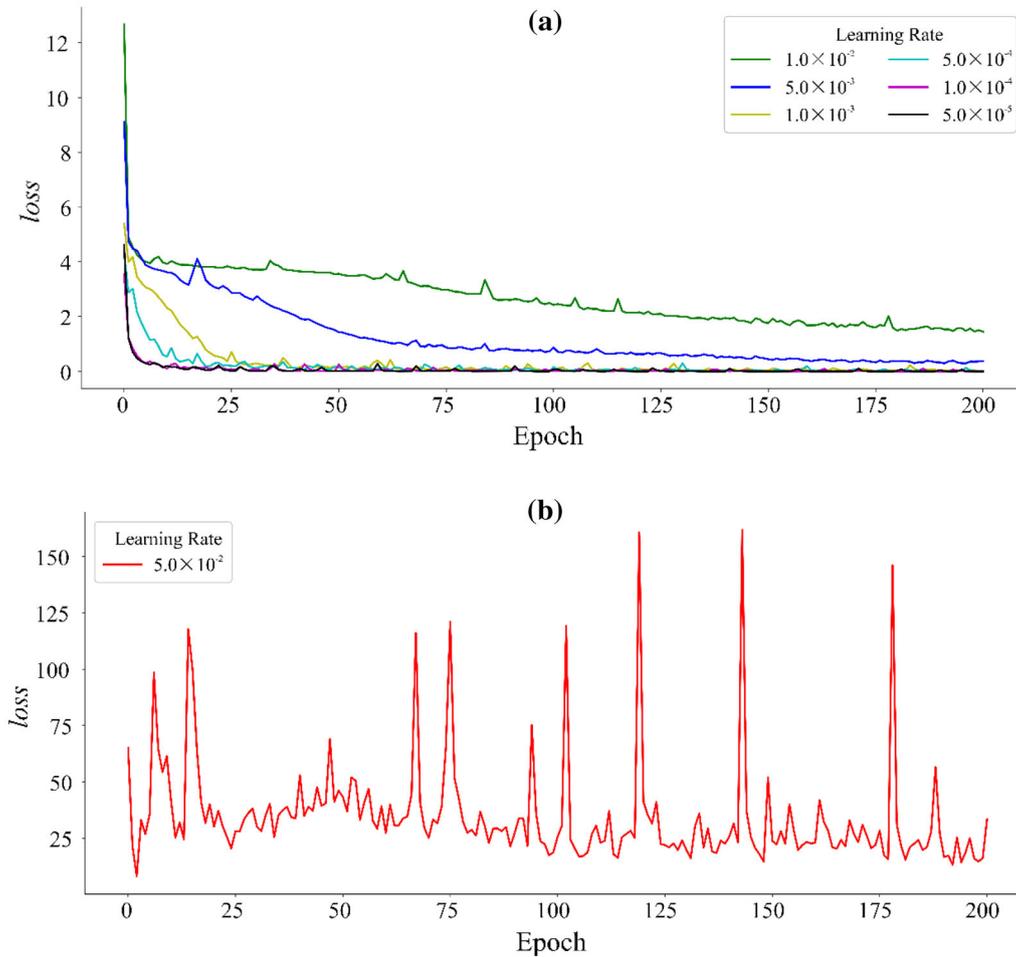


Figure 8. Values of *loss* obtained during the training with different learning rates: (a) learning rate = $5 \times 10^{-5} - 1 \times 10^{-2}$; (b) learning rate = 5×10^{-2} .

convergence rate was faster when learning rate was 1×10^{-4} than when the learning rate was 5×10^{-5} . Figure 8b shows that when the learning rate was 5×10^{-2} , *loss* increased and decreased repeatedly, which could even be larger than the initial value. The value of *loss* finally reached a large value (33.407). This proved that learning rate had a major effect on the convergence and fitting of *loss*, and the learning rate should not be larger than 1×10^{-3} .

Based on the accuracies of the improved SSD (Fig. 9a), the final accuracies increased to varying degrees compared with the initial values. When the learning rates were 5×10^{-2} , 1×10^{-2} , 5×10^{-3} , 1×10^{-3} , 5×10^{-4} , 1×10^{-4} , and 5×10^{-5} , the final accuracies were 9.0%, 11.9%, 27.1%, 30.6%, 24.5%, 38.1%, and 43.5%, respectively. The corre-

sponding final accuracies of DGS were 31.9%, 34.8%, 55.5%, 57.4%, 55.8%, 67.4%, and 73.5%, respectively (Fig. 9b). These indicated that, generally, the smaller the learning rate, the higher the final accuracy. This is because large *lr* may have difficulty converging (Buduma and Locascio, 2017), and small *lr* can approximate optimal solution. In addition, the accuracies of DGS were higher than those of the improved SSD. It was demonstrated that introducing DB and GIS technologies can greatly improve accuracy. The curve shapes of the accuracies of the improved SSD and DGS were similar and showed a small tendency to increase because the improved SSD was used for identification in DGS. Moreover, how learning rate affects the values of *loss*, accuracy, and convergence rate

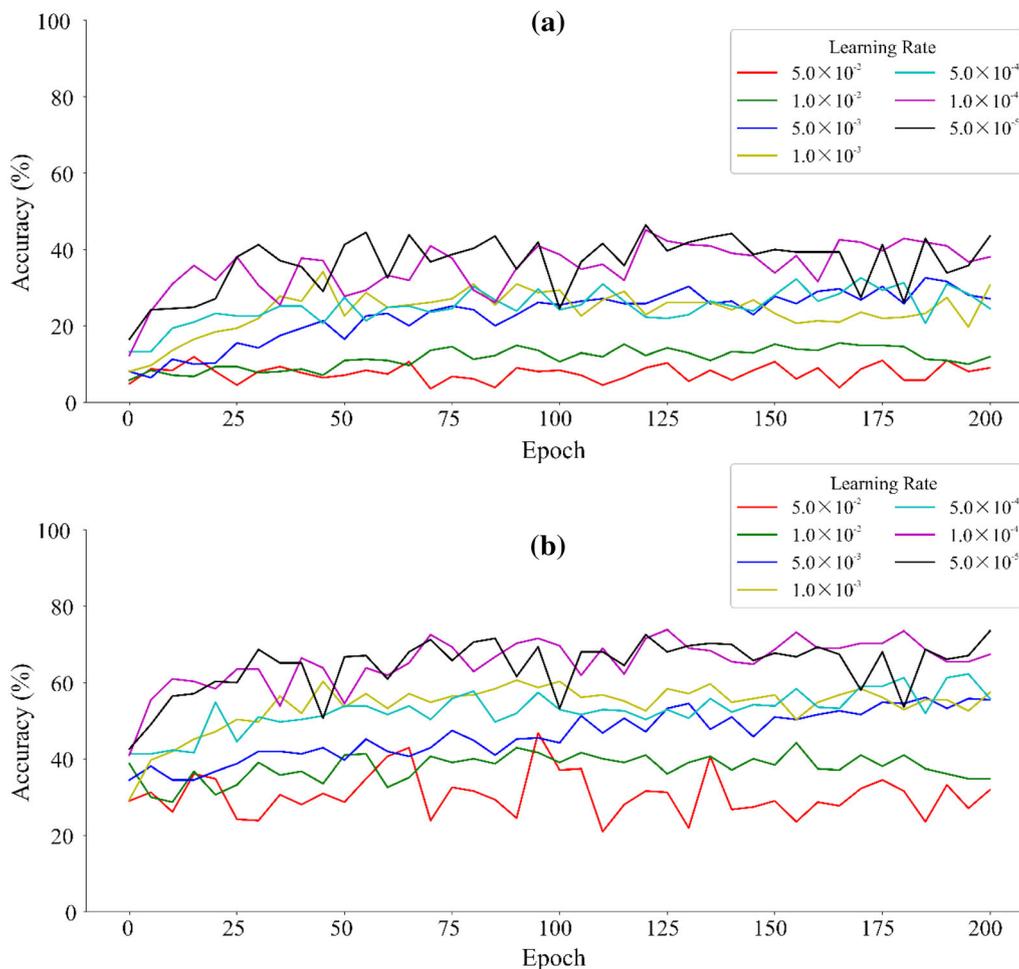


Figure 9. Accuracies with different learning rates using (a) improved SSD and (b) DGS.

should be considered for the optimal solution. Hence, 1×10^{-4} was regarded as the learning rate used for the best identification results.

Next, we set the learning rate to 1×10^{-4} and batch sizes to 4, 8, 16, 32, 64, and 128, whereas the data and other parameters were the same. The effect of batch size on identification was discussed. The results are shown in Figures 10 and 11.

As shown in Figure 10, in the first 25 epochs, the values of all the curves showed a fast convergence. As the epoch number increased, *loss* fluctuated. The larger the batch size, the larger the amplitude of the fluctuation. When batch sizes were 4, 8, 16, 32, 64, and 128, the final values of *loss* were 0.016, 0.010, 0.005, 0.004, 0.003, and 0.004, respectively. These indicated that the final values of *loss* decreased initially and then increased as batch size increased.

Therefore, batch size should be more than 32 if a small value of *loss* is desired.

When analyzing the effect of batch size on accuracy, we still found that the accuracies of DGS were higher than those of the improved SSD (Fig. 11). When batch sizes were 4, 8, 16, 32, 64, and 128, the accuracies of the improved SSD were 89.4%, 59.0%, 38.1%, 24.5%, 22.3%, and 16.8%, respectively. The corresponding accuracies of DGS were 98.4%, 84.2%, 67.4%, 56.8%, 49.0%, and 50.6%, respectively. Generally, the accuracies decreased as batch size increased. For all batch sizes, the accuracy tended to be stable after training within 25 epochs. When batch size was 4, the highest accuracy was obtained and the curve had the fastest ascending rate. Because the methods with large batch sizes tended to converge to sharp minimizers

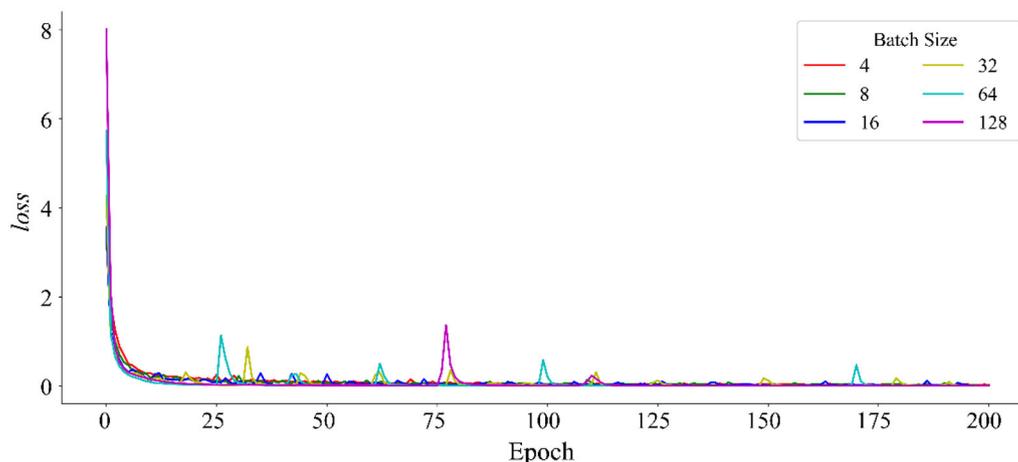


Figure 10. Values of *loss* obtained during the training with different batch sizes.

and led to poor generalization, the methods with small batch sizes converged to flat minimizers (Keskar et al., 2016). In addition, a small batch size can result in more trainings so that more image features can be extracted. Although small batch size caused a large value of *loss*, it had better generalization for identifications of testing dataset. A proper batch size should provide high accuracy instead of a low value of *loss*. Hence, 4 was regarded as the optimal batch size. In Figure 11, the curve shapes of the accuracies of the improved SSD and DGS were similar. This also appears in Figure 9. Overall, although the learning rate of 1×10^{-4} and batch size of 4 could not provide minimum *loss* and maximum accuracy simultaneously, they were still considered the optimal parameters for the training of the improved SSD and DGS.

Finally, the improved SSD and DGS were compared with YOLOv5s to further verify the proposed methods' superiority in in situ lithology identification. The dataset and training parameters were the same for the improved SSD, DGS and YOLOv5s. The metrics in the section of Assessment Methodology were used for evaluation (Table 2), where the values behind @ represent the thresholds of intersection over union when computing mAP. As shown in Table 2, compared with YOLOv5s, the improved SSD had higher accuracy, precision, F_1 -score, $mAP@0.5:0.95$, and close recall and $mAP@0.5$. The metrics of DGS showed that identification was improved to varying degrees compared with the improved SSD. As DGS was developed by adding constraints based on the improved SSD, their

recalls were the same. Except for the recall, the accuracy, precision, F_1 -score, and mAP of DGS were superior to those of YOLOv5s. YOLOv5s has the advantage of a shorter identification time; however, the identification times of DGS and the improved SSD were also fast. Faster-RCNN was also chosen for comparison, but its metrics were very low regardless of the identification speed. Hence, Faster-RCNN was considered unsuitable for identification in this study, and its metrics are thus not shown. We demonstrated that DGS and the improved SSD were very effective in in situ lithology identification compared with other methods.

CONCLUSIONS

Aiming at solving low-accuracy problems of in situ lithology identification in the field, an improved SSD was proposed and was combined with DB and GIS to propose a method called DGS. The methods were applied to the images of typical rocks in Xingcheng area. The average accuracies of the improved SSD and DGS were 89.4% and 98.4%, respectively, and the maximum accuracies could reach 100%. The two proposed methods could identify lithology accurately. Moreover, DGS with constraints helped improve the accuracy based on the improved SSD and support future identifications by building a lithology database. A series of tests were designed and implemented to discuss the effects of various parameters on identification. Different learning rates and batch sizes were used to

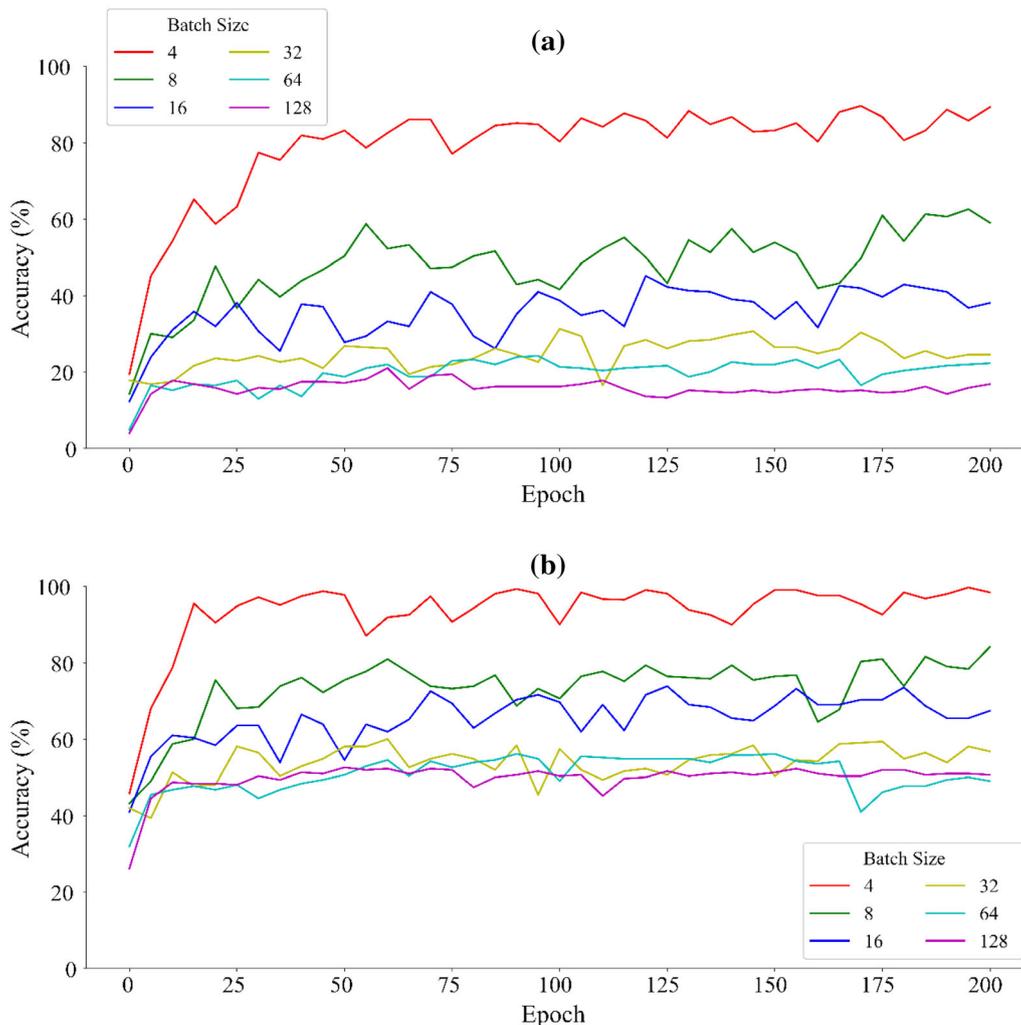


Figure 11. Accuracies with different batch sizes using (a) improved SSD and (b) DGS.

Table 2. Evaluation records of different identification methods

Metrics	Improved SSD	DGS	YOLOv5s
Accuracy (%)	89.4	98.4	87.6
Precision (%)	66.2	84.5	63.7
Recall (%)	96.8	96.8	98.0
F ₁ -score	0.786	0.902	0.772
mAP@0.5	0.780	0.883	0.788
mAP@0.5:0.95	0.762	0.847	0.713
Identification time per image (Unit: ms)	64.7	67.8	20.2

train the model, and the values of *loss* and accuracies were recorded and analyzed. Generally, the smaller the learning rate, the smaller the value of *loss*, the

faster the convergence rate, and the higher final accuracy; the smaller the batch size, the larger the value of *loss* and the higher final accuracy. Fur-

thermore, learning rate had a small impact on the ascending rate of accuracy, whereas batch size had a small impact on the convergence rate. The values of learning rate and batch size should be determined correctly, or else *loss* increased. Hence, the appropriate parameter combination was provided for the optimal model. In this study, the optimal parameters were a learning rate of 1×10^{-4} and batch size of 4. The superiority of the proposed methods was further proven based on various metrics, including accuracy, precision, recall, F_1 -score, and mAP. Compared with YOLOv5s, DGS had a stronger identification ability; however, the former had a faster identification speed. Moreover, the improved SSD and DGS were effective and feasible, and could provide new insights into and support for in situ lithology identification in the field.

ACKNOWLEDGMENTS

We will thank the editors and the anonymous reviewers for their work. This work was supported by Liaoning Provincial Natural Science Foundation of China (2022-MS-107), and the National Natural Science Foundation of China (42204140).

DECLARATIONS

Conflict of Interest The authors declare that there are no conflicts of interest.

OPEN ACCESS

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

- Asfahani, J., Ghani, B. A., & Ahmad, Z. (2015). Basalt identification by interpreting nuclear and electrical well logging measurements using fuzzy technique (case study from southern Syria). *Applied Radiation and Isotopes*, *105*, 92–97.
- Bosch, M., & McGaughey, J. (2001). Joint inversion of gravity and magnetic data under lithologic constraints. *The Leading Edge*, *20*(8), 877–881.
- Buduma, N., & Locascio, N. (2017). *Fundamentals of deep learning: Designing next-generation machine intelligence algorithms*. O'Reilly.
- Chatterjee, S. (2013). Vision-based rock-type classification of limestone using multi-class support vector machine. *Applied Intelligence*, *39*(1), 14–27.
- Chen, Z., Guo, H., Yang, J., Jiao, H., Feng, Z., Chen, L., & Gao, T. (2022). Fast vehicle detection algorithm in traffic scene based on improved SSD. *Measurement*, *201*, 111655.
- Cheng, W., Cheng, H., Yu, H., Zhang, S., Wang, Z., Li, X., & Johnson, L. (2022). X-Ray Fluorescence for Laminated Silty Shale Reservoirs in Ordos Basin, China: Implications for lithology identification. *Geofluids*, *2022*, 3207575.
- Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, *12*, 2121–2159.
- Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440–1448).
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580–587). Columbus, Ohio, USA.
- Gleeson, M. L., Gibson, S. A., & Williams, H. M. (2020). Novel insights from Fe-isotopes into the lithological heterogeneity of Ocean Island Basalts and plume-influenced MORBs. *Earth and Planetary Science Letters*, *535*, 116114.
- Han, L., Fuqiang, L., Zheng, D., & Weixu, X. (2018). A lithology identification method for continental shale oil reservoir based on BP neural network. *Journal of Geophysics and Engineering*, *15*(3), 895–908.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–780). Las Vegas, NV, USA.
- Izadi, H., Sadri, J., & Bayati, M. (2017). An intelligent system for mineral identification in thin sections based on a cascade approach. *Computers & Geosciences*, *99*, 37–49.
- Keskar, S. N., Mudigere D., Nocedal J., Smelyanskiy, M., & Tang, P. T. P. (2016). On large-batch training for deep learning: Generalization gap and sharp minima. In *Proceedings of the International Conference on Learning Representations* (pp. 1–16). Palais des Congrès Neptune, Toulon, France.
- Kingma, D. P., & Ba, L. J. (2014). Adam: A method for stochastic optimization. <https://arxiv.org/abs/1412.6980>. Retrieved January 30, 2017.
- Kumar, C. V., Vardhan, H., Murthy, C. S., & Karmakar, N. C. (2019). Estimating rock properties using sound signal dominant frequencies during diamond core drilling operations. *Journal of Rock Mechanics and Geotechnical Engineering*, *11*(4), 850–859.
- Li, M., Huang, B., & Tian, G. (2022). A comprehensive survey on 3D face recognition methods. *Engineering Applications of Artificial Intelligence*, *110*, 104669.
- Liang, C., Liu, Y., Neubauer, F., Bernroider, M., Jin, W., Li, W., Zeng, Z., Wen, Q., & Zhao, Y. (2015). Structures, kinematic analysis, rheological parameters and temperature-pressure estimate of the Mesozoic Xingcheng-Taili ductile shear zone

- in the North China Craton. *Journal of Structural Geology*, 78, 27–51.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference* (pp.21–37). Amsterdam, Netherlands.
- Lu, X., Ji, J., Xing, Z., & Miao, Q. (2021). Attention and feature fusion SSD for remote sensing object detection. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–9.
- Marmo, R., Amodio, S., Tagliaferri, R., Ferreri, V., & Longo, G. (2005). Textural identification of carbonate rocks by image processing and neural network: Methodology proposal and examples. *Computers & Geosciences*, 31(5), 649–659.
- Młynarczuk, M., Górszczyk, A., & Ślipek, B. (2013). The application of pattern recognition in the automatic classification of microscopic rock images. *Computers & Geosciences*, 60, 126–133.
- Nguyen, D. M., Nguyen, T. T., Vu, H., Pham, Q., Nguyen, M. D., Nguyen, B. T., & Sonntag, D. (2022). TATL: task agnostic transfer learning for skin attributes detection. *Medical Image Analysis*, 78, 102359.
- Pour, A. B., Park, Y., Park, T. Y. S., Hong, J. K., Hashim, M., Woo, J., & Ayoobi, I. (2018). Regional geology mapping using satellite-based remote sensing approach in Northern Victoria Land, Antarctica. *Polar Science*, 16, 23–46.
- Qin, J., Pan, W., Xiang, X., Tan, Y., & Hou, G. (2020). A biological image classification method based on improved CNN. *Ecological Informatics*, 58, 101093.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788). Las Vegas, NV, USA.
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. <https://arxiv.org/abs/1409.1556>. Retrieved April 10, 2015.
- Singh, N., Singh, T. N., Tiwary, A., & Sarkar, K. M. (2010). Textural identification of basaltic rock mass using image processing and neural network. *Computational Geosciences*, 14(2), 301–310.
- Singh, V., & Rao, S. M. (2005). Application of image processing and radial basis neural network techniques for ore sorting and ore classification. *Minerals Engineering*, 18(15), 1412–1420.
- Song, F., Li, M., Yang, Y., Yang, K., Gao, X., & Dan, T. (2018). Small UAV based multi-viewpoint image registration for monitoring cultivated land changes in mountainous terrain. *International Journal of Remote Sensing*, 39(21), 7201–7224.
- Sun, J., Li, Q., Chen, M., Ren, L., Huang, G., Li, C., & Zhang, Z. (2019). Optimization of models for a rapid identification of lithology while drilling-A win-win strategy based on machine learning. *Journal of Petroleum Science and Engineering*, 176, 321–341.
- Sun, X., Gu, J., & Huang, R. (2020). A modified SSD method for electronic components fast recognition. *Optik*, 205, 163767.
- Tieleman, T., & Hinton, G. (2012). Lecture 6.5-rmsprop: Divide the Gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning*, 4, 26–31.
- Xie, Y., Zhu, C., Hu, R., & Zhu, Z. (2021). A coarse-to-fine approach for intelligent logging lithology identification with extremely randomized trees. *Mathematical Geosciences*, 53(5), 859–876.
- Xu, Z., Shi, H., Lin, P., & Liu, T. (2021). Integrated lithology identification based on images and elemental data from rocks. *Journal of Petroleum Science and Engineering*, 205, 108853.
- Xu, Z., Shi, H., Lin, P., & Ma, W. (2022). Intelligent on-site lithology identification based on deep learning of rock images and elemental data. *IEEE Geoscience and Remote Sensing Letters*, 19, 6511205.