Journal of Computer Science and Electrical Engineering
Print ISSN: 2663-1938
Online ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3104

ADVANCES IN INTELLIGENT ROCK IMAGE RECOGNITION BASED ON CONVOLUTIONAL NEURAL NETWORKS

He Ma
Yellow River Engineering Consulting Co., Ltd., Zhengzhou 450003, Henan, China.
Corresponding Email: mahe026@163.com

Abstract: Lithology identification is a fundamental task in resource exploration and engineering geology, yet traditional methods face bottlenecks such as low efficiency and high subjectivity. In recent years, intelligent rock image recognition techniques based on convolutional neural network (CNN) have demonstrated remarkable advantages. This paper systematically reviews the research progress of CNN in intelligent rock image recognition and explores their application potential and technical challenges in this field. First, the basic architecture and working principles of CNN are introduced, including the synergistic interactions among convolutional layers, pooling layers, and fully connected layers. Subsequently, the criteria for model selection and optimization pathways in rock image recognition are analyzed, covering task-specific model adaptation strategies and multi-model comparative evaluation and selection strategies. Additionally, the roles of data augmentation strategies, resolution enhancement techniques, and model architecture innovations in improving model performance are discussed. Finally, this paper summarizes the limitations of current research and proposes future research directions, aiming to provide theoretical support and practical guidance for overcoming existing technical bottlenecks.
Keywords: Convolutional neural network; Rock image; Lithology identification; Identification mode; Optimization path

1 INTRODUCTION

Lithology identification serves as a core foundational task in resource exploration and engineering geology. Precise identification results not only provide critical data support for deep mineral exploration and hydrocarbon reservoir evaluation but also offer a scientific basis for design optimization and construction safety in major engineering projects, such as mining, tunneling, and hydraulic infrastructure development [1]. Although traditional methods—including macroscopic observation, thin-section identification, and laboratory analysis—have been widely applied [2-7], these workflows are characterized by high labor intensity, substantial costs, prolonged duration, and significant subjectivity [8-11].

In recent years, the rapid advancement of artificial intelligence has provided technical support for the intelligent detection, classification, and segmentation of rock imagery, offering a novel pathway to mitigate the excessive reliance on expert experience inherent in traditional lithology identification. Current intelligent recognition technologies based on rock images are primarily categorized into two distinct paradigms according to their automation levels: traditional machine learning (ML)-based classification and deep learning (DL)-based intelligent recognition [12]. Regarding ML approaches, Marmo and Amodio utilized a multilayer perceptron neural network to identify mudstones and clastic rocks based on 23 distinct features [13], such as grayscale percentages and edge pixel counts. Similarly, Chatterjee applied support vector machines (SVM) to limestone identification using 40 attributes involving color, texture [14], and morphology; Cheng and Yin employed SVM with 13 spatial and textural features to classify four rock types [15]; and Izadi et al. conducted intelligent identification of 23 igneous rocks using artificial neural networks (ANN) based on 12 color parameters [16]. Although ML-based classification can effectively differentiate lithologies, it necessitates the manual extraction of features for rock delineation. Consequently, these techniques often suffer from diminished efficiency and accuracy when processing large-scale or diverse datasets [17-18].

Deep learning algorithms, such as CNNs, utilize unique hierarchical abstraction mechanisms to adaptively extract multi-scale spatial features directly from raw pixel data. This establishes an end-to-end learning framework that eliminates the need for manual feature selection, thereby further attenuating subjectivity in lithology identification [12, 19-21]. For instance, Zhang and Li constructed an intelligent identification model using Inception-V3 for rocks such as granite [20], phyllite [22], and breccia, achieving a classification accuracy exceeding 90% on the test set. Feng and Gong developed a Siamese CNN model based on AlexNet for 28 rock types, reporting a test accuracy of 89.4%. Furthermore, Ran and Xue successfully classified granite, limestone [23], conglomerate, sandstone, shale, and mylonite using a custom RTCNN architecture, attaining an accuracy of 97.96%. Building upon these foundations, researchers have deployed such models onto mobile devices, enabling rapid in-situ lithology identification for field geological surveys [24-26].

Despite the immense potential of CNNs in the domain of rock image recognition, practical application remains constrained by multi-dimensional technical bottlenecks. First, the proliferation of CNN architectures has intensified the dilemma of model selection [27-31], where traditional trial-and-error methods struggle to adapt to the multi-scale feature requirements of rock imagery. Second, data scarcity and the "annotation paradox" restrict model generalization; rock sampling is limited by geological conditions, and high-precision labeling relies on professional geologists,

resulting in a paucity of large-scale annotated datasets. Furthermore, class imbalance and inter-class similarity exacerbate decision risks. The natural distribution of rock types is uneven (e.g., sedimentary rocks significantly outnumber metamorphic rocks), and subtle textural differences between similar lithologies (such as argillaceous sandstone versus sandy mudstone) lead to persistent misclassification rates [8, 10, 32-34]. Finally, image degradation caused by environmental interference—such as uneven lighting and surface contamination during field acquisition—diminishes feature validity, as low-resolution images fail to preserve critical textural details [35-40]. Collectively, these challenges severely impede the generalization capability and engineering applicability of AI models.

Against this background, constructing a robust intelligent rock identification framework that transcends the synergistic constraints of data, models, and environment has become a focal point for both academia and industry. This paper systematically reviews the research progress of CNN-based rock image identification, focusing on model selection criteria and optimization pathways. The objective is to provide theoretical underpinning for overcoming existing technical bottlenecks.

2 INTRODUCTION TO CONVOLUTIONAL NEURAL NETWORKS

Since its inception in the 1980s, the CNN has established itself as a cornerstone tool in the field of image recognition. Its canonical architecture comprises convolutional layers, pooling layers, and fully connected layers, which synergize to realize end-to-end learning from raw input images to final classification results (Figure 1). The convolutional layer, acting as the backbone of the CNN, extracts local features through convolution operations between kernels and local receptive fields of the image. This mechanism captures critical low-level information, such as edges and textures, providing a fundamental basis for subsequent tasks. Interspersed between successive convolutional layers, the pooling layer performs down-sampling to progressively diminish the spatial dimensionality of feature representations. This process minimizes the number of model parameters and computational load while serving to mitigate overfitting. Positioned at the terminus of the architecture, the fully connected layer integrates and maps the extracted features to derive high-level semantic information. By establishing dense connectivity where each neuron links to all neurons in the preceding layer, it applies non-linear transformations via activation functions to output the final classification results.

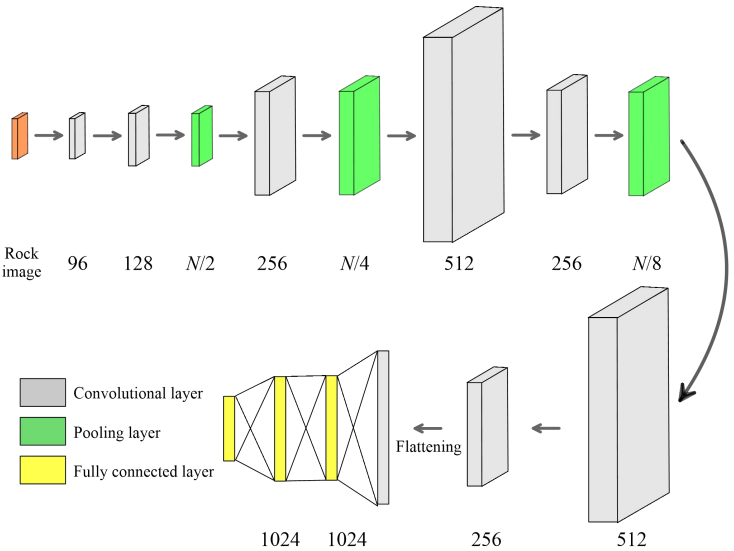

Figure 1 Architectural Diagram of the Convolutional Neural Network

Driven by the continuous evolution of the ILSVRC competition in recent years, numerous deep learning architectures demonstrating superior image classification performance have emerged, including AlexNet [27], VGG [31], Inception-V3 [29], ResNet [30], and Xception [28]. The structural and algorithmic innovations within these models have solidified the foundation for deep learning applications in image classification. Specifically, relative to conventional neural networks, AlexNet incorporated Dropout and ReLU activation functions, significantly accelerating training convergence and enhancing performance (Figure 2a). VGG16 deepened the network architecture by employing small-sized 3×3 convolution kernels while maintaining manageable parameter counts (Figure 2b). Inception-V3 introduced auxiliary classifiers and label smoothing techniques to effectively alleviate the vanishing gradient problem and ensure gradient stability, thereby improving model generalizability and preventing overfitting (Figure 2c). ResNet, through the introduction of residual units, optimized parameter configuration while achieving comparable accuracy; this design enables the model to attain desired performance levels with fewer iterations, substantially boosting training efficiency (Figure 2d). Finally, Xception adopted depthwise separable convolutions to drastically reduce the number of model parameters while preserving high performance (Figure 2e).

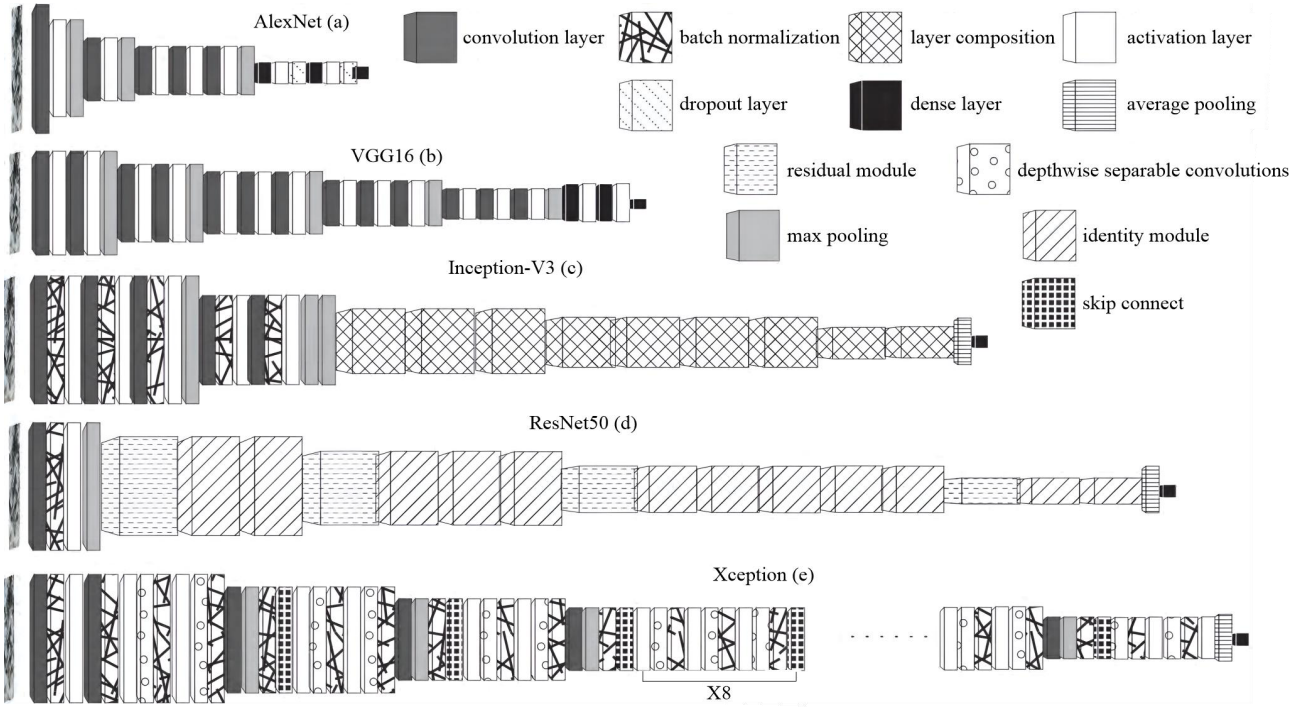

Figure 2 Schematic Diagram of a Typical Convolutional Neural Network Architecture

3 OPTIMIZATION PARADIGMS FOR CONVOLUTIONAL NEURAL NETWORK ARCHITECTURES

In the realm of deep learning, the relationship between the complexity of model architectures and their feature extraction capabilities exhibits non-linear characteristics, a phenomenon particularly pronounced in research on intelligent rock image recognition. Classical theory posits that increasing network depth facilitates the extraction of high -order abstract features, thereby enhancing image recognition performance [30]. However, recent investigations indicate that when the number of layers exceeds a critical threshold, models may experience accuracy saturation or even performance degradation [21, 36]. This finding has been validated within the domain of rock lithology identification: comparative studies involving AlexNet, ResNet, Inception, and VGG [22], as well as evaluations across AlexNet, VGG16, ResNet50, and Xception [41], and layer-depth contrast experiments with ResNet-50/101/152 [38], have all demonstrated that models with lower structural complexity often exhibit superior classification performance. These results suggest that relying solely on network depth as a selection criterion has limitations, necessitating a comprehensive trade-off between computational efficiency and engineering applicability [42]. Currently, academia primarily adopts two optimization strategies: (1) adaptation based on model characteristics, and (2) multi-model comparative selection. The advantages of these approaches in practical applications are discussed below.

3.1 Adaptation Strategies Based on Model Characteristics

In research on CNN-based intelligent rock image recognition, the selection of model architectures often involves multi-dimensional technical considerations. For instance, Bai and Yao [8] pioneered the use of the Inception-V3 architecture for constructing a rock image transfer learning model, a decision primarily grounded in the model's historical success in general image recognition tasks. As research deepened, Hu and Ye [40] proposed a dual criterion for model selection: prioritizing network architectures like ResNet-50, which possess moderate parameter counts and high training efficiency, while ensuring recognition accuracy; this approach effectively balances computational resource consumption with model performance. Notably, the preference for ResNet models by Yang and Xiong and Wang and Liu was largely driven by the architecture's ability to mitigate issues such as vanishing gradients [36, 43], exploding gradients, and network degradation during training. In contrast, distinct technical priorities exist among different research groups. For example, despite acknowledging the risks of saturation and degradation associated with increased depth, some researchers selected the ResNet-101 model [21]. Systematic experiments revealed that for complex lithology recognition tasks, extending network depth to ResNet-101 yielded significantly better feature representation capabilities than ResNet-50. These studies indicate that in the field of intelligent rock image recognition, model selection requires not only an assessment of baseline performance metrics but also a comprehensive integration of specific geological scenario requirements, computational efficiency, network depth, and gradient optimization.

3.2 Multi-Model Comparative Selection Strategy

Comparative analysis represents another pivotal methodology for selecting appropriate architectures. As illustrated in Table 1, Alzubaidi and Mostaghimi [34] systematically evaluated three typical CNN models—ResNeXt-50, Inception-

v3, and ResNet-18. They discovered that ResNeXt-50 demonstrated significant superiority in core image recognition tasks, improving accuracy by 10 percentage points compared to Inception-v3 while reducing training time by 43%; consequently, it was selected as the core architecture for their intelligent lithology identification system. This finding resonates with the large-scale model screening conducted by Ma and Ma [38], who initially filtered 13 mainstream models (including VGG16, MobileNet, and Xception) before conducting an in-depth comparison of five candidates, such as ResNet50 and Inception-v3. Their experimental results indicated that ResNet50 achieved the optimal balance between model complexity and recognition performance, establishing it as the underlying pre-trained model for lithology identification. It is worth noting that the superiority of the ResNet series has been validated across multiple independent studies. Ren and Zhang [44] compared current mainstream deep learning algorithms—including AlexNet, VGGNet, GoogleNet, and ResNet—in rock and mineral sample identification tasks, similarly finding that the ResNet50 model yielded the highest recognition accuracy, leading to its selection as the foundational model.

However, recent scholarship has revealed the scenario-dependent nature of model selection. Comparative experiments by Li targeting fresh rock section images demonstrated that the myDenseNet-all model [9], based on the DenseNet architecture, achieved optimal performance on the test set with an F1-score of 89.84% and an accuracy of 94.48%. These metrics exceeded those of the improved myResNet-all model by 2.01 and 1.72 percentage points, respectively, suggesting that dense connectivity structures possess a stronger capacity for capturing the features of fresh sections. This discrepancy was foreshadowed in early research by Feng and Gong [22], where comparative experiments showed that the accuracy (0.66) and F1-score (0.77) of ResNet-vl-50 were significantly lower than the accuracy (0.79) and F1-score (0.87) achieved by AlexNet. Moreover, Zhang and Tang [32] further validated this pattern when constructing a centimeter-scale intelligent identification system for continental shale lithology. They found that EfficientNet-B0 surpassed ResNeXt-50 in both Top-1 (60%) and Top-5 (95%) accuracy, as its compound scaling strategy was better adapted to the microstructural features of centimeter-scale shale thin sections.

A comprehensive analysis indicates that current research on intelligent rock image recognition exhibits a dual development trend characterized by "benchmark model selection + scenario-based adaptation." Although multi-model comparisons can identify high-precision architectures for specific tasks, significant disparities remain in the performance of identical or similar models across different rock image datasets.

Table 1 Comparative analysis of Convolutional Neural Network in Research Citations

Tested model	Optimal model	References
ResNeXt-50, Inceptoin-v3, ResNet-18	ResNeXt-50	[34]
VGGl6, VGG19, MobileNet, AlexNet, LeNet, ZF_Net, ResNet18, ResNet34, ResNet50, Inception-V3, ResNet152, ResNet101, Xception	ResNet50	[38]
AlexNet, VGGNet, GoogleNet, ResNet	ResNet50	[44]
DenseNet121, EfficientNet-BO, EfficientNet-B8, MobileNet-v2, MVIT-v2, ResNext50	EfficientNet	[32]
AlexNet, ResNet-vl-50, Inception-V2, VGG-19	AlexNet	[22]
VGG, ResNet, DenseNet	myDenseNet-all	[9]

4 OPTIMIZATION PATHWAYS FOR INTELLIGENT ROCK IMAGE IDENTIFICATION MODEL PERFORMANCE

Building upon diverse selection criteria, scholars have established multiple architectural frameworks for intelligent rock image identification. However, existing research indicates that model accuracy on test sets generally possesses room for optimization. The primary constraints can be categorized into three aspects: first, the insufficient scale of training data fails to meet big data volume requirements, significantly impacting model generalization capabilities [8, 10, 20, 33, 36, 39, 40, 44-49]; second, low image resolution leads to inadequate feature extraction, resulting in identification errors for lithologies with similar mineral compositions or macroscopic characteristics [8, 10, 33, 36, 38-40, 43, 50]; and third, the discriminative capability of model architectures requires enhancement through further optimization or the adoption of more refined models [10, 22, 37, 38, 43, 44, 46, 47, 51]. Based on this analysis, the following sections systematically review the latest research progress across three dimensions: data augmentation strategies, resolution enhancement technologies, and model architecture innovation.

4.1 Data Augmentation Strategies: From Limited Samples to Robust Feature Learning

The core of Convolutional Neural Networks lies in learning intrinsic feature representations of rock images via a data-driven approach. Consequently, the scale and quality of training data directly influence the model's ability to generalize lithological features. When sample sizes are insufficient, models are prone to overfitting local features, leading to a significant decline in lithological discrimination capability [20]. Currently, two primary data augmentation paradigms are employed [10, 32, 45, 48]: (1) Physical augmentation, which involves acquiring incremental raw data through multi-angle rock sampling (e.g., 3D rotational scanning, multi-spectral imaging); and (2) Digital augmentation, which expands datasets using traditional methods such as geometric transformations (flipping/rotation/translation), photometric adjustments (brightness/contrast perturbation), and noise injection (Gaussian/Salt-and-Pepper noise).

Regarding intelligent identification models based on Inception-V3, Zhang and Li observed that classification

probability values for two granite images and one breccia image in the test set were below 70% [20]. By augmenting training data through local image cropping, model accuracy rose to over 85%. This study suggests that data augmentation targeting local textural features of specific rock types can effectively improve fine-grained classification performance. However, in a study by Bai and Yao [8] expanded to 15 lithologies (approximately 1000 images per class), validation accuracy plummeted to 63%, likely due to feature confusion effects caused by crossing mineral compositions in multi-class scenarios. Notably, Xiong and Liu achieved a validation accuracy of 76.31% in a study of 8,514 mesoscopic images of typical rocks (≥1,371 per class) from the main urban area of Chongqing [39]. Although identification accuracy showed an upward trend with increasing single-class sample sizes compared to the study by Bai and Yao [8], it remained significantly lower than that achieved by Zhang and Li [20], who utilized fewer samples per class (Table 2).

In applications involving the ResNet50 model (Table 2), Hu and Ye achieved 90% test accuracy based on 1,200 samples across 8 classes (~150 per class) [40], potentially due to the effective representation of mineral paragenetic associations by deep residual networks under limited categories. However, subsequent research presents contradictory trends: Wang and Liu [36], investigating 4 lithologies (596 per class), found that even after excluding broken rock masses with developed structural joints, model identification accuracy ranged only between 75% and 90%. Similarly, Ren and Zhang [44], in a study of over 60,000 samples covering more than 100 types, observed a significant drop in accuracy to just over 50%, despite increasing single-class samples to ~600. Furthermore, Zhang and Yi expanded single -class samples to 3,912 (27,384 total across 7 classes) [46], yet test accuracy reached only 74.1%.

The VGG-16 model exhibits similar patterns (Table 2). Yang and Xiong obtained 91.6% accuracy in a study of 221 cutting images across 5 classes (≥33 per class) [43]. However, when Dong and Zhang extended the research to 3,526 cutting images across 18 classes (up to 195 per class) [33], the test set lithology identification accuracy was 87.3%, again lower than the study by Yang and Xiong which used fewer samples per class [43]. Interestingly, Zhang and Tang [32], based on 6 classes with 300 samples each, reported a test set accuracy of 94.56%, reaffirming the trend where larger single-class sample sizes correlate with higher identification accuracy.

In summary, although some studies indicate that data augmentation can enhance model performance in specific scenarios, comprehensive analysis across single models like ResNet50, Inception-V3, and VGG-16, as well as cross-model synthesis (Figure 3), reveals that increasing the number of single-class rock images yields diminishing marginal returns on identification accuracy. When the number of categories exceeds the model's representational capacity, sample size growth may even lead to accuracy degradation. This contradictory phenomenon exposes the limitations of traditional data augmentation strategies: without synchronous optimization of model capacity and feature decoupling capability, merely increasing sample quantity is insufficient to breakthrough the theoretical bottlenecks of multi-class lithology identification**.**

Table 2 Performance Comparison of Convolutional Neural Networks in Rock Image Recognition Tasks

Model	Total images	Number of images per class	Test accuracy (%)	Reference	No.
Inception-V3	9 images (3 classes)	3	87~97%	[20]	1
Inception-V3	~15,000 images (15 classes)	~1000	63 %	[8]	2
Inception-V3	8,514 images (4 classes)	＞1371	76.31%	[39]	3
ResNet50	1,200 images (8 classes)	~150	~90%	[40]	4
ResNet50	2,384 images (4 classes)	596	75%~90%	[36]	5
ResNet50	60,000 images (>100 classes)	~600	50%	[44]	6
ResNet50	27,384 images (7 classes)	~3912	74.1%	[46]	7
VGG-16	221 images (5 classes)	35~64	91.6%	[43]	8
VGG-16	3,526 images (18 classes)	195	87.3%	[33]	9
VGG16	1,800 images (6 classes)	300	94.56%	[32]	10

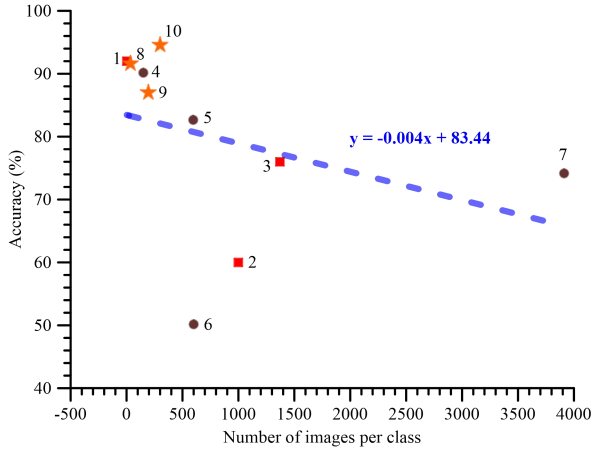

Figure 3 Correlation between Intelligent Rock Image Recognition Accuracy and sample size per class

Note: The blue dashed line represents the fitted trend for the aggregate data. Refer to Table 2 for data sources and corresponding serial numbers.

4.2 Resolution Enhancement Technologies: Capture and Limitations of Mesoscopic Features

In intelligent rock image recognition, the convergence of mineral compositions and the similarity of macroscopic characteristics constitute a dual challenge. For instance, Zhang and Tang found that models tend to misclassify shale as dolomite due to color similarity [32]. Alzubaidi and Mostaghimi discovered that ResNet-18 erroneously classified approximately 50 limestone images as sandstone and about 40 limestone images as shale [34]. Dong and Zhang noted that the identification accuracy for gray argillaceous siltstone was only 34.7% [33], with half being misclassified as dark gray silty mudstone. Tan and Tian observed that three types of tuff (RLCSF-Tuff, RSB-Tuff, RVMBV-Tuff) were frequently misjudged as the characteristically similar RCG-Tuff [10]. Bai and Yao found a 10% mutual misjudgment rate among dolomite [8], limestone, and marble due to close mineral compositions, and a 5%–10% misjudgment rate between gabbro and basalt.

To address these issues, considering that rocks with similar macroscopic features or mineral compositions often exhibit differences in mesoscopic features such as texture, roundness, and grain size [32, 38, 40], scholars have attempted to acquire high-precision rock image data to display these details, thereby improving model accuracy. For example, Xiong and Liu utilized a measuring electron microscope at 90x magnification to capture mesoscopic images of four typical rock samples—mudstone, sandy mudstone [39], argillaceous sandstone, and sandstone—from the main urban area of Chongqing, which showed obvious differences in color, flatness, grain prominence, and cementation forms (Figure 4 a-b). Building on this, a deep learning model for mesoscopic rock images was established using Inception-V3 and transfer learning. Results showed that sandstone identification accuracy in the validation set reached 97.28%. However, this study also illuminates the limitations of resolution enhancement: the identification accuracies for argillaceous sandstone and sandy mudstone, which have similar major components, were only 72.59% and 72.35%, respectively. Furthermore, argillaceous sandstone had a 14.02% probability of being mistaken for sandy mudstone, while sandy mudstone had a 12.18% probability of being mistaken for argillaceous sandstone, with mutual error rates exceeding 10%. Similarly, a recent study indicated that for limestone and tuff, which are visually close and difficult to distinguish with the naked eye (Figure 4 c-d), model differentiation remains problematic even with high-resolution images acquired via multiple magnifications and supplementary lighting [43], where the misidentification probability for limestone reached 11%.

Overall, for rock images with similar macroscopic features or mineral compositions, elevating image resolution does provide models with more learnable detailed features. However, judging from the aforementioned research results, substantial room for improvement remains in using resolution enhancement to increase identification accuracy, particularly compared to the accuracy achievable for rock images with significantly distinct macroscopic features.

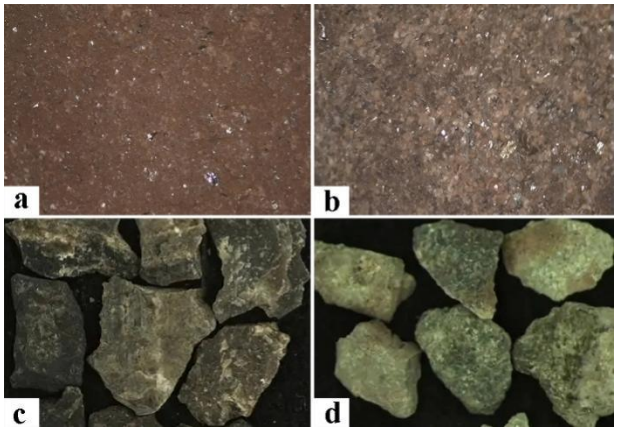

Figure 4 Examples of High-Resolution Rock Images: (a) Sandy mudstone and (b) argillaceous sandstone [39]; (c) tuff cuttings and (d) limestone cuttings [43]

4.3 Model Architecture Improvement: Adaptive Design for Complex Scenarios

Despite continuous optimization of rock image recognition models, the classification precision of existing methods still struggles to meet practical engineering requirements. Particularly in complex geological scenarios, model robustness and generalization capability face severe challenges [38, 44]. As shown in Figure 5, scholars have proposed various innovative improvement schemes tailored to different application scenarios.

In terms of feature extraction optimization, Zhang and Zhang introduced multi-scale dilated convolution attention blocks into ResNet-101 to address the need for rapid in-situ rock classification at tunnel faces (Figure 5a) [35], effectively enhancing fine-grained feature capture capabilities. To handle complex image processing demands, Zhang and Ye innovatively combined VGG16 with Principal Component Analysis (PCA) to construct a PCA-VGG16 model (Figure 5b) [2], significantly reducing computational complexity while ensuring accuracy. Feng and Gong built a Siamese convolutional neural network model based on AlexNet to balance global image information and local textural information of rock data [22]. Ran and Xue successfully resolved the issue of interference elements affecting classification results by constructing a deep CNN model [23]. Additionally, improved models can achieve substantial accuracy increases for images with similar macroscopic features. For instance, Liu and Wang adopted a simplified

VGG16 as the image feature extraction network within a Faster R-CNN deep learning object detection framework [11]; results showed that probability scores for visually similar basalts (surface amygdaloidal structures) and conglomerates (surface rounded shapes) exceeded 99%.

Regarding engineering application adaptation, Yang and He improved the AlexNet model (Figure 5c) [52], significantly enhancing identification accuracy for large-sized rock fragments during tunnel boring processes. Xu and Ma integrated a Region Proposal Network and detector within the Faster R-CNN framework [53], developing an intelligent lithology identification system suitable for on-site detection. Addressing the limitations of 2D images, Xu and Shi proposed an intelligent lithology identification method based on deep learning of rock images and elemental information [54], markedly improving identification accuracy for weathered rocks. Considering the typically small sample size of rock specimens, Zhang and Yi designed a new neural network model [46], MyNet, targeted at small-sample databases, which demonstrated accuracy superior to ResNet50 and VGG16 models with or without transfer learning (Figure 5d).

In the direction of model lightweighting, Tan and Tian improved the Xception model by introducing residual connection mechanisms [10], significantly reducing parameter count while maintaining performance. Xiao and Li integrated technologies such as ResNetv1d and deformable convolutions to improve Mask R-CNN (Figure 5e) [51], realizing real-time online detection of ore types. Yang and Xiong employed depthwise separable convolutions to improve the ResNet model (Figure 5f) [43], significantly increasing the classification efficiency of sedimentary rock cuttings.

A synthesis of existing research reveals that whether improving existing models (e.g., optimization of AlexNet/VGG16) or constructing new ones (e.g., multi-modal identification systems), the average accuracy of improved models on test sets has indeed achieved significant enhancement, despite potential concerns regarding generalization capabilities**.**

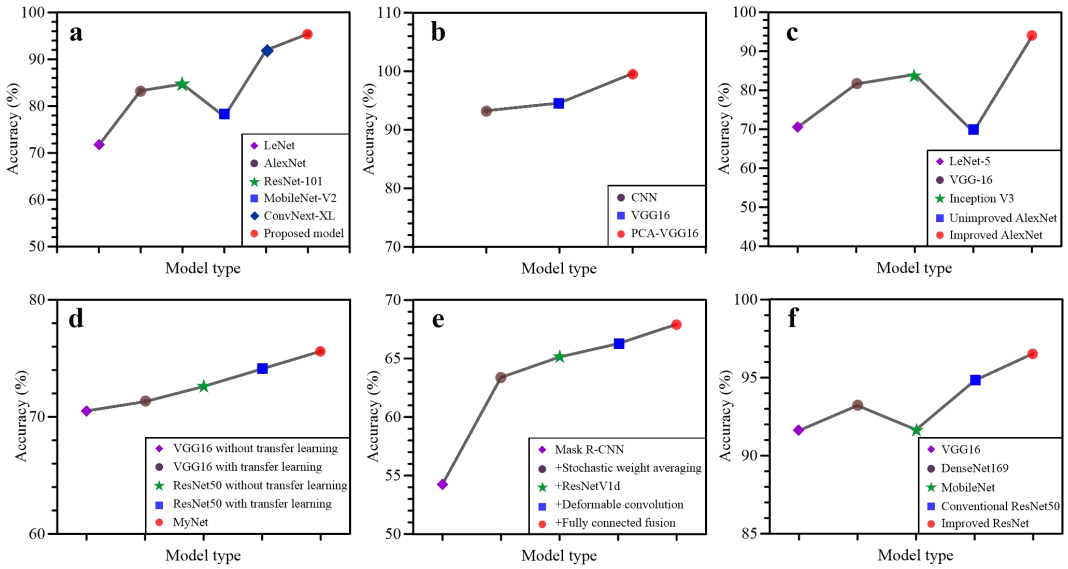

Figure 5 Performance Comparison between Optimized CNN Architectures and Benchmark Models for Rock Image Classification

Note: Data for panels (a)–(f) are derived from Zhang et al. [35], Zhang and Ye [2], Yang and He [52], Zhang and Yi [46], Xiao and Li [51], and Yang and Xiong [43], respectively.

5 CONCLUSION AND PERSPECTIVES

Research on intelligent rock image identification based on Convolutional Neural Networks has achieved significant progress, with CNNs demonstrating powerful feature extraction and classification capabilities that offer efficient solutions for this field. However, technical bottlenecks—including model selection, data scarcity, class imbalance, and environmental interference—continue to constrain model generalization capabilities and engineering applicability. While performance can be notably enhanced through data augmentation strategies, resolution enhancement technologies, and model architecture innovations, existing methods still possess limitations in complex lithology identification tasks. Therefore, constructing a robust intelligent rock image identification framework that overcomes the synergistic constraints of data, models, and the environment remains a shared challenge for both academia and industry.

Future development of intelligent rock image identification technology should focus on breakthrough research in the following dimensions:

(1) Construction of a multi-dimensional model selection assessment system. It is recommended to establish a three-dimensional decision model integrating task complexity, data distribution characteristics, and model structural parameters to form quantifiable model adaptation standards.

(2) Development of more efficient data augmentation strategies. This involves addressing data scarcity to enhance model generalization, while simultaneously fusing multi-modal data—such as rock images, elemental composition, and physical properties—to build a multi-dimensional feature representation system that improves discrimination of similar lithologies.

(3) Research on environmental adaptability technologies. This includes developing image enhancement and denoising algorithms, and designing robust models resistant to uneven illumination and weathering contamination, thereby mitigating the impact of environmental interference on image quality.
(4) Exploration of novel hybrid network architectures. Building upon the local feature extraction advantages of convolutional networks, research should introduce the global modeling capabilities of Transformers to construct hybrid architectures with multi-scale feature fusion**.**

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Liu X, Wang H, Jing H, et al. Research on intelligent identification of rock types based on Faster R-CNN method. IEEE Access, 2020, 8: 21804-21812.
[2] Zhang Y, Ye Y L, Guo D J, et al. PCA-VGG16 model for classification of rock types. Earth Science Informatics, 2024, 17(2): 1553-1567.
[3] Wu Lei, Zhai Xinwei, Wang Erteng, et al. Geochemical characteristics and formation environment of Jijitaizi ophiolite in Beishan area. Northwestern Geology, 2025, 58(1): 27-42.
[4] Meng Wuyi, Zhang Zhen, Gao Yongbao, et al. Mineral composition and geological significance of the newly discovered Wangzhuang gold deposit in Southern Qinling. Northwestern Geology, 2024, 57(4): 157-169.
[5] Liu Hao, Cui Junping, Jin Wei, et al. Geochemical characteristics and geological significance of granites in the eastern Songliao Basin. Northwestern Geology, 2024, 57(2): 46-58.
[6] Dai Xinyu, Zhou Bin, Li Xinlin, et al. Geochronology, geochemistry and tectonic significance of Miocene quartz monzonite intrusions in the north of Qitaidaban, West Kunlun. Northwestern Geology, 2024, 57(4): 191-205.
[7] Chen Yangyang, Duan Jun, Xu Gang, et al. Geochemical characteristics and tectonic significance of Late Triassic lamprophyres in Beishan area, Gansu Province. Northwestern Geology, 2024, 57(6): 78-94.
[8] Bai Lin, Yao Yu, Li Shuangtao, et al. Mineral composition analysis of rock images based on deep learning feature extraction. China Mining Magazine, 2018, 27(7): 178-182.
[9] Li Yan. Rock image recognition based on deep learning. Beijing: Beijing Forestry University, 2020.
[10] Tan Yongjian, Tian Miao, Xu Dexin, et al. Research on rock image classification and recognition based on Xception network. Geography and Geo-Information Science, 2022, 38(3): 17-22.
[11] Liu X, Wang H, Jing H, et al. Research on intelligent identification of rock types based on Faster R-CNN method. IEEE Access, 2020, 8: 21804-21812.
[12] Yuan Hang. Research on intelligent lithology identification method based on rock image feature learning. 2023.
[13] Marmo R, Amodio S, Tagllaferri R, et al. Textural identification of carbonate rocks by image processing and neural network: methodology proposal and examples. Computers & Geosciences, 2005, 31(5): 649-659.
[14] Chatterjee S. Vision-based rock-type classification of limestone using multi-class support vector machine. Applied Intelligence, 2013, 39: 14-27.
[15] Cheng Guojian, Yin Juanjuan. Rock thin section image classification based on SVM. Technology Innovation and Application, 2015, 5(1): 38.
[16] Izadi H, Sadri J, Bayati M. An intelligent system for mineral identification in thin sections based on a cascade approach. Computers & Geosciences, 2017, 99: 37-49.
[17] Chai H, Li N, Xiao C, et al. Automatic discrimination of sedimentary facies and lithologies in reef-bank reservoirs using borehole image logs. Applied Geophysics, 2009, 6: 17-29.
[18] Zhang F, Liu J, Lu X, et al. Spatial weighted graph-driven fault diagnosis of complex process industry considering technological process flow. Measurement Science and Technology, 2023, 34(12): 125143.
[19] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. Science, 2006, 313(5786): 504-507.
[20] Zhang Ye, Li Mingchao, Han Shuai. Automatic identification and classification method of lithology based on deep learning of rock images. Acta Petrologica Sinica, 2018, 34(2): 333-342.
[21] Xu Zhenhao, Ma Wen, Lin Peng, et al. Intelligent lithology identification based on transfer learning of rock images. Journal of Basic Science and Engineering, 2021, 29(5): 1075-1092.
[22] Feng Yaxing, Gong Xi, Xu Yongyang, et al. Lithology identification method based on fresh rock surface images and Siamese convolutional neural networks. Geography and Geo-Information Science, 2019, 35(5): 89-94.
[23] Ran X, Xue L, Zhang Y, et al. Rock classification from field image patches analyzed using a deep convolutional neural network. Mathematics, 2019, 7: 755.
[24] Fan G, Chen F, Chen D, et al. Recognizing multiple types of rocks quickly and accurately based on lightweight CNNs model. IEEE Access, 2020, 8: 55269-55278.
[25] Wang C, Li Y, Fan G, et al. Quick recognition of rock images for mobile applications. Journal of Engineering Science and Technology Review, 2018, 11(4): 111-117.
[26] Fan G, Chen F, Chen D, et al. A deep learning model for quick and accurate rock recognition with smartphones. Mobile Information Systems, 2020, 2020: 1-14.

[27] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017, 60(6): 84-90.
[28] Chollet F. Xception: deep learning with depthwise separable convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[29] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[30] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[31] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv, 2015.
[32] Zhang Z, Tang J, Fan B, et al. An intelligent lithology recognition system for continental shale by using digital coring images and convolutional neural networks. Geoenergy Science and Engineering, 2024, 239: 212909.
[33] Dong Wenhao, Zhang Huai. Lithology identification of cuttings based on transfer learning. Journal of University of Chinese Academy of Sciences, 2023, 40(6): 743-750.
[34] Alzubaidi F, Mostaghimi P, Swietojanski P, et al. Automated lithology classification from drill core images using convolutional neural networks. Journal of Petroleum Science and Engineering, 2021, 197: 107933.
[35] Zhang W, Zhang W, Zhang G, et al. Hard-rock tunnel lithology identification using multi-scale dilated convolutional attention network based on tunnel face images. Frontiers of Structural and Civil Engineering, 2024, 17(12): 1796-1812.
[36] Wang Xiaobing, Liu Lin, Wang Junqing, et al. Research on lithology identification of rock images based on convolutional neural network ResNet50 residual network. Geotechnical Engineering Technique, 2024, 38(3): 294-302.
[37] Xiong Feng, Liao Yifan, Cao Weiteng, et al. Automatic lithology identification based on convolutional neural network-deep transfer learning. Safety and Environmental Engineering, 2023, 30(4): 26-34.
[38] Ma Zedong, Ma Lei, Li Ke, et al. Multi-scale lithology identification based on deep learning of rock images. Bulletin of Geological Science and Technology, 2022, 41(6): 316-322.
[39] Xiong Yuehan, Liu Dongyan, Liu Dongsheng, et al. Automatic lithology classification method based on deep learning of mesoscopic images of rock samples. Journal of Jilin University (Earth Science Edition), 2021, 51(5): 1597-1604.
[40] Hu Qicheng, Ye Weimin, Wang Qiong, et al. Research on lithology identification based on big data of geological images. Journal of Engineering Geology, 2020, 28(6): 1433-1440.
[41] Chen Zhongliang, Yuan Feng, Li Xiaohui, et al. Interpretability study of deep transfer learning for images of plutonic intrusive rocks from Dabie Mountain. Geological Review, 2023, 69(6): 2263-2273.
[42] Xu Z, Ma W, Lin P, et al. Deep learning of rock microscopic images for intelligent lithology identification: neural network comparison and selection. Journal of Rock Mechanics and Geotechnical Engineering, 2022, 14(4): 1140-1152.
[43] Yang Lei, Xiong Chang, Liu Wenchao, et al. Research on lithology identification of cuttings based on improved ResNet deep residual network. Journal of Yangtze University (Natural Science Edition), 2023, 20(2): 11-19.
[44] Ren Wei, Zhang Sheng, Qiao Jihua, et al. Intelligent identification of rock and minerals based on deep learning. Geological Review, 2021, 67(S1): 1281-1282.
[45] Han Xinhao, He Yueshun, Chen Jie, et al. Research on intelligent identification of rock lithology based on Swin Transformer. Modern Electronics Technique, 2024, 47(7): 37-44.
[46] Zhang Chaoqun, Yi Yunheng, Zhou Wenjuan, et al. Small sample rock classification based on deep learning and data augmentation technology. Science Technology and Engineering, 2022, 22(33): 14786-14794.
[47] Liu Xiaobo, Wang Huaiyuan, Wang Liancheng. Faster R-CNN method for intelligent identification of rock types. Modern Mining, 2019, 35(5): 60-64.
[48] Xu Shuteng, Zhou Yongzhang. Experimental research on intelligent identification of microscopic ore minerals based on deep learning. Acta Petrologica Sinica, 2018, 34(11): 3244-3252.
[49] Theodoridis S. Machine learning: a Bayesian and optimization perspective. Academic Press, 2015.
[50] Bai Lin, Wei Xin, Liu Yu, et al. Rock thin section image recognition based on VGG model. Geological Bulletin of China, 2019, 38(12): 2053-2058.
[51] Xiao Chengyong, Li Qing, Li Hui, et al. Ore type detection algorithm based on improved Mask R-CNN. Sintering and Pelletizing, 2024, 49(2): 65-73, 106.
[52] Yang Z, He B N, Liu Y, et al. Classification of rock fragments produced by tunnel boring machine using convolutional neural networks. Automation in Construction, 2021, 125: 103612.
[53] Xu Z, Ma W, Lin P, et al. Deep learning of rock images for intelligent lithology identification. Computers & Geosciences, 2021, 154: 104799.
[54] Xu Z, Shi H, Lin P, et al. Intelligent on-site lithology identification based on deep learning of rock images and elemental data. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.