Research on Iron Ore Classification Method based on Improved Shufflenet-v2 Network

Jiahao Lin, Jiyang Wang

Abstract—**Conventional iron ore sorting methods are intricate and time-intensive, impeding precise characterization and resulting in misclassifications. This study introduces a novel iron ore image detection model utilizing the ShuffleNetV2 architecture, incorporating a hybrid attention mechanism merging Selective Kernel Attention (SK) and Efficient Channel Attention (ECA). The H-Swish activation function is implemented in lieu of the standard ReLU, alongside depthwise separable convolutions within the lightweight network design, with modifications to the stacked cell quantity. Assessments conducted on four in-house iron ore datasets demonstrate the efficacy of the SEH-ShuffleNetV2s model, boasting a streamlined, efficient structure. The enhanced model accuracy reaches 94.2%, a 2.1% enhancement over the original ShuffleNetV2, with reduced model parameters. Comparative analysis reveals that the SEH-ShuffleNetV2s model outperforms counterparts in terms of parameter efficiency, accuracy, and expedited detection, meeting the demands for real-time iron ore identification.**

Index Terms—**Selective Kernel Attention, ECA attention, Image classification, SEH-ShuffleNetV2 network**

I. Introduction

ORE sorting plays a pivotal role in the mining process. Traditional ore sorting methods are often characterized by complexity and inefficiency, leading to issues such as high energy consumption, considerable costs, poor environmental sustainability, and suboptimal efficiency[1]. These factors have resulted in relatively low ore production efficiency in China. Traditional ore sorting methods not only require significant time and labor but also involve cumbersome processes that hinder the overall sorting speed, making it challenging to meet the requirements of rapid and real-time production. Therefore, implementing intelligent sorting methods can accelerate the rapid sorting of iron ore, significantly reducing subsequent production and labor costs.

In recent years, traditional machine learning and deep learning techniques have been widely used in automated feature extraction during the analysis of mineral images. Regarding established machine learning algorithms, Patel et al. [2] utilized a machine vision-based support vector machine to classify iron ores according to grades, they extracted 18 features from 2,200 images and achieved a classification error rate of 0.27%. However, this model depends on manually provided features and requires labor-intensive feature extraction techniques. Singh et al. [3] employed radial basis function neural networks to classify iron ores in manganese smelting plants, achieving an accuracy of only 88.71%. When traditional machine learning algorithms are applied to complex datasets, they may introduce biases that influence recognition results and demand substantial processing time. In contrast, deep learning can extract image features with greater precision, thereby reducing information bias. Deep learning has recently made remarkable progress in the field of mineral classification. For example, Apel et al. [4] applied transfer learning. They froze all convolutional layers of VGG16 and customized the fully connected layers, achieving an ore classification accuracy of 82.5%. However, the limited sample size limits its generalization ability. Wang et al. [5] utilized the Wu-VGG19 transfer network to conduct binary classification of black tungsten ore and surrounding rock, achieving a recognition rate of 97.51%. Baraboshkin et al. [6] and Bai Lin et al. [7] employed Inception v3 to classify 5 and 15 types of ores. Xiao D et al. [8] utilized an infrared spectrometer to acquire spectral images of ores. These images were then fed into a custom convolutional neural network (CNN) for training, achieving an overall accuracy of 98.11% for hematite, granite, magnetite, chrysotile, and chlorite. Nevertheless, these methodologies often involve complex architectures and a large number of parameters, resulting in suboptimal training efficiency and limited practicality. To address this issue, the present study proposes a simplified iron ore image detection model based on ShuffleNetV2. The improved model effectively overcomes the limitations imposed by traditional methods in the detection of iron ore.

The primary contributions of this research are outlined below:

The model integrates SK and ECA attention mechanisms to improve its feature extraction capabilities, allowing it to selectively amplify relevant features while suppressing irrelevant information.

The H-Swish activation function was employed instead of ReLU to address the "neuronal deactivation" phenomenon present in the original model and to prevent the vanishing of negative gradients.

The number of stacked network units is reduced, this simplification leads to a less complex network structure and replaces the traditional max-pooling layer with depthwise separable convolution. A Dropout layer with a dropout rate of

Manuscript received November 26, 2024; revised May 17, 2025.

This work was supported by Science and Technology Plan Natural Science Foundation of Liaoning Province. The project number is 2024-MSLH-347.

Jiahao Lin is a postgraduate student of School of Artificial Intelligence, Shenyang University of Technology, Liaoning Shenyang, 110000, China (e-mail: 2089465899@qq.com).

Jiyang Wang is an associate professor of School of Artificial Intelligence, Shenyang University of Technology, Liaoning Shenyang, 110000, China (corresponding author to provide e-mail: wjystu@163.com).

0.3 is added before the final fully connected layer, effectively mitigating model overfitting.

II. Related Work

A. CNNs and Lightweight Neural Networks

Convolutional Neural Networks (CNNs) have been widely applied in various fields, such as image processing, video analysis, and speech recognition. These networks are designed to extract features by performing convolutional operations. The following section will discuss some of the more prominent CNN architectures, such as ResNet [9], DenseNet [10], and Transformer. These architectures have attracted much attention. The development of sophisticated architectures has brought about remarkable progress in feature extraction capabilities. However, these architectures often require significantly greater depth and width than previous networks, leading to millions or even billions of parameters. This increased complexity demands larger storage capacity and substantial computational resources. In platforms with limited computational resources, such as those used in iron ore detection, the necessity for efficient and lightweight algorithms is utmost importance. The goal of lightweight networks is to optimize the network architecture and reduce the number of parameters. In this way, high efficiency and low resource consumption can be achieved while maximizing performance. The following will discuss classic lightweight neural networks, such as MobileNet [11]-[13], ShuffleNet [14],[15], and GhostNet [16]. These networks mainly use depthwise separable convolution (DWConv).

B. ShuffleNet V2

Ma et al. suggested enhancing algorithm efficiency by advocating for the preservation of a consistent channel count in the ShuffleNetV2 network, an evolution of ShuffleNetV1. This network minimizes convolutional computations and group numbers while integrating channel partitioning and shuffling techniques. Serving as a lightweight architecture for high-performance Convolutional Neural Networks (CNNs), ShuffleNetV2 adeptly harmonizes speed and accuracy.

In the ShuffleNetV2 network model, convolutional blocks are divided into two modules: the basic module (a) unit1 and the downsampling module (b) unit2. In the network architecture, the input feature map is first partitioned into channels, generating two branches, as shown in Figure 1a. The right branch incorporates three convolution operations: one 3×3 depthwise-separable convolution and two 1×1 convolutions. Conversely, the left branch consists of an identity mapping. As illustrated in Figure 1b, in the absence of channel partitioning, the convolution operations of the right branch are characterized by a depthwise-separable convolution with a stride of two. Conversely, the left branch incorporates a depth-separable convolution with a stride of two and a standard convolution. Finally, the outputs of the two branches are concatenated and then subjected to channel shuffling to enable information exchange. The authors proposed the concept of channel shuffle, which involves a sparsely connected channel approach. This approach divides the input feature map into multiple subgroups and uses different convolution kernels to perform group convolutions. This process enables information exchange among different groups.

The fundamental feature extraction module of the ShuffleNetV2 model consists of several key components. "Conv" is an abbreviation for "standard convolution," while "BN" denotes "batch normalization". The term "DWConv" represents "depthwise separable convolution," which includes both depthwise and pointwise convolutions. "ReLU" represents the term activation function, and "Concat" denotes channel splicing. As shown in Figure 1, unit 1 corresponds to the Basic module, and unit 2 is designated as the Downsampling module. The ShuffleNetV2 unit is employed in stage 2, stage 3, and stage 4.

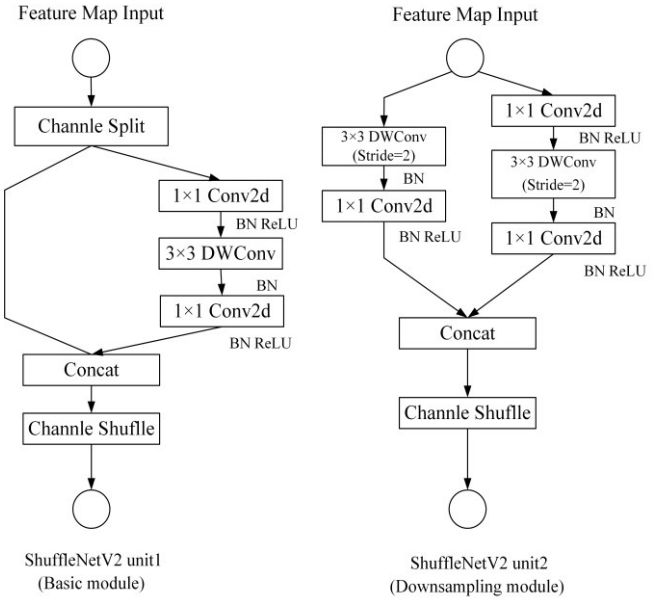


Fig. 1. ShuffleNetV2 unit.

It can be argued that the ShuffleNetV2 network is most distinct from other traditional deep learning networks because of its relatively small network parameter scale. The parameter counts of commonly used deep learning networks, such as ResNet50, GoogLeNet, and EfficientNet, are tabulated in Table I.

TABLE I
Comparison Of Parameter Counts For Different Networks

Network Name	Number of Parameters/M	Computational Load Flop/G
ResNet50	25.66	4.11
GooleNet	10.31	1.50
EfficientNet	6.54	0.59
ShuffleNetv2 2×	7.40	0.59
ShuffleNetv2 1×	2.30	0.146
ShuffleNetv2 0.5×	0.35	0.041

C. Attention Mechanism

The attention mechanism is a sophisticated technique that emulates the selective focusing process inherent in human vision. It has been widely applied in the field of deep learning, particularly when handling sequential and image data. The schematic diagram of the SK attention and ECA attention module is shown in Figure 2 and Figure 3.

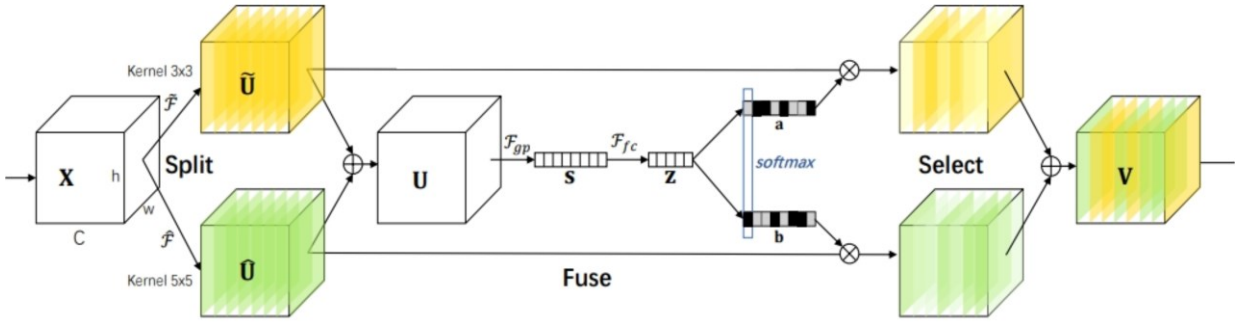


Fig. 2. Selective Kernel Convolution.

(1) Selective Kernel Attention

The main features and functional aspects of the SK module[17] are described below, the schematic diagram of the SK attention module is shown in Figure 2.

Adaptive rescaling: Initially, the SK module acquires the global feature representation for each channel through global average pooling, commonly referred to as the "squeeze" operation. Subsequently, it employs a fully connected layer to generate channel weights. These weights are utilized to rescale the features adaptively, and this process is termed "excitation".

Depth separable convolution: Conventionally, the SK module utilizes depthwise separable convolutions. The purpose is to decrease computational complexity and reduce the number of parameters, all the while maintaining the feature representation capability.

Channel selection: Through the application of channel weighting mechanisms, the SK module is able to automatically identify which channels are more crucial for the current task. Consequently, this enhances the quality of the feature representation.

Improved information flow: The SK module adjusts channel information dynamically. This adjustment enhances the model's capability to capture vital details across various feature levels.

The SK module substantially enhances both performance and efficiency of the model. This is particularly true for visual tasks such as image classification, where it is widely used.

(2) ECA attention mechanism

The ECA attention mechanism[18] is an extended version of the SE variant. It is incorporated into the relevant model. Subsequently, in the ECA module, a 1×1 convolutional layer is added after the global average pooling layer, thereby eliminating the necessity of a fully connected layer. This approach prevents dimensionality reduction, captures cross-channel relationships, and attains high performance while utilizing parameters efficiently. Specifically, the ECA module employs 1D convolutions to expedite cross-channel information exchange, and the kernel size is adaptively determined by a function. This enables layers with a larger number of channels to have more extensive cross-channel interactions. The primary objective of the ECA attention mechanism is to adaptively rebalance the weights of channel features. This enables a more focused emphasis on prominent features while suppressing less significant ones. This enhances the network's representational ability without substantially increasing the number of parameters or computational complexity. The ECA module achieves remarkable performance while keeping the model complexity relatively low. The ECA attention mechanism contributes to a further enhancement of the performance of ShuffleNetV2. The structure of the ECA module is shown in Figure 3. The operation of the ECA module will be described in detail below.

After acquiring the aggregated features through Global Average Pooling (GAP), the ECA module produces channel weights by applying a fast 1D convolution with a kernel size of k. Specifically, the value of k is adaptively determined based on a meticulously designed mapping of the channel dimension C. The operational mechanism of the ECA module relies on a specific mapping function. The details of the adaptive function will be described as follows.

$$k = \varphi(C) = \left| \frac{\log_2(C) + b}{\gamma} \right|_{odd} \tag{1}$$

Where $|\ |$ represents the nearest odd number to t. In the experiments, $\gamma = 2$ and $b = 1$. Through this mapping, high-dimensional channels have longer interactions, while low-dimensional channels use non-linear mapping for shorter interactions.

Finally, the channel weights are calculated using the one - dimensional convolution (1D) with a kernel size of k. The formula is as follows.

Finally, calculate the channel weight using the one-dimensional volume(1D) of the volume kernel k, the formula is as follows.

$$w = \sigma\left(C1D_k(y)\right) \tag{2}$$

Here, w represents the channel weight, σ denotes the sigmoid function, and C1D stands for a one-dimensional convolution with the sigmoid function applied to it.

The ECA module has fewer parameters than SE module,which consequently reduces the model's complexity. This design choice of the ECA module ensures excellent performance and remarkable efficiency.

The attention mechanism enables the model to focus autonomously on more crucial information during the input data processing stage, thereby enhancing the model's performance and efficiency.

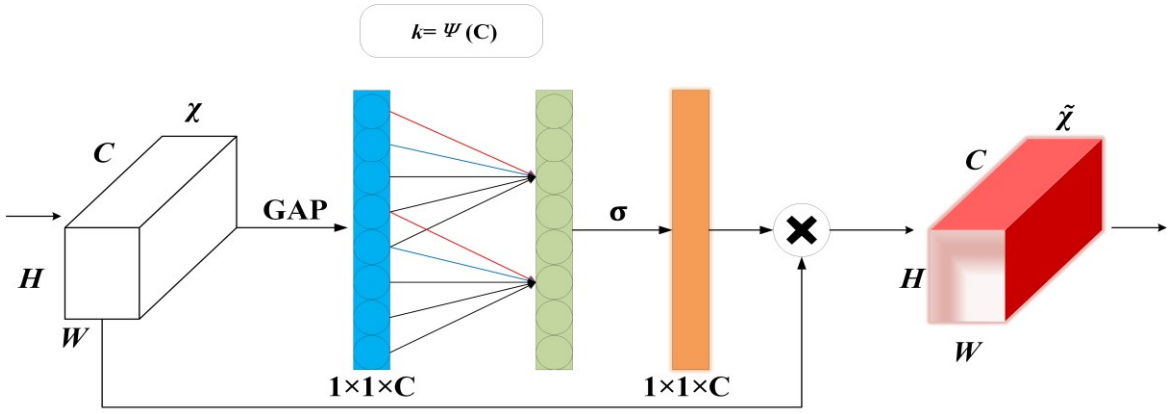


Fig. 3. Efficient Channel Attention.

D. H-Swish and ReLU Activation Function

As depicted in Equation (3), the Swish activation function surpasses the ReLU activation function and notably enhances the accuracy of the network. However, the computational and differentiable processes of the Swish activation function are intricate, posing a challenge to quantization. To address this issue, this paper employs the H-Swish activation function. The H-Swish activation function was introduced in MobileNetV3 as an approximation of the Swish activation function. The corresponding formulas are presented in Equation (4) and Equation (5).

Compared with the Swish activation function, H-Swish activation function reduces computational complexity and is more suitable for hardware acceleration. In contrast to the traditional ReLU activation function, the H-Swish activation function offers a smoother gradient flow and circumvents the common "dead neuron" issue. For negative input values, the H-Swish activation function provides smooth transitions rather than causing the gradient to vanish.

Owing to its straightforward computational architecture, the H-Swish activation function operates with higher efficiency on actual hardware platforms, reducing the dependency on computational resources, especially in the context of low-power devices.

$$Sigmoid(x) = \sigma(x) = \frac{1}{1+e^{-x}} \tag{3}$$

$$Swish(x) = x \cdot \sigma(x) \tag{4}$$

$$H-Swish(x) = x \cdot \frac{ReLU6(x+3)}{6} = \begin{cases} 0, x \le -3 \\ x, x \ge +3 \\ x \cdot (x+3)/6, otherwise \end{cases} \tag{5}$$

$$ReLU6(x) = min(max(x,0),6) \tag{6}$$

The initial ShuffleNet V2 network employs the ReLU activation function. The primary advantages of the ReLU activation function include its computational simplicity, high operational efficiency, and its capability to alleviate issues like gradient vanishing and overfitting. However, one limitation of the ReLU activation function is that it sets negative gradients to zero. This can potentially cause neurons to become inactive since not all input data is activated by it. The formula for ReLU is given in equation (7).

$$ReLU(x) = max(0,x) = \begin{Bmatrix} x & x > 0 \\ 0 & others \end{Bmatrix} \tag{7}$$

The graph of Swish and H-Swish activation functions are shown in Figure 4.

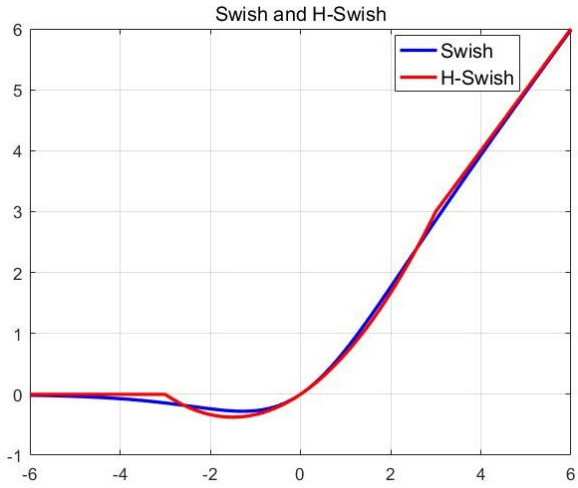


Fig. 4. Swish and H-Swish activation functions.

By effectively replacing the ReLU activation function with the H-Swish activation function, the model's nonlinear modeling capability is strengthened. In comparison with ReLU, the H-Swish activation function shows smoother behavior at the boundaries. This helps alleviate the gradient vanishing problem. Additionally, it demonstrates higher computational efficiency, which in turn improves the overall training speed and stability. Consequently, the adoption of the H-Swish activation function results in a substantial improvement in both the overall training speed and stability.

III. Proposed Method

A. SEH-ShuffleNetV2s Model Structure

The original architecture of the ShuffleNetV2 1× network and the improved network are depicted in Figure 5.

In Figure 5(a), the input feature map has dimensions of 224×224×3. Initially, 24 standard 3×3 convolutions with a stride of 2 are employed to extract features. Subsequently, downsampling is carried out via a max-pooling layer. The network is then split into three distinct stages, each contained

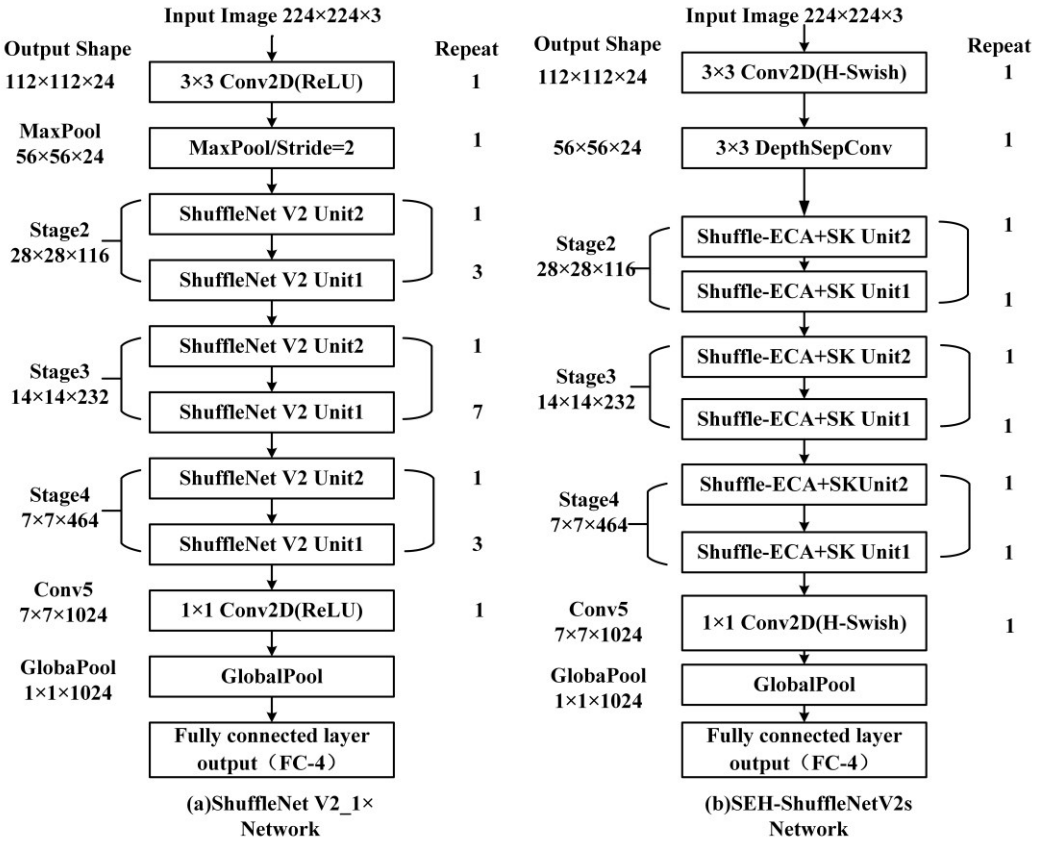


Fig. 5.The original and improved network structure.

multiple ShuffleNet V2 units. The repeat of unit2 and unit1 are 1:3, 1:7, 1:3. At the end of each stage, a standard 1×1 convolution is applied to increase the number of channels. This is combined with a global pooling layer to integrate feature information and prevent overfitting. The number of channels in each stage is 116, 232, and 464.

The improved network structure is shown in Figure 5(b). In contrast to the original ShuffleNet V2 1× presented in Figure 5(a), the enhanced network integrates SK modules into Unit1 and Unit2, thus improving the network's capacity to capture image details. Meanwhile, upon the introduction of the ECA attention mechanism, the interdependence among the feature map channels is reinforced. This enables the intelligent enhancement or suppression of feature information and further boosts the network performance. To tackle the problem of "neural deactivation", the H - Swish activation function is employed to substitute the traditional ReLU activation function in the standard 3×3 and 1×1 convolution operations. Since we are only classifying four types of iron - ore images, the number of times Unit1 and Unit2 are stacked in each stage is reduced to one to simplify the network structure while ensuring performance. Additionally, for more precise feature extraction, a 3×3 depth separable convolution with a stride of 2 is employed to substitute the conventional max-pooling layer. Before the final fully - connected layer, a dropout layer (with a dropout rate of 0.3) is also added to effectively mitigate the model's overfitting problem.

B. Hybrid Attention Mechanism

SK module: For the convolution layer within each stage, multi-scale convolution kernels (including 3x3, 5x5, etc.) can be used. Moreover, the size of the convolution kernel that is suitable for the current input can be dynamically chosen through the selection mechanism of the SK module. This, in turn, improves the capability to extract information across different scales.

ECA module: After the convolution operation in each stage, the ECA module is incorporated to carry out weighted processing for the channel attention mechanism. It dynamically assigns a weight to each channel with the aim of enhancing the features of important channels while suppressing the features of unimportant channels.

The module diagrams after the conversion are shown in Figure 6.

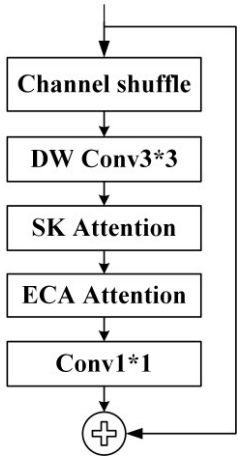


Fig. 6. The module after the conversion.

The SK module facilitates multi-scale feature extraction. Through the adaptive selection of the convolution kernel size, the model's capacity to represent information across various

scales can be improved. The ECA module offers an effective channel attention mechanism. This mechanism not only reduces the computational cost but also efficiently captures the interdependencies among channels and boosts the feature representation ability. When these two attention mechanisms are integrated with ShuffleNetV2, the performance of the model can be further enhanced. This combination allows maintaining efficient computational performance while improving accuracy.

C. Activate function replacement module

The H-Swish activation function integrates the ReLU6 function with constant multiplication to eliminate the need for exponent calculations and sigmoid operations, particularly beneficial in hardware acceleration contexts like GPUs and TPUs, reducing computational complexity. This activation function offers a seamless activation mode akin to Swish but is more streamlined than Sigmoid. By addressing the "dead neuron" issue of ReLU and maintaining a favorable gradient flow with a smoother activation pattern that avoids the abrupt "dead zone" of traditional ReLU in negative intervals, H-Swish enhances model stability during training, effectively preventing the dead neuron problem. Its simple hardware implementation efficiently enhances inference speed, especially in edge computing, reducing energy consumption and improving performance under resource constraints while maintaining computational efficiency.

The ReLU activation function following the initial layer's convolution operation has been substituted with H-Swish. Within each bottleneck module, the ReLU comprises the ReLU following the 1×1 convolution and the ReLU at the module's output. Furthermore, the ReLU preceding the fully connected layer has also been exchanged.

D. Network Structure Adjustment

Since this paper only aims at classifying four different types of iron ore images, the classification task is relatively straightforward. Consequently, the depth of the required network model does not have to be excessive. Therefore, in order to decrease both the parameter count and the computational complexity, the number of ShuffleNetV2 Unit1 stacks in Stage 2, Stage 3, and Stage 4 of the original network is decreased to one. After the 3×3 conventional convolution, the original network employs the max-pooling layer for downsampling to decrease the parameter number. In contrast, the enhanced network substitutes the max-pooling layer with a 3×3 depth-separable convolution. This convolution has a relatively small parameter count and a stride of 2. During the training process, the majority of the convolution layers are frozen, and only the last fully connected layer or some of the upper layers of the network are updated. The weights derived from the pre-training of ShuffleNetV2 can be utilized to reduce the training time and avoid the necessity of training the entire network from the beginning.

After the improvement and experimental verification of the methods mentioned above, it was determined that these enhancements were effective and could significantly boost the accuracy and speed of iron ore image classification. In the following section, the experimental results will be presented together with the corresponding analysis.

IV. Experimental And Analysis

A. Experiment Details

(1) Dataset and environment

This study utilizes an iron ore dataset assembled by our research group via web crawling and on-site imaging. The dataset is composed of images of four common types of iron ore, namely hematite, magnetite, siderite, and chlorite, and shows a substantial data imbalance. To address the problem of data imbalance, this study initially conducts data equalization and subsequently expands the datasets by applying simple random cropping and mirroring techniques. Seventy percent of the dataset is allocated for training, and thirty percent is set aside for testing. The training set contains 3947 iron ore images, and the test set contains 1692 iron ore images.

The network model for classifying iron ore images was developed in the Pycharm integrated development environment (IDE) by leveraging the Pytorch framework. It was trained on an NVIDIA GeForce GTX 1060Ti graphics processing unit (GPU) with the utilization of Pytorch version 2.3.1 and Python version 3.11.

(2) Parameter Settings

In the experiment, the Stochastic Gradient Descent with Momentum(SGD) optimizer was employed. It had a momentum value of 0.9, an initial learning rate of 0.001, and a weight decay coefficient of 3E-4. The training process comprised one hundred iterations, where the batch size for each iteration was set to 32. The dropout rate was set to 0.3. The loss function employed was the cross-entropy loss. The images were randomly rotated horizontally, cropped to a size of 224×224 pixels, and uniformly normalized. Subsequently, their final dimensions were standardized to 224×224×3.

B. Experiment Result

During the training of deep learning network models, accuracy serves as a key metric for measuring the correctness of model predictions, while the loss value acts as an indicator to quantify the discrepancy between predicted results and actual outcomes. These two metrics are fundamental in evaluating model performance: accuracy reflects the proportion of correctly classified instances, directly indicating the model's prediction precision; the loss value, calculated through specific cost functions (e.g., cross-entropy loss or mean squared error), measures the cumulative difference between predictions and ground-truth labels, guiding the optimization process by providing gradients for parameter updates. Together, they offer complementary insights into model behavior, enabling researchers to balance prediction accuracy with the minimization of systematic errors during iterative training.

The experimental results of the original network training are depicted in Figure 7. It can be noticed that the loss value for the training set is comparatively low, and the accuracy of the test set is high, reaching 92.1%.

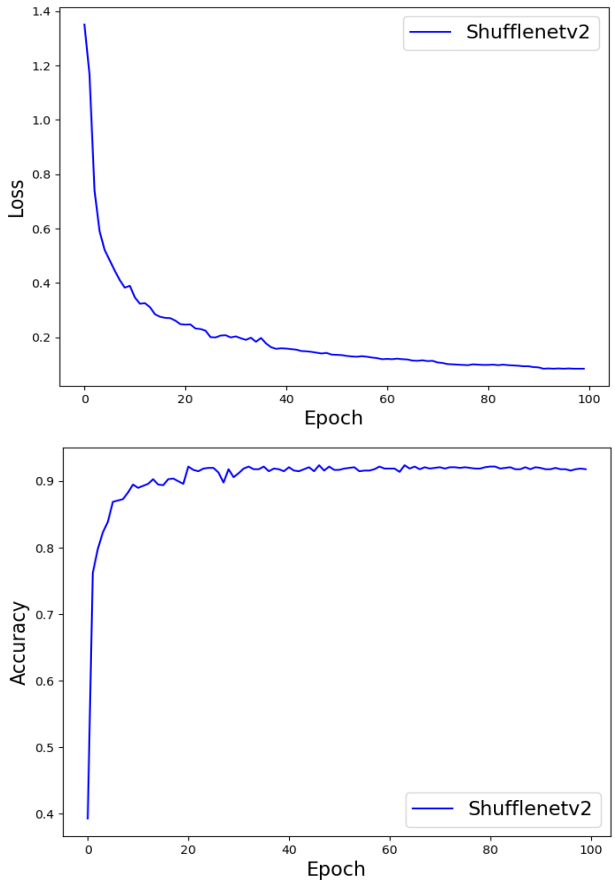


Fig. 7. Loss and accuracy curves for original Shufflenetv2 1×

The results of the modifications made to the original network are illustrated in Figure 8.

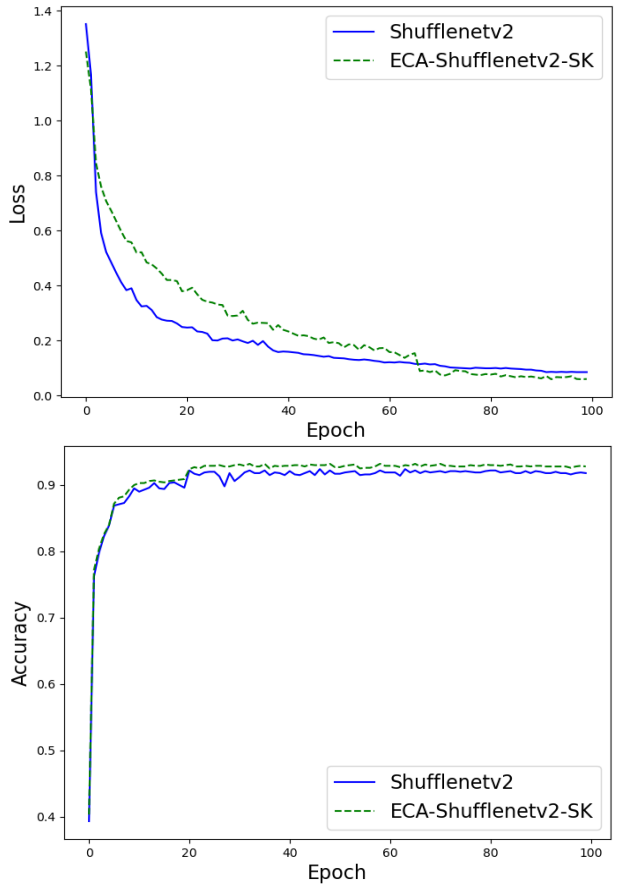


Fig. 8. Loss and accuracy curves for ECA-Shufflenetv2-SK and original network.

In Figure 8 Shufflenetv2 represent the original network, whereas ECA-Shufflenetv2-SK denotes the network after integrating the combined attention mechanism. As is evident, compared with the original network, the network incorporating the SK and ECA attention mechanisms in Unit 1 and Unit 2 exhibits higher accuracy, faster convergence, and smaller fluctuations during the convergence process. Compared with the original network, the training set loss of the network with the addition of hybrid attention mechanism is 0.16 lower than that of the original network; The accuracy of the test set is 1.1% higher than that of the original network. Moreover, the loss value is notably lower, which further validates the effectiveness and reliability of the improved network, as well as the enhanced performance brought about by these alterations.

This study presents several modifications to the existing network, including replacing the ReLU activation function with the H-Swish activation function and adjusting the stacked unit ratio in Stages 2, 3, and 4 to 1:1. The efficacy of these modifications is depicted in Figure 9,

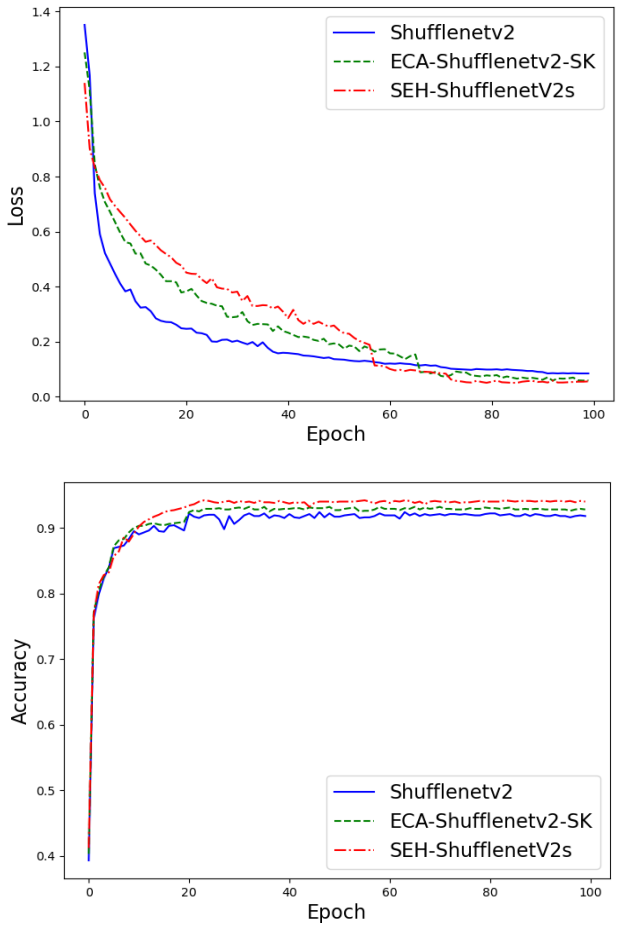


Fig. 9. A comparison chart of the loss values and accuracy of three models.

In Figure 9 Shufflenetv2 represents the original network, ECA-Shufflenetv2-SK denotes the network with the hybrid attention mechanism, and SEH-Shufflenetv2s signifies the enhanced network. The improved network demonstrates notably superior accuracy, reduced loss, and quicker convergence. Compared with the original network, the improved SEH-Shufflenetv2s network has a training set loss 0.21 lower than the original network, the accuracy of the test set is 2.1% higher than that of the original network. These

outcomes validate the effectiveness of the proposed enhancements, enhancing network performance and classification accuracy. The application of the proposed methodologies and subsequent comparative experimental analysis has led to a marked decrease in training set loss and a significant enhancement in test set accuracy. Subsequent sections will present various evaluation metrics for both the original and enhanced networks, thereby affirming the effectiveness and feasibility of the implemented enhancements.

C. *Analysis and Evaluation*

This study employs precision, recall, and accuracy as the metrics to assess network performance. In the confusion matrix, TP (True Positive) refers to the number of samples correctly classified as positive, FP (False Positive) refers to the number of samples incorrectly classified as positive, FN (False Negative) refers to the number of samples incorrectly classified as negative, and TN (True Negative) refers to the number of samples correctly classified as negative. The specific formula for the classification evaluation index can be obtained from the confusion matrix shown in Table II.

TABLE II
CONFUSION MATRIX

Confusion matrix	Actual results	
Forecast	TP	FP
results	FN	TN

Precision measures the proportion of correctly classified positive samples out of the total number of positive samples identified by the model.

$$P = \frac{TP}{TP + FP} \tag{8}$$

Recall is a measure of the proportion of correctly classified true positive samples out of all true positive samples.

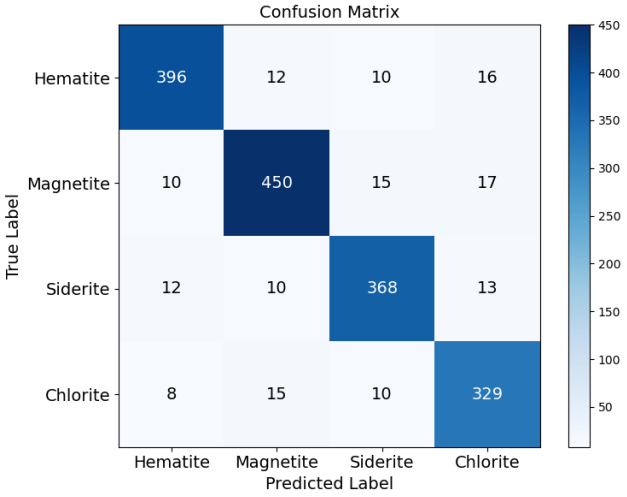
$$R = \frac{TP}{TP + FN} \tag{9}$$

Accuracy refers to the proportion of correctly identified samples (both positive and negative) among the total number of samples.

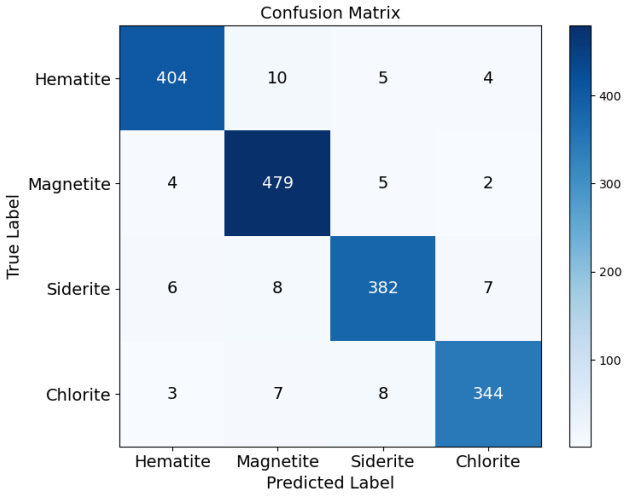
$$Acc = \frac{TP + TN}{TP + TN + FN + FP} \tag{10}$$

These metrics are employed to assess the classifier's ability to correctly classify both positive and negative samples in terms of accuracy and overall performance. By leveraging these quantitative metrics, researchers can gain a more comprehensive understanding of the classifier's performance across different scenarios. This understanding facilitates the optimization of the classification algorithm and enhances its effectiveness.

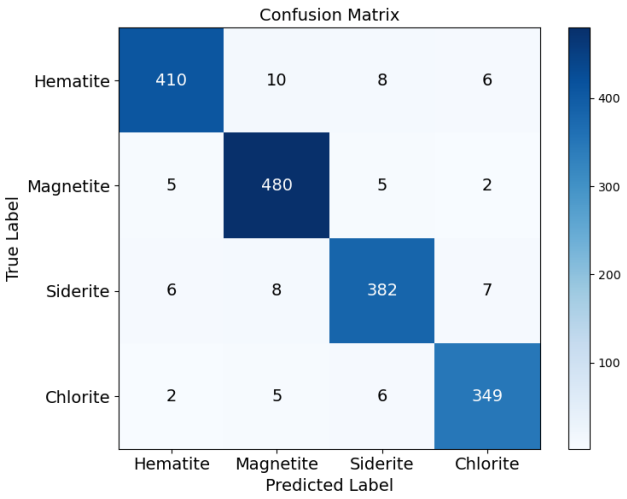
The confusion matrix of the three models in this paper is shown in Figure 10. The following three images are Shufflenetv2 network, ECA-Shufflenetv2-SK, and SEH-Shufflenetv2s network, respectively. In this matrix, the rows represent the predicted classes by the model, the columns represent the actual classes, and the values on the diagonal indicate correct predictions by the model, i.e., the predicted classes are consistent with the actual classes. The larger the number of predicted samples on the diagonal of the confusion matrix, the better the model's performance.



(a) Shufflenetv2 confusion matrix



(b) ECA-Shufflenetv2-SK confusion matrix



(c) SEH-Shufflenetv2s confusion matrix

Fig. 10. Confusion matrix of the three models.

Through observation, it can be found that the overall classification accuracies of the original model, the model with hybrid attention mechanism, and the improved model are 91.4%, 93.27%, and 94.2%. The improved SEH-ShuffleNetV2 correctly classified 78 more images compared to the original network The improved model performs best in detection hematite, following closely behind are siderite, chlorite, and magnetite. The model demonstrates superior performance in the iron ore image classification task, with high overall accuracy and good recognition ability for each iron ore type, but there is some misclassification phenomenon, which may be related to the feature similarity of iron ore images.

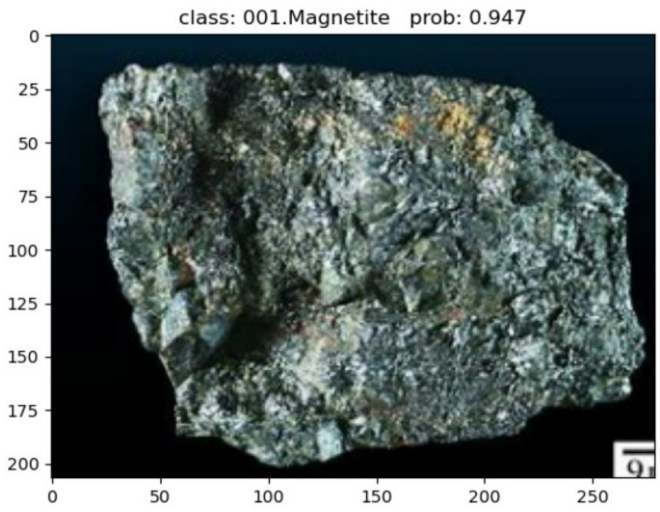
A comparison of the evaluation metrics among the improved network and other networks is presented in Table III.

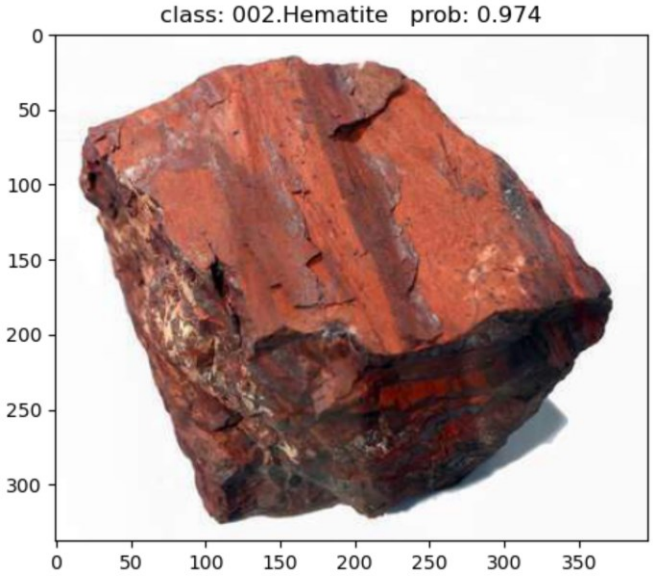
TABLE III
EVALUATION METRICS FOR DIFFERENT NETWORKS

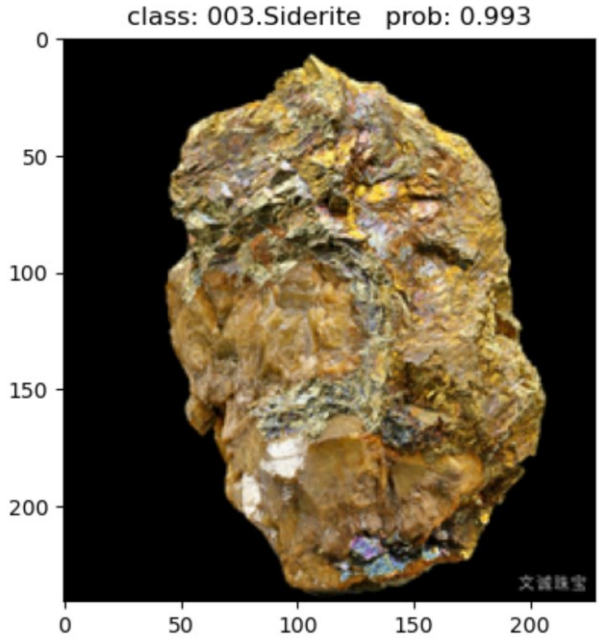
Method	Recall(%)	Precision(%)	Accuracy (%)
VGG16	88.82	90.12	89.16
GLCM+SVM	91.19	92.38	92.15
Moiblenetv2	91.83	91.81	91.92
Shufflenetv2	92.04	91.95	92.10
ECA-Shufflenetv2-SK	93.22	93.10	93.21
SEH-Shufflenetv2s	94.11	94.04	94.20

The table provided illustrates that the Shufflenetv2 network exhibits superior performance in recall, precision, and accuracy compared to alternative networks, while maintaining a lightweight architecture with minimal parameters. Specifically, the ECA-Shufflenetv2-SK network demonstrates a 1.16% higher recall, 1.15% higher precision, and 1.1% higher accuracy than the baseline Shufflenetv2 network. Furthermore, the improved Shufflenetv2 network shows enhancements of 2.07% in recall, 2.09% in precision, and 2.1% in accuracy over the original network. These notable advancements in performance metrics underscore the efficacy and validity of the enhanced model, surpassing the original model for tasks related to iron ore image classification.

The subsequent section presents the results of the iron ore image classification, as illustrated in Figure 11.







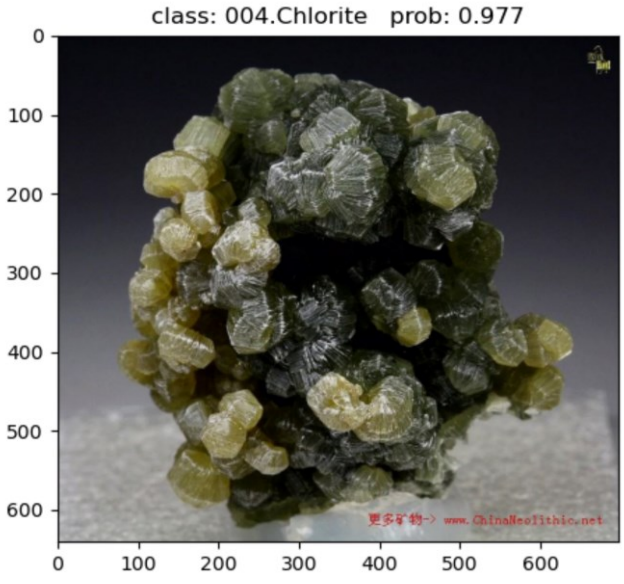


Fig. 11. Some iron ore image classification results.

The test images demonstrated a mean recognition accuracy of above 94%, with mean recognition speeds ranging from 0.1 to 0.5 seconds. These results indicate that the enhanced model in this study exhibits superior performance in terms of classification accuracy and speed when compared to other networks.

V. Conclusion

In the realm of iron ore sorting, conventional methods and manual detection face obstacles stemming from the similarity of ore characteristics, occlusion, and complex environmental factors. This study introduces a streamlined detection model for classifying iron ore images. The model integrates an SK module at each phase to determine an appropriate convolution kernel size for the current input, thereby improving the extraction of feature information across different scales. To enhance the prominence of crucial feature information while suppressing less significant features, an ECA module is introduced. The ShuffleNetV2 activation function module is adapted to enhance the efficiency of the H-Swish activation function, mitigating neuron "inactivation" and promoting model training stability and performance. Furthermore, the network structure is optimized to reduce both parameter count and recognition time effectively. The efficacy of this approach is validated by achieving a recognition accuracy of 94.20% across four self-compiled iron ore image datasets. The enhanced model showcases versatility in diverse production settings, exhibiting notable enhancements in classification accuracy and inference time reduction. Future research will focus on addressing additional challenges in iron ore image classification within more complex contexts. Comprehensive optimization and iterative enhancements are anticipated to further augment the model's practical utility, ensuring alignment with the varied demands of real-world applications.

References

[1] Jiyang Wang, Yang, C.et al. An Iron Ore Identification Method Based on Improved Bilinear Network. International Annual Conference on Complex Systems and Intelligent Science 2023. 20-22 October,2023, Shenzhen, China, pp. 421-426.

[2] Patel A K, Chatterjee S, Gorai A K. Development of Machine Vision-Based Ore Classification Model Using Support Vector Machine (SVM) Algorithm. Arabian Journal of Geosciences 2017, vol. 10, no.5, 2017, pp. 107-114.

[3] Singh V, Rao S M. Application of Image Processing and Radial Basis Neural Network Techniques for Ore Sorting and Ore Classification. Minerals Engineering 2005, vol. 18, no.15, 2005, pp.1412-1420.

[4] Pu Y, Apel D B, Szmigiel A, et al. Image Recognition of Coal and Coal Gangue Using a Convolutional Neural Network and Transfer Learning. Energies 2019, vol. 12,no.9, 2019, pp. 1735.

[5] Liguan Wang, Sijia Chen, Mingtao Jia. et al. Beneficiation Method of Wolframite Image Recognition Based on Deep Learning. The Chinese Journal of Nonferrous Metals, 2020, vol. 30, no.5, pp. 1192-1201.

[6] Baraboshkin, Evgeny E. et al. "Deep Convolutions for In-Depth Automated Rock Typing." Comput. Geosci. Vol.135, 2019, 104330.

[7] Bai Lin, Yao Yu, Li Shuangtao, Xu Dongjing, Wei Xin. Mineral Compositionanalysis of Rock Image Based on Deep Learning Feature Extraction. China Mining Magazine, 2018, 27(7): 178-182. DOI: 10.12075/j.issn.1004-4051.2018.07.038

[8] Xiao, D., Le, B.T., & Ha, T.T. Iron Ore Identification Method Using Reflectance Spectrometer and A Deep Neural Network Framework. Spectrochimica Acta. Part A, Molecular and Biomolecular Spectroscopy, 2021, pp. 119-168.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, 27-30 June, 2016, pp. 770–778.

[10] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger, "Densely Connected Convolutional Networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, 21-26 July, 2017, pp. 4700–4708.

[11] Andrew G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv:1704.04861, 2017.

[12] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018, 18-22 June, 2018, pp. 4510–4520.

[13] Andrew Howard et al., "Searching for MobileNetV3," Proceedings of the IEEE/CVF International Conference on Computer Vision 2019, 27 October - 2 November, 2019, pp. 1314–1324.

[14] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018, 18-22 June, 2018, pp.6848–6856.

[15] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun, "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design," Proceedings of the European Conference on Computer Vision 2018, 8-14 September, 2018, pp. 116–131.

[16] Kai Han et al., "GhostNet: More Features from Cheap Operations," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020, 14-19 June, 2020, pp. 1580–1589.

[17] Li X, Wang W, Hu X, et al. "Selective kernel networks", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019, 16–20 June, pp. 510-519.

[18] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020, 14–19 June, pp. 11531-11539.