Article

Rock Image Classification Based on EfficientNet and Triplet Attention Mechanism

Zhihao Huang, Lumei Su *, Jiajun Wu and Yuhan Chen

School of Electrical Engineering and Automation, Xiamen University of Technology, Xiamen 361024, China
* Correspondence: sulumei@163.com

Featured Application: The work presents an image classification algorithm for rock-type recognition, which can provide reliable guidance for geological surveys.

Abstract: Rock image classification is a fundamental and crucial task in the creation of geological surveys. Traditional rock image classification methods mainly rely on manual operation, resulting in high costs and unstable accuracy. While existing methods based on deep learning models have overcome the limitations of traditional methods and achieved intelligent image classification, they still suffer from low accuracy due to suboptimal network structures. In this study, a rock image classification model based on EfficientNet and a triplet attention mechanism is proposed to achieve accurate end-to-end classification. The model was built on EfficientNet, which boasts an efficient network structure thanks to NAS technology and a compound model scaling method, thus achieving high accuracy for rock image classification. Additionally, the triplet attention mechanism was introduced to address the shortcoming of EfficientNet in feature expression and enable the model to fully capture the channel and spatial attention information of rock images, further improving accuracy. During network training, transfer learning was employed by loading pre-trained model parameters into the classification model, which accelerated convergence and reduced training time. The results show that the classification model with transfer learning achieved 92.6% accuracy in the training set and 93.2% Top-1 accuracy in the test set, outperforming other mainstream models and demonstrating strong robustness and generalization ability.

Keywords: rock image; EfficientNet; image classification; transfer learning

1. Introduction

Rock classification is an essential and critical task in various fields, such as geology, resource exploration, geotechnical investigation, rock mechanics, mineral resource prospecting [1], and constructional engineering [2]. It plays a vital role in supporting mineral and petroleum resource exploration, and in guiding design scheme optimization, safety assessment, and risk assessment in geotechnical engineering. Traditional methods of rock classification can be broadly categorized into physical tests and numerical statistical analysis. Physical tests analyze rock samples using techniques such as X-ray powder diffraction, scanning electron microscopy, and infrared spectroscopy [3], while numerical statistical analysis employs mathematical methods like the nearest-neighbor algorithm [4] and principal component analysis [5] to extract rock classification features. These traditional methods heavily rely on the expertise of professionals and specific equipment to extract useful information from rocks [6]. Therefore, the accuracy of rock classification can be greatly affected by poor experimental conditions or low-quality personnel, resulting in significant fluctuations. In addition, traditional methods are cumbersome and time-consuming, and they are incompatible with the trend of the widespread use of remote exploration devices such as drones in geological surveys [7] since they cannot meet the needs of geological survey personnel to classify rocks directly from images collected by the equipment.

With the advancement of computer technology, computer vision methods based on machine learning and deep learning have begun to show excellent performance in image classification and have been applied in the field of rock classification. Researchers initially applied traditional machine learning methods, which train classifiers using artificially extracted rock image features for automatic classification. For example, Marmo et al. [8] extracted image feature values from more than 1000 carbonate slices using image processing and input them into a multi-layer perceptron neural network model to achieve the intelligent classification of carbonate rocks, with an accuracy of 93.3%. Similarly, Singh et al. [9] extracted 27 features from thin-section images of basalt rock and were able to recognize and classify 140 thin-section images of rock samples with an accuracy of 92.22%. Although these traditional machine learning methods can produce satisfactory results, they cannot automatically extract features directly from unprocessed images due to the limited capacity of the shallow models. [10] Therefore, before training the classification model, manual pre-processing of images is necessary to extract features such as color, shape, and texture. This significantly reduces the level of automation in type classification.

In recent years, deep learning has been applied in rock-type classification to overcome the limitations of traditional machine learning methods in image feature extraction [11]. Researchers have utilized various deep learning models to automatically extract image features, which leads to more intelligent image classification. Xu et al. [12] designed a U-net convolutional neural network for ore mineral recognition under the microscope with a test set success rate of above 90%. Zhang et al. [13] developed a rock image recognition model based on the Inception-V3 deep convolutional neural network for various rock types, achieving a classification accuracy of over 85%. Cheng et al. [14] used the ResNet50 and ResNet101 models for automatic feature extraction and classification of rock slice images, with an accuracy of 90.24% and 91.63% in the test set, respectively. Chen et al. [15] proposed a deep residual neural network model with transfer learning to establish an automatic rock classification model with over 90% accuracy. Koeshidayatullah et al. [16] proposed a transformer-based model for automatic core-face classification, eliminating the need for pre-processing and manual feature extraction. The deep learning method offers the advantage of automatically extracting image features, which eliminates the influence of subjective factors on experimental results and greatly reduces the workload of rock classification. In addition, deep learning models can extract more abstract and complex image features to classify a wider range of rock types. [17] However, the existing deep learning models have typically been built by manually designing a network module and stacking it [18–20], resulting in an irrational network structure and redundant parameters. Moreover, for objects like rock images with strong interference and cluttered information, the existing models without attention mechanisms suffer from dispersed attention during image processing, which makes it challenging to effectively capture useful features. These factors ultimately contribute to the limited accuracy of current models in rock image classification.

Current methods for rock image classification each have their limitations. Traditional methods can achieve acceptable accuracy under specific conditions but are heavily dependent on manual effort and lack stability in their accuracy. Although machine learning methods have made initial strides towards intelligent rock image classification, they still require manual feature extraction, limiting the degree of automation in image recognition. Existing deep learning methods have realized end-to-end automatic image recognition but still suffer from issues related to redundant network parameters and scattered attention, resulting in limited accuracy in rock image classification. To overcome the limitations of current methods and achieve accurate end-to-end classification of rock images, a rock image classification model based on EfficientNet and a triplet attention mechanism is proposed. The contributions of this study can be summarized as follows:

- Deep learning methods were introduced to eliminate the dependence of traditional methods and machine learning methods on human intervention, so as to realize end-to-end automatic identification of rock images without requiring additional manual operations.

- A rock image classification model was constructed based on EfficientNet, which overcomes the issue of parameter redundancy and scattered attention in previous deep learning models, as well as achieving an efficient and attention-focused network structure through Neural Architecture Search, resulting in higher model accuracy compared to its predecessors.
- In view of the problem that EfficientNet neglects spatial attention information of rock images, the triplet attention mechanism [21] was introduced to improve EfficientNet and enhance its ability to extract effective rock features, further improving the accuracy of the rock classification model.
- The transfer learning method was utilized in the training process to accelerate the model convergence and significantly enhance its training performance, so as to obtain a classification model with higher accuracy using fewer rock images and less training time.
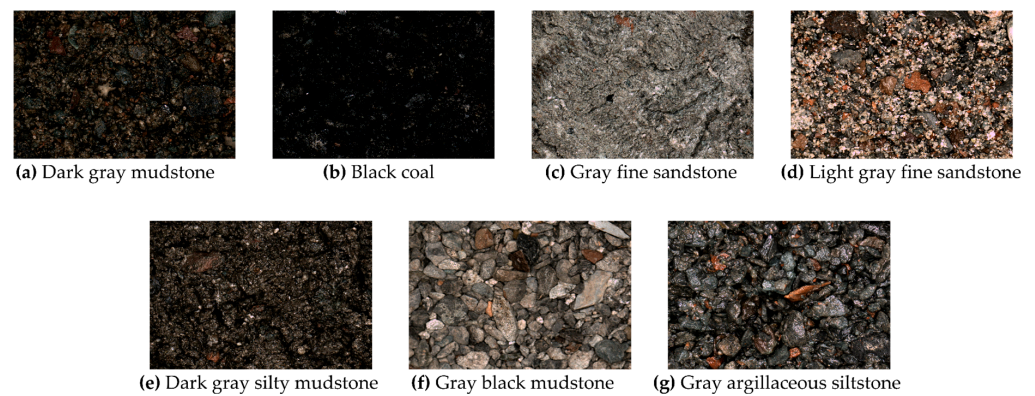
2. Materials

The rock image dataset was provided by Guangdong TipDM Intelligent Technology Co., Ltd. (Wuhan, China) and includes 315 rock images. These rock images were high-resolution photographs of rock fragments taken by geological staff using an industrial camera under white light conditions at the well logging site. The images were captured using a fixed-height lens. In addition, most of the images had an initial size of 4096 × 3000 pixels, while a small number of images with backgrounds had a size of 2448 × 2048 pixels. The rock fragment image collection process generally involves the following steps:

(1) Collect rock fragments: Geological personnel collect rock fragments at the wellhead or drilling rig, using manual or mechanical tools for collection.
(2) Prepare samples: After collecting rock fragments, preliminary processing such as cleaning and sieving is necessary to remove impurities and unwanted parts and obtain a sufficient number of samples of the same rock type.
(3) Capture images: For each rock fragment sample, capture its image using an industrial camera or other equipment. During image capture, it is important to keep the sample in the same position and angle to ensure the comparability and repeatability of the images.

Rock fragments images are often affected by the diverse physical and chemical properties of rocks, and typically exhibit small sizes, complex and diverse shapes, and variations in texture and color. In addition, noise and background interference in rock fragment images may adversely affect the quality of the images. All of these factors present challenges for the identification of rock types in this type of image.

The dataset contains rock images of seven types in total, including 21 black coal images, 30 gray black mudstone images, 46 gray argillaceous siltstone images, 18 gray fine sandstone images, 85 light gray fine sandstone images, 40 dark gray silty mudstone images, and 75 dark gray mudstone images. Figure 1 shows seven rock images of different types.

Different rocks possess distinct characteristics such as shape, color, and texture. Table 1 illustrates the key features of the various rock images included in the dataset. The serial number of each row in the table corresponds to Figure 1. The seven types of rocks in the table can be classified into four types based on their particle size: sandstone, siltstone, mudstone, and coal. Sandstone typically features a gray or light gray color, a non-blocky shape, and a rough surface [22]. Siltstone is mostly gray in color, with a flaky and blocky shape and a smooth surface [23]. Mudstone can be gray, dark gray, or gray black in color and tends to have a flaky and blocky shape [24]. Coal is the most identifiable rock type in the dataset, being black in color.

(a) Dark gray mudstone **(b)** Black coal **(c)** Gray fine sandstone **(d)** Light gray fine sandstone

(e) Dark gray silty mudstone **(f)** Gray black mudstone **(g)** Gray argillaceous siltstone

Figure 1. Seven rock images of different types in the rock image datasets.

Table 1. Characteristics of various rocks in rock image dataset.

Serial Number of Image	Type	Particle Size (mm)	Color	Shape Characteristics
a	Dark gray mudstone	<0.005	Dark gray	Lamellar, Micrite structure
b	Black coal	<6	Black	Granular, Asymmetrical
c	Gray fine sandstone	0.05–2	Gray	Fine grain structure, Rough surface, Uneven
d	Light gray fine sandstone	0.05–2	Light gray	Fine sand structure, Thin-layered structure
e	Dark gray silty mudstone	<0.005	Dark gray	Silty argillaceous structure, Bedding structure
f	Gray black mudstone	<0.005	Gray black	Cryptocrystalline structure, Massive structure
g	Gray argillaceous siltstone	0.005–0.05	Gray	Silty structure, Massive structure

3. Method

In this study, we propose a rock-type classification method based on EfficientNet and a triplet attention mechanism. The method focuses on the establishment of a classification model and integrates various methods, such as transfer learning and data augmentation, to achieve accurate automatic classification of rock images. Figure 2 is the flow diagram of the proposed method. The detail of the method is shown as a pseudo-code in Algorithm 1.

3.1. EfficientNet Neural Network

Model scaling has been widely used to improve the accuracy of convolutional neural networks. In previous work, the most common way is to only change the network depth, width, or input image resolution of the baseline neural networks in a single dimension, as shown in Figure 3b–d. For example, Huang et al. [25] greatly improved the accuracy of GPipe on the ImageNet dataset by scaling up the depth of the baseline network. Though it is possible to scale up the model in multiple dimensions at the same time, multidimensional scaling requires tedious manual tuning. To realize simple yet efficient model scaling, Tan et al. [26] proposed that EfficientNet, which is obtained by a Neural Architecture Search (NAS) [27] technology and a compound scaling method, is one of the best classification performance networks on the ImageNet dataset. They first searched the structure of the baseline network using NAS technology and then scaled up the baseline network in multiple dimensions by the compound scaling method. This scaling method allows for uniform changes in the network depth, width, and input image resolution, as shown in Figure 3e, resulting in higher classification accuracy without the need for additional fine-tuning.


Algorithm 1: A Rock-type Classification Method Based on EfficientNet and Triplet Attention Mechanism


Input: 315 rock images containing seven types of rocks

1:   Perform data pre-processing.

2:    Randomly apply the following nine data augmentation operations to each image: Rotation, Salt-and-pepper noise addition, Brightening, Darkening, Enlargement, Vertical flip, Horizontal flip, Gaussian noise addition, and Translation. (The number of rock images is augmented to 6949 after augmentation.)

3:    Divide the augmented images randomly into a training set, a validation set, and a test set with a ratio of 60%, 20%, and 20%. (The number of samples for them is 4169, 1389, and 1389.)

4:   Construct a rock-type classification model based on EfficientNet and a triplet attention mechanism.

5:    Select the EfficientNet-B7 model as the baseline model.

6:    Replace each SE attention module of the EfficientNet-B7 model with the triplet attention module to construct the Triplet-Efficient model.

7:    Build a classification model based on the Triplet-EfficientNet model as the backbone network.

8:    Add 1×1 convolutional layer, pooling layer, fully connected layer, and softmax classifier after Triplet-EfficientNet.

9:    Set the number of types for the softmax classifier to 7.

10: Start model training.

11: Use the transfer learning method: load the parameters of the pre-trained model trained on ImageNet dataset into an untrained model.

12:    Set training hyperparameters: set the learning rate to 0.01, the epoch to 60, and the batch size to 16.

13:    Select the Swish function as the activation function, the cross entropy function as the loss function, and Adaptive Moment Estimation(Adam) as the optimizer.

14:    Uniformly scale the images in the training set to the size of $600 \times 600 \times 3$ and randomly package them into the model to start the training.

15: Train all parameters of the transferred model for 60 epochs.

16: Output the final model after training.

17:   Start testing: input a randomly selected rock image.

Output: The probabilities of this rock image being classified as each type of rock. (The rock type corresponding to the maximum probability is the final identification result.)
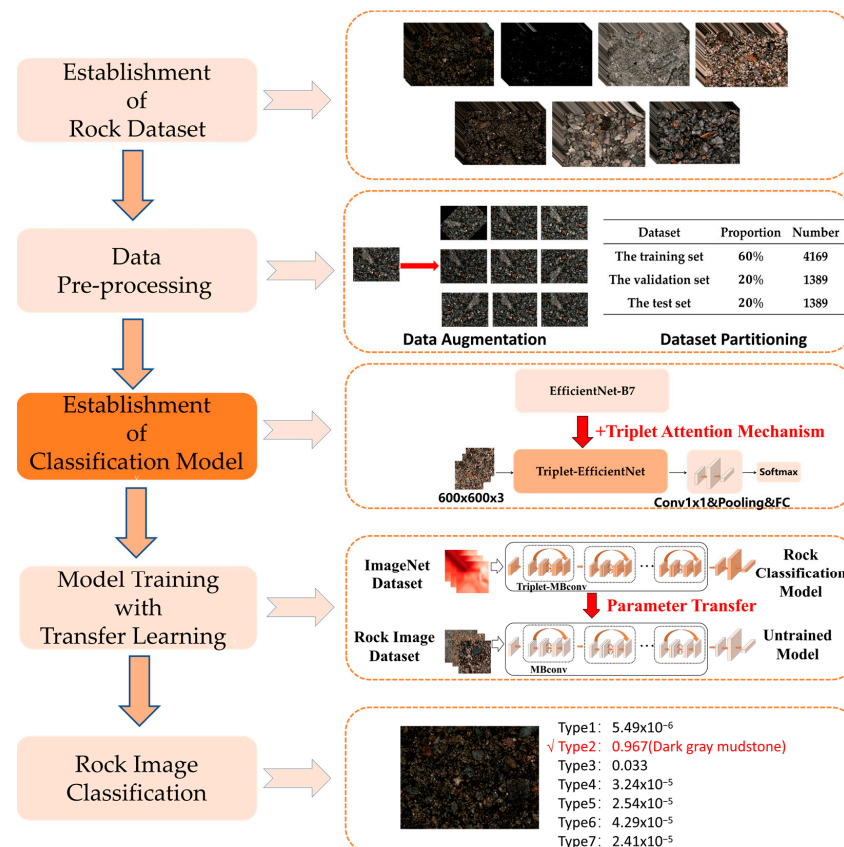
Figure 2. Flowchart of the classification method proposed in this study.
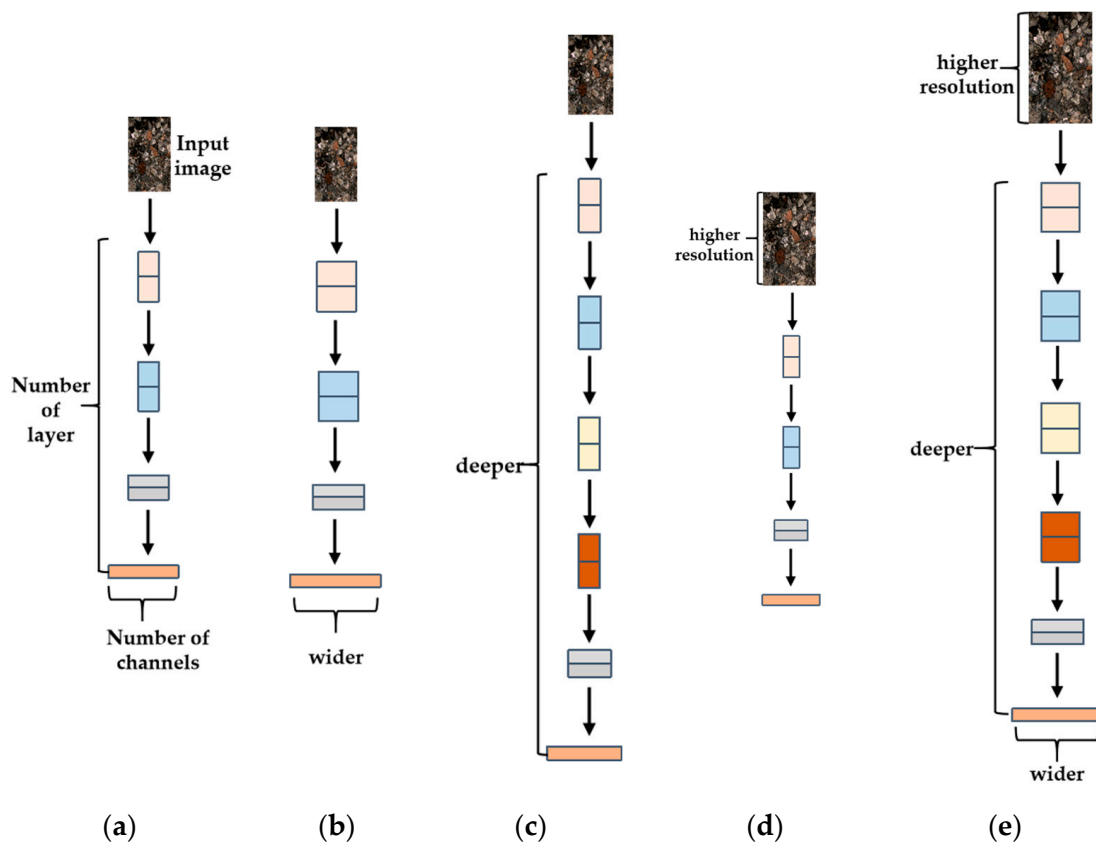
Figure 3. Diagrams of different model scaling methods. (**a**) a baseline network example; (**b**–**d**) model scaling methods that only change one dimension of network width, depth, or resolution. (**e**) a compound scaling method that uniformly scales three dimensions.

The key to the compound scaling method based on EfficientNet is to find a set of compound coefficients of depth, width, and image resolution to maximize the network's performance. This optimization problem is mathematically formulated as in Equation (1).

$$\max_{d,w,r} \text{Accuracy}(\text{N}(d,w,r)) \tag{1}$$

where, d, w, r are the scaling coefficients of network depth, width, and image resolution respectively, $\text{N}(d,w,r)$ is the classification model, and $\max_{d,w,r} \text{Accuracy}$ is the maximum accuracy of the model.

To realize uniform scaling of d, w, r, this method introduces φ, which is a user-specified coefficient that controls model scale, as shown in Equation (2). The implementing steps of this scaling method are as follows: After determining the structure of the baseline network, this method first fixes the control coefficient φ as 1, then uses NAS technology to search coefficients d, w, r that maximize the classification accuracy, resulting in the final baseline model called EfficientNet-B0; Finally, this method specifies different φ from 2 to 7 and obtains corresponding models of different sizes, referred to as EfficientNet-B1, EfficientNet-B2, . . . , EfficientNet-B7, respectively.

$$d = \alpha^{\varphi}, \ w = \beta^{\varphi}, \ r = \gamma^{\varphi}$$
$$\alpha \geqslant 1, \ \beta \geqslant 1, \ \gamma \geqslant 1 \tag{2}$$

where, α, β, γ are constants that can be determined by NAS technology.

3.2. EfficientNet-B7 Model

EfficientNet-B7 is a high-precision model obtained by scaling up EfficientNet-B0, its input image resolution is 600×600, width multiplier factor (w) is 2.0, and depth multiplier

factor (*d*) is 3.1. In this study, EfficientNet-B7 was selected as the benchmark model in order to achieve the best possible classification accuracy in the rock image dataset.

As shown in Figure 4, the EfficientNet-B7 model was built by stacking multiple Mobile Inverted Bottleneck Convolution (MBConv) modules. The structure of the MBConv module is different from traditional residual modules as its input and output feature maps are both wider than the middle. As shown in Figure 5, the MBConv module includes a convolutional layer of kernel size 1×1, depth-separable convolution, Squeeze(SE) attention module, and Dropout layer. The MBConv module applies Batch Normalization (BN) and Swish activation function after the convolutional layers. BN can normalize the data and speed up the model convergence during training, while the Swish activation function can introduce non-linearity to the data and avoid overfitting. In addition, the SE attention module in the MBConv module incorporates the SE [28] attention mechanism, which enhances the model's feature representation capability by capturing channel attention information in the input feature map.
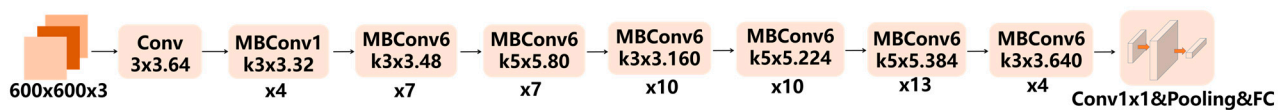

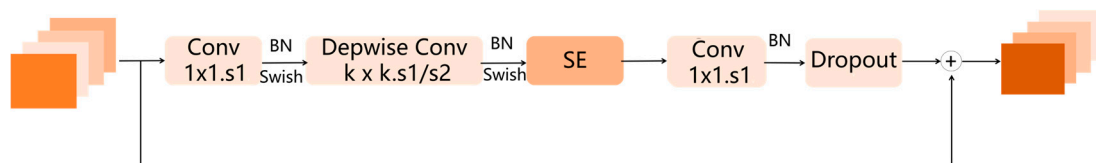**Figure 4.** EfficientNet-B7 network structure.

Figure 5. MBConv module structure.

The SE attention module utilizes the SE attention mechanism to calculate the importance of each channel in the input feature map for the current task, as well as weighting them. This is achieved through the following three operations:

(1) Squeeze: The input feature map of size $C \times H \times W$ is globally average pooled into a feature map of $1 \times 1 \times C$, thus squeezing each two-dimensional feature channel into a single value to represent the global distribution of responses on each channel.

(2) Excitation: A fully connected neural network is used to nonlinearly transform the squeezed map, generating activated weights through ReLU and Sigmoid activation functions.

(3) Scale: The activated weights are used to weight each channel in the input feature map by performing dot multiplication.

As shown in Figure 6, a rock feature map of size $C \times H \times W$ is input into the SE attention module. Through the above three operations, the module assigns weights to each channel based on their impact on rock image classification accuracy, enhancing effective rock feature channels and suppressing weak ones. In this figure, the different channels in the output tensor have different color borders, indicating that different weights have been assigned to each channel after passing through the SE module.
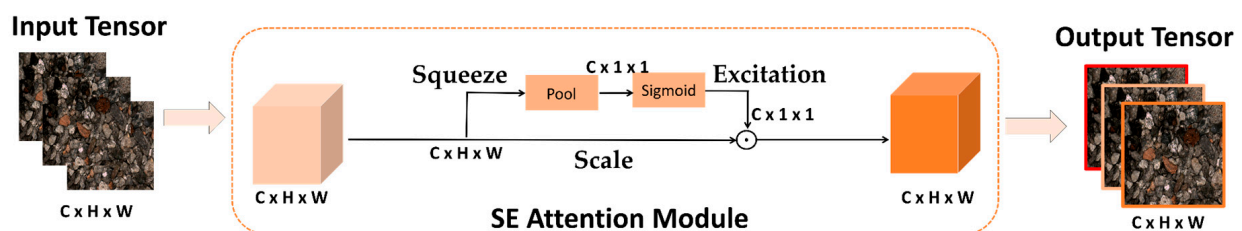

Figure 6. Structure of SE attention module in EfficientNet.

3.3. Triplet Attention Mechanism

In this study, we aim to classify images of various unpretreated rocks taken under natural conditions. Although each rock image in the dataset contains only one rock type, the features of rock morphology, texture, and color may vary slightly at different spatial positions within the same image, resulting in an uneven distribution of rock features in rock images. The SE attention module of EfficientNet only pays attention to the importance of different rock feature channels and is unable to calculate the importance of rock features at different spatial positions, thereby limiting the accuracy of the rock image classification model based on EfficientNet. To address this, we introduce a triplet attention module that can capture both spatial and channel attention information to replace the SE module in EfficientNet, thus proposing an improved EfficientNet model—Triplet-EfficientNet.

As shown in Figure 7, the triplet attention module not only assigns weights to different feature channels of the input rock feature tensor, but also assigns weights to different spatial positions on each channel. The output rock feature tensor in Figure 7 not only has colored borders representing different weights for each feature channel, but also different colors at different positions on each channel. The parts of each feature channel that are closer to blue represent spatial features that have less impact on classification accuracy and will be assigned smaller weights by the module, while the parts that are closer to red represent more effective rock spatial features, and thus receive more attention from the module and are assigned larger weights. Therefore, by incorporating a triple attention module in the classification model, we can effectively address the issue of imbalanced spatial feature distribution in rock images, significantly improve the model's capability to extract effective features, and ultimately enhance its overall accuracy.
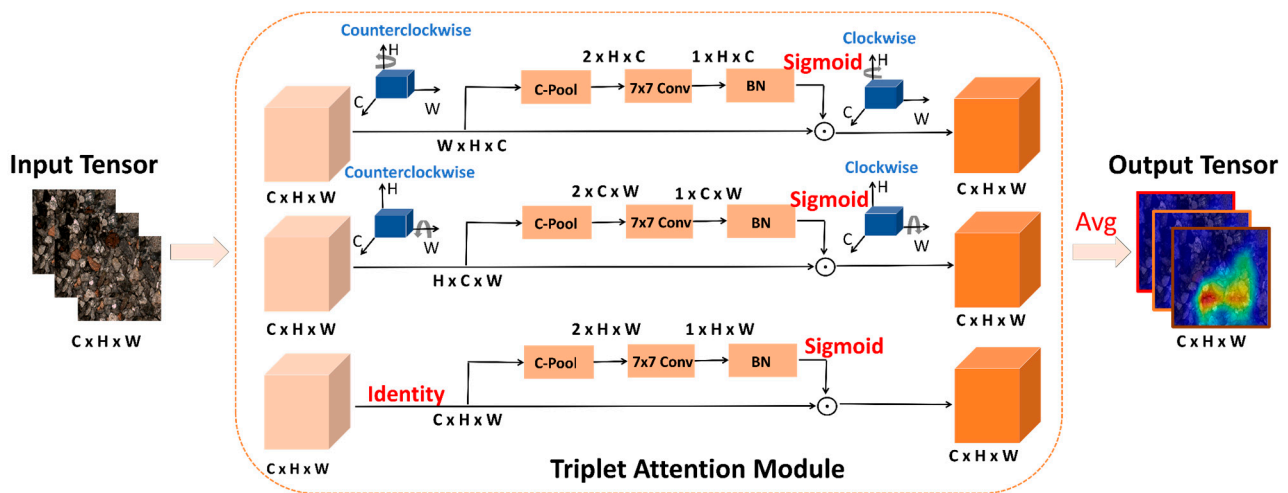

Figure 7. Structure of Triplet attention module.

The triplet attention module consists of three parallel branches, as shown in Figure 7, which takes in an input rock feature tensor and outputs a refined tensor of the same shape. Given an input rock feature tensor $X \in R^{C \times H \times W}$, the module first passes it to each attention branch respectively to capture the cross-dimension interaction information among channel dimension C, spatial dimension H, and W.

In the first branch, the triplet attention module builds the interactions between the spatial height (H) dimension and the channel (C) dimension. To achieve so, the input X is first rotated 90 degrees counterclockwise along the H axis to obtain a rotated tensor $X_H \in R^{W \times H \times C}$. Then X_H would be squeezed into a two-dimensional tensor of size $1 \times H \times C$ through a C-Pool layer, a convolutional layer of kernel size 7×7, and a batch normalization layer. Next, the tensor of size $1 \times H \times C$ passes through the Sigmoid activation function to generate the resultant attention weights. Following this, the generated attention weights are applied to X_H, namely the feature parameters with the same width would be weighted, generating weighted feature maps. Finally, the weighted feature maps
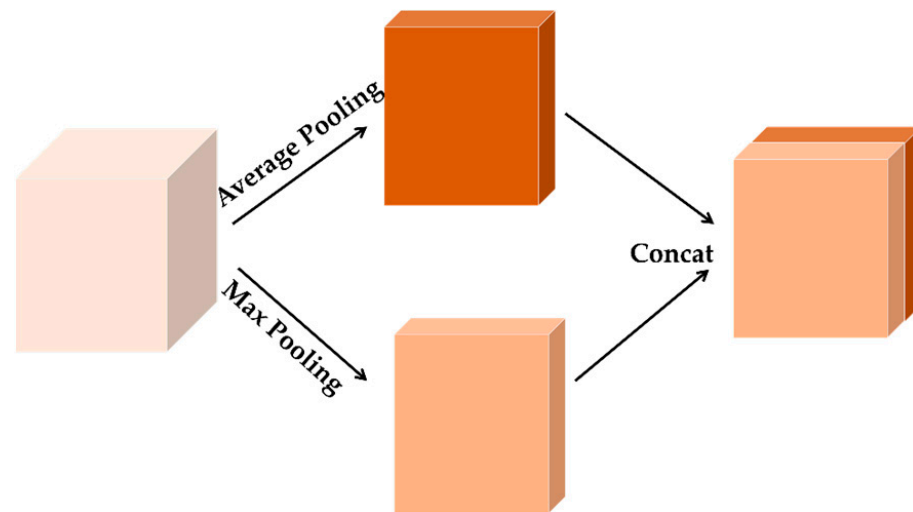
are rotated 90 degrees clockwise along axis H to output X_{H+} of the same shape as the input X. The calculation process of this branch can be represented by the following equation:

$$X_{H+} = R^{H+}(X_H \cdot \sigma(\text{ConvBN}(\text{C-pool}(X_H))))\tag{3}$$

where, R^{H+} represents clockwise rotation of 90 degrees along the H axis, σ represents Sigmoid activation function, ConvBN represents combination operation of convolution and batch normalization, and C-pool represents compound pooling.

The calculation process of the compound pooling layer is shown in the following Figure 8. In this layer, the input tensor is processed through both max-pooling and average-pooling operations along the channel dimension, and the resulting features are combined through concatenation.

Figure 8. Structure of compound pooling layer.

Similarly, in the second branch, the input X is first rotated 90 degrees counterclockwise along the W axis to obtain a rotated tensor $X_W \in R^{H \times C \times W}$. Then, X_W is input into the attention branch to generate the weighted feature maps. Finally, the weighted maps are rotated 90 degrees clockwise along the W axis. The calculation process of this branch can be represented by the following equation:

$$X_{W+} = R^{W+}(X_W \cdot \sigma(\text{ConvBN}(\text{C-pool}(X_W))))\tag{4}$$

where, R^{W+} represents a clockwise rotation of 90 degrees along the W axis.

In the last branch, rotation is not carried out on the input tensor X, but directly weighting the feature parameters in the same channel and generating weighted feature maps is performed. The calculation process is shown in Equation (5).

$$X_C = X \cdot \sigma(\text{ConvBN}(\text{C-pool}(X)))\tag{5}$$

After the calculation of each branch, as shown in Equation (6), the triplet attention module would aggregate the refined tensors generated by each branch using simple averaging, so as to realize the fusion of channel attention and spatial attention information.

$$y = \frac{1}{3}(X_{H+} + X_{W+} + X_C)\tag{6}$$

3.4. Classification Model Based on Triplet-EfficientNet

In this study, we propose Triplet-EfficientNet, a modified version of EfficientNet-B7 that incorporates the triplet attention mechanism, and establish the rock classification model based on Triplet-EfficientNet. The classification model is shown in Figure 9, and its backbone is Triplet-EfficientNet, which is similar to EfficientNet-B7. The difference is that Triplet-EfficientNet consists of Triplet-MBConv, an improved MBConv module with

stronger characterization capability. The Triplet-MBConv module is shown in Figure 10, it was improved by replacing the SE module in the original MBConv module with the triplet attention module. The classification model based on Triplet-EfficientNet can not only capture the long-term dependence between network channels but also retain the precise location information, so as to further improve the accuracy of rock image classification.
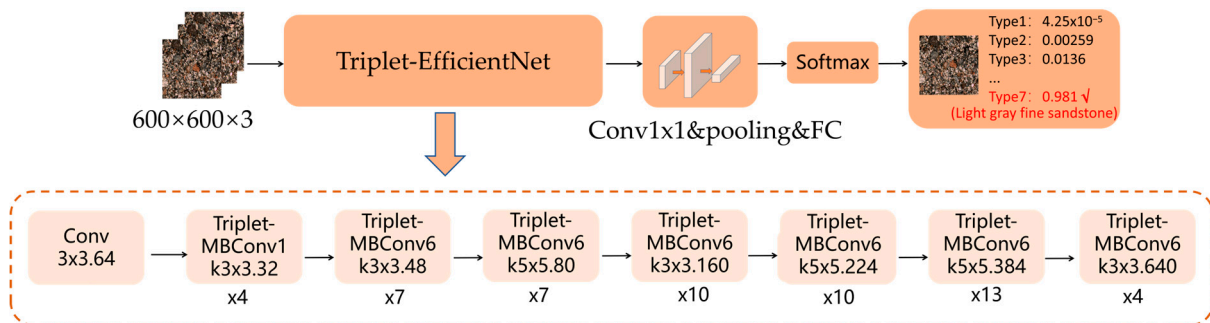

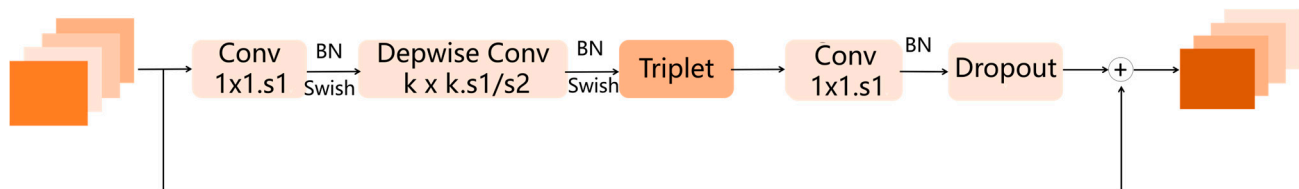**Figure 9.** Structure of classification model based on Triplet-EfficientNet.

Figure 10. Structure of Triplet-MBConv module.

The process of rock-type classification using the classification model based on Triplet-EfficientNet is as follows: First, the input rock images are pre-processed using methods of data enhancement, and converted into images of size $600 \times 600 \times 3$ as the input of the classification model. Then the first convolution layer of kernel size 3×3 will downsample the input image to achieve space squeeze and channel expansion. Subsequently, the seven stages of Triplet-MBConv layers containing the triplet attention modules will further extract high-dimensional features from the rock images. The high-dimensional feature maps are then squeezed into a two-dimensional tensor through a convolution layer of kernel size 1×1, a pooling layer, and a fully connected layer. Finally, through the Softmax function, the model will output the prediction probabilities, which represent the probability value of the input rock image belonging to each type. The type corresponding to the maximum probability is the final classification result.

3.5. Transfer Learning

Although the classification algorithm based on deep learning models has overcome the disadvantages of traditional methods based on machine learning in feature extraction, it needs sufficient images as training input to achieve high accuracy In addition, it will take a lot of time to build a dataset which contains a large variety of rock images. Even if the dataset is completed, it will also cost a lot of time and computational resources to train a high-precision deep learning model from scratch. To solve the above problems, the transfer learning method was introduced for model training in this study.

This method applies the parameters and weights of the pre-trained model trained on an existing large-scale annotated image dataset to a specific model oriented to a similar problem, and then re-trains and fine-tunes the specific model to obtain the final model. Through this method, we can obtain a classification model with higher accuracy using fewer rock images and less training time.

In this study, we applied the transfer learning method to model training by training all model parameters after loading the pre-trained weights, so as to make the final model have stronger feature-extraction capability. The implementing steps are shown in Figure 11,

we first pre-train the rock classification model on ImageNet [29], which includes 27 types and more than 20,000 fine-classified images, and the model can learn common image information on ImageNet. Then we transfer the shared parameters and weights on the pre-trained model into an untrained model. Finally, by re-training and fine-tuning all weights and parameters on the transferred model using the rock image dataset, we can obtain the final model. This method allows us to leverage the knowledge learned from the pre-trained model to significantly accelerate model convergence and improve overall model accuracy with fewer data and computational resources.
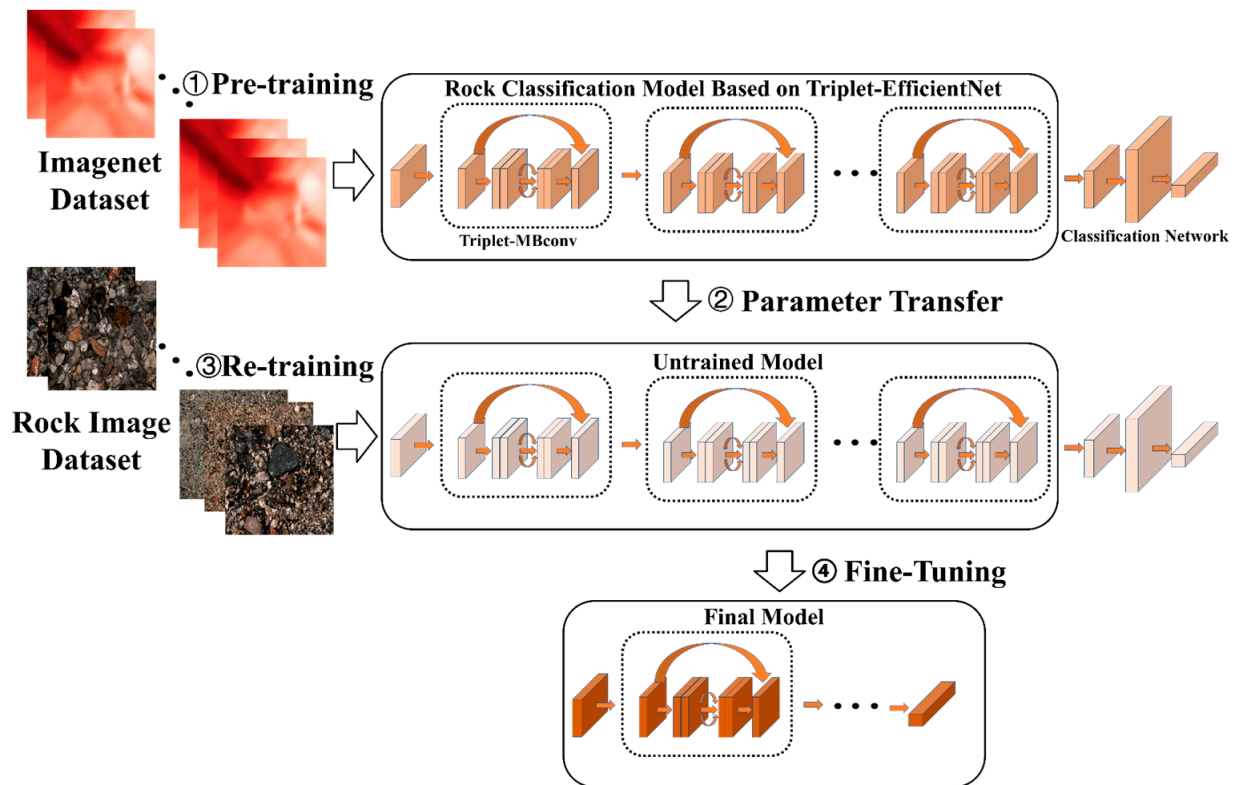

**Figure 11.** Schematic diagram of the transfer learning method.

4. Experiments and Results

4.1. Data Pre-Processing

As the number of samples in the rock image dataset is too small and the quantity distribution among various types is not balanced, the model trained with such a dataset has low accuracy and a risk of overfitting [30]. In order to solve the above problems and make the trained model have higher classification accuracy and stronger generalization ability, it is necessary to augment the number of samples in the dataset. In addition, in order to meet the needs of model training, verification, and testing, the augmented dataset needs to be partitioned into the training set, the verification set, and the test set in a certain proportion.

4.1.1. Data Augmentation

In this study, we augment the number of samples in the dataset with nine data enhancement operations, including Rotation, Salt-and-pepper noise addition, Brightening, Darkening, Enlargement, Vertical flip, Horizontal flip, Gaussian noise addition, and Translation. The schematic of the data augmentation is shown in Figure 12a. The final result is shown in Figure 12b. After data enhancement, the number of samples in the dataset is augmented to 6949, and the number of different samples is basically even.
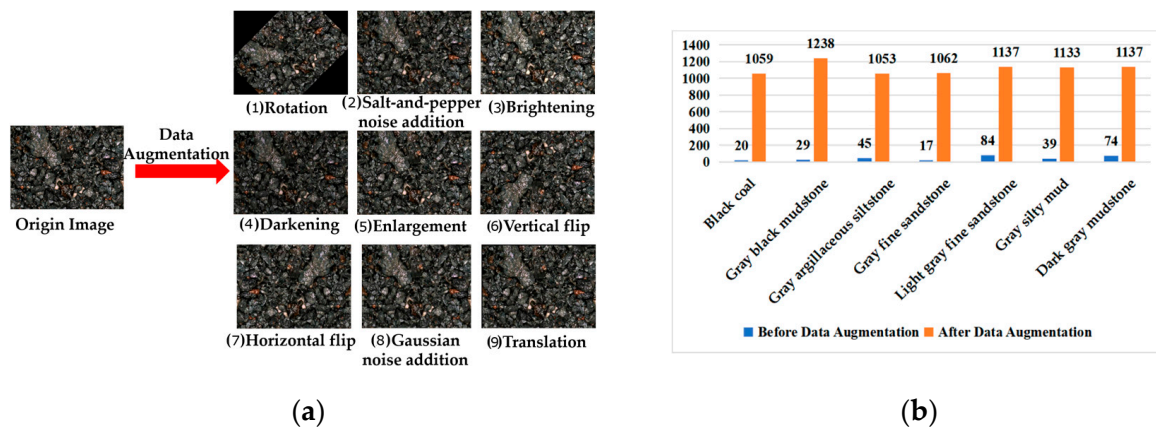
Figure 12. Data augmentation. (**a**) Data augmentation methods; (**b**) Data augmentation result.

Rotation is a transformation that rotates an image by a specified angle around its center; salt-and-pepper noise is a type of image noise that is characterized by random white and black pixels. In the process of the noise addition for each pixel in the image, if the pixel is not randomly selected, its value in the new image will be the same as its value in the original image. If the pixel is randomly selected, its value in the new image will be 0 or 255, and its color will be black or white. Brightening is a transformation that increases the brightness of an image. Darkening is a transformation that decreases the brightness of an image. Enlargement is a transformation that increases the size of an image. Vertical flip is a transformation that flips an image vertically. Horizontal flip is a transformation that flips an image horizontally. Translation is a transformation that shifts an image by a specified amount on the x-axis and y-axis. In addition, Gaussian noise is a type of image noise that is characterized by random values that follow a Gaussian distribution. The addition of Gaussian noise increases the complexity of the image and makes it more similar to the noise present in actual rock images. This allows for better simulation of real-world scenarios and improves the robustness of the model to noise. In rock classification, this method allows the model to adapt better to different environments, thereby improving recognition accuracy. The formula for adding Gaussian noise to an image is:

$$I'(x,y) = I(x,y) + N(\mu, \sigma^2) \tag{7}$$

where, I is the original image, I' is the brightened image, N is the function of gaussian noise, μ is the mean, and σ^2 is the variance. In this study, μ is set as 0 and σ^2 is set as 0.01.

4.1.2. Dataset Partitioning

In deep learning, datasets are typically partitioned into the training set, the validation set, and the test set to meet the needs of model training, validation, and testing. The training set is a collection of data used to train the model. By learning from the training set, the model can learn the features and patterns of the data, thus improving the accuracy of predictions. The training set includes input data and corresponding labels, and the model continuously updates its weights to improve prediction accuracy. The training set is the foundation of deep learning model training and has a crucial impact on the model's performance, therefore the training set occupies the largest amount of data. The validation set is a dataset used to evaluate the model's performance. The validation set can help us detect the degree of overfitting of the model and adjust the model's hyperparameters in a timely manner. At the same time, it can also help us choose the best model parameters, such as epoch, learning rate, and batch size. The test set is a dataset used to evaluate the final performance of the model. After the training and validation process was completed, we usually use the test set to evaluate the final performance of the model. In this study, the rock dataset was also randomly divided into the training set, the validation set, and the

test set with a ratio of 60%, 20%, and 20%, respectively. Therefore, the number of samples for them is 4169, 1389, and 1389, respectively.

4.2. Experiment Details

The model was trained and tested on a high-performance workstation with Windows 10 operating system, which was configured with a 2.10 GHz Intel Xeon Silver 4110 CPU (16 GB memory) and NVIDIA GeForce RTX 2080 Ti GPU. The software environment is as follows: Windows 10 operating system based on 64-bit, Pytorch deep learning framework, CUDA11.0, OpenCV2 library and VS Code integrated development environment.

During network training, the learning rate was set to 0.01, the epoch was set to 60, and the batch size was set to 16. The Swish function was selected as the activation function, the cross entropy function was selected as the loss function, and Adaptive Moment Estimation(Adam) was selected as the optimizer.

Before the model training, the parameters of the pre-trained model were loaded to the classification model by the transfer learning method. The images in the training set were then uniformly scaled to the size of $600 \times 600 \times 3$ and randomly packaged into the model to start the training. In the process of training, we randomly selected 4169 images for training, and each image would be used several times. The training set and verification set were evaluated once per iteration, and the process of training accuracy, verification accuracy, and cross-entropy loss changes in each generation were saved as a log file, and then uploaded to Tensorboard for review.

4.3. Evaluation Metrics

Accuracy and loss value are the two most common evaluation indexes for image classification. The accuracy represents the proportion of correctly classified samples in all samples, which is the evaluation index that most directly reflects the performance of the classification model. It is formulated by Equation (8):

$$\text{Accuracy} = \frac{t}{T} \tag{8}$$

where, t is the number of samples correctly classified, and T is the total number of samples.

In this study, the cross-entropy loss function was used to quantitatively evaluate the difference between the predicted value and the real value. Through the calculation of the loss function, the parameters of our model were updated. It is formulated by Equation (9):

$$\text{Loss} = \frac{1}{N} \sum_i \text{L}_i = -\frac{1}{N} \sum_i \sum_{c=1}^{M} y_{ic} \log(p_{ic}) \tag{9}$$

where M is the total number of types. y_{ic} is the indicator variable, and if the type is as same as the type of the sample i, y_{ic} is 1, otherwise it is 0. p_{ic} is the prediction probability of the sample i belonging to the type c.

4.4. Results Analysis

4.4.1. The Effectiveness of Data Augmentation

In this study, we sought to enhance the performance of our classification model by expanding the original dataset through some methods of data augmentation. To evaluate the effectiveness of this pre-processing step, we conducted a series of ablation experiments. The classification model based on Triple-EfficientNet was used to conduct experiments on the training set and the test set before and after expansion. As can be seen from the results in Table 2, the application of data augmentation resulted in a significant increase in classification accuracy for both the training set and the test set. Specifically, the accuracy of the training set increased by 31.4%, while the accuracy of the test set increased by 22.4%. These results demonstrate that pre-processing a small sample dataset through data augmentation can improve the network's ability to extract more comprehensive rock features, thereby enhancing the model's overall generalization capability.

Table 2. The results of the ablation experiments on the data augmentation methods.

Method	Number of Images in the Rock Dataset	Final Accuracy in the Training Set (Epoch = 60)	Top-1 Accuracy in the Test Set
Triplet-EfficientNet + Original dataset	315	61.2%	70.8%
Triplet-EfficientNet + Augmented dataset	6949	92.6%	93.2%

4.4.2. The Effectiveness of Transfer Learning

In order to evaluate the effect of the training strategy with transfer learning on the convergence of network training, we respectively trained a classification model based on Triplet-EfficientNet with the transfer learning method and another identical model but without the transfer learning method. Except for the different training strategies, the settings of the other hyperparameters were the same for these two models.

The changes in accuracy and loss value in the training process are shown in Figure 13a,b, respectively. As can be seen from Figure 13a, the model with the transfer learning method can obtain an initial accuracy of up to 75.2%, and its final accuracy reaches 92.6% after 60 epochs. However, the initial accuracy of the model without the transfer learning method is only one-quarter of the former, and the final accuracy is only 44.2%. As can be seen from Figure 13b, the initial loss value of the model with the transfer learning method is only 1.225, then it rapidly converges in 60 epochs, and the final loss value converges to 0.214. However, the overall loss value of the model without the transfer learning method is higher than one, and its convergence speed is much lower than that of the former.

(**a**)

(**b**)
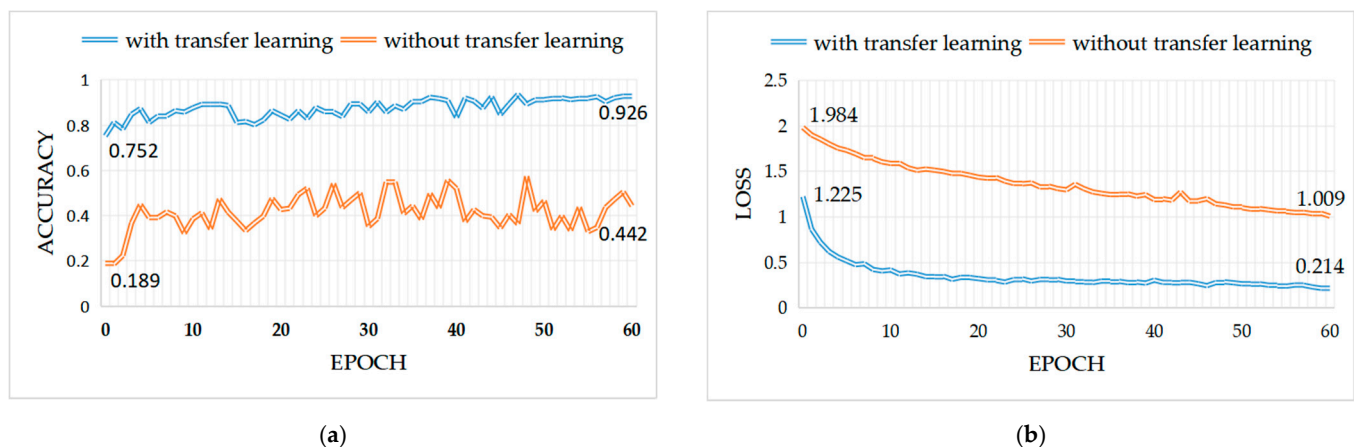
Figure 13. Performance comparison of model with transfer learning versus model without transfer learning. (**a**) Accuracy; (**b**) Loss.
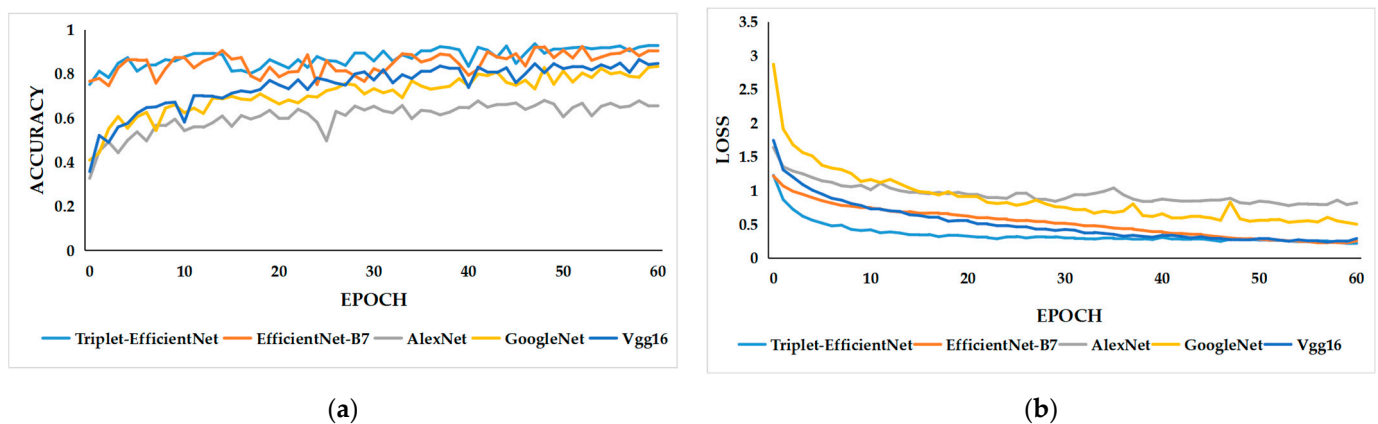
By comparing the changes in accuracy and loss value for the same model with different training strategies, it can be seen that the transfer learning method greatly speeds up the convergence process of the classification model and improves the overall accuracy of the model in the training process. This is because the transfer learning method pre-loaded the common model parameter information, which is for image feature extraction and obtained in the large-scale image dataset, to the trained model. Thus, the transfer learning method endows the model with strong image feature-characterization ability at the beginning of training, leading to improved performance and faster training convergence.

4.4.3. Evaluation of Model Training Performance

In order to evaluate the training convergence performance of the classification model based on Triplet-EfficientNet, we trained five models to do a comparative study. These models are respectively EfficientNet-B7, VGG16, GoogleNet, AlexNet [31], and Triplet-

EfficientNet. The training strategies and hyperparameters used in the training of each model were consistent.

The changes in accuracy and loss value in the training process are shown in Figure 14a,b, respectively. As can be seen from Figure 14a, the average accuracy in the training set of the five models in 60 epochs from high to low is as follows: Triplet-EfficientNet, EfficientNet-B7, VGG16, GoogleNet, and AlexNet. As shown in Figure 14b, the rank of loss values is reversed. The following Table 3 shows the concrete accuracy and loss values of each model. As can be seen from the table, the classification model based on Triplet-EfficientNet proposed in this study shows higher accuracy and lower loss value in the training set compared with EfficientNet-B7 and other models. The results show that the compound model scaling method based on EfficientNet obtains better network structure and more reasonable parameter configuration by NAS technology, and the introduction of a triplet attention mechanism further improves the overall accuracy and training convergence performance of the model.

(**a**)   (**b**)

Figure 14. The changes in accuracy and loss value of different model during training. (**a**) Accuracy; (**b**) Loss.

Table 3. Accuracy and loss value of different model in the training set.

Model	Initial Accuracy	Initial Loss	Final Accuracy (Epoch = 60)	Final Loss (Epoch = 60)
AlexNet	32.7%	1.638	65.5%	0.818
GoogleNet	40.9%	2.870	83.3%	0.501
VGG16	35.7%	1.744	84.6%	0.289
EfficientNet-B7 (+SE attention mechanism)	75.2%	1.225	90.4%	0.255
Triplet-EfficientNet (+Triplet attention mechanism)	76.6%	1.215	92.6%	0.214

4.4.4. Performance Comparison for Different Models

In this study, we aimed to evaluate the robustness and generalization ability of the classification models based on Triplet-EfficientNet and to demonstrate the superiority of our proposed method by comparing its performance with other models. We not only introduced mainstream classification models such as AlexNet, GoogleNet, and VGG16, but also replicated recent similar image classification methods, including a deep residual network model (ResNet34) proposed by Chen et al. [15] and CA-EfficientNet proposed by Gan et al. [32]. CA-EfficientNet is an improved EfficientNet model by incorporating the coordinate attention mechanism. Seven models, including the Triplet-EfficientNet model

proposed in this paper, were tested with consistent training sets, training strategy, and input image size.

In performance testing, we employed seven trained models for image inference in the test set and used the widely accepted Top-1 accuracy metric for evaluating the inference results. Top-1 accuracy refers to the accuracy with which the type with the highest probability of prediction matches the actual result [33]. As shown in Table 4, for input images of the same size, EfficientNet-B7, CA-EfficientNet, and Triplet-EfficientNet models greatly outperform other models in Top-1 accuracy due to the efficient network structure of EfficientNet. Among them, CA-EfficientNet and Triplet-EfficientNet models benefit from the introduction of the coordinate attention mechanism and triplet attention mechanism, respectively, and exhibit better performance. Compared with EfficientNet-B7, they have stronger spatial feature characterization ability and can obtain more effective feature information in rock images, thus further improving their Top-1 accuracy. Through further comparison, it was found that the triplet attention mechanism of Triplet-EfficientNet outperforms the coordinate attention mechanism of CA-EfficientNet in terms of model performance. This is because the triplet attention mechanism has more attention branches compared with the coordinate attention mechanism, allowing the model to more comprehensively extract feature information across image dimensions. Therefore, Triplet-EfficientNet which incorporates the triplet attention mechanism has the highest Top-1 accuracy.

Table 4. Performance comparison for different network models in the test set.
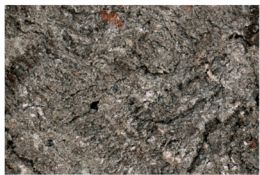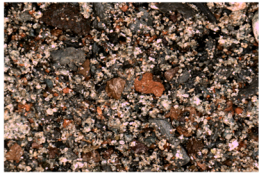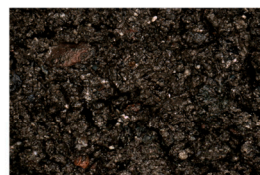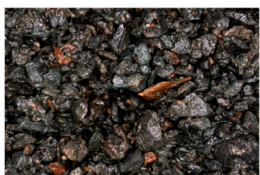
Model	Input Image Size	Top-1 Accuracy	Parameters	FLOPs
AlexNet [31]	600 × 600	71.9%	61 MB	5 G
GoogleNet [18]	600 × 600	80.6%	13 MB	10 G
VGG16 [19]	600 × 600	88.1%	138 MB	110 G
ResNet34 [15]	600 × 600	86.3%	36 MB	26 G
EfficientNet-B7 [26] (+SE attention mechanism)	600 × 600	92.0%	66 MB	38 G
CA-EfficientNet [32] (+Coordinate attention mechanism)	600 × 600	92.6%	67 MB	39 G
Triplet-EfficientNet (+Triplet attention mechanism)	600 × 600	93.2%	64 MB	36 G

In addition to testing the performance of the models, we also used two indicators, Parameters and FLOPs, to calculate the computational complexity of each model. Parameters refer to the total number of parameters that need to be trained during model training and are used to measure the computational space complexity of the model. FLOPs (Floating-point Operations) refer to the number of floating-point operations that need to be performed in a neural network model, which is used to measure the computational time complexity of the model. As shown in Table 4, EfficientNet-B7, CA-EfficientNet, and Triplet-EfficientNet achieved higher accuracy with moderate parameters and FLOPs, thanks to the efficiency of the EfficientNet network structure and its parameters. Furthermore, the triplet attention mechanism of Triplet-EfficientNet is more efficient compared to the coordinate attention mechanism of CA-EfficientNet and the SE attention mechanism of EfficientNet-B7. It not only reduces the computational complexity of the original EfficientNet model but also further improves the model's performance.

4.4.5. Reality Testing

In order to evaluate the prediction effect of the classification model based on Triplet-EfficientNet proposed in this study, seven rock images of various types are randomly selected from the rock dataset for prediction, and the prediction probability of seven rock-type labels in each image is output, as shown in Figure 15. The rock-type label corresponding to the maximum predicted probability value is the final classification result. The results show that the classification effect of this model on all kinds of selected images achieved a high level, with an accuracy of more than 95%, and the prediction probability of black coal, dark gray silty mudstone, and gray argillaceous siltstone is close to 100%. The results show that the classification model proposed in this study has great robustness and generalization ability.

(a) Dark gray mudstone
Type1: 5.49×10^{-6}
Type2: 0.967(Dark gray mudstone)
Type3: 0.033
Type4: 3.24×10^{-5}
Type5: 2.54×10^{-5}
Type6: 4.29×10^{-5}
Type7: 2.41×10^{-5}

(b) Black coal
Type1: 1.0(Black coal)
Type2: 1.48×10^{-7}
Type3: 3.17×10^{-6}
Type4: 1.87×10^{-7}
Type5: 1.42×10^{-8}
Type6: 2.46×10^{-7}
Type7: 2.64×10^{-9}

(c) Gray fine sandstone
Type1: 6.36×10^{-5}
Type2: 0.00183
Type3: 0.0397
Type4: 9.45×10^{-6}
Type5: 0.000121
Type6: 0.958(Gray fine sandstone)
Type7: 0.000285

(d) Light gray fine sandstone
Type1: 4.25×10^{-5}
Type2: 0.00259
Type3: 0.0136
Type4: 0.000542
Type5: 0.00132
Type6: 0.00125
Type7: 0.981(Light gray fine sandstone)

(e) Dark gray silty mudstone
Type1: 1.39×10^{-9}
Type2: 1.08×10^{-5}
Type3: 1.0(Dark gray silty mudstone)
Type4: 4.78×10^{-6}
Type5: 8.36×10^{-7}
Type6: 4.99×10^{-7}
Type7: 2.96×10^{-6}

(f) Gray black mudstone
Type1: 7.04×10^{-7}
Type2: 0.00457
Type3: 0.000723
Type4: 0.992(Gray black mudstone)
Type5: 0.00317
Type6: 2.35×10^{-5}
Type7: 1.63×10^{-5}

(g) Gray argillaceous siltstone
Type1: 4.22×10^{-6}
Type2: 0.000159
Type3: 3.29×10^{-5}
Type4: 1.76×10^{-5}
Type5: 1.0 (Gray argillaceous siltstone)
Type6: 6.9×10^{-6}
Type7: 2.79×10^{-5}

Type1: black coal
Type2: dark gray mudstone
Type3: dark gray silty mudstone
Type4: gray black mudstone
Type5: gray argillaceous siltstone
Type6: gray fine sandstone
Type7: light gray fine sandstone

Figure 15. The prediction probability of the classification model for each rock-type sample.

4.4.6. Comprehensive Analysis and Discussion

In this section, we validated the effectiveness and advancedness of our proposed rock image classification method based on EfficientNet and a triplet attention mechanism through a series of experiments.

First, we conducted ablation experiments to verify the effectiveness of data augmentation. This pre-processing method significantly increased the sample size of the dataset through various image transformations, effectively addressing the problem of insufficient sample size and uneven distribution among types in the original dataset, allowing the model to capture enough data patterns, and thus improving and enhancing the recognition accuracy and generalization ability of the model.

Next, we fully validated the effectiveness of the transfer learning in model training by training a classification model with transfer learning and another identical model without transfer learning. The transfer learning method loads pre-trained weights into an untrained model during the initial training phase, endowing the model with strong image feature extraction capabilities, and enabling the model to achieve a higher accuracy with less time and data samples.

We then trained five models, including EfficientNet-B7, VGG16, GoogleNet, AlexNet, and Triplet-EfficientNet, respectively, to compare their model training performance, thus

validating the effectiveness of the EfficientNet network structure and Triplet attention mechanism in improving model training performance.

To further validate the robustness and generalization ability of our proposed model, as well as its superiority over other mainstream and cutting-edge models, we trained six models (AlexNet, GoogleNet, VGG16, ResNet34, EfficientNet-B7, CA-EfficientNet) and our proposed Triplet-EfficientNet. We applied the same training strategy and parameter settings to all models. After training, we conducted performance testing and computational complexity testing on these models. The results showed that the high efficiency of the EfficientNet network structure allowed EfficientNet-B7, CA-EfficientNet, and Triplet-EfficientNet to achieve higher model accuracy with moderate Parameters and FLOPs. Meanwhile, the triplet attention mechanism of Triplet-EfficientNet was more efficient in image inference compared to the SE attention mechanism of EfficientNet-B7 and the coordinate attention mechanism of CA-EfficientNet, allowing EfficientNet models to further improve accuracy while reducing Parameters counts and FLOPs.

Finally, we assessed the actual performance of our rock image classification model on seven images containing various types of rocks. The results demonstrated that the model achieved accurate classification of all test images with a prediction accuracy of over 95%. These outcomes showcase the exceptional ability of our model in classifying rock images.

5. Conclusions

In this study, a rock image classification model based on EfficientNet and a triplet attention mechanism is proposed to achieve accurate end-to-end rock image classification. To begin, we expanded the rock image dataset through various data augmentation methods to prevent overfitting of model training and improve model performance. In building the model, we utilized EfficientNet as the benchmark network, which boasts an efficient network structure thanks to NAS technology and a compound model scaling method. On this basis, the triplet attention mechanism was introduced to improve the original EfficientNet and enhance the model's ability to extract spatial features of rock images. The experimental results demonstrate that the classification model in this study outperforms other mainstream models on both the training set and the test set, the accuracy reached 92.6% and 93.2% respectively. In training the model, we employed the transfer learning method during the training process to accelerate model convergence and significantly enhance the model's training performance. The training accuracy of the model with transfer learning increased by 48.4% compared with that without transfer learning.

Through further research, it has been revealed that the number of samples and rock types in the rock dataset have a significant impact on the number of rock types that can be recognized by the model and the final classification accuracy. With this in mind, our future research endeavors will concentrate on expanding the variety of rock types and the quantity of rock images, while ensuring that the classification accuracy is further enhanced with the addition of more rock types. Furthermore, given the difficulty of obtaining timely feedback on rock-type recognition results through networks in field exploration, we intend to deploy our model on mobile devices in addition to improving it. This will allow geological surveyors to conveniently identify rocks using the classification model under offline conditions.

Author Contributions: Conceptualization, Z.H. and L.S.; methodology, Z.H.; software, Z.H.; validation, Z.H., J.W. and Y.C.; formal analysis, Z.H.; investigation, Z.H.; resources, Z.H.; data curation, Z.H.; writing—original draft preparation, Z.H.; writing—review and editing, Z.H.; visualization, Z.H.; supervision, L.S.; project administration, Z.H.; funding acquisition, L.S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

References

1. Fu, G.; Yan, J.; Zhang, K.; Hu, H.; Luo, F. Current status and progress of lithology identification technology. *Prog. Geophys.* **2017**, *32*, 26–40.
2. Zhang, S.; Bogus, S.M.; Lippitt, C.D.; Kamat, V.; Lee, S. Implementing Remote-Sensing Methodologies for Construction Research: An Unoccupied Airborne System Perspective. *J. Constr. Eng. Manag.* **2022**, *148*, 3122005. [CrossRef]
3. Guo, Q.; Zhou, Y.; Cao, S.; Qiu, Z.; Xu, Z.; Zhang, Y. Study on Mineralogy of Guangning Jade. *Acta Sci. Nat. Univ. Sunyatseni* **2010**, *49*, 146–151.
4. Młynarczuk, M.; Górszczyk, A.; Ślipek, B. The application of pattern recognition in the automatic classification of microscopic rock images. *Comput. Geosci.* **2013**, *60*, 126–133. [CrossRef]
5. Xiao, F.; Chen, J.; Hou, W.; Wang, Z. Identification and extraction of Ag-Au mineralization associated geochemical anomaly in Pangxitong district, southern part of the Qinzhou-Hangzhou Metallogenic Belt, China. *Acta Petrol. Sin.* **2017**, *33*, 779–790.
6. Xu, Z.; Ma, W.; Lin, P.; Shi, H.; Liu, T.; Pan, D. Intelligent Lithology Identification Based on Transfer Learning of Rock Images. *J. Basic Sci. Eng.* **2021**, *29*, 1075–1092.
7. Lippitt, C.D.; Zhang, S. The impact of small unmanned airborne platforms on passive optical remote sensing: A conceptual perspective. *Int. J. Remote Sens.* **2018**, *39*, 4852–4868. [CrossRef]
8. Marmo, R.; Amodio, S.; Tagliaferri, R.; Ferreri, V.; Longo, G. Textural identification of carbonate rocks by image processing and neural network: Methodology proposal and examples. *Comput. Geosci.* **2005**, *31*, 649–659. [CrossRef]
9. Singh, N.; Singh, T.; Tiwary, A.; Sarkar, K.M. Textural identification of basaltic rock mass using image processing and neural network. *Comput. Geosci.* **2010**, *14*, 301–310. [CrossRef]
10. Yen, H.H.; Tsai, H.Y.; Wang, C.C.; Tsai, M.C.; Tseng, M.H. An Improved Endoscopic Automatic Classification Model for Gastroesophageal Reflux Disease Using Deep Learning Integrated Machine Learning. *Diagnostics* **2022**, *12*, 2827. [CrossRef]
11. Dimitrovski, I.; Kitanovski, I.; Kocev, D.; Simidjievski, N. Current trends in deep learning for Earth Observation: An open-source benchmark arena for image classification. *ISPRS J. Photogramm. Remote Sens.* **2023**, *197*, 18–35. [CrossRef]
12. Xu, S.; Zhou, Y. Artificial intelligence identification of ore minerals under microscope based on deep learning algorithm. *Acta Petrol. Sin.* **2018**, *34*, 3244–3252.
13. Zhang, Y.; Li, M.; Han, S. Automatic identification and classification in lithology based on deep learning in rock images. *Acta Petrol. Sin.* **2018**, *34*, 333–342.
14. Cheng, G.; Li, P. Rock thin-section image classification based on residual neural network. In Proceedings of the 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 9–11 April 2021; pp. 521–524.
15. Chen, W.; Su, L.; Chen, X.; Huang, Z. Rock image classification using deep residual neural network with transfer learning. *Front. Earth Sci.* **2023**, *10*, 1079447. [CrossRef]
16. Koeshidayatullah, A.; Al-Azani, S.; Baraboshkin, E.E.; Alfarraj, M. Faciesvit: Vision transformer for an improved core lithofacies prediction. *Front. Earth Sci.* **2022**, *10*, 992442. [CrossRef]
17. Zhang, W.; Zhang, Q.; Liu, S.; Pan, X.; Lu, X. A Spatial–Spectral Joint Attention Network for Change Detection in Multispectral Imagery. *Remote Sens.* **2022**, *14*, 3394. [CrossRef]
18. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 8–10 June 2015; pp. 1–9.
19. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
20. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
21. Misra, D.; Nalamada, T.; Arasanipalai, A.U.; Hou, Q. Rotate to attend: Convolutional triplet attention module. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Online, 5–9 January 2021; pp. 3139–3148.
22. Okada, H. Classification of sandstone: Analysis and proposal. *J. Geol.* **1971**, *79*, 509–525. [CrossRef]
23. Haimson, B.; Rudnicki, J.W. The effect of the intermediate principal stress on fault formation and fault angle in siltstone. *J. Struct. Geol.* **2010**, *32*, 1701–1711. [CrossRef]
24. Vaniman, D.; Bish, D.; Ming, D.; Bristow, T.; Morris, R.; Blake, D.; Chipera, S.; Morrison, S.; Treiman, A.; Rampe, E. Mineralogy of a mudstone at Yellowknife Bay, Gale crater, Mars. *Science* **2014**, *343*, 1243480. [CrossRef]
25. Huang, Y.; Cheng, Y.; Bapna, A.; Firat, O.; Chen, D.; Chen, M.; Lee, H.; Ngiam, J.; Le, Q.V.; Wu, Y. Gpipe: Efficient training of giant neural networks using pipeline parallelism. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 103–112.

26. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 6105–6114.
27. Yao, Q.; Wang, M.; Chen, Y.; Dai, W.; Li, Y.F.; Tu, W.W.; Yang, Q.; Yu, Y. Taking human out of learning applications: A survey on automated machine learning. *arXiv* **2018**, arXiv:1810.13306.
28. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
29. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
30. Santoro, A.; Bartunov, S.; Botvinick, M.; Wierstra, D.; Lillicrap, T. Meta-learning with memory-augmented neural networks. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 1842–1850.
31. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
32. Gan, Y.; Guo, Q.; Wang, C.; Liang, W.; Xiao, D.; Wu, H. Recognizing crop pests using an improved EfficientNet model. *Trans. Chin. Soc. Agric. Eng.* **2022**, *38*, 203–211.
33. Wei, Y.; Wang, Z.; Qiao, X.; Zhao, C. Lightweight rice disease identification method based on attention mechanism and EfficientNet. *J. Chin. Agric. Mech.* **2022**, *43*, 172–181.