

文章

基于EfficientNet和三元组注意力机制的岩石图像分类

黄志豪, 苏绿梅*, 吴佳俊, 陈宇涵

厦门理工学院电气工程与自动化学院, 福建 厦门 361024* 通信作者: sulumei@163.com

特色应用: 该工作提出了一种用于岩石类型识别的图像分类算法, 可为地质调查提供可靠的指导。

摘要: 岩石图像分类是地质调查工作中的一项基础且关键的任务。传统的岩石图像分类方法主要依赖人工操作, 成本高且精度不稳定。虽然现有的基于深度学习模型的方法克服了传统方法的局限性, 实现了图像智能分类, 但由于网络结构欠佳, 仍存在精度较低的问题。本研究提出了一种基于 EfficientNet 和三重注意力机制的岩石图像分类模型, 以实现准确的端到端分类。该模型基于 EfficientNet 构建, 借助神经架构搜索 (NAS) 技术和复合模型缩放方法, 拥有高效的网络结构, 能够实现较高的岩石图像分类精度。此外, 引入三重注意力机制, 弥补了 EfficientNet 在特征表达方面的不足, 使模型能够充分捕捉岩石图像的通道和空间注意力信息, 进一步提高了分类精度。在网络训练过程中, 采用迁移学习方法, 将预训练模型参数加载到分类模型中, 加速了收敛速度, 减少了训练时间。结果表明, 采用迁移学习的分类模型在训练集上的准确率达到 92.6%, 在测试集上的 Top-1 准确率达到 93.2%, 优于其他主流模型, 表现出较强的鲁棒性和泛化能力。

关键词: 岩石图像; 高效网络 (EfficientNet); 图像分类; 迁移学习



引用: 黄, Z.; 苏, L.; 吴, J.; 陈, Y. 基于EfficientNet和三元组注意力机制的岩石图像分类。《应用科学》2023年, 第13卷, 第 3180页。

<https://doi.org/10.3390/app13053180>

学术编辑: 安德里亚·普拉蒂

接收日期: 2023年2月13日

修订日期: 2023年2月24日

接受日期: 2023年2月27日

发表时间: 2023年3月1日



版权所有: © 2023 作者保留所有权利。许可方: 瑞士巴塞尔 MDPI。本文为开放获取文章, 根据知识共享署名 (CC BY) 许可协议 (<https://creativecommons.org/licenses/by/4.0/>) 的条款和条件分发。

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.



Citation: Huang, Z.; Su, L.; Wu, J.; Chen, Y. Rock Image Classification Based on EfficientNet and Triplet Attention Mechanism. *Appl. Sci.* **2023**, *13*, 3180. <https://doi.org/10.3390/app13053180>

Academic Editor: Andrea Prati

Received: 13 February 2023

Revised: 24 February 2023

Accepted: 27 February 2023

Published: 1 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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.

随着计算机技术的发展，基于机器学习和深度学习的计算机视觉方法在图像分类中开始展现出优异的性能，并已应用于岩石分类领域。研究人员最初采用传统的机器学习方法，即利用人工提取的岩石图像特征训练分类器以实现自动分类。例如，Marmo等人[8]通过图像处理从1000多个碳酸盐切片中提取图像特征值，并将其输入到多层感知器神经网络模型中，实现了碳酸盐岩的智能分类，准确率达到93.3%。同样，Singh等人[9]从玄武岩薄片图像中提取了27个特征，能够识别和分类140个岩石样本薄片图像，准确率为92.22%。尽管这些传统的机器学习方法能够取得令人满意的结果，但由于浅层模型的能力有限，它们无法直接从未处理的图像中自动提取特征[10]。因此，在训练分类模型之前，需要对图像进行人工预处理，以提取颜色、形状和纹理等特征。这显著降低了类型分类的自动化程度。

近年来，深度学习已应用于岩石类型分类，以克服传统机器学习方法在图像特征提取方面的局限性[11]。研究人员利用各种深度学习模型自动提取图像特征，从而实现更智能的图像分类。徐等人[12]设计了一种U型卷积神经网络，用于显微镜下的矿石矿物识别，测试集成功率达90%以上。张等人[13]基于Inception - V3深度卷积神经网络开发了一种适用于各种岩石类型的岩石图像识别模型，分类准确率超过85%。程等人[14]使用ResNet50和ResNet101模型对岩石切片图像进行自动特征提取和分类，测试集准确率分别为90.24%和91.63%。陈等人[15]提出了一种采用迁移学习的深度残差神经网络模型，以建立准确率超过90%的自动岩石分类模型。科希达亚图拉等人[16]提出了一种基于Transformer的自动岩心表面分类模型，无需进行预处理和手动特征提取。深度学习方法具有自动提取图像特征的优势，消除了主观因素对实验结果的影响，大大减轻了岩石分类的工作量。此外，深度学习模型可以提取更抽象、更复杂的图像特征，以对更多类型的岩石进行分类[17]。然而，现有的深度学习模型通常是通过手动设计网络模块并进行堆叠来构建的[18 - 20]，导致网络结构不合理且参数冗余。此外，对于像岩石图像这样干扰强、信息杂乱的对象，现有的无注意力机制的模型在图像处理过程中注意力分散，难以有效捕捉有用特征。这些因素最终导致当前模型在岩石图像分类中的准确率有限。

目前的岩石图像分类方法各有其局限性。传统方法在特定条件下可以达到可接受的精度，但严重依赖人工操作，且精度缺乏稳定性。虽然机器学习方法在岩石图像智能分类方面取得了初步进展，但仍需要人工进行特征提取，限制了图像识别的自动化程度。现有的深度学习方法实现了端到端的自动图像识别，但仍存在网络参数冗余和注意力分散的问题，导致岩石图像分类的精度有限。为了克服现有方法的局限性，实现岩石图像的准确端到端分类，本文提出了一种基于EfficientNet和三重注意力机制的岩石图像分类模型。本研究的贡献总结如下：

- 引入深度学习方法，消除传统方法和机器学习方法对人工干预的依赖，从而实现岩石图像的端到端自动识别，无需额外的人工操作。

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.

- 基于EfficientNet构建了岩石图像分类模型，该模型克服了以往深度学习模型中参数冗余和注意力分散的问题，并通过神经架构搜索实现了高效且聚焦注意力的网络结构，与前代模型相比，提高了模型的准确性。
- 针对EfficientNet忽略岩石图像空间注意力信息的问题，引入三元组注意力机制[21]对其进行改进，增强其提取有效岩石特征的能力，进一步提高岩石分类模型的准确性。
- 在训练过程中采用迁移学习方法，加速模型收敛，显著提升训练性能，从而用更少的岩石图像和更短的训练时间获得更高精度的分类模型。

2. 材料

岩石图像数据集由广东天枢智能科技有限公司（中国武汉）提供，包含315张岩石图像。这些岩石图像是地质人员在测井现场使用工业相机在白光条件下拍摄的岩石碎片高分辨率照片。图像使用固定高度的镜头拍摄。此外，大多数图像的初始尺寸为 4096×3000 像素，少数带有背景的图像尺寸为 2448×2048 像素。岩石碎片图像采集过程通常包括以下步骤：

- (1) 收集岩石碎片：地质人员在井口或钻机处收集岩石碎片，使用手动或机械工具进行采集。
- (2) 制备样本：收集岩石碎片后，需要进行清洗和筛分等初步处理，以去除杂质和不需要的部分，获得足够数量的同一岩石类型的样本。
- (3) 采集图像：对于每个岩石碎块样本，使用工业相机或其他设备采集其图像。在图像采集过程中，保持样本处于相同的位置和角度非常重要，以确保图像的可比性和可重复性。

岩石碎块图像通常受岩石多样的物理和化学性质影响，且一般尺寸较小、形状复杂多样，纹理和颜色也存在差异。此外，岩石碎块图像中的噪声和背景干扰可能会对图像质量产生不利影响。所有这些因素都给此类图像中岩石类型的识别带来了挑战。

该数据集总共包含七种类型的岩石图像，其中包括21张黑煤图像、30张灰黑色泥岩图像、46张灰色泥质粉砂岩图像、18张灰色细砂岩图像、85张浅灰色细砂岩图像、40张深灰色粉砂质泥岩图像和75张深灰色泥岩图像。图1展示了七种不同类型的岩石图像。

不同的岩石具有独特的特征，如形状、颜色和质地。表1展示了数据集中各种岩石图像的关键特征。表中每行的序号对应图1。表中的七种岩石根据粒度可分为四类：砂岩、粉砂岩、泥岩和煤。砂岩通常呈灰色或浅灰色，形状非块状，表面粗糙[22]。粉砂岩大多为灰色，呈片状和块状，表面光滑[23]。泥岩颜色可为灰色、深灰色或灰黑色，且往往呈片状和块状[24]。煤是数据集中最易识别的岩石类型，颜色为黑色。

- 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.

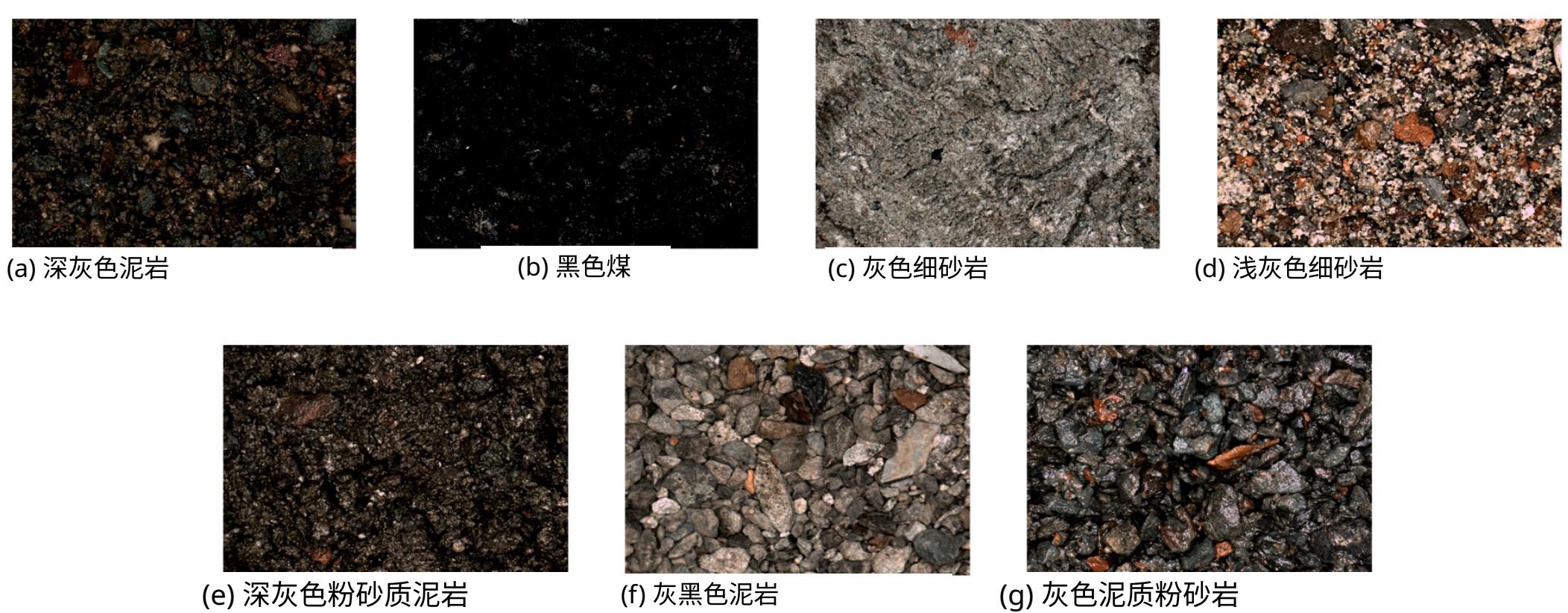


图1. 岩石图像数据集中不同类型的七张岩石图像。

表1. 岩石图像数据集中各类岩石的特征。

序号 图像数量	类型	粒径 (毫米)	颜色	形状特征
a	深灰色泥岩	<0.005	深灰色	薄片状, 泥晶结构
b	黑煤	<6	黑色	颗粒状、不对称
c	灰色细砂岩	0.05–2	灰色	细粒结构、表面粗糙 不均匀
d	浅灰色细砂岩	0.05–2	浅灰色	细砂结构, 薄层状结 构
e	深灰色粉砂质泥岩	<0.005	深灰色	粉质泥质结构, 层理结构
f	灰黑色泥岩	<0.005	灰黑色	隐晶质结构, 块状结构
g	灰色泥质粉砂岩	0.005 - 0.05	灰色	粉质结构, 块状 结构

3. 方法

在本研究中，我们提出了一种基于EfficientNet和三元组注意力机制的岩石类型分类方法。该方法侧重于建立分类模型，并集成了迁移学习和数据增强等多种方法，以实现岩石图像的准确自动分类。图2是所提方法的流程图。该方法的详细信息以伪代码形式在算法1中展示。

3.1. EfficientNet神经网络

模型缩放已被广泛用于提高卷积神经网络的准确性。在以往的工作中，最常见的方法是仅在单一维度上改变基准神经网络的网络深度、宽度或输入图像分辨率，如图3b - d所示。例如，Huang等人[25]通过增加基准网络的深度，极大地提高了GPipe在ImageNet数据集上的准确性。虽然可以同时在多个维度上对模型进行缩放，但多维缩放需要繁琐的手动调优。为了实现简单而高效的模型缩放，Tan等人[26]提出，EfficientNet是通过神经架构搜索（NAS）[27]技术和复合缩放方法得到的，它是ImageNet数据集上分类性能最佳的网络之一。他们首先使用NAS技术搜索基准网络的结构，然后通过复合缩放方法在多个维度上对基准网络进行放大。这种缩放方法允许网络深度、宽度和输入图像分辨率进行统一变化，如图3e所示，无需额外的微调即可获得更高的分类准确率。

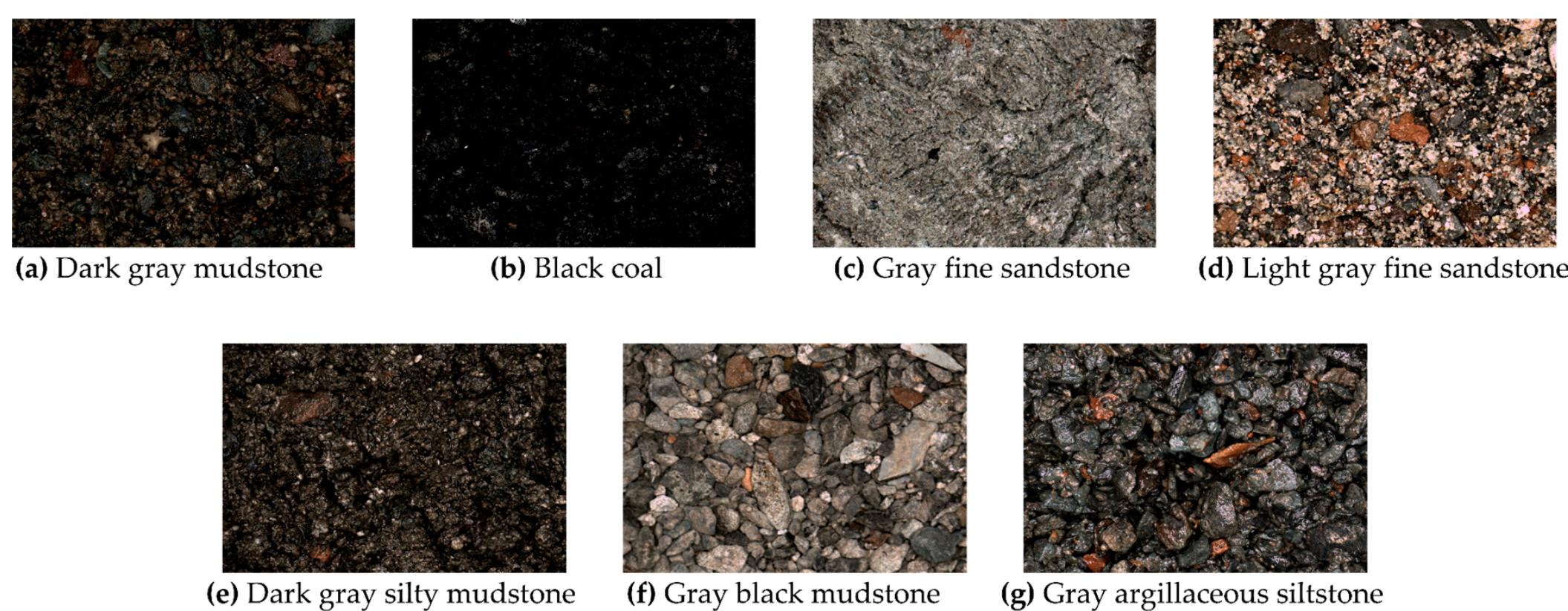


Figure 1. Seven rock images of different types in the rock image datasets.

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.

算法1：基于EfficientNet和三元组注意力机制的岩石类型分类方法

输入：包含七种岩石类型的315张岩石图像
 1: 进行数据预处理。
 2: 对每张图像随机应用以下九种数据增强操作：旋转、添加椒盐噪声、增亮、变暗、放大、垂直翻转、水平翻转、添加高斯噪声和平移。（增强后岩石图像数量增加到6949张。）
 3: 将增强后的图像按照60%、20%和20%的比例随机划分为训练集、验证集和测试集。（它们的样本数量分别为4169、1389和1389。）
 4: 构建基于EfficientNet和三元组注意力机制的岩石类型分类模型。
 5: 选择EfficientNet - B7模型作为基线模型。
 6: 用三元组注意力模块替换EfficientNet - B7模型的每个SE注意力模块，以构建三元组Efficient模型。
 7: 基于三元组高效网络（Triplet-EfficientNet）模型作为骨干网络构建分类模型。
 8: 在三元组高效网络（Triplet-EfficientNet）之后添加 1×1 卷积层、池化层、全连接层和 Softmax 分类器。
 9: 将 Softmax 分类器的类别数量设置为 7。
 10: 开始模型训练。
 11: 使用迁移学习方法：加载在 ImageNet 数据集上预训练的模型参数到未训练的模型中。
 12: 设置训练超参数：将学习率设置为 0.01，轮数设置为 60，批量大小设置为 16。
 13: 选择 Swish 函数作为激活函数，交叉熵函数作为损失函数，以及自适应矩估计（Adam）作为优化器。
 14: 将训练集中的图像统一缩放到 $600 \times 600 \times 3$ 大小，并随机打包输入模型以开始训练。
 15: 对迁移模型的所有参数进行 60 轮训练。
 16: 训练后输出最终模型。
 17: 开始测试：输入一张随机选取的岩石图像。
 输出：该岩石图像被分类为每种岩石类型的概率。（最大概率对应的岩石类型即为最终识别结果。）

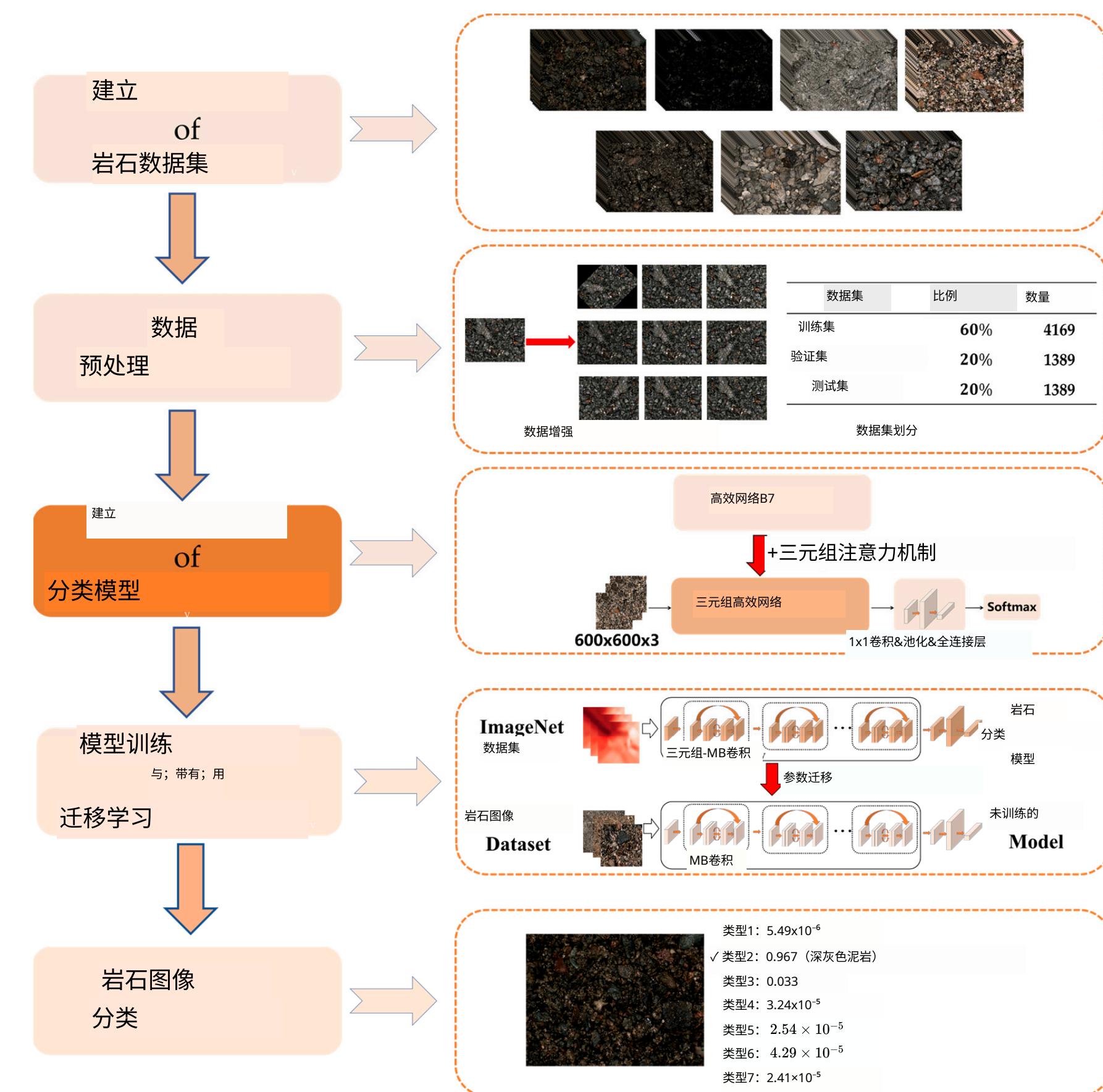


图2. 本研究提出的分类方法流程图。

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.)

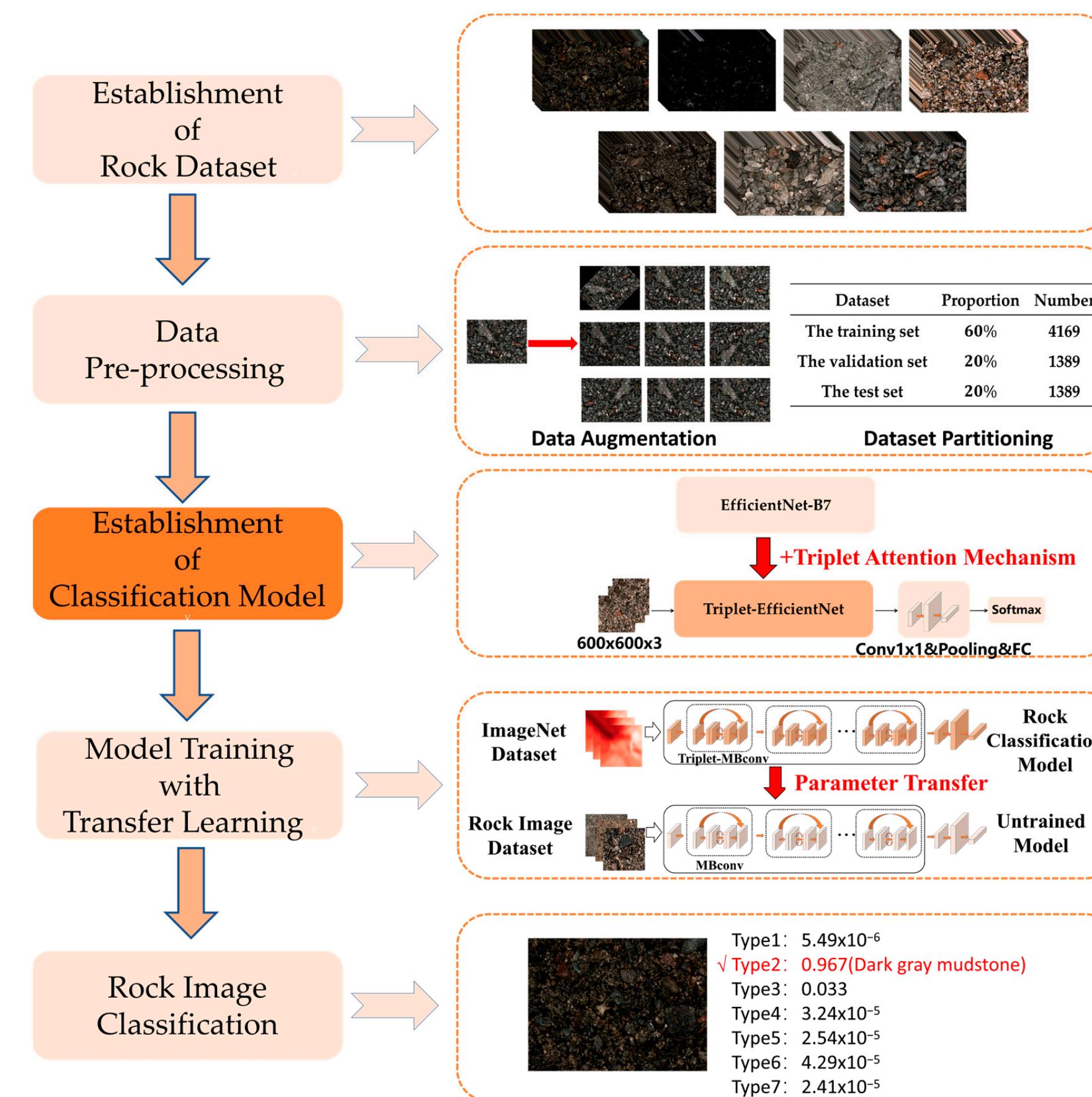


Figure 2. Flowchart of the classification method proposed in this study.

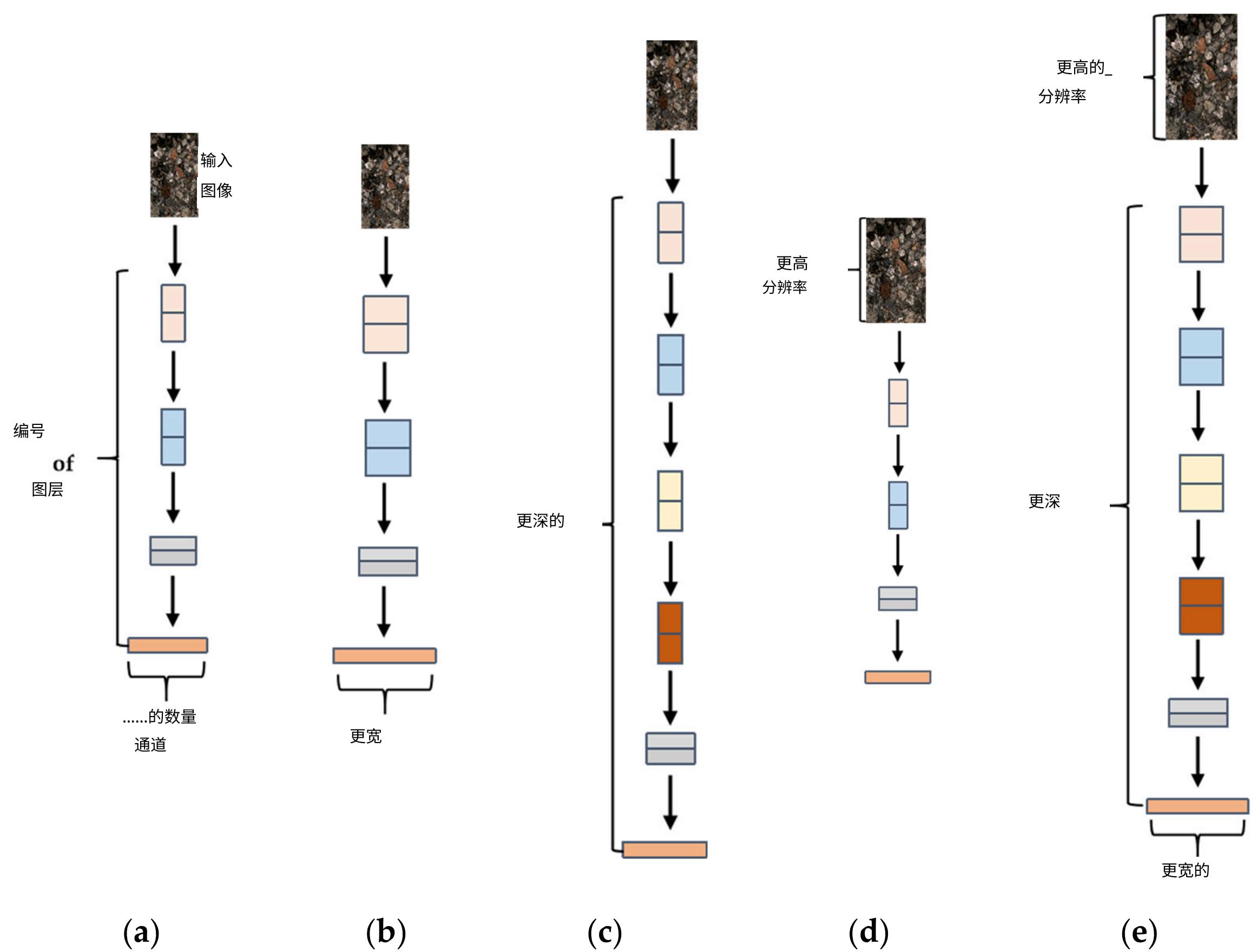


图3. 不同模型缩放方法的示意图。(a)一个基准网络示例; (b - d)仅改变网络宽度、深度或分辨率一个维度的模型缩放方法。(e)一种统一缩放三个维度的复合缩放方法。

基于EfficientNet的复合缩放方法的关键是找到一组深度、宽度和图像分辨率的复合系数，以使网络性能最大化。这个优化问题在数学上如公式(1)所示。

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

其中， d, w, r 分别是网络深度、宽度和图像分辨率的缩放系数， $N(d, w, r)$ 是分类模型， $\max_{d,w,r}$ 准确率是模型的最大准确率。

为实现 d, w, r 的均匀缩放，该方法引入了 φ ，它是一个用户指定的控制模型规模的系数，如公式(2)所示。这种缩放方法的实现步骤如下：在确定基准网络的结构后，该方法首先将控制系数 φ 固定为1，然后使用神经架构搜索 (NAS) 技术搜索使分类准确率最大化的系数 d, w, r ，从而得到最终的基准模型，称为EfficientNet - B0；最后，该方法指定从2到7的不同 φ 值，并分别获得不同大小的对应模型，分别称为 EfficientNet - B1、EfficientNet - B2、……、EfficientNet - B7。

$$d = \alpha^\varphi, w = \beta^\varphi, r = \gamma^\varphi \quad (2)$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

其中， α, β, γ 是可以通过神经架构搜索 (NAS) 技术确定的常数。

3.2. EfficientNet - B7模型

EfficientNet - B7是通过对EfficientNet - B0进行放大得到的高精度模型，其输入图像分辨率为 600×600 ，宽度乘数因子(w)为2.0，深度乘数

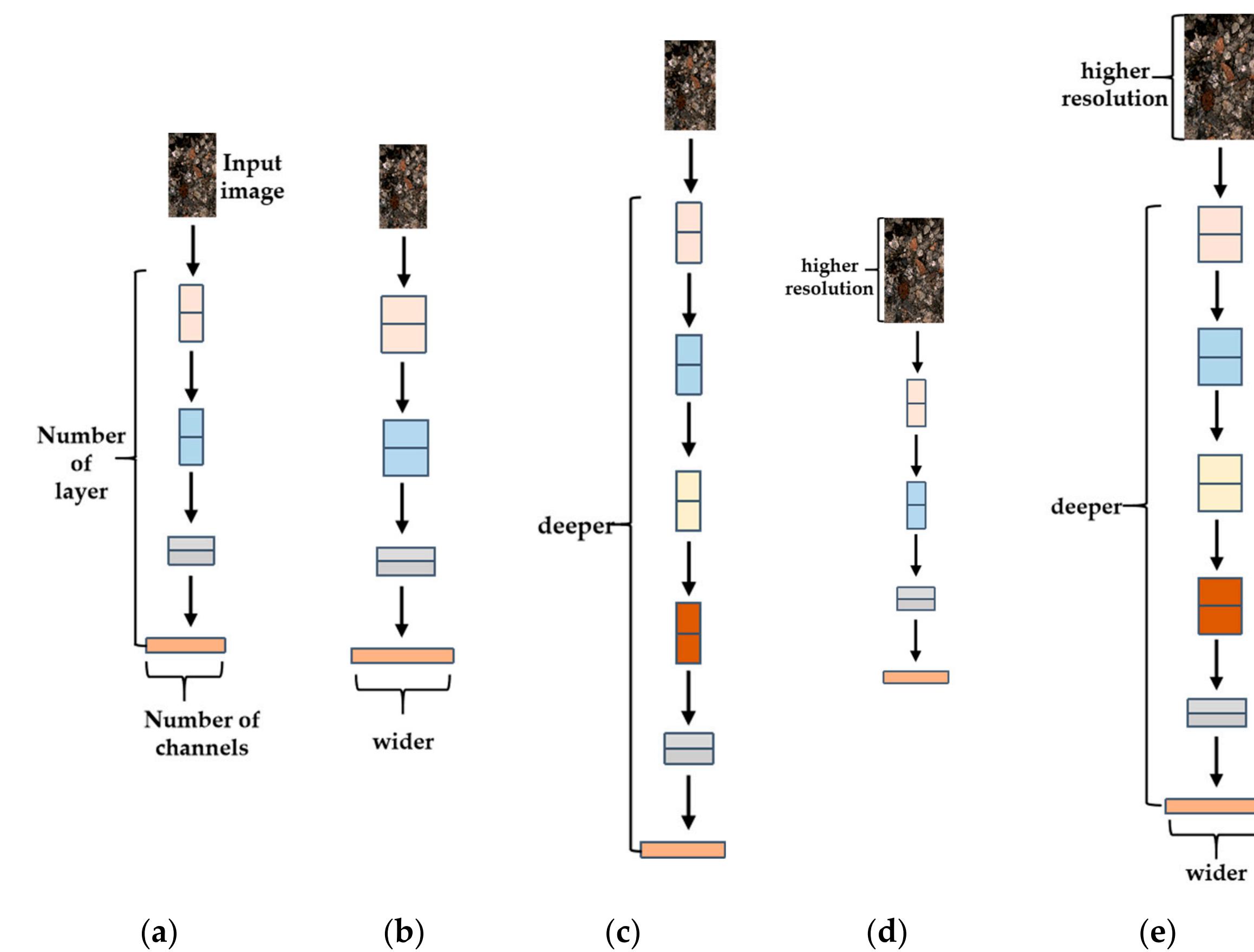


Figure 3. Diagrams of different model scaling methods. (a) a baseline network example; (b-d) model scaling methods that only change one dimension of network width, depth, or resolution. (e) a compound scaling method that uniformly scales three dimensions.

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).

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

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

$$d = \alpha^\varphi, w = \beta^\varphi, r = \gamma^\varphi \quad (2)$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

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

因子(d)为3.1。在本研究中，为了在岩石图像数据集上实现尽可能高的分类准确率，选择了 EfficientNet - B7作为基准模型。

如图4所示，EfficientNet - B7模型是通过堆叠多个移动倒置瓶颈卷积 (MBConv) 模块构建而成的。MBConv模块的结构与传统残差模块不同，其输入和输出特征图都比中间部分更宽。如图5所示，MBConv模块包括一个核大小为 1×1 的卷积层、深度可分离卷积、挤压 (SE) 注意力模块和Dropout层。MBConv模块在卷积层之后应用了批量归一化 (BN) 和 Swish 激活函数。BN可以对数据进行归一化处理，并在训练过程中加快模型收敛速度，而 Swish 激活函数可以为数据引入非线性特性，避免过拟合。此外，MBConv模块中的SE注意力模块采用了SE [28] 注意力机制，该机制通过捕获输入特征图中的通道注意力信息来增强模型的特征表示能力。

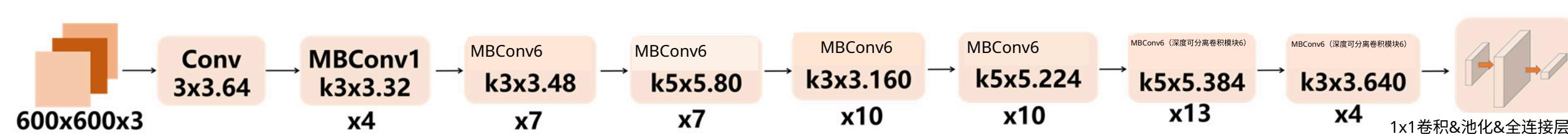


图4. EfficientNet - B7网络结构。

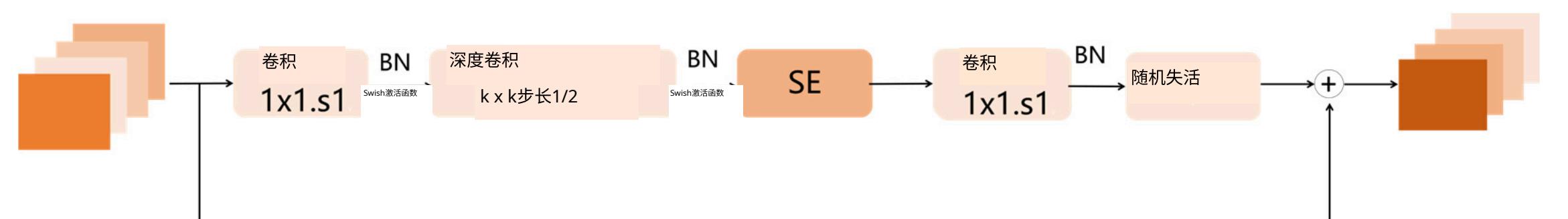


图5. 移动倒置瓶颈卷积 (MBConv) 模块结构。

SE注意力模块利用SE注意力机制来计算输入特征图中每个通道对于当前任务的重要性，并对其进行加权。这通过以下三个操作实现：

- (1) 压缩：将大小为 $C \times H \times W$ 的输入特征图进行全局平均池化，得到大小为 $1 \times 1 \times C$ 的特征图，从而将每个二维特征通道压缩为一个单一值，以表示每个通道上响应的全局分布。
- (2) 激励：使用全连接神经网络对压缩后的特征图进行非线性变换，通过ReLU和Sigmoid激活函数生成激活权重。
- (3) 缩放：通过点乘操作，使用激活后的权重对输入特征图中的每个通道进行加权。

如图6所示，将大小为 $C \times H \times W$ 的岩石特征图输入到 SE 注意力模块中。通过上述三个操作，该模块根据每个通道对岩石图像分类准确率的影响为其分配权重，增强有效的岩石特征通道并抑制较弱的通道。在该图中，输出张量中的不同通道具有不同颜色的边框，这表明经过SE模块后，每个通道被分配了不同的权重。

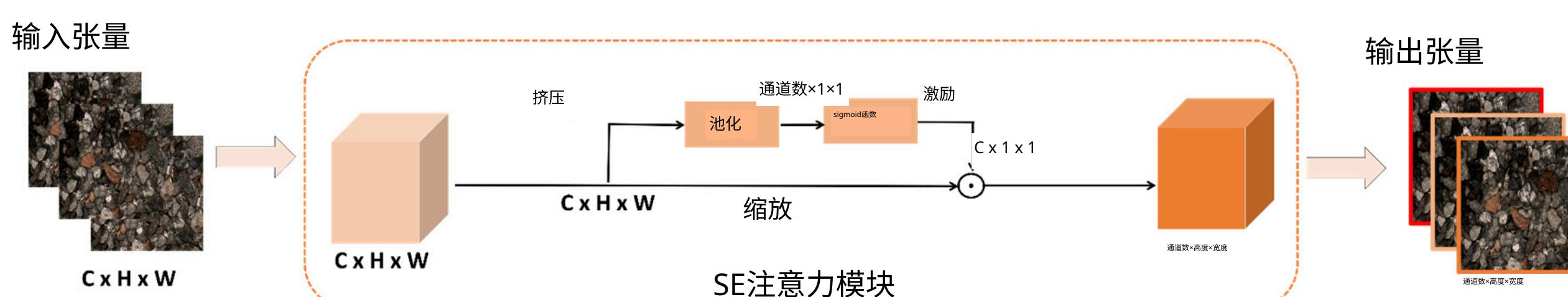


图6. EfficientNet中SE注意力模块的结构。

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.

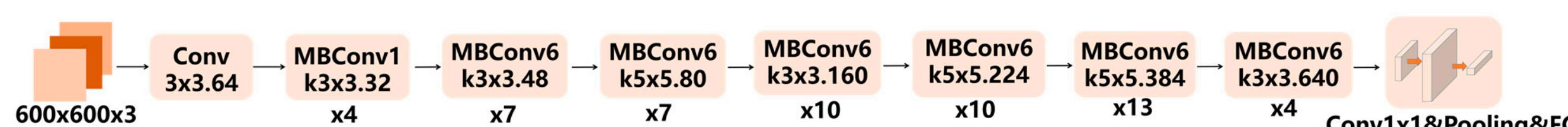


Figure 4. EfficientNet-B7 network structure.

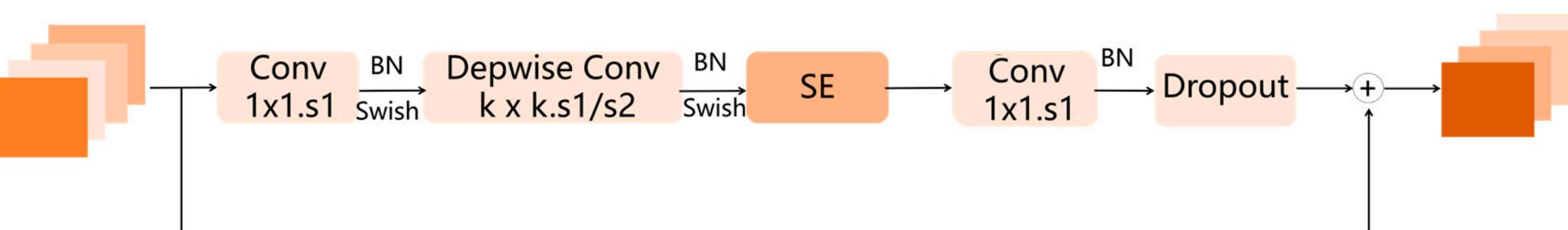


Figure 5. MBConv module structure.

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.

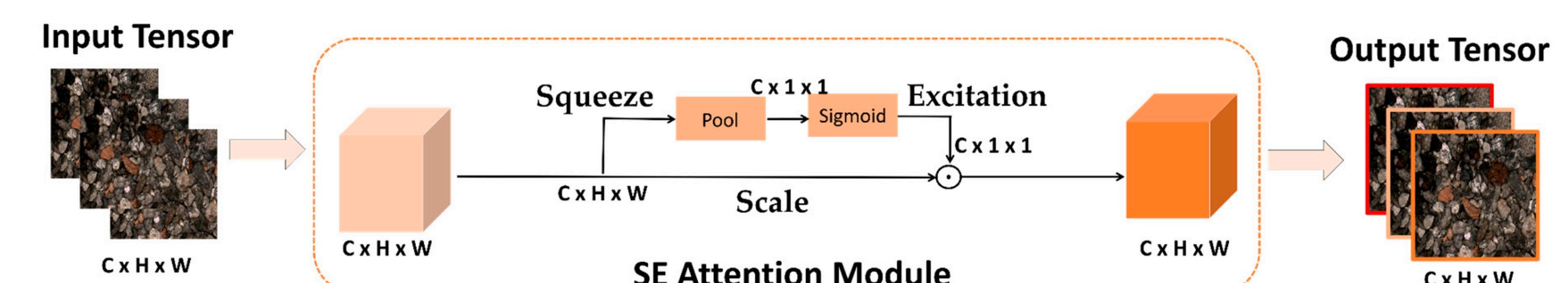


Figure 6. Structure of SE attention module in EfficientNet.

3.3. 三元组注意力机制

在本研究中，我们旨在对自然条件下拍摄的各种未经预处理的岩石图像进行分类。尽管数据集中的每幅岩石图像仅包含一种岩石类型，但在同一图像的不同空间位置，岩石的形态、纹理和颜色特征可能会略有差异，导致岩石图像中岩石特征分布不均。EfficientNet的SE注意力模块仅关注不同岩石特征通道的重要性，无法计算不同空间位置岩石特征的重要性，从而限制了基于EfficientNet的岩石图像分类模型的准确性。为解决这一问题，我们引入了一个能够同时捕捉空间和通道注意力信息的三元组注意力模块，以取代EfficientNet中的SE模块，从而提出了一种改进的EfficientNet模型——三元组EfficientNet。

如图7所示，三元组注意力模块不仅为输入的岩石特征张量的不同特征通道分配权重，还为每个通道上的不同空间位置分配权重。图7中的输出岩石特征张量不仅有彩色边框表示每个特征通道的不同权重，而且每个通道上不同位置的颜色也不同。每个特征通道中更接近蓝色的部分代表对分类准确率影响较小的空间特征，模块会为其分配较小的权重；而更接近红色的部分代表更有效的岩石空间特征，因此会受到模块更多的关注，并被分配更大的权重。因此，通过在分类模型中引入三元组注意力模块，我们可以有效解决岩石图像中空间特征分布不均衡的问题，显著提高模型提取有效特征的能力，最终提升其整体准确率。

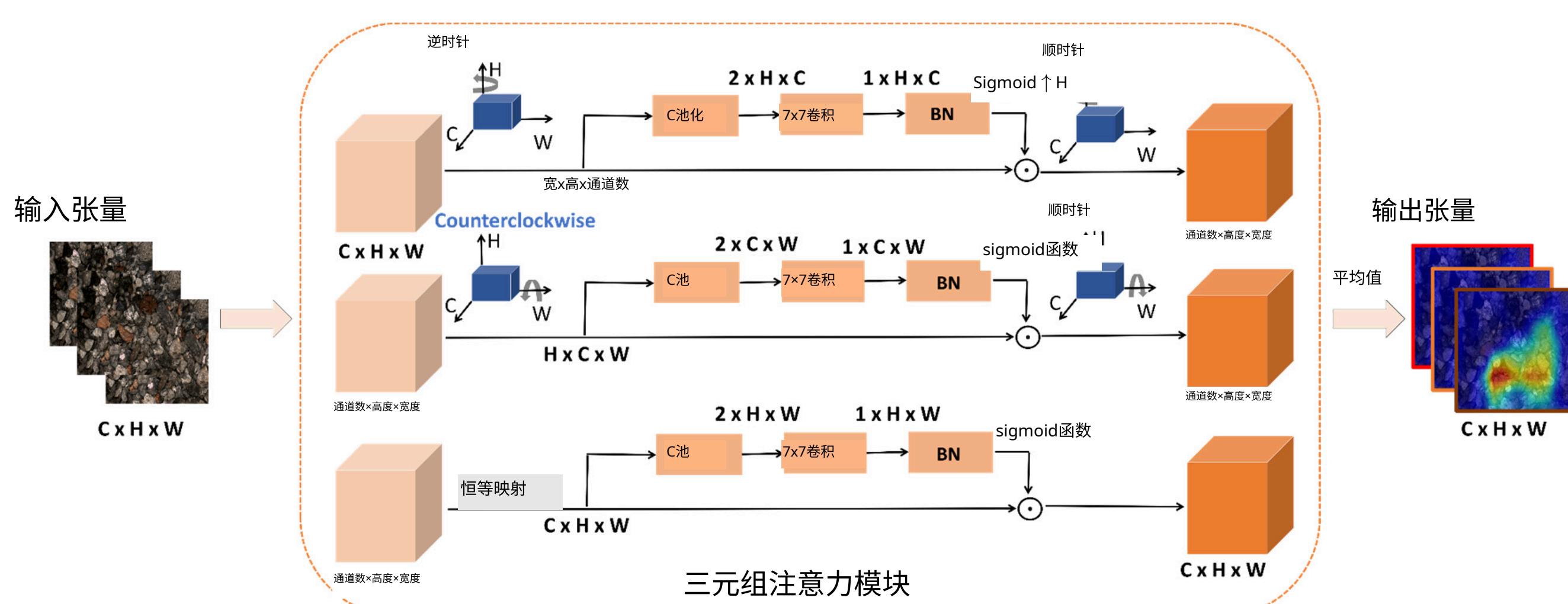


图7. 三元组注意力模块的结构。

三元组注意力模块由三个并行分支组成，如图7所示，该模块接收一个输入的岩石特征张量，并输出一个相同形状的精炼张量。给定一个输入的岩石特征张量 $X \in \mathbb{R}^{C \times H \times W}$ ，该模块首先将其分别传递到每个注意力分支，以捕获通道维度 C、空间维度 H 和 W 之间的跨维度交互信息。

在第一个分支中，三元组注意力模块构建了空间高度 (H) 维度和通道 (C) 维度之间的交互。为实现这一点，输入 X 首先沿 H 轴逆时针旋转 90 度，以获得旋转后的张量 $X_H \in \mathbb{R}^{W \times H \times C}$ 。然后， X_H 将通过一个 C - 池化层、一个核大小为 7×7 的卷积层和一个批量归一化层，被压缩成大小为 $1 \times H \times C$ 的二维张量。接下来，大小为 $1 \times H \times C$ 的张量通过 Sigmoid 激活函数，生成最终的注意力权重。随后，将生成的注意力权重应用于 X_H ，即对具有相同宽度的特征参数进行加权，生成加权特征图。最后，加权特征图

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.

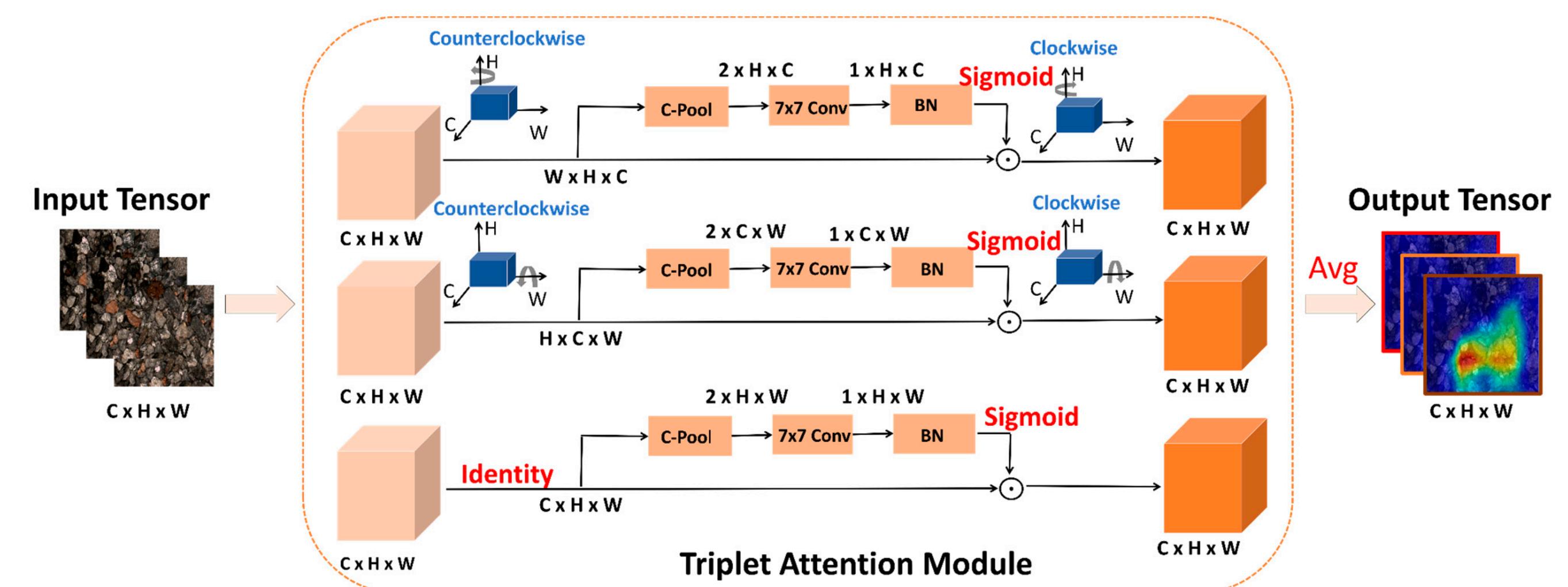


Figure 7. Structure of Triplet attention module.

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 $X \in \mathbb{R}^{C \times H \times W}$, 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 counter-clockwise along the H axis to obtain a rotated tensor $X_H \in \mathbb{R}^{W \times H \times C}$. 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

沿轴 H 顺时针旋转 90 度，以输出与输入 X 形状相同的 X_{H+} 。此分支的计算过程可以用以下方程表示：

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

其中， R^{H+} 表示沿 H 轴顺时针旋转 90 度， σ 表示 Sigmoid 激活函数，ConvBN 表示卷积和批量归一化的组合操作，C-pool 表示复合池化。

复合池化层的计算过程如图 8 所示。在该层中，输入张量沿通道维度同时进行最大池化和平均池化操作，并通过拼接将得到的特征组合起来。

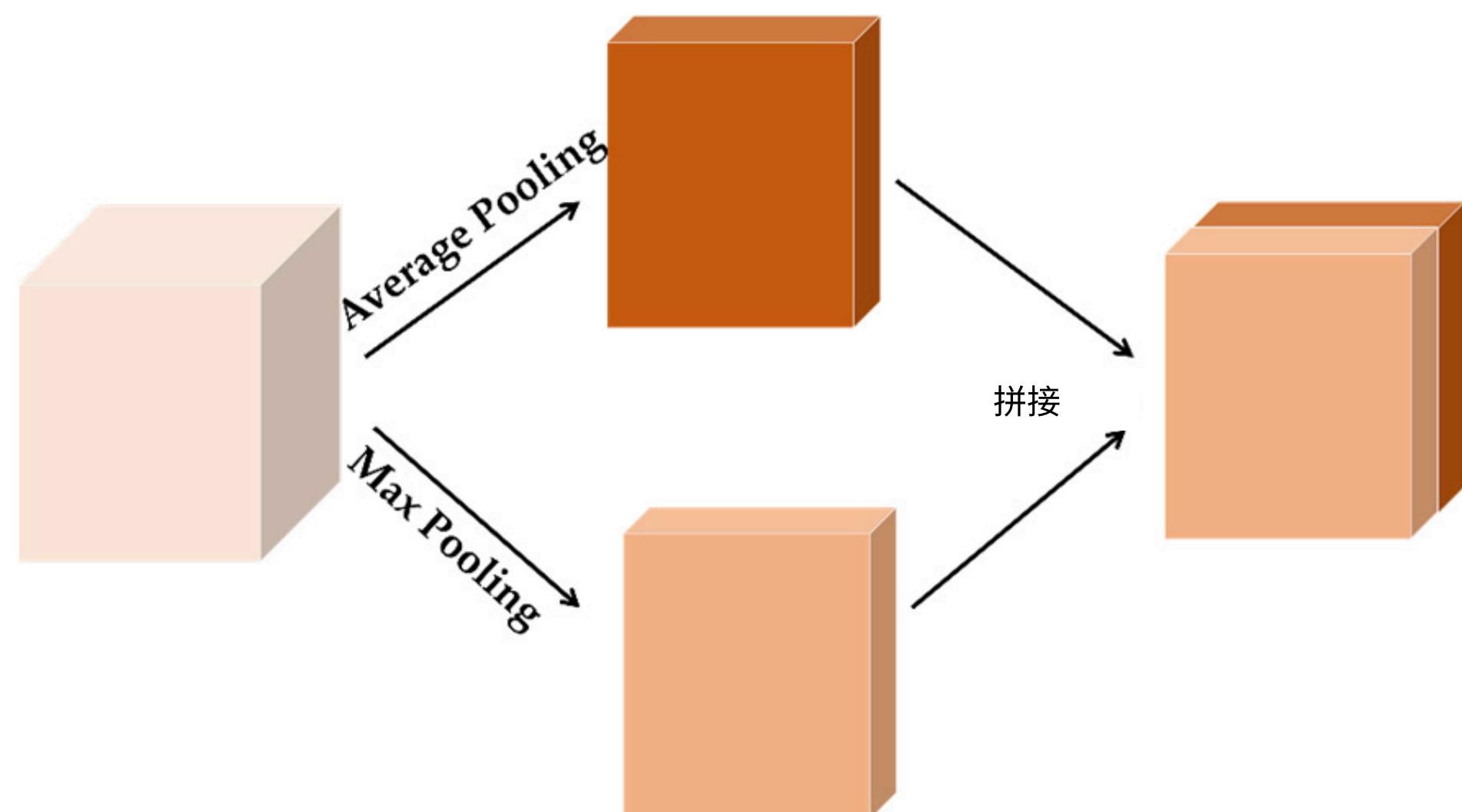


图 8. 复合池化层的结构。

同样，在第二个分支中，输入 X 首先沿 W 轴逆时针旋转 90 度，以获得旋转后的张量 $X_W \in R^{H \times C \times W}$ 。然后，将 X_W 输入到注意力分支中以生成加权特征图。最后，加权图沿 W 轴顺时针旋转 90 度。该分支的计算过程可以用以下方程表示：

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

其中， R^{W+} 表示沿 W 轴顺时针旋转 90 度。

在最后一个分支中，不对输入张量 X 进行旋转，而是直接对同一通道中的特征参数进行加权并生成加权特征图。计算过程如公式 (5) 所示。

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

在计算完每个分支后，如公式(6)所示，三元组注意力模块将使用简单平均的方法聚合每个分支生成的精炼张量，以实现通道注意力和空间注意力信息的融合。

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

3.4 基于三元组高效网络的分类模型

在本研究中，我们提出了三元组高效网络 (Triplet - EfficientNet)，它是高效网络B7 (EfficientNet - B7) 的改进版本，融入了三元组注意力机制，并建立了基于三元组高效网络的岩石分类模型。分类模型如图9所示，其骨干网络是三元组高效网络，与高效网络 B7 类似。不同之处在于，三元组高效网络由三元组移动瓶颈卷积 (Triplet - MBConv) 组成，这是一种改进的移动瓶颈卷积 (MBConv) 模块，

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:

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

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.

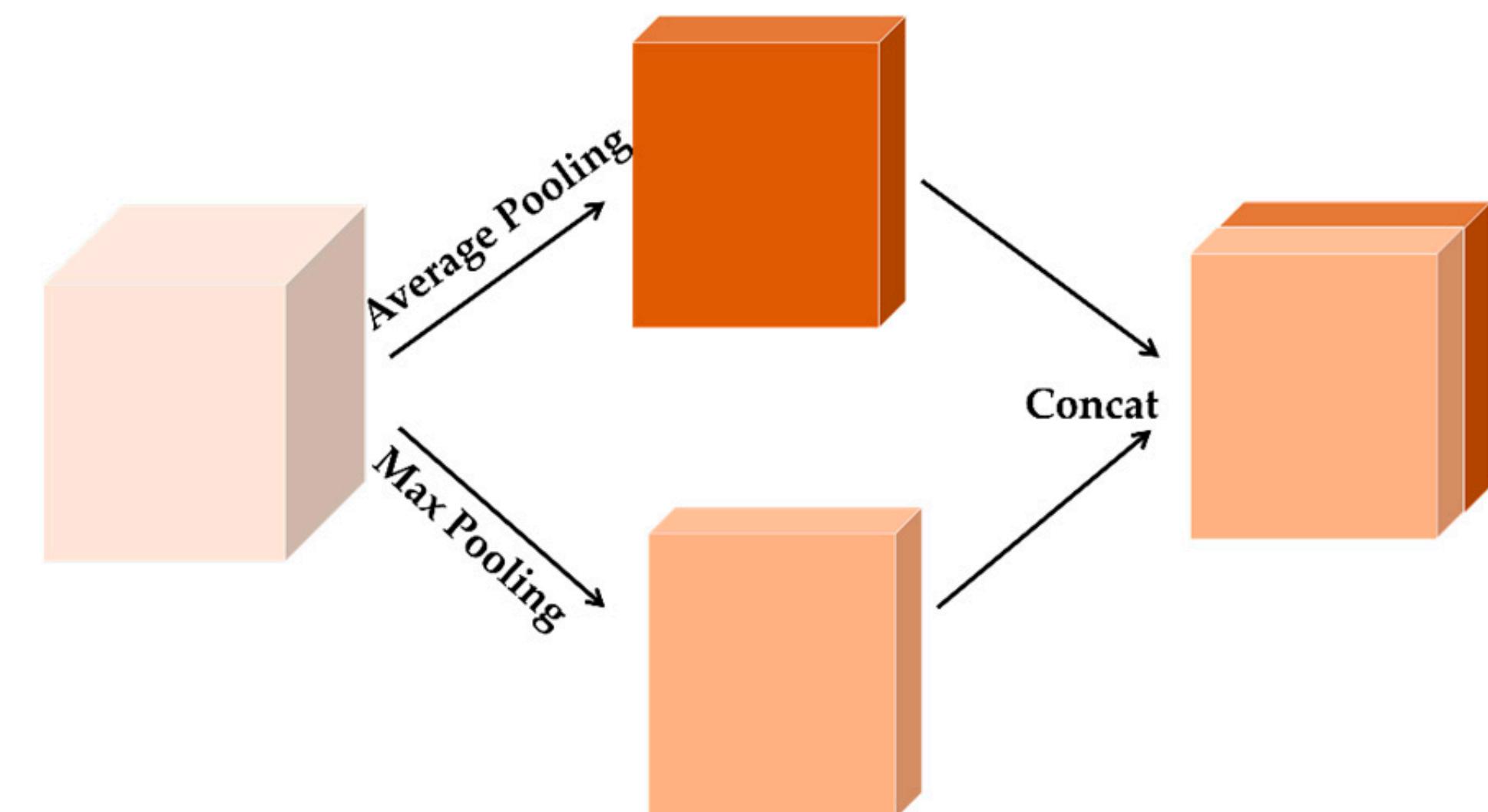


Figure 8. Structure of compound pooling layer.

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:

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

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).

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

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.

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

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

更强的特征表征能力。三元组MBConv模块如图10所示，它是通过将原始MBConv模块中的SE模块替换为三元组注意力模块改进而来的。基于三元组EfficientNet的分类模型不仅可以捕捉网络通道之间的长期依赖关系，还能保留精确的位置信息，从而进一步提高岩石图像分类的准确性。

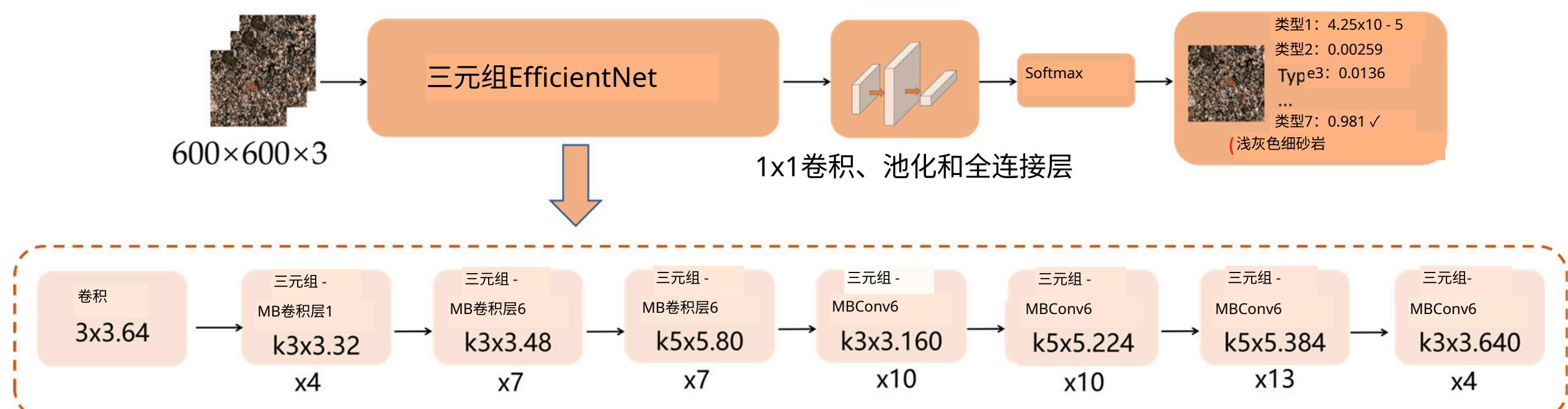


图9. 基于三元组高效网络的分类模型结构。

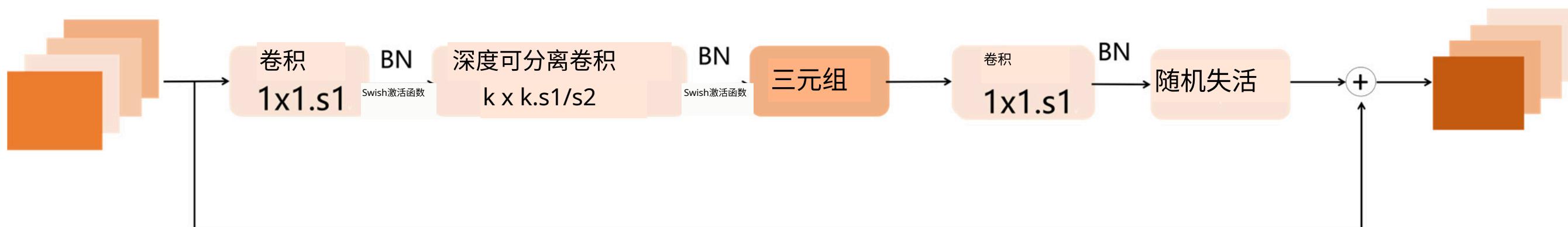


图10. 三元组MBConv模块的结构。

使用基于三元组EfficientNet的分类模型进行岩石类型分类的过程如下：首先，使用数据增强方法对输入的岩石图像进行预处理，并将其转换为大小为 $600 \times 600 \times 3$ 的图像，作为分类模型的输入。然后，核大小为 3×3 的第一个卷积层将对输入图像进行下采样，以实现空间压缩和通道扩展。随后，包含三元组注意力模块的七个阶段的三元组MBConv层将进一步从岩石图像中提取高维特征。然后，通过核大小为 1×1 的卷积层、池化层和全连接层，将高维特征图压缩为二维张量。最后，通过Softmax函数，模型将输出预测概率，该概率表示输入岩石图像属于每种类型的概率值。最大概率对应的类型即为最终分类结果。

3.5. 迁移学习

尽管基于深度学习模型的分类算法克服了传统机器学习方法在特征提取方面的缺点，但它需要足够的图像作为训练输入才能实现高精度。此外，构建一个包含大量不同岩石图像的数据集需要花费大量时间。即使数据集构建完成，从头开始训练一个高精度的深度学习模型也会耗费大量时间和计算资源。为解决上述问题，本研究引入迁移学习方法进行模型训练。

该方法将在现有大规模标注图像数据集上训练好的预训练模型的参数和权重应用于针对类似问题的特定模型，然后对该特定模型进行重新训练和微调以获得最终模型。通过这种方法，我们可以使用更少的岩石图像和更短的训练时间获得更高精度的分类模型。

在本研究中，我们通过在加载预训练权重后训练所有模型参数，将迁移学习方法应用于模型训练，以使最终模型具有更强的特征提取能力。实施步骤如图11所示。

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.

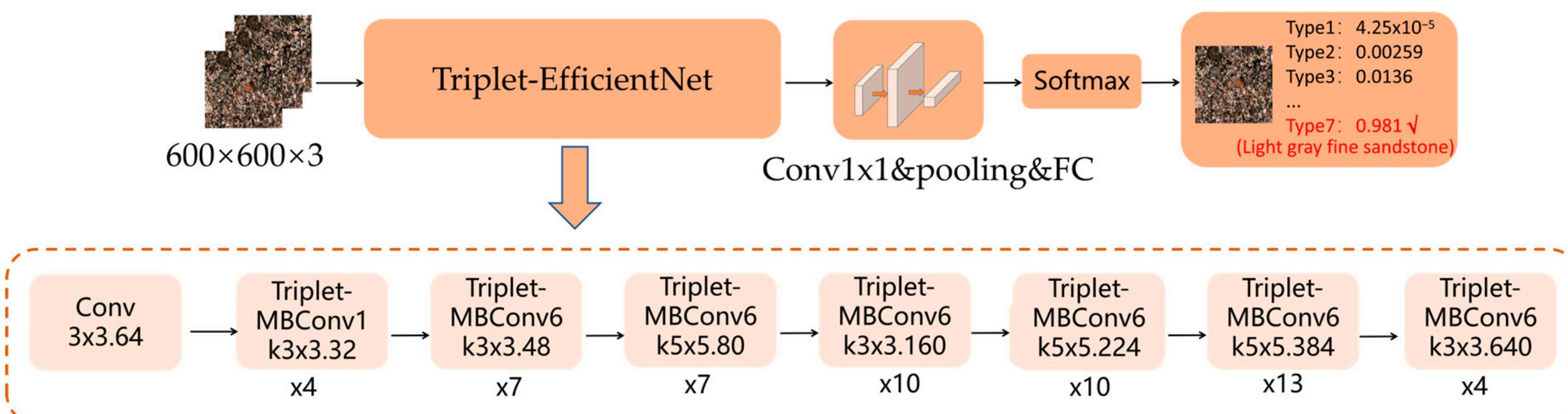


Figure 9. Structure of classification model based on Triplet-EfficientNet.

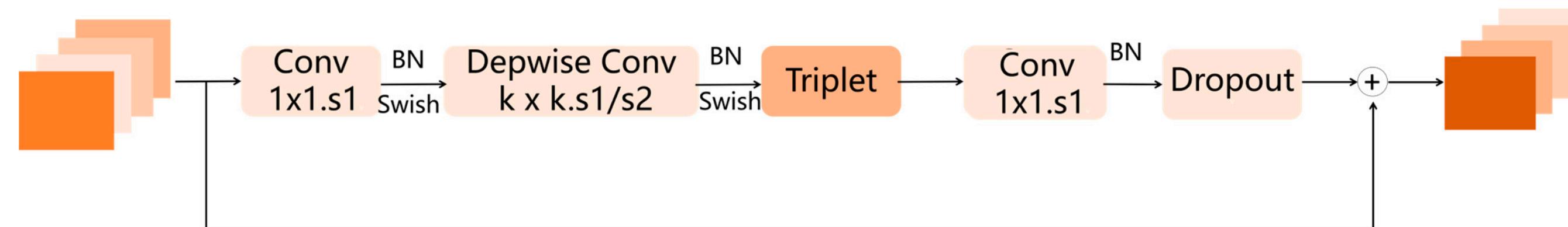


Figure 10. Structure of Triplet-MBConv module.

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,

我们首先在ImageNet [29]上预训练岩石分类模型，该数据集包含27种类型和超过20000张精细分类的图像，模型可以在ImageNet上学习常见的图像信息。然后，我们将预训练模型上的共享参数和权重转移到未训练的模型中。最后，通过使用岩石图像数据集对转移后的模型的所有权重和参数进行重新训练和微调，我们可以得到最终模型。这种方法使我们能够利用从预训练模型中学到的知识，以更少的数据和计算资源显著加速模型收敛并提高整体模型的准确性。

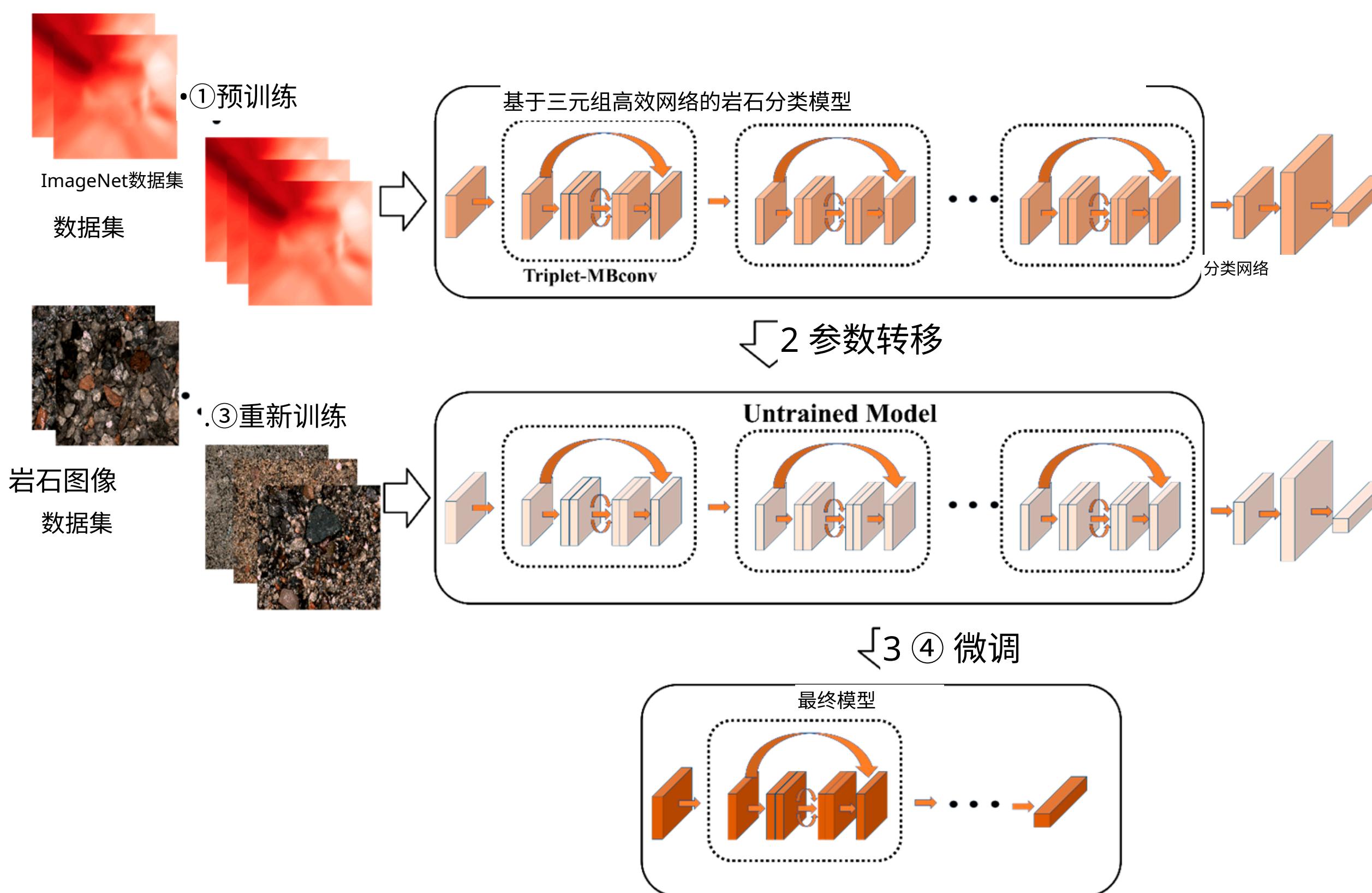


图11. 迁移学习方法示意图。

4. 实验与结果

4.1. 数据预处理

由于岩石图像数据集的样本数量过少，且各类别之间的数量分布不均衡，使该数据集训练的模型准确率较低，且存在过拟合风险[30]。为了解决上述问题，使训练的模型具有更高的分类准确率和更强的泛化能力，需要增加数据集中的样本数量。此外，为了满足模型训练、验证和测试的需求，需要将扩充后的数据集按一定比例划分为训练集、验证集和测试集。

4.1.1. 数据增强

在本研究中，我们通过九种数据增强操作增加了数据集中的样本数量，这些操作包括旋转、添加椒盐噪声、变亮、变暗、放大、垂直翻转、水平翻转、添加高斯噪声和平移。数据增强的示意图如图12a所示。最终结果如图12b所示。经过数据增强后，数据集中的样本数量增加到6949个，不同样本的数量基本均衡。

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.

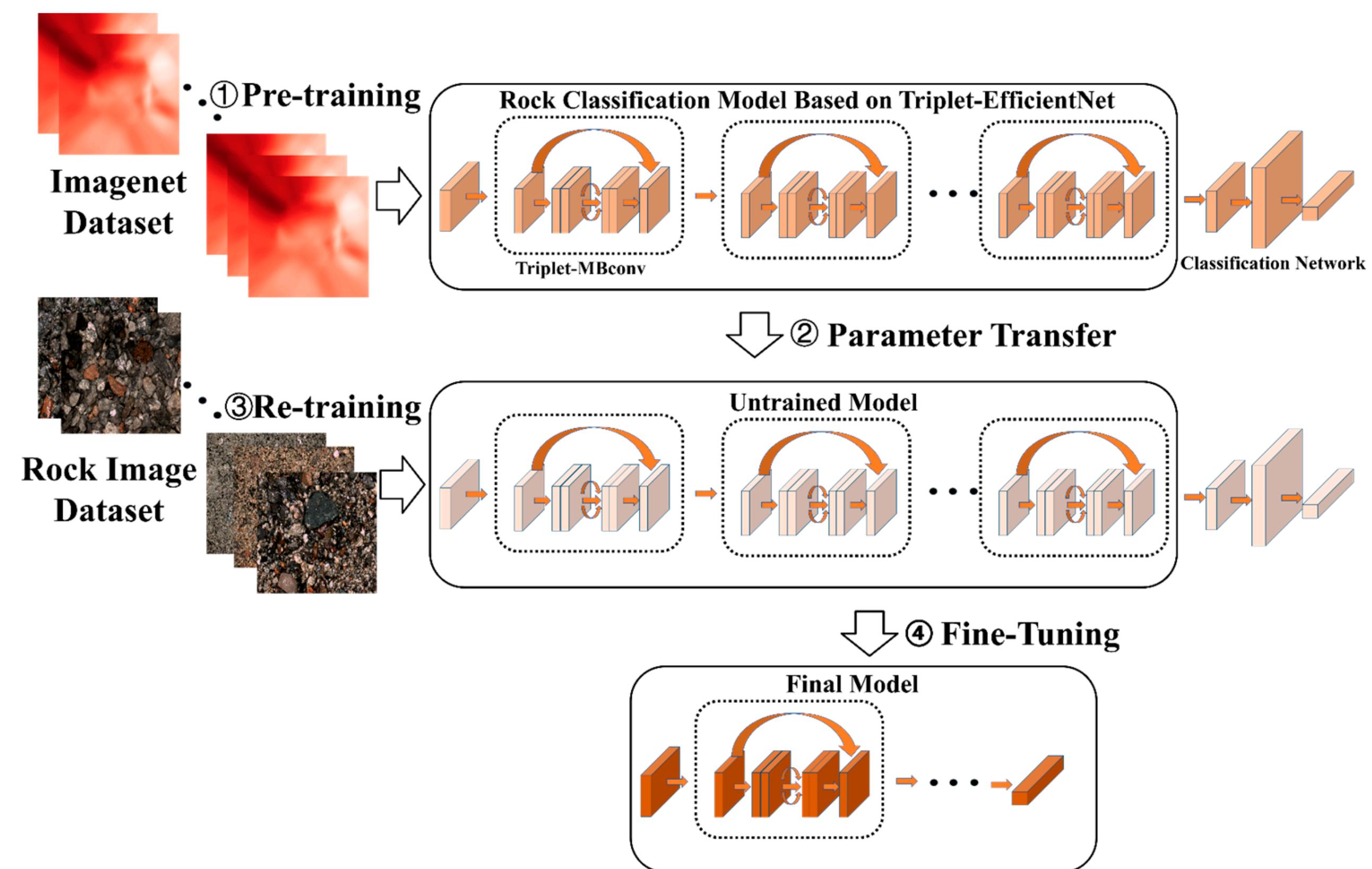


Figure 11. Schematic diagram of the transfer learning method.

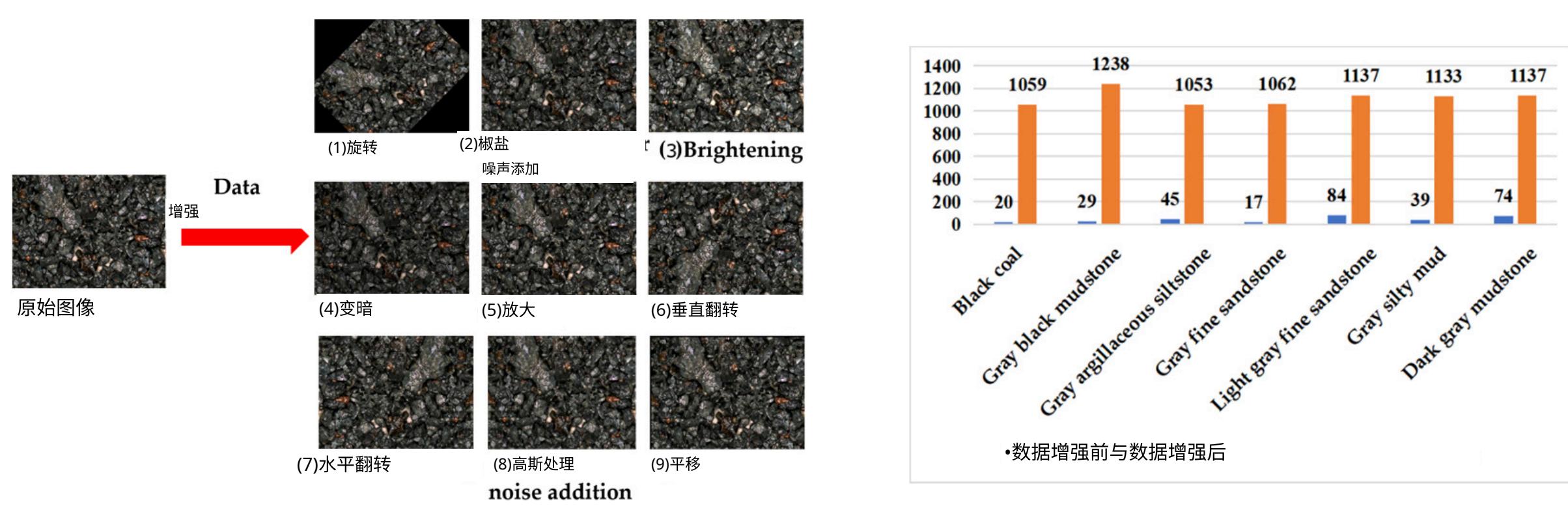
4. Experiments and Results

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.



(a)

(b)

Figure 12. Data augmentation. (a) Data augmentation methods; (b) Data augmentation result.

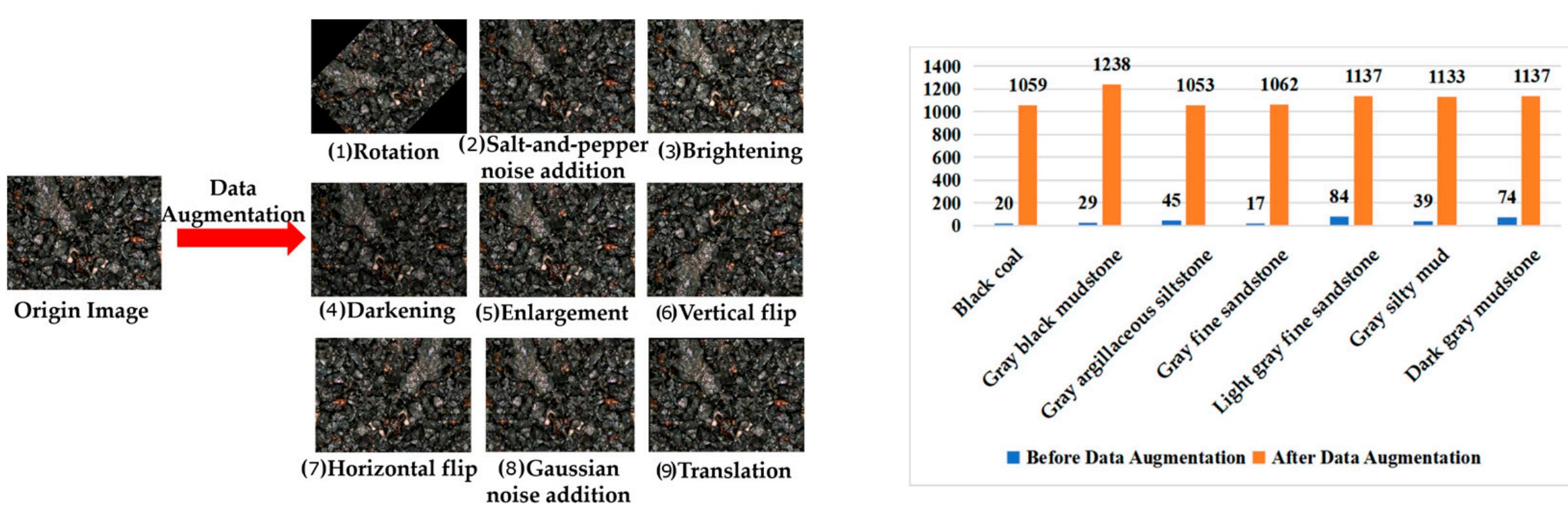
旋转是一种将图像围绕其中心按指定角度旋转的变换；椒盐噪声是一种图像噪声，其特征是随机出现的白色和黑色像素。在为图像中的每个像素添加噪声的过程中，如果未随机选择该像素，则其在新图像中的值将与原始图像中的值相同。如果随机选择了该像素，则其在新图像中的值将为0或255，颜色将为黑色或白色。亮度增强是一种增加图像亮度的变换。亮度降低是一种降低图像亮度的变换。放大是一种增大图像尺寸的变换。垂直翻转是一种将图像垂直翻转的变换。水平翻转是一种将图像水平翻转的变换。平移是一种将图像在x轴和y轴上按指定量移动的变换。此外，高斯噪声是一种图像噪声，其特征是随机值遵循高斯分布。添加高斯噪声会增加图像的复杂度，使其更接近实际岩石图像中存在的噪声。这有助于更好地模拟现实场景，并提高模型对噪声的鲁棒性。在岩石分类中，这种方法使模型能够更好地适应不同环境，从而提高识别准确率。为图像添加高斯噪声的公式为：

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

其中， I 是原始图像， I' 是增亮后的图像， N 是高斯噪声函数， μ 是均值， σ^2 是方差。在本研究中， μ 设为 0， σ^2 设为 0.01。

4.1.2. 数据集划分

在深度学习中，数据集通常会划分为训练集、验证集和测试集，以满足模型训练、验证和测试的需求。训练集是用于训练模型的数据集合。通过从训练集中学习，模型可以学习到数据的特征和模式，从而提高预测的准确性。训练集包含输入数据和相应的标签，模型会不断更新其权重以提高预测精度。训练集是深度学习模型训练的基础，对模型的性能有着至关重要的影响，因此训练集占用的数据量最大。验证集是用于评估模型性能的数据集。验证集可以帮助我们检测模型的过拟合程度，并及时调整模型的超参数。同时，它还可以帮助我们选择最佳的模型参数，如轮数、学习率和批量大小。测试集是用于评估模型最终性能的数据集。在完成训练和验证过程后，我们通常使用测试集来评估模型的最终性能。在本研究中，岩石数据集也被随机划分为训练集、验证集和



(a)

(b)

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:

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

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

测试集的比例分别为 60%，20% 和 20%。因此，它们的样本数量分别为 4169、1389 和 1389。

4.2. 实验细节

该模型在装有 Windows10 操作系统的高性能工作站上进行训练和测试，该工作站配置了 2.10GHz Intel Xeon Silver 4110 CPU (16GB 内存) 和 NVIDIA GeForce RTX 2080 Ti GPU。软件环境如下：基于 64 位的 Windows 10 操作系统、Pytorch 深度学习框架、CUDA11.0、OpenCV2 库和 VS Code 集成开发环境。

在网络训练期间，学习率设置为 0.01，轮数设置为 60，批量大小设置为 16。选择 Swish 函数作为激活函数，选择交叉熵函数作为损失函数，选择自适应矩估计 (Adam) 作为优化器。

在模型训练前，通过迁移学习方法将预训练模型的参数加载到分类模型中。然后将训练集中的图像统一缩放到 $600 \times 600 \times 3$ 大小，并随机打包到模型中开始训练。在训练过程中，我们随机选取 4169 张图像进行训练，每张图像会被使用多次。每迭代一次对训练集和验证集进行一次评估，并将每一代的训练准确率、验证准确率和交叉熵损失变化过程保存为日志文件，然后上传到 TensorBoard 进行查看。

4.3. 评估指标

准确率和损失值是图像分类中最常用的两个评估指标。准确率表示所有样本中分类正确的样本所占的比例，是最直接反映分类模型性能的评估指标。其计算公式如式(8)所示：

$$\text{Accuracy} = \frac{t}{T} \quad (8)$$

其中， t 为分类正确的样本数量， T 为样本总数。

在本研究中，使用交叉熵损失函数来定量评估预测值与真实值之间的差异。通过计算损失函数，更新了我们模型的参数。其公式如式(9)所示：

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

其中 M 是类型总数。 y_{ic} 是指示变量，如果类型与样本类型相同，则 i, y_{ic} 为 1，否则为 0。 p_{ic} 是样本 i 属于类型 c 的预测概率。

4.4. 结果分析

4.4.1. 数据增强的有效性

在本研究中，我们试图通过一些数据增强方法扩展原始数据集，以提高分类模型的性能。为评估这一预处理步骤的有效性，我们进行了一系列消融实验。基于 Triple-EfficientNet 的分类模型被用于在扩展前后的训练集和测试集上进行实验。从表 2 的结果可以看出，应用数据增强使训练集和测试集的分类准确率显著提高。具体而言，训练集的准确率提高了 31.4%，而测试集的准确率提高了 22.4%。这些结果表明，通过数据增强对小样本数据集进行预处理可以提高网络提取更全面岩石特征的能力，从而增强模型的整体泛化能力。

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):

$$\text{Accuracy} = \frac{t}{T} \quad (8)$$

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):

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

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.

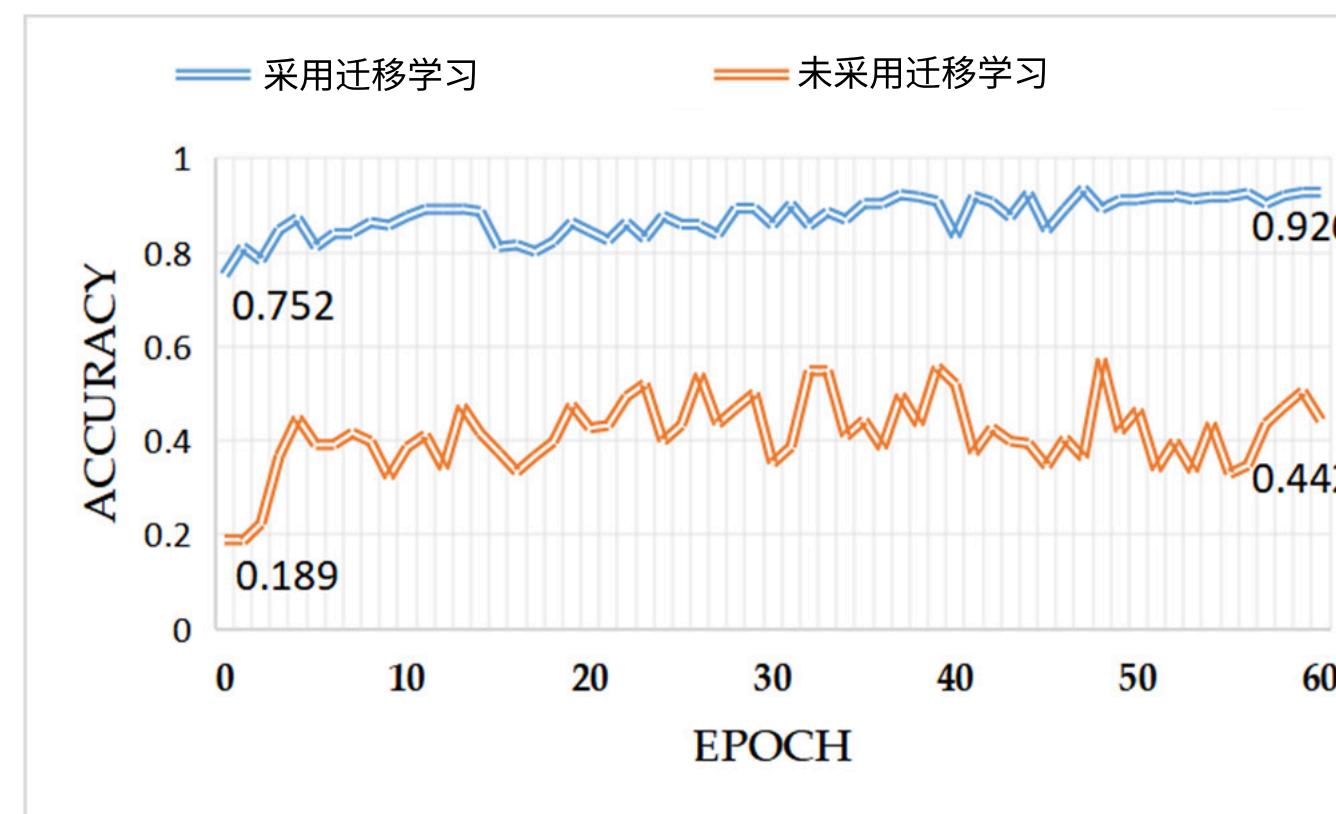
表2. 数据增强方法的消融实验结果。

方法	岩石数据集中的图像数量	训练集的最终准确率 (轮数 = 60)	测试集的Top - 1准确率
三元组高效网络+原始数据集	315	61.2%	70.8%
三元组高效网络+增强数据集	6949	92.6%	93.2%

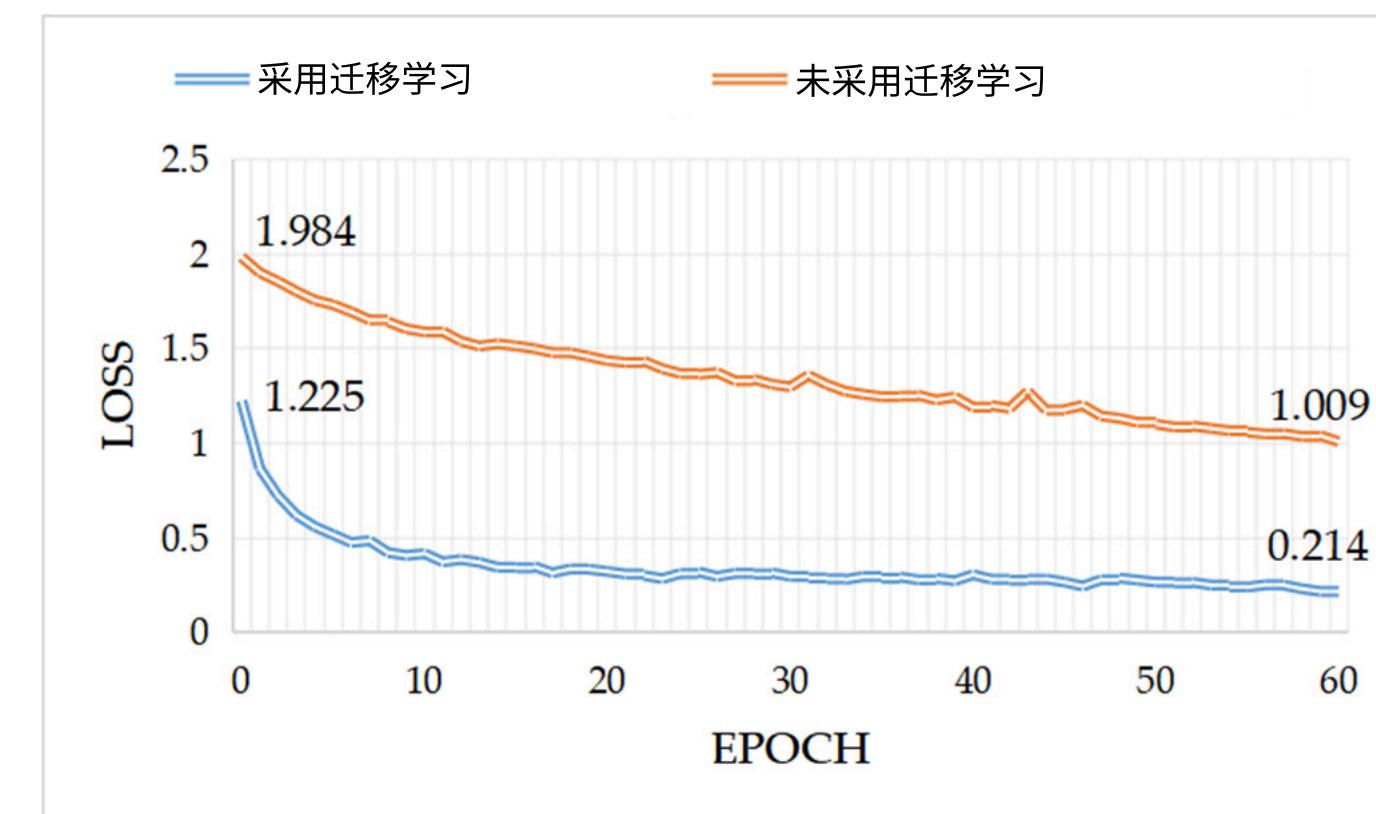
4.4.2. 迁移学习的有效性

为了评估采用迁移学习的训练策略对网络训练收敛性的影响，我们分别使用迁移学习方法训练了一个基于三元组高效网络的分类模型，并使用非迁移学习方法训练了另一个相同的模型。除训练策略不同外，这两个模型的其他超参数设置均相同。

训练过程中准确率和损失值的变化分别如图13a、b所示。从图13a可以看出，采用迁移学习方法的模型初始准确率最高可达75.2%，经过60个训练周期后最终准确率达到92.6%。然而，未采用迁移学习方法的模型初始准确率仅为前者的四分之一，最终准确率仅为44.2%。从图13b可以看出，采用迁移学习方法的模型初始损失值仅为1.225，然后在60个训练周期内迅速收敛，最终损失值收敛到0.214。然而，未采用迁移学习方法的模型整体损失值大于1，且收敛速度远低于前者。



(a)



(b)

图13. 采用迁移学习的模型与未采用迁移学习的模型的性能比较。(a) 准确率; (b) 损失。

通过比较同一模型在不同训练策略下的准确率和损失值变化，可以看出迁移学习方法极大地加快了分类模型的收敛过程，并提高了模型在训练过程中的整体准确率。这是因为迁移学习方法将在大规模图像数据集中获得的用于图像特征提取的通用模型参数信息预加载到训练好的模型中。因此，迁移学习方法在训练开始时就赋予了模型强大的图像特征表征能力，从而提高了性能并加快了训练收敛速度。

4.4.3. 模型训练性能评估

为了评估基于三元组高效网络 (Triplet - EfficientNet) 的分类模型的训练收敛性能，我们训练了五个模型进行对比研究。这些模型分别是高效网络B7 (EfficientNet - B7)、VGG16、谷歌网络 (GoogleNet)、亚历克斯网络 (AlexNet) [31]和三元组 -

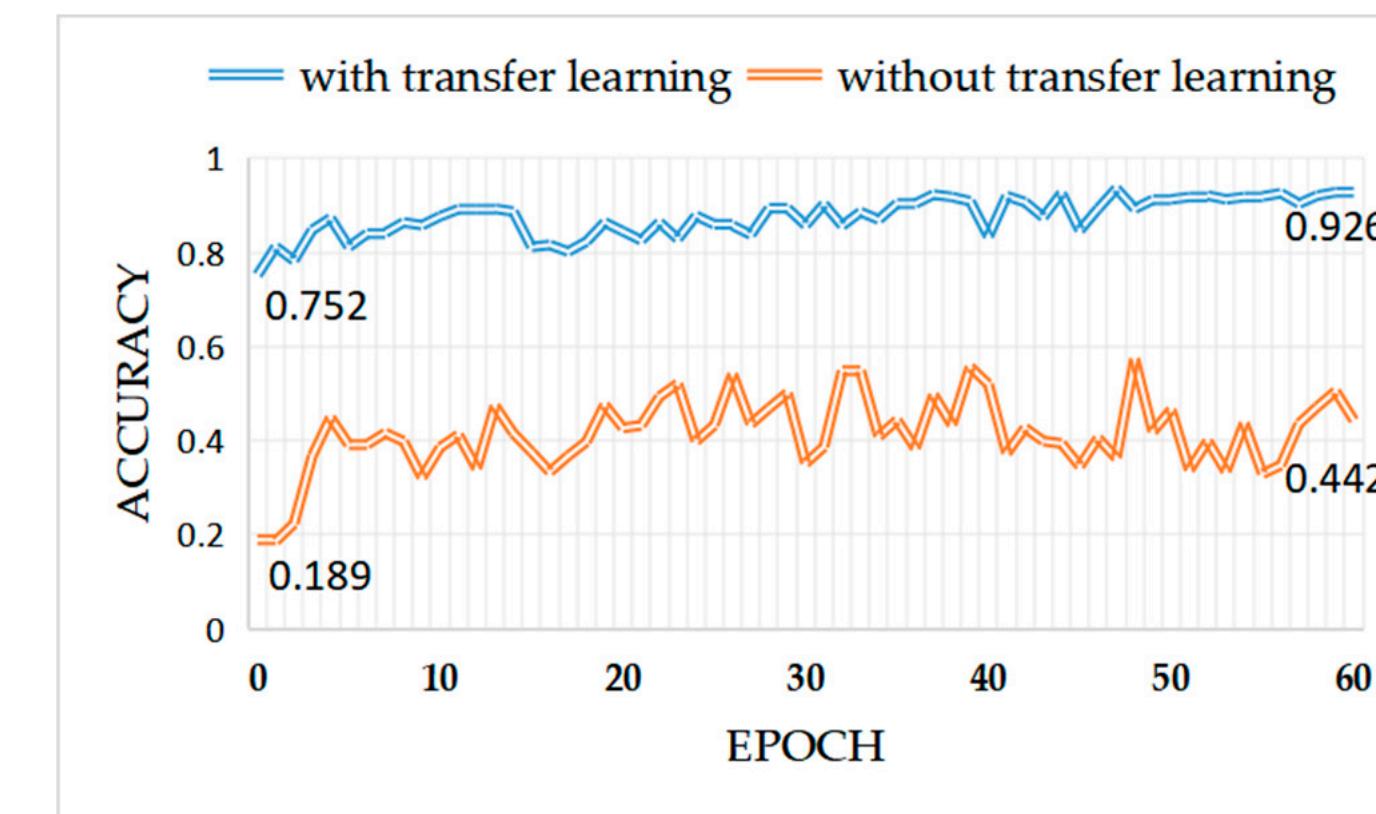
Table 2. The results of the ablation experiments on the data augmentation methods.

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%

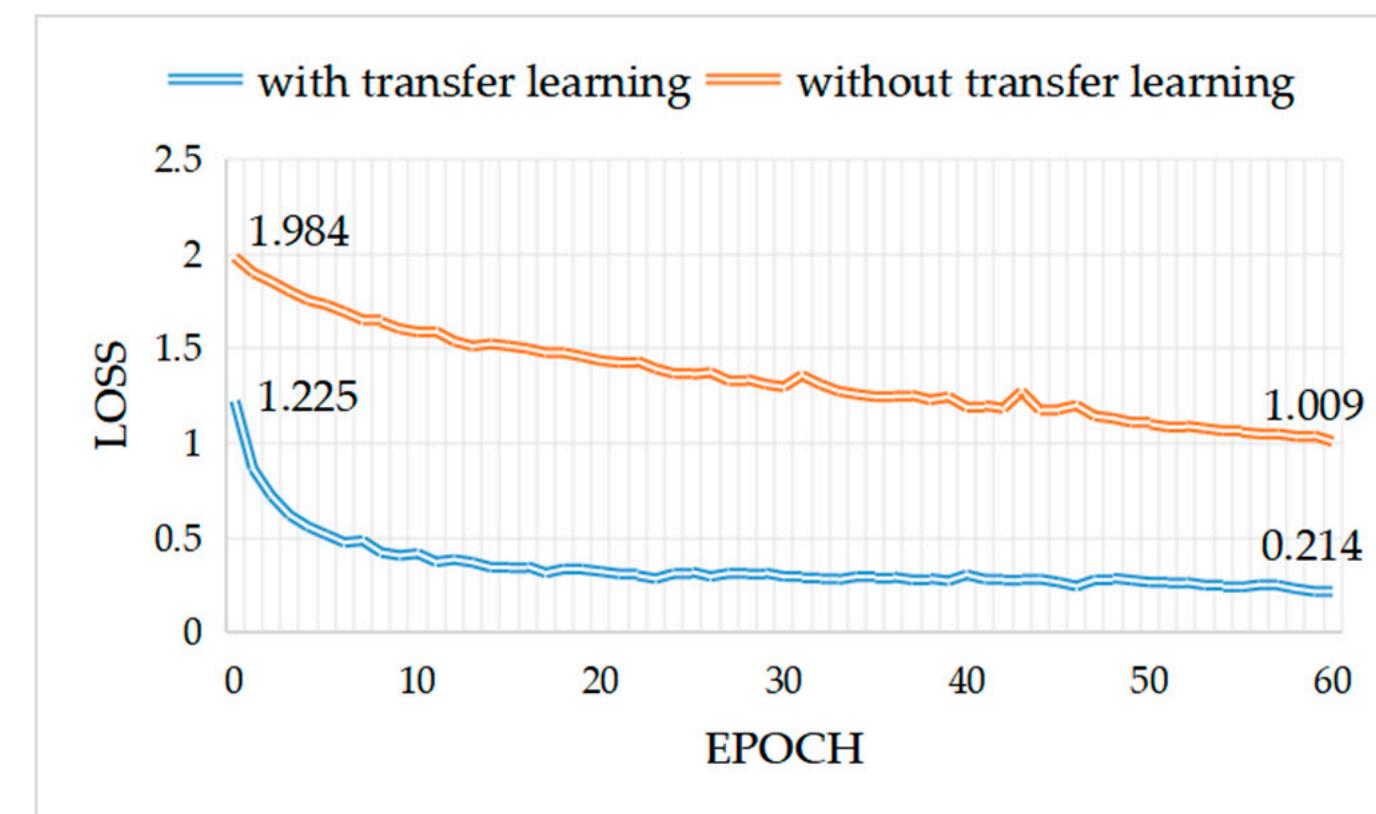
4.4.2. The Effectiveness of Transfer Learning

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.

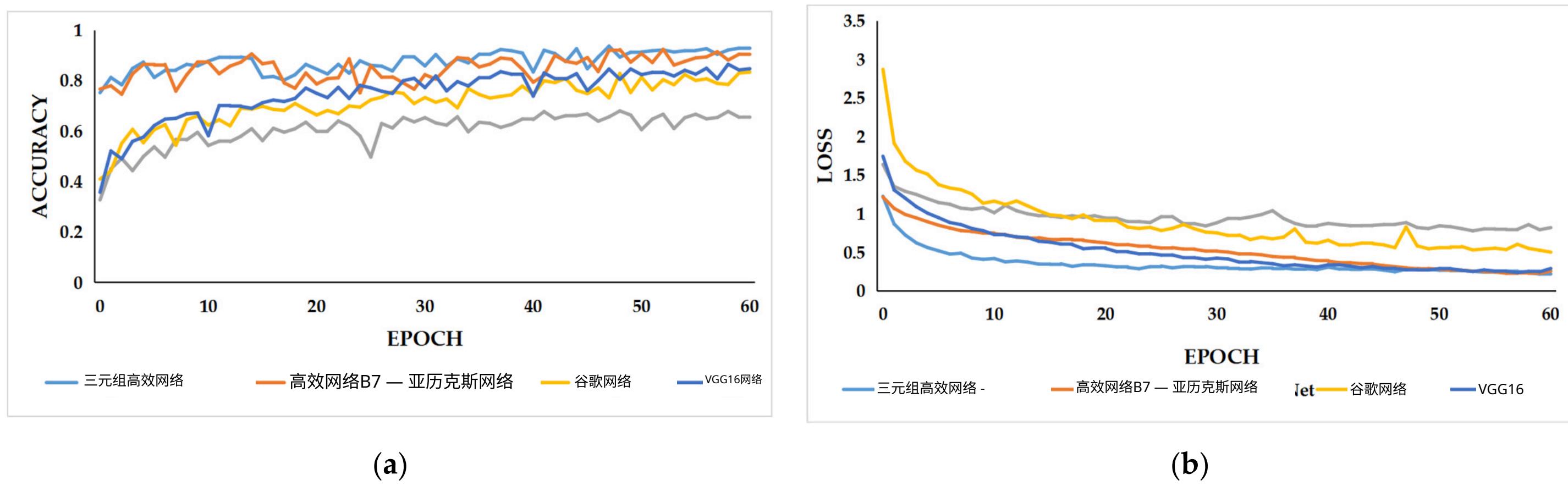
By comparing the changes in accuracy and loss value for the same model with different training strategies, it can be seen that the transfer learning method greatly speeds up the convergence process of the classification model and improves the overall accuracy of the model in the training process. This 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-

高效网络 (EfficientNet)。每个模型训练时使用的训练策略和超参数保持一致。

训练过程中准确率和损失值的变化分别如图14a、b所示。从图14a可以看出，五个模型在60个训练周期内训练集的平均准确率从高到低依次为：三元组高效网络 (Triplet - EfficientNet)、高效网络B7 (EfficientNet - B7)、VGG16、谷歌网络 (GoogleNet) 和亚历克斯网络 (AlexNet)。如图14b所示，损失值的排名则相反。下表3展示了每个模型的具体准确率和损失值。从表中可以看出，本研究提出的基于三元组高效网络的分类模型在训练集中与高效网络B7和其他模型相比，表现出更高的准确率和更低的损失值。结果表明，基于高效网络的复合模型缩放方法通过神经架构搜索 (NAS) 技术获得了更好的网络结构和更合理的参数配置，而引入三元组注意力机制进一步提高了模型的整体准确率和训练收敛性能。



(a)

(b)

图14. 不同模型在训练过程中准确率和损失值的变化。(a) 准确率; (b) 损失值。

表3. 不同模型在训练集上的准确率和损失值。

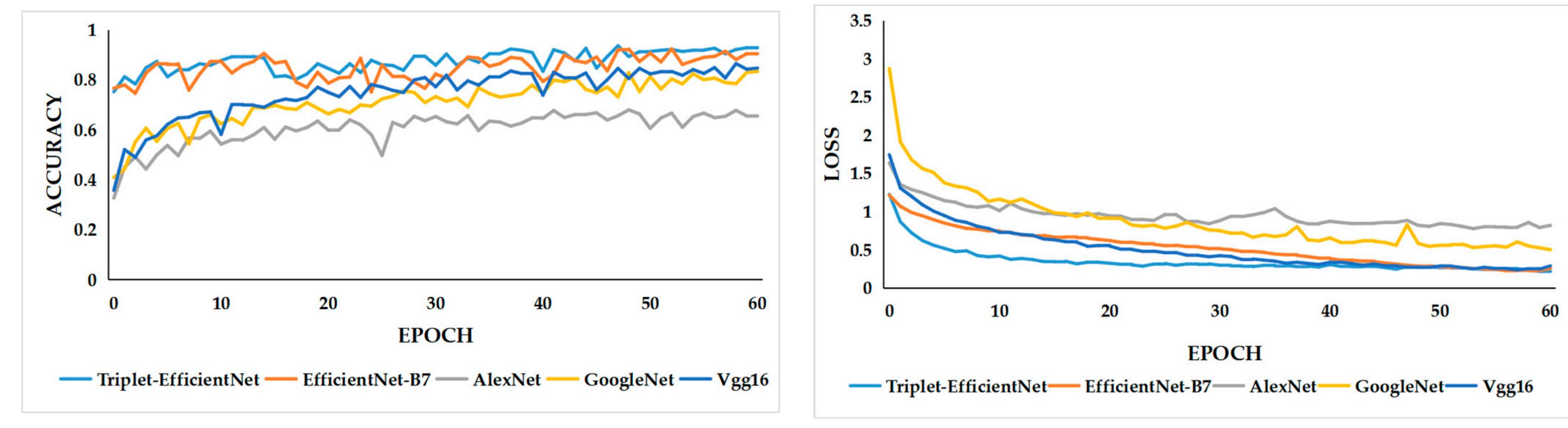
模型	初始准确率	初始损失	最终准确率 (轮次 = 60)	最终损失 (轮次 = 60)
亚历克斯网络	32.7%	1.638	65.5%	0.818
谷歌网络	40.9%	2.870	83.3%	0.501
VGG16	35.7%	1.744	84.6%	0.289
高效网络 - B7 (+SE注意力机制)	75.2%	1.225	90.4%	0.255
三元组高效网络 (+三元组注意力机制)	76.6%	1.215	92.6%	0.214

4.4.4. 不同模型的性能比较

在本研究中，我们旨在评估基于三元组高效网络的分类模型的鲁棒性和泛化能力，并通过将其性能与其他模型进行比较来证明我们所提出方法的优越性。我们不仅引入了AlexNet、GoogleNet和VGG16等主流分类模型，还复现了近期类似的图像分类方法，包括Chen等人[15]提出的深度残差网络模型 (ResNet34) 和Gan等人[32]提出的CA - 高效网络。CA - 高效网络是一种通过引入坐标注意力机制改进的高效网络模型。包括三元组高效网络模型在内的七种模型

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

本文提出的方法，使用一致的训练集、训练策略和输入图像大小进行了测试。

在性能测试中，我们在测试集上使用了七个训练好的模型进行图像推理，并采用广泛认可的Top-1准确率指标来评估推理结果。Top-1准确率是指预测概率最高的类别与实际结果相符的准确率[33]。如表4所示，对于相同大小的输入图像，由于EfficientNet高效的网络结构，EfficientNet-B7、CA-EfficientNet和Triplet-EfficientNet模型在Top-1准确率方面大大优于其他模型。其中，CA-EfficientNet和Triplet-EfficientNet模型分别受益于坐标注意力机制和三元组注意力机制的引入，表现出更好的性能。与EfficientNet-B7相比，它们具有更强的空间特征表征能力，能够在岩石图像中获取更有效的特征信息，从而进一步提高其Top-1准确率。通过进一步比较发现，Triplet-EfficientNet的三元组注意力机制在模型性能方面优于CA-EfficientNet的坐标注意力机制。这是因为与坐标注意力机制相比，三元组注意力机制有更多的注意力分支，使模型能够更全面地跨图像维度提取特征信息。因此，引入了三元组注意力机制的Triplet-EfficientNet具有最高的Top-1准确率。

表4. 测试集中不同网络模型的性能比较。

模型	输入图像尺寸	Top-1 准确率	参数	浮点运算次数
亚历克斯网络 [31]	600 × 600	71.9%	61 MB	5 G
谷歌网络 [18]	600 × 600	80.6%	13 MB	10 G
VGG16网络 [19]	600 × 600	88.1%	138 MB	110 G
残差网络34 [15]	600 × 600	86.3%	36 MB	26 G
高效网络-B7 [26] (+SE注意力机制)	600 × 600	92.0%	66 MB	38 G
CA-EfficientNet [32] (+坐标注意力机制)	600 × 600	92.6%	67 MB	39 G
三元组- EfficientNet (+三 元组注意力机制)	600 × 600	93.2%	64 MB	36 G

除了测试模型的性能外，我们还使用了参数和浮点运算次数 (FLOPs) 这两个指标来计算每个模型的计算复杂度。参数是指模型训练期间需要训练的参数总数，用于衡量模型的计算空间复杂度。FLOPs (浮点运算) 是指神经网络模型中需要执行的浮点运算次数，用于衡量模型的计算时间复杂度。如表4所示，由于EfficientNet网络结构及其参数的高效性，EfficientNet - B7、CA-EfficientNet和三元组 - EfficientNet在参数和FLOPs适中的情况下实现了更高的准确率。此外，与CA - EfficientNet的坐标注意力机制和EfficientNet - B7的SE注意力机制相比，三元组 - EfficientNet的三元组注意力机制效率更高。它不仅降低了原始EfficientNet模型的计算复杂度，还进一步提高了模型的性能。

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. 现实检验

为了评估本研究提出的基于三元组高效网络 (Triplet - EfficientNet) 的分类模型的预测效果，从岩石数据集中随机选取七张不同类型的岩石图像进行预测，并输出每张图像中七种岩石类型标签的预测概率，如图15所示。预测概率最大值对应的岩石类型标签即为最终分类结果。结果表明，该模型对各类选取图像的分类效果达到了较高水平，准确率超过95%，且对黑煤、深灰色粉砂质泥岩和灰色泥质粉砂岩的预测概率接近100%。结果显示，本研究提出的分类模型具有很强的鲁棒性和泛化能力。

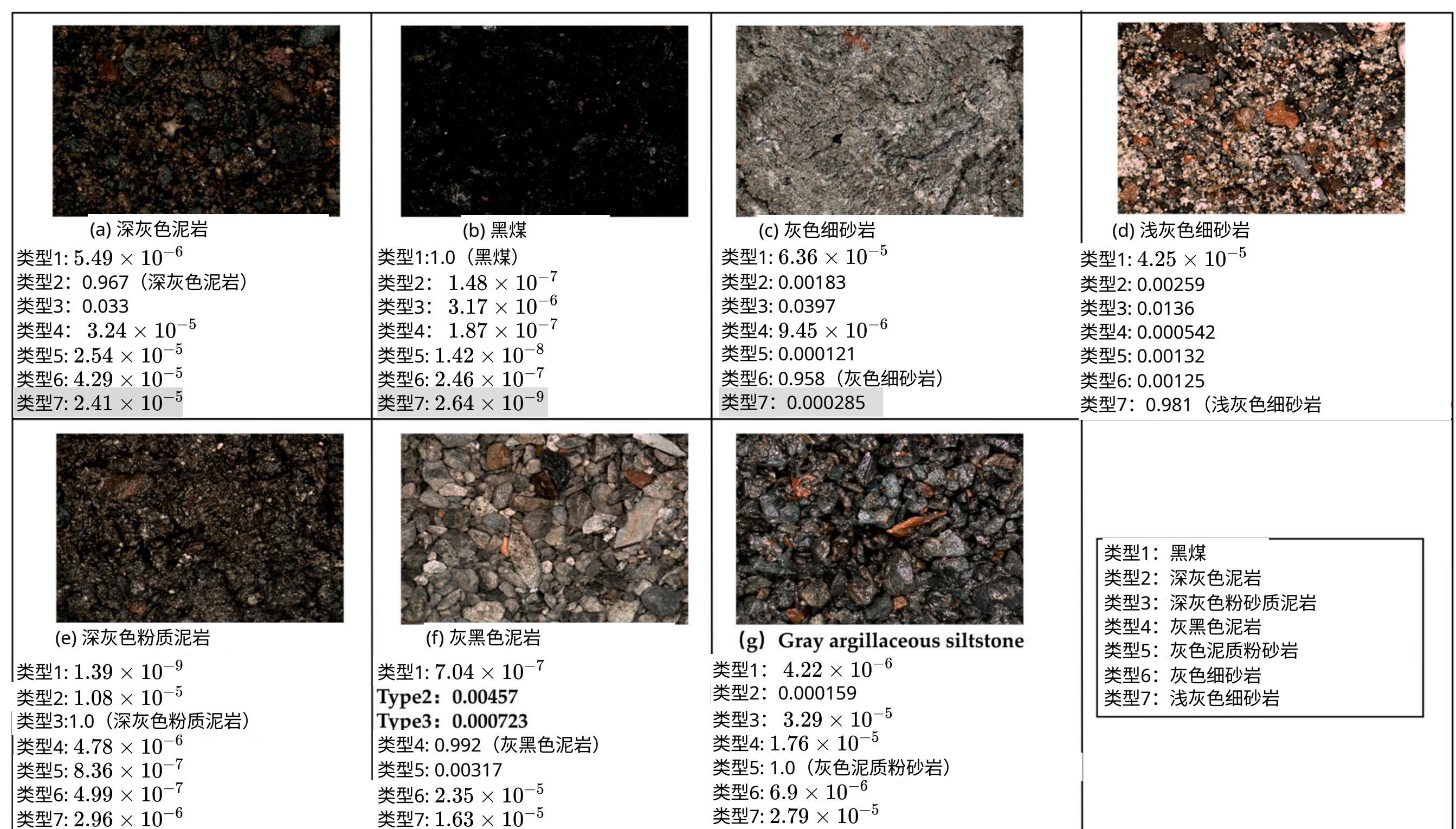


图15. 分类模型对每种岩石类型样本的预测概率。

4.4.6. 综合分析与讨论

在本节中，我们通过一系列实验验证了我们提出的基于EfficientNet和三重注意力机制的岩石图像分类方法的有效性和先进性。

首先，我们进行了消融实验以验证数据增强的有效性。这种预处理方法通过各种图像变换显著增加了数据集的样本量，有效解决了原始数据集中样本量不足和类型间分布不均的问题，使模型能够捕捉到足够的数据模式，从而提高和增强了模型的识别准确率和泛化能力。

接下来，我们通过使用迁移学习训练一个分类模型和另一个不使用迁移学习的相同模型，充分验证了迁移学习在模型训练中的有效性。迁移学习方法在初始训练阶段将预训练权重加载到未训练的模型中，使模型具备强大的图像特征提取能力，并使模型能够用更短的时间和更少的数据样本达到更高的准确率。

然后，我们分别训练了五个模型，包括EfficientNet - B7、VGG16、GoogleNet、AlexNet和Triplet - EfficientNet，以比较它们的模型训练性能，从而

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.

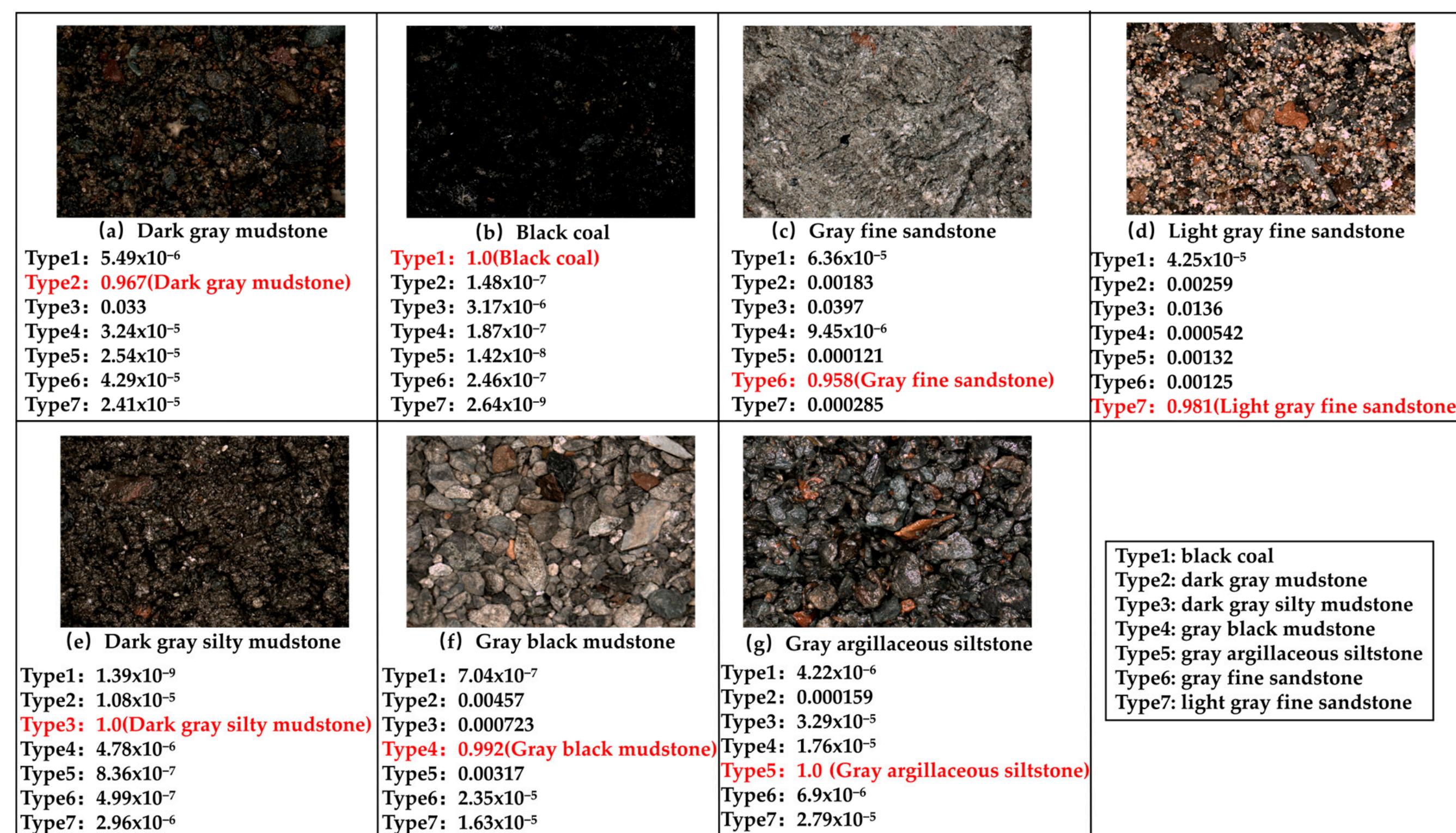


Figure 15. The prediction probability of the classification model for each rock-type sample.

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

验证EfficientNet网络结构和三元组注意力机制在提高模型训练性能方面的有效性。

为进一步验证我们提出的模型的鲁棒性和泛化能力，以及其相对于其他主流和前沿模型的优越性，我们训练了六个模型（AlexNet、GoogleNet、VGG16、ResNet34、EfficientNet - B7、CA - EfficientNet）和我们提出的Triplet - EfficientNet。我们对所有模型应用了相同的训练策略和参数设置。训练完成后，我们对这些模型进行了性能测试和计算复杂度测试。结果表明，EfficientNet网络结构的高效性使EfficientNet - B7、CA - EfficientNet和Triplet - EfficientNet能够以适中的参数数量和浮点运算次数（FLOPs）实现更高的模型准确率。同时，与EfficientNet - B7的SE注意力机制和CA - EfficientNet的坐标注意力机制相比，Triplet - EfficientNet的三元组注意力机制在图像推理方面更高效，使EfficientNet模型在减少参数数量和FLOPs的同时进一步提高准确率。

最后，我们在七张包含各类岩石的图像上评估了我们的岩石图像分类模型的实际性能。结果表明，该模型对所有测试图像实现了准确分类，预测准确率超过95%。这些结果展示了我们的模型在岩石图像分类方面的卓越能力。

5. 结论

在本研究中，提出了一种基于EfficientNet和三重注意力机制的岩石图像分类模型，以实现准确的端到端岩石图像分类。首先，我们通过各种数据增强方法扩充了岩石图像数据集，以防止模型训练过拟合并提高模型性能。在构建模型时，我们以EfficientNet作为基准网络，由于采用了神经架构搜索（NAS）技术和复合模型缩放方法，该网络具有高效的网络结构。在此基础上，引入三重注意力机制对原始的EfficientNet进行改进，增强模型提取岩石图像空间特征的能力。实验结果表明，本研究中的分类模型在训练集和测试集上均优于其他主流模型，准确率分别达到了92.6%和93.2%。在模型训练过程中，我们采用了迁移学习方法来加速模型收敛，显著提高了模型的训练性能。采用迁移学习的模型训练准确率比未采用迁移学习的模型提高了48.4%。

通过进一步研究发现，岩石数据集中的样本数量和岩石类型对模型可识别的岩石类型数量以及最终分类准确率有显著影响。考虑到这一点，我们未来的研究工作将集中在增加岩石类型的多样性和岩石图像的数量上，同时确保随着更多岩石类型的加入，分类准确率能进一步提高。此外，鉴于野外勘探中通过网络及时获取岩石类型识别结果反馈存在困难，我们打算在改进模型的基础上，将其部署到移动设备上。这将使地质调查人员能够在离线条件下方便地使用分类模型识别岩石。

作者贡献：概念构思，Z.H.和L.S.；方法学，Z.H.；软件，Z.H.；验证，Z.H.、J.W.和Y.C.；正式分析，Z.H.；调查，Z.H.；资源，Z.H.；数据整理，Z.H.；撰写——初稿准备，Z.H.；撰写——审核与编辑，Z.H.；可视化，Z.H.；监督，L.S.；项目管理，Z.H.；资金获取，L.S.。所有作者都已阅读并同意该论文的发表版本。

资助信息：本研究由国家自然科学基金（项目编号61903315）和福建省科学技术厅自然科学基金（项目编号2022J1011255）资助。

机构审查委员会声明：不适用。

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.; 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.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 61903315, and the Natural Science Foundation of the Department of Science and Technology of Fujian Province, grant number 2022J1011255.

Institutional Review Board Statement: Not applicable.

知情同意声明: 不适用。

数据可用性声明: 本研究中呈现的数据可应要求从通讯作者处获取。

致谢: 我们感谢所有审稿人的评论，也感谢广东天枢智能科技有限公司提供的岩石图像。

利益冲突: 作者声明无利益冲突。

参考文献

- 傅, G.; 严, J.; 张, K.; 胡, H.; 罗, F. 岩性识别技术的现状与进展。《地球物理学进展》2017, 32, 26 – 40。
- 张, S.; 博古斯, S.M.; 利皮特, C.D.; 卡马特, V.; 李, S. 为建筑研究实施遥感方法: 从无人机系统的视角。《建设工程与管理杂志》2022年, 148卷, 3122005期。[CrossRef]
- 郭, Q.; 周, Y.; 曹, S.; 邱, Z.; 徐, Z.; 张, Y. 广宁玉的矿物学研究。《中山大学学报(自然科学版)》2010年, 49卷, 146 - 151页。
- 姆林纳丘克, M.; 戈尔什奇克, A.; 斯利佩克, B. 模式识别在微观岩石图像自动分类中的应用。《计算机与地球科学》2013年, 60卷, 126 - 133页。[CrossRef]
- 肖, F.; 陈, J.; 侯, W.; 王, Z. 欧杭成矿带南段旁西峒地区与银 - 金成矿相关的地球化学异常识别与提取。《岩石学报》2017年, 33卷, 779 - 790页。
- 徐, Z.; 马, W.; 林, P.; 施, H.; 刘, T.; 潘, D. 基于岩石图像迁移学习的智能岩性识别。《基础科学与工程学报》2021年, 29卷, 1075 - 1092页。
- 利皮特, C.D.; 张, S. 小型无人航空平台对被动光学遥感的影响: 概念视角。《国际遥感学报》2018年, 39卷, 4852 - 4868页。[CrossRef]
- 马尔莫 (Marmo), R.; 阿莫迪奥 (Amodio), S.; 塔利亚费里 (Tagliaferri), R.; 费雷里 (Ferreri), V.; 隆戈 (Longo), G. 利用图像处理和神经网络对碳酸盐岩进行纹理识别: 方法建议与实例。《计算机与地球科学》2005年, 31卷, 649 - 659页。[交叉引用]
- 辛格 (Singh), N.; 辛格 (Singh), T.; 蒂瓦里 (Tiwary), A.; 萨卡尔 (Sarkar), K.M. 利用图像处理和神经网络对玄武岩岩体进行纹理识别。《计算机与地球科学》2010年, 14卷, 301 - 310页。[交叉引用]
- 严 (Yen), H.H.; 蔡 (Tsai), H.Y.; 王 (Wang), C.C.; 蔡 (Tsai), M.C.; 曾 (Tseng), M.H. 一种利用深度学习集成机器学习的改进型胃食管反流病内镜自动分类模型。《诊断学》2022年, 12卷, 2827页。[交叉引用]
- 迪米特罗夫斯基, I.; 基塔诺夫斯基, I.; 科切夫, D.; 西米迪耶夫斯基, N. 用于地球观测的深度学习当前趋势: 一个用于图像分类的开源基准平台。《国际摄影测量与遥感学会摄影测量与遥感杂志》2023年, 第197卷, 第18 - 35页。[CrossRef]
- 徐, S.; 周, Y. 基于深度学习算法的显微镜下矿石矿物人工智能识别。《石油学报》2018年, 第34卷, 第3244 - 3252页。
- 张, Y.; 李, M.; 韩, S. 基于深度学习的岩石图像岩性自动识别与分类。《石油学报》2018年, 第34卷, 第333 - 342页。
- 程, G.; 李, P. 基于残差神经网络的岩石薄片图像分类。见《2021年第六届智能计算与信号处理国际会议 (ICSP) 论文集》, 中国西安, 2021年4月9 - 11日; 第521 - 524页。
- 陈, W.; 苏, L.; 陈, X.; 黄, Z. 利用迁移学习的深度残差神经网络进行岩石图像分类。《地球科学前沿》2023年, 第10卷, 第1079447期。[CrossRef]
- 科希达亚图拉, A.; 阿扎尼, S.; 巴拉博什金, E.E.; 阿尔法拉吉, M. Faciesvit: 用于改进岩心岩相预测的视觉变换器。《地球科学前沿》2022年, 第10卷, 第992442期。[CrossRef]
- 张, W.; 张, Q.; 刘, S.; 潘, X.; 陆, X. 用于多光谱图像变化检测的空间 - 光谱联合注意力网络。《遥感》2022年, 第14卷, 第3394期。[CrossRef]
- 塞格迪 (Szegedy), C.; 刘 (Liu), W.; 贾 (Jia), Y.; 塞尔马内特 (Sermanet), P.; 里德 (Reed), S.; 安格洛夫 (Anguelov), D.; 埃尔汗 (Erhan), D.; 万霍克 (Vanhoucke), V.; 拉比诺维奇 (Rabinovich), A. 卷积神经网络的深度探索。见《电气与电子工程师协会计算机视觉与模式识别会议论文集》, 美国马萨诸塞州波士顿, 2015年6月8 - 10日; 第1 - 9页。
- 西蒙扬 (Simonyan), K.; 齐斯曼 (Zisserman), A. 用于大规模图像识别的深度卷积神经网络。预印本2014, arXiv:1409.1556。
- 何 (He), K.; 张 (Zhang), X.; 任 (Ren), S.; 孙 (Sun), J. 用于图像识别的深度残差学习。见《电气与电子工程师协会计算机视觉与模式识别会议论文集》, 美国内华达州拉斯维加斯, 2016年6月26日 - 7月1日; 第770 - 778页。
- 米斯拉 (Misra), D.; 纳拉马达 (Nalamada), T.; 阿拉萨尼帕莱 (Arasanipalai), A.U.; 侯 (Hou), Q. 旋转以聚焦: 卷积三元注意力模块。见《IEEE 冬季计算机视觉应用会议论文集》, 线上会议, 2021年1月5 - 9日; 第3139 - 3148页。
- 冈田 (Okada), H. 砂岩分类: 分析与建议。《地质学杂志》1971年, 79卷, 509 - 525页。[参考文献链接]
- 海姆森 (Haimson), B.; 鲁德尼基 (Rudnicki), J.W. 中间主应力对粉砂岩断层形成和断层角度的影响。《构造地质学杂志》2010年, 32卷, 1701 - 1711页。[参考文献链接]
- 瓦尼曼 (Vaniman), D.; 比什 (Bish), D.; 明 (Ming), D.; 布里斯托 (Bristow), T.; 莫里斯 (Morris), R.; 布莱克 (Blake), D.; 奇佩拉 (Chipera), S.; 莫里森 (Morrison), S.; 特雷曼 (Treiman), A.; 兰普 (Rappe), E. 火星盖尔陨石坑耶洛奈夫湾泥岩的矿物学。《科学》2014年, 343卷, 1243480页。[参考文献链接]
- 黄, Y.; 程, Y.; 巴普纳, A.; 菲拉特, O.; 陈, D.; 陈, M.; 李, H.; 尼亚姆, J.; 乐, Q.V.; 吴, Y. Gpipe: 使用流水线并行性高效训练巨型神经网络。《神经信息处理系统进展》2019年, 第32卷, 第103 - 112页。

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We thank all reviewers for their comments and Guangdong TipDM Intelligent Technology Co., Ltd. for the rock images.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Fu, G.; Yan, J.; Zhang, K.; Hu, H.; Luo, F. Current status and progress of lithology identification technology. *Prog. Geophys.* **2017**, 32, 26–40.
- 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]
- 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.
- Mlynarczuk, 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]
- 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.
- 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.
- 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]
- 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]
- 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]
- 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]
- 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]
- Xu, S.; Zhou, Y. Artificial intelligence identification of ore minerals under microscope based on deep learning algorithm. *Acta Petrol. Sin.* **2018**, 34, 3244–3252.
- 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.
- 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.
- 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]
- 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]
- 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]
- 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.
- Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
- 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.
- 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.
- Okada, H. Classification of sandstone: Analysis and proposal. *J. Geol.* **1971**, 79, 509–525. [CrossRef]
- 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]
- 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]
- 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. 谭 (Tan), M.; 勒 (Le), Q. 高效网络 (EfficientNet)：重新思考卷积神经网络的模型缩放。见《国际机器学习会议论文集》，美国加利福尼亚州长滩，2019年6月10 - 15日；第6105 - 6114页。
27. 姚 (Yao), Q.; 王 (Wang), M.; 陈 (Chen), Y.; 戴 (Dai), W.; 李 (Li), Y.F.; 涂 (Tu), W.W.; 杨 (Yang), Q.; 余 (Yu), Y. 让人类脱离学习应用：自动机器学习综述。预印本arXiv 2018, arXiv:1810.13306。
28. 胡 (Hu), J.; 沈 (Shen), L.; 孙 (Sun), G. 挤压与激励网络。见《电气与电子工程师协会计算机视觉与模式识别会议论文集》，美国犹他州盐湖城，2018年6月18 - 22日；第7132 - 7141页。
29. 邓, J.; 董, W.; 索切尔, R.; 李, L.J.; 李, K.; 费菲, L. ImageNet：一个大规模分层图像数据库。见《2009年IEEE计算机视觉与模式识别会议论文集》，美国佛罗里达州迈阿密，2009年6月20 - 25日；第248 - 255页。
30. 桑托罗, A.; 巴尔图诺夫, S.; 博特维尼克, M.; 维斯特拉, D.; 利利克拉普, T. 基于记忆增强神经网络的元学习。见《国际机器学习会议论文集》，美国纽约州纽约市，2016年6月19 - 24日；第1842 - 1850页。
31. 克里兹夫斯基, A.; 苏斯克韦弗, I.; 辛顿, G.E. 基于深度卷积神经网络的ImageNet分类。《美国计算机协会通讯》2017年, 60卷, 84 - 90页。[CrossRef]
32. 甘, Y.; 郭, Q.; 王, C.; 梁, W.; 肖, D.; 吴, H. 利用改进的EfficientNet模型识别农作物害虫。《农业工程学报》2022年, 38卷, 203 - 211页。
33. 魏, Y.; 王, Z.; 乔, X.; 赵, C. 基于注意力机制和EfficientNet的轻量级水稻病害识别方法。《中国农业机械学报》2022年, 43卷, 172 - 181页。

免责声明/出版商说明：所有出版物中包含的声明、观点和数据仅代表个别作者和贡献者的意见，而非MDPI和/或编辑的意见。MDPI和/或编辑对因内容中提及的任何想法、方法、说明或产品导致的人员伤害或财产损失不承担责任。

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.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.