

Article

Classification of Rice Seeds Grown in Different Geographical Environments: An Approach Based on Improved Residual Networks

Helong Yu ^{1,2} , Zhenyang Chen ², Shaozhong Song ^{3,*}, Mojun Chen ⁴ and Chenglin Yang ¹

¹ College of Information Technology, Jilin Agricultural University, Changchun 130118, China; yuhelong@aliyun.com (H.Y.); yang_chenglin_1@163.com (C.Y.)

² Smart Agriculture Research Institute, Jilin Agricultural University, Changchun 130118, China; 20231615@mails.jlau.edu.cn

³ School of Data Science and Artificial Intelligence, Jilin Engineering Normal University, Changchun 130052, China

⁴ Jilin Academy of Agricultural Sciences, Changchun 130033, China; chmj123@126.com

* Correspondence: songsz@jleu.edu.cn; Tel.: +86-185-0440-6677

Abstract: Rice is one of the most important crops for food supply, and there are multiple differences in the quality of rice in different geographic regions, which have a significant impact on subsequent yields and economic benefits. The traditional rice identification methods are time-consuming, inefficient, and delicate. This study proposes a deep learning-based method for fast and non-destructive classification of rice grown in different geographic environments. The experiment collected rice with the name of Ji-Japonica 830 from 10 different regions, and a total of 10,600 rice grains were obtained, and the fronts and backsides of the seeds were photographed with a camera in batches, and a total of 30,000 images were obtained by preprocessing the data. The proposed improved residual network architecture, High-precision Residual Network (HResNet), was used to compare the performance of the models. The results showed that HResNet obtained the highest classification accuracy result of 95.13%, which is an improvement of 7.56% accuracy with respect to the original model, and validation showed that HResNet achieves a 98.7% accuracy in the identification of rice grown in different soil classes. The experimental results show that the proposed network model can effectively recognize and classify rice grown in different soil categories. It can provide a reference for the identification of other crops and can be applied for consumer and food industry use.

Keywords: rice region classification; residual network; rice soil; deep learning



Citation: Yu, H.; Chen, Z.; Song, S.; Chen, M.; Yang, C. Classification of Rice Seeds Grown in Different Geographical Environments: An Approach Based on Improved Residual Networks. *Agronomy* **2024**, *14*, 1244. <https://doi.org/10.3390/agronomy14061244>

Academic Editor: Yanbo Huang

Received: 8 May 2024

Revised: 23 May 2024

Accepted: 6 June 2024

Published: 7 June 2024



Copyright: © 2024 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

Rice, the staple food for nearly half of the global human population, is a rich source of nutrients [1]. Its yield and quality are crucial to global food security, soil conservation, and genetic diversity [2]. Moreover, rice cultivation serves as a vital income generator for farmers, creating employment opportunities. The cultivation and sale of rice provide farmers with economic stability, enabling them to improve their livelihoods and enhance their living standards. However, rice quality varies significantly across geographical regions, primarily due to natural factors such as climate, soil, and water conditions, as well as anthropogenic factors, including farming techniques and cultivar choices. These factors interact to influence rice's growth and development, ultimately determining its quality, yield, taste, and nutritional value. Typically, rice with superior quality commands a higher market preference, translating into increased economic benefits.

In many countries, seed production and management have been commercialized, harboring immense economic interests [3]. Nevertheless, low-quality rice is frequently misused by unethical traders, who disguise it as rice from popular cultivation regions in the market to maximize profits [4]. This practice does not only violate breeders' rights to

their varieties but also inflicts significant economic losses on seed producers, managers, and farmers. Furthermore, the widespread use of hybrid technology has compromised the purity of high-quality seeds [5]. Traditionally, rice classification methods have often relied on biological testing [6] and chemical detection [7], which are not only time-consuming but also often involve damaging the rice itself, posing potential hazards. Consequently, there is an urgent demand for a rapid and non-destructive method that can accurately classify rice grown in diverse geographical environments.

Adopting computer vision technology for classifying rice grown in diverse geographical environments offers numerous advantages compared to traditional classification methods. Computer vision technology can automatically extract image features from rice, enabling the swift processing of vast amounts of rice image data. Through algorithmic models, it achieves precise classification, significantly enhancing both efficiency and accuracy [8]. Moreover, this technology can effectively handle the vast diversity of rice grown in different geographical environments. Significant variations in climate, soil, and water conditions across regions lead to notable differences in rice morphology, color, and texture [9]. Traditional methods often struggle to accurately and quickly capture these differences in categories, and computer vision techniques can learn and adapt to the characteristics of rice in different environments, allowing for more accurate and faster classification [10]. Additionally, the utilization of computer vision technology for classifying rice grown in various geographical environments can propel the development of agricultural modernization and intelligence [11]. Machine vision technology typically enhances the accuracy of seed classification by extracting key features such as texture, morphology, and color [12,13].

Before that, many scholars carried out research on rice variety recognition by various means and methods. Qi et al. [14] studied the near-infrared hyperspectral imaging technology in the detection of rice seed vitality and achieved significant results in combination with transfer learning methods. The CNN model of Yongyou12, which was constructed with MixStyle transfer knowledge, was used to classify the vitality of Yongyou 1540, Su Xiangjing 100, and Longjingyou 1212. The accuracy rate reached 90.00%, 80.33%, and 85.00%, respectively, which achieved very good results. Joshi et al. [15] used deep neural networks and optical coherence tomography to achieve marker-free and lossless classification of rice seeds, achieving good classification results. Ge et al. [16] proposed a public hyperspectral image of rice seed data set and a case difficulty weighted K-nearest neighbor algorithm (IDKNN), which has achieved good results in the classification of different varieties of rice. Bi et al. [17] proposed a seed classification model based on Swin Transformer. The self-attention mechanism of this model can effectively extract image information, and it focuses on the study of feature attention and multi-scale feature fusion learning, which can accurately and efficiently classify seeds and achieve a very good effect. Jin et al. used near infrared hyperspectral imaging technology combined with deep learning to effectively distinguish different varieties of rice seeds and achieved very good results [18]. They also used IR-HSI, LeNet, GoogLeNet, and ResNet to identify rice seed varieties, among which the classification effect of the ResNet model is the best, and the classification accuracy of the test set is 86.08%. Diaz-Martinez et al. [19] proposed a framework for rice seed hyperspectral image processing and classification that combined two powerful tools of hyperspectral imaging and deep learning. 3D-CNN was used to classify five seeds with different processing and six seeds with different high temperature treatments. The average classification accuracy was 91.33% and 89.50%, respectively. The average accuracy of DNN for five different treatments was 94.83%, and the average accuracy of DNN for six different high-temperature exposure durations of each treatment was 91%, which achieved very good results. However, the research of hyperspectral and near-infrared spectroscopy is relatively expensive and low in efficiency, which is difficult to realize for large-scale application. In addition, they are all classifications of different rice varieties, while the research on rice growing in different regions and the tracing of rice origin of the same variety are limited. By using deep learning and machine vision technology, this study provides a fast,

lossless, and inexpensive classification method to classify rice seeds of the same variety growing in different regions and can effectively realize the trace of rice origin.

The improved network model in this study is ResNet50 [20], and ResNet50 is selected as the architecture of the improved model mainly for several considerations. First of all, ResNet50 is already a very classic and stable network architecture, which still shows excellent performance in a variety of computer vision tasks. Its deep residual structure allows the network to be deeper, so that more levels of feature information can be extracted. This can be very beneficial for many tasks. Second, the number of parameters in ResNet50 is moderate, neither too large to cause excessive consumption of computing resources, nor too small to limit its performance. Finally, ResNet50 has been extensively researched and applied, and there are a lot of resources and experiences in the community to refer to. This makes it easier to get technical support and solutions when making improvements using ResNet50 as the base model. Before that, many scholars had carried out a lot of research on ResNet. Elena Limonova et al. [21] introduced a bipolar ResNet (BM-ResNet) model, which can obtain a more complex ResNet architecture by converting its layers into bipolar layers. The performance of the model is verified on MNIST and CIFAR-10 image classification problems, and the results show that the model achieves good results. Sasha Targ et al. [22] proposed a new architecture, namely, ResNet (RiR) in ResNet, which incorporated these generalized residual blocks into the framework of ResNet, and demonstrated the latest performance of RiR architecture on CIFAR-100, showing that very good results were achieved. WEI-JIAN HU et al. [23] proposed a multi-dimensional feature compensation residual neural network (MDFC-ResNet) model for crop disease diagnosis. MDFC-ResNet identified species, coarse-grained diseases, and fine-grained diseases from three dimensions and set compensation layers, achieving very good results. Jie Hu et al. [24] built SE-ResNet50, which not only improved the model performance but also increased the calculation amount of the model, which still needs further research and improvement. Although many new network structures have appeared in recent years, they may still be in constant development and improvement, and there may be certain uncertainties and risks compared with classical models like ResNet50. Therefore, in the pursuit of stability and reliability, the selection of ResNet50 as an improved model is a relatively conservative but very effective choice. The research contents and methods of this paper include:

- (1) Phenotypic differences were revealed by collecting data from rice seed images of Jiji-Japonica 830 grown in 10 different regions.
- (2) The generalization ability of the model was improved through preprocessing and data amplification [25].
- (3) Construct the residual network model HResNet, and train and optimize the model.
- (4) The classification performance of the model was evaluated and compared with ResNeXt50 [26] proposed by Xie et al. in 2017 based on ResNet50, Res2Net50 [27] proposed by Gao et al. in 2019, MobileNetV3 [28], a lightweight network model proposed by Howard et al. in 2019 and 2021. Ding et al. proposed the classical network model RepVggNet_B0 [29] and DenseNet121 [30], a classical network model proposed by Iandola et al. in 2014, for comparison.
- (5) According to the confusion matrix and loss curve, the improvement process of the model is revealed, and the visualization tool Grad-CAM [31] is used to show the convolution effect of the model, and the role of the SCE block, activation function, and learning rate adjustment strategy is revealed.
- (6) Verify the recognition ability of the model for rice growing in different soil categories.

Through this study, we aspire to provide novel methods and perspectives for computer vision in classifying and identifying rice grown in diverse geographical environments. Additionally, we aim to offer more precise and scientific guidance for agricultural production practices, thereby making positive contributions to the development of agricultural intelligence and enhancing the efficiency and profitability of rice cultivation.

2. Materials and Methods

2.1. Sample Preparation

2.1.1. Data Acquisition

The samples were selected from the Rice Research Institute of Jilin Academy of Agricultural Sciences. Ji-Japonica 830 was selected in the experiment. JiB2008-1853 (Yunlangxiang) was used as the maternal parent, and Tongke 17 was used as the male parent. Ji-Japonica 830 belonged to the late-maturing variety with a growth period of about 144 days and a plant height of about 104.1 cm, which was affected by the growth environment. Ji-Japonica 830 varieties of rice from ten different regions were collected for research. The longitude, latitude, and soil type of the ten different regions were different, as shown in Figure 1 and Table 1, and the naming rules of the different regions are abbreviated by the name of the specific region where the rice is grown, for example, Dagushan, Siping City (DGS), laying a solid foundation for further research on rice growing in different geographical environments.

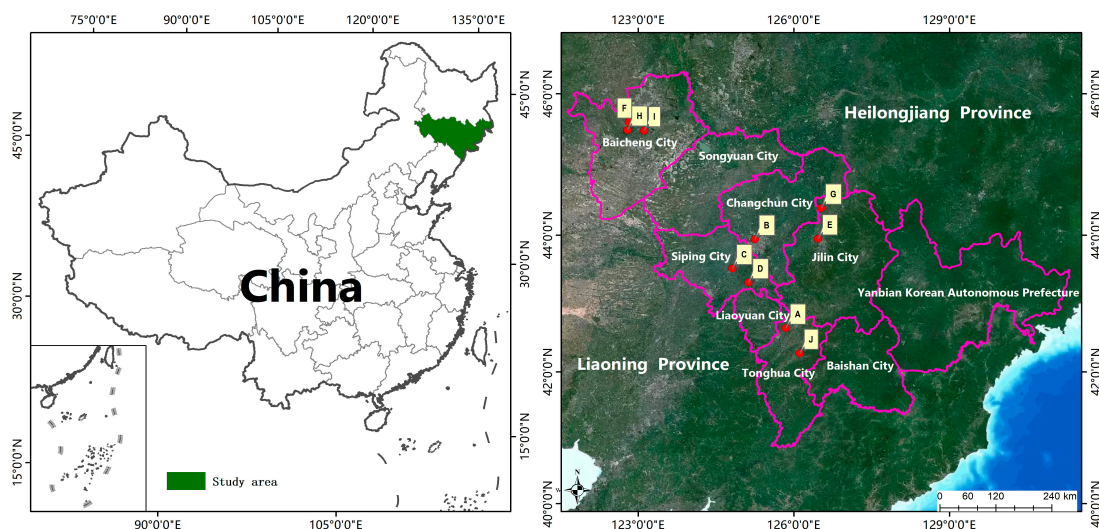


Figure 1. Regional distribution of Ji-Japonica 830.

Table 1. Latitude and longitude and soil types of ten regions in Jilin Province. (The Label in Table 1 represents the serial number in Figure 1).

Label	Name	Growing Area	Longitude and Latitude	Soil Type
A	THS	Tonghua City	125.84°, 42.49°	Albic Soil
B	JG	Changchun City	125.25°, 43.94°	Phaeozem
C	GZL	Siping City	124.82°, 43.52°	Phaeozem
D	DGS	Siping City	125.13°, 43.31°	Phaeozem
E	JZ	Jilin City	126.47°, 43.95°	Meadow Soil
F	BSC	Baicheng City	122.83°, 45.60°	Chernozem
G	BQC	Jilin City	126.43°, 44.39°	Meadow Soil
H	CCKN	Baicheng City	122.79°, 45.49°	Chernozem
I	BGNC	Baicheng City	123.12°, 45.48°	Meadow Soil
J	JJD	Tonghua City	126.12°, 42.28°	Dark Brown Forest Soil

2.1.2. Image Acquisition and Preprocessing

The imaging system, as shown in Figure 2a, consists of a NikonD7100 camera (Nikon, Tokyo, Japan) and a 18–105 mm lens; two lights are controlled by the lamp source control system, a load stage is placed at the bottom to place the seeds, and the camera is shot vertically. In the data collection process, 200 rice seeds from the same region were randomly selected and arranged in a 10 × 20 template. Then, without overlapping or sticking, the seeds were placed on the black grinding plate, RGB images were obtained through

the camera, and the JPG format of the images was saved to the corresponding folder. Then, the grinding plate was buttoned on the absorbent cloth, the backs of the seeds were photographed, and the above process was repeated. As shown in Figure 2c, the obtained image is processed with grayscale, and then a binary mask is performed, as shown in Figure 2d, where the threshold is set to 0.3. According to the binary image and the original image, the image with the background removed is obtained, as shown in Figure 2e. Finally, each image with the background removed is divided into an image containing the target region. As shown in Figure 2f, each seed was defined as the target region in this study, and the size of the image was controlled to be 224×224 . Save the image of the target area in JPG format to the corresponding folder. As shown in Figure 3, rice seeds of the same category in different regions have very diverse forms.

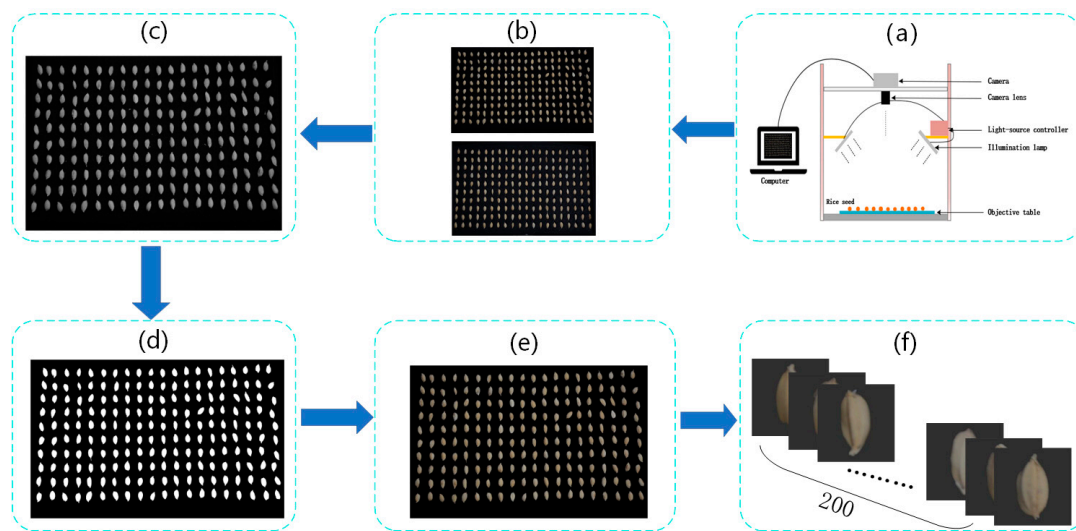


Figure 2. Schematic diagram of image acquisition and preprocessing process, (a) is the image capturing system, (b) is the front and back image of the original image, (c) is the grayscale image, (d) is the binary image, (e) is the original image with the background removed, and (f) is the segmented image.

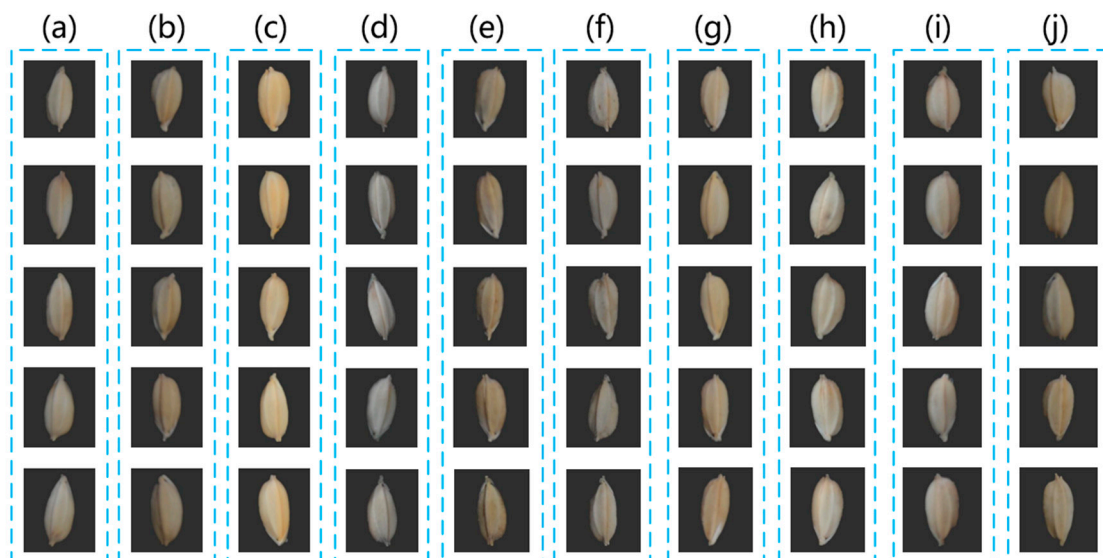


Figure 3. The segmented original image shows that (a) is BGNC, (b) is BQC, (c) is BSC, (d) is CCKN, (e) is DGS, (f) is GZL, (g) is JG, (h) is JJD, (i) is JZ, and (j) is THS.

2.1.3. Data Expansion and Segmentation

As shown in Table 2, a total of 10,600 rice grains were collected in the experiment. In this experiment, useless images, such as those blurred, damaged, or completely inconsistent with the requirements of the experiment, need to be removed after shooting the grains, which may interfere with the training and evaluation process of the model, resulting in the model learning wrong or irrelevant features. These images not only fail to provide valuable information to the model but may also degrade the performance of the model.

In this experiment, data were amplified for categories with a small number of samples, and samples with a large number of data were randomly culled to control the size of the data set and make the training process more efficient. As shown in Table 2, after removing useless images, the minimum number was only 1048, while the maximum number was 3944, indicating a very unbalanced data set. Controlling the size of the data set can prevent overfitting to some extent. At the same time, by introducing more transformations, the robustness of the model can be increased, making it less sensitive to small changes in the input data, thus reducing the risk of overfitting.

Table 2. Data processing shows details.

Name	BGNC	BQC	BSC	CCKN	DGS	GZL	JG	JJD	JZ	THS
Seed number	600	1000	2000	1200	1000	800	1000	1000	1000	1000
Number of images	1200	2000	4000	2400	2000	1600	2000	2000	2000	2000
Eliminate useless images	1048	1902	3944	2245	1942	1539	1916	1970	1904	1945
Data amplification and deletion	3000	3000	3000	3000	3000	3000	3000	3000	3000	3000
Training set	2400	2400	2400	2400	2400	2400	2400	2400	2400	2400
Validation set	600	600	600	600	600	600	600	600	600	600

As shown in Figure 4, image mirroring and random rotation were used in this experiment for data amplification. Data amplification is a very effective technique that can generate more training samples through a series of transformations on the original image while taking into account a small number of data sets. This not only increases the diversity of the data set, but also helps the model learn more different features, thus improving its generalization ability. For deep learning models, more training samples means more learning opportunities, which helps to improve the accuracy and stability of the model.

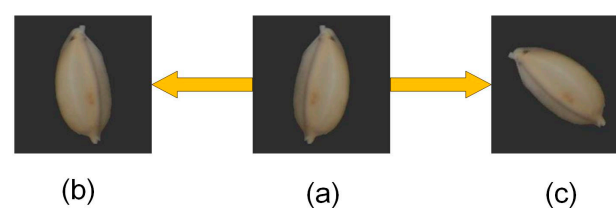


Figure 4. Data set expansion diagram (a) is the original image, (b) is the image after image mirroring, and (c) is the image after image rotation.

Finally, we split the training set and the verification set by 8:2, without a dedicated test set. This is because we chose to use part of the data set to verify the performance of the model, while the rest of the data were used entirely to train the model. This setup helps us to monitor and adjust the training process of the model in time to ensure that it performs well on previously unseen data. In this experiment, the validation set is divided and evaluated several times to simulate the role of the test set to obtain a reliable estimate of the model's performance. Our experimental setup is a choice based on specific needs and resource considerations, and through the rational use of training and validation sets, we are able to effectively train and tune the model and evaluate its performance as best as possible.

2.1.4. Data Sets by Soil Type

To locate the producing soil of rice, we mixed rice growing in the same soil together to conduct corresponding research, disrupting the corresponding order, and randomly selected 3000 samples from each soil category, totaling 15,000 samples, to explore the model's classification and recognition performance of rice growing in different soil categories. Details are shown in Table 3. On this basis, continue to divide the training set and the verification set in an 8:2 ratio.

Table 3. The data set is divided into details by rice producing soil type.

Soil Type	Locality of Growth	Name	Quantity per Class	Total Sample Size	Randomly Selected Sample
Chernozem	Baicheng City	BSC	3000	6000	3000
	Baicheng City	CCKN	3000		
Meadow soil	Baicheng City	BGNC	3000	9000	3000
	Jilin City	JZ	3000		
	Jilin City	BQC	3000		
Dark brown forest soil	Tonghua City	JJD	3000	3000	3000
Albic soil	Tonghua City	THS	3000	3000	3000
Phaeozem	Changchun City	JG	3000	9000	3000
	Siping City	GZL	3000		
	Siping City	DGS	3000		

2.2. Model Building

2.2.1. The Structure of ResNet50 and HResNet

The improvement process of the model is shown in Figure 5. Figure 5a represents the original structure of the ResNet50 model, Figure 5b represents the bottleneck of the ResNet50 model, and Figure 5c represents the bottleneck of the improved HResNet model, where ReLU and Mish are two different activation functions, BN is a batch normalization layer, Max pool maximum pooling, Avg pool is average pooling, Conv is an ordinary convolutional layer, GConv is a group convolution, DSConv is a depth separable convolution, Fc stands for fully connected layer, and SCEBlock is an improved module in this study.

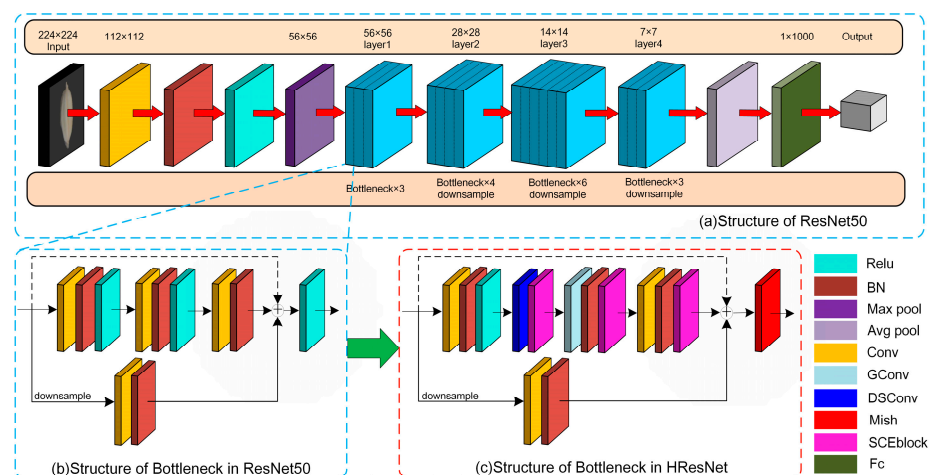


Figure 5. Structure diagram of the original model ResNet50 and the improved model HResNet.

2.2.2. Different Convolutions

To improve the classification performance without greatly increasing the number of parameters, the model introduced deep separable convolution [32], as shown in Figure 5c. Standard convolutional neural networks mainly include the convolutional layer, pooling layer, and fully connected layer. However, standard convolution obtains a large number of parameters and high computational costs, which puts high demands on the computing

platform. By adding depth-separable convolution, the accuracy is improved correspondingly, and the number of parameters is not greatly increased. Therefore, this study uses depth-separable convolution instead of standard convolution to improve our network. The specific calculation parameters are shown in Table 4. It decomposes the traditional convolution layer into a deep convolution layer and a point convolution layer [33], as shown in Figure 6.

In order to further improve the classification performance, the model replaced the 3×3 ordinary convolution in the trunk with group convolution [34] and defined the group size as 32, as shown in Figure 5c. Group convolution can effectively reduce the computational load of convolution, mainly because it groups the output eigenvalues and allows different groups to use different convolution cores for convolution operations. This grouping and selective convolution method, compared with the traditional convolution with one convolution kernel for all input features, greatly reduces the computational load and speeds up the training of the model. Group convolution can increase the learning ability and expressiveness of the model. By grouping convolution, the model can learn more information about the interactions between channels, which helps the model better understand and process complex input data. In fact, group convolution and depthwise convolution are very similar, but there are also differences. When the size of the group is equal to the size of the input channel, it is a depthwise convolution; when the size of the group is smaller than the size of the input channel and greater than 1, it is a group convolution, as shown in Figure 6.

Through the decomposition of the convolution kernel and convolution process, separable convolution and group convolution have higher efficiency than standard convolution. This can greatly control the number of parameters and the calculation cost while improving the model performance (see Table 4). As shown in Figure 6, assume that the number of input and output channels are C_{in} and C_{out} , respectively, K^2 is the kernel size of the convolution, G is the size of the group, and the size of the picture is $H \times W$.

Table 4. Comparison of parameters and computation amount of standard convolution, group convolution, and depthwise separable convolution.

Convolution	Parameters	Computational Cost
Canonical convolution	$K^2 \times C_{in} \times C_{out}$	$K^2 \times H \times W \times C_{out} \times C_{in}$
Group convolution	$K^2 \times \frac{C_{in}}{G} \times C_{out}$	$K^2 \times \frac{C_{in}}{G} \times H \times W \times C_{out}$
Separable convolution	$K^2 \times C_{in} + 1 \times 1 \times C_{out}$	$(C_{in} \times K^2 + C_{in} \times C_{out}) \times H \times W$

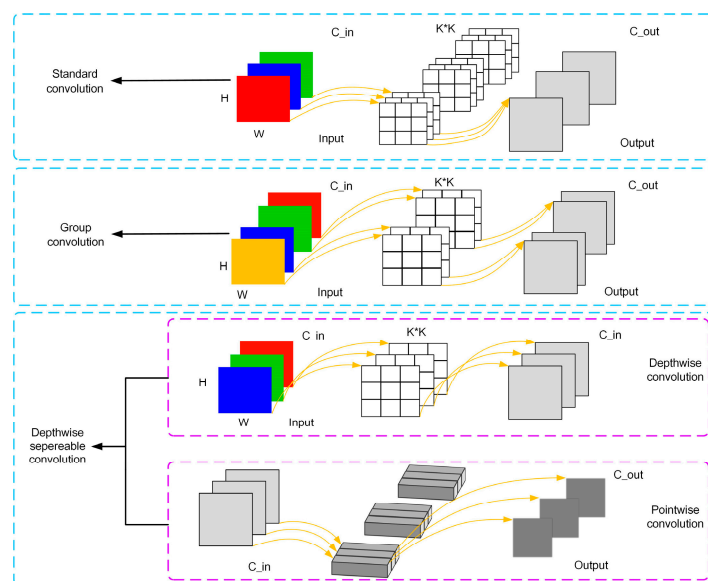


Figure 6. Schematic diagram of standard convolution, group convolution, and separable convolution.

2.2.3. The Activation Function Used in This Study

To enhance model convergence and smoothness during training, we incorporated the Mish activation function [35] into the model, illustrated in Figure 5c. Compared to the ReLU activation function [36], Mish offers several advantages. First, as depicted in Figure 7, Mish is smoother than ReLU, mitigating issues like gradient disappearance and explosion, thereby enhancing the stability and convergence speed of the neural network. Second, both Mish and ReLU have infinite upper bounds, allowing positive values to reach any height, thus circumventing saturation problems associated with capping. While ReLU tends to truncate at negative values, Mish exhibits a slight tolerance for negatives, promoting better gradient flow and alleviating the hard zero boundary problem in ReLU. Finally, owing to Mish's smoothness, it facilitates deeper penetration of information into the neural network, thereby improving model accuracy and generalization.

Mish encompasses the Tanh activation function [37] and the Softplus activation function [38]. Due to its higher computational demand compared to ReLU, we opted to introduce Mish selectively. The formulas for the ReLU and Mish functions are outlined in Table 5.

Table 5. Comparison of ReLU activation functions and Mish activation functions.

Activation Function	Calculation Formula
ReLU	$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$
Mish	$f(x) = x \times \tanh(\text{softplus})$
Tanh	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
Softplus	$f(x) = \ln(1 + e^x)$

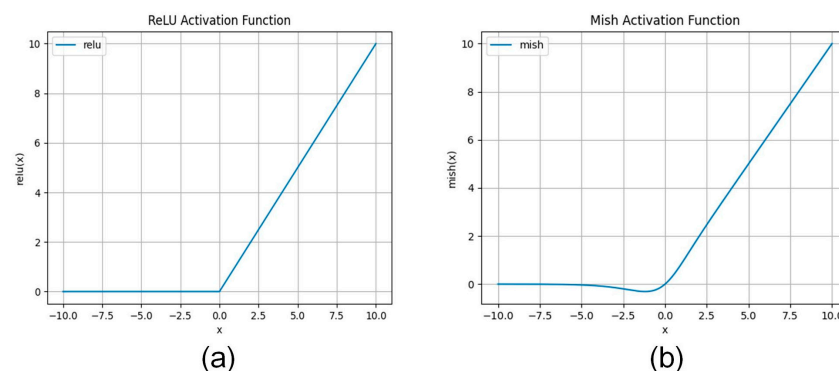


Figure 7. Curves of ReLU and Mish functions, (a) is the curve for ReLU, (b) is the curve for Mish.

2.2.4. SCE Block

To enhance recognition accuracy in small areas, we prioritize exploring the relationship between feature channels, enabling the network to construct informative features across various image locations. To achieve this, we introduce the SCE block, as depicted in Figure 8, an enhancement inspired by SE Net [39]. This involves replacing all full connection layers with 1×1 convolution layers and substituting the ReLU activation function with the Mish activation function, which effectively emphasizes the relationship between feature channels.

The approach entails automatically learning the importance of each feature channel, facilitating the promotion of useful features based on their significance while suppressing less relevant features for the current task. The adaptability of the SCE block allows for its seamless integration into existing networks, thereby enhancing model accuracy. Notably, as depicted in Figure 5c, our study strategically placed SCE blocks in three distinct locations, resulting in superior performance.

Moreover, 1×1 convolution offers unique advantages over fully connected layers. While there is no difference in calculation between a fully connected layer and 1×1 convolution

when the input feature graph size is 1×1 , differences arise when the input size exceeds this. A key distinction is that the input size of a fully connected layer is fixed, whereas a convolutional layer has arbitrary input sizes, granting 1×1 convolution greater flexibility in handling inputs of various sizes. Moreover, 1×1 convolution serves to decrease or increase nonlinearity. It can effectively reduce the number of input channels, thereby cutting down on computation and model parameters. Simultaneously, it introduces nonlinearities that facilitate the learning of complex feature representations, which is particularly beneficial for large datasets or scenarios requiring efficient computation. Additionally, 1×1 convolution can be utilized for feature combination and interaction. By employing an appropriate activation function in 1×1 convolution, the model can grasp the nonlinear relationships between channels, thus enhancing the network's expressiveness. In contrast, a fully connected layer may compromise spatial information when processing data with spatial structures like images. However, 1×1 convolution retains this spatial information, enabling cross-channel information interaction. In summary, 1×1 convolution has greater flexibility, the ability to downscale and increase nonlinearity, and the ability to combine and interact features compared to fully connected layers. These advantages position 1×1 convolution as a powerful tool in deep neural networks, enhancing model performance and expressiveness.

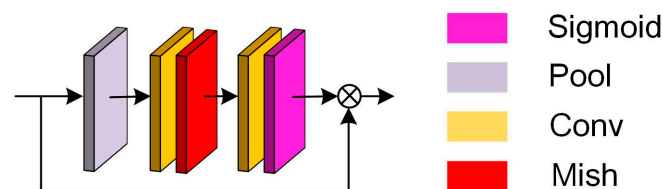


Figure 8. Structure of the SCE block.

2.2.5. Cross Entropy Loss

The loss function used in this study is Cross Entropy Loss [40], which is robust to the probability distribution predicted by the model. Even if the model has a small deviation in the predicted probability of some categories, it will not affect the overall loss too much. This makes the model more stable during training and less susceptible to noise or outliers. The binary classification Cross Entropy Loss is shown below:

$$L = -[y \log p + (1 - y) \log(1 - p)] \quad (1)$$

where y denotes the sample label and p denotes the probability that the corresponding sample label is predicted to be positive. In the multiclassification task, each sample may have more than one possible category, and the model output is the probability distribution of each sample belonging to each category. Cross Entropy Loss can measure the distance between the probability distribution of the model output and the true labels so as to guide the model optimization. The multcategory Cross Entropy Loss formula is as follows:

$$L = - \sum_{c=1}^M y_c \log p_c \quad (2)$$

where p_c denotes the probability that the label is predicted to be c .

2.3. Evaluation Indexes of Model

In the field of machine learning, confusion matrices are often used to compare the results of model classification in supervised learning. Each column of the matrix represents the predicted class, and each row represents the actual class. The binary classification problem is taken as an example. The actual result is defined as positive and the predicted result is defined as positive, denoted as TP; if the actual result is negative, the predicted result is positive, denoted as FP; if the actual result is positive, the predicted result is

negative, denoted as FN; if the actual result is negative, the predicted result is negative, denoted as TN. The specific structure of the confusion matrix is shown in Table 6.

Table 6. Confusion matrix of binary classification problem.

Confusion Matrix		Actual Results	
		Positive	Negative
Forecast results	Positive	TP	FP
	Negative	FN	TN

Accuracy (ACC), precision (P), recall (R), F1 score (F1), and specificity (SP) can be derived from the confusion matrix's data and serve as evaluation metrics for assessing the classification performance of the model. Table 7 presents the formulas and concise descriptions of these evaluation indicators.

Table 7. Formula and short description of evaluation indicators.

Evaluation Metric	Calculation Formula	Short Description
Accuracy (ACC)	$ACC = \frac{TP+TN}{TP+FP+FN+TN}$	The ratio of the number of correctly predicted positive and negative samples to the total number of samples.
Precision (P)	$P = \frac{TP}{TP+FP}$	The ratio of the number of correctly predicted positive samples to the total number of samples predicted to be positive.
Recall (R)	$R = \frac{TP}{TP+FN}$	The ratio of the number of correctly identified positive samples to the total number of actual positive samples.
F1-score (F1)	$F1 = \frac{2TP}{2TP+FP+FN}$	The harmonic mean of precision and recall.
Specificity (SP)	$SP = \frac{TN}{FP+TN}$	The proportion of predictions that are actually negative samples that are correctly predicted by the model to be negative

2.4. Training Hyperparameter Information

Detailed information regarding the specific experimental parameters utilized in training the novel network model proposed in this paper is provided. We standardized the input size of the dataset to 224×224 pixels, set the number of training rounds to 100, and ensured a sample count of 30,000. The learning rate plays a pivotal role in the learning process of the model, and setting it incorrectly can have profound consequences. If the learning rate is excessively high, the model may rapidly update its weights, leading to a faster exploration of the solution space. However, this can also result in the model skipping over the optimal solution due to oversized steps, causing vibrations or even divergence during training, and hampering convergence to a stable solution. Conversely, a learning rate set too low ensures a more stable convergence, but at a significantly slower rate. While this allows for finer adjustments to the weights, leading to a more precise solution, it significantly extends the training time and increases the risk of the model settling into a local, rather than global, optimal solution. Recognizing the importance of balancing these factors, we employed an exponential decay strategy to dynamically adjust the learning rate during training, aiming for optimal training outcomes. We initiated the learning rate at 0.01 and set the decay rate to 0.91, naming this approach “Nine One Decay” in this study. Details pertaining to this and other experimental parameters are outlined in Table 8.

Table 8. Training hyperparameter information details.

Set of Parameter	Value or Name
Input size	224×224
Batch size	64
Epoch	100
Learning rate	0.01
Number of samples	30,000
Optimizer	SGD
Learning rate scheduler	$LR_{epoch} = LR_0 \times decay\ rate^{epoch}$
Decay rate	0.91

2.5. Experimental Environment Configuration

The experiment was conducted on a computer equipped with an Intel® Xeon® Gold 6246R CPU (Intel, Santa Clara, CA, USA) operating at 3.4 GHz and an NVIDIA Quadro RTX 8000 GPU (NVIDIA, Santa Clara, CA, USA) boasting 48 GB of memory, running on a Windows 10 operating system. The software setup was composed of Anaconda 3-2021.11 for Windows, utilizing the PyCharm compiler and the Python 3.8 programming language, which was integrated with Pytorch 1.12.1. It is noteworthy that all algorithms were executed within this uniform and standardized environment.

3. Results

3.1. Comparative Experiment with Image Amplification

To guarantee that data amplification genuinely influenced the model's performance, we employed the original ResNet50 model as the starting point for our investigation. To ensure consistency, all experimental parameters were maintained identically. After implementing the sample data amplification strategy, we observed a noteworthy 7.8% increase in the model's recognition accuracy. This significant enhancement demonstrates that data amplification can effectively broaden the diversity of data and subsequently strengthen the model's generalization capabilities. Consequently, we selected the rice sample enhanced with data amplification for further experimentation. Table 9 provides a comprehensive overview of the model's recognition accuracy and loss under the influence of data amplification.

Table 9. Model recognition accuracy under data amplification.

Data Amplification	Sample Size	ACC (%)	Loss
Yes	30,000	87.57	0.356
No	20,355	81.23	0.498

3.2. Ablation Experiments

Ablation experiments serve as a pivotal research approach in deep learning models, aimed at exploring and validating the significance and role of various components or functionalities within the model. As evident from the experimental data presented in Table 10, the addition of group convolution, depth separable convolution, Mish activation function, and SCE block individually led to improvements in accuracy. Notably, the incorporation of the SCE block resulted in the highest accuracy of 91.47%, while the inclusion of the Mish activation function alone yielded the lowest increase in accuracy, reaching 88.96%. This can be attributed to the Mish activation function's effectiveness in accelerating convergence and mitigating gradient disappearance during training. Comparatively, combining two components yielded a relative improvement in accuracy. Specifically, the combination of DSConv and SCE blocks demonstrated the most significant enhancement, achieving 93.47%. Furthermore, the integration of all three components—SCE block, DSConv, and Mish activation function—proved to be the most beneficial, achieving a remarkable accuracy of 94.73%. Ultimately, when all components were jointly employed, optimal performance

was attained, with an accuracy of 95.13%. These findings highlight the synergistic effect of various components in enhancing the model's overall performance.

Table 10. Ablation results.

GConv	DSConv	Mish	SCE Block	ACC (%)	P (Avg)	R (Avg)	SP (Avg)	F1 (Avg)
				87.57	0.876	0.876	0.986	0.875
✓				89.03	0.890	0.890	0.988	0.890
	✓			89.78	0.898	0.898	0.989	0.898
		✓		88.96	0.889	0.890	0.989	0.889
			✓	91.47	0.915	0.915	0.990	0.914
✓	✓			90.23	0.903	0.902	0.989	0.902
✓		✓		91.08	0.911	0.911	0.990	0.911
✓			✓	91.97	0.920	0.920	0.991	0.919
	✓	✓		91.17	0.912	0.912	0.990	0.912
	✓		✓	93.47	0.935	0.935	0.993	0.935
		✓	✓	93.02	0.930	0.930	0.992	0.930
✓	✓	✓		91.95	0.920	0.920	0.991	0.919
✓	✓		✓	93.37	0.934	0.934	0.993	0.934
✓		✓	✓	93.15	0.932	0.932	0.993	0.931
	✓	✓	✓	94.73	0.947	0.947	0.994	0.947
✓	✓	✓	✓	95.13	0.951	0.951	0.995	0.951

3.3. Comparison with Classical CNN Models

To comprehensively assess the performance of the enhanced model on targeted tasks, authenticate its effectiveness and advancements, and benchmark it against established methodologies, we have devised a series of comparative experiments that aim to thoroughly evaluate the model's robustness. Through this experiment, we compare the improved network model with five classical convolutional neural networks, namely, ResNeXt50, Res2Net50, RepVggNet_B0, MobileNetV3, and DenseNet121. The results are shown in Table 11 and Figure 9.

Table 11. Comparative experimental results.

Model	ACC (%)	P (Avg)	R (Avg)	SP (Avg)	F1 (Avg)
ResNeXt50	90.43	0.905	0.904	0.989	0.905
Res2Net50	85.92	0.859	0.859	0.984	0.860
RepVggNet_B0	90.15	0.902	0.902	0.989	0.901
MobileNetV3	89.40	0.894	0.894	0.988	0.894
DenseNet121	90.48	0.905	0.905	0.989	0.905
HResNet	95.13	0.951	0.951	0.995	0.951

As shown in Figure 9, when comparing the confusion matrix of the other five models, we can find that each model has a different performance on the classification task. Looking at the confusion matrix of the improved model, we can see that the value on the diagonal is relatively high, which means that the model performs well in correctly classifying rice from different regions. The higher the value on the diagonal, the higher the prediction accuracy of the model for the corresponding category. At the same time, the relatively low value on the non-diagonal line means that there are fewer cases in which the model misclassifies the sample into other categories. In contrast, the confusion matrices of the other five models show varying degrees of difference. Some models have lower values on the diagonal of the confusion matrix than HResNet, indicating that they are less accurate than HResNet in the classification task. Other models do not perform well in specific categories. As can be seen from the figure, DGS category classification performs poorly, but HResNet has a significantly higher ability to classify DGS than other models. By comparing the confusion matrix of the five models, we can find that the improved model has obvious advantages

in classification performance. It can not only accurately distinguish rice from different producing areas but also maintain a high accuracy rate on the whole. This advantage makes the improved model more reliable and effective for the rice origin tracing task.

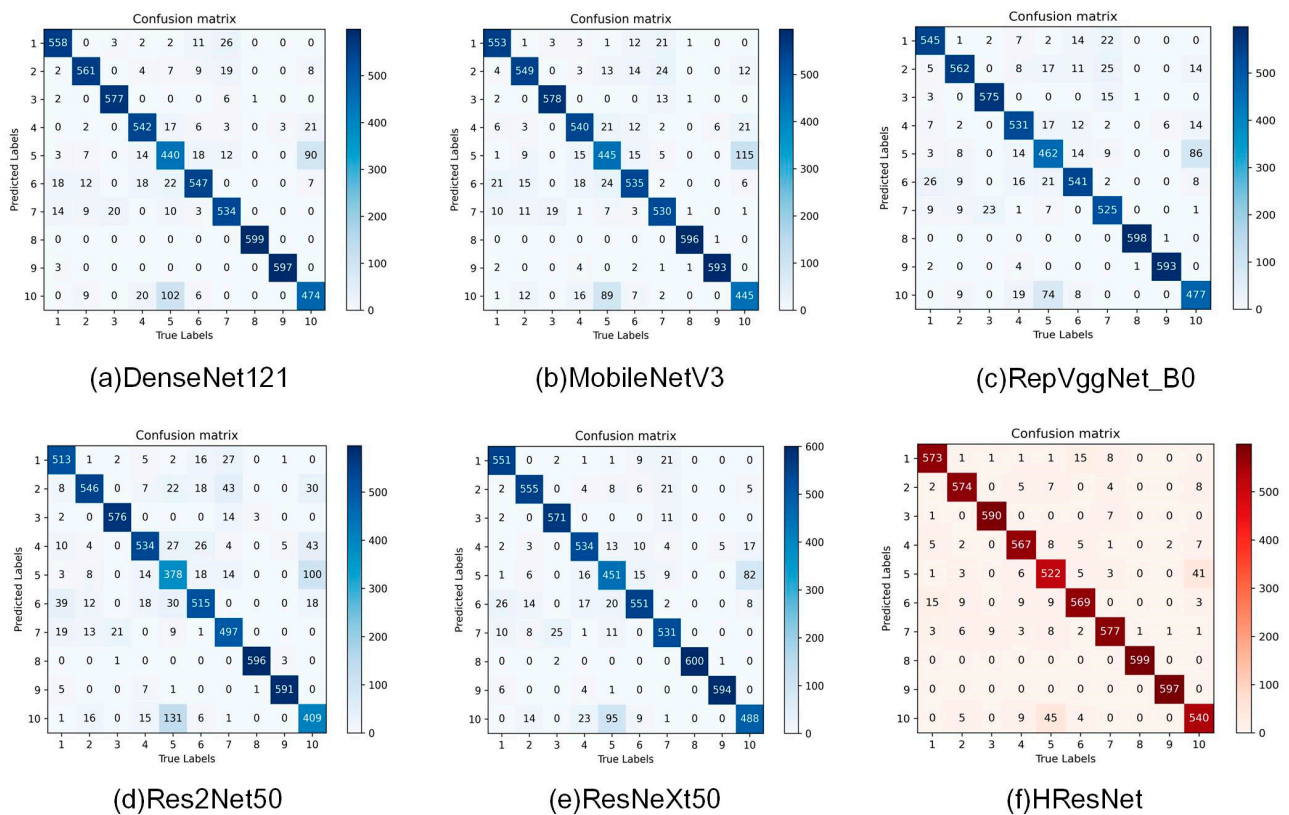


Figure 9. Compare the confusion matrix of the model in the comparative experiment.

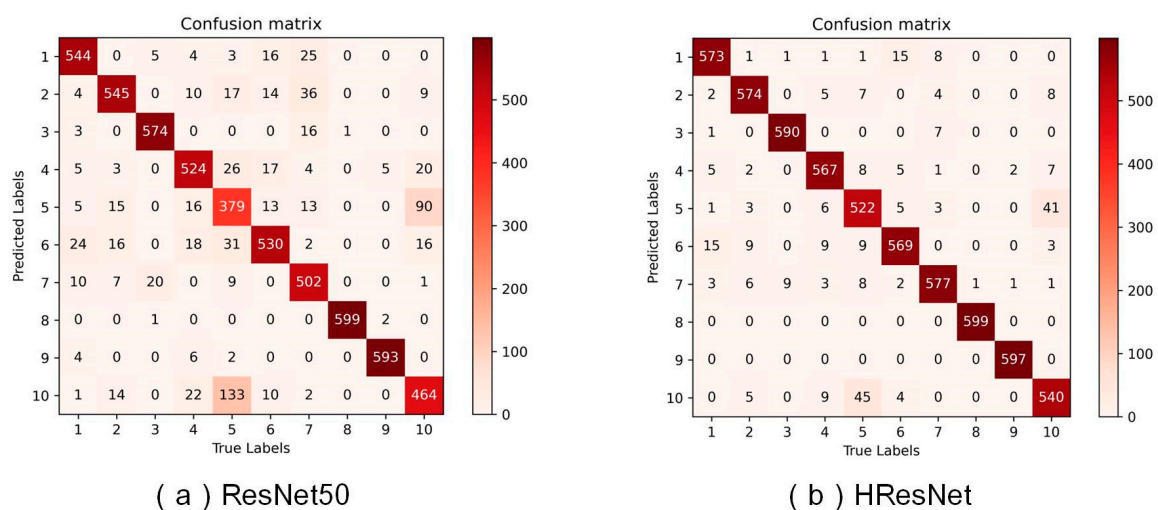
3.4. Comparison of Results before and after Improvement

The detailed comparison results of various indicators between ResNet50 and the improved model HResNet are shown in Table 12. Before the improvement, the accuracy of the model was relatively low, which means that the model often makes mistakes in prediction and cannot well identify samples that really belong to a certain category. After optimization and adjustment, the accuracy of the model has been significantly improved. Now the model can more accurately predict the categories of samples and reduce the possibility of misjudgment. Especially, the categories of DGS and THS have been greatly improved, from 0.714 to 0.898 and from 0.718 to 0.896, respectively, and JJD and JZ have reached 1.0. Before the model improvement, the overall recall rate was not high, resulting in the omission of the model in identifying the samples that really belonged to DGS and THS, and the model could not correctly identify all the positive samples. By improving the structure and parameters of the model, the recall rate has been significantly improved. The model is now able to cover positive samples more comprehensively, reducing omissions. The low specificity of the model before improvement means that the model is prone to misjudgment when identifying samples that do not belong to a certain class and misclassifies negative samples as positive samples. After the improvement, the specificity of the model has improved. The model can now more accurately distinguish negative samples, reducing the number of cases of false negative samples. Before the improvement, the F1 index of the model was also low due to the relatively low accuracy and recall rate, and the comprehensive performance was not ideal. With the improvement in accuracy and recall rate, the F1 index of the model was also significantly improved. This shows that the model has achieved a better balance between accuracy and recall, and the overall performance has been optimized.

Table 12. Comparison of results for ten categories before and after improvement.

Label	Name	Precision		Recall		Specificity		F1 Score	
		Before	After	Before	After	Before	After	Before	After
1	BGNC	0.911	0.955	0.907	0.955	0.990	0.995	0.909	0.955
2	BQC	0.858	0.957	0.908	0.957	0.983	0.995	0.882	0.957
3	BSC	0.966	0.987	0.957	0.983	0.996	0.999	0.961	0.985
4	CCKN	0.868	0.950	0.873	0.945	0.985	0.994	0.870	0.947
5	DGS	0.714	0.898	0.632	0.870	0.972	0.989	0.671	0.884
6	GZL	0.832	0.927	0.883	0.948	0.980	0.992	0.857	0.937
7	JG	0.914	0.944	0.837	0.962	0.991	0.994	0.874	0.953
8	JJD	0.995	1.000	0.998	0.998	0.999	1.000	0.996	0.999
9	JZ	0.980	1.000	0.988	0.995	0.998	1.000	0.984	0.997
10	THS	0.718	0.896	0.773	0.900	0.966	0.988	0.744	0.898

As illustrated in Figure 10, a comparison of the confusion matrices before and after the improvement reveals a substantial enhancement in the model's performance. Specifically, examining the confusion matrix for HResNet, it is evident that the classification capabilities of the model have undergone a noticeable upgrade across all categories. Notably, the misclassification rate for DGS, which was previously a significant challenge, has been significantly reduced, enabling the model to more precisely identify samples belonging to this category. Furthermore, the diagonal values have undergone a marked improvement, resulting in a darker coloration that indicates a significant enhancement in the recognition rates across each category. Additionally, the overall clarity of the confusion matrix has been improved, resulting in a higher degree of differentiation between classes. In summary, the enhanced deep learning model exhibits superior performance in classification tasks, delivering more reliable and accurate classification outcomes for a range of related applications.

**Figure 10.** The confusion matrix of the model before and after improvement.

As depicted in Figure 11, the accuracy curve of the original model exhibited fluctuations during the initial stages of training, but its growth gradually decelerated and stabilized towards the later phases. Despite achieving a relatively consistent accuracy level, there was a potential risk of overfitting throughout the training process. Conversely, the accuracy curve of the improved model demonstrated a more pronounced upward trend. In the early stages of training, the accuracy rose rapidly and subsequently maintained a steady growth pattern, ultimately reaching its peak accuracy level and stabilizing. By comparing the accuracy curves before and after the enhancement, it becomes apparent

that the optimized model exhibits significant improvements in both optimization capacity and accuracy.

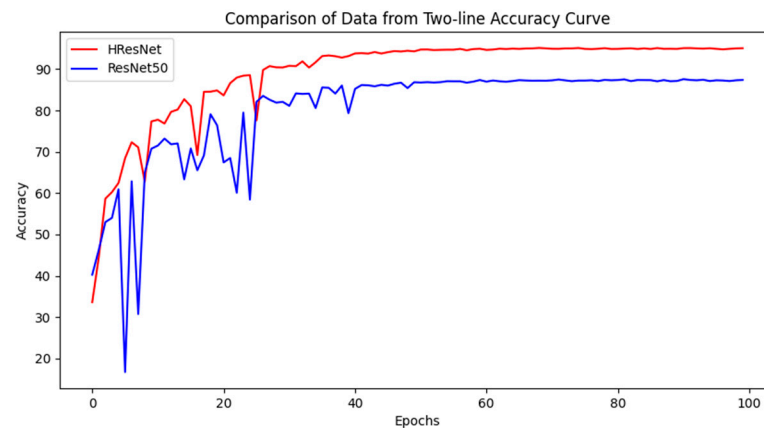


Figure 11. Accuracy curve of model validation before and after improvement.

As depicted in Figure 12, the loss curve prior to improvement, while gradually flattening out in the later stages of training, exhibits significant fluctuations during the initial phase. This suggests that the model encounters optimization challenges during the training process, making it challenging to further minimize losses. Furthermore, the relatively high final value of the loss curve indicates limited fitting capabilities of the model, potentially indicating the presence of overfitting or underfitting issues. However, the improved loss curve exhibits a more desirable downward trend. During the initial stages of training, the loss value rapidly decreases and subsequently stabilizes, with minimal fluctuations. This demonstrates that the enhanced model exhibits greater stability during the optimization process and is more effective at reducing losses. Ultimately, the lower value of the loss curve signifies a significant improvement in the model's fitting capabilities, enabling it to better adapt to the training data.

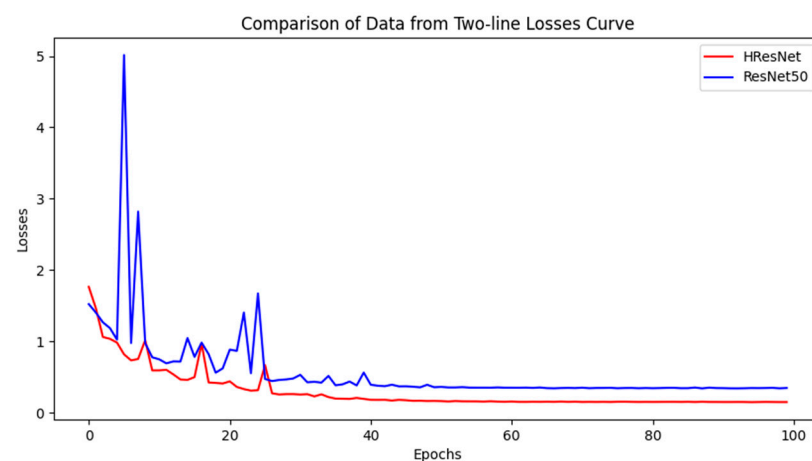


Figure 12. The loss curve of model validation before and after improvement.

The visualizations generated by the Grad-CAM tool offer a more intuitive visualization of the interested regions within each layer of the feature map, comparing the model's state before and after improvements. As depicted in Figure 13, although the Grad-CAM heatmap of the original model can localize regions in the rice images that contribute significantly to the classification outcomes during the rice origin tracing task, its precision and clarity in identifying activated regions remain inadequate. Prior to the enhancement, the Grad-CAM heatmap often struggled to pinpoint crucial information, thus limiting the accuracy of classification. However, with the advancement of the residual network, the new Grad-

CAM heatmap has undergone significant improvements in terms of precision and feature representation. The refined model excels at extracting key features from rice images and mapping them onto heatmaps with unprecedented precision. On the heatmap, the crucial feature areas are clearly highlighted, with sharper boundaries that more accurately reflect the underlying feature information. Comparing the Grad-CAM heatmaps of each layer, both before and after the improvements, the enhanced model's prowess in feature extraction becomes evident. Notably, the transition from Layer 4 reveals a shift in the model's focus, from a preponderance on local features to a stronger emphasis on global features. This transformation not only augments the model's tracing accuracy but also enhances its resilience against complex backgrounds and images with subtle features, providing a more robust technical backbone for the rice origin traceability task.

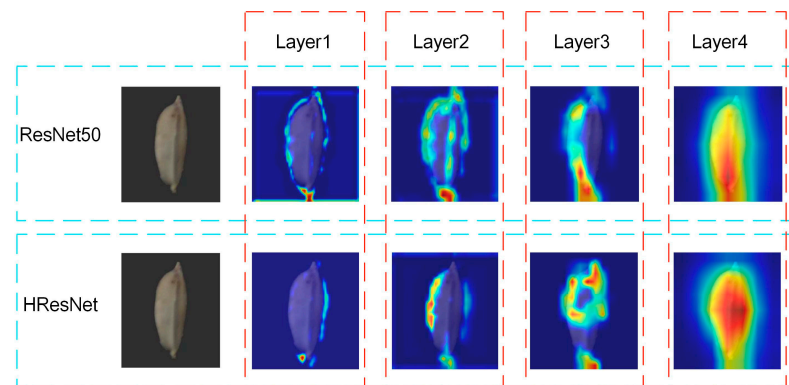


Figure 13. Grad-CAM schematic of each layer before and after improvement.

3.5. SCE Block Verification Result

To investigate the performance disparities between the SCE block and the SE block, we conducted a series of experiments in this study. Specifically, we substituted the SCE block in HResNet with the SE block, resulting in HResNet_SE, and subsequently performed comparable experiments. As depicted in Table 13, the accuracy, training loss, and parameter size of HResNet augmented with the SE block are inferior to those of HResNet enhanced with the SCE block. Furthermore, the parameter size of HResNet_SE is greater than that of ResNet50, mainly because the parameters of the fully connected layer are larger than those of 1×1 convolution. At the same time, FLOPs of HResNet are significantly reduced compared with ResNet50, mainly because depthwise separable convolution and group convolution will greatly reduce the number of parameters and calculation amount compared with ordinary convolution.

Table 13. Results of accuracy, loss, and parameter size comparison between HResNet, HResNet added with SE, and the original model.

Model	Accuracy (%)	Loss	Parameter Size (MB)	FLOPs (G)
ResNet50	87.57	0.356	89.75	4.132
HResNet (SE)	92.62	0.214	96.06	2.698
HResNet (SCE)	95.13	0.162	63.77	2.690

Although the HResNet equipped with the SE block demonstrates commendable performance when compared to the original model, it falls short when benchmarked against the SCE block. According to Table 14, the SE block exhibits weak advantages in terms of accuracy rate, recall rate, specificity, and F1 index. Consequently, overall, the SCE block outperforms the SE block in terms of its comprehensive performance.

Table 14. Comparison results of ten categories of HResNet and HResNet_SE.

Name	Precision		Recall		Specificity		F1 Score	
	SE	SCE	SE	SCE	SE	SCE	SE	SCE
BGNC	0.952	0.955	0.935	0.955	0.995	0.995	0.943	0.955
BQC	0.929	0.957	0.942	0.957	0.992	0.995	0.935	0.957
BSC	0.985	0.987	0.970	0.983	0.998	0.999	0.977	0.985
CCKN	0.925	0.950	0.930	0.945	0.992	0.994	0.927	0.947
DGS	0.800	0.898	0.795	0.870	0.978	0.989	0.797	0.884
GZL	0.915	0.927	0.938	0.948	0.990	0.992	0.926	0.937
JG	0.929	0.944	0.937	0.962	0.992	0.994	0.933	0.953
JJD	1.000	1.000	0.995	0.998	1.000	1.000	0.997	0.999
JZ	0.995	1.000	0.988	0.995	0.999	1.000	0.991	0.997
THS	0.832	0.896	0.832	0.900	0.981	0.988	0.832	0.898

3.6. The Effect of Nine One Decay and Activation Function on the Model

To meticulously investigate the impact of the learning rate adjustment strategy and the Mish activation function within the HResNet architecture, we conducted a comparative analysis by substituting the Mish function with the ReLU function and eliminating the Nine One Decay mechanism. This allowed us to delve into the convergence effects of Mish and Nine One Decay, as well as their capacity to mitigate gradient vanishing issues.

As evident in Figure 14, upon removing Nine One Decay, the accuracy fluctuations in the model training process become pronounced. This excessive step size may hinder the discovery of the optimal solution, leading to jitters or even divergence during training and compromising the ability to converge to a stable solution. However, the inclusion of Nine One Decay offers a remedy. As training progresses, the learning rate gradually tapers off, enabling the model to refine its step size as it nears the optimal solution, thus minimizing volatility. Consequently, the convergence curve exhibits smoother behavior in the later stages and gradually gravitates towards the optimal solution.

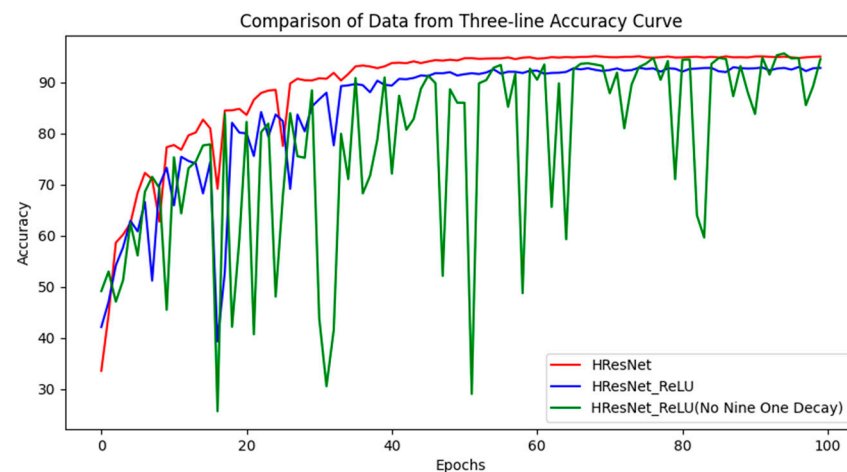


Figure 14. Accuracy comparison of HResNet, HResNet_ReLU, and HResNet_ReLU (No Nine One Decay) verification processes.

This strategy not only accelerates the convergence speed of the model but also enhances its stability, facilitating satisfactory performance in fewer iterations. Nevertheless, initial training stages exhibited significant fluctuations, which the Mish activation function effectively addressed. A comparative analysis revealed that HResNet with the Mish function avoided this initial volatility, fostering faster convergence, higher accuracy, and a smoother training process than when utilizing the ReLU function.

In conclusion, the Mish function and Nine One Decay play pivotal roles in HResNet, jointly contributing to its superior performance.

3.7. Results of Soil Verification in the Rice Producing Area

To delve into the remarkable performance of HResNet in classifying rice grown in diverse soil categories, this study conducted pertinent experiments, as outlined in Table 15. Overall, the model demonstrated outstanding classification capabilities, particularly for Dark Brown Forest Soil, achieving a perfect score of 1 across all metrics. First, the model exhibited exceptional accuracy, accurately categorizing samples into the correct soil type when predicting the rice varieties associated with different types of soils, with minimal instances of misclassification. Second, the model's recall rate was also impressive, achieving a perfect score of 1 in both Dark Brown Forest Soil and Meadow Soil. This indicates that the model effectively identifies and correctly classifies the majority, or even all, of the samples from each origin, minimizing the occurrence of overlooked cases. Additionally, the model's specificity was exceptionally high, reaching a maximum of 1 in both Dark Brown Forest Soil and Meadow Soil. This high specificity ensures that the model accurately distinguishes non-target origin samples, thereby reducing the likelihood of misclassification. This feature significantly enhances the model's stability and reliability. Lastly, the F1 score, a composite metric reflecting both accuracy and recall, also demonstrated a high level. The high F1 score further validates the model's excellent performance in balancing accuracy and recall, effectively showcasing its superior capabilities in origin tracing tasks.

Table 15. Classification results of five soil types for rice growth.

Label	Soil Type	Precision	Recall	Specificity	F1 Score
1	Dark brown forest soil	1.000	1.000	1.000	1.000
2	Albic soil	0.983	0.992	0.996	0.987
3	Meadow soil	0.998	1.000	1.000	0.999
4	Chernozem	0.980	0.980	0.995	0.980
5	Phaeozem	0.973	0.963	0.993	0.968

As Figure 15 clearly illustrates, the confusion matrix demonstrates an impressive accuracy rate of 98.7% for the HResNet model in classifying rice grown in five distinct soil categories, indicating its outstanding performance. The high values along the diagonal elements of the matrix signify that the vast majority of samples have been accurately categorized into their respective soil types. This remarkable achievement underscores the superior capability of HResNet in distinguishing rice varieties cultivated under different soil conditions. The relatively low values in the non-diagonal elements indicate that misclassifications are infrequent and, if they occur, their proportion is minimal, having a negligible impact on the overall accuracy of soil classification in rice-producing areas. This further validates the high stability and reliability of the HResNet model. Furthermore, upon careful examination of the confusion matrix, we did not detect any discernible trends of misclassification among rice varieties grown in different soils. This suggests that our model exhibits consistent performance across various production regions and is resilient to potential classification errors that might arise due to similarities between soil types. Consequently, the analysis of the five soil categories for rice production in the confusion matrix confirms that HResNet has achieved exceptional results, not only in terms of high accuracy but also in maintaining stability and balance in its performance.

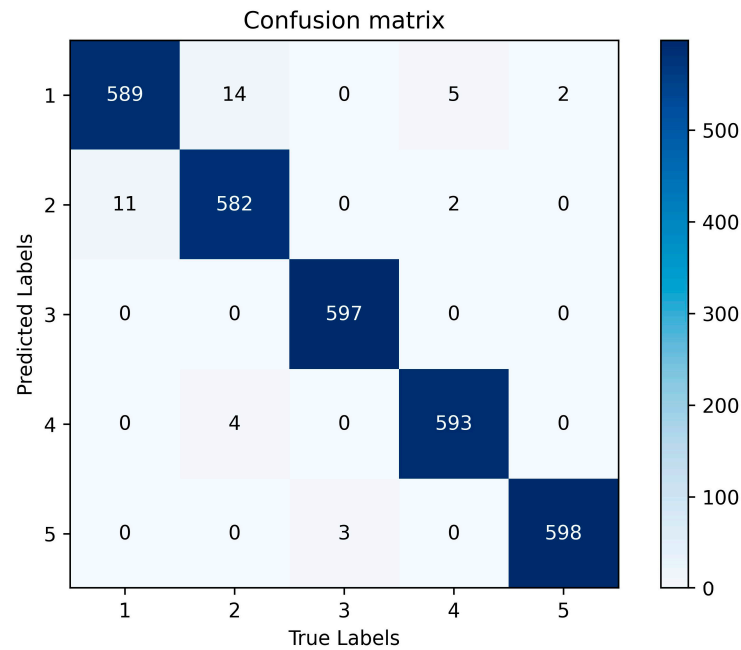


Figure 15. Confusion matrix of rice-producing soils from HResNet.

As depicted in Figure 16, rice images representing five distinct soil categories and their corresponding heatmaps through each layer of the HResNet model are exhibited. In the original images, features like color, texture, and shape are prominently displayed, serving as fundamental inputs for subsequent layers of processing. In the initial layer, HResNet initiates the extraction of original image features, encompassing edges, corners, and basic textures. The heatmap of this layer reveals that the model has already begun abstracting and extracting these original features. Differences in these original features among rice grown in various soil types start to manifest in this layer, establishing a solid foundation for more intricate feature extraction in subsequent layers. Moving to the second layer, the model delves deeper into extracting the subtler characteristics of the rice. In the third layer, the model proceeds to abstract and integrate original features, resulting in higher level feature representations. The heatmap of this layer illustrates that the model is now capable of distinguishing more intricate and specific attributes, such as the rice's shape, color, and texture, which are crucial in differentiating rice grown in different soil conditions. Finally, in the fourth layer's heatmap, the model has accomplished the comprehensive recognition and feature integration of rice from the five different origins. The color distribution and intensity on the heatmap reflect the model's confidence and discriminatory ability to distinguish rice grown in diverse soils. It is evident that the model demonstrates a clear distinction between rice from different origins, underscoring its efficiency and precision in the task of origin tracing. Through the sequential processing of these four layers, the model transparently exhibits how it extracts, abstracts, and combines features from the raw image data, ultimately achieving the precise identification and classification of rice grown in diverse soil conditions. The heatmaps at each layer vividly illustrate the distinct stages and transformations undergone during the feature extraction and classification processes of HResNet, underscoring its impressive capabilities and effectiveness in the crucial task of origin tracing.

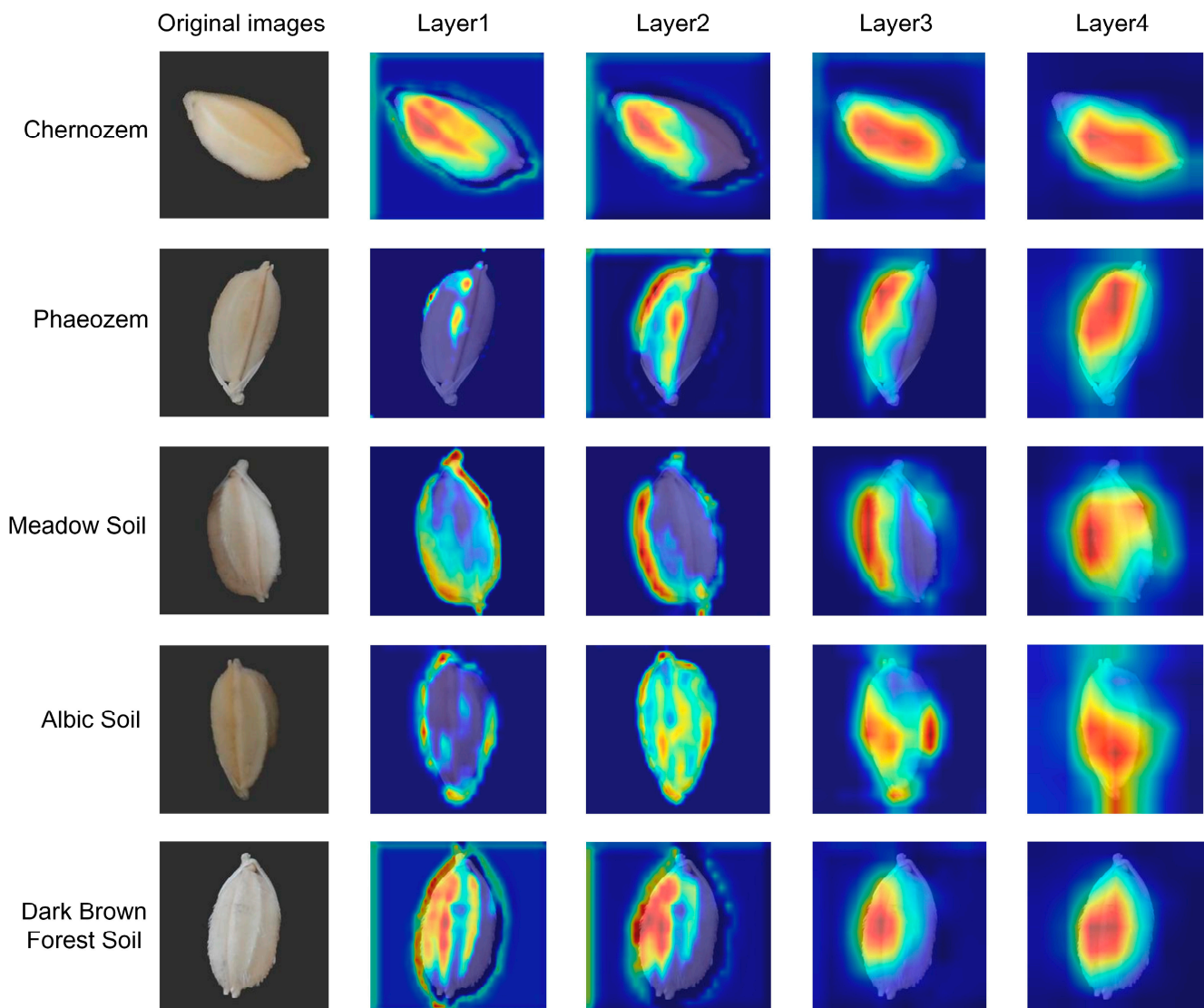


Figure 16. Heat maps of each layer of rice in HResNet for five soil categories.

4. Discussion

In this study, improvements were made to the ResNet50 model for the purpose of classifying and identifying rice grains grown in diverse geographical environments and soil types, achieving exceptional classification performance. The introduction of depthwise separable convolutions, grouped convolutions, SCE blocks, and the Mish activation function, tailored specifically to the characteristics of rice grain images, significantly enhanced the model's ability to extract crucial feature information. Furthermore, data augmentation techniques were utilized to enrich the dataset and bolster the model's generalization capabilities. These advancements allowed the model to effectively learn the distinguishing features of rice grains across varying geographical environments during the training process, resulting in superior classification accuracy on the test set. Notably, the HResNet model achieved high precision without an increase in the size of parameters and FLOPs. This was achieved through the utilization of depthwise separable convolutions and grouped convolutions, which divide the computational process into multiple groups, substantially reducing computational costs. Additionally, experimental results demonstrated that the parameters of fully connected layers tend to be larger compared to 1×1 convolutions.

However, this study has some limitations. The impact of five distinct soil types on rice growth was examined, yet the potential influences of other environmental factors such as climate and lighting were not considered. Future research should expand the

dataset to include a wider range of environmental factors to provide a more comprehensive evaluation of rice grain classification performance. Moreover, while the focus was primarily on rice seed classification, other related applications, such as real-time monitoring and prediction of rice growth status, were not explored. Extending the methodology to these additional areas could yield broader applications.

In conclusion, the improvements to the ResNet50 model and the incorporation of various optimization measures have effectively bolstered its capability in classifying and identifying rice grains cultivated in varying geographical environments and soil types. Despite the existing limitations, it is anticipated that further research and technological advancements will address these issues, paving the way for wider application prospects.

5. Conclusions

In this study, the model underwent enhancements in four crucial areas. First, by introducing depthwise separable convolution operations, we not only boosted the model's performance but also managed to maintain the same number of parameters. Second, the integration of the SCE block significantly improved the recognition accuracy of smaller regions. Third, we replaced ordinary convolution with group convolution, effectively reducing the computational burden. Lastly, the switch from the ReLU activation function to the Mish function significantly improved the model's convergence, effectively preventing issues like gradient disappearance and gradient explosion. Furthermore, we validated the effectiveness of Nine One Decay, which significantly mitigated the risk of overfitting. This comprehensive approach not only enhances the model's performance but also boosts its fine extraction and overall recognition capabilities. This integration further expanded the network model's scale feature extraction abilities and strengthened its expressive power. Backed by the characteristics of this improved model, we conducted rigorous experiments across datasets from 10 diverse production areas. The results revealed that our proposed method achieved a remarkable accuracy rate of 95.13%, significantly surpassing the original network model. Additionally, when compared to five commonly used image classification network models, our HResNet model demonstrated superior performance. Experimental verification further confirmed that HResNet excelled in rice classification across various soil categories, achieving an accuracy of 98.7%.

In the future, this method can be seamlessly integrated with other technological advancements, including remote sensing technology, the Internet of Things, and other cutting-edge techniques, to further enhance the precision and comprehensiveness of agricultural product traceability. In essence, the enhanced residual network's application to rice origin traceability has introduced innovative ideas and methods for tracing agricultural products, offering a novel solution for tracing other agricultural products as well. This advancement holds significant theoretical value and practical importance.

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

Funding: This research was funded by the National Key R&D Program of China (2022YFD2001602) and the Natural Science Foundation of Jilin Province (No. 2020122348JC). This study was also supported in part by the Jilin Provincial Department of Science and Technology (20220202032NC) and the Innovation Capacity Project on Development and Reform Commission of Jilin Province (2020C019-6).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: We thank Jilin Agricultural University for providing equipment support and the reviewers for their suggestions.

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

References

1. Fukagawa, N.K.; Ziska, L. Rice: Importance for global nutrition. *J. Nutr. Sci. Vitaminol.* **2019**, *65*, S2–S3. [\[CrossRef\]](#)
2. Custodio, M.C.; Demont, M.; Laborte, A.; Ynion, J. Improving food security in Asia through consumer-focused rice breeding. *Glob. Food Secur.* **2016**, *9*, 19–28. [\[CrossRef\]](#)
3. Amerit, B.; Ntayi, J.M.; Ngoma, M.; Bashir, H.; Echegu, S.; Nantongo, M. Commercialization of biofuel products: A systematic literature review. *Renew. Energy Focus* **2023**, *44*, 223–236. [\[CrossRef\]](#)
4. Van der Laan, H.L. *The Lebanese Traders in Sierra Leone*; Walter de Gruyter GmbH & Co KG: Berlin, Germany, 2019.
5. Gaffney, J.; Anderson, J.; Franks, C.; Collinson, S.; MacRobert, J.; Woldemariam, W.; Albertsen, M. Robust seed systems, emerging technologies, and hybrid crops for Africa. *Glob. Food Secur.* **2016**, *9*, 36–44. [\[CrossRef\]](#)
6. Li, P.; Chen, Y.-H.; Lu, J.; Zhang, C.-Q.; Liu, Q.-Q.; Li, Q. Genes and their molecular functions determining seed structure, components, and quality of rice. *Rice* **2022**, *15*, 18. [\[CrossRef\]](#)
7. Cui, S.; Ma, X.; Wang, X.; Zhang, T.A.; Hu, J.; Tsang, Y.F.; Gao, M.T. Phenolic acids derived from rice straw generate peroxides which reduce the viability of *Staphylococcus aureus* cells in biofilm. *Ind. Crops Prod.* **2019**, *140*, 111561. [\[CrossRef\]](#)
8. Tian, H.; Wang, T.; Liu, Y.; Qiao, X.; Li, Y. Computer vision technology in agricultural automation—A review. *Inf. Process. Agric.* **2020**, *7*, 1–19. [\[CrossRef\]](#)
9. Patindol, J.A.; Siebenmorgen, T.J.; Wang, Y. Impact of environmental factors on rice starch structure: A review. *Starch-Stärke* **2015**, *67*, 42–54. [\[CrossRef\]](#)
10. Patrício, D.I.; Rieder, R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* **2018**, *153*, 69–81. [\[CrossRef\]](#)
11. Shaikh, T.A.; Rasool, T.; Lone, F. Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Comput. Electron. Agric.* **2022**, *198*, 107119. [\[CrossRef\]](#)
12. Cinar, I.; Koklu, M. Identification of rice varieties using machine learning algorithms. *J. Agric. Sci.* **2022**, *9*. [\[CrossRef\]](#)
13. Saxena, P.; Priya, K.; Goel, S.; Aggarwal, P.K.; Sinha, A.; Jain, P.J. Rice varieties classification using machine learning algorithms. *J. Pharm. Negat. Results* **2022**, 3762–3772. [\[CrossRef\]](#)
14. Qi, H.; Huang, Z.; Sun, Z.; Tang, Q.; Zhao, G.; Zhu, X.; Zhang, C.J. Rice seed vigor detection based on near-infrared hyperspectral imaging and deep transfer learning. *Front. Plant Sci.* **2023**, *14*, 1283921. [\[CrossRef\]](#)
15. Joshi, D.; Butola, A.; Kanade, S.R.; Prasad, D.K.; Mithra, S.A.; Singh, N.; Bisht, D.S.; Mehta, D.S. Label-free non-invasive classification of rice seeds using optical coherence tomography assisted with deep neural network. *Opt. Laser Technol.* **2021**, *137*, 106861. [\[CrossRef\]](#)
16. Ge, Y.; Song, S.; Yu, S.; Zhang, X.; Li, X. Rice seed classification by hyperspectral imaging system: A real-world dataset and a credible algorithm. *Comput. Electron. Agric.* **2024**, *219*, 108776. [\[CrossRef\]](#)
17. Bi, C.; Hu, N.; Zou, Y.; Zhang, S.; Xu, S.; Yu, H. Development of deep learning methodology for maize seed variety recognition based on improved swin transformer. *Agronomy* **2022**, *12*, 1843. [\[CrossRef\]](#)
18. Jin, B.; Zhang, C.; Jia, L.; Tang, Q.; Gao, L.; Zhao, G.; Qi, H. Identification of rice seed varieties based on near-infrared hyperspectral imaging technology combined with deep learning. *ACS Omega* **2022**, *7*, 4735–4749. [\[CrossRef\]](#)
19. Díaz-Martínez, V.; Orozco-Sandoval, J.; Manian, V.; Dhatt, B.K.; Walia, H. A deep learning framework for processing and classification of hyperspectral rice seed images grown under high day and night temperatures. *Sensors* **2023**, *23*, 4370. [\[CrossRef\]](#)
20. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
21. Limonova, E.; Alfonso, D.; Nikolaev, D.; Arlazarov, V.V. ResNet-like architecture with low hardware requirements. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 6204–6211.
22. Targ, S.; Almeida, D.; Lyman, K. Resnet in resnet: Generalizing residual architectures. *arXiv* **2016**, arXiv:1603.08029.
23. Hu, W.-J.; Fan, J.; Du, Y.-X.; Li, B.-S.; Xiong, N.; Bekkering, E. MDfC-ResNet: An agricultural IoT system to accurately recognize crop diseases. *IEEE Access* **2020**, *8*, 115287–115298. [\[CrossRef\]](#)
24. 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, 18–23 June 2018; pp. 7132–7141.
25. Shorten, C.; Khoshgoftaar, T. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [\[CrossRef\]](#)
26. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
27. Gao, S.-H.; Cheng, M.-M.; Zhao, K.; Zhang, X.-Y.; Yang, M.-H.; Torr, P. Res2net: A new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 652–662. [\[CrossRef\]](#)
28. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
29. Ding, X.; Zhang, X.; Ma, N.; Han, J.; Ding, G.; Sun, J. Repvgg: Making vgg-style convnets great again. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13733–13742.

30. Iandola, F.; Moskewicz, M.; Karayev, S.; Girshick, R.; Darrell, T.; Keutzer, K. Densenet: Implementing efficient convnet descriptor pyramids. *arXiv* **2014**, arXiv:1404.1869.
31. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
32. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
33. Huang, Z.; Wang, R.; Cao, R.; Zheng, Y.; Teng, S.; Wang, Y.; Wang, F.; Du, L. Deep learning based soybean seed classification. *Comput. Electron. Agric.* **2022**, *202*, 107393. [[CrossRef](#)]
34. Cohen, T.; Welling, M. Group equivariant convolutional networks. In Proceedings of the International Conference on Machine Learning, New York City, NY, USA, 19–24 June 2016; pp. 2990–2999.
35. Misra, D. Mish: A self regularized non-monotonic activation function. *arXiv* **2019**, arXiv:1908.08681.
36. Ramachandran, P.; Zoph, B.; Le, Q. Searching for activation functions. *arXiv* **2017**, arXiv:1710.05941.
37. Karlik, B.; Olgac, A. Performance analysis of various activation functions in generalized MLP architectures of neural networks. *Int. J. Artif. Intell. Expert Syst.* **2011**, *1*, 111–122.
38. Liu, Q.; Furber, S. Noisy softplus: A biology inspired activation function. In *Neural Information Processing: Proceedings of the 23rd International Conference, ICONIP 2016, Kyoto, Japan, 16–21 October 2016, Proceedings, Part IV 23*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 405–412.
39. Cheng, D.; Meng, G.; Cheng, G.; Pan, C. SeNet: Structured edge network for sea–land segmentation. *IEEE Geosci. Remote Sens. Lett.* **2016**, *14*, 247–251. [[CrossRef](#)]
40. Zadeh, S.G.; Schmid, M. Bias in cross-entropy-based training of deep survival networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3126–3137. [[CrossRef](#)]

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.