Article

Research on the Classification of Sun-Dried Wild Ginseng Based on an Improved ResNeXt50 Model

Dongming Li [1,2], Zhenkun Zhao [1], Yingying Yin [1] and Chunxi Zhao [3,*]

1　College of Information Technology, Jilin Agricultural University, Changchun 130118, China; ldm0214@163.com (D.L.); zzkzgr2022@163.com (Z.Z.); yyy@jlau.edu.cn (Y.Y.)
2　College of Internet of Things Engineering, Wuxi University, Wuxi 214105, China
3　Information Center of Jilin Agricultural University, Changchun 130118, China
*　Correspondence: zcx@jlau.edu.cn

Abstract: Ginseng is a common medicinal herb with high value due to its unique medicinal properties. Traditional methods for classifying ginseng rely heavily on manual judgment, which is time-consuming and subjective. In contrast, deep learning methods can objectively learn the features of ginseng, saving both labor and time. This experiment proposes a ginseng-grade classification model based on an improved ResNeXt50 model. First, each convolutional layer in the Bottleneck structure is replaced with the corresponding Ghost module, reducing the model's computational complexity and parameter count without compromising performance. Second, the SE attention mechanism is added to the model, allowing it to capture feature information more accurately and precisely. Next, the ELU activation function replaces the original ReLU activation function. Then, the dataset is augmented and divided into four categories for model training. A model suitable for ginseng grade classification was obtained through experimentation. Compared with classic convolutional neural network models ResNet50, AlexNet, iResNet, and EfficientNet_v2_s, the accuracy improved by 10.22%, 5.92%, 4.63%, and 3.4%, respectively. The proposed model achieved the best results, with a validation accuracy of up to 93.14% and a loss value as low as 0.105. Experiments have shown that this method is effective in recognition and can be used for ginseng grade classification research.

Keywords: deep learning; image processing; ginseng classification; ResNeXt50; Ghostnet; convolutional neural network

1. Introduction

Sun-dried wild ginseng is an essential medicinal herb in traditional Chinese medicine. It possesses various pharmacological activities and nutritional components. It is believed to have a solid qi-boosting effect, enhancing the body's resistance and immune function. Therefore, it is often used to prevent and treat various diseases, particularly chronic illnesses, in individuals with weakened immunity. It is also regarded as an effective remedy for combating fatigue and aging. It is suitable for those experiencing physical and mental fatigue, helping restore energy and delay aging, positively impacting human health. The primary part of ginseng is its root, which has unique medicinal value [1]. The root is long and conical, with a light yellow or brown surface and a firm texture [2].

Traditional classification of wild ginseng relies mainly on morphological characteristics, such as root shape, epidermal color, and internal structure. These methods are time-consuming, labor-intensive, and require high expertise from the identifier, making them susceptible to subjective factors. Although chemical composition analysis can provide accurate classification, the process is complex, costly, and destructive to samples, making it unsuitable for large-scale applications. By employing image classification technology, one can utilize features such as shape, color, and texture from images to assist in determining the variety and quality of ginseng, thus improving classification accuracy and efficiency. Additionally, wild ginseng cultivation is a vast industry where different

varieties and origins significantly vary in market value. Therefore, image classification technology can rapidly identify wild ginseng products, helping consumers purchase products that meet their needs, preventing counterfeit and inferior products, and protecting consumer rights [3]. Finally, as an essential medicinal herb, wild ginseng has significant application value in the informatization management of medicinal materials, production quality control, and scientific research in medicinal material cultivation. By analyzing and processing image data of sun-dried wild ginseng, more botanical features, pharmacological active ingredients, and their correlations can be discovered, aiding in the in-depth study of the pharmacological effects and medicinal value of sun-dried wild ginseng, promoting scientific advancement in related fields [4].

Li et al. enhanced a self-built dataset by introducing an attention mechanism and using a Focal Loss function, addressing the issue of dataset imbalance, ultimately improving the recognition accuracy by 1.72% compared to the original model, with promising experimental results [5]. In 2023, Zhai et al. used a self-built dataset and proposed replacing the original activation function GELU with the PReLU activation function [6]. This change enhanced the nonlinear representation capability of the neural network, significantly improving the model's performance and efficiency and validating the importance of deep learning in medicinal material classification. Liao et al. proposed a rice disease image classification method based on transfer learning built on the VGG19 convolutional neural network [7]. By utilizing the VGG19 network pre-trained on the ImageNet dataset, they transferred and adjusted relevant parameters, establishing a technical process for rice disease image classification and achieving high model accuracy. Wang et al. improved fine-grained feature extraction by integrating features from EfficientNet-B0 and DenseNet121 models and introducing a Focal Loss function combined with label smoothing, enabling accurate identification of small and similar characteristic apple leaf spots in natural environments [8]. Although the improved model is more significant and inference time is longer than a single model, its average precision increased by 12.29%.

In 2017, Howard et al. introduced MobileNets, a lightweight deep neural network based on a streamlined architecture using depthwise separable convolutions. They introduced two simple global hyperparameters for practical trade-offs between latency and accuracy, maintaining strong performance compared to other ImageNet classification models at the time [9]. Xie et al. improved ResNet with the concept of grouped convolutions, resulting in the iterative convolutional neural network ResNeXt, which uses fewer parameters and achieves lower error rates with the same computational load, yielding better results in image classification [10]. V. Krishna Pratap et al. adapted the final fully connected layer and output neurons of EfficientNetB4 to differentiate various chili leaf diseases meticulously, employing data preparation techniques like scaling, pixel normalization, and augmentation to enhance model resilience, enabling the model to recognize image variations in chili leaves due to light, angles, and disease severity, achieving an improved average accuracy of 91.2%, demonstrating deep learning's potential in addressing specific agricultural issues and extending its applications to different crops [11].

In 2023, Li et al. used AlexNet as a base model to study 10 types of medicinal herbs, employing ridge regression and transfer learning to effectively alleviate overfitting and analyze multiple common data, achieving a high recognition accuracy of 95.4% [12]. Han et al. improved the DenseNet-201 model, using 50 types of medicinal slices as a dataset [13]. To prevent overfitting due to excessive parameters in the fully connected layer, they used dropout, randomly setting 50% of the elements, enriching feature learning diversity, enhancing each feature's contribution to model prediction, and strengthening the regularization effect during training. Wang et al. proposed a method for identifying medicinal materials based on an improved TCM-Net, introducing an attention mechanism and improved mobile inverted bottleneck convolution modules, ensuring network lightness while significantly enhancing the accuracy of medicinal material recognition [14]. Zhang et al. used SE-MobileNetV2 as a base model, employing the Momentum optimization algorithm and ReLU6 activation function, training the network model for the

recognition task of traditional Chinese medicine powder microscopic identification images [15]. Experimental results showed an overall recognition accuracy of 97.5% in the eight tested types of medicinal powder microscopic identification images.

Based on previous experiments and literature review, we determined that convolutional neural networks can classify and identify medicinal materials, addressing issues such as excessive reliance on manual classification and traditional methods' relatively low accuracy and slow classification speed. Considering computational load, parameter count, and model size, the ResNeXt50 model was chosen as the base model. This experiment designed an improved ResNeXt50 model to achieve fast and accurate classification of sun-dried wild ginseng grades. The critical improvement involves replacing the three-layer convolution in the Bottleneck structure with the Ghost module, reducing computational load and increasing classification efficiency while ensuring accuracy. Additionally, the SE (Squeeze and Excitation) attention mechanism was used to highlight essential features of sun-dried wild ginseng and suppress irrelevant or unimportant features [16]. The ELU (Exponential Linear Unit) activation function was employed to extract finer-grained features, enhancing the overall recognition accuracy of the model and thereby addressing the challenging problem of classifying sun-dried wild ginseng [17].

2. Data Preprocessing

2.1. Dataset Establishment

After observing various types of ginseng on the market, it was found that under-forest ginseng has diverse shapes and is commonly available across different years, with significant market demand. Therefore, under-forest ginseng was chosen as the experimental subject. The dataset was collected in October 2023 from a herbal medicine store in Changchun. The classing standards referenced the "Authentic Jilin Medicinal Ginseng", including premium-class, first-class, second-class, and ordinary ginseng, subsequently verified by ginseng experts from the College of Traditional Chinese Medicine at Jilin Agricultural University. A Canon EOS R10 mirrorless camera with a 50 mm fixed-focus lens was used. The camera was mounted 1.5 m above the ginseng on a tripod under consistent ambient lighting. Photos were taken at ISO 400, a shutter speed of 1/200 s, and an aperture of 2.8, resulting in 4000 × 6000 high-definition images. The details of the under-forest ginseng are visible, and photos were captured from multiple angles. A white A3-sized wallpaper was used as the background. In total, 152 premium-class, 154 first-class, 154 s-class, and 154 ordinary sun-dried wild ginseng images were collected. The grading standards for sun-dried wild ginseng are shown in Table 1.

Table 1. Rating criteria.

Project	Premium-Class Ginseng	First-Class Ginseng	Second-Class Ginseng	Ordinary Ginseng
Main Root	Cylindrical shape with a thinner top and a thicker bottom			
Cross-section	Yellowish white in section, powdery, with resinous tract visible			
Texture	Harder, powdery, non-hollow			
Damage, scars	Almost no injury	No significant injury	Minor injury	More obvious
Section	Section neat, clear	Segment is obvious	Segment is a little bit obvious	Segment is not obvious
Odor	Unique aroma, taste slightly bitter, sweet			

2.2. Data Augmentation

In deep learning research, data augmentation can expand the dataset by generating additional training samples. This is particularly important when the data are limited, providing more samples for training the model. Augmenting the data introduces more variability and diversity, allowing the model to better adapt to different inputs. This helps improve the model's generalization ability and performance on real-world data. It can also mitigate overfitting by introducing randomness and variation.

During the dataset collection, capturing the complete environment was challenging, and the total dataset size was relatively small, so data augmentation was performed. This study

used offline augmentation, employing random brightness, contrast, Gaussian noise, and salt-and-pepper noise to expand the dataset. Random contrast factor ranges from 0.1 to 2, random brightness factor ranges from 0.5 to 1.5, random salt-and-pepper noise percentage ranges from 0.1 to 0.5, random mean for Gaussian noise ranges from 0.01 to 0.1, and random standard deviation ranges from 0.05 to 0.1. After offline augmentation, the data were expanded to 3070 images. Information about the dataset is shown in Figure 1.
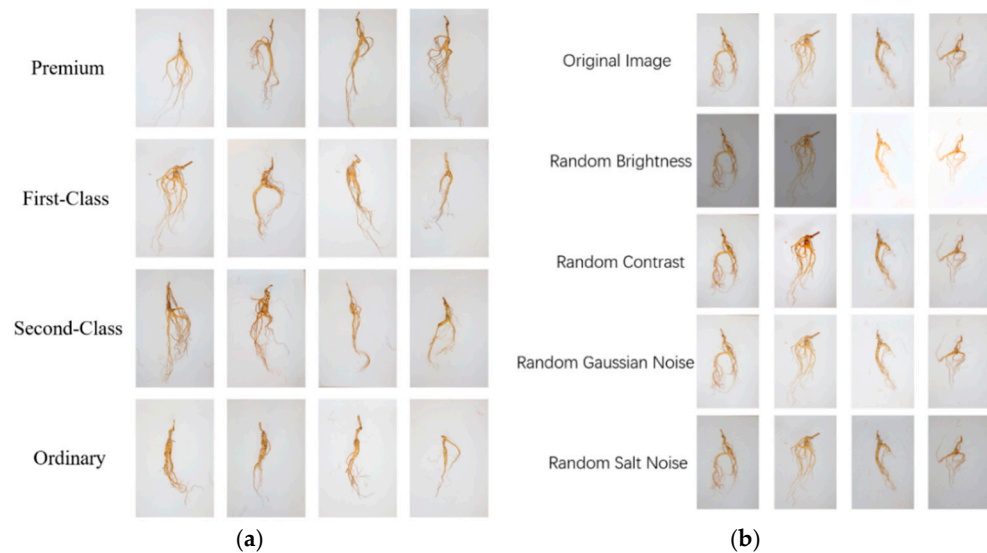
Figure 1. Ginseng dataset. (**a**) Different levels of ginseng images; (**b**) sample image after data enhancement.

2.3. Data Partitioning

This study used five-fold cross-validation to evaluate the model's performance and prevent the results of a single experiment from being coincidental. This approach also provides a more comprehensive assessment of the model's generalization ability, randomly selecting 80% of the images from the dataset as the training set and 20% as the test set, each time using 614 images for testing and 2456 images for training, ensuring no overlap in test data. After partitioning, the training and test sets for each grade of ginseng are as follows: 608 and 152 images for particular grade ginseng; 616 and 154 images for first-, second-, and common-grade ginseng. The final result of this experiment is the average of the five experiments. Since the dataset is relatively small, dividing out a validation set could lead to insufficient data in both the training and test sets, affecting the model's training effectiveness and evaluation accuracy, so no validation set was created.

3. To Construct the Network Model

3.1. ResNeXt50 Model

ResNeXt is an improved version of ResNet, enhancing network expressiveness by introducing the concept of cardinality [18]. Unlike traditional ResNet, which relies on increasing depth to improve performance, ResNeXt focuses on increasing network width. This approach enhances performance while avoiding a significant increase in the number of parameters and computational complexity. Specifically, ResNeXt enhances expressiveness by increasing the number of branches in each residual block without significantly increasing parameters and computational complexity. In the ResNeXt architecture, a 256-channel input feature is divided into 32 groups; each is first compressed to 4 channels before processing. The processed 32 groups are then summed and output through a residual connection with the original features. In this context, cardinality refers to the number of identical branches in each block. ResNeXt employs grouped convolution, dividing feature maps into multiple groups and performing convolution operations on each group separately,

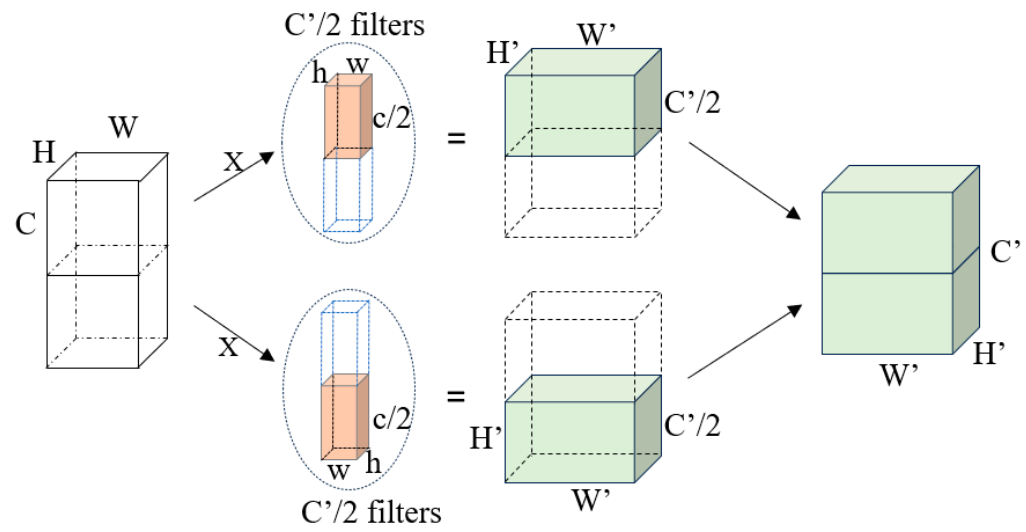effectively reducing computational complexity. The principle of grouped convolution is illustrated in Figure 2.

Figure 2. Group convolution.

3.2. Ghost Module

The Ghost module is derived from the GhostNet model released by Han et al. in 2020, which achieves a good balance between computational efficiency and model accuracy [19]. When designing deep neural networks, many feature maps are typically included, even with some redundancy, to ensure a comprehensive analysis of the input data. Analyzing the feature maps processed by the first residual block of ResNeXt50 shows noticeable similarities among some feature maps. This indicates that specific feature maps can be generated through simple operations from another, meaning one can be considered a "ghost" of another. Therefore, it can be inferred that not all feature maps need to be generated through convolution operations; some "ghost" feature maps can be obtained through lower-cost operations. Assuming the input feature is $X \in \mathbf{R}^{c \times h \times w}$, the convolution kernel is represented as $f \in \mathbf{R}^{c \times k \times k \times n}$, and the output feature is $Y \in \mathbf{R}^{h' \times w' \times n}$, the regular convolution operation can be expressed as follows:

$$Y = X * f + b \tag{1}$$

h' and w' are the height and width of the output data, and $k \times k$ is the kernel size of the convolution kernel f. In this convolution process, the number of kernels n and channels c are typically large (e.g., 256 or 512). The computational cost of this operation is $n \cdot h' \cdot w' \cdot c \cdot k \cdot k$. According to the Formula (1), the computational cost of the entire operation is closely related to the dimensions of the input and output feature maps. However, considering the previous findings, it can be concluded that the output features contain many similar "ghost" features. Therefore, using a large number of floating-point operations (FLOPs) and parameters to generate these redundant feature maps is unnecessary. Assume that only m original feature maps $Y' \in \mathbf{R}^{h' \times w' \times m}$ are generated using a single convolution:

$$Y' = X * f' \tag{2}$$

$f' \in \mathbf{R}^{c \times k \times k \times m}$ is the convolution kernel used, where $m \leq n$. For simplicity, the bias term is omitted here. Additionally, all hyperparameters (e.g., filter size, stride, padding) remain consistent with those in regular convolution to ensure that the spatial dimensions h' and w' of the output features remain unchanged. At this point, we obtain only m feature

maps. To extend this to n feature maps, each feature in Y′ undergoes a low-cost linear transformation to generate s "ghost" feature maps.

$$y_{ij} = \Phi_{i,j}(y_i'), \forall i = 1, \ldots, m, j = 1, \ldots, s, \tag{3}$$

yi′ represents the i-th original feature map, which undergoes a linear transformation $\Phi_{i,j}$ to produce the j-th "ghost" feature map y_{ij}. This means each original feature map *yi′* can generate s "ghost" feature maps $\{y_{ij}\}_{j=1}^s$. The final $\Phi_{i,s}$ is an identity mapping used to retain the original feature map, as shown in Figure 3.
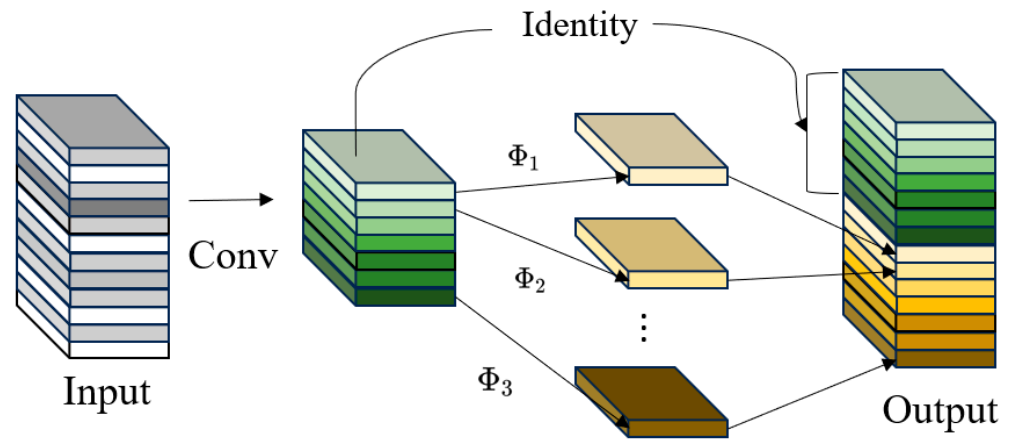

Figure 3. Ghost module.

3.3. Using the ELU Activation Function

The ReLU (Rectified Linear Unit) activation function was proposed to address the vanishing gradient problem [20]. The gradient of the ReLU function is either 0 or 1, and it truncates negative values to 0, introducing sparsity into the network and further enhancing computational efficiency. Its function is expressed in Equation (3).

$$\text{ReLU}(x) = \begin{cases} 0, x \leq 0 \\ x, x > 0 \end{cases} \tag{4}$$

Although sparsity can improve computational efficiency, it may also hinder training. As shown in Figure 4, the input value of the activation function includes a bias term. Suppose this bias term becomes too small, causing the input to the activation function to always be negative. In that case, the gradient through this node during backpropagation will always be 0, preventing the related weights and bias parameters from being updated. If the input values for all samples in this activation function are negative, the neuron will no longer be able to learn. This phenomenon is known as the "neuron death" problem.

The ELU is an activation function used for neuron activation in neural networks. It is an improvement upon existing functions. Its function is expressed in Equation (4).

$$\text{ELU}(x) = \begin{cases} \alpha(e^x - 1) & , x < 0 \\ x & , x \geq 0 \end{cases} \tag{5}$$

In the ELU function, α is a hyperparameter that controls the slope in the negative value region. For inputs more significant than 0, the gradient is 1 and for inputs less than 0, it asymptotically approaches $-\alpha$, with α typically set to 1. Compared to ReLU, the ELU function is smoother in the negative region, meaning its gradient does not experience abrupt changes, which helps enhance the stability of the training process. Additionally, the output in the negative region can approach negative infinity, which helps reduce neuron

output saturation and avoids the vanishing gradient problem. Moreover, the mean of the ELU function over the entire input range is close to zero, accelerating the network's convergence speed.
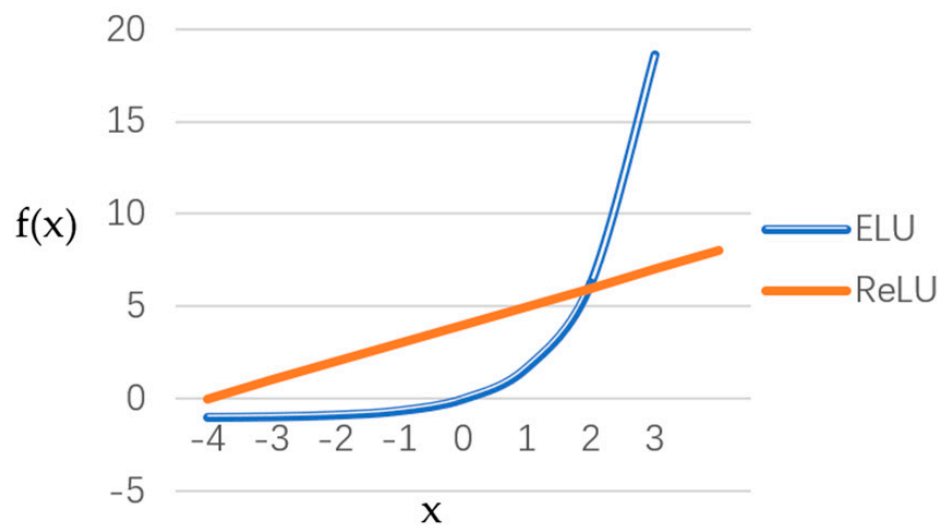

Figure 4. Comparison of the ReLU and Leaky ELU functions.

3.4. SE Attention Mechanism

The core idea of the SE attention mechanism is to adaptively adjust channel weights, enabling the model to focus more on essential feature channels. This mechanism mainly consists of two core steps: squeeze and excitation. In the squeeze phase, the SE attention mechanism uses global average pooling to convert each channel's feature map into a single scalar, reflecting the channel's importance across the entire feature map. In the excitation phase, a fully connected layer maps the scalar generated in the squeeze phase to a new weight vector, representing the importance of each channel. Finally, this weight vector is multiplied element-wise with the original feature map to produce a weighted feature map. This approach allows the model to focus more on essential feature channels, enhancing classification performance. The SE attention mechanism is advantageous due to its simple and efficient structure, making it easy to integrate into existing convolutional neural network (CNN) architectures. It has significantly improved performance in various image classification tasks, becoming a widely adopted attention mechanism. Figure 5 illustrates its working principle.
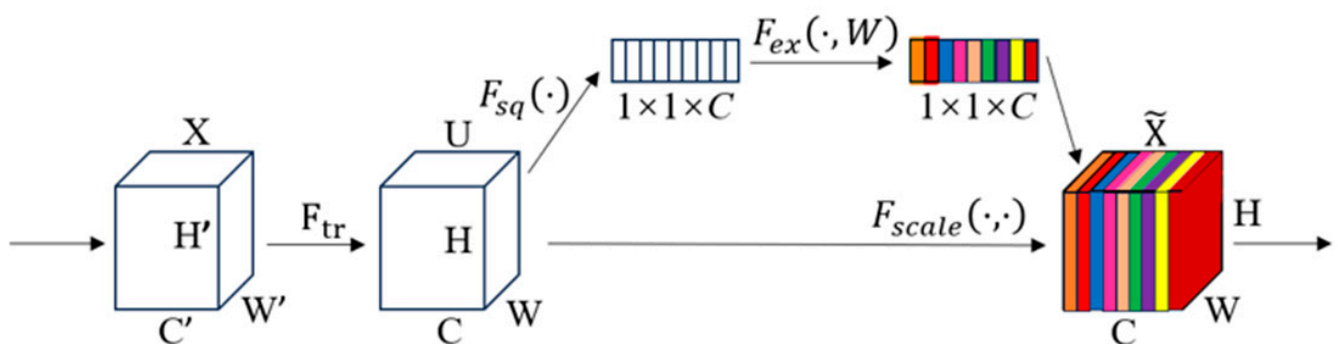

Figure 5. Squeeze and excitation networks.

(1) Input Features: For any given transformation, such as a convolution, which maps the input X to $U\left(U^{H \times W \times C}\right)$, a corresponding SE block can be constructed to perform feature recalibration.

(2) Calculation Method: In the squeeze phase, the feature map U undergoes a squeezing operation, aggregating feature mappings over its spatial dimensions (H × W) to generate

channel descriptors. The primary function of this descriptor is to create a global distribution embedding of channel feature responses, allowing global receptive field information from the network to be utilized by all layers. Specifically, this process involves applying global average pooling to the input feature map to achieve compression along the channel dimension. Assuming the shape of the input feature map is (C, H, W), where C represents the number of channels, and H and W are the height and width of the feature map, respectively, a tensor of shape (C, 1, 1) is obtained through global average pooling. The calculation formula for the squeeze phase is as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j) \tag{6}$$

In the excitation phase, a simple adaptive channel selection mechanism is employed. This mechanism uses the embedding as input to generate a set of modulation weights for each channel. These weights are then applied to the feature map U to produce the output of the SE block, which can be directly passed to subsequent layers of the network. Specifically, the tensor of shape (C, 1, 1) obtained from the previous step is fed into a fully connected layer, reducing the number of channels to a smaller value, referred to as middle channels. After compression, the features are input into a ReLU activation function, resulting in a vector of shape (middle_channels). The calculation formula for the excitation phase is as follows:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \tag{7}$$

(3) During the output feature process, the previously obtained excitation vector is expanded into a tensor of shape (C, 1, 1) and multiplied element-wise with the original feature map X. This operation results in a weighted feature map for each channel, effectively enhancing feature representation capabilities. The calculation formula for the output features is as follows:

$$\widetilde{X}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \tag{8}$$

Wild sun-dried ginseng roots have dense and varied shapes and sizes, which can affect a network model's ability to extract detailed features. The experiment chose to place the SE attention mechanism near the input layer of the network model. This allows the model to focus on and learn essential features in the input data at an earlier stage.

3.5. Improvement in ResNeXt50 Model Structure

The Bottleneck structure in the ResNeXt50 model is the core of its repeating blocks and involves multiple convolution operations. Typically, the Bottleneck structure includes three layers of convolution: a 1×1 convolution for dimensionality reduction, a 3×3 convolution for feature extraction, and a final 1×1 convolution for dimensionality restoration. Considering the design intention of the Ghost Module is to generate equivalent feature maps with less computation, we propose replacing each convolution layer in the Bottleneck structure with a corresponding Ghost Module. First 1×1 Convolution (Dimensionality Reduction): since this layer aims to reduce the number of channels in the feature map, we replace it with a Ghost Module to achieve the same purpose but with less computation. Middle 3×3 Convolution: This layer is the core of the Bottleneck, used for feature extraction. We replaced it with a Ghost Module, utilizing group convolution to enhance efficiency further. Final 1×1 Convolution (Dimensionality Restoration): This layer restores the number of channels in the feature map to its original size. We also replaced it with a Ghost Module to restore dimension while maintaining low computational cost. The ReLU activation function is replaced after batch normalization in these three layers with the ELU activation function for more stable gradient calculation, which aids in optimizing optimization algorithms, allowing the model to learn and converge faster. The SE attention mechanism is inserted after layer 2 and layer 3 of ResNeXt50 to dynamically adjust the importance of each channel. This highlights more helpful

information and suppresses irrelevant or redundant data. The structure of the improved ResNeXt50 model proposed in this study is shown in Figure 6.
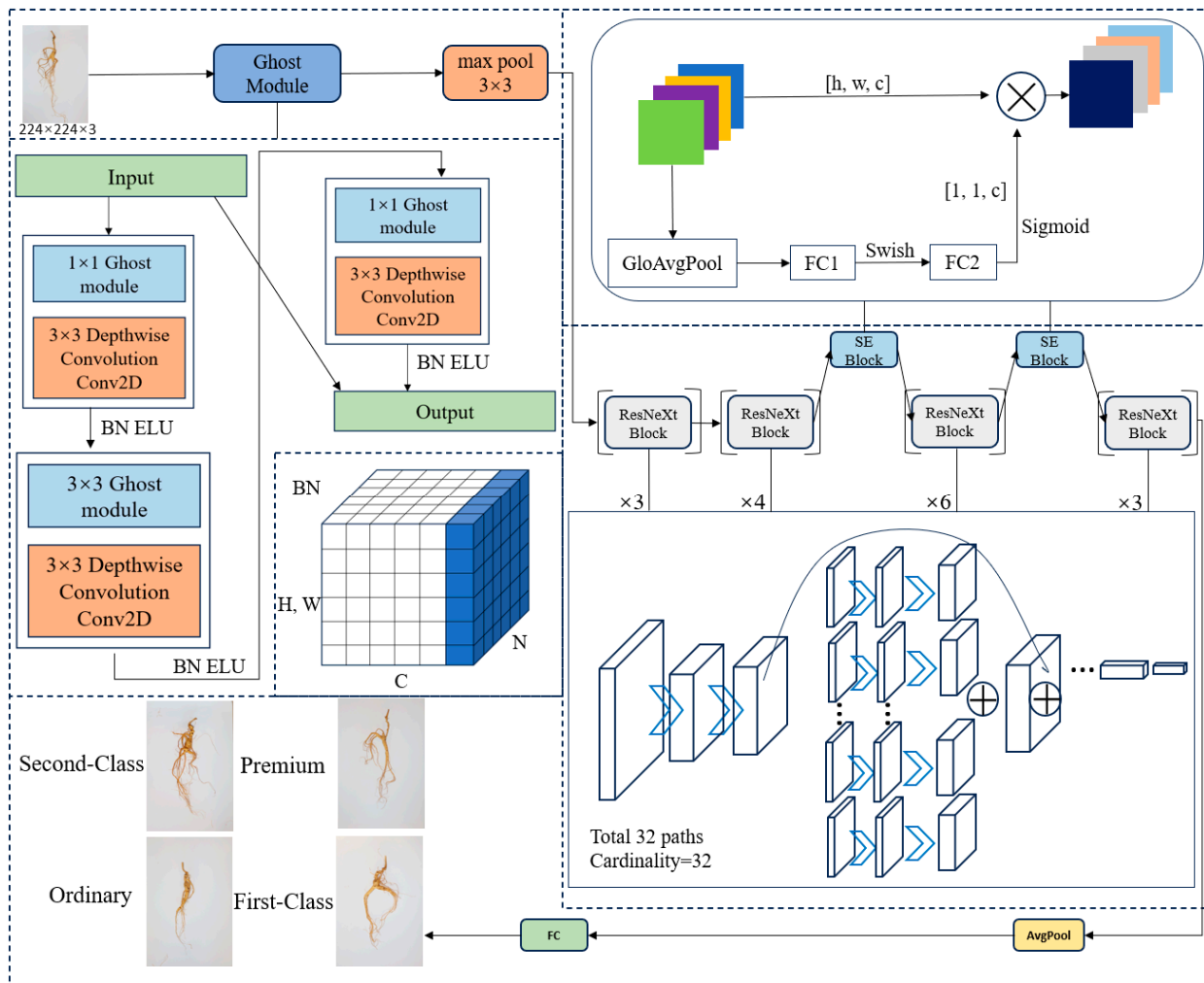

Figure 6. Improved ResNeXt50 model structure.

3.6. Experimental Setup

This experiment was conducted using PyTorch. The experimental setup includes a 13th Gen Intel (R) Core (TM) i9-13980HX processor, an NVIDIA GeForce RTX 4080 Laptop GPU, and 32GB of RAM. The software environment was configured with Windows 11, Python 3.9.0, PyTorch 2.0.0, and CUDA 11.8. The input image size was fixed at 224×224 pixels. Each experiment ran for 200 epochs with a batch size of 32. The loss function used was Focal Loss [21], the optimizer was Adam [22], and the activation function was ELU.

4. The Influence of Different Factors on Experimental Results

4.1. The Impact of Attention Mechanism on Model Performance

By controlling other variables, different attention mechanisms were integrated separately: CBAM (Convolutional Block Attention Module) [23], ECA (Efficient Channel Attention) [24], SK (Selective Kernel) [25], and SE attention mechanism. The loss values decreased by 0.105, 0.258, 0.222, and 0.279 points, respectively, while the recognition accuracy increased by 0.48, 2.32, 3.26, and 6.43 percentage points, respectively. As shown in Table 2, the SE attention mechanism resulted in the highest recognition accuracy and the lowest loss value, providing the greatest overall improvement to the network model. Compared to other attention mechanisms, its effect was more significant. Therefore, it is necessary to compare various mechanisms to identify the most suitable attention mechanism for this experiment.

Table 2. Comparison of experimental results on multiple attention mechanisms.

Attention Mechanism	Accuracy %	Loss	Convergence Rounds	Training Time per Round/s	Parameter/M
No Attention	86.71	0.384	77	83	74.47
CBAM	87.18	0.279	72	80	89.93
ECA	89.02	0.126	71	79	84.54
SK	89.96	0.162	69	77	95.48
SE	93.14	0.105	63	76	82.17

4.2. The Impact of Activation Function on Model Performance

Before adding activation functions, the inputs and outputs are linear combinations, similar to having no hidden layers, which makes the network less likely to converge and limits its learning ability, much like the original perceptron. After adding activation functions, neural networks can learn smooth curves to segment the plane, rather than using complex linear combinations to approximate smooth curves, thus enhancing the network's expressive capability and better fitting the target function. In this experiment, with other conditions unchanged, the ELU activation function was replaced with ReLU, Tanh [26–28], Sigmoid [29,30], and Leaky ReLU [31], respectively. As shown in Table 3, when using Sigmoid and Leaky ReLU, there were noticeable differences in recognition accuracy and loss values compared to using ReLU. Compared to ReLU, using the ELU activation function increased recognition accuracy by 6.26% and reduced the loss value by 0.143, indicating a significant improvement. This improvement is due to ELU's non-zero output for negative inputs, which avoids the problem of dead neurons, keeping neurons active throughout training. Additionally, ELU's exponential growth in the negative region results in a smoother curve.

Table 3. Comparison of experimental results of multiple activation functions.

Activation Function	Accuracy %	Loss	Parameter/M
Sigmoid	79.11	0.857	85.59
Leaky ReLU	84.55	0.681	86.74
ReLU	86.88	0.248	83.27
Tanh	89.72	0.173	87.29
ELU	93.14	0.105	82.17

4.3. The Impact of Ghost Module on Model Performance

When using the original Bottleneck of the ResNeXt50 model, the parameter count is significantly more significant, as all feature maps need to be generated through convolution operations, resulting in higher computational cost and lower efficiency. After replacing the Ghost module, feature maps can be formed from another feature map through simple operations, referred to as ghost feature maps, which enhance computational efficiency and reduce the parameter count. Table 4 compares the data before and after replacing the Ghost module while keeping other improvements unchanged. The model's performance significantly improved after the replacement.

Table 4. Comparison before and after replacing Ghost module.

| Bottleneck	Accuracy %	Loss	Training Time per Round/s	Parameter/M
ResNeXt50	87.73	0.131	79	92.53
Ghost Module	93.14	0.105	76	82.17

5. Results
5.1. Experimental Result

To demonstrate the effectiveness of the experimental model, comparisons were made with several classic convolutional neural network models: ResNet50, AlexNet [32], iRes-

Net50 [33], and EfficientNet_v2_s [34]. As shown in Table 5, the improved model shows significant advantages in terms of convergence epochs and training time per epoch. Its accuracy increased by 10.22%, 5.92%, 4.63%, and 3.4% compared to these four classic models, while the loss values decreased by 0.261, 0.103, 0.03, and 0.007, respectively. The parameter count was reduced by 68% compared to AlexNet, and both convergence epochs and training time per epoch were reduced. Figure 7 compares the accuracy and loss values of each model.

Table 5. Comparison of experimental results of multiple models.

| Model	Accuracy/%	Loss	Converg-Ence Rounds	Training Time per Round/s	Parameter/M
ResNet50	82.92	0.366	80	78	98.75
AlexNet	87.22	0.208	96	77	233.08
iResNet50	88.51	0.135	85	81	97.49
Effcientnet_v2_s	89.74	0.112	77	79	83.46
Our model	93.14	0.105	63	76	82.17
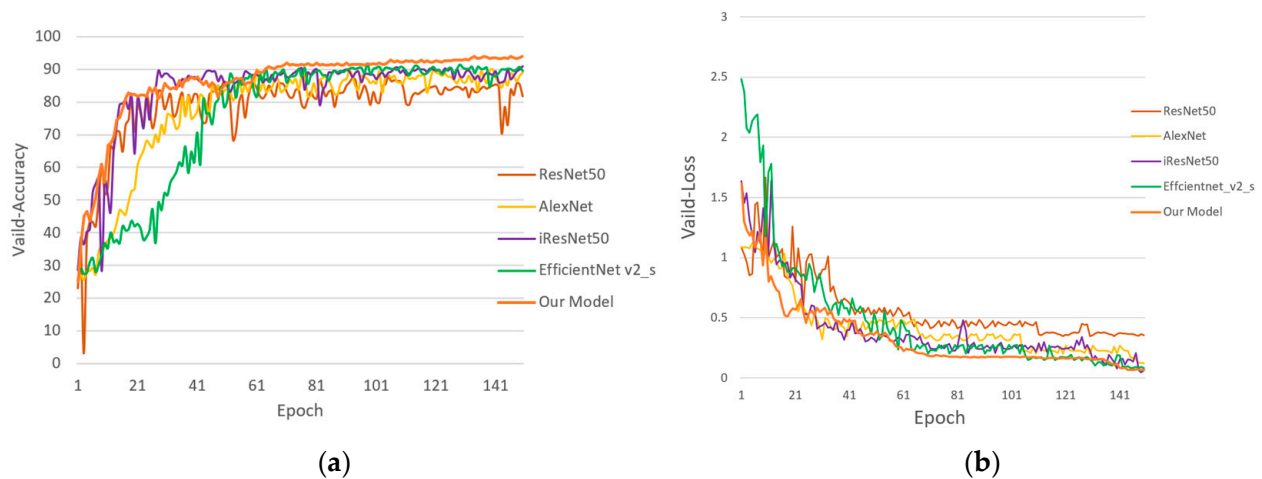
Figure 7. Experimental results of each model: (**a**) Model Accuracy; (**b**) Model Loss.

To make the model's position more transparent and convincing, we used the Saposhnikovia divaricata dataset from Li et al., consisting of 2001 images categorized into five types based on their origin [35]. Below are the conclusions drawn from comparisons: experiments demonstrate that the model proposed in this study is superior to the improved iResNet model. Furthermore, the morphological characteristics of sun-dried wild ginseng are more complex and variable, with interwoven roots, providing sufficient evidence that the improved ResNeXt50 model has better expressive ability and performance. The experimental results are shown in Table 6.

Table 6. Comparison results using the Saposhnikovia divaricata dataset.

Model	Accuracy	Loss	FLOPs	Inference Time	FPS
Our Model	94.91%	0.1478	3.21 Gmac	6.05 ms	165.2 fps
Improved iResNet	93.72%	0.2192	3.32 Gmac	6.29 ms	158.9 fps

Heatmaps generated using the Grad-CAM [36] tool can visually display the areas the model focuses on when processing images, aiding in understanding the model's decision-making process. Figure 8 compares the heatmaps of the improved model. The improved model exhibits more refined feature extraction and focus, with more accurate attention to regions and a stronger ability to extract ginseng features.
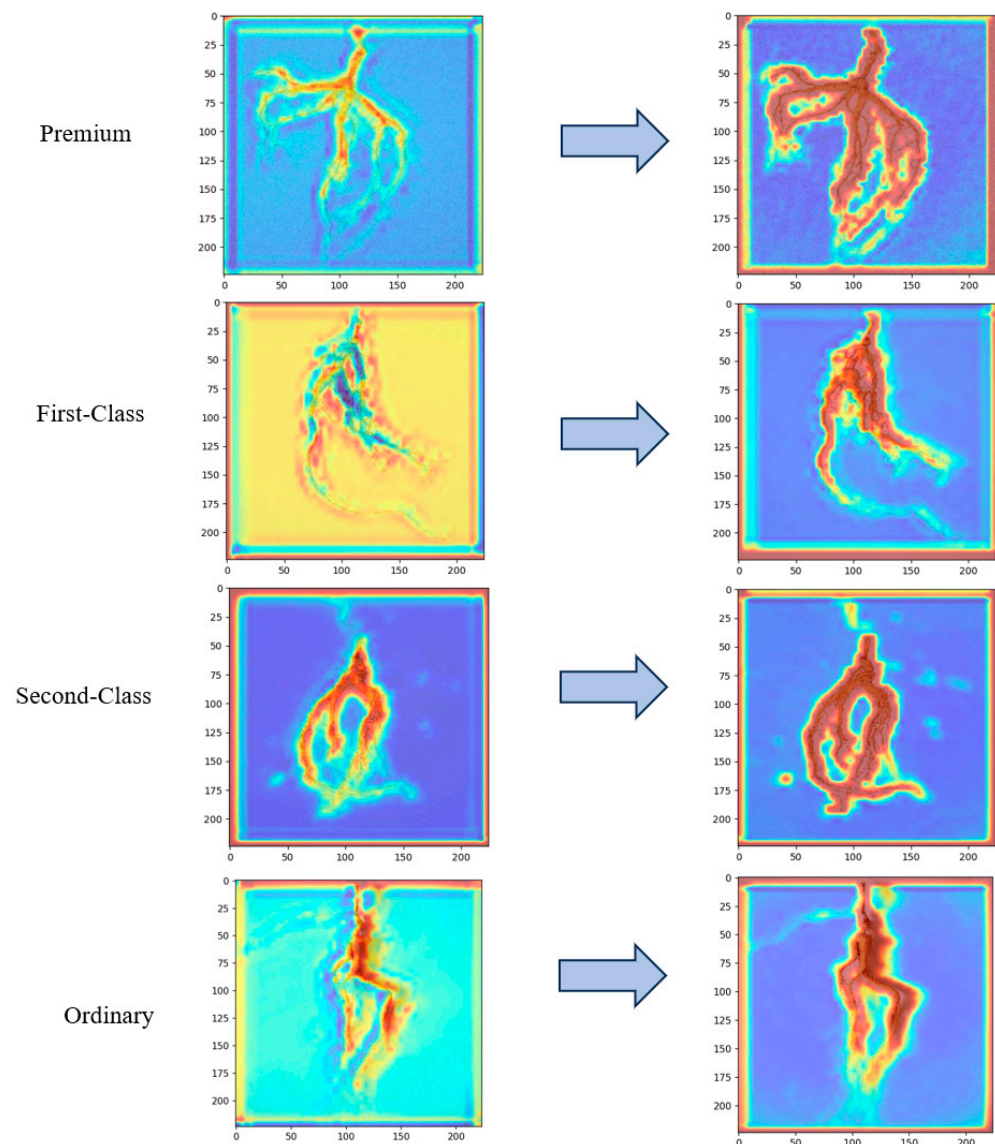
Figure 8. Comparison of thermal characteristic maps before and after model improvement.

5.2. Model Evaluation

The prediction results of the classification task for sun-dried wild ginseng grades in this experiment will be represented using a confusion matrix. The confusion matrix provides a comprehensive evaluation metric for model performance. It clearly shows the classification results for each category, including True Positives, False Positives, True Negatives, and False Negatives. These metrics help assess the model's accuracy, recall, precision, and other performance indicators.

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \qquad (9)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (10)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (11)$$

In the above equation, TP (True Positive) is the number of correctly identified sun-dried wild ginseng samples, FP (False Positive) is the number of incorrectly identified samples, FN (False Negative) is the number classified as other types, and TN (True Negative) is the number of other correctly identified ginseng samples. In the confusion matrix of Figure 9,

0 represents premium sun-dried wild ginseng, 1 represents first-grade, 2 represents second-grade, and 3 represents regular. The horizontal axis indicates the true category, while the vertical axis indicates the predicted category. From the confusion matrix in Figure 8, it can be seen that the original ResNeXt50 model achieved an accuracy of 83.38% and a recall of 83.45% on the test set. The improved model achieved an accuracy of 93.14% and a recall of 91.75%. As shown in Table 7, the improved model significantly enhances the recognition accuracy, recall, and F1 score for the four types of sun-dried wild ginseng compared to the original model. The confusion matrix shows that the misclassification rate for ordinary ginseng is higher than for other types of ginseng. By visualizing many misclassified samples using the trained model weights, we identified a set of representative misclassified samples [37]. As illustrated in Figure 10, ordinary ginseng was misclassified once as second-class ginseng and twice as premium-class ginseng. The main reasons for this include ginseng's complex and variable shape, making it challenging to accurately and effectively capture some root features. Additionally, the ELU introduces a hyperparameter α that requires tuning and optimization [38]. Although this process can be adjusted through experimentation, it still adds complexity to the model tuning and hyperparameter optimization process. In subsequent experiments, narrow the tuning range of the α value and lock in on a threshold with good experimental performance. Restrict the search range to this interval to reduce search space and time.
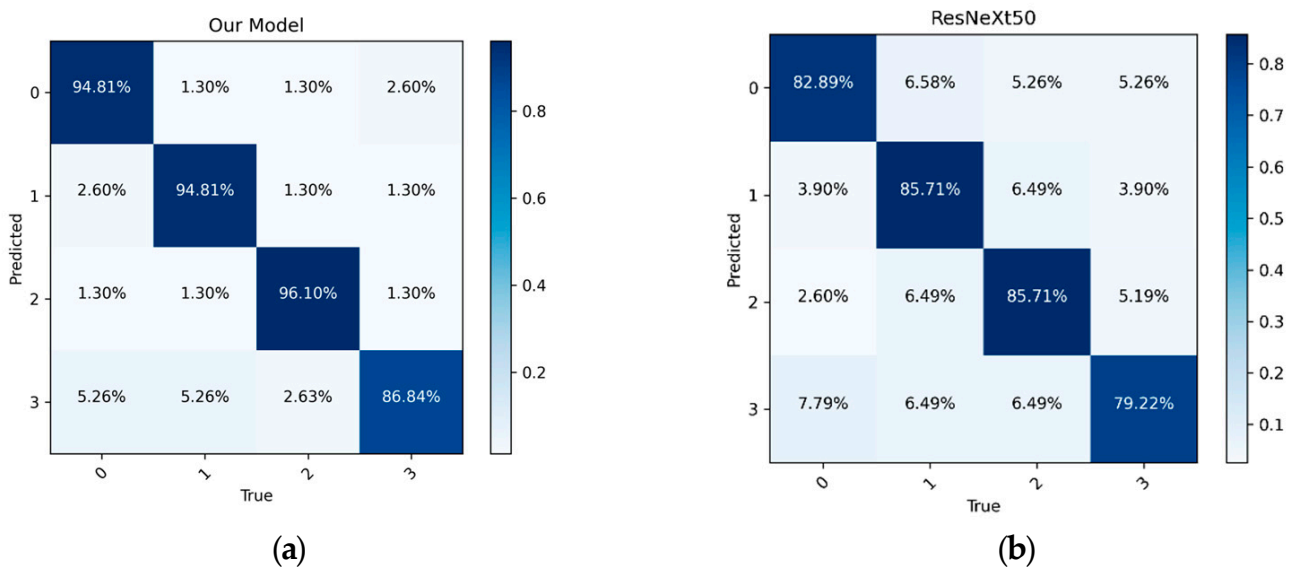

(a)

(b)

Figure 9. Confusion matrix before and after model improvement. (**a**) The confusion matrix of improved model; (**b**) confusion matrix of ResNeXt50.
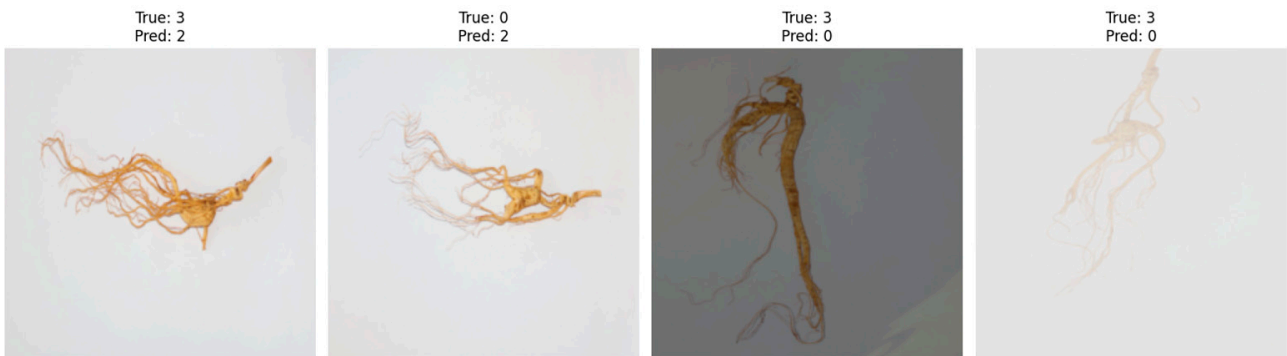

Figure 10. Visualization of misclassified samples.

Table 7. Recall, Precision, and F1 Scores for each type of ginseng before and after model improvement.

Classification	Model	Recall%	Accuracy %	F1/%
Premium	ResNeXt50	85.13	82.89	83.99
	Our Model	86.38	94.81	90.39
First-Class	ResNeXt50	81.48	85.71	83.54
	Our Model	97.34	94.81	96.05
Second-Class	ResNeXt50	82.50	85.71	84.07
	Our Model	92.21	96.10	94.11
Ordinary	ResNeXt50	84.72	79.22	81.87
	Our Model	91.10	86.84	88.91

6. Discussion

By classifying ginseng, a scientific quality evaluation system can be established to help the ginseng industry with quality control, ensuring product quality and safety. Using deep learning technology dramatically saves labor and avoids the subjectivity of manual classification, thereby improving the efficiency of ginseng classification and enabling sustainable development of the ginseng industry. This paper proposes an improved ResNeXt50 network model tailored to the features of forest-grown ginseng. The conclusions drawn from the experiments are as follows:

1. A ginseng dataset was constructed and classified into four different grades: premium-class, first-class, second-class, and ordinary ginseng. Data augmentation was performed on the dataset to enhance its diversity. ResNeXt50 was chosen as the base model, with the Ghost module replacing the three-layer convolution of the Bottleneck structure, maintaining excellent performance while reducing computational load. The ReLU activation functions after the three-layer convolution were replaced with ELU activation functions to avoid dead neurons and accelerate convergence. SE attention mechanisms were added to the second and third layers of the model to capture key ginseng features more accurately, improving the model's accuracy and generalization ability.

2. The improved ResNeXt50 model achieved an accuracy of 93.14% and a recall of 91.75% on the self-constructed dataset. Its parameter size was 74.47MB, and each training epoch took only 76 s. The accuracy and F1 score improved by 9.76% and 8.99%, respectively.

3. The misclassification rate for ordinary ginseng is relatively high. Through an objective model analysis, the primary direction for future experiments is to enhance the model's expressive capabilities further and balance the quantity of each class in the dataset, thereby increasing the correct classification rate for ordinary ginseng.

The improved model shows significant accuracy parameter size advantages and commendable convergence speed, contributing to the development of ginseng classification. In summary, the improvements positively enhance the model, demonstrating the feasibility of modifying the original model.

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

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The dataset was collaboratively created by the team and is also a unique aspect of this study, therefore, it is not publicly available.

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

References

1. Yu, X.; Feng, X.; Zhang, J.; Huang, J.; Zhang, Q. Research progress on chemical constituents and pharmacological effects of Panax ginseng. *Res. Ginseng* **2019**, *31*, 47–51.
2. Wang, F. Research on Processing Technology and Products of Ginseng Root. *China Food Semimonthly Mag.* **2024**, *53*, 124–126.
3. Kong, F.; Xu, S.; Lu, H.; Cao, S.; Liu, J.; Li, Z.; Sun, W. Exploring Key Technologies for Intelligent Production of Authentic Ginseng, Rooted in Its Three Major Values. *Spec. Wild Econ. Anim. Plant Res.* **2023**, *62*, 1–6.
4. Zhou, C.; Zhao, F.; Di, J.; Cao, S.; Zhang, C.; Zhang, H.; Kong, F. The Application of Mechanized Production in Ginseng Planting and Origin Processing. *Spec. Res.* **2022**, *44*, 161–163.
5. Li, D.; Piao, X.; Lei, Y.; Li, W.; Zhang, L.; Ma, L. A grading method of ginseng (Panax ginseng CA Meyer) appearance quality based on an improved ResNet50 model. *Agronomy* **2022**, *12*, 2925. [CrossRef]
6. Zhai, M.; Zhang, L.; Piao, X.; Li, W.; Li, D. A Ginseng Appearance Quality Grading Method Based on Modified ConvNeXt Model. *J. Jilin Agric. Univ.* **2023**, *45*, 791–802.
7. Liao, L.; Han, C.; He, C. Rice Diseases Image Classification Method Based on VGG19 and Transfer Learning. *Surv. Mapp.* **2023**, *46*, 153–157.
8. Wang, R.; Chen, F.; Zhu, X.; Zhang, X. Identifying apple leaf diseases using improved EfficientNet. *Trans. Chin. Soc. Agric. Eng.* **2023**, *39*, 201–210.
9. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
10. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
11. Pratap, V.K.; Kumar, N.S. High-precision multiclass classification of chili leaf disease through customized EffecientNetB4 from chili leaf images. *Smart Agric. Technol.* **2023**, *5*, 100295. [CrossRef]
12. Li, W.; Wu, L. Image Recognition of Chinese Medicinal Materials Based on Improved AlexNet. *Softw. Eng.* **2023**, *26*, 38–41.
13. Han, Y.; Lan, J.; Guo, R.; Xing, C.; Huang, X.; Huo, Y. Identification of Chinese Herbal Medicine Slices Based on Deep Learning. *Acta Agric. Boreali-Occident. Sin.* **2023**, *32*, 1859–1867.
14. Wang, Q.; Dong, N.; Yang, Y.; Liu, P. Key TCM identification techniques based on image processing and deep learning. *Autom. Instrum.* **2023**, *43*, 30–35.
15. Zhang, X.; Wang, L.; Li, L. Key techniques of image processing and in-depth learning for identification of Chinese medicinal materials. *Autom. Instrum.* **2023**, *43*, 143–147.
16. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 2011–2023. [CrossRef]
17. Clevert, D.; Unterthiner, T.; Hochreiter, S. Fast and accurate deep network learning by exponential linear units (elus). *arXiv* **2015**, arXiv:1511.07289.
18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
19. Han, K.; Wang, Y.H.; Tian, Q.; Guo, J.Y.; Xu, C.J.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
20. Agarap, A.F. Deep learning using rectified linear units (relu). *arXiv* **2018**, arXiv:1803.08375.
21. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
22. Bock, S.; Goppold, J.; Weiß, M. An improvement of the convergence proof of the ADAM-Optimizer. *arXiv* **2018**, arXiv:1804.10587.
23. Woo, S.; Park, J.; Lee, J.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the 2018 European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
24. Wang, Q.L.; Wu, B.G.; Zhu, P.F.; Li, P.H.; Zuo, W.M.; Hu, Q.H. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
25. Li, X.; Wang, W.H.; Hu, X.L.; Yang, J. Selective kernel networks. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
26. Xu, Z.; Li, Z. Extended Tanh-function Method and its Applications. *J. OF Guangxi Univ. Natl. (Nat. Sci. Ed.)* **2009**, *15*, 54–56.
27. Peng, Z.; Xie, B. Wavelet Threshold Denoising Algorithm Based on Tanh Function. *J. Hunan Inst. Eng. (Nat. Sci. Ed.)* **2022**, *32*, 35–40.
28. Gu, X.; Guan, Q.; Yu, Z. Absolute Value Circuit for Tanh Activation Function in Computing in Memory. *J. Electron. Inf. Technol.* **2023**, *45*, 3350–3358.

29. Fan, K.; Qiu, H. Robust Nonnegative Least Mean Square Algorithm Based on Sigmoid Framework. *J. Electron. Inf. Technol.* **2021**, *43*, 349–355.

30. Pan, B.; Zhang, Z.; Zhou, Q.; Ma, Z.; Wang, C. Continuous diversion spacing of ramps based on sigmoid lane-changemodel. *J. Chang Univ. (Nat. Sci. Ed.)* **2023**, *43*, 37–48.

31. Xu, J.; Li, Z.S.; Du, B.W.; Zhang, M.M.; Liu, J. Reluplex made more practical: Leaky ReLU. In Proceedings of the 2020 IEEE Symposium on Computers and Communications (ISCC), Rennes, France, 7–10 July 2020.

32. Alom, M.Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S.; Van Esesn, B.C.; Awwal, A.A.S.; Asari, V.K. The history began from alexnet: A comprehensive survey on deep learning approaches. *arXiv* **2018**, arXiv:1803.01164.

33. Duta, I.C.; Liu, L.; Zhu, F.; Shao, L. Improved residual networks for image and video recognition. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021.

34. Tan, M.; Le, Q. Efficientnetv2: Smaller models and faster training. In Proceedings of the 2021 International Conference on Machine Learning, Virtual, 18–24 July 2021.

35. Li, D.; Yang, C.; Yao, R.; Ma, L. Origin identification of Saposhnikovia divaricata by CNN Embedded with the hierarchical residual connection block. *Agronomy* **2023**, *13*, 1199. [CrossRef]

36. 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 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.

37. Alsallakh, B.; Hanbury, A.; Hauser, H.; Miksch, S.; Rauber, A. Visual methods for analyzing probabilistic classification data. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 1703–1712. [CrossRef]

38. Kim, J.; Cho, S. Evolutionary optimization of hyperparameters in deep learning models. In Proceedings of the 2019 IEEE Congress on Evolutionary Computation (CEC), Wellington, New Zealand, 10–13 June 2019.