



An Enhanced Convolutional Neural Network for Accurate Classification of Grape Leaf Diseases



Yinglai Huang¹, Ning Li¹, Zhenbo Liu^{2*}

¹ College of Information and Computer Engineering, Northeast Forestry University, 150040 Harbin, China

² Material Science and Engineering College, Northeast Forestry University, 150040 Harbin, China

* Correspondence: liu.zhenbo@nefu.edu.cn

Received: 01-18-2023

Revised: 03-02-2023

Accepted: 03-06-2023

Citation: Y. L. Huang, N. Li, and Z. B. Liu, “An enhanced convolutional neural network for accurate classification of grape leaf diseases,” *Inf. Dyn. Appl.*, vol. 2, no. 1, pp. 8-18, 2023. <https://doi.org/10.56578/ida020102>.



© 2023 by the authors. Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: Grape leaf diseases can significantly reduce grape yield and quality, making accurate and efficient identification of these diseases crucial for improving grape production. This study proposes a novel classification method for grape leaf disease images using an improved convolutional neural network. The Xception network serves as the base model, with the original ReLU activation function replaced by Mish to improve classification accuracy. An improved channel attention mechanism is integrated into the network, enabling it to automatically learn important information from each channel, and the fully connected layer is redesigned for optimal classification performance. Experimental results demonstrate that the proposed model (MS-Xception) achieves high accuracy with fewer parameters, achieving a recognition accuracy of 98.61% for grape leaf disease images. Compared to other state-of-the-art models such as ResNet50 and Swim-Transformer, the proposed model shows superior classification performance, providing an efficient method for intelligent diagnosis of grape leaf diseases. The proposed method significantly improves the accuracy and efficiency of grape leaf disease diagnosis and has potential for practical application in the field of grape production.

Keywords: Grape disease; Image classification; Deep learning; Attentional mechanisms; Xception

1. Introduction

Grapes are a globally important cash crop, prized for their sweet taste and nutritional value. However, the incidence of grape leaf diseases has been increasing due to climate and environmental changes, resulting in common diseases such as black rot, whorl spot, and brown spot, which severely impact grape yield and quality. Precise identification of grape leaf disease species is essential to enable precise treatment of grape leaves. Yet, the current manual visual determination of grape leaf disease species in most vineyards is inefficient, costly, and prone to errors.

In recent years, deep learning has become a popular tool in computer vision applications, owing to its ability to extract image features automatically. Among the various deep learning techniques, convolutional neural network (CNN) has made significant strides in image recognition, including target detection [1-4], image segmentation [5, 6], and autonomous driving [7]. Researchers have also applied CNN to crop disease recognition, where it has shown promising results.

For instance, Sladojevic et al. [8] proposed a plant disease recognition method using deep convolutional networks, achieving recognition rates between 91% to 98%. This was the first application of deep learning methods to plant disease classification. Bi et al. [9] developed a leaf disease classification method based on MobileNet networks, which demonstrated better recognition efficiency for apple leaf diseases than InceptionV3 and ResNet152. Hameed and Üstündağ [10] proposed a method to detect apple leaf disease species using deep neural networks (DNN), utilizing accelerated robust features (SURF) for feature extraction and grasshopper optimization algorithm (GOA) for feature optimization, achieving better classification accuracy. Krishnamoorthy et al. [11] used the InceptionResNetV2 model combined with a migration learning approach to identify diseases in rice leaf images, achieving 95.67% recognition accuracy. Luo et al. [12] proposed a multi-scale feature fusion-based apple

disease classification network by altering the batch normalization and ReLU position, improving the ResNet network, and optimizing the network using methods such as pyramidal convolution instead of 3×3 convolution, achieving 94.24% recognition accuracy. Zhang et al. [13] combined null convolution with global pooling and proposed a global pooled null convolution neural network that can effectively discriminate cucumber diseases. Hu et al. [14] developed a lightweight adaptive feature extraction network model GKFENet based on the SqueezeNet model, achieving an average recognition accuracy of 97.90% for tomato diseases.

While previous studies have achieved good results in plant leaf disease classification, the small and variable spots in grape leaf disease images present a unique challenge, making it difficult to obtain accurate classification results. To address this challenge and achieve precise classification of grape leaf disease species, an improved convolutional neural network based on the Xception network [15] is proposed in this study. The main contributions of this article are as follows:

Firstly, the acquired dataset is expanded using data enhancement methods, including luminance brightness adjustment, rotation, and Gaussian noise addition, to prevent model overfitting and improve the robustness of the network.

Secondly, the Mish activation function is used to replace the ReLU activation function of the Xception network, improving the classification accuracy of the network and avoiding the problem of neuron death.

Thirdly, the network is enhanced with the Squeeze-and-Excitation (SE) module, which enables the network to focus more on extracting important information between feature channels, improving the network's ability to extract small lesion features.

Finally, the fully connected layer of the Xception network is improved using 1×1 convolution instead of the fully connected layer, enhancing the classification performance of the network.

2. Data Acquisition and Processing

To accomplish the grape leaf disease classification task, a certain number of grape leaf disease images were collected, followed by data enhancement and division of the dataset into a training set and a test set.

2.1 Data Acquisition

The experimental data for this study were collected from the plant village dataset, consisting of 2000 grape leaf images, including 500 images of each grape leaf black rot, whorl spot, brown spot, and healthy leaves.

2.2 Data Enhancement

In real-life situations, grape leaves grow in complex environments, and images taken in such environments may be impacted by various factors such as weather conditions, equipment clarity, and shooting angles. As a result, the photographs of grape leaves taken in real-life situations may exhibit complex backgrounds with varied angles, levels of clarity, etc., which can impact the classification results.

To simulate the real-life scenario of taking photos and make the classification task more suitable for application in real environments, the data were processed through data enhancement techniques. Specifically, image brightness was enhanced and reduced, flipping was applied, and Gaussian noise was added to expand the dataset. The resulting processed dataset comprises 10,000 images, which were divided into a training set and a validation set in a 4:1 ratio. Some of the original and processed images are shown in Figure 1.

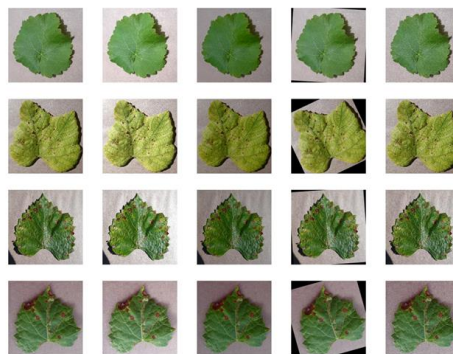


Figure 1. Some images of the dataset adopted in this study, from top to bottom, are healthy grape leaves, brown spot, whorl spot, and black rot; and from left to right, are original image, high brightness, low brightness, random rotation, and Gaussian noise

3. Classification Model of Grape Leaf Diseases

3.1 Xception Network

The basic structure of the Xception network is introduced in this section.

3.1.1 Inception module

Before 2014, most convolutional neural networks improved their performance by increasing the depth or width of the network, which was computationally expensive and could lead to overfitting. To address this problem, GoogLeNet [16] proposed the Inception module, which merges 1×1 convolution, 3×3 convolution, 5×5 convolution, 3×3 pooling, and dimensionality reduction using 1×1 convolution in parallel to reduce computation. The network can choose the appropriate combination of convolution layers to use, and the output feature map shape remains unchanged. The Inception module increases the network depth while enhancing the network's adaptability to different scale features and significantly reducing computational effort (See Figure 2).

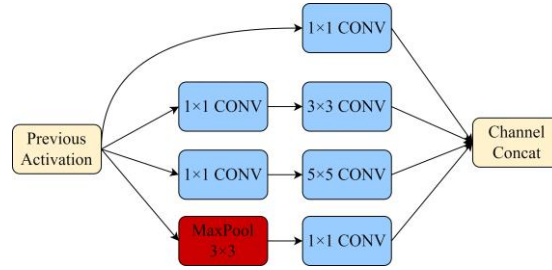


Figure 2. Inception module

3.1.2 Depthwise Separable Convolution

Depthwise Separable Convolution (DSC) comprises Depthwise Convolution (DC) and pointwise convolution (PC), which significantly reduce the number of parameters and computational costs of the network compared to normal convolution operations.

The DC operation performs spatial convolution for each input channel, and the results are restacked to obtain the output. Each convolution kernel corresponds to a separate channel for convolution. The PC operation performs a second convolution of the feature map obtained after the DC operation, performing channel fusion, changing the number of output channels, and combining the outputs of the DC operation.

Eq. (1) shows the computation required for the ordinary convolution operation.

$$Conv = H \times W \times f \times f \times C \times N \quad (1)$$

where, H and W denote the height and width of the feature map, respectively, f denotes the size of the convolution kernel, and C denotes the number of channels.

And the calculation required to perform the DSC operation is shown in Eq. (2):

$$DSC = H \times W \times f \times f \times C + C \times N \times H \times W \quad (2)$$

The schematic diagram of the DSC operation is shown in Figure 3.

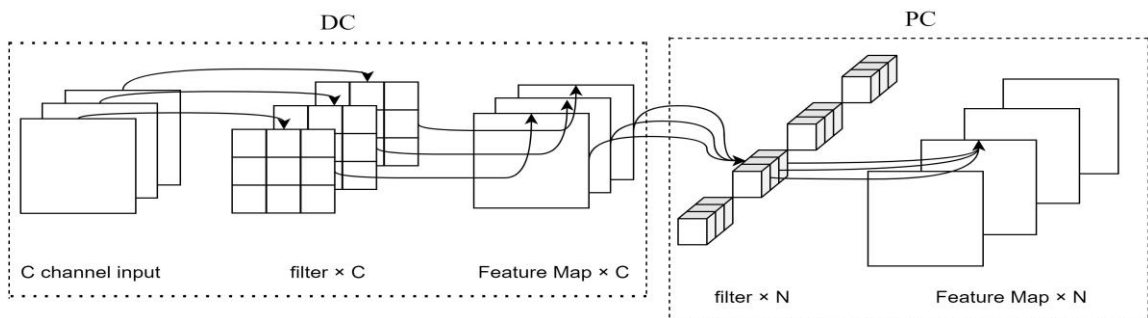


Figure 3. Depthwise Separable Convolution

The ratio of the computational effort of the DSC operation to the ordinary convolution is:

$$\frac{DSC}{Conv} = \frac{1}{N} + \frac{1}{f^2} \quad (3)$$

Using Depthwise Separable Convolution can significantly reduce computation, as shown by the reduced ratio of computational effort in Eq. (3).

3.1.3 Xception Network

The Xception network is a convolutional neural network that replaces the 3×3 convolution in the Inception v3 [17] network with DSC and combines the residual structure of the ResNet [18] network. It comprises three parts: Entry flow, Middle flow, and Exit flow, and includes a total of 14 blocks, each containing the DSC structure. The Xception network processes spatial and channel information separately, which enables it to extract image features more comprehensively.

3.2 Model Improvement

This section focuses on the improvement made to the Xception network in this study. The experimental results demonstrate that the proposed improved Xception model can effectively extract the disease spot features of grape disease leaves and achieve better classification performance.

3.2.1 Replacing the activation function

The activation function plays a crucial role in convolutional neural networks by mapping the input of neurons to the output and introducing nonlinearity into the network. This enhances the network's ability to fit various nonlinear models and increases the expressiveness of the model, thereby making deep networks effective.

The Xception network uses ReLU as its activation function, and its formula is shown in Eq. (4):

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

The ReLU activation function sets all positive values as constant and all non-positive values as 0. This function can activate only some neurons at a time, introducing sparsity in the network, improving computational efficiency, and reducing parameter dependence. The ReLU function's non-negative interval has a constant gradient, which avoids the problem of gradient disappearance. However, the ReLU activation function may result in the "permanent death of neurons" problem because the output may be zero, and the gradient cannot be updated.

The Mish function's graph is similar to that of ReLU, but smoother and does not have a value of 0 in the negative interval. The equation of the Mish function is as follows:

$$Mish(x) = x \cdot \tanh(\ln(1 + e^x)) \quad (5)$$

The Mish activation function also eliminates the problem of gradient disappearance and has a non-zero gradient in the negative interval, preventing the problem of neuron death and enabling the network to learn more features. Additionally, the Mish activation function can speed up the training process and improve the network's accuracy. The smoother nature of the Mish function allows information to penetrate better into the network, resulting in better accuracy and generalization. Figure 4 displays the graphs of the two activation functions.

Replacing the ReLU activation function with the Mish activation function improves the recognition accuracy of the classification network.

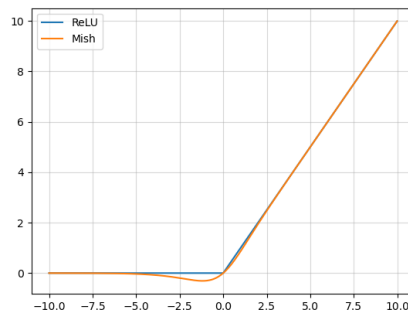


Figure 4. Graph of Mish activation function and ReLU activation function

3.2.2 Introduction of an improved SE attention module

The Attention Mechanism (AM) is a method of focusing on relevant information by automatically calculating the importance of input information to the network's output. This enables the network to prioritize effective information and ignore irrelevant information, improving the network's efficiency. The channel attention mechanism focuses on relevant information within channels, enabling the network to learn the importance of each channel and use resources more efficiently to extract more effective information.

SENet [19] (Squeeze-and-Excitation Networks) proposed a channel attention structure called the SE module, which comprises two parts: Squeeze and Excitation. Figure 5 shows the structure of the SE module. The SE module is divided into three parts: Squeeze operation, Excitation operation, and Scale operation.

The Squeeze operation performs global average pooling on feature maps within each path to compress channel features into a real number that represents each channel with a numerical value. The formula for this operation is shown in Eq. (6):

$$z = f_{Sq}^k = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W v^k(i, j), k = 1, 2, \dots, C \quad (6)$$

where, z is the result of the Squeeze operation performed on the input features in spatial dimension $H \times W$, and $v^k(i, j)$ is the feature map after a series of convolutions, and C is the number of channels of v .

The Excitation operation learns the feature weights of each channel. This operation reduces dimensionality through a fully connected layer, applies a ReLU activation function, raises the dimensionality through another fully connected layer, and finally generates a weight coefficient between 0 and 1 through a sigmoid activation function. This process predicts the importance of each channel. The calculation formula for the Excitation operation is shown in Eq. (7):

$$s = f_{Ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (7)$$

where, σ denotes the sigmoid function, δ denotes the ReLU activation function, $W_1 \in R^{\frac{C}{r} \times C}$ and $W_2 \in R^{C \times \frac{C}{r}}$ are the parameters of the two fully connected layers, and r is used to reduce the dimensionality of the fully connected layers.

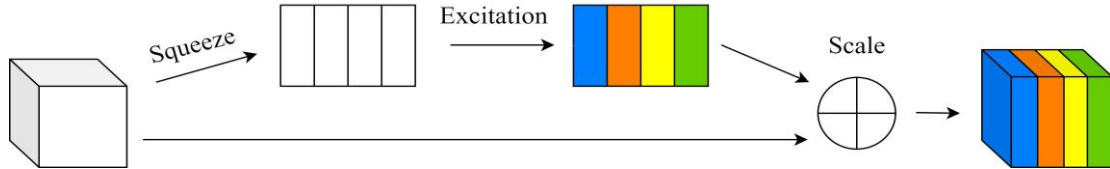


Figure 5. SE module

The Scale operation weights the weight coefficients obtained from Excitation to the original features channel by channel, labeling the importance of each channel. The formula is shown in Eq. (8):

$$\tilde{X} = f_{Scale} = s_k \times v_k, k = 1, 2, \dots, C \quad (8)$$

To extract disease spot information more comprehensively from disease images, this study proposes an improved SE module that uses a parallel structure of global maximum pooling and global average pooling instead of the original global average pooling. This improves the network's classification capability. Figure 6 compares the principle of the SE module before and after the improvement.

The global average pooling and global max pooling are used to process the feature maps separately, compressing them from space (N, H, W, C) to space $(N, 1, 1, C)$. The two compressed feature maps are then fused to fully extract the texture information of the disease images.

The improved SE module is integrated into the Middle flow section of the Xception network to enhance its ability to extract disease spot features from disease images.

3.2.3 Improving the fully connected layer

In convolutional neural networks, the fully connected layer can combine local features obtained in the previous

layer, reduce the impact of feature position on classification results, and improve the network's robustness. However, using fully connected layers can result in excessive network parameters.

By contrast, the 1×1 convolution can represent the entire image information while greatly reducing the number of parameters. Using 1×1 convolution instead of fully connected layers can effectively reduce the network's parameter count and improve its performance.

The overall structure of the improved network is presented in Figure 7.

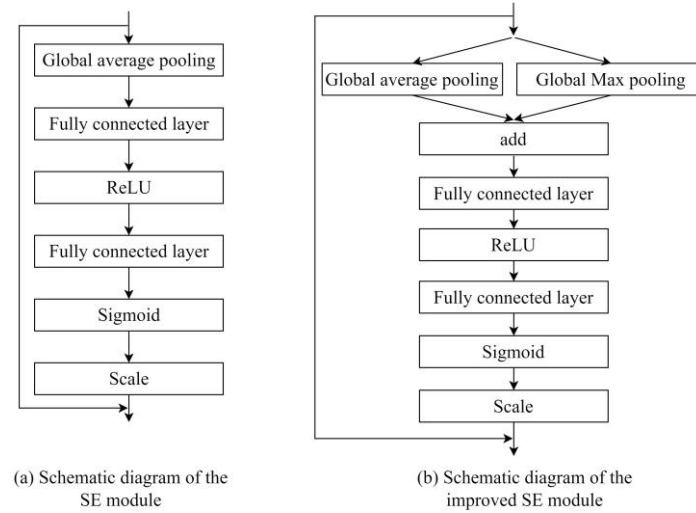


Figure 6. Comparison of the principle of SE module before and after improvement

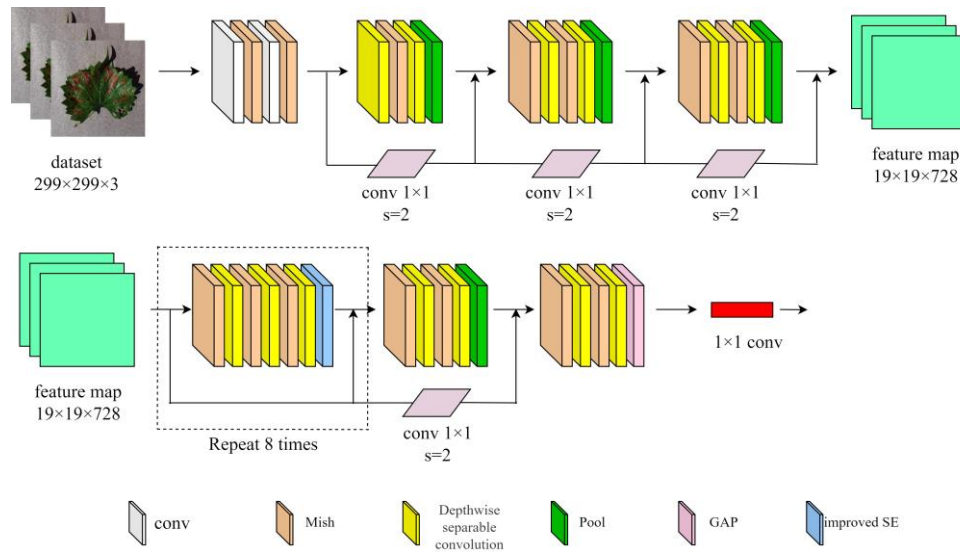


Figure 7. Network structure

4. Experimental Results and Analysis

4.1 Experimental Platform

The experimental platform used Windows 10 operating system with an AMD Ryzen 7 5800H processor with Radeon Graphics 3.20 GHz and 16 GB memory. The classification model was based on the Pytorch deep learning framework in Python 3.8, and the experimental software used was PyCharm. The GPU used was NVIDIA GeForce RTX 3060 Laptop GPU.

4.2 Experimental Design

4.2.1 Hyperparameter setting

The experimental parameters used in this network were as follows: a batch size of 12, 100 rounds of

experimentation, Adam optimizer, weight decay to suppress overfitting with a decay coefficient of 0.0002, and the cross-entropy loss function as the loss function.

To investigate the effect of learning rate on the experimental results, comparison experiments were designed to verify the classification effects at learning rates of 0.0001, 0.0005, and 0.001, respectively. The optimal learning rate was determined to be 0.001.

4.2.2 Network evaluation methods

Accuracy, recall (R), precision (P), and F1-score are commonly used metrics to evaluate the effectiveness of a classification model. They assist in determining how well the model is able to correctly identify and classify samples.

Accuracy rate is a metric that measures the proportion of correctly predicted samples to the total number of samples. The formula for accuracy rate is shown in Eq. (9):

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

Recall rate, also known as sensitivity or true positive rate, measures the proportion of actual positive samples that are correctly predicted as positive. The formula for recall rate is as follows:

$$Recall(R) = \frac{TP}{TP + FN} \quad (10)$$

Precision rate measures the proportion of positive predictions that are truly positive. Its formula is shown in Eq. (11):

$$Precision(P) = \frac{TP}{TP + FP} \quad (11)$$

F1-score is a combined metric that takes into account both precision and recall. It is the harmonic mean of precision and recall. The formula for *F1-score* is as follows:

$$F1-Score = 2 \times \frac{P \times R}{P + R} \quad (12)$$

In the formula for *F1-score*, *TP* represents the number of true positives, *TN* represents the number of true negatives, *FP* represents the number of false positives (i.e., negative samples predicted as positive), and *FN* represents the number of false negatives (i.e., positive samples predicted as negative).

4.3 Analysis of Experimental Results

4.3.1 Classification results

The confusion matrix is a widely used evaluation metric in multi-classification tasks. After testing the trained model on a test set of 2000 images, the resulting confusion matrix is presented in Figure 8. From the confusion matrix, the single test accuracy of the model was calculated to be 99.3%. Table 1 shows the recall, precision, and F1-score for each category.

Table 1. Test results of classification

	Precision/%	Recall/%	F1-Score
black rot	99.4	97.8	0.986
brown spot	99.6	100	0.98
healthy	100	100	1
whorl spot	98.2	99.4	0.988

Table 1 indicates that the trained model in this study has achieved an accuracy of 98% or higher for recognizing three types of diseased leaves, with an F1-score above 0.985. This further confirms the effectiveness of the proposed model.

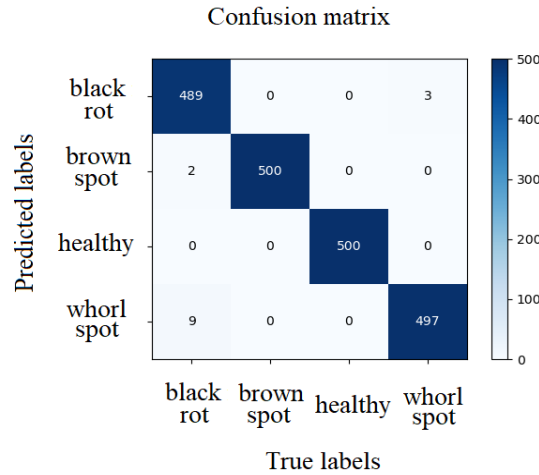


Figure 8. Confusion matrix

4.3.2 Effect of learning rate on recognition accuracy

Selecting an appropriate learning rate is crucial because a low learning rate slows down network convergence while a high learning rate may result in the gradient explosion problem and make model convergence difficult. To determine the optimal learning rate, we conducted a comparison experiment with learning rates of 0.0001, 0.0005, and 0.001, respectively. The results are presented in Figure 9, which shows that the classification performance is best when the learning rate is set to 0.001.

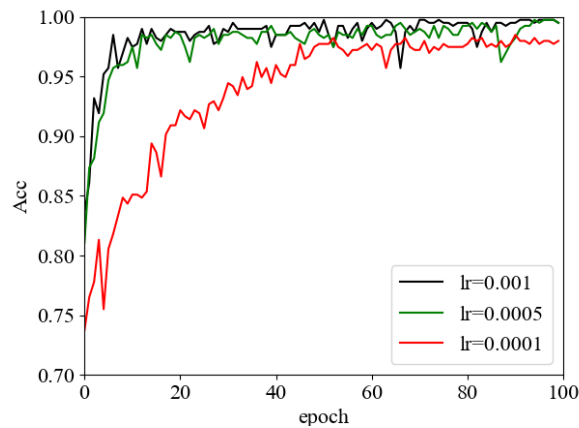


Figure 9. Effect of learning rate on classification accuracy

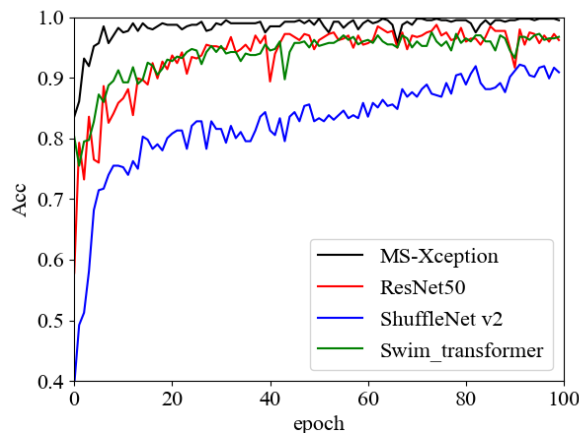


Figure 10. Comparison of experimental results of different models

4.3.3 Classification performance of different models

To assess the effectiveness of the proposed method, we conducted comparative experiments using ResNet50, ShuffleNet V2, and Swim Transformer classification models. The experimental results are presented in Figure 10. The classification accuracy of the ShuffleNet V2 model is comparatively lower, as it is a lightweight convolutional neural network model that may not perform as well as other deep convolutional models in terms of classification results. The ResNet50 and Swim Transformer models both exhibit good classification results, but their models have more parameters and are computationally complex. The MS-Xception model proposed in this study exhibits the best performance in the classification task, with faster convergence and higher accuracy, and significant advantages over other classification models.

4.3.4 Ablation experiments

Ablation experiments were conducted to verify the effectiveness of the proposed improved method. Table 2 presents the experimental results, where P1 denotes the replacement activation function, P2 denotes the introduction of the improved SE module, and P3 denotes the improved fully connected layer. The experiments demonstrate that the proposed improved method significantly enhances the performance of the original network, resulting in an average test accuracy increase of 2.38% to 98.61%.

Table 2. Comparison of experimental results of improved models

Model	Average Test Accuracy/%
Xception	96.23
Xception+P1	97.54
Xception+P1+P2	97.87
Xception+P1+P2+P3	98.61

The performance of the network was enhanced by using the Mish activation function to mitigate the issue of "neuron necrosis". The introduction of the improved SE module enabled the network to focus more effectively on the critical information in disease images, resulting in improved classification accuracy with only a small increase in computation. Improving the fully connected layer using 1×1 convolution reduced the number of network parameters and improved the classification performance. The proposed MS-Xception network demonstrated better results than the original network for the grape leaf disease classification task.

4.3.5 Model feature visualization

Feature visualization of convolutional neural networks can provide visual insight into the learning ability of classification models. In this study, we utilized Grad-CAM [20] for feature visualization of the classification model, and the resulting heat map is presented in Figure 11. The darker the color, the more capable the model is at learning features. From the heat map, it can be observed that the MS-Xception network focuses its attention on the disease spot region of the leaf images, validating its ability to identify grape leaf disease features.

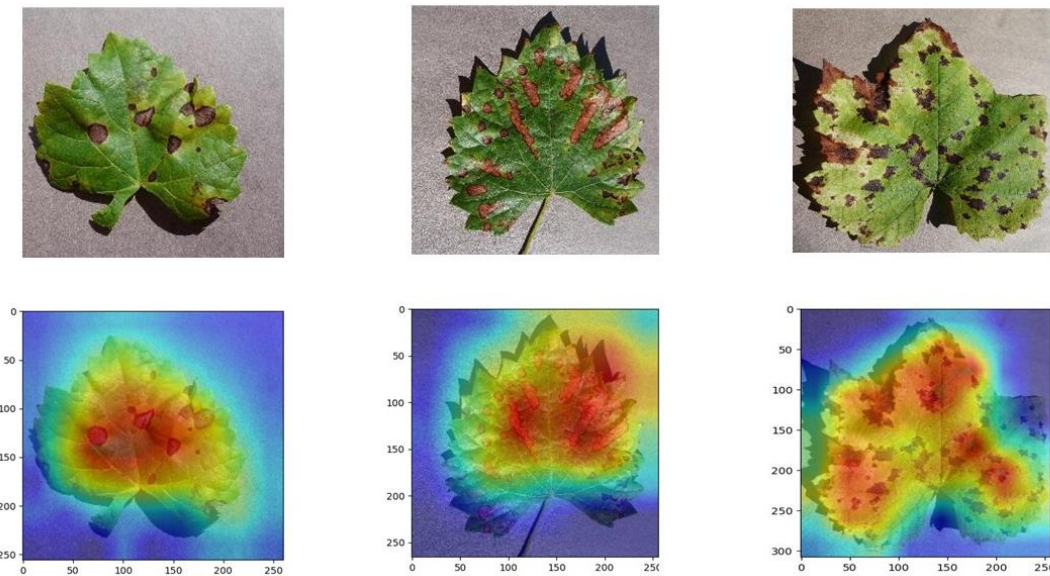


Figure 11. The images from left to right are heat maps of grape leaf black rot, whorl spot and brown spot

5. Conclusion

This study proposes an improved lightweight convolutional network (MS-Xception) for enhancing the accuracy of grape leaf disease species identification. The network is based on the Xception network, and modifications were made to the activation function, the introduction of an attention mechanism, and the improvement of the fully connected layer to enhance the classification accuracy of the network. The experimental results demonstrate that the proposed model outperforms other networks, providing an effective method for grape leaf disease classification and serving as a reference for crop pest and disease identification.

However, the proposed method has certain limitations. The model parameters are relatively large, making it unsuitable for deployment on mobile devices. Furthermore, identifying the degree of grape leaf disease is a crucial task that requires further exploration. Future work will focus on developing lightweight models and improving the identification of disease extent.

Author Contributions

Significant contributions to the design of the experiment and manuscript revision were made by Yinglai Huang. Ning Li made significant contributions to the experiment design, data collection and processing, execution of the experiment, as well as manuscript writing and revision. Zhenbo Liu contributed significantly to the project provision and manuscript revision. All authors have read and agreed to the final version of the manuscript that was published.

Funding

This research was funded by the National Natural Science Foundation of China (Grant No.: 61902059) and the Natural Science Foundation of Heilongjiang Province (Grant No.: LH2020C051).

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Acknowledgments

The hard work and valuable comments of the anonymous reviewers are greatly appreciated, as they have contributed to the improvement in the quality of this paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] V. Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," *Information Processing in Agriculture*, vol. 4, no. 1, pp. 41-49, 2017. <https://doi.org/10.1016/j.inpa.2016.10.005>.
- [2] A. Kumar, A. Kalia, and A. Kalia, "ETL-YOLO v4: A face mask detection algorithm in era of COVID-19 pandemic," *Optik*, vol. 259, Article ID: 169051, 2022. <https://doi.org/10.1016/j.ijleo.2022.169051>.
- [3] C. Jiang, H. Ren, X. Ye, J. Zhu, H. Zeng, Y. Nan, M. Sun, X. Ren, and H. Huo, "Object detection from UAV thermal infrared images and videos using YOLO models," *Int. J. Appl. Earth Obs. Geoinformation*, vol. 112, Article ID: 102912, 2022. <https://doi.org/10.1016/j.jag.2022.102912>.
- [4] R. Xia, G. Li, Z. Huang, H. Meng, and Y. Pang, "Bi-path combination YOLO for real-time few-shot object detection," *Pattern Recognit. Lett.*, vol. 165, pp. 91-97, 2023. <https://doi.org/10.1016/j.patrec.2022.11.025>.
- [5] S. A. Güven and M. F. Talu, "Brain MRI high resolution image creation and segmentation with the new GAN method," *Biomed. Signal Process. Control.*, vol. 80, Article ID: 104246, 2023. <https://doi.org/10.1016/j.bspc.2022.104246>.
- [6] Q. Han, H. Wang, M. Hou, T. Weng, Y. Pei, Z. Li, G. Chen, Y. Tian, and Z. Qiu, "HWA-SegNet: Multi-channel skin lesion image segmentation network with hierarchical analysis and weight adjustment," *Computers in Biology and Medicine*, vol. 152, Article ID: 106343, 2023. <https://doi.org/10.1016/j.compbimed.2022.106343>.
- [7] C. C. Pham and J. W. Jeon, "Robust object proposals re-ranking for object detection in autonomous driving using convolutional neural networks," *Signal Processing: Image Commun.*, vol. 53, pp. 110-122, 2017.

- <https://doi.org/10.1016/j.image.2017.02.007>.
- [8] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, vol. 2016, Article ID: 3289801, 2016. <https://doi.org/10.1155/2016/3289801>.
 - [9] C. Bi, J. Wang, Y. Duan, B. Fu, J. Kang, and Y. Shi, "MobileNet based apple leaf diseases identification," *Mobile Netw. Appl.*, vol. 27, pp. 172-180, 2022. <https://doi.org/10.1007/s11036-020-01640-1>.
 - [10] J. S. Hameed and B. B. Üstündağ, "Evolutionary feature optimization for plant leaf disease detection by deep neural networks," *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, pp. 12, 2020. <http://dx.doi.org/10.2991/ijcis.d.200108.001>.
 - [11] N. Krishnamoorthy, L. N. Prasad, C. P. Kumar, B. Subedi, H. B. Abraha, and V. E. Sathishkumar, "Rice leaf diseases prediction using deep neural networks with transfer learning," *Environmental Research*, vol. 198, p. 111275, 2021.
 - [12] Y. Luo, J. Sun, J. Shen, X. Wu, L. Wang, and W. Zhu, "Apple leaf disease recognition and sub-class categorization based on improved multi-scale feature fusion network," *IEEE Access*, vol. 9, pp. 95517-95527, 2021. <https://doi.org/10.1109/ACCESS.2021.3094802>.
 - [13] S. Zhang, S. Zhang, C. Zhang, X. Wang, and Y. Shi, "Cucumber leaf disease identification with global pooling dilated convolutional neural network," *Comput. Electron. Agric.*, vol. 162, pp. 422-430, 2019. <https://doi.org/10.1016/j.compag.2019.03.012>.
 - [14] L. Hu, T. Zhou, Y. Liu, W. Xu, R. Gai, X. Li, Y. Pei, and Z. Wang, "Tomato disease recognition based on lightweight network auto-adaptive feature extraction," *Jiangsu Journal of Agricultural Sciences*, vol. 3, pp. 696-705, 2022.
 - [15] F. Chollet, "Xception: Deep learning with Depthwise Separable Convolutions," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017, pp. 1251-1258. <https://doi.org/10.1109/CVPR.2017.195>.
 - [16] C. Szegedy et al., "Going deeper with convolutions," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015, pp. 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>.
 - [17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 2818-2826. <https://doi.org/10.1109/CVPR.2016.308>.
 - [18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 770-778.
 - [19] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 7132-7141. <https://doi.org/10.1109/CVPR.2018.00745>.
 - [20] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017, pp. 618-626. <https://doi.org/10.1109/ICCV.2017.74>.