

Retraction

Retracted: Multidimensional Attention-Based CNN Model for Identifying Apple Leaf Disease

Journal of Food Quality

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

Copyright © 2024 Journal of Food Quality. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] K. Perveen, S. Kumar, S. Kansal et al., "Multidimensional Attention-Based CNN Model for Identifying Apple Leaf Disease," *Journal of Food Quality*, vol. 2023, Article ID 9504186, 12 pages, 2023.

Research Article

Multidimensional Attention-Based CNN Model for Identifying Apple Leaf Disease

Kahkashan Perveen,¹ Sanjay Kumar,² Sahil Kansal ,³ Mukesh Soni,⁴ Najla A. Alshaikh,¹ Shanzeh Batool,⁵ Mehrun Nisha Khanam,⁶ and Bernard Osei ⁷

¹Department of Botany & Microbiology, College of Science, King Saud University, Riyadh 11495, Saudi Arabia

²Computer Science Engineering Department, Chandigarh Group of Colleges, Jhanjeri (Mohali) 140307, Punjab, India

³IT, JIMS, Rohini, Delhi, India

⁴Department of CSE, University Centre for Research & Development Chandigarh University, Mohali 140413, Punjab, India

⁵School of Computer Science Engineering (SCSE), Vellore Institute of Technology, Bhopal 466114, India

⁶School of Biological Sciences, College of Natural Sciences, Seoul National University, Gwanak-Gu, Seoul 08826, Republic of Korea

⁷Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Correspondence should be addressed to Bernard Osei; bosei26@st.knust.edu.gh

Received 25 June 2022; Revised 6 August 2022; Accepted 9 August 2022; Published 8 February 2023

Academic Editor: Rijwan Khan

Copyright © 2023 Kahkashan Perveen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To prevent the spread of illnesses and guarantee the steady and healthy growth of the apple sector, the proper diagnosis of apple leaf diseases is of utmost importance. The subtle interclass variations and enormous intraclass variances among apple leaf disease features, together with the uniformity of disease spots and the complicated background environment, make apple leaf disease diagnosis extremely challenging. A unique dual-branch apple leaf disease diagnosis system (DBNet) was put out to address the aforementioned issues. An attention branch with many dimensions and a multiscale joint branch (MS) make up the dual-branch network topology of the DBNet (DA). In this study, the MS branch and the DA branch are combined to create a DBNet, which successfully improves recognition accuracy while mitigating the negative impacts of complicated backdrop environments and lesion similarities. The accuracy of the DBNet network increases by 0.02843, 0.02412, 0.0144, and 0.0125, respectively, when compared to previous leaf disease detection models. This makes it evident that the suggested DBNet model has certain benefits over others in terms of identifying apple leaf disease.

1. Introduction

Plant diseases are the main factors that endanger the development of agriculture worldwide and cause severe economic losses every year. Owing to advancements in machine learning technologies, plant disease identification has become an important research field in pattern recognition and modern agricultural development [1]. Early plant disease identification methods extract the diseased area from the image manually and use the *K*-means clustering algorithm, an artificial neural network (ANN), and a support vector machine (SVM) according to the segmented plant disease spots and the SVM method for disease diagnosis [2, 3].

However, due to the complexity of plant disease spots, it is not easy to extract helpful disease spot features [4]. Even if compelling lesion features are extracted for diagnosis, the method has the disadvantages of a low recognition rate and weak generalization ability [5, 6]. The proposal of AlexNet in 2012 made the deep convolutional neural network (CNN) more widely used in image recognition [7, 8]. [9] used transfer learning to adjust VGG-16 and realized the disease identification of *Camellia oleifera*. [10] proposed an ARNet combining attention mechanism and residual to learn tomato disease identification. [11] proposed a dual-channel convolutional neural network that connects the grayscale image of the lesion and the LBP (local binary pattern) feature

map to realize the disease identification of pumpkin. The above methods have achieved good results in plant disease identification. However, there is still room for further research on the intra- and interclass gaps caused by complex background environments and the similarity of disease spots.

B-CNN record localized feature interactions in a translationally invariant way and represent an image as a pooled outer product of features acquired from two CNNs. Although they may be trained end-to-end, B-CNNs are similar to orderless texture representations constructed on deep features. When two variant extractors gather detailed features with varied morphological and geographic properties rather than using bottleneck features, B-CNNs can be informative. Transferability of weights would be provided by bottleneck features obtained from several state-of-the-art models, but a fine-tuning model created from scratch would provide a general interpretation of features. It is necessary to build second-order pooling techniques that offer 11 precise self-attention processes. In the future, it is intended to use aerial imaging technology to construct a model to extract several properties from many plant leaf photos for different classes instead of doing experiments on single leaf images. To better solve this problem, based on the bilinear convolution neural network B-CNN (bilinear CNN) [12], this paper proposes a new apple leaf disease identification method, DBNet (dual branch network), which is used to identify five common and high incidence mosaic diseases: rust, grey spot, spotted leaf spot, and brown spot-on apple leaves.

To accomplish weak-feature adaptive extraction, DB-Net employs a unique dual-branch separation and fusion approach. The Brain Compression Assessment Branch and Infarct Assessment Branch of the DB-Net encoder sub-modules are suggested to use lightweight encoding structures with variable receptive fields that are tailored to the features of the lesion region. DB-Net reduces duplicate information and clearly explains the severity of lesions using its keyframe selection method and area guiding expertise. In summary, DB-Net combines global and local characteristics to gather multiscale and multicategory brain status information, therefore improving noncontrast CT's weak points and achieving accurate HT prediction. DB-Net assists clinicians in secondary diagnosing the HT risk of cerebral stroke patients and can prevent doctors from making incorrect HT risk diagnoses. DBNet contains a multiscale branch (MS) and a multidimensional attention branch (DA). The multiscale joint branch (MS) improves the structure of inception [13] to extract image features from different receptive fields and fuses the feature information of different levels through cross-layer connections. Multidimensional attention branch (DA) is a new attention mechanism proposed based on SENet [14], which can extract attention features of 3 different dimensions. Then DBNet fuses the two branches' multiscale elements and multidimensional features and achieves high recognition accuracy on the apple leaf disease dataset [15] released by Northwest A&F University.

The detection of apple leaf diseases has been quite challenging because of their complicated history and the

similarities of their lesions. This study suggests a new double-branch network technique for identifying apple leaf diseases (DBNet). This article further goes into great length on the DBNet network topology, the multiscale joint branch (MS) model building, and the multidimensional attention branch (DA) principle in the coming sections.

2. Problem Description and Solution Ideas

A dual-branch network detection approach for apple leaf diseases is presented in the study. By combining the multiscale and multidimensional attention elements of lesion pictures, the proposed technique may successfully reduce the complexity of the background environment and the resemblance of lesions to apple leaf diseases. Due to the complex background and the similarity of lesions, the identification of apple leaf diseases has caused great difficulties. This paper proposes a new double-branch network apple leaf disease identification method (DBNet). Next, the DBNet network structure, the model construction of the multiscale joint branch (MS), and the multidimensional attention branch (DA) principle are studied further in this paper in detail.

2.1. DBNet Network Structure. Identifying apple leaf diseases is a more detailed subcategory division of the same primary category. Affected by the similarity of lesions, the interclass gap between subclasses is small and influenced by the complex background environment. The intraclass gap between subclasses is large. As shown in Figure 1, the color and texture of rust, grey spot, and leaf spot are similar in color and texture, resulting in a small gap between classes.

On the other hand, the influence of leaves and a cluster of leaves results in a sizeable intraclass gap. Therefore, the convolutional neural network has a high probability of misjudgment in the recognition process.

B-CNN is mainly used for fine-grained image classification tasks. Compared with traditional convolution neural networks, its unique feature is that the architecture adopts a shunt structure with two branches, each of which uses the primary convolution neural network as feature extraction. The two branches can be either symmetrical or asymmetrical. After the feature extraction, the bilinear pooling operation is used to realize feature fusion. B-CNN adopts a dual-branch structure to extract more object features and local features but does not consider the attention mechanism and the impact of multiscale information on network performance. Based on B-CNN, this paper proposes a novel dual-branch convolution neural network model (DBNet). The starting point of DBNet network design is to alleviate the adverse effects of complex background environment and lesion similarity on disease identification, so a dual-branch structure is designed, and the weights of each branch are not shared. The key to its network structure is to use a dual-branch network to extract more effective lesion features, and because of the different branch network structures, the extracted features are also different. DB-Net offers a novel dual-branch

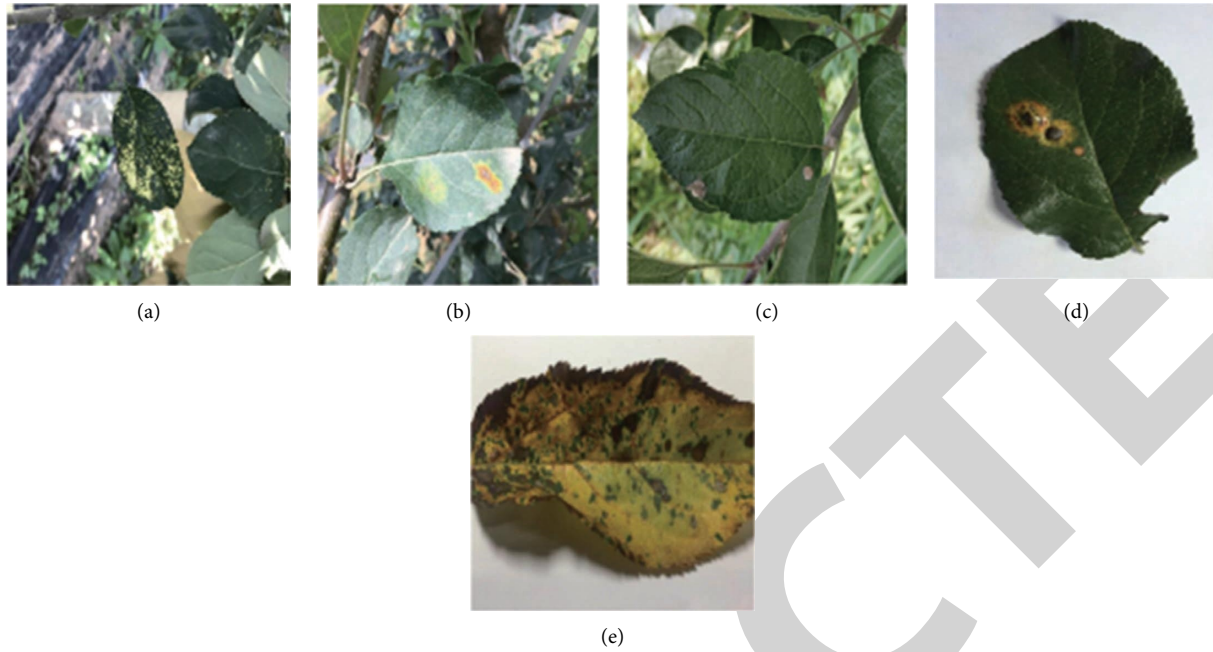


FIGURE 1: Image of apple leaf disease. (a) Mosaic, (b) rust, (c) grey spot, (d) Alternaria leaf spot, and (e) brown spot.

separation and fusion technique to provide weak-feature adaptive extraction. The DB-Net encoder submodules, brain compression assessment branch and infarct assessment branch are advised to utilize lightweight encoding structures with variable receptive fields that are suited to the peculiarities of the lesion location. DB-keyframe Net's selection process and expertise remove redundant information and effectively convey the severity of lesions. In summary, DB-Net integrates global and local features to collect multiscale and multicategory brain state information, therefore addressing the shortcomings of noncontrast CT and attaining accurate HT prediction. DB-Net helps physicians identify the secondary HT risk of cerebral stroke patients and can help doctors avoid making inaccurate HT risk diagnoses. The multiscale joint branch (MS) uses different convolutional kernels such as standard and atrous convolution to obtain various receptive fields in parallel so that both overall global and detailed local information can be extracted. The multidimensional attention branch (DA) adopts a combination of standard convolution and attention mechanisms to pay more attention to the lesion area of the image and the minor differences between the diseases during disease identification, thereby improving the recognition efficiency of the network accuracy. The network structure of DBNet is shown in Figure 2, where \odot represents the Concat connection, \otimes represents “*,” \oplus means “+,” FC refers to the fully connected layer, and Softmax refers to an activation function.

The DBNet network structure can be categorized into three parts: the multiscale joint branch (MS), the multidimensional attention branch (DA), and the final feature fusion part. Input a disease image of size $3 \times 224 \times 224$, which extracts features through the MS and DA branches,

respectively. Both units use VGG-16 as the base network and adjust the VGG-16 network according to the different functions of the two branches in DBNet. The MS branch obtains image features under various receptive fields through convolution kernels of different sizes, aggregates the extracted features, and connects with the low-level image information of the external network through cross-layer connections to obtain multilevel features, thereby achieving the purpose of multiscale information fusion. The DA branch combines spatial and channelization through the attention mechanism in the three dimensions of channel, width, and height. It performs three-dimensional attention adjustment for each position of the feature map so that the convolutional neural network can better focus on the prominent lesion area of the image. Finally, the features extracted by the MS and DA branches are subjected to global average pooling. While reducing the model parameters, the concept method is used for feature fusion, and finally, the Softmax function is used to achieve apple leaf disease identification.

2.2. Multiscale Joint Branch (MS). Due to the differences in disease spot images of different plant diseases, for example, mosaic disease and brown spot disease on apple leaves are different from rust, grey spot disease, and spotted leaf disease. The lesion area of the latter three only occupies a small part of the leaves. Moreover, the influence of complex backgrounds such as a single leaf and a cluster of leaves will also cause significant difficulties in the image recognition of the same type of disease. Therefore, it is not optimal to obtain features of multiscale objects under only one convolution branch. The primary purpose of constructing the MS branch is to get different receptive fields by

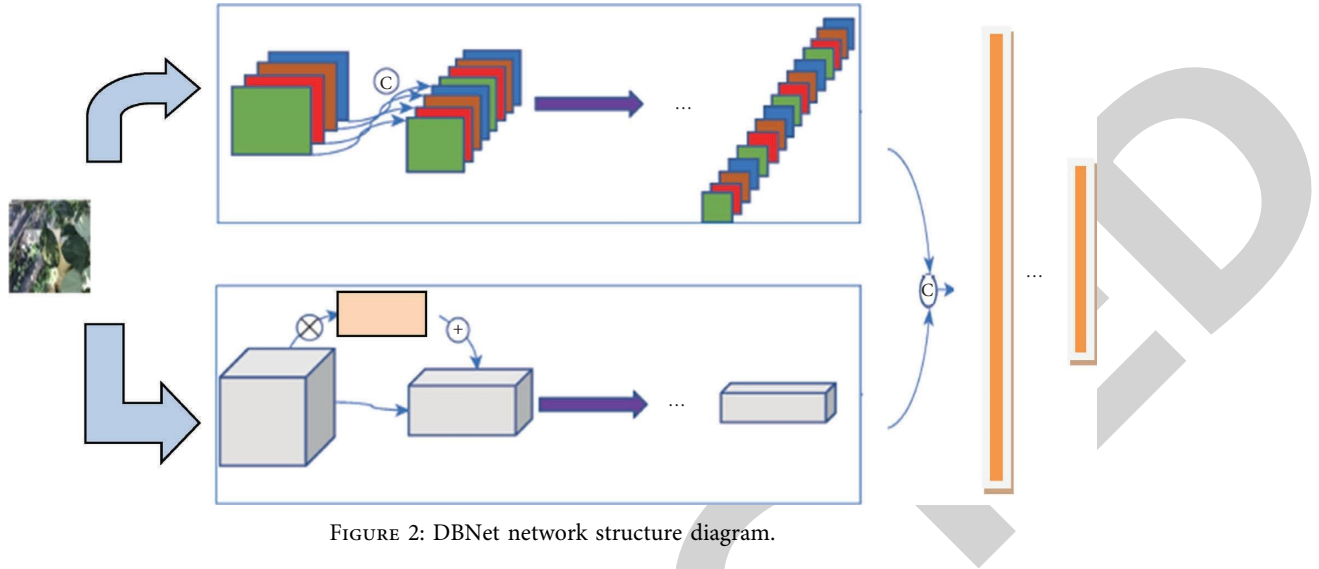


FIGURE 2: DBNet network structure diagram.

using convolution kernels of various sizes. When the lesion information in the image is scattered, the large kernel convolution can obtain more global features, while the lesion information can be obtained.

On the other hand, the small kernel convolution can get more local parts when the information distribution is relatively concentrated. The structure of the MS branch module is shown in Figure 3. C refers to the number of channels, W refers to the width, H refers to the height, d refers to the void ratio, Concat connection, and \oplus means “+.” The MS branch can be divided into multiscale convolutional and cross-layer links. The multisize convolution part consists of 6 convolution kernels of different sizes, consisting of 1×1 and 3×3 convolutions, respectively, atrous convolutions with atrous rates of 2 and 3 [14], and 1×3 and 3×1 nonconvolutions composed of symmetric convolutions [16].

The distance between the kernel locations is governed by the dilation factor. The atrous method is another name for the convolution carried out in this manner. The architecture, in this case, uses the atrous algorithm in conjunction with FC-reduced VGG16. By introducing zero-values into the convolution kernels, dilated convolutions or atrous convolutions increase the window size without increasing the number of weights. Among them, the atrous convolution is a hyperparameter with a whole rate added to the standard convolution, and the number 0 can be used to fill the convolution kernel during implementation. As shown in Figure 4(a): the 3×3 standard convolution has a hole rate of 1, and the receptive field is 3×3 . The convolution with a hole rate of 2 is filled with a 0 in the 3×3 standard convolution. The receptive field is 5×5 ; the convolution with a dilation rate of 3 is filled with two 0s in the 3×3 standard convolution, and the receptive field is 7×7 . In the deep convolutional neural network, down-sampling is usually used to increase the receptive field, but it will reduce the image

resolution and may cause loss of spatial detail information. The dilated convolution expands the receptive field by setting the dilation rate, and setting different dilation rates can also capture multiscale context information. It can be seen from Figure 4(a) that the atrous convolution with atrous rates of 2 and 3 expands the receptive field of the 3×3 convolution to 5×5 and 7×7 without adding additional parameters, capturing the multistage information.

Asymmetric convolutions operate by separating the image’s x and y axes. Using a convolution with an $(n \times 1)$ kernel before one with a $(1 \times n)$ kernel, for instance. Fewer parameters and less processing are required for the asymmetric 3D convolutions. For a more accurate analysis of the human activity in movies, 3D convolution extends 2D convolution to the spatial-temporal domain. However, compared to 2D convolution, 3D convolution uses a much larger number of parameters. The standard $n \times n$ convolution can be decomposed into two $one \times n$ and $n \times one$ asymmetric convolutions. The calculation amount of $n \times n$ standard convolution is $n \times n$, and the calculation amount becomes $2 \times n$ after decomposing into asymmetric convolution. Asymmetric convolutions 1×3 and 3×1 were initially designed to reduce the amount of computation, but there will be a slight loss of accuracy, so the results of the 3×3 convolution and 1×3 , 3×1 convolution are added. Based on the original 3×3 convolution, the horizontal and vertical kernels are also learned, as shown in Figure 4(b). The main idea is to use 3 parallel layers to operate the input feature with the standard convolution and two asymmetric convolutions, and then use the additive properties of the convolution to fuse the asymmetric convolution into the standard convolution equivalently. The three convolution kernels can be regarded as enhanced convolution kernels. That is, the kernel skeleton of the standard convolution is enhanced by the asymmetric convolution.

$$(\text{out}) = \text{sum}(g_{3 \times 3}(y), g_{3 \times 1}(y)) \odot g_{1 \times 1}(y), g_{3 \times 3(2)}(y)) \odot g_{3 \times 3(3)}(y). \quad (1)$$

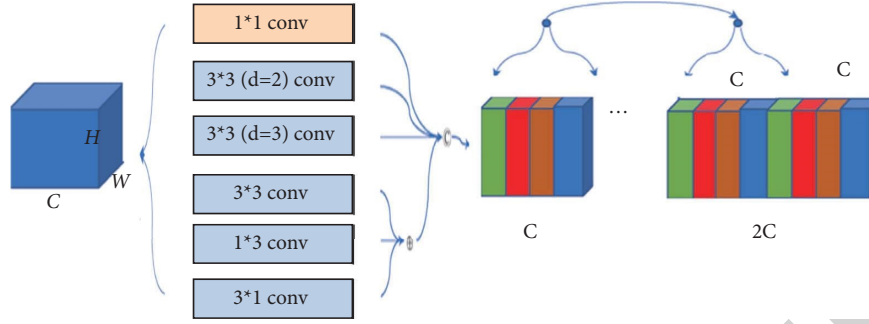


FIGURE 3: MS branch structure diagram.

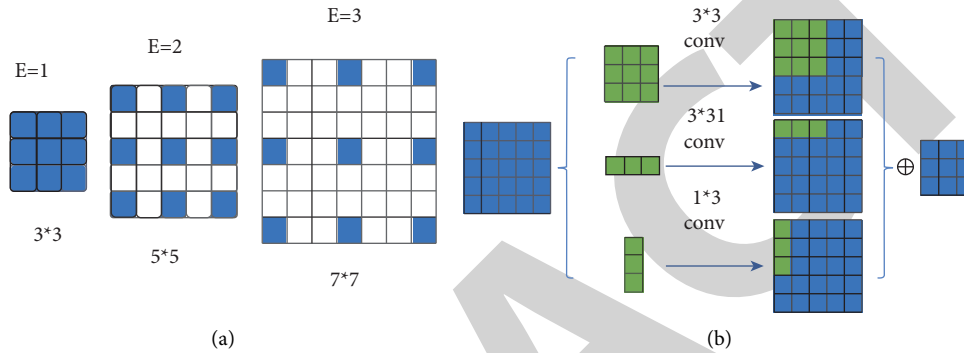


FIGURE 4: Atrous convolution and asymmetric convolution. (a) Dilated convolution and (b) asymmetric convolution.

In the formula, g represents different types of convolution operations: x represents the input; \oplus represents the addition operation; and \otimes represents the concept connection.

The input image size of the MS branch is $C \times H \times W$, the number of channels is C , and 4 sets of feature maps are generated through 6 convolution kernels of different sizes. First, the number of product kernels is set to $C/4$. The concept connection aggregates 4 groups of feature maps with a channel number of $C/4$. The channel number of the feature map is C , the same as the original input size. The cross-layer connection combines the low-level and high-level information of the CNN, which not only retains the low-level texture information but also improves the high-level semantic information of the CNN. As shown in Figure 5, the cross-layer connection connects the feature map output by the upper layer of the CNN with the feature map output by the lower layer of the convolution neural network, the number of channels will be doubled, and this feature map is used like the following feature map—layer input.

2.3. Multidimensional Attention Branch (DA). SENet is built with a “Squeeze-and-Excitation” (SE) block that explicitly models interdependencies between channels to adaptively recalibrate channel-wise feature responses. A compression method suggested by the squeeze and excitation (SE) methodology is used for condensable abstraction of numerous contiguous slices of a volume into

a single slice to consolidate the information across layers of an image volume. By updating the concurrent SE block, we additionally augment the network with spatial and channel-wise attention in addition to the increased input for the 2D backbone network. The resultant network gains from the 2D network’s cheap computing cost as well as the enhanced data from picture volumes and carefully observed characteristics. The multidimensional attention branch net (MDA-Net) simulates the manual segmentation procedure for 3D images. When working with an image volume, we frequently choose one main view to successively segment 2D slices, periodically checking the third dimension across slices to get more data.

Compared with traditional image classification, the gaps between different types of plant diseases are only contained in small local details. They are also affected by the lesions’ background and shape, making identification more difficult. The MS branch can deal with the intraclass gap caused by complex environments and the uneven distribution of lesions through the fusion of multiscale information. However, for the small interclass gap between some lesions, the effect is not apparent, and in the process of down-sampling, important local news will be lost. Based on SENet, this article proposes a new attention mechanism, multidimensional attention, and constructs a multidimensional attention branch (DA) through the multidimensional attention mechanism. The DA branch employs an attention mechanism to pinpoint the lesion location in the image, allowing for more effective lesion

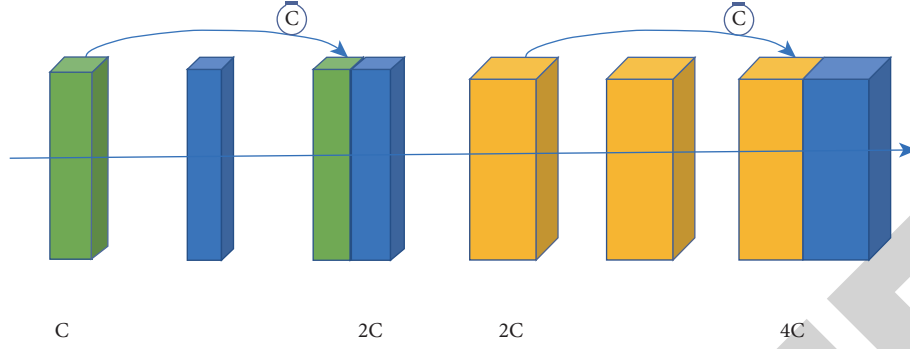


FIGURE 5: Cross-layer connections.

features and mitigating the negative effects of the image size change. The structure of the multidimensional attention module is shown in Figure 6.

Among them, C refers to the number of channels, H refers to the height, W refers to the width, 1×1 Conv represents 1×1 convolution, and $\text{Softmax}()$ refers to the Softmax activation operation on a particular dimension, \otimes describes “*,” and \oplus means “+.”

The input size of the DA branch module is $C \times H \times W$. After 1×1 convolution, Softmax operations are performed in the three dimensions of channel, height, and width, and x_i represents different states. Input f_1 refers to the convolution operation, and $\text{Softmax}()$ refers to the Softmax operation on a particular dimension in the following formula:

$$\begin{cases} y_C = \text{Softmax}(C)(g_1(y)), \\ y_H = \text{Softmax}(H)(g_1(y)), \\ y_W = \text{Softmax}(W)(g_1(y)). \end{cases} \quad (2)$$

After the Softmax operation is multiplied by a dimension proportion coefficient, the coefficient represents the proportion of each dimension to obtain the overall balance of each dimension’s attention. A_i represents attention to different sizes, and sum represents the sum operation in the following formula:

$$\begin{cases} B_C = \frac{C}{\text{sum}(C, H, W)} * y_C, \\ B_H = \frac{H}{\text{sum}(C, H, W)} * y_H, \\ B_W = \frac{W}{\text{sum}(C, H, W)} * y_W. \end{cases} \quad (3)$$

Then, the attention of different dimensions is added, multiplied with the input image to obtain attention features, and finally accumulated with the input image to get more significant image features.

$$y_{\text{out}} = y + \text{sum}(B_C, B_H, B_W). \quad (4)$$

The principle of the DA branch attention module is shown in Figure 7. Among them, the size of the input matrix

$X_{in}(C \times W \times H)$ is $2 \times 3 \times 4$, and the value of each element is as follows:

$$Y_{in} = \begin{bmatrix} 2 & 3 & 4 & 5 \\ 1 & 4 & 5 & 6 \\ 7 & 8 & 8 & 9 \end{bmatrix} \begin{bmatrix} 2 & 2 & 3 & 4 \\ 5 & 4 & 7 & 6 \\ 5 & 7 & 8 & 9 \end{bmatrix}. \quad (5)$$

Spatial attention adjusts the feature map at each position of the (H, W) two dimensions, and channel attention changes the channel dimension C . The multidimensional attention module treats the feature map in the convolutional neural network as a three-dimensional matrix, and each positioning element is independent of the other. The attention matrices of the three dimensions C , H , and W are calculated by the Softmax function, but because the three dimensions C , H , and W are not the same size, the proportion of each dimension in the whole is not the same. As the number of layers of the convolution neural network continues to deepen, the three dimensions C , W , and H will also change, so they need to be multiplied by a dimension proportion coefficient as shown in formula (3). In the shallow CNN, the number of channels is small due to the large height and width of the image, and the two dimensions of width and size play a crucial role. With the down-sampling of the convolution neural network, the number of channels increases and the width and height decrease, so the channel dimension plays a significant role in the deep convolution neural network. Therefore, the attention mechanism of the DA branch can also change the proportion of attention in each dimension as the number of network layers continues to deepen.

3. Experimental Results and Analysis

3.1. Experiment Preparation. The experiment uses the open-source apple leaf disease data set of Northwest A&F University, as shown in Figure 1. The data set uses a BM-500GE color camera to shoot in outdoor and laboratory environments, comprising 26,377 images, including 4,875 mosaic disease images, 5 655 pieces of brown spot disease, 5 694 pieces of rust disease, 4 810 pieces of grey leaf spot disease, and 5 343 pieces of leaf spot disease. Take 70% of the data set as training and 30% as a test, the training set to the test set is 7 : 3, and keep the identical data distribution of the training

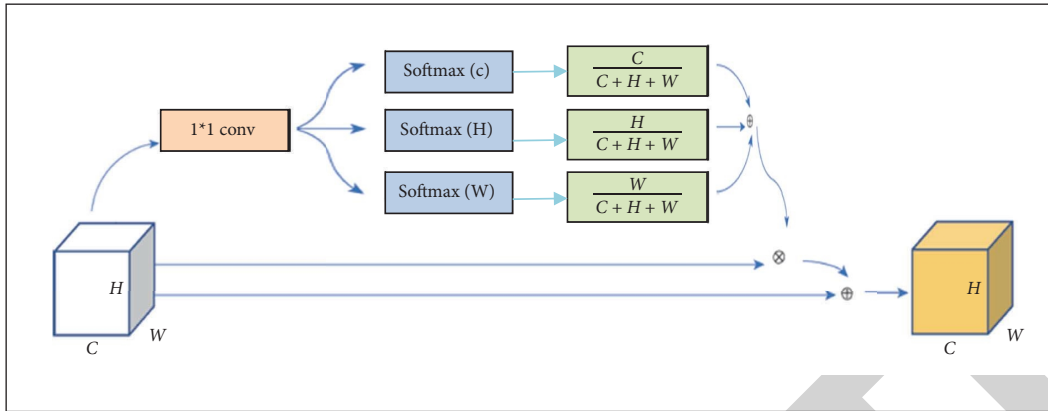


FIGURE 6: Structure diagram of the DA branch attention module.

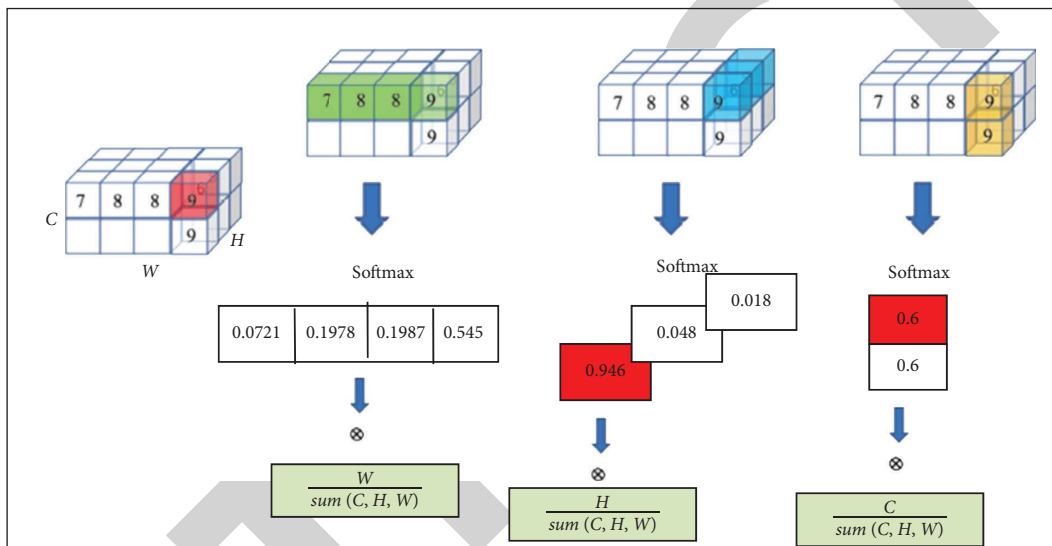


FIGURE 7: Schematic diagram of multidimensional attention.

set and test set. The specific data distribution is shown in Table 1. The training set contains 18462 images, and the validation set contains 7,915 photographs. Table 2 depicts the experimental environment's specific setup. NVIDIA Tesla V100 graphics card with 32 GB video memory is used, and the Paddle 1.8.3 version is used for the deep learning framework. The rate is 0.001, the loss function adopts cross-entropy loss, and finally, the neural network is trained with 60 images per batch (batch size), and the training period (epoch) is 30 rounds. Table 3 shows the experimental parameters.

The NVIDIA V100 Tensor Core is the most sophisticated data center GPU yet created, capable of accelerating AI, high-performance computing (HPC), and graphics. It is driven by NVIDIA Volta architecture, is available in 16 and 32 GB variants, and can execute up to 100 CPUs in a single GPU. Data scientists, academics, and engineers may now focus on building the next AI breakthrough rather than minimizing memory utilization. Data scientists are tackling more sophisticated AI tasks, such as voice recognition, virtual personal assistant training, and teaching self-driving cars. Solving this kind of challenge necessitates training deep

TABLE 1: Data distribution.

Apple leaf disease	Data set	Training set	Validation set
Mosaic disease	5875	4412	2463
Brown spot disease	6655	4958	2697
Rust	6694	4985	2709
Gray spot disease	5810	4367	2443
Spotted leaf disease	6343	4740	2603
Total	31377	23462	12915

learning models with exponentially increasing complexity in a reasonable period. The V100 GPU is the first in the world to achieve the deep learning performance of 100 teraFLOPS (TFLOPS) thanks to its 640 Tensor Cores. The most powerful processing servers in the world are made possible by the newest iteration of NVIDIA NVLINK, which links several V100 GPUs at speeds of up to 300 GB/s. On earlier systems, training AI models would take weeks of computer time. Now, it just takes a few days. With this significant decrease in training time, AI will now be able to solve a vast array of issues.

TABLE 2: Experimental configuration.

Configuration items	Parameter values
CPU	Intel® Xeon® gold 6271C CPU @ 2.60 GHz
Graphics card	NVIDIA tesla V100 GPU 32 GB
Memory	32 GB
Disk	100 GB
Deep learning framework	Paddle 1.8.3
Operating system	Ubuntu 16.04 LTS
Other tools	Python 3.7, CUDA10.1

TABLE 3: Experimental parameter.

Parameter names	Parameter values
Size of image	224 × 224
Size of batch	70
Rate of learning	0.001
Training period (epoch)	40

3.2. Experimental Results and Analysis. To verify the effectiveness of the proposed DBNet method, this method is compared with the commonly used convolutional neural network models, namely, VGG-16, AlexNet, ResNet-50, and B-CNN. On the apple leaf disease dataset, each model is trained from scratch until the model converges, and the training conditions of each model are guaranteed to be the same. In the model training process, each time the training set completes a training iteration, a test is performed on the validation set, which can more intuitively display the performance changes of the model during the training process. Figure 8 with Tables 4–7 shows the changes in the loss function and accuracy of various network models on the apple leaf disease dataset. Among them, loss refers to the loss function, Epoch refers to the training period, and accuracy (acc) refers to the accuracy rate.

It can be seen from Figure 8 that the models of DBNet, AlexNet, VGG-16, ResNet-50, and B-CNN all converged within 30 rounds of training cycles. Although the DBNet model has more attention modules and multiscale parts than other models, it does not affect the model loss function's decline speed and convergence speed. From the training and validation data results, it can be seen that DBNet does not have serious overfitting due to its model structure in the apple leaf disease data.

The accuracy rates of VGG-16, DBNet, ResNet-50, AlexNet, and B-CNN network models are shown in Table 8. The DBNet proposed in this article has attained the highest recognition accuracy of 97.662% on the apple leaf disease dataset. Compared with VGG-16, ResNet-50, AlexNet, and B-CNN, the recognition accuracy is improved by 0.10764, 0.05407, 0.04649, respectively, from 0.0312. The effectiveness of the DBNet method in apple leaf disease identification has been verified.

Figure 9 with table data in 6 depicts the variance in the accuracy of different models. In the figure, each network model shows an overall upward trend. In the first 20 training cycles, the accuracy of each network model fluctuates wildly.

The accuracy improves significantly in the last 10 training cycles and gradually stabilizes. AlexNet contains five convolutional layers and three fully connected layers; the network depth is eight layers; and the accuracy rate on the dataset is only 86.898%. The VGG-16 network depth is sixteen layers, which is 0.05357 higher than that of AlexNet, indicating that the network depth significantly influences the accuracy. Experimenting with the deeper ResNet-50 network, the accuracy rate is improved by 0.00758 compared with VGG-16, and the improvement is insignificant. Therefore, after the network depth reaches a certain level, the neural network is deepened again, and the effect is not expected to be considerable. Still, with the web, as the model's depth increases, the model becomes more complex. Compared with the single-branch network ResNet-50 with the highest accuracy, the B-CNN network composed of two VGG-16s in the experiment is improved by 0.01529, indicating that the dual-branch network is better than the single-branch in the task of apple leaf disease identification. The dual-branch network can extract more image features. The DBNet proposed, which includes the MS branch and DA branch, can better alleviate the adverse effects of complex background environment and the similarity of lesions compared with the two VGG-16 branches of the B-CNN, and the accuracy rate is also high. It reached 97.662%.

To verify the complexity of the DBNet model, this paper analyses it in terms of model parameters (Params), floating-point operations (FLOPs), and CPU prediction time, as shown in Table 9. AlexNet is affected by the depth of the network, and the model parameters, the number of floating-point operations, and the prediction time are the least affected, but its accuracy is also affected. ResNet-50 adopts the bottleneck module, which reduces the number of parameters compared with VGG-16, and B-CNN makes the model too complicated due to the dual branch network and bilinear pooling. Although the DBNet in this article uses two VGG-16 as the primary network, the DA branch in DBNet replaces the standard convolution in VGG-16 with a depth-wise separable convolution, and the MS branch uses a combination of shallow information and in-depth information. The cross-layer connection dramatically reduces the number of model parameters, floating-point operations, and prediction time.

3.3. Ablation Experiment. To verify the influence of different combinations of convolution types on the MS branch, 9 groups of ablation experiments were set up, and the results are shown in Table 10. When a single 3×3 convolution kernel is increased to multiple convolution kernels, the network width in the MS branch increases. The hole convolution experiment with 3 different hole rates is set to have an upward trend; when the hole rate $d = 4$, the accuracy rate decreases, and finally, the whole convolution with hole rate $d = 2, 3$ is selected. The 3×1 and 1×3 asymmetric convolutions boost the nonlinearity of the network, improve the expressiveness of the model, and additionally learn the features of the horizontal and vertical kernels, increasing the diversity of captured features and thereby improving the

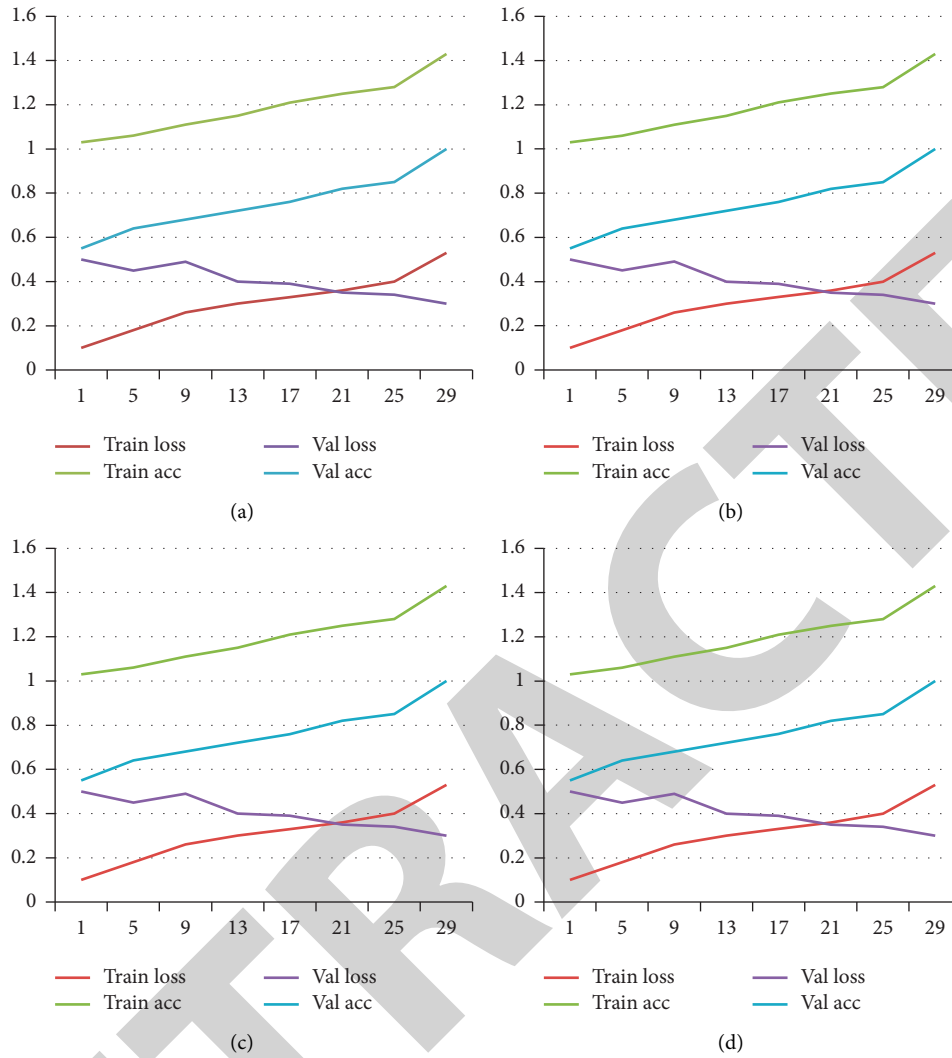


FIGURE 8: (a) AlexNet loss and accuracy, (b) VGG-16 loss and accuracy, (c) ResNet-50 loss and accuracy, and (d) training results of different network models.

TABLE 4: AlexNet loss and accuracy.

Serial	Train loss	Train acc	Val loss	Val acc
1	0.3	0.4	0.33	0.39
5	0.35	0.48	0.36	0.44
9	0.4	0.52	0.41	0.48
13	0.45	0.56	0.45	0.52
17	0.5	0.59	0.51	0.56
21	0.55	0.62	0.55	0.62
25	0.6	0.66	0.58	0.65
29	0.65	0.69	0.63	0.7

TABLE 5: VGG-16 loss and accuracy.

Serial	Train loss	Train acc	Val loss	Val acc
1	0.2	0.3	0.23	0.29
5	0.25	0.38	0.26	0.34
9	0.3	0.42	0.31	0.38
13	0.35	0.46	0.35	0.42
17	0.4	0.49	0.41	0.46
21	0.45	0.52	0.45	0.52
25	0.5	0.56	0.48	0.55
29	0.55	0.69	0.63	0.7

classification accuracy. To verify the influence of the MS branch and DA branch on DBNet network performance, four kinds of network structures VGG + VGG, VGG + MS, VGG + DA, and MS + DA ablation experiments were set up, respectively. After 30 rounds of training, the accuracy changes of the four network structures in the validation set are shown in Figure 9. The accuracy rates of the VGG + MS structure and VGG + DA structure are better than those of

the VGG + VGG structure, and the MS + DA structure has the highest accuracy, indicating that the MS branch and DA branch have a promoting effect on apple leaf disease identification.

Table 11 identifies different apple leaf diseases by the four network structures of VGG + VGG, VGG + MS, VGG + DA, and MS + DA. In Table 11, the accuracy of the VGG + MS structure in the identification of five kinds of apple leaf

TABLE 6: ResNet-50 loss and accuracy.

Serial	Train loss	Train acc	Val loss	Val acc
1	0.1	0.03	0.09	0.19
5	0.18	0.06	0.14	0.24
9	0.22	0.11	0.18	0.28
13	0.26	0.15	0.22	0.32
17	0.29	0.21	0.26	0.36
21	0.32	0.25	0.32	0.42
25	0.36	0.28	0.35	0.45
29	0.49	0.43	0.5	0.6

TABLE 7: Training results of different network models.

Serial	Train loss	Train acc	Val loss	Val acc
1	0.1	1.03	0.5	0.55
5	0.18	1.06	0.45	0.64
9	0.26	1.11	0.49	0.68
13	0.3	1.15	0.4	0.72
17	0.33	1.21	0.39	0.76
21	0.36	1.25	0.35	0.82
25	0.4	1.28	0.34	0.85
29	0.53	1.43	0.3	1

TABLE 8: Accuracy of the network model.

Network model	Accuracy (accuracy) %
AlexNet	96.898
VGG16	82.255
ResNet50	83.013
B-CNN	84.542
DBNet	87.662

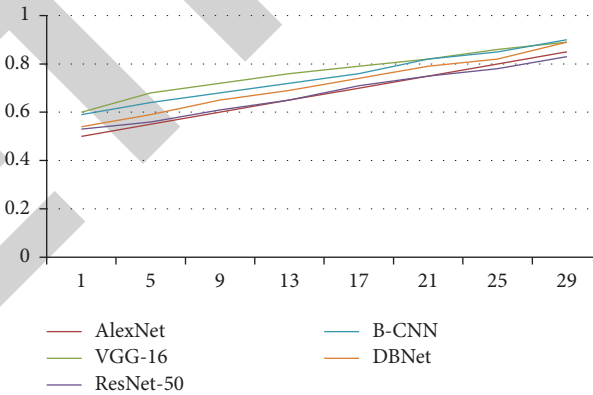


FIGURE 9: Comparison of accuracy changes of different network models.

TABLE 9: Network complexity.

Network model	Model parameters (10^6)	Number of floating-point operations (10^9)	CPU prediction time (ms)
AlexNet	6.7748	2.17	29
ResNet50	33.5183	9.19	180
VGG16	238.35 6	31.84	502
B-CNN	391.58 8	62.69	1139
DBNet	24.032	18.89	226

TABLE 10: Comparison of accuracy changes of different network models.

Serial	AlexNet	VGG-16	ResNet-50	B-CNN	DBNet
1	0.5	0.6	0.53	0.59	0.54
5	0.55	0.68	0.56	0.64	0.59
9	0.6	0.72	0.61	0.68	0.65
13	0.65	0.76	0.65	0.72	0.69
17	0.7	0.79	0.71	0.76	0.74
21	0.75	0.82	0.75	0.82	0.79
25	0.8	0.86	0.78	0.85	0.82
29	0.85	0.89	0.83	0.9	0.89

TABLE 11: Ablation experiments.

Apple leaf disease	VGG + VGG	VGG + MS	VGG + DA	MS + DA
Mosaic disease	98.658	100	100	100
Brown spot disease	97.994	99.055	99.232	99.939
Rust	94.68	95.499	95.675	96.846
Gray spot disease	96.357	96.564	98.228	98.643
Spotted leaf disease	94.637	95.448	96.757	98.005
Five apple leaf diseases	95.049	97.058	97.791	98.662

TABLE 12: Comparison of experiment results.

Models	Accuracy (Accuracy) %
[18]	95.819
[19]	96.25
[20]	97.222
[21]	97.412
DBNet	96.662

diseases was higher than that of the VGG + VGG structure, indicating that MS branches can extract multiscale features and alleviate the adverse effects of complex background environments and uneven distribution of lesion shape [17]. On the contrary, the VGG + DA structure has high recognition accuracy for rust, grey spot, and spotted leaf disease, indicating that the DA branch can use a multidimensional attention method to alleviate the adverse effects of the similarity of lesions effectively. Therefore, this article combines the MS branch and the DA branch to form a DBNet, which can refine the recognition accuracy effectively and alleviate the adverse effects of complex background environments and lesion similarity.

As given in Table 12, when comparing the DBNet network to other leaf disease detection models, its accuracy improves by 0.02843, 0.02412, 0.0144, and 0.0125, respectively. This clearly shows that the proposed DBNet model has certain advantages as compared to others in detecting apple leaf disease.

4. Conclusion

The article presented a dual-branch network identification method for apple leaf diseases. The presented method can effectively combine the multiscale and multidimensional attention features of lesion images to effectively alleviate the complex background environment and the similarity of lesions to apple leaf diseases. It is verified through experiments that the presented method has high recognition

accuracy. At the same time, a new attention mechanism is proposed, which can extract attention features in three dimensions: channel, width, and height. The experimental comparison demonstrated that the presented method may enhance the recognition accuracy effectively. Future work may include investigations incorporating the multidimensional attention mechanism into additional computer vision tasks such as semantic segmentation.

Data Availability

The data are made available on upon reasonable request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to acknowledge the support provided by Researchers Supporting Project Number (RSP2023R358) King Saud University, Riyadh, Saudi Arabia.

References

- [1] V. V. Srinidhi, A. Sahay, and K. Deeba, "Plant pathology disease detection in apple leave using deep convolutional neural networks: apple leaves disease detection using EfficientNet and DenseNet," in *Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1119–1127, Erode, India, April 2021.
- [2] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks," *IEEE Access*, vol. 7, pp. 59069–59080, 2019.
- [3] 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.
- [4] K. Sangeetha, P. Vishnu Raja, P Rima, M. Pranesh Kumar, and S. Preethees, “Apple leaf disease detection using deep learning,” in *Proceedings of the 2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1063–1067, Erode, India, March 2022.
- [5] C. Chen, T. Zhu, S. Li, and B. Liu, “Apple leaf disease recognition method base on improved ShuffleNet V2,” in *Proceedings of the 2021 3rd International Conference on Advances in Computer Technology, Information Science and Communication (CTISC)*, pp. 276–282, Shanghai, China, April 2021.
- [6] M. Agarwal, R. K. Kaliyar, G. Singal, and S. K. Gupta, “Fcnn-lda: a faster convolution neural network model for leaf disease identification on apple’s leaf dataset,” in *Proceedings of the 2019 12th International Conference on Information & Communication Technology and System (ICTS)*, pp. 246–251, Surabaya, Indonesia, July 2019.
- [7] X. Yuan, C. Yu, B. Liu, H. Sun, and X. Zhu, “CGAN-IRB: a novel data augmentation method for apple leaf diseases,” in *Proceedings of the 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 192–200, Madrid, Spain, July 2021.
- [8] S. S. Chouhan, A. Kaul, U. P. Singh, and S. Jain, “Bacterial foraging optimization based radial basis function neural network (BRBFNN) for identification and classification of plant leaf diseases: an automatic approach towards plant pathology,” *IEEE Access*, vol. 6, pp. 8852–8863, 2018.
- [9] Y. Liu, M. A. Shah, A. Pljonkin, M. A. Iqbal, and M. Shabaz, “Design and research on the intelligent system of urban rail transit Project based on BIM+GIS,” *Scalable Computing: Practice and Experience*, vol. 22, no. Issue 2, 2021.
- [10] S. Mohanasundaram, E. Ramirez-Asis, A. Quispe-Talla, M. W. Bhatt, and M. Shabaz, “Experimental replacement of hops by mango in beer: production and comparison of total phenolics, flavonoids, minerals, carbohydrates, proteins, and toxic substances,” *International Journal of System Assurance Engineering and Management*, vol. 13, 2021.
- [11] N. Garrett, M. Nouri, A. Herrero, A. Gowen, and J. Roger, “Early detection of the fungal disease “apple scab” using SWIR hyperspectral imaging,” in *Proceedings of the 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4, Amsterdam, Netherlands, September 2019.
- [12] K. Phasinam, T. Kassanuk, and M. Shabaz, “Applicability of internet of things in smart farming,” *Journal of Food Quality*, vol. 2022, Article ID 7692922, 7 pages, 2022.
- [13] A. Kp and J. Anitha, “Plant disease classification using deep learning,” in *Proceedings of the 2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, pp. 407–411, Coimbatore, India, May 2021.
- [14] P. Ratta, A. Kaur, S. Sharma, M. Shabaz, and G. Dhiman, “Application of Blockchain and Internet of Things in healthcare and medical sector: applications, challenges, and future perspectives,” *Journal of Food Quality*, vol. 2021, Article ID 7608296, 20 pages, 2021.
- [15] T. Vijaykanth Reddy and K. Sashi Rekha, “Deep leaf disease prediction framework (DLDPF) with transfer learning for automatic leaf disease detection,” in *Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1408–1415, Erode, India, April 2021.
- [16] J. Lee, S. Y. Lee, R. E. Wijesinghe et al., “On-field in situ inspection for marssonina coronaria infected apple blotch based on non-invasive bio-photonics imaging module,” *IEEE Access*, vol. 7, pp. 148684–148691, 2019.
- [17] M. Yang, P. Kumar, J. Bhola, and M. Shabaz, “Development of image recognition software based on artificial intelligence algorithm for the efficient sorting of apple fruit,” *International Journal of System Assurance Engineering and Management*, vol. 13, 2021.
- [18] S. Divakar, A. Bhattacharjee, and R. Priyadarshini, “Smote-DL: a deep learning based plant disease detection method,” in *Proceedings of the 2021 6th International Conference for Convergence in Technology (I2CT)*, pp. 1–6, Maharashtra, India, April 2021.
- [19] D. V. Shivling, S. Kumar, C. Ghanshyam et al., “Low-cost sensor-based embedded system for plant protection and pest control,” in *Proceedings of the 2015 International Conference on Soft Computing Techniques and Implementations (ICS)*, pp. 179–184, Faridabad, India, October 2015.
- [20] Z. Zheng, S. Pan, and Y. Zhang, “Fruit tree disease recognition based on convolutional neural networks,” in *Proceedings of the 2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS)*, pp. 118–122, Shenyang, China, October 2019.
- [21] K. Swetha, V. Venkataraman, G. D. Sadhana, and R. Priyatharshini, “Hybrid approach for anthracnose detection using intensity and size features,” in *Proceedings of the 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, pp. 28–32, Chennai, India, July 2016.