An Improved DeepLab v3+ Deep Learning Network Applied to the Segmentation of Grape Leaf Black Rot Spots

Hongbo Yuan, Jiajun Zhu, Qifan Wang, Man Cheng* and Zhenjiang Cai

College of Mechanical and Electrical Engineering, Hebei Agricultural University, Baoding, China

The common method for evaluating the extent of grape disease is to classify the disease spots according to the area. The prerequisite for this operation is to accurately segment the disease spots. This paper presents an improved DeepLab v3+ deep learning network for the segmentation of grapevine leaf black rot spots. The ResNet101 network is used as the backbone network of DeepLab v3+, and a channel attention module is inserted into the residual module. Moreover, a feature fusion branch based on a feature pyramid network is added to the DeepLab v3+ encoder, which fuses feature maps of different levels. Test set TS1 from Plant Village and test set TS2 from an orchard field were used for testing to verify the segmentation performance of the method. In the test set TS1, the improved DeepLab v3+ had 0.848, 0.881, and 0.918 on the mean intersection over union (mIOU), recall, and F1-score evaluation indicators, respectively, which was 3.0, 2.3, and 1.7% greater than the original DeepLab v3+. In the test set TS2, the improved DeepLab v3+ improved the evaluation indicators mIOU, recall, and F1-score by 3.3, 2.5, and 1.9%, respectively. The test results show that the improved DeepLab v3+ has better segmentation performance. It is more suitable for the segmentation of grape leaf black rot spots and can be used as an effective tool for grape disease grade assessment.

Keywords: grape black rot, semantic segmentation, DeepLab V3+, channel attention, feature pyramid network

INTRODUCTION

Grapes are one of the most grown economic fruits in the world. Grapes are often used in the production of wine, fermented beverages, and raisins (Kole et al., 2014). In the cultivation of grapes, the larger the area planted, the larger the scale of damage when a disease occurs as well as the greater the economic losses caused. Black rot, which is a fungal disease, is one of the most important grape diseases in the world (Molitor and Berkelmann-Loehnertz, 2011). Black rot spots are black in color and have a small spot area compared to grape leaves. Generally, the assessment of black rot damage on grapes is done by judging the size of the spot on the leaves. This operation is currently performed mainly by hand. However, the manual assessment of spot size and leaf damage area is highly subjective, difficult to quantify, and inefficient. The use of computers and image processing techniques for the identification and segmentation of black rot spots on grapevine leaves can facilitate rapid and

accurate assessment of damage for targeted treatment, which is important for ensuring grapevine yield and growers' economic incomes.

The methods of image segmentation have experienced three basic stages from classic segmentation methods, machine learning method, and deep learning method with the development of image processing and computer technology. These methods have been applied in agricultural disease detection. The classical image segmentation, such as threshold segmentation (Mehl et al., 2002; Kim et al., 2005), usually uses color and texture features (Samajpati and Degadwala, 2016) to separate the disease spots from the background. Chaudhary et al. (2012) transformed the RGB image into CIELAB, HIS, and YCbCr color space according to the different color features between the disease spots and leaf, respectively. Then the disease spots were segmented with threshold calculated by the OTSU method based on color features. Ma et al. (2017) achieved segmentation of disease spots from the background by fusion features of the super red index, the H-component of HSV, and the b-component of color space for the greenhouse vegetable images with 97% accuracy. Jothiaruna et al. (2019) proposed a method that integrated color features and region growing for the segmentation of leaves disease spots with an average segmentation accuracy of 87%. Sinha and Shekhawat (2020) segmented peacock disease spots on olive leaves according to the different textures of the leaves and spots, and the purpose of disease detection was realized. The classical image segmentation methods require high image quality, and the recognition result will be poor or even invalid if the environmental conditions changed when the image acquiring. Therefore, the generality and robustness of those methods are unsatisfactory, and the accuracy in practical application is not guaranteed.

With the development of machine learning, many researchers began to try to apply it to disease spots segmentation to improve the accuracy and robustness of segmentation. Zhou et al. (2014) inputted the color histogram of the image into the support vector machine (SVM) model to segment the Cercospora disease spots for sugar beet, and the average accuracy, recall, and F value were more than 0.87. Bai et al. (2017) used a fuzzy C-means algorithm for segmentation of cucumber leaves spots disease in complex backgrounds, and the experimental results showed that the average error did not exceed 0.12%. Pan et al. (2019) segmented pear blackspot disease in hyperspectral images using SVM with an overall accuracy of 97.5%. Singh (2019) applied a particle swarm optimization algorithm for the segmentation of downy mildew spots in sunflower leaves with an average accuracy of 98%. Appeltans et al. (2021) removed soil pixels from hyperspectral images by linear discriminant analysis classification and used a logistic regression supervised machine learning classifier for pixel classification of leek leaves to segment the spots of leek white tip disease with an accuracy of 96.74%. Machine learning methods can achieve satisfactory segmentation results using small sample size, but these methods require multiple steps of image preprocessing and are relatively complex to execute. In addition, the machine learning-based segmentation methods are relatively weakly adapted to unstructured environments

and need researchers to manually design feature extraction and classifiers, which makes the work more difficult.

With the improvement of computer hardware performance, deep learning has been developed rapidly (Lecun et al., 2015). Common deep learning algorithms are full convolutional neural network algorithm (FCN; Long et al., 2015), DeepLab (Chen et al., 2017), U-Net (Ronneberger et al., 2015), V-Net (Milletari et al., 2016), USE-Net (Rundo et al., 2019), SegNet (Badrinarayanan et al., 2017), etc. Lin et al. (2019) designed a semantic segmentation model based on convolutional neural network (CNN) for pixel-level segmentation of cucumber leaves powdery mildew disease spots, which provided a valuable tool for cucumber breeders to assess the severity of powdery mildew. Jiang et al. (2020) combined deep learning and SVM to segment the leaves disease images of four rice species with an accuracy of 96.8%. Wang et al. (2021) used DeepLab v3+ and U-Net methods to segment disease spots from cucumber leaves, and calculate their damage levels with an average accuracy of 92.85%. Lin et al. (2019) constructed a U-Net-based semantic segmentation model for cucumber powdery mildew spots segmentation with an average accuracy of 96.08%. Wspanialy and Moussa (2020) used U-Net neural network for segmentation of tomato leaves and spots in leaves with an average accuracy of 98% and then assessed the disease hazard level. Hu et al. (2021) segmented tea leaves and disease spots using a CNN and assessed the damage level. Liang et al. (2019) used PD²SE-Net neural network to segment plant disease spots areas and assessed their damage levels with an overall accuracy of more than 91%. The deep learning approach has all the work done by the CNN, which does not require too much pre-processing process or artificial selection of potential features compared to classical image processing methods and machine learning methods. The deep learning approach not only reduces the difficulty of plant leaves spots segmentation but also has higher accuracy and robustness.

Our group has developed a method to improve the recognition accuracy for grape leaf black rot by combine image enhancement technology and a deep learning network (Zhu et al., 2021). It can recognize the disease spots and calculate the number, but cannot segment the disease spots from the background. To realize the spot segmentation of grape leaf black rot, this paper designs a CNN based on an improved DeepLab v3+.

MATERIALS AND METHODS

Dataset and Test Environment Setup

The open dataset Plant Village (Hughes and Salathe, 2016) was used to perform experiments in this work, which provides symptoms of 26 common diseases on leaves of 14 plant species with a total of 54,309 RGB images. We selected 1,180 images of grape leaves infected with black rot as test subjects, and all these images were confirmed by researchers studying grape diseases. The selected images were taken in an indoor environment with a uniform gray background, and each image included only one frontal view of a grape leaf with 256×256 pixels. The areas of disease spots were manually labeled by

LabelMe (Russell et al., 2008) software. The average number of diseases present in an image was around 15, with more than 17,000 segmentation targets present in total. Before the experimental training, 1,180 data images were divided into training and test sets, and 1,072 images were selected for training the network and 108 images were selected as the test set for evaluating the network, which was named TS1. Furthermore, to increase the credibility of the model, a large number of images of grape leaves with disease spots from orchard sites were collected *via* the Internet. A total of 108 images of grape leaves with black rot spots in natural environments were selected by researchers studying grape diseases for an extra test set, which was named TS2. During the process of network training, the training set was divided into two parts in the form of training and validation data. The division ratio of training and validation data was 9:1. The training data were used for model fitting, and the validation data were used to adjust the super parameters of the model and to preliminarily evaluate the ability of the model. The test set was used to evaluate the generalization ability of the final model. In this study, the number of epochs was 120, the input batch was four, the learning rate was 0.001, and the size of the input image was 512×512. The VOC 2007 dataset format was used for the dataset. The experiments were conducted on Windows 10 with the Pytorch deep learning framework. The test computer contained an 8 GB GPU GeForce GTX 1070Ti and an AMD Ryzen 51600X Six-Core processor. Python language was used for programming.

Segmentation Method of Grape Leaf Black Rot Spots

To improve the segmentation performance of grapevine leaf black rot spots, a deep learning network based on the DeepLab v3+ was constructed. It is the third version of DeepLab, with high segmentation effectiveness and speed. In the improved DeepLab v3+ network constructed in this paper, the residual part in the backbone network ResNet101 incorporates a plug-and-play attention mechanism module. This improves the performance of various CNNs without increasing the complexity of the model. Moreover, a feature fusion branch based on a feature pyramid network (FPN) was added to the DeepLab v3+ encoder, which performs feature fusion on high-resolution and low-resolution feature maps. Finally, in the improved DeepLab v3+, one 4-fold up-sampling is replaced with two 2-fold up-sampling. Furthermore, the continuity of pixels in the obtained images is stronger and the network segmentation effect is improved.

Channel Attention Module

The efficient channel attention (ECA; Wang et al., 2020) module is a local cross-channel interaction strategy without dimensionality reduction, which can be efficiently implemented *via* one-dimensional (1D) convolution. The ECA module is obtained by improving on Squeeze-and-Excitation (SE; Hu et al., 2020), which is an effective channel attention learning method. It predicts a weight to be weighted for each output channel. The SE method

first uses global average pooling (GAP) for each feature channel individually to reduce the two-dimensional feature channel to a real number. Then, two fully-connected layers capture the non-linear cross-channel interaction. Finally, a Sigmoid function generates the channel weights with a value between 0 and 1. This weight is added to the feature channel as a weight to generate the next level of input data. The characteristic of SE is to use the correlation between channels instead of the correlation in the spatial distribution. By controlling the magnitude of the weight, the important features are enhanced and the unimportant features are weakened so that the extracted features are more directional. Compared with SE, the improvement of ECA is that the GAP operation of feature channels does not reduce the dimensionality. Instead, it captures local cross-channel interaction information by considering each channel and its K nearest neighbors. The ECA module can be used as a very lightweight plug-and-play module to improve the performance of various CNNs (Gao et al., 2020; Wang et al., 2020). Its implementation process is shown in **Figure 1**. The blue part uses GAP to aggregate convolutional features without performing dimensionality reduction operations. The ECA module can be efficiently implemented *via* a 1D convolution of size *k*, where the size of the convolution kernel *k* represents the coverage of local cross-channel interaction, that is, how many neighbors near the channel participate in the attention prediction of this channel. Wang et al. (2020) studied the *k* value of the CNN network with ResNet-101 as the backbone, and the *k* of the ECA module was set to 3, 5, 7, and 9 for training. The accuracy value was used to evaluate the effect of *k*. The experimental results showed that the accuracy was 78.47%, 78.58%, 78.0%, and 78.57% corresponding to the *k* value of 3, 5, 7, and 9, respectively. Therefore, *k* was set to 5 in this paper. The yellow part is the result of implementation *via* 1D convolution, and then the Sigmoid function can be used to generate the channel weights to obtain the normalized weights between 0 and 1. Finally, the original feature image X, whose matrix size is H×W×C, is multiplied by the weight generated by the Sigmoid function to obtain a new feature image X′, and the matrix size is H×W×C.

In this method, the backbone network of DeepLab v3+ is constructed using ResNet101, and an ECA module is inserted into the residual (Bottleneck; He et al., 2016) module of ResNet101. This method can realize the adaptive adjustment of the convolution kernel size in the channel of each residual block. The purpose is to improve the segmentation effect of the model. **Figure 2** shows a schematic diagram of the insertion of ECA in the residual module of ResNet101.

Feature Fusion Branching Based on a FPN

In the process of learning image features by CNNs, the resolution of the image is gradually reduced due to the deep convolution operation, resulting in low-resolution deep features at the output. In this way, there will be recognition errors for objects with a relatively small proportion of pixels in the image. The accuracy of multi-scale detection can be improved if the features at different levels of the network training process can be combined. An FPN (Lin et al., 2017) is a method that can fuse the feature maps of different layers. Feature maps that can reflect

FIGURE 1 | Efficient channel attention module.
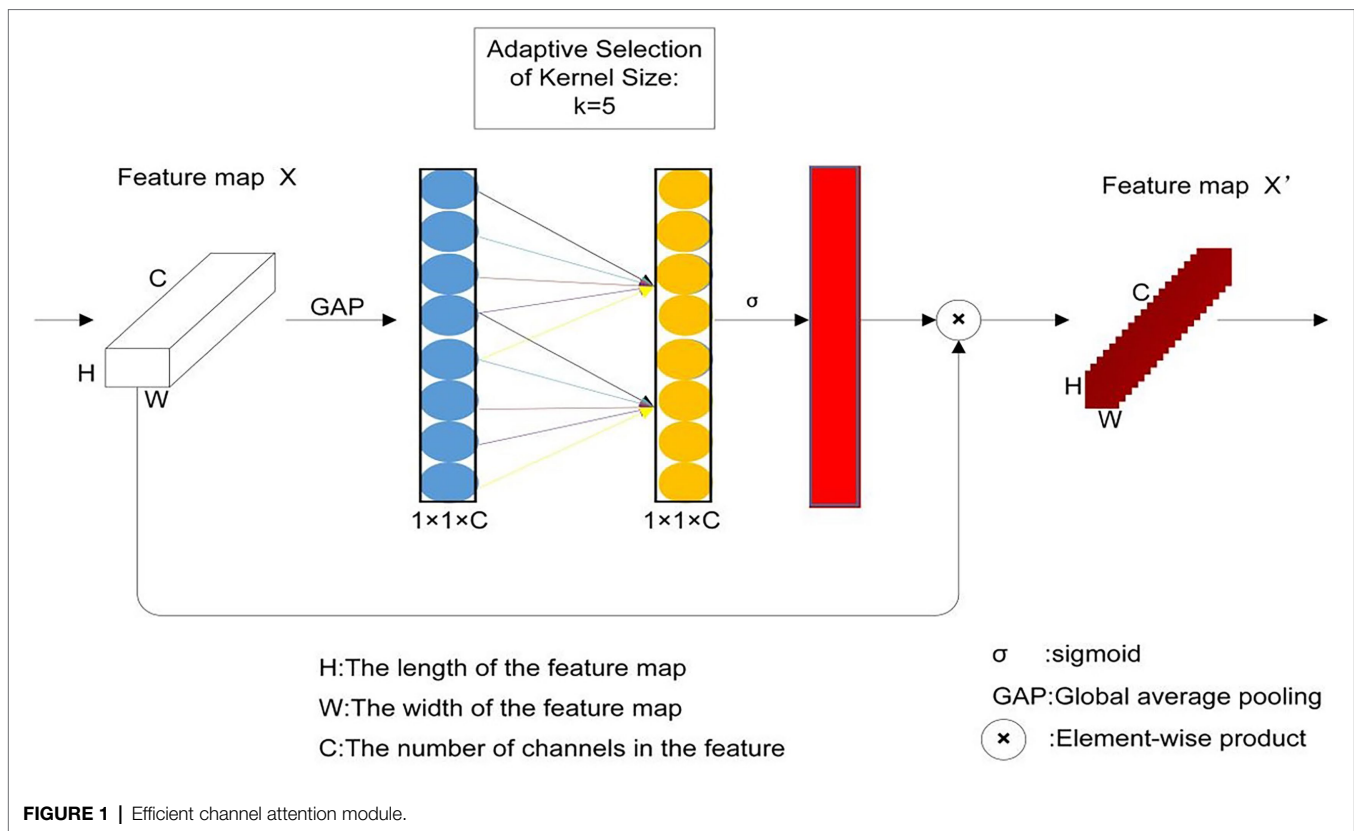
semantic information at different scales can be obtained through the fusion of FPNs. The feature fusion process of feature pyramids is shown in **Figure 3**. As shown, the left side is the feature maps of three different layers, whose resolutions become smaller from the bottom to the top. The middle part is the FPN, which can up-sample the deep-level features to convert them to the size of the shallow-level feature map and then fuses them with the shallow-level features. The right side is the feature map obtained after the FPN, which contains not only the deep level features but also the features of different levels. Here, the feature maps generated by Block3 and Block2 in the backbone network ResNet101 of DeepLab v3+ were fused. The feature map sizes of Block3 and Block2 were 1/16 and 1/8, and the number of channels was 1,024 and 512. In the FPN, the feature maps in Block3 and Block2 were subjected to 1×1 convolutional dimension reduction. The number of feature map channels in Block3 was changed from 1,024 to 256, and the number of feature map channels in Block2 was changed from 512 to 256. Then, the feature map of Block3 was up-sampled by a factor of 2 to change the size of the feature map from 1/16 to 1/8. Finally, the feature maps of Block3 and Block2 were combined to obtain the fused feature maps. The fused feature map has richer semantic and spatial information because it contains features from both levels, which can improve the segmentation effect of DeepLab v3+ network.

Improved DeepLab v3+ Network Structure

The improved DeepLab v3+ network consists of two parts, an encoder and decoder (Chen et al., 2018), which shows

in **Figure 4**. The encoder part trains the network, progressively obtains the feature maps, and captures higher-level semantic information. The decoder part semantically projects the features learned by the encoder part into the pixel space to achieve pixel segmentation. In the encoder, the backbone network is constructed using ResNet101 and the ECA module is inserted in its residual module. Moreover, to enhance the semantic information of the feature map, the feature maps of Block2 and Block3 of the ResNet101 network are fused. Atrous Spatial Pyramid Pooling (ASPP; Chen et al., 2018) is connected behind the ResNet101 backbone network. Dilated convolution with different sampling rates can be sampled in parallel by ASPP, which is equivalent to capturing the context of images at multiple scales. Dilated convolution (Yu et al., 2017) adds atrous to the convolution map during the convolution operation to expand the reception field so that each convolution output can contain a larger range of information. In addition to the convolution kernel, the dilated convolution also has a hyper-parameter dilation rate. It refers to the number of intervals between the convolution kernel during convolution mapping, that is, the number of atrous inserted. **Figure 5** shows the execution process of convolution. Here, **Figure 5A** is the standard convolution process and **Figure 5B** is the process of dilated convolution.

The encoder module has three outputs. The first is the low-level feature (LF) output by Block1 in the backbone network. The second is the fusion feature (FF) of Block2 and Block3 output by the FPN. The last one is the high-level
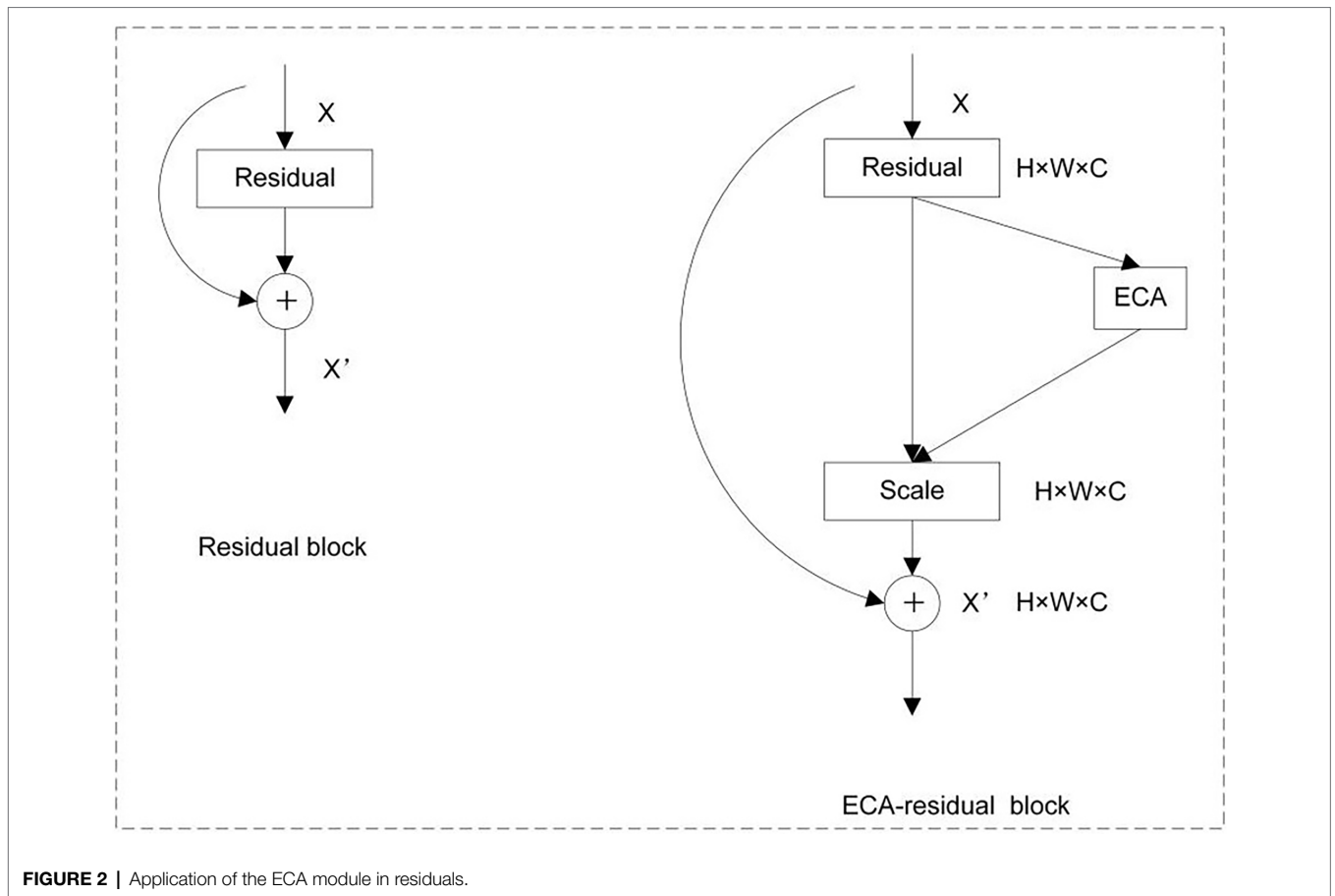
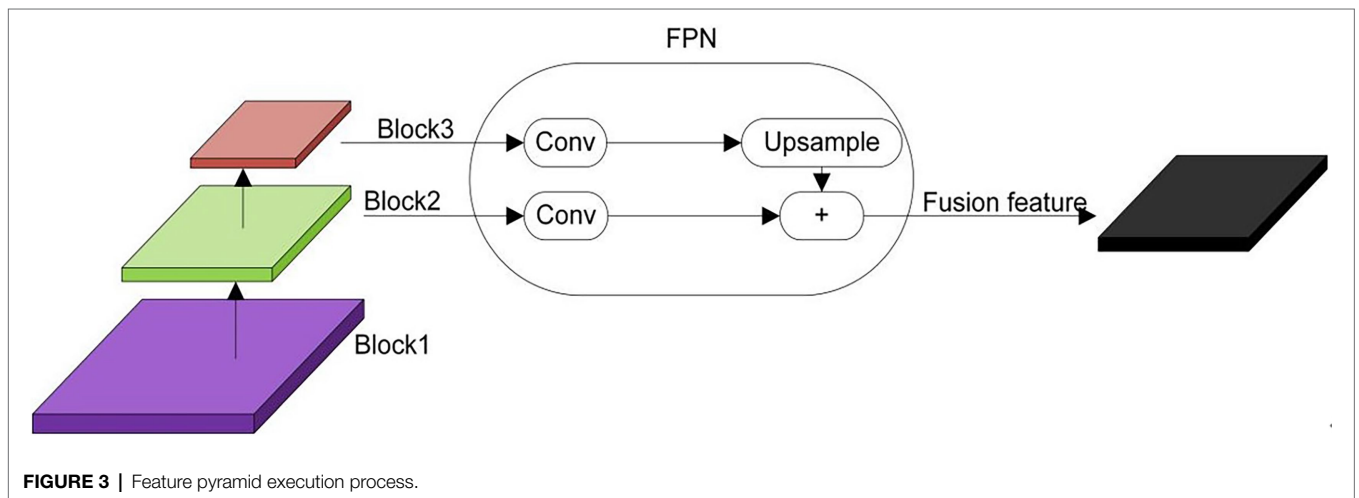**FIGURE 2 |** Application of the ECA module in residuals.

FIGURE 3 | Feature pyramid execution process.

feature (HF) output by the ASPP module after 1×1 convolution. High-level feature output concatenates to FF after it has undergone 2-fold up-sampling, and then the second 2-fold up-sampling is performed. The result of this operation is concatenated to the LF, which has been convoluted by 1×1 convolution. A 3×3 convolution is performed after the above operation, and then a single four-fold up-sampling is performed.

Then, the dense classification of pixels is obtained, which is image segmentation.

Parameters Setting of Improved DeepLab v3+ Network

The stochastic gradient descent method was applied to the end-to-end training of the deep learning network, and the
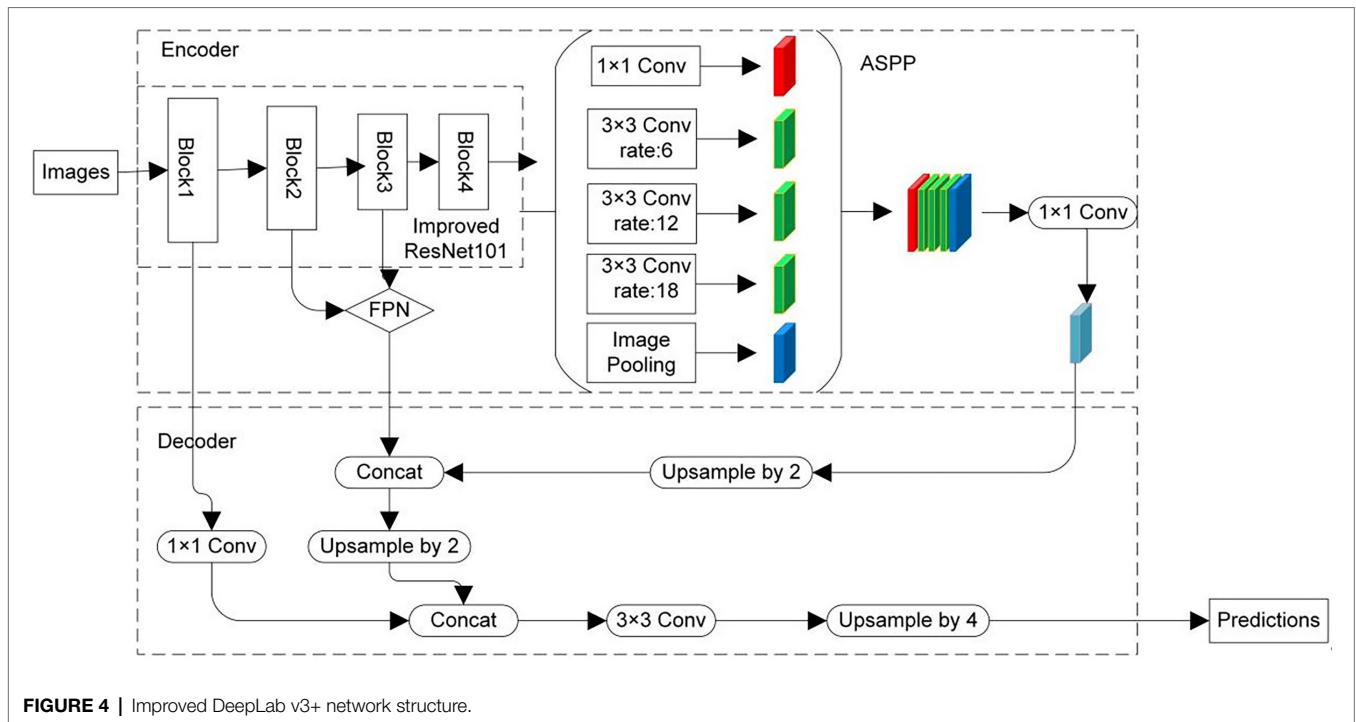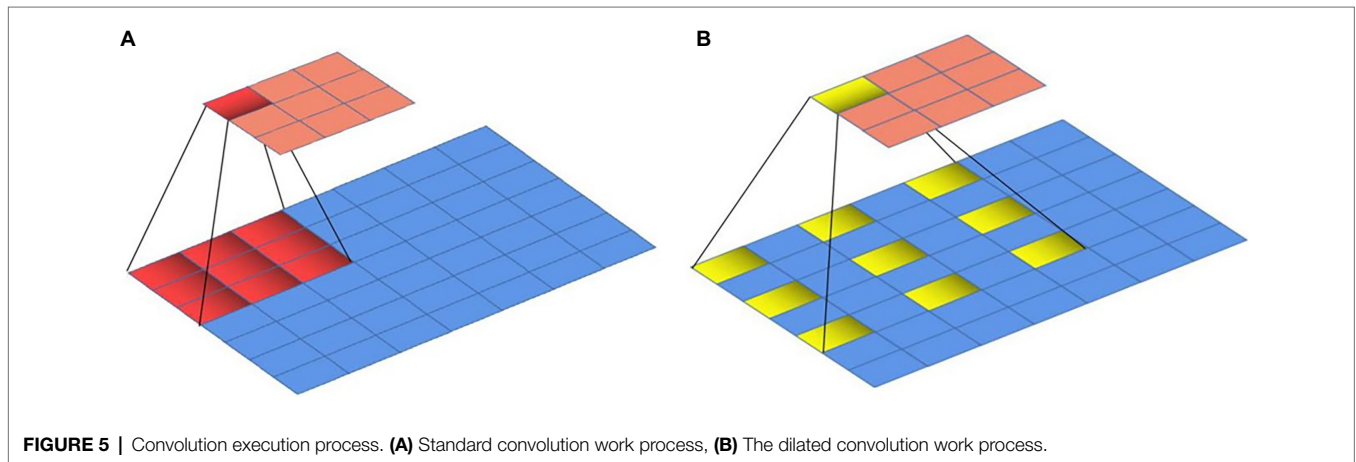
FIGURE 4 | Improved DeepLab v3+ network structure.

FIGURE 5 | Convolution execution process. **(A)** Standard convolution work process, **(B)** The dilated convolution work process.

loss function was set to Dice_Loss as shown in Equation (1). The weight decay rate was set to 0.001, and the kinetic energy factor was set to 0.8. The initial learning rate was set to 0.001, the learning rate decay mode was exponential decay, and the Batch_size was set to 4. The maximum iteration period (Epochs) was set to 120, and the network input size was set to 512×512. The data set was stored in the format of the VOC 2007 data set, and pre-trained model weights were loaded in the experiment to speed up the convergence of the model.

$$Dice\_Loss = \frac{FP + FN}{FP + 2TP + FN} \qquad (1)$$

where TP represents the true positives, indicating that the black rot area of grape leaves automatically segmented by the model overlaps with the real disease area; FP represents the false positives, indicating that the model misidentified the background area as a black rot spot area and segmented it; TN represents the true negatives, indicating that the model identified the real background area as the background area; and FN represents the false negatives, indicating that the model misidentified the real black rot area as the background area.

Evaluation Indicators

In this study, to evaluate the performance of the improved DeepLab v3+ network segmentation, the mean intersection over union (mIOU), the dice coefficient (Dice), the pixel accuracy (ACC), precision (*P*), recall (*R*), and *F1*-score were selected as evaluation metrics.

The mIOU is a common evaluation metric in semantic segmentation methods. In semantic segmentation, the predicted and true regions are obtained by pixel operation, and Equation (2) is as follows:

$$mIOU = \frac{1}{2} \sum_{i=0}^{1} \frac{p_{ii}}{\sum_{j=0}^{1} p_{ij} + \sum_{j=0}^{1} p_{ji} - p_{ii}} \qquad (2)$$

where p_{ij} denotes the number of pixels that originally belonged to class i but are predicted to be class j, p_{ii} denotes the number of pixels whose true label is class i predicted to be class i, and p_{ji} denotes the number of pixels that originally belonged to class j but are predicted to be class i. In this study, the pixels in each image were classified into two classes: black rot spots and background.

The Dice value is usually used to calculate the similarity of two samples, and the value range is (0,1). A Dice value close to 1 indicates a high set similarity, that is, the target is better segmented from the background; while a Dice value close to 0 indicates that the target cannot be effectively segmented from the background. The dice value equation is as follows:

$$Dice = \frac{2TP}{FP + 2TP + FN} \qquad (3)$$

The ACC is the ratio of the number of correctly predicted pixels to the total number of pixels in the category, and its equation is as follows:

$$ACC = \frac{TP + TN}{TP + FN + FP + TN} \qquad (4)$$

The P, R, and $F1$-score were calculated by the following equation:

$$\begin{cases} P = \dfrac{TP}{TP + FP} \\ R = \dfrac{TP}{TP + FN} \\ F1 - score = 2 \times \dfrac{P \cdot R}{P + R} \end{cases} \qquad (5)$$

Comparison of the Effects of Different Improvements of DeepLab v3+

To verify the effectiveness of the neural network constructed in this paper for grape leaf spot segmentation, eight sets of comparison experiments with different improvements were designed. These eight different improvements were named from Imp1 to Imp8, as shown in **Table 1**. In Imp1, the three dilated convolutions of the ASPP model of the original DeepLab v3+ network were modified to four dilated convolutions, and their dilated rate sizes were 4, 8, 12, and 16, respectively. Theoretically, the increase of dilated convolutions and the change of dilated

rate sizes will improve the fusion effect of semantic features. In Imp2, the ResNet 101, backbone of the DeepLab v3+, was replaced with Wide ResNet (Zagoruyko and Komodakis, 2016), which can improve the network segmentation performance by changing the width of the network without changing the network depth. The residual module of the backbone ResNet101 was inserted into the ECA module in Imp3, and the ECA model can adaptively adjust the convolutional kernel size in each channel of the residual block, which can improve the segmentation effect of the network. The coding side of the DeepLab v3+ network was added with a feature fusion branch based on the FPN in Imp4. The FPN can fuse different levels of feature maps and can obtain feature maps that can reflect semantic information at different scales. In imp5, the ASPP part of DeepLab v3+ was combined with DenseNet (Yang et al., 2018) to form DenseASPP, and the new module had a larger receiver field and more densely sampled points. Imp1, Imp3, and Imp4 were combined as Imp6. Imp3 and Imp5 were combined as Imp7. Imp3 and Imp4 were combined as Imp8, which is the improvement method used in this paper.

RESULTS

The Segmentation Results of Improved DeepLab v3+ for Grape Leaves Black Rot

The training dataset with annotation information was fed into the improved DeepLab v3+ network for training. The network was trained for 120 epochs, which required around 8.3 h. During the training process, the training model was saved once every 1 epoch, and a total of 120 completed models were saved. The convergence of the model can be reflected by the loss values generated during the training process. **Figure 6** shows the changes in the loss values of the training data and validation data in the training set during the training process. The training loss and validation loss gradually converged to stability during the training process, and the final training loss and validation loss values stabilized at 0.132.

TABLE 1 | Different DeepLab v3+ improvement methods.

Improvement methods	Improvement content
Imp1	Modify the three dilated convolutions of ASPP in the original network to four dilated convolutions with a dilated rate size of 4, 8, 12, and 16, respectively
Imp2	Replace the ResNet backbone in the original network with wider ResNet
Imp3	Insert the ECA module in the residual module of the backbone ResNet101
Imp4	A feature fusion branch based on an FPN is added to the coding side of the original network
Imp5	Combine the ASPP part of the original network with DenseNet to form DenseASPP
Imp6	Imp1 + Imp3 + Imp4
Imp7	Imp3 + Imp5
Imp8	Imp3 + Imp4

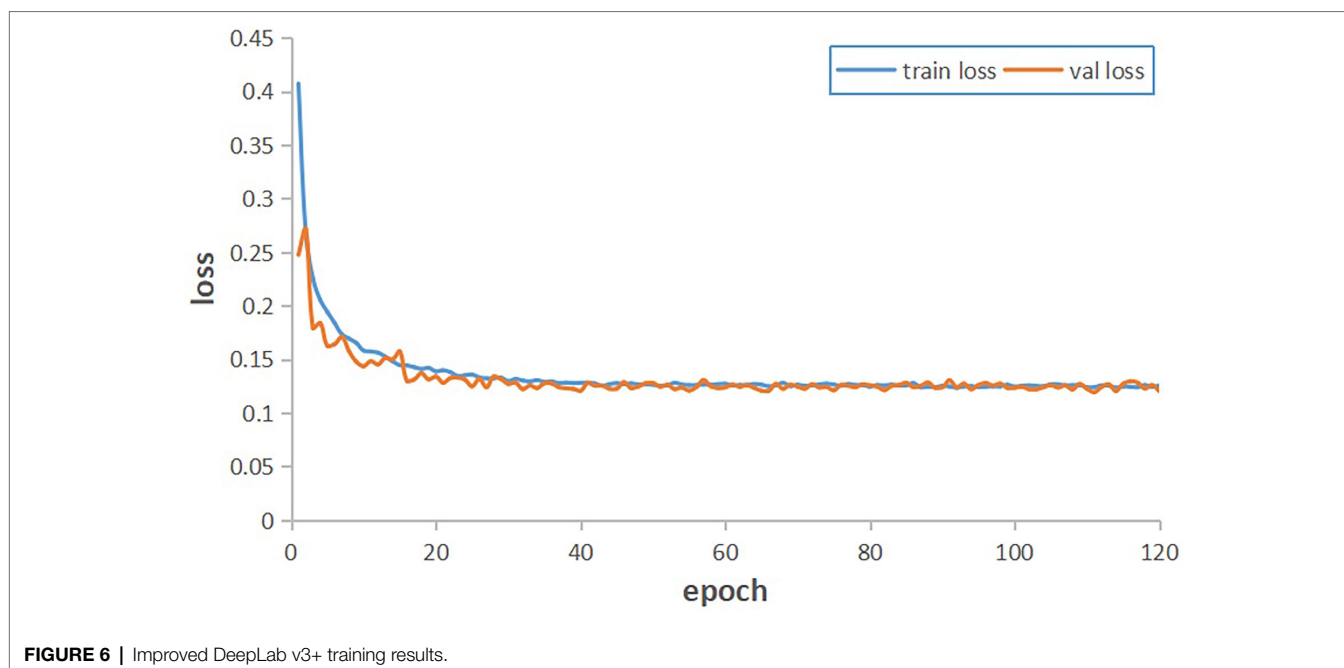**FIGURE 6 |** Improved DeepLab v3+ training results.

TABLE 2 | Statistics of the segmentation results of the test set TS1 by the before and after improved DeepLab v3+.

Algorithm	Evaluation indicators					
	mIOU	**ACC**	**Dice**	**P**	**R**	**F1-score**
DeepLab v3+	0.823	0.984	0.903	0.949	0.861	0.903
DeepLab v3+ (improved)	0.848	0.987	0.918	0.957	0.881	0.918

To verify the performance of the model, the optimal model at the end of training was selected to be used for segmentation trials on test set TS1. The statistical results of the experiment before and after improved DeepLab v3+ are shown in **Table 2**. As can be seen from **Table 2**, the improved DeepLab v3+ outperforms the pre-improvement DeepLab v3+ in all evaluation metrics. In particular, it improved 3.0, 2.3, and 1.7% in mIOU, R, and F1-score, respectively. The effects of the segmentation are shown in **Figure 7**.

Figure 8 shows the segmentation results of DeepLab v3+ before and after improvement applied to black rot spots of grape leaves in test set TS1. **Figure 8A** shows the original image, **Figure 8B** shows the manually labeled and segmented image, **Figure 8C** shows the segmentation results of DeepLab v3+ before improvement, and **Figure 8D** shows the segmentation results of DeepLab v3+ after improvement. The blue markers in **Figure 8** indicate the small spots targeted in the original image that were not identified and segmented by the original network model but were correctly segmented by the improved network model. The yellow markers indicate that the semantic segmentation network correctly identified and segmented some small spots in the original image even though they were not manually labeled and segmented due

to human oversight. This also demonstrates that the use of deep learning methods can reduce subjective errors caused by manual segmentation. The red markers indicate that the leaf edges were misidentified as spots and segmented by the network model due to shadows. This indicates that there is a requirement for background conditions for disease spot recognition using deep learning. Furthermore, **Figure 8** shows that although the improved network model could segment the spots at the same location, the improved network model was more accurate and the segmented spots overlapped more with the actual spots.

Experiments with the Plant Village dataset demonstrated that the improved DeepLab v3+, which incorporates an attention mechanisms and feature pyramids, could improve the segmentation of black rot spots on grape leaves. An additional dataset, TS2, with 108 images from photos taken in different orchard fields was used for testing to verify the effectiveness of the method in an orchard field setting. The TS2 dataset was tested experimentally using the DeepLab v3+ network before and after the improvement. **Figure 9** shows the experimental results of the DeepLab v3+ algorithm before and after the improvement on TS2. **Figure 9A** is the original image, **Figure 9B** is the unimproved DeepLab
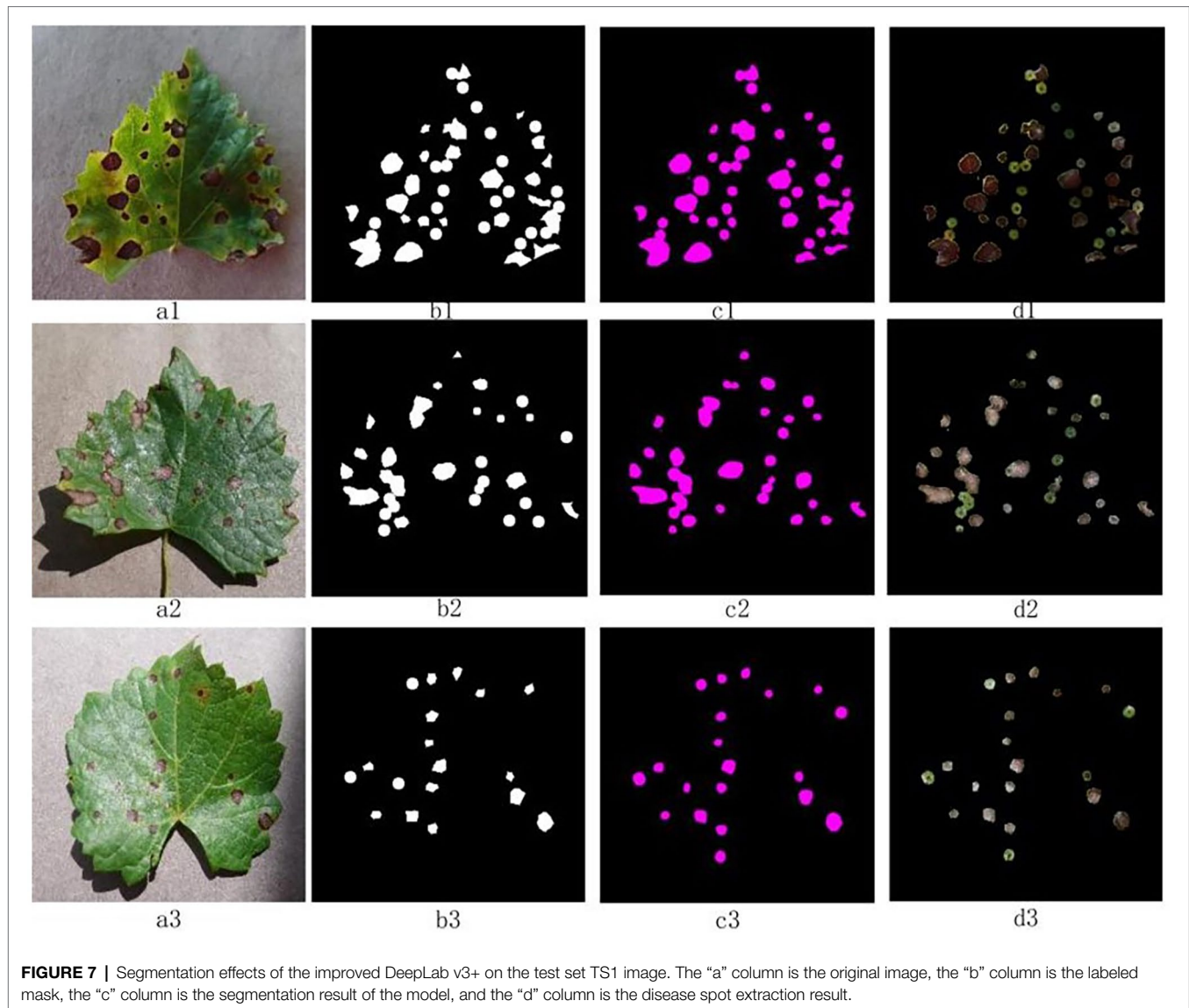
FIGURE 7 | Segmentation effects of the improved DeepLab v3+ on the test set TS1 image. The "a" column is the original image, the "b" column is the labeled mask, the "c" column is the segmentation result of the model, and the "d" column is the disease spot extraction result.

v3+ segmentation result, and **Figure 9C** is the improved DeepLab v3+ segmentation result. To show the network segmentation effect before and after the improvement, different colors are marked in **Figures 9B,C**. The yellow markers show that the improved network was more comprehensive in terms of the segmentation effect. The red markers show that the improved network was more accurate in segmentation. The blue markers show that the improved network was less affected by the background under the interference of complex background. The experimental results show that the improved DeepLab v3+ network performed better than the unimproved DeepLab v3+ network. Moreover, comparing the experimental segmentation effects shows that the improved DeepLab v3+ network can be applied to an actual orchard situation.

The statistical results of DeepLab v3+ before and after the improvement are shown in **Table 3** for test set TS2.

Table 3 shows that the improved DeepLab v3+ did not segment as well as TS1 for grape leaf black rot spots in a natural environment. This is because the images in TS1 were indoor environments, and the grape leaves were tiled with a single and simple background. In contrast, there were negative effects, such as overlapping leaves, gaps formed by shading, and lighting in the orchard field environment, which caused interference for accurate segmentation. Moreover, for large and dense spot areas, the network model would segment the dense spot areas as a whole; thus incorrectly classifying some backgrounds as spot areas. However, segmentation using the improved DeepLab v3+ still outperformed the one before the improvement, especially reaching scores of 0.756, 0.734, and 0.805 in mIOU, R, and $F1$-score, respectively, which were 3.3, 2.5, and 1.9% higher than those before improvement. This indicates that the proposed method improves the segmentation performance
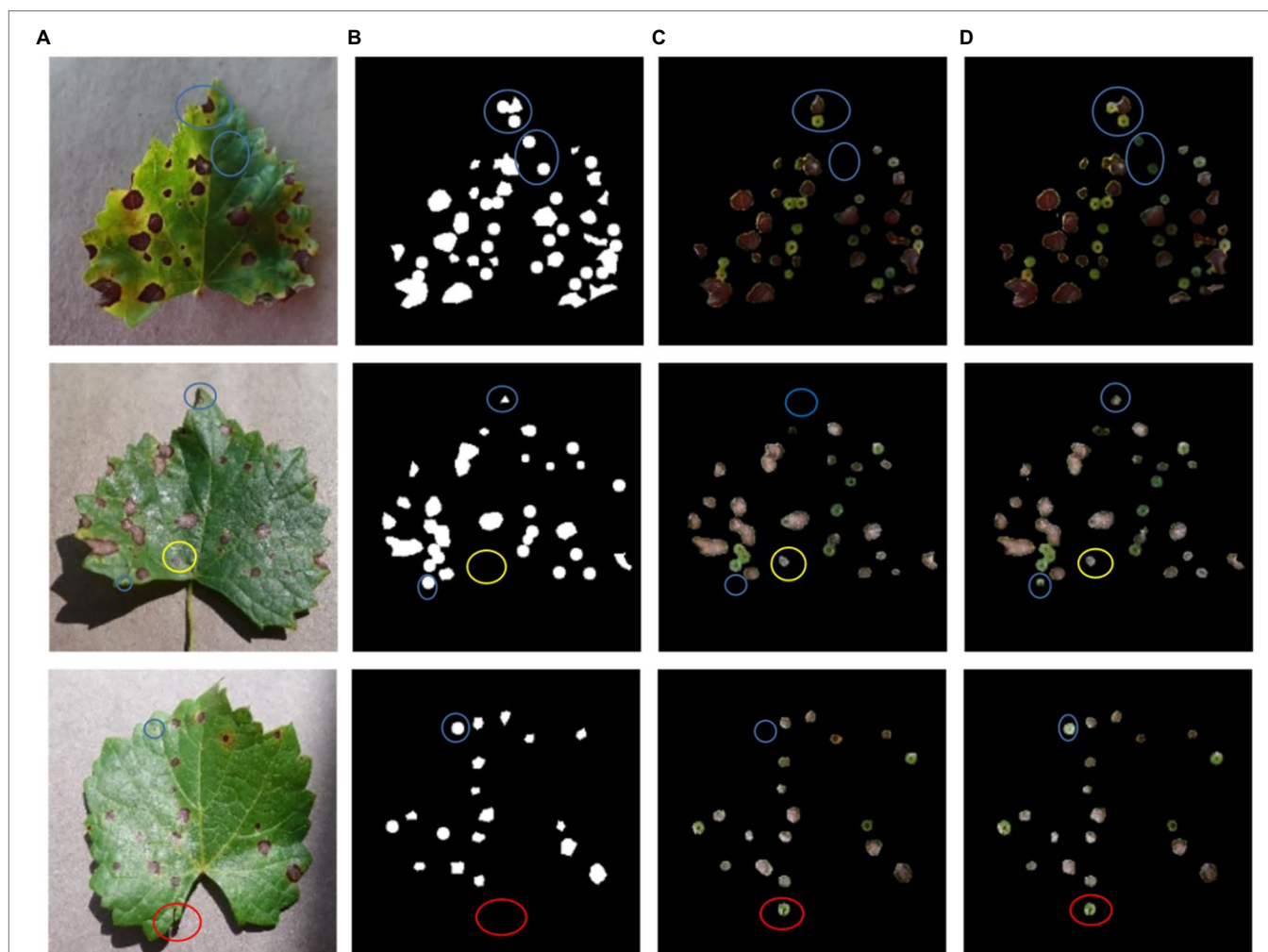
FIGURE 8 | A comparison of network training results before and after DeepLab v3+ improvement. **(A)** The original image, **(B)** the manually labeled and segmented image, **(C)** the DeepLab v3+ segmentation results, **(D)** the improved DeepLab v3+ segmentation results.

of DeepLab v3+, and its ubiquity and adaptability for application in a real environment are better compared with the unimproved network model.

Comparison of the Effects of Different Improvements of DeepLab v3+

For the above eight DeepLab v3+ improvement methods, the same training set was used for training, and the performances were tested with the test set TS1. To compare the results of different improvement methods, the parameters of the network, such as the learning rate, epoch, and batch size, were kept consistent during the experiments. The test results are shown in **Table 4**, where the four parameters mIOU, ACC, Dice, *P*, *R*, *F1*-score, and Pt are used for comparison. The Pt is the storage space occupied by the weight file generated after network training. **Table 4** shows that the performance indicators of the unimproved DeepLab v3+ on the test set TS1 were 0.823, 0.984, and 0.811 for mIOU, ACC, and Dice, respectively.

Table 5 shows that, compared with the DeepLab v3+ network before improvement, the scores of mIOU, ACC, and Dice were higher for the other six of the eight improved methods, except for Imp1 and Imp2. Compared with the DeepLab v3+ before improvement, Imp3 and Imp4 were 1.6% and 1.3% higher in mIOU and 0.5% and 1.3% higher in Dice, respectively. This indicates that fusing ECA or adding FPN in DeepLab v3+ network could improve the segmentation performance of the model. Although the improved method of Imp5 had improved mIOU and Dice by 1.4% and 1%, respectively. The Pt generated by this method required more memory space than that of Imp3 and Imp4. Moreover, Imp6 is a fusion of Imp1, Imp2, and Imp3, but its mIOU and Dice were lower than Imp3 and Imp4. This shows that the additional change of the dilated rate of the dilated convolution did not improve the performance of the network, which was consistent with the test results of Imp1. Besides, Imp7 is a fusion of Imp3 and Imp5, because fusing ECA in Imp3 alone or modifying ASPP to DenseASPP in Imp5 alone could improve network
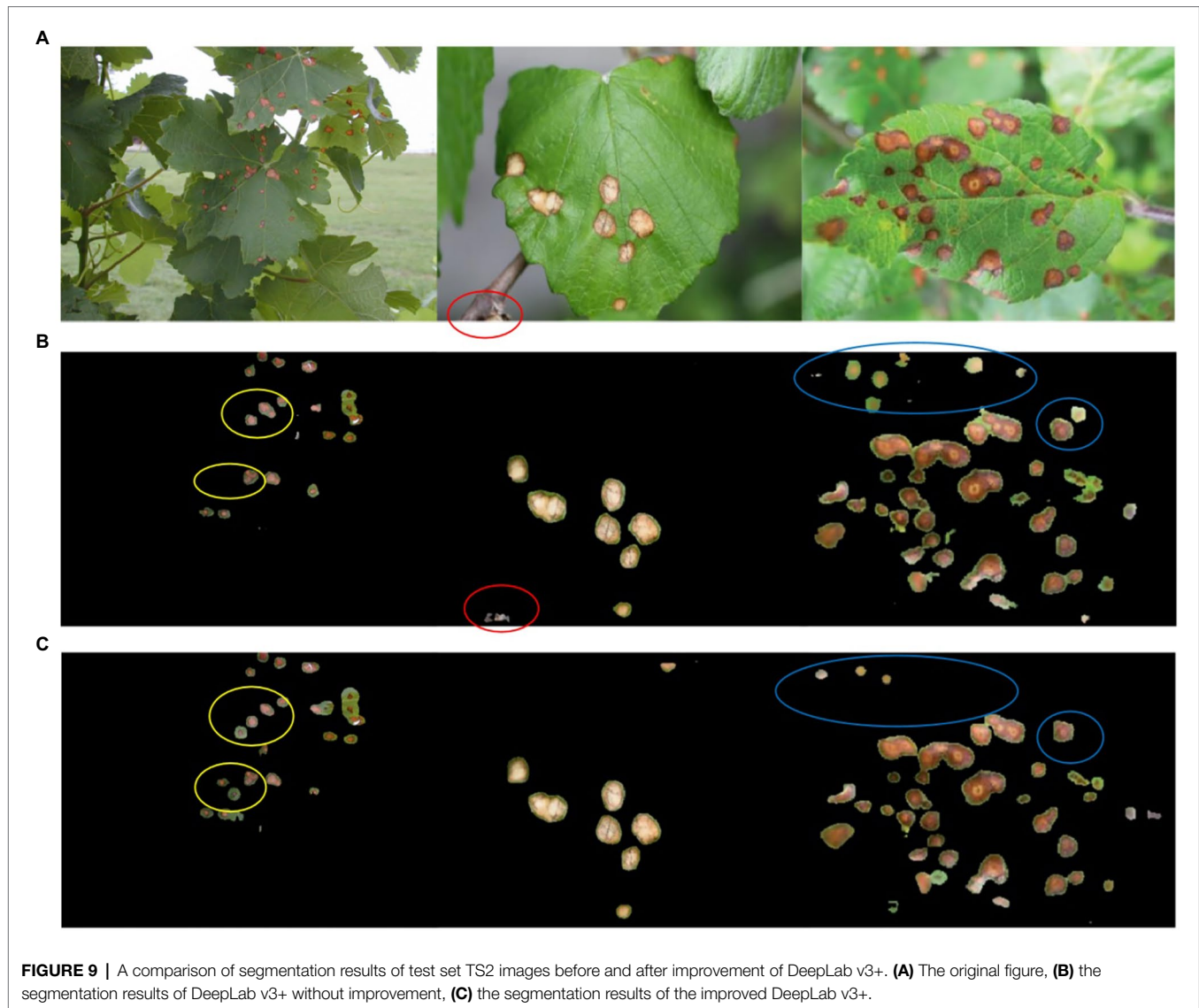
FIGURE 9 | A comparison of segmentation results of test set TS2 images before and after improvement of DeepLab v3+. **(A)** The original figure, **(B)** the segmentation results of DeepLab v3+ without improvement, **(C)** the segmentation results of the improved DeepLab v3+.

TABLE 3 | Statistics of the segmentation results of test set TS2 images before and after DeepLab v3+ improvement.

Algorithm	Evaluation indicators					
	mIOU	ACC	Dice	*P*	*R*	*F1*-score
DeepLab v3+	0.732	0.874	0.845	0.916	0.785	0.845
DeepLab v3+ (improved)	0.756	0.889	0.861	0.925	0.805	0.861

performance. Thus, Imp7 scored higher in mIOU than Imp3 and Imp5, and the Dice value was in line with Imp5 and higher than Imp3. However, the introduction of DenseASPP led to a larger computation within the network and its obtained weight file was relatively large, which was consistent with the performance of Imp5. The final improved method adopted in this paper was Imp8, which fuses Imp3 and Imp4 and adds both ECA and FPN in the DeepLab v3+ network. Here, Imp8 scored 0.848, 0.987, 0.918, 0.957, 0.881, and 0.918 for mIOU,

ACC, Dice, P, *R*, and *F1*-score, respectively, after the same test set test, and it received the highest scores among all eight methods. Moreover, its weight file occupied 241,553 kb of space, which was in the middle level among the eight improved methods. This indicates that the Imp8 method used in this paper has a better overall performance compared to other improvement methods.

A comparison of the training performance of the unimproved DeepLab v3+ and the improved network using

TABLE 4 | Comparison of the test results of different improvement methods of DeepLab v3+.

Type	Evaluation indicators						
	mIOU	ACC	Dice	*P*	*R*	*F1*-score	Pt (kb)
Imp1	0.812	0.982	0.896	0.945	0.852	0.896	572,794
Imp2	0.818	0.982	0.900	0.947	0.857	0.900	554,101
Imp3	0.839	0.985	0.912	0.954	0.874	0.912	232,841
Imp4	0.836	0.987	0.911	0.953	0.872	0.911	241,541
Imp5	0.837	0.986	0.911	0.954	0.873	0.911	310,161
Imp6	0.833	0.986	0.909	0.952	0.869	0.909	241,533
Imp7	0.841	0.986	0.914	0.955	0.876	0.914	310,173
Imp8	0.848	0.987	0.918	0.957	0.881	0.918	241,553

TABLE 5 | Detection statistics results of the two methods for the grape leaves in **Figure 11**.

| Leaf	Number of real disease spots			Pixels of real disease spots		
	Actual number	Detected by the detection method	Detected by the segmentation method	Actual pixels	Segmented by the detection method	Segmented by the segmentation method
Left	18	16	18	2,301	/	2,237
Middle	17	10	16	2,328	/	2,228
Right	14	12	12	2,132	/	2,066

The disease spots were manually segmented using LabelMe and then the number of pixels was counted by a self-developed python program. All these operations were carried out under the guidance and supervision of the grape disease specialist.

the Imp8 method is shown in **Figure 10**. The training set loss curves are shown in **Figure 10A**, where the red curve is before improvement and the blue curve is after improvement. When training until the model converged, the value of the red curve was about 0.17 and the value of the blue curve is about 0.132, which indicates that the improved model fit better on the training set than before improvement. **Figure 10B** shows the validation set loss curves, where the red curve is before improvement and the blue curve is after improvement. When training until the model converged, the value of the red curve was about 0.16, while the value of the blue curve was about 0.13, which indicates that the generalization ability of the model after the improvement was better than that before the improvement. Therefore, the improved DeepLab v3+ always converged faster and had better model fitting ability than the pre-improvement one whether on the training set or the validation set.

DISCUSSION

Effect Comparison Between Detection and Segmentation for Disease Spots

The grape leaf black rot disease spots can be recognized in the previous research of our group, and the spots were accurately segmented from the background in this paper. The effect of disease spots detection and segmentation for test set TS1 is compared in **Figure 11**. **Figure 11A** shows
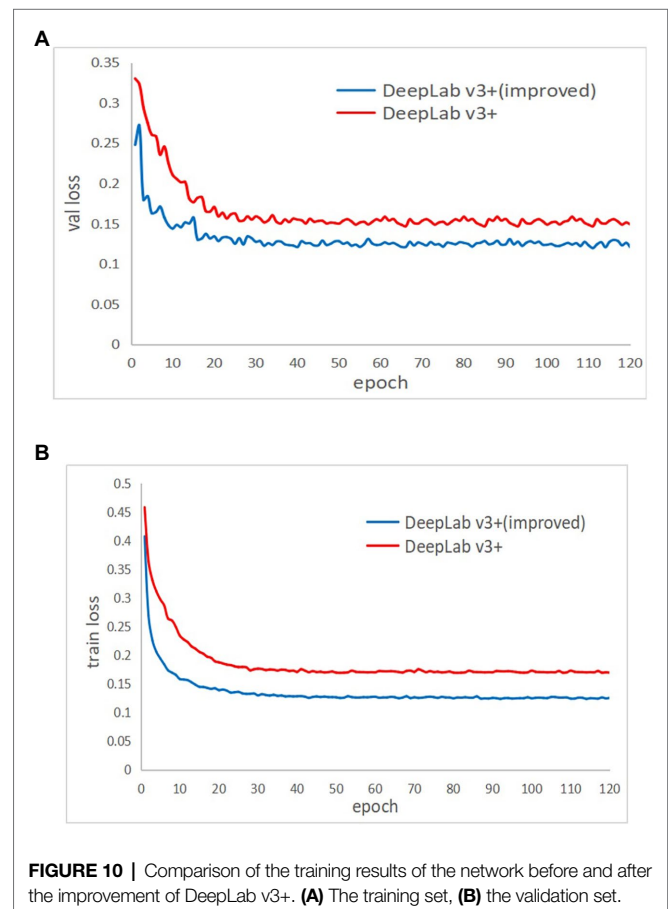
FIGURE 10 | Comparison of the training results of the network before and after the improvement of DeepLab v3+. **(A)** The training set, **(B)** the validation set.
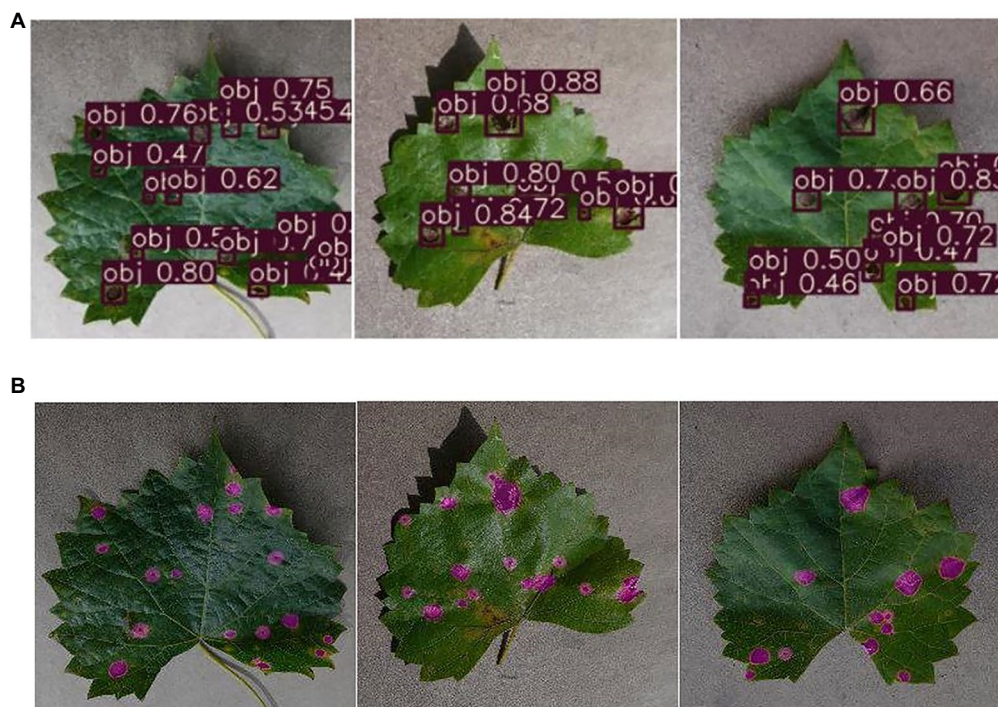
FIGURE 11 | The effect comparison between detection and segmentation on diseased spot. **(A)** The results of disease spots detection, **(B)** the results of disease spots segmentation.

the result of detection using the previous recognition method (Zhu et al., 2021), the number and location of the disease spots can be recognized, but cannot be segmented from the background. **Figure 11B** shows the result of segmentation using the method in this paper. The disease spots are not only recognized but also segment from the background according to their contour shape. **Table 5** shows the detection statistics results of the two methods for the grape leaves in **Figure 11**. As shown in **Table 5**, the segmentation method not only recognizes the number of disease spots but also obtains the number of pixels of spots. In addition, the segmentation method also detects and segments some tiny spots, which shows that this method is also better than the previous methods in recognition performance.

Comparison of Different Segmentation Algorithms

In this paper, DeepLab v3+ was chosen as the base algorithm to be improved for the segmentation of grape leaf black rot spots. This choice was based on the comparison of three common current mainstream deep learning segmentation algorithms. Pyramid Scene Parsing Network (PSP Net; Zhao et al., 2017) and U-Net are the other two common deep learning segmentation methods besides DeepLab v3+. PSPNet consists of a ResNet backbone that imposes a dilated convolution and a pyramid pooling module, which can mine

global contextual information for fast network training. U-Net is an FCN with a simple structure, which can obtain very accurate segmentation results using few training images and is widely used in medical image analysis.

In this study, these three semantic segmentation networks were trained using the same dataset, and segmentation experiments of black rot spots were conducted on the test set TS1. **Figure 12** shows the segmentation results of three different networks. As shown, PSPNet could segment the black rot spots, but the network performed poorly for the segmentation of connected spots, and it mistakenly segmented the leaf part between two spots. The segmentation effect of U-net was better than PSPNet, which could separate the lesion area independently, but the segmentation was not fine enough. Improved DeepLab v3+ was better than the other two methods.

Table 6 shows the experimental statistical results of the different segmentation methods. In terms of ACC, there was no significant difference between the three methods, but in the mIOU metric, improved DeepLab v3+ was 10.6 and 4.4% higher than PSPNet and U-net, respectively. In terms of the R value, improved DeepLab v3+ was 8.2 and 3.4% higher than PSPNet and U-net, respectively. The experimental results showed that the improved DeepLab v3+ had better segmentation performance compared with PSPNet and U-net, and the improved DeepLab v3+ could further improve the segmentation performance of black rot spots on grape leaves.
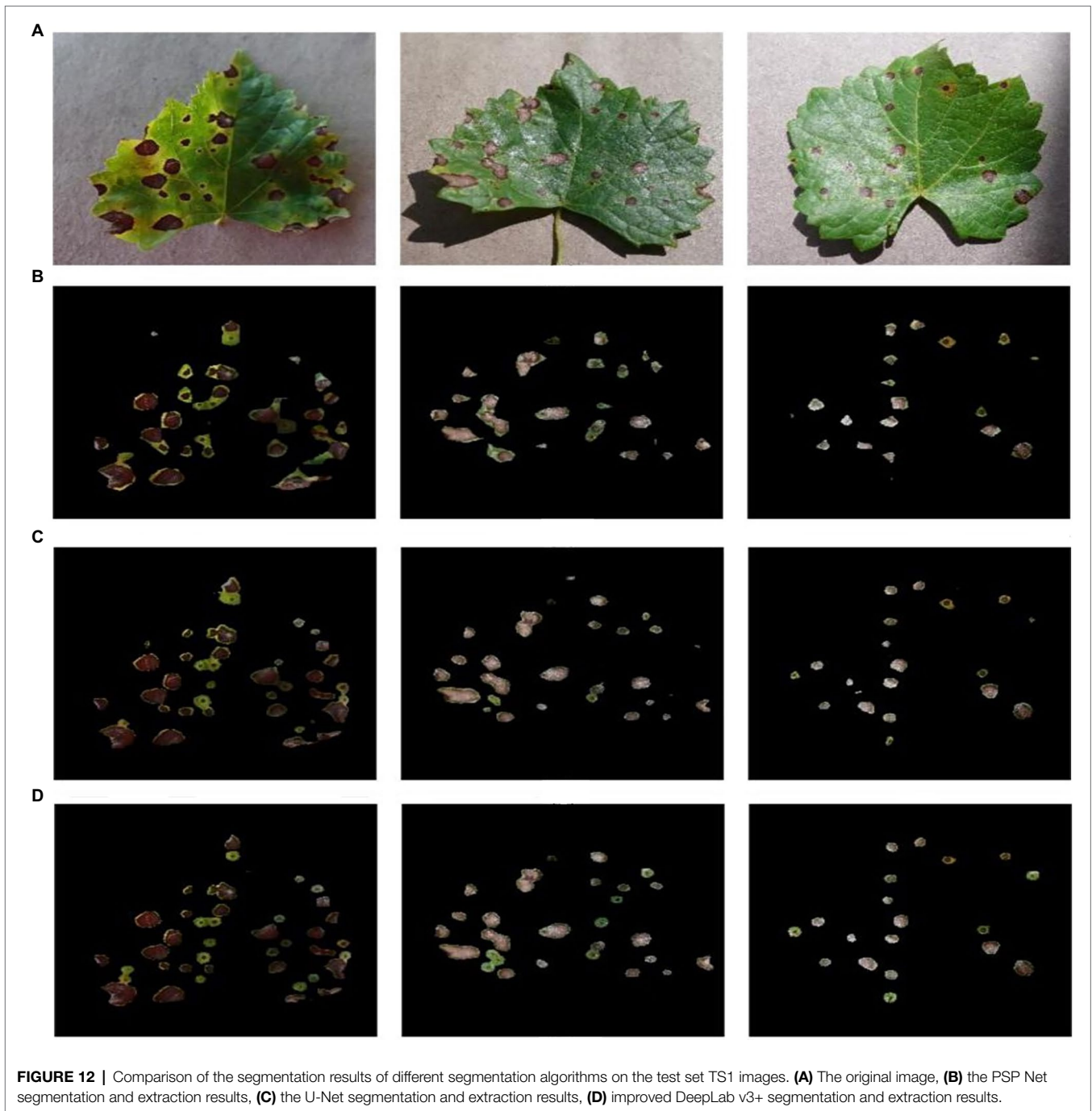
FIGURE 12 | Comparison of the segmentation results of different segmentation algorithms on the test set TS1 images. **(A)** The original image, **(B)** the PSP Net segmentation and extraction results, **(C)** the U-Net segmentation and extraction results, **(D)** improved DeepLab v3+ segmentation and extraction results.

TABLE 6 | Statistical segmentation results of different segmentation algorithms on the test set TS1 images.

Algorithm	Evaluation indicators					
	mIOU	ACC	Dice	*P*	*R*	*F1*-score
PSP Net	0.767	0.972	0.868	0.929	0.814	0.868
U-Net	0.812	0.98	0.896	0.945	0.852	0.896
DeepLab v3+ (improved)	0.848	0.987	0.918	0.957	0.881	0.918

CONCLUSION

This paper proposes an improved DeepLab v3+ network model for the segmentation of black rot spots on grape leaves. This method inserts the ECA module into the residual module of the original DeepLab v3+ backbone network. Moreover, a feature fusion branch based on a FPN is added at the encoder end. One 4-fold up-sampling to two 2-fold up-sampling are modified in the original network. To verify the performance of the improved network model, two test sets based on Plant Village and an orchard field environment were constructed for experiments. The experimental results showed that the improved DeepLab v3+ network model exhibited better performance on both test sets than before improvement, and the improved model could be applied to the segmentation of black rot spots on grapes in real production environments. This approach can not only provide an effective tool for classifying grape disease extent classes but also be applied to the evaluation of other plant leaf and fruit diseases. In future work, we will attempt to combine super-resolution image enhancement with this approach to further improve the effect of small target recognition and segmentation.

REFERENCES

Appeltans, S., Pieters, J. G., and Mouazen, A. M. (2021). Detection of leek white tip disease under field conditions using hyperspectral proximal sensing and supervised machine learning. *Comput. Electron. Agric.* 190:106453. doi: 10.1016/j.compag.2021.106453

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495. doi: 10.1109/TPAMI.2016.2644615

Bai, X., Li, X., Fu, Z., Lv, X., and Zhang, L. (2017). A fuzzy clustering segmentation method based on neighborhood grayscale information for defining cucumber leaf spot disease images. *Comput. Electron. Agric.* 136, 157–165. doi: 10.1016/j.compag.2017.03.004

Chaudhary, P., Chaudhari, A. K., Cheeran, A. N., and Godara, S. (2012). Color transform based approach for disease spot detection on plant leaf. *Int. J. Comp. Sci. Telecom.* 3, 4–9.

Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. (2017). DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Match. Intell.* 40, 834–848. doi: 10.1109/TPAMI.2017.2699184

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Pertanika J. Trop. Agric. Sci.* 11211, 137–143. doi: 10.1007/978-3-030-01234-2_49

Gao, C., Cai, Q., and Ming, S. (2020). YOLOv4 object detection algorithm with efficient Channel attention mechanism. in *2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, December 25, 2020, 1764–1770.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. December 2016, 770–778.

Hu, J., Shen, L., Albanie, S., Sun, G., and Wu, E. (2020). Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 2011–2023. doi: 10.1109/TPAMI.2019.2913372

Hu, G., Wei, K., Zhang, Y., Bao, W., and Liang, D. (2021). Estimation of tea leaf blight severity in natural scene images. *Precis. Agric.* 22, 1239–1262. doi: 10.1007/s11119-020-09782-8

Hughes, D. P., and Salathe, M. (2016). An open access repository of images on plant heath to enable the development of mobile disease diagnostics. Available at: http://arxiv.org/abs/1511.08060v2

Jiang, F., Lu, Y., Chen, Y., Cai, D., and Li, G. (2020). Image recognition of four rice leaf diseases based on deep learning and support vector machine. *Comput. Electron. Agric.* 179:105824. doi: 10.1016/j.compag.2020.105824

Jothiaruna, N., Sundar, K. J. A., and Karthikeyan, B. (2019). A segmentation method for disease spot images incorporating chrominance in comprehensive color feature and region growing. *Comput. Electron. Agric.* 165:104934. doi: 10.1016/j.compag.2019.104934

Kim, M. S., Lefcourt, A. M., Chen, Y. R., and Tao, Y. (2005). Automated detection of fecal contamination of apples based on multispectral fluorescence image fusion. *J. Food Eng.* 71, 85–91. doi: 10.1016/j.jfoodeng.2004.10.022

Kole, D. K., Ghosh, A., and Mitra, S. (2014). "Detection of downy mildew disease present in the grape leaves based on fuzzy set theory," in *Advanced Computing, Networking and Informatics. Vol. 1.* eds. M. K. Kundu, D. P. Mohapatra, A. Konar and A. Chakraborty (Cham: Springer International Publishing), 377–384.

Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Liang, Q., Xiang, S., Hu, Y., Coppola, G., Zhang, D., and Sun, W. (2019). PD 2 SE-net: computer-assisted plant disease diagnosis and severity estimation network. *Comput. Electron. Agric.* 157, 518–529. doi: 10.1016/j.compag.2019.01.034

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern recognition*, *CVPR*, January 2017, 936–944.

Lin, K., Gong, L., Huang, Y., Liu, C., and Pan, J. (2019). Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network. *Front. Plant Sci.* 10:155. doi: 10.3389/fpls.2019.00155

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, 3431–3440.

Ma, J., Du, K., Zhang, L., Zheng, F., Chu, J., and Sun, Z. (2017). A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing. *Comput. Electron. Agric.* 142, 110–117. doi: 10.1016/j.compag.2017.08.023

Mehl, P. M., Chao, K., Kim, M., and Chen, Y. R. (2002). Detection of defects on selected apple cultivars using hyperspectral and multispectral image analysis. *J. Agric. Saf. Health* 18, 219–226. doi: 10.13031/2013.7790

Milletari, F., Navab, N., and Ahmadi, S. A. (2016). V-net: fully convolutional neural networks for volumetric medical image segmentation. in *2016 4th International Conference on 3D Vision (3DV)*, IEEE, 2016, October 25, 2016, 565–571.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

HY, JZ, and MC conceived the idea and proposed the method. JZ and QW contributed to the preparation of equipment and acquisition of data. JZ wrote the code and tested the method. JZ, QW, and MC validated results. HY and JZ wrote the paper. HY, JZ, and ZC revised the paper. All authors have read and approved the final manuscript.

FUNDING

Molitor, D., and Berkelmann-Loehnertz, B. (2011). Simulating the susceptibility of clusters to grape black rot infections depending on their phenological development. *Crop Prot.* 30, 1649–1654. doi: 10.1016/j.cropro.2011.07.020

Pan, T. T., Chyngyz, E., Sun, D. W., Paliwal, J., and Pu, H. (2019). Pathogenetic process monitoring and early detection of pear black spot disease caused by *Alternaria alternata* using hyperspectral imaging. *Postharvest Biol. Technol.* 154, 96–104. doi: 10.1016/j.postharvbio.2019.04.005

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: convolutional networks for biomedical image segmentation. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, October 5, 2015, 12–20.

Rundo, L., Han, C., Nagano, Y., Zhang, J., Hataya, R., Militello, C., et al. (2019). USE-net: incorporating squeeze-and-excitation blocks into U-net for prostate zonal segmentation of multi-institutional MRI datasets. *Neurocomputing* 365, 31–43. doi: 10.1016/j.neucom.2019.07.006

Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. (2008). LabelMe: A database and web-based tool for image annotation. *Int. J. Comput. Vis.* 77, 157–173. doi: 10.1007/s11263-007-0090-8

Samajpati, B. J., and Degadwala, S. D. (2016). Hybrid approach for apple fruit diseases detection and classification using random forest classifier. in *2016 International Conference on Communication and Signal Processing (ICCSP)*, April 6, 2016, 1015–1019.

Singh, V. (2019). Sunflower leaf diseases detection using image segmentation based on particle swarm optimization. *Artif. Intell. Agric.* 3, 62–68. doi: 10.1016/j.aiia.2019.09.002

Sinha, A., and Shekhawat, R. S. (2020). Olive spot disease detection and classification using analysis of leaf image textures. *Procedia Comput. Sci.* 167, 2328–2336. doi: 10.1016/j.procs.2020.03.285

Wang, C., Du, P., Wu, H., Li, J., Zhao, C., and Zhu, H. (2021). A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-net. *Comput. Electron. Agric.* 189:106373. doi: 10.1016/j.compag.2021.106373

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. (2020). ECA-net: efficient channel attention for deep convolutional neural networks. in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 13, 2020, 11531–11539.

Wspanialy, P., and Moussa, M. (2020). A detection and severity estimation system for generic diseases of tomato greenhouse plants. *Comput. Electron. Agric.* 178:105701. doi: 10.1016/j.compag.2020.105701

Yang, M., Yu, K., Zhang, C., Li, Z., and Yang, K. (2018). DenseASPP for semantic segmentation in street scenes. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2018, 3684–3692.

Yu, F., Koltun, V., and Funkhouser, T. (2017). "Dilated residual networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, 636–644.

Zagoruyko, S., and Komodakis, N. (2016). Wide residual networks. in *British Machine Vision Conference (BMVC)*, September 2016, 87.1–87.12.

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). Pyramid scene parsing network. in *Proceedings of 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, January 2017, 6230–6239.

Zhou, R., Kaneko, S., Tanaka, F., Kayamori, M., and Shimizu, M. (2014). Disease detection of Cercospora leaf spot in sugar beet by robust template matching. *Comput. Electron. Agric.* 108, 58–70. doi: 10.1016/j.compag.2014.07.004

Zhu, J., Cheng, M., Wang, Q., Yuan, H., and Cai, Z. (2021). Grape leaf black rot detection based on super-resolution image enhancement and deep learning. *Front. Plant Sci.* 12, 1–16. doi: 10.3389/fpls.2021.695749