IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 13, 2021, accepted January 15, 2021, date of publication January 19, 2021, date of current version January 27, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3052769

SPEDCCNN: Spatial Pyramid-oriented Encoder-Decoder Cascade Convolution Neural Network for Crop Disease Leaf Segmentation

YUXIA YUAN[1], ZENGYONG XU[2], AND GANG LU[3]
[1]School of Electrical Engineering, Zhengzhou University of Science and Technology, Zhengzhou 450000, China
[2]School of Automotive Studies, Henan College of Transportation, Zhengzhou 450000, China
[3]School of Electrical Engineering and Automation, Luoyang Institute of Science and Technology, Luoyang 471023, China

Corresponding author: Yuxia Yuan (publicgj163.com)

This work was supported by the 2021 Henan University of Higher Education Key Research Project, Project Name: Research and Design of Low Frequency Vibration Energy Capture Device Based on Piezoelectric Technology under Grant 21B510012.

ABSTRACT Disease is one of the main factors affecting crop growth. How to reflect the external morphological features of the disease and completely retain the color and texture information of the disease area is one of the key research issues for crop disease segmentation. Meanwhile, aiming at the problem of low segmentation accuracy with traditional convolution neural network-based methods in the crop disease leaf image, this paper proposes a spatial pyramid-oriented encoder-decoder cascade convolution neural network for crop disease leaf segmentation. The network consists of a region disease detection network and a region disease segmentation network. Region disease detection network is a kind of network combining cascade convolution neural network with spatial pyramid. This method connects the three-level convolution neural network model, where the structure of the three-level neural network model varies from simple to complex. Different crop disease leaf features are extracted from the different neural network levels. And images are screened to complete the detection of crop disease leaf. What's more, a space pyramid pooling layer is added to each network level. This pooling strategy does not require fixed size input, which increases the size selection of input model. The region segmentation network is established based on the Encoder-Decoder structure. The multi-scale convolution kernel is used to improve the local receptive field of the original convolution kernel and accurately segment the crop disease leaf area. Finally, we conduct experiments on the crop disease leaf images under different conditions, the results show that the proposed method has higher segmentation accuracy. In terms of Precision, Correct segmentation, over-segmentation and under-segmentation indexes, etc., the average values of proposed method are more than 90%. The average dice similarity coefficient is over 95% under different background. Moreover, it can meticulously reflect the external morphological features of the crop disease leaf and relatively better retain the color and texture information.

INDEX TERMS Crop disease leaf segmentation, spatial pyramid, encoder-decoder cascade convolution neural network, pooling strategy.

I. INTRODUCTION

Crop disease is one of the important factors affecting crop yield and quality [1]. In order to improve the quality and yield of the main crops in production, it is necessary to make effective diagnosis and treatment for the crop diseases encountered in the process of crop growth. Traditional crop disease diagnosis requires experts to go into the field, which is often time-consuming and laborious. At the same time, due to the influence of various external environmental conditions and subjective factors, artificial diagnosis and treatment of many diseases can easily lead to subjective misjudgment. Since the symptoms of most diseases are mainly on crop leaves, the occurrence and types of diseases can be determined by observing the characteristics of crop leaves. The traditional way of judging diseases by artificial observation has some problems such as subjectivity, blindness and low efficiency [2], [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

In recent years, with the continuous development of computer vision technology, it has become a trend to use automatic detection methods to judge crop diseases, which can effectively solve the shortcomings of artificial disease judgment.

In the process of crop disease leaf image detection, the accurately segmentation of disease area directly affects the accuracy of disease recognition [4], [5].

There are many segmentation methods for crop disease leaf image, which can be divided into traditional image segmentation method and deep learning-based segmentation method. Traditional crop disease leaf image segmentation methods contain four categories: threshold-based, clustering-based, classification-based and graph theory-based segmentation [6]–[9]. In threshold-based method, the key threshold point is found to segment the crop disease leaf image. This method is suitable for the crop disease leaf image segmentation under simple background. The clustering-based method classifies the disease region according to the different pixels of the image, but it requires many iterations to determine the optimal clustering point during the segmentation process, which reduces the timeliness of the segmentation method. The classification-based method uses a variety of image features to train the classifier, and adopts the trained classifier to classify the pixel points to achieve segmentation. It needs to complex feature extraction algorithm in the feature extraction stage, so this method is less practical. The graph theory-based method maps the original image to an undirected weight graph and performs image segmentation by solving the optimal solution of the function. This method is suitable for the condition with simple background or low pixel similarity. The above traditional image segmentation method is only applicable to the crop disease leaf image with a simple background. When the color of crop disease area is similar to the background or the boundary is fuzzy, the traditional image segmentation methods are difficult to distinguish the object and background, so the segmentation effect is poor.

Recently, due to the development of deep learning technology, it has become a trend to use convolution neural network (CNN) to perform image semantic segmentation. Many scholars have tried to use CNN to solve the problem of crop disease leaf image segmentation [10]–[13].

Liu [14] combined Markov and CNN to complete the segmentation of cotton disease image. This method used CNN to extract image deep semantic features and combined with the relative relationship between the feature map pixels to construct the conditional random field energy function and conduct optimization training. Liu [15] trained the rice disease leaf image segmentation model based on the full convolution neural network (FCN) using the data set of rice disease to complete the rice disease leaf image segmentation under a complex background. Zhao [16] combined FCN and Conditional Random Field (CRF) to segment grape disease leaves. In this method, CRF was used as a probability graph to describe the detailed information of the image in the segmentation process, so the training time in this network model is long. Xiong [17] combined super pixel segmentation with CNN to achieve field rice spike segmentation in different growing periods. However, the segmentation effect were poor under complex background. Li [18] designed a greenhouse cucumber disease recognition system based on CNN. In the pretreatment stage, the system combined the compound color features with CNN to realize the complete segmentation of different diseases. Ahmad [19] proposed a strawberry leaf disease detection method based on AlexNet, which was effective in the strawberry leaf disease detection, but the detection effect was poor in the case of occlusion. Pound [20] used CNN to segment different parts of wheat, such as root tip, spike tip and spike base, and achieved good results. Since CNN needed to preprocess the original image, the universality of this method was poor. The above deep learning-based image segmentation methods have achieved high segmentation accuracy in the different crop disease leaf images. Wen [21] proposed an end-to-end detection-segmentation system to implement detailed face labeling. Yin [22] introduced a deep guidance network to segment the biomedical image, which consisted of a guided image filter module to restore the structure information through the guidance image. Opbroek [23] investigated kernel learning as a way to reduce differences between training and test data and explore the added value of kernel learning for image weighting. Also a new image weighting method was proposed that minimized maximum mean discrepancy between training and test data, which enabled the joint optimization of image weights and kernel. Wang [24] developed a new class of hierarchical stochastic image models called spatial random trees (SRTs) which admitted polynomial-complexity exact inference algorithms.

However, the existing segmentation methods based on deep learning can only complete the segmentation task for a single plant disease, and cannot solve the segmentation problem with many plant diseases, which reduces the practicability of the segmentation method. Moreover, some network models are complex, which is difficult to extend them to other crop disease recognition tasks. The established deep learning network model needs many parameters to train the network, which is difficult to satisfy the real-time demand for crop disease leaf segmentation.

In the image processing with multi-convolution neural networks, a single convolution network model is often designed. In this way, the features collected by the neural network are relatively single, which makes the generalization performance of the network model be poor. It usually has good effects for some specific problems. As a result, a cascade convolution neural network was proposed in CVPR2015. By connecting convolution neural network models with different structures, the model can gradually extract features. The model structure varies from simple to complex. It is relatively easy to start the convolution neural network, i.e., rough feature extraction process, which is equivalent to roughly carry out the first-level classification of

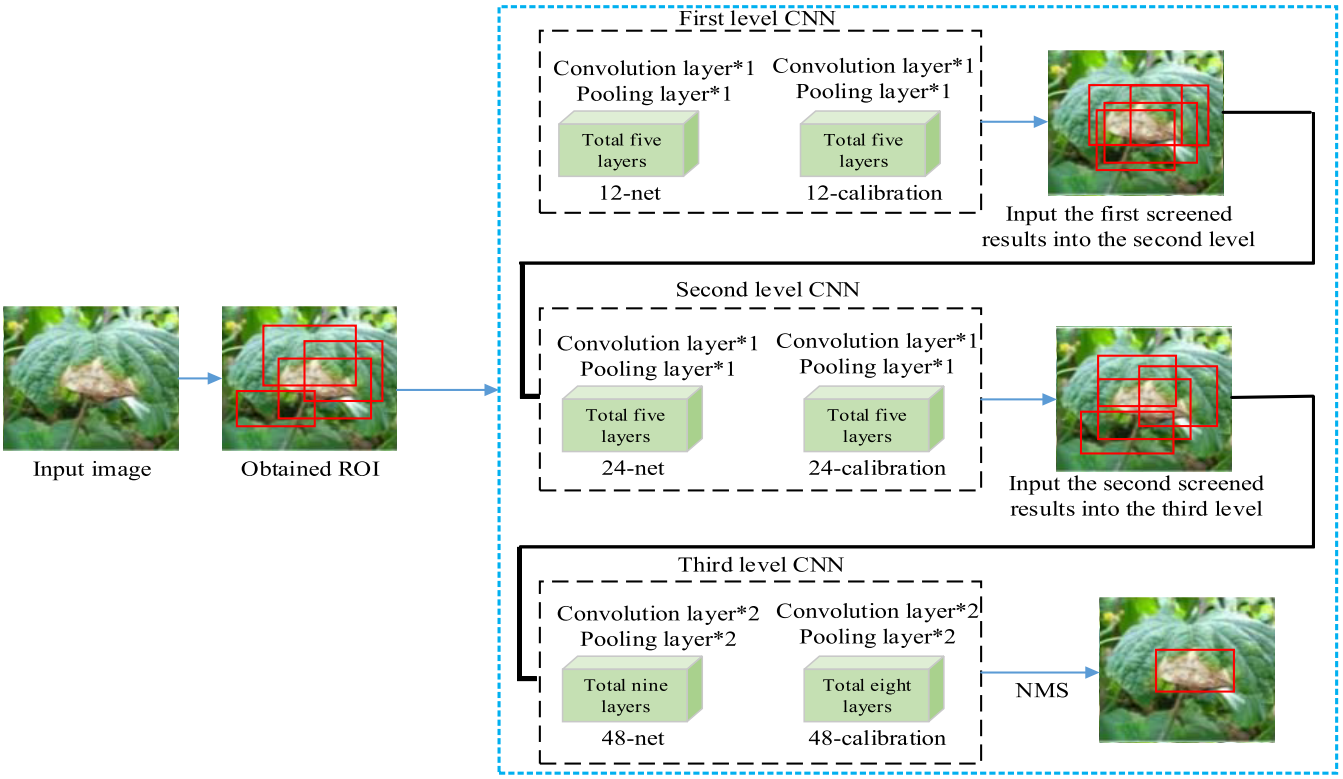


FIGURE 1. Structure of cascade CNN.

the input images. The classification results are input into the second-level convolution neural network model, which is more complex than the previous level. The extraction process of image features will also be more detailed. After this image classification, the required results are input into the most complex network (third-level) to carry out the final feature extraction. Finally, the classification results are obtained. The model uses different models to extract features, which reduces the detection time and improves the accuracy.

Therefore, aiming at the problem of crop disease leaf segmentation, this paper proposes a spatial pyramid-oriented encoder-decoder cascade convolution neural network (abbreviated to SPEDCCNN) and applies it to the crop disease leaf image segmentation under different environments. Therefore, our main contributions are as follows:

- ➢ This paper proposes a spatial pyramid-oriented encoder-decoder cascade convolution neural network for crop disease leaf segmentation.
- ➢ The network consists of a region disease detection network and a region disease segmentation network. Region disease detection network is a kind of network combining cascade convolution neural network with spatial pyramid. This method connects the three-level convolution neural network model
- ➢ A space pyramid pooling layer is added to each network level. This pooling strategy does not require fixed size input, which increases the size selection of input model.
- ➢ The region segmentation network is established based on the Encoder-Decoder structure.
- ➢ The multi-scale convolution kernel is used to improve the local receptive field of the original convolution kernel and accurately segment the crop disease leaf area.
- ➢ Finally, experiments on the crop disease leaf images under different conditions show that the proposed method has higher segmentation accuracy.

II. RELATED WORKS

In recent years, the research in the object detection field mainly focuses on the uncontrollable part, which will affect the final effect of object detection. In addition, it is difficult to generate better generalization ability only by using the model with a single structure, which makes the robustness of the model in practical application low. Therefore, the three main difficulties in object detection are as follows:

a. The object has too many variable cases in the complex scene;

b. There are too many possible pseudo-objects in the image;

c. The single structure has weak robustness with variable conditions.

In view of the above problems, the cascade convolution neural network model proposed by Liu [25] effectively solves the main difficulties. The structure of the cascade convolution neural network is shown in figure 1.

It can be seen from figure 1 that the network model is mainly composed of a three-level convolution neural

network. The candidate box (ROI) is obtained through the simple fixed size window. Each level includes a dichotomous network (12-net, 24-net, 48-net), and a calibration network (12-calibration, 24-calibration, 48-calibration). The three network levels have the different resolutions of the input image. The gradually increasing of resolution is mainly to improve the recognition accuracy, which can reduce the running time and improve the efficiency of the model. The structure of this network varies from simple to complex. The simple network is used for rough feature extraction, while the complex network is devoted to accurate classification.

For most of the current CNNs, the size of the input image is required to be fixed, which requires the CNN to resize the image before training or testing. However, in data preprocessing, the original image is cropped to a uniform size. When the scale of the input image changes, the traditional CNN will not be able to realize the multi-scale input. Meanwhile, more data information will be lost in this process compared with the multi-scale pre-processing process, which will have a certain impact on the subsequent training and testing.

To solve this problem, He [26] proposed a spatial pyramiding (SP) pooling method to solve the scale variation problem of input data. Since the convolution layer and the pooling layer do not require fixed-size inputs, the full connection layer requires fixed-size input. Assuming that the size of the input image is 100×100, and the $5 \times 98 \times 98$ feature graph will be generated after five 3×3 convolution kernels. When the size of the input image becomes 102×102, the same operation will result in a $5 \times 100 \times 100$ feature map. Then the input of the two sizes will get 25×25 and 26×26 feature graphs after 2×2 pooling respectively. Therefore, it can be seen that the convolution layer and the pooling layer can process images with any input size. However, the full connection layer has a requirement for the input size. Suppose that the last convolution layer has 50 outputs and the next full connection layer has 1000 neurons, the dimension of this connection matrix is 50×1000. If the input image size is different at each time, then the matrix dimension connected to the full connection layer will be changed, making it impossible to train or test the network. Therefore, the space pyramid pooling algorithm is to add a space pyramid pooling layer before the full connection layer, which ensures that the input images with any size is processed as the same dimension. The convolution neural network with pyramidal pooling is shown in figure 2.

It can be clearly seen that the input image does not need to be preprocessed after introducing pyramid pooling, which realizes the multi-scale input of convolution neural network.

Figure 3 shows a traditional network architecture model. The fully connection layer is behind the convolution layer. Adding a pyramid pooling layer before the fully connection layer solves the problem of different input image sizes. It can be seen that the pyramidal pooling layer is to carry out three pooling operations on the feature map obtained from the previous convolution layer. The uppermost pooling operation is pooling the original feature map. In the middle, the feature map is divided into four parts for pooling. At the bottom, the feature map is divided into 16 parts for pooling. This can form a 21-eigenvector input to the fully connection layer, which solves the problem of inconsistent input image size.

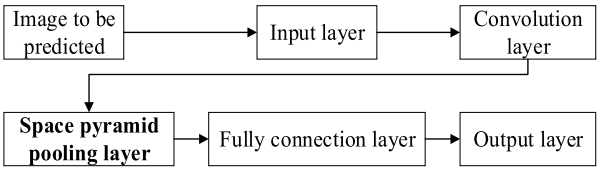


FIGURE 2. CNN with SP.

Algorithm 1 SP Pooling

Input: eigenmatrix after convolution pooling $X \in R^{a \times b}$.
Output: eigenvector after space pyramid pooling F.
Step 1. calculating $w = cell(a/n)$, $h = cell(b/n)$, $stride1 = floor(a/n)$,$stride2 = floor(b/n)$,$n = 1,2,3\ldots$;
Step 2. The obtained parameters are used to pool the eigenmatrix X, then it obtains the features *f1*, *f2*, and *f3*;
Step 3. Connecting the obtained features and getting the new feature F.

In the case of different size of image input, assuming that the obtained size of the feature map after convolution pooling is a × a. The number of pyramid pooling is n, so the side length of each window is $win = cell(a/n)$. The pooling strides are $strides = floor(a/n)$. Finally, after pyramid pooling, n pooling operations will be formed as shown in figure 5. These pooling operations will adopt the basic pooling method (such as maximum pooling), but with different window sizes and moving steps.

This paper is organized as follows. Section 2 introduces the related works for this paper. In Section 3, we present the proposed image segmentation method. In Section 4, we describe experimental results. A conclusion with open problems ends the paper in Section 5.

III. PROPOSED CROP DISEASE LEAF SEGMENTATION METHOD

The structure of the proposed segmentation model is shown in figure 4. Because the crop disease leaf area has the higher pixel similarity with the normal area of the leaf. And the acquired image contains a lot of background noise. Therefore, the cascade convolution neural network is combined with the spatial pyramid to establish the region disease localization network (named RDL-NET), which is used to detect the leaf disease area to reduce the influence of background information on the segmentation effect. The region disease segmentation network based on Encoder-Decoder model architecture (named RSED-NET) is established for crop leaf disease segmentation.

In SPEDCCNN, the scale-independent cascade convolution neural network is used to obtain the features in different crop disease leaf image scales. The output result of RDL-NET is input into RSED-NET. Then the localized disease leaf

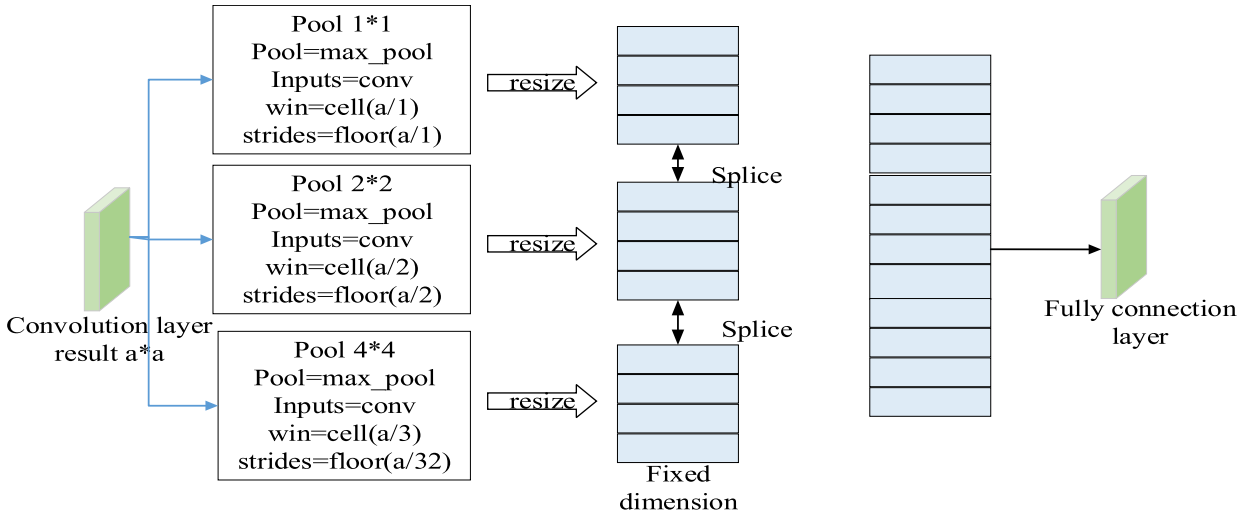


FIGURE 3. Process of SP pooling algorithm.

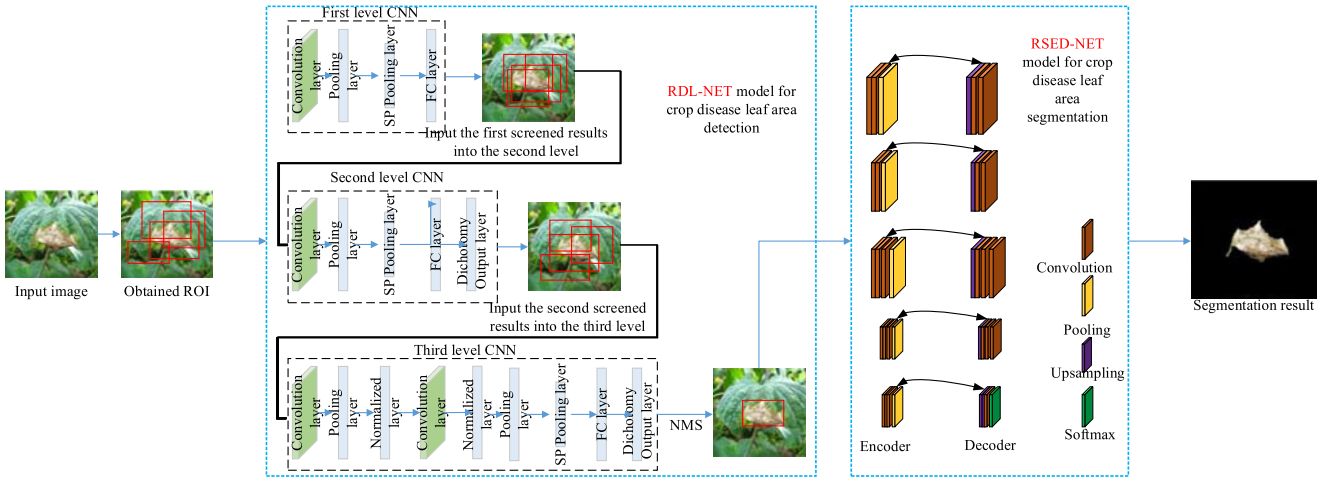


FIGURE 4. Structure of SPEDCCN model in this paper.

area is segmented. Most of the existing deep learning-based segmentation methods adopt the end-to-end segmentation structure, which is directly used for the image segmentation task, but it is unable to localize the specific location to be segmented in the image, this approach reduces the segmentation efficiency. In this study, the proposed method combines RDL-NET and RSED-NET to improve the segmentation accuracy. After the localization of the crop disease leaf area in the image by RDL-NET, the visual saliency of the SPEDCCNN network is enhanced, which promotes the segmentation speed of the new model.

The image to be detected is input in the network for feature classification, if it is judged as the crop disease leaf image, then the image will be input into the first correction network; otherwise, it will delete this image directly and judge the next image. In the second-level dichotomy network, the same operation is carried out as before. The obtained crop disease leaf is processed by the non-maximum suppression (NMS) and finally the crop disease leaf area is obtained in the original image. The model has better recognition accuracy when testing the standard data sets. In addition, due to the simple network structure in the first two layers, the detection rate of the overall model is also improved, and the detection time is significantly reduced compared with other traditional networks.

A. IMAGE ACQUISITION

The experimental images were collected at the Datian Experimental Base in Zhengzhou city, Henan Province from May to August, 2017. In order to fully consider the impact of natural light on the segmentation effect, shooting was conducted at 8:00, 10:00, 16:00 and 18:00 every day. The leaves of main crop diseases included corn great spot, corn round spot, wheat stripe rust, wheat anthracnose, cucumber target spot and cucumber brown spot [27]. A total of 900 crop disease leaf images were taken with 150 images of each disease

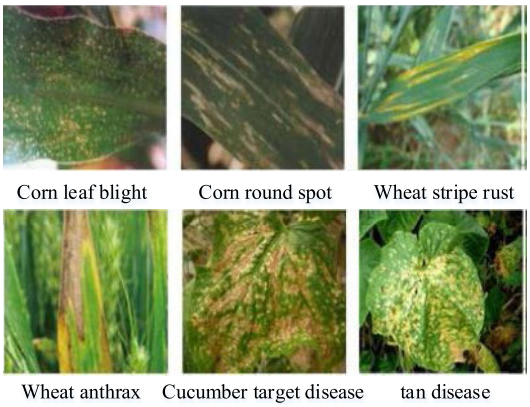


FIGURE 5. Crop disease leaf images.

leaf. It was collected with a Cannon EOS650D digital camera, without supplementary light source, and was shot about 15-30cm away from the crop leaf. In order to improve the training efficiency of the network model, the images were re-sized as 256 × 256 pixel. The instances of crop disease leaf images are shown in figure 5.

The crop disease leaves data sets contain 650 training sets, 125 verification sets and 125 testing set. The training set is input into the segmentation network model and the weight parameters of the model are trained. The validation set is used to determine the structure of the model. And the complexity of the model is controlled and the model parameters are optimized. The testing set is utilized to test the segmentation effect of the proposed model.

B. RDL-NET MODEL

Although the cascade convolution neural network has excellent performance in the field of crop disease leaf detection, it does not support the multi-scale input, which leads to the information loss of the images in the pre-processing stage of the model. The spatial pyramid pooling algorithm solves the multi-scale input problem of convolution neural network. Therefore, this paper proposes a scale-independent cascade convolution neural network leaf detection algorithm and combines with the advantages of space pyramid pooling algorithm in the RDL-NET model.

The structure of the model is shown in figure 6. Pyramid pooling layer is added before the full connection layer of the convolution neural network. The unified channel number of pyramid pooling is 5, so that each level of convolution neural network supports multi-scale input, and the structural complexity of the overall network is from simple to complex. In order to carry out image detection faster, the proposed algorithm in this paper does not set up the calibration network. In this way, the whole model only has three convolution neural networks, which will greatly speed up the training and detection speed.

RDL-NET network mainly consists of 4 convolution layers, 4 pooling layers, 3 SPP pooling layers, 3 FC layers and 3 output layers. All convolution kernels are set as 3 × 3, and the stride is set as 1. In order to keep the same dimension with the input image, the parameter pad is set to 1 with the way of boundary expansion to complete the edge of the convolution layer. The Pooling layer size is set as 3 × 3, the Pooling window is set as 2 × 2, and the sliding step size is set as 2. The sliding window of n × n is used to carry out the convolution operation on the feature map output by the convolution layer. The convolution layer obtained through the sliding window is mapped into a feature vector, and the feature vector is input to the classification layer and the regression layer. The classification layer is used to determine whether it is a normal area or a disease area. The regression layer is used to locate the the disease area.

In the process of detecting disease leaf areas, The center point of the sliding window is defined as the anchor point, and each slide of the window corresponds to three scales and three aspect ratios respectively. Therefore, each slide will generate n = 9 anchor points and also generate the same number of area detection boxes. The detected disease leaf image is compared with the original manual labeled image, and the error between the detected value and the real labeled value is calculated by using two different loss functions.

$$L_{cls}(p_i, p_i^*) = -\lg[p_i^* p_i + (1 - p_i^*)(1 - p_i)] \tag{1}$$

Position loss function is:

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*) \tag{2}$$

where, R(x) is the regularization loss function, i.e,

$$R(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & others \end{cases} \tag{3}$$

The overall loss function is:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \tag{4}$$

where λ is the weight coefficient. p_i represents the disease probability of the detection area. p_i^* is the real labeled class, where $p_i^* = 0$ represents the background and $p_i^* = 1$ represents the disease area. N_{cls} and N_{reg} are regular terms used to avoid over-fitting. i is the index value of the detected candidate region. t_i is the position parameter of the original coordinate box, and t_i^* is the actual coordinate of the candidate region.

C. RSED-NET MODEL

The structure of RSED-NET is similar to SegNet, which is mainly composed of encoder network, decoder network and Softmax classifier. The encoder network is constructed based on VGG16 network structure. However, because the detailed features of crop disease leaves are relatively complex, it is difficult to extract complex features by using traditional convolution kernel. In order to extract the features of complex leaves diseases in the image, it is necessary to

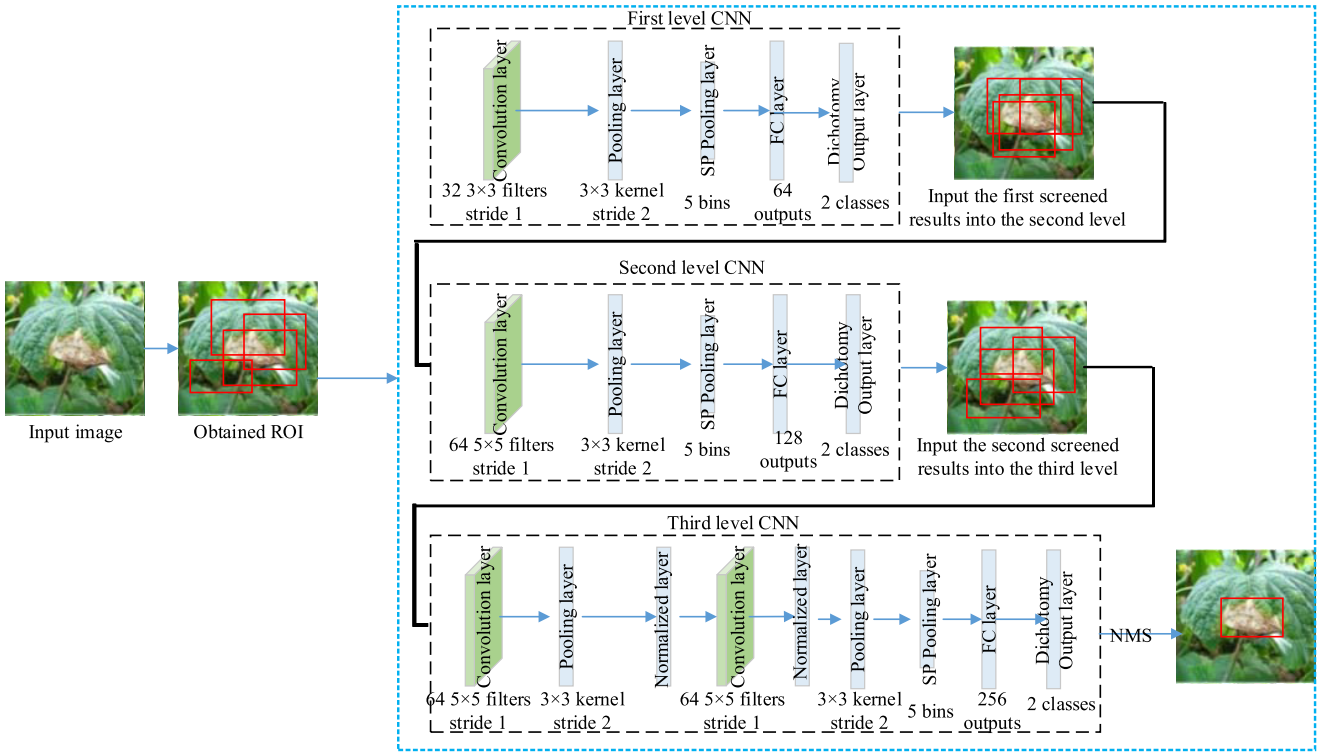


FIGURE 6. RDL-NET model.

increase the number of convolution layers or introduce multiple dimensional convolution kernels. However, only stacking the simple convolution layer will reduce the nonlinear change of the model, so that the image features extracted by each convolution layer are the same. The introduction of multiple dimensional convolution kernels will only greatly increase the amount of parameter calculation, lead to the increase of training time and reduce the efficiency of the model. Therefore, the multi-scale convolution kernel is used in the encoding network instead of the original convolution kernel. By introducing different scale convolution kernels into the same convolution layer, different scale features of the image are obtained, and then these features are fused to obtain the multi-scale feature map of the original image.

SegNet network can obtain accurate semantic segmentation results mainly due to the design of its decoding network. The SegNet network model has a higher segmentation precision than the fully convolution neural network (FCN) by recording the maximum pooling level index in the encoder network and using the index factor to directly up-sample in the decoder network. RSED-NET uses the same decoder network as SegNet, mainly including convolution layer (Conv1-Conv14) and up-sampling layer (up-sampling1-up-sampling5). The convolution layer is used to extract the deep feature of the multi-scale feature map obtained by the encoder network, and the up-sampling layer is used to restore the image resolution. The crop disease leaf segmentation is equivalent to a dichotomy problem. The purpose of leaf segmentation is to distinguish the normal part from the diseased part. Therefore, the channel number of Conv14 in the decoding network is set as 2, followed by a pixel level classifier Softmax. The number of classifiers is also set as 2.

On the basis of the above analysis, the leaf disease segmentation network (RSED-NET) is constructed as shown in figure 7, which mainly includes two sub-networks, encoder network and decoder network. Where, the convolution layer number in encoder network and decoder network is 13, and the convolution layer size corresponding to the pooling index is the same. The stride of each convolution layer is set as 1. In order to keep the dimensions of input and output images unchanged, the boundary extension parameter (Pad) is set as 1. The maximum pooling operation is adopted for pooling layer (pooling1-pooling5). The size of pooling layer is 2×2, the sliding step is set as 2. The sampling scale is set as 2.

D. MODEL TRAINING

During CNN training process, a large number of training samples are needed to update the network model parameters to improve the performance of the network. SPEDCCNN is trained by using the constructed training data set, which can obtain a better network with high segmentation accuracy and fast operation speed. However, when the number of training samples is small or the content of training sets is similar, the model is prone to over-fitting or poor segmentation effect. In order to improve the training efficiency and segmentation performance of the network model, SPEDCCNN adopts transfer learning to train the network. By using transfer learning to initialize model parameters, the extracted features of the model

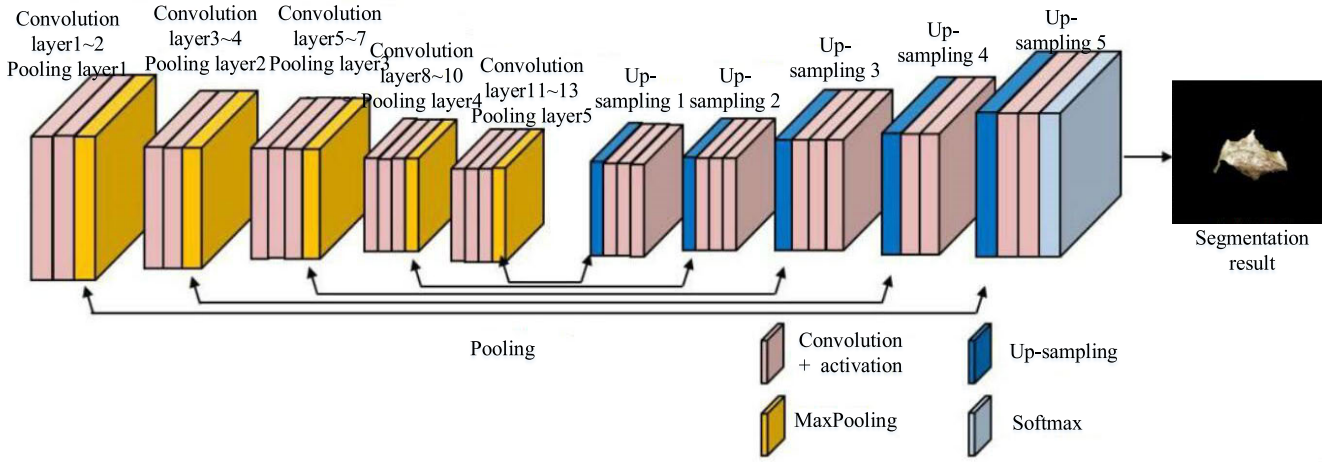


FIGURE 7. RSED-NET model.

have the level-layer characteristics. Shallow network is used to extract low-level features of the image, and deep network is used to extract high-level semantic features of the image. The low level features include color, edge and other primary information. The low-level features obtained by training the network with different data sets are similar. Therefore, when the data set size is small, the low level information obtained from the large-scale data set is transferred to the network model to be trained by transfer learning, which not only solves the over-fitting problem of the model, but also reduces the training time of the model.

The performance of the model is evaluated mainly by calculating the difference between the segmented image and the labeled image. The training is evaluated by calculating the cross entropy between the labeled image and the segmented image. The cross entropy function is a commonly used model performance evaluation index in neural networks. The smaller cross entropy denotes the better training effect. In the training process of SPEDCCNN, the cross entropy of each pixel point between the labeled image and the segmented image is calculated. The average pixel cross entropy loss function is used to evaluate the training effect of SPEDCCNN. The average cross entropy loss function is denoted as follows:

$$loss(p, q) = \sum_{N} (-\sum_{X} p(x) \lg q(x)) \tag{5}$$

where $p(x)$ is the pixel classification vector of the labeled image. $q(x)$ is the pixel classification vector of the segmented image. N is the total number of pixels contained in the image. X is the feature vector of the input image, and x is the feature vector of each pixel of the input image.

After the pixel loss value is obtained according to equation (1), the loss value is transmitted back to each convolution layer of the network model using the back propagation algorithm. The weight parameters of the convolution layer are updated. After many repeated iterations, the training is completed until the output loss value is stable.

E. ALGORITHM DESCRIPTION

Next, the technical procedures of the proposed method are given as below:

Step 1: training stage. The datasets are from public crop leaf data set and collected from real field crops. 80% and 20% of the samples are used for model training and testing, respectively.

Step 2: testing stage. In the testing stage, the image to be detected needs to be preprocessed. First, Pyramid image processing is used to scale images into groups containing different sizes. Then all the images are operated with the sliding window mechanism, where the sliding window size is 24 × 24 to generate all candidate boxes. The position information of all candidate boxes in the original image is labeled. And all candidate boxes are put into the trained neural network model. At this time, the model will delete the non-disease images in the candidate box and keep the disease images in the candidate box. Following networks perform this operation in turn until the last layer reaches the final candidate box. Finally, the candidate boxes with high overlapped areas are deleted by non-maximum suppression (NMS) to obtain the final disease leaf region, and then the detected disease leaf region is marked in the original image according to the position information of the candidate boxes.

Input: The images to be detected.

Output: The detected crop disease leaf image area.

Step 2.1: The training samples are used for model training.

Step 2.2: The image is preprocessed to get ROI.

Step 2.3: In RDL-NET model, all obtained candidate windows are put into the trained first-level convolution neural network. The screened candidate windows at the first level are put into the second level convolution neural network. The obtained candidate window from the second-level is put into the third-level convolution neural network, which is used

to detect the leaf disease area to reduce the influence of background information on the segmentation effect.

Step 2.4: The region disease segmentation network RSED-NET is established for crop leaf disease segmentation.

Step 2.5: The final candidate window is processed by the NMS, and the overlap greater than 0.5 is removed.

Step 2.6: According to the position information of the candidate window, the detected disease leaf region is marked in the original image.

IV. EXPERIMENTS AND ANALYSIS

The proposed model is verified on the crop disease leaves image data set and compared with the state-of-the-art models including Hseg [28], FCN8s [29], SegNet [30], JointSeg [31] and DeepLab [32] segmentation algorithms. The experimental software environment is Ubuntu 16.04 LTS, Matlab2017a. The hardware environment is Intel core i7-7550k, CPU@3.60GHz, RAM32GB and GTX1060Ti GPU. The deep learning development framework is Matconvent.

A. EXPERIMENT PARAMETER SETTING AND TRAINING PROCESS ANALYSIS

Batch training is used in the experiment. The training set, validation set and testing set are input into the network model with batches. The training set contains 650 images. Ten images are used as a batch. So 65 batches are required to complete the training of the model. In order to ensure the training efficiency and segmentation accuracy of the model, the Epoch is set as 700, and the iteration number is set as 250000 times. Gradient descent and back propagation algorithms are used to update the weight parameters of the model. The learning rate is set as 1×10^{-10}. In order to prevent over-fitting of the SPEDCCNN model, the momentum factor is set as 0.99. The weight parameters are modified locally by using the validation set. The change curves of loss value and accuracy of training set and verification set are shown in figure 8.

As can be seen from figure 8, the loss value of training set and verification set decreases with the increase of iteration number. However, the variation range of the training set is relatively large, because the parameters need to be trained many times in the model initial stage. When the iteration number increases continuously, the loss value of the training set tends to be stable. The curve of the verification set changes slightly. And with the increase of iteration, the loss value of the verification set decreases gradually. When the iteration number is 200000, the loss value tends to be stable. The main reason is that the validation set optimizes the model parameters and improves the model performance. The accuracy curve generally tends to increase. With the increase of iteration, the pixel classification accuracy is improved. Where, the top-1 accuracy represents the accuracy of dividing pixels into two categories, namely the normal leaf area and the disease leaf area. The top-5 accuracy represents the accuracy rate of dividing pixels into multiple categories, that is, the misclassification probability of disease leaf area and normal

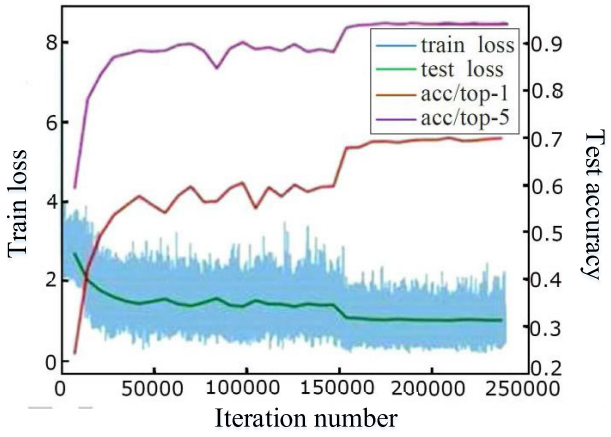


FIGURE 8. The curve of loss and accuracy.

leaf area. From figure 8, the pixel segmentation accuracy is increasing. When the loss value is basically converged, the top-1 accuracy of the model reaches the highest level, indicating that the model has achieved the best training effect.

In the training stage, firstly, the training convergence and the training speed of the model are displayed when the size of input training image patch is 12×12 and 24×24 with the spatial pyramid pooling layer in the model. Then the training convergence and speed of the model with or without a spatial pyramid pooling layer are compared when the input image size is 24×24.

When the input image size is 12×12 and 24×24, the training convergence is shown in figure 9. As can be seen from figure 9, the convergence of the two input sizes is almost the same and there is not much difference. However, there is a big difference in the training time. The model with 12×12 will train about 3000 samples per second. When the input size is 24×24, the model will only train about 750 samples per second. The main reason is that when the model structure and parameters are exactly the same, if the image size is smaller, the data amount is smaller too, and the model is easier to calculate the parameters.

It can be inferred that with the spatial pyramid pooling layer, the convergence is almost the same when the size of the input image is different. It will converge smoothly when the training number is about 8000. In other words, the size of the input image will not affect the convergence of the model.

Then, when the size of the input image is 24×24, the training effect of the model with pyramid pooling layer and without pyramid pooling layer is compared as shown in figure 10. As can be seen from the figure, with the same training data, the convergence of the model will be relatively earlier after adding the pyramid pooling method. The main reason for this is that pyramid pooling layer is a feature extraction process from coarse to fine, which can screen out key information from the whole image more quickly and has better generalization ability for data. Therefore, adding pyramid pooling layer will make convolution neural network model converge earlier.

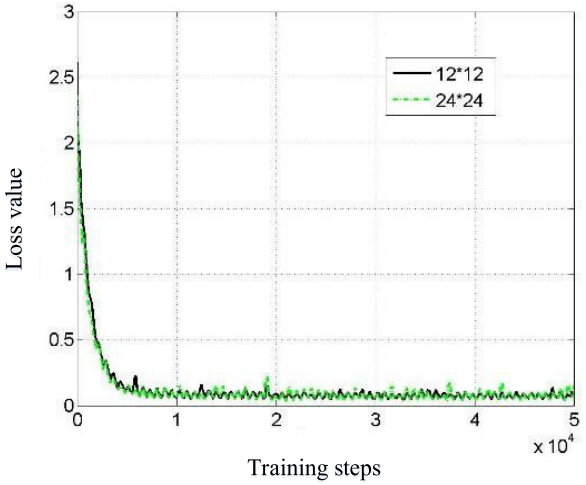


FIGURE 9. Training convergence comparison when the size of input training image patch is 12 × 12 and 24 × 24.

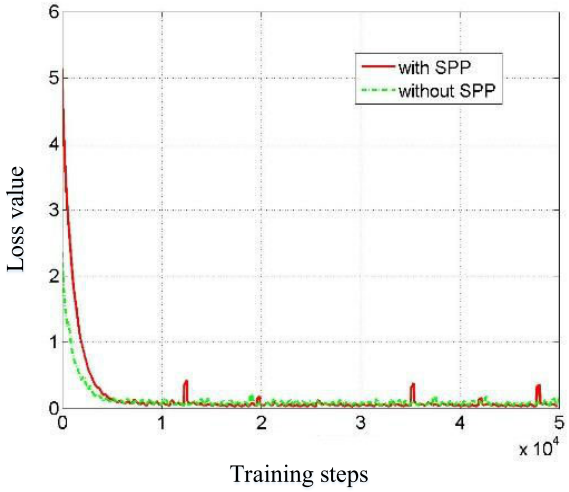


FIGURE 10. Training comparison with/without SPP.

B. VISUAL ANALYSIS

In the training process of the convolution neural network, the convolution layer is used to extract the features of the input image, and the features extracted by different convolution kernels are different. Where, the shallow convolution kernel mainly extracts the primary features of the image including color and contour, etc,. Deep convolution kernel extracts more semantic features of the input image including texture and detail features. To better display the different features extracted by different convolution kernels, the convolution kernel and feature map are visualized. Figure 11 and figure 12 show the convolution kernel and the feature map obtained by different convolution layers, respectively.

As can be seen from figure 11, the convolution kernel (Kernel1∼Kernel3) shows relatively rough information, while the convolution kernel (Kernel4∼kernel5) shows more detailed features. In figure 12, the feature graphs obtained by the convolution layer (Conv1∼Conv3) are mainly contour features, while the feature graphs obtained by the convolution layer (Conv4∼Conv5) mainly include texture features. The visualization results in figure 11 and figure 12 fully reflect the weight sharing in convolution neural network. With the

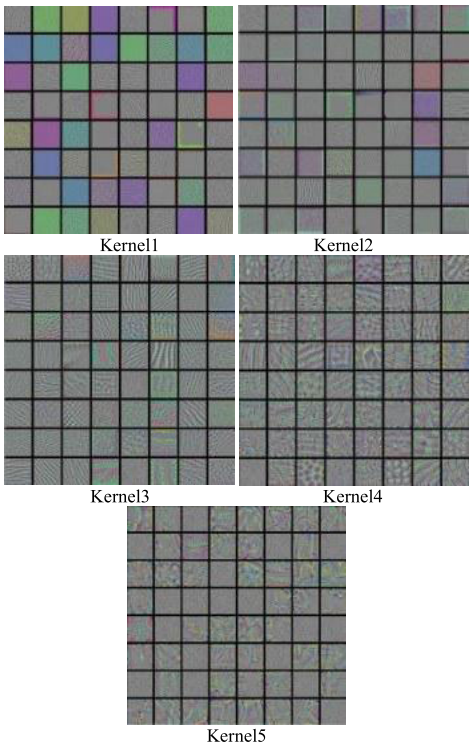


FIGURE 11. Convolution kernel of Kernel1∼Kernel5.

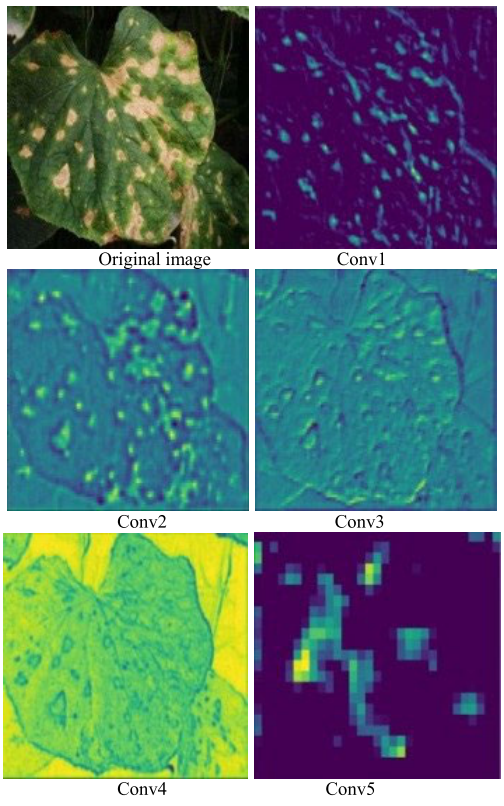


FIGURE 12. Feature map of Conv1∼Conv5.

convolution layers increasing, the model acquires more detailed features of the input image. The different convolution kernels can obtain different features during the model training. Each convolution kernel pays attention to the different parts of the image, and can fully learn the significant area, which plays a significantly role in image accurate segmentation.

C. COMPARISON RESULTS

200 crop disease leaves images from natural scenes are used as test sets to verify the effect of SPEDCCNN. In order to more objectively evaluate the segmentation effect, Precision (P), Recall (R) and Average F value are introduced as evaluation indexes to measure the difference between the segmentation results and the actual labeled images [33].

Where, the Precision is the consistency between the detected disease leaf area and the real disease leaf area. Recall is the proportion of correctly segmented samples in the total sample. The average F value takes segmentation precision P and recall rate R into comprehensive consideration to reflect the overall accuracy. Because the segmentation time is the key technical index to evaluate the practicability of the model, the single image segmentation time is used as the time index to measure the segmentation speed. P, R and F are calculated as follows:

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

$$F = \frac{2 \cdot P \cdot R}{P + R} \tag{8}$$

where, TP is the overlap part between the segmentation results obtained by the network and the original disease leaf area. TN is the overlap part between the segmented normal leaf area and the original normal leaf area. FP is the segmented result that does not belong to the disease leaf area. FN is the segmented result that does not belong to the normal leaf area.

The segmentation results obtained by different segmentation methods are shown in figure 13. It can be seen from the segmentation results that the Hseg segmentation method mistakenly segments most of the normal areas as the disease areas and fails to complete the segmentation task of crop disease leaf. FCN8s network model has a better segmentation effect than Hseg. It is not difficult to see from the segmentation results that this method can be used to segment the general contour of the disease, but the segmentation effect of the small disease leaf is poor. SegNet segmentation network can achieve better segmentation effect. However, SegNet has a poor segmentation effect on the disease leaf edge area. The JointSeg network model has a better segmentation effect than SegNet in terms of the edge area, and it can realize complete segmentation of the detailed parts in the disease leaf image. However, the segmentation results are greatly influenced by the light. It obviously mis-segments the

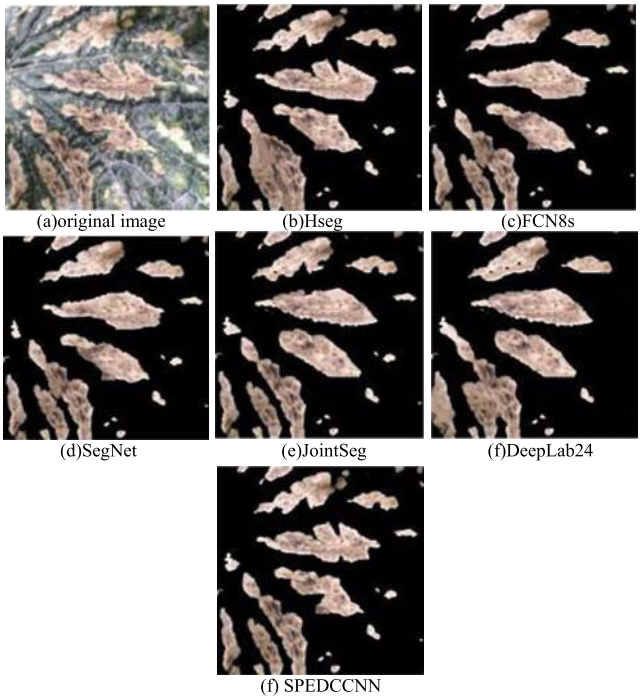


FIGURE 13. The segmentation effect with different methods.

disease leaf with strong light. DeepLab24 network model can completely segment the disease leaf areas and has a good effect on the detailed edge areas. But it cannot segment the disease leaf if the disease areas are adhesive, which results in poor segmentation effect. In the proposed method, the RDL-NET model is used to locate the disease leaf area. Then RSED-NET model based on encoder-decoder network is used to segment the localized disease leaf area to ensure the integrity of the segmentation. Since the edge part of the disease in the leaves is difficult to be segmented and it is easy to be mis-segmented, the morphology optimization layer is used to optimize the segmentation results, which not only ensures the segmentation integrity of the disease leaf area, but also enhances the segmentation effect of the detailed parts.

In order to quantify the segmentation performance of different segmentation methods, P, R, F and single image segmentation time are calculated as shown in table 1. In here, we also compare the famous machine learning methods Gaussian process regression (GPR), support vector machine (SVM), extreme learning machine (ELM) and artificial neural network (ANN) [34]. Note that all the best values are bolder.

As can be seen from table 1, the P, R and F of SPEDCCNN method are the highest (87.59%, 72.61% and 88.83%, respectively). The values of GPR, SVM, ELM and ANN are relatively low. Although DeepLab model uses the same segmented network as SPEDCCNN, SPEDCCNN has higher performance evaluation indexes. Because the region localization network is built by SPEDCCNN in the first phase, which can accurately locate the disease leaf part and reduce the missegmentation rate. FCN8s uses the

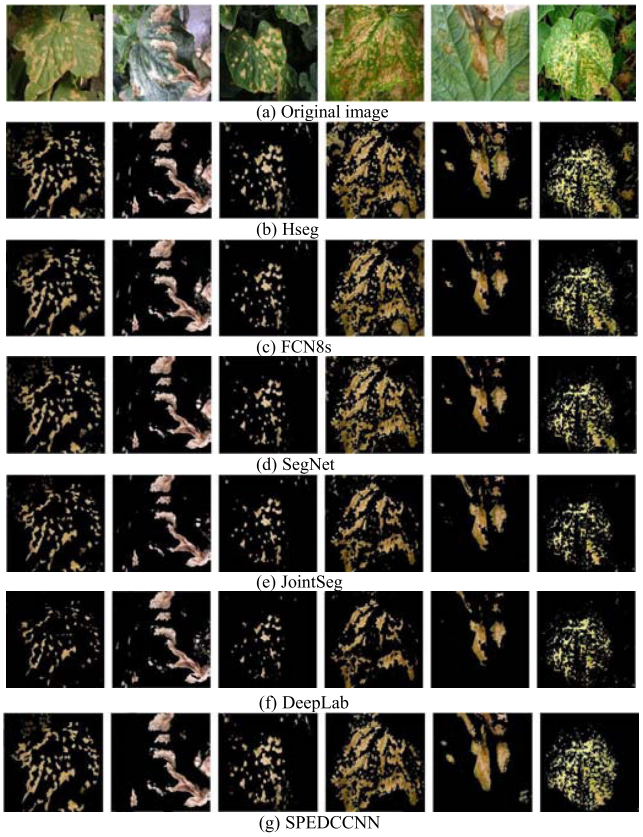


FIGURE 14. The segmentation effect with different methods in different complex backgrounds.

deconvolution process to restore the image resolution optimization segmentation results. However, the down-sampling operation in this method weakens the feature extraction capability, resulting in poor segmentation effect on the small disease leaf areas. Therefore, the segmentation performance value is poor. Compared with the traditional threshold-based Hseg segmentation method, SPEDCCNN is far superior to Hseg in terms of P, R, F. The main reason is that Hseg sets a fixed threshold for image segmentation, so the segmentation result is poor. SegNet and JointSeg are both based on cavity convolution. Cavity convolution can enlarge the local receptive field of the original convolution kernel. However, some disease leaf areas take up a small proportion in the whole leaf, so the two segmentation methods have lower performance than SPEDCCNN.

According to the above results, it can be judged that SPEDCCNN has a good segmentation effect, the segmentation result is close to the real value, which can satisfy the high-precision requirements of crop disease leaves image segmentation in the natural state.

However, the boundary of crop disease leaf area image is relatively complex and occupies only a small area. The similarity of some disease leaf areas and normal leaf areas is high, which leads to the difficulty in distinguishing the boundary between normal area and disease leaf area. Therefore, it can be seen from table 1 that the segmentation performance with the SPEDCCNN algorithm is still low.

TABLE 1. Performance comparison with different methods.

Method	P/%	R/%	F/%	Time/s
GPR	58.67	59.71	59.13	0.48
SVM	60.41	62.55	61.62	0.51
ELM	62.86	61.26	61.09	0.44
ANN	61.77	60.83	61.12	0.38
Hseg	69.83	63.97	83.52	0.33
FCN8s	63.90	61.64	81.63	0.56
SegNet	72.94	66.82	84.27	0.82
JointSeg	74.71	68.35	85.38	0.59
DeepLab	79.02	70.96	86.46	0.71
SPEDCCNN	**87.59**	**72.61**	**88.83**	**0.19**

TABLE 2. Performance comparison with GPR methods in different complex backgrounds.

Background condition	P/%	R/%	F/%	Time/s
soil background	53.62	52.71	53.14	0.33
leaf shading	54.78	55.63	54.96	0.34
multiple leaves	55.67	53.21	54.58	0.36
leaf defect	54.93	55.21	54.86	0.34
disease adhesion	53.68	52.19	53.22	0.35
complicated lighting	54.78	53.08	5.33	0.33

TABLE 3. Performance comparison with SVM methods in different complex backgrounds.

Background condition	P/%	R/%	F/%	Time/s
soil background	58.69	87.25	58.31	0.29
leaf shading	59.36	58.74	58.99	0.31
multiple leaves	60.37	59.22	59.68	0.44
leaf defect	59.38	58.25	59.16	0.42
disease adhesion	58.94	58.66	58.52	0.43
complicated lighting	59.87	57.29	58.16	0.44

TABLE 4. Performance comparison with ELM methods in different complex backgrounds.

Background condition	P/%	R/%	F/%	Time/s
soil background	63.87	64.57	64.12	0.31
leaf shading	64.93	62.56	63.88	0.34
multiple leaves	65.91	63.47	64.57	0.33
leaf defect	64.57	62.78	63.63	0.35
disease adhesion	66.72	63.84	64.89	0.35
complicated lighting	65.28	62.45	63.71	0.32

In terms of the single image segmentation time, the segmentation time of SPEDCCNN is less than that of other segmentation algorithms. Because SegNet and JointSeg are constructed by cavity convolution, the network model requires a large amount of training time, resulting in a lower segmentation efficiency. The segmentation time of the SegNet and JointSeg is 0.82s and 0.59s, respectively. Hseg method does not need to train the model, but it needs to perform morphology operation on each image after segmentation, so much segmentation time is needed. FCN8s requires

TABLE 5. Performance comparison with ANN methods in different complex backgrounds.

Background condition	P/%	R/%	F/%	Time/s
soil background	71.65	69.28	70.32	0.29
leaf shading	72.96	70.55	71.34	0.31
multiple leaves	72.65	70.91	71.82	0.28
leaf defect	73.09	71.57	72.24	0.31
disease adhesion	72.38	69.88	71.21	0.32
complicated lighting	72.99	71.64	71.38	0.33

TABLE 6. Performance comparison with Hseg methods in different complex backgrounds.

Background condition	P/%	R/%	F/%	Time/s
soil background	78.63	68.91	80.91	0.37
leaf shading	77.54	67.49	80.04	0.40
multiple leaves	75.73	66.63	78.97	0.41
leaf defect	73.96	65.42	77.43	0.51
disease adhesion	74.93	66.49	75.36	0.57
complicated lighting	72.72	66.05	76.43	0.48

TABLE 7. Performance comparison with FCN8s methods in different complex backgrounds.

Background condition	P/%	R/%	F/%	Time/s
soil background	79.98	70.94	81.63	0.36
leaf shading	78.68	69.83	81.08	0.39
multiple leaves	78.17	69.06	80.36	0.44
leaf defect	77.49	67.49	80.09	0.48
disease adhesion	77.02	66.01	80.45	0.52
complicated lighting	77.13	67.08	78.79	0.45

TABLE 8. Performance comparison with SegNet methods in different complex backgrounds.

Background condition	P/%	R/%	F/%	Time/s
soil background	81.76	72.27	83.52	0.52
leaf shading	80.71	71.56	83.07	0.50
multiple leaves	79.47	70.23	81.84	0.54
leaf defect	78.68	71.16	81.02	0.61
disease adhesion	78.21	72.43	80.27	0.57
complicated lighting	79.13	71.07	79.74	0.59

deconvolution to restore the resolution of the image during segmentation, thus increasing the segmentation time of a single image.

The training in SPEDCCNN is carried out with a cascade way. The region localization network and the region segmentation network are trained separately, and the transfer learning is used, which not only reduces the hardware requirements, but also reduces the training time. Therefore, the SPEDCCNN method can meet the real-time requirement of image segmentation for disease leaf image.

In order to verify the robustness of the proposed model, different segmentation algorithms are used to perform segmentation tests on crop disease leaf images under a variety

TABLE 9. Performance comparison with JointSeg methods in different complex backgrounds.

Background condition	P/%	R/%	F/%	Time/s
soil background	83.68	74.69	84.27	0.78
leaf shading	82.53	73.93	83.71	0.79
multiple leaves	80.71	72.76	82.49	0.84
leaf defect	78.64	72.09	82.05	0.81
disease adhesion	80.07	71.09	81.07	0.77
complicated lighting	80.13	72.47	83.48	0.79

TABLE 10. Performance comparison with Deeplab methods in different complex backgrounds.

Background condition	P/%	R/%	F/%	Time/s
soil background	86.84	76.96	85.38	0.59
leaf shading	85.42	75.58	84.87	0.61
multiple leaves	83.75	74.43	83.71	0.63
leaf defect	83.13	74.09	82.79	0.66
disease adhesion	83.07	75.94	81.14	0.59
complicated lighting	85.39	73.73	83.08	0.68

TABLE 11. Performance comparison with SPEDCCNN methods in different complex backgrounds.

Background condition	P/%	R/%	F/%	Time/s
soil background	**87.52**	**79.17**	**89.27**	**0.19**
leaf shading	**87.16**	**78.36**	**90.12**	**0.21**
multiple leaves	**87.14**	**78.59**	**89.34**	**0.23**
leaf defect	**86.92**	**78.71**	**91.25**	**0.27**
disease adhesion	**87.38**	**78.61**	**89.67**	**0.17**
complicated lighting	**87.71**	**78.25**	**89.92**	**0.24**

of complex backgrounds. The complex background mainly includes the soil background, leaf shading, multiple leaves, leaf defect, disease adhesion and complicated lighting conditions in the image to be segmented.

For each complex background, 20 images are selected for robust comparison test, and the segmentation effect is shown in figure 14. Tables 2-11 show the segmentation results with different methods in different complex background images.

As shown in figure 14, the SPEDCCNN model is almost unaffected by the environment and can effectively segment the crop disease leaf area with a variety of complex backgrounds. SegNet misclassifies a large area of background areas as disease leaf areas under soil background. DeepLab has good segmentation effect, but it is not accurate enough for small disease leaf area. Under leaf occlusion, the above segmentation methods show false segmentation results, where Hseg method is the most serious, it cannot segment the occluded disease leaves. Under multiple leaves condition,

the above segmentation methods cannot segment the occurred disease leaf areas in multiple leaves completely. In the case of leaf defect, different segmentation methods show different missegmentation situations. DeepLab method mistakenly segments the disease leaf areas into normal areas. However, Hseg method mistakenly segments the normal region as the disease region. For the disease adhesion condition, due to the high similarity of different disease leaf areas, the above segmentation methods cannot segment the leaf with disease adhesion. For complex light, because the light intensity is enhanced, the segmentation complexity is increased, which results in the poor disease leaf area segmentation.

SPEDCCNN can effectively segment disease leaf areas in different scenes and has higher stability. The segmentation effect is less affected by natural conditions, which can meet the segmentation task of crop disease leaves in different scenes.

It can be seen from the above tables, the SPEDCCNN has the best performance, the maximum and the minimum of P, R, F have little difference, which shows that SPEDCCNN has strong robustness under different scenarios. The other segmentation algorithms have higher segmentation accuracy under the soil background condition, which indicates that the soil background has little influence on the segmentation results of different methods. In the leaf occlusion, multiple leaves and leaf adhesion conditions, the index value is unstable, but the overall index value is fluctuated within a small range, it indicates that the other segmentation methods can adjust to a slightly complex segmentation scene. When the leaves in disease adhesion and complex light, the value is decreased, which indicates that disease adhesion and complex light have a great influence on the segmentation effect. In terms of the segmentation time on a single image, SPEDCCNN is lower than other segmentation methods. The average segmentation time on a single image in multiple scenes is 0.21s.

The following is the segmentation result for the disease adhesion of cotton leaf image to further explain the advantage of SPEDCCNN as shown in figure 15. In here, we use the different evaluation indexes due to the specificity of disease adhesion. It contains Correct segmentation (C), over-segmentation (O) and under-segmentation (U).

$$A = \frac{Num_A}{Num_{ALL}} \times 100\% \tag{9}$$

$$O = \frac{Num_O}{Num_{ALL}} \times 100\% \tag{10}$$

$$U = \frac{Num_U}{Num_{ALL}} \times 100\% \tag{11}$$

where Num_{ALL} is the total number of disease spots, Num_A is the number of correct disease spots. Num_O is the over-segmented number of disease spots. Num_U is the under-segmented number of disease spots. Table 12 shows the comparative segmentation statistical results.

It also demonstrates that SPEDCCNN can eliminate the noise and fine structure in the original image. It can suppress

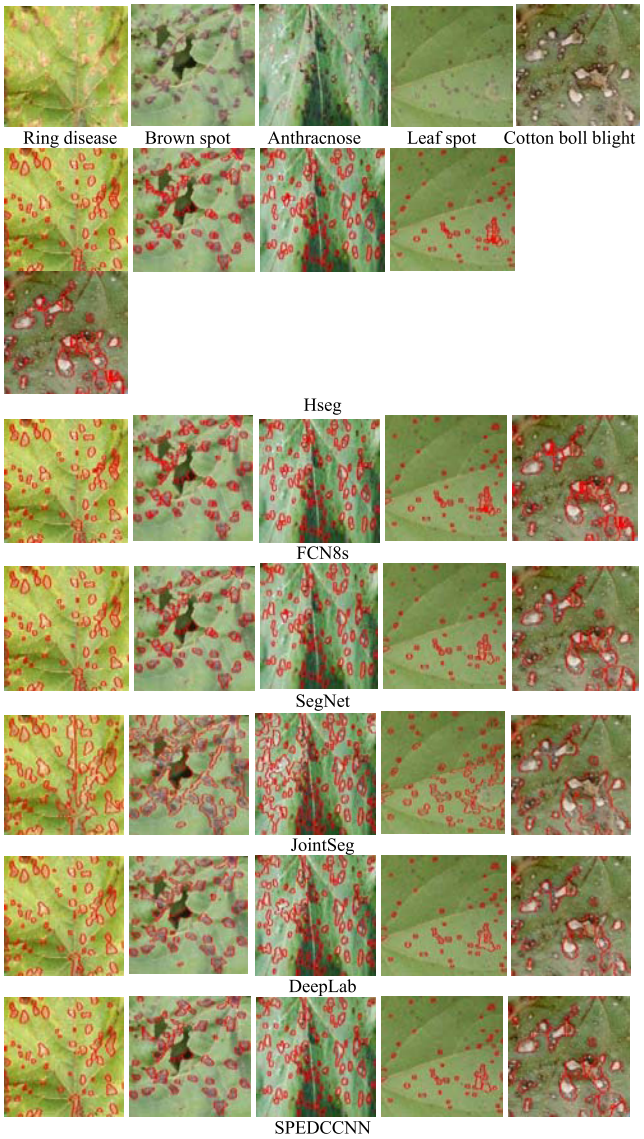


FIGURE 15. Comparison results of cotton leaf adhesions between proposed method and other five methods.

under-segmentation and over-segmentation, which has the best segmentation effect compared with other methods.

Ring disease Brown spot Anthracnose Leaf spot Cotton boll blight

In this paper, there are also segmentation error cases in the leaf adhesion segmentation as shown in figure 16. As shown in figure 16(a), in the rectangular box, there are actually three disease spots, but the SPEDCCNN method only detects two disease spots, so it exists under-segmentation case. In figure 16(b), the disease spot in the rectangular box is one. However, the SPEDCCNN method detects two disease spots resulting in over-segmentation case.

In the next paragraph, we make comparison with other state-of-the-art methods including IHI [35], ABC-FCM [36], MRCRG [37] on the public disease image sets: Arkansas Plant Diseases Database (https://www.uaex.edu/yardgarden/

TABLE 12. Performance comparison with different methods in different complex backgrounds.

Method	Background condition	A	O	U	Time/s
Hseg	Ring disease	70.2	24.3	5.9	0.33
	Brown spot	61.7	31.8	6.9	0.32
	Anthracnose	66.4	26.9	6.9	0.34
	Leaf spot	73.3	19.8	7.3	0.39
	Cotton boll blight	72.1	18.4	9.6	0.33
	Average	68.7	24.3	7.3	0.34
FCN8s	Ring disease	35.8	56.4	7.5	0.36
	Brown spot	35.9	58.5	5.2	0.35
	Anthracnose	37.1	49.9	12.6	0.33
	Leaf spot	40.6	50.3	8.8	0.37
	Cotton boll blight	32.2	57.2	10.3	0.38
	Average	36.3	54.4	8.9	0.36
SegNet	Ring disease	84.6	16.9	2.3	0.43
	Brown spot	78.5	18.1	3.3	0.49
	Anthracnose	86.1	10.3	3.6	0.44
	Leaf spot	88.3	9.8	1.9	0.48
	Cotton boll blight	81.6	12.5	5.8	0.43
	Average	83.8	13.4	3.4	0.45
JointSeg	Ring disease	71.9	4.6	31.6	0.31
	Brown spot	71.3	3.7	25.8	0.32
	Anthracnose	72.2	6.8	21.7	0.39
	Leaf spot	67.9	5.1	27.8	0.36
	Cotton boll blight	80.4	4.5	15.9	0.41
	Average	71.3	4.9	24.5	0.36
DeepLab	Ring disease	81.5	3.8	14.7	0.11
	Brown spot	80.8	4.1	15.1	0.12
	Anthracnose	85.1	3.9	10.9	0.11
	Leaf spot	81.6	2.5	15.8	0.13
	Cotton boll blight	82.5	5.7	11.8	0.17
	Average	82.3	4.1	13.6	0.18
SPEDCCNN	Ring disease	**95.9**	**2.1**	**1.9**	**0.08**
	Brown spot	**93.3**	**3.1**	**3.5**	**0.09**
	Anthracnose	**90.4**	**4.8**	**6.7**	**0.11**
	Leaf spot	**97.9**	**0.7**	**1.4**	**0.12**
	Cotton boll blight	**91.1**	**3.4**	**5.4**	**0.14**
	Average	**93.7**	**2.7**	**3.8**	**0.12**

resourcelibrary/diseases/) and NARO (https://www.gene.affrc.go.jp/index_en.php). The samples are shown in figure 17.

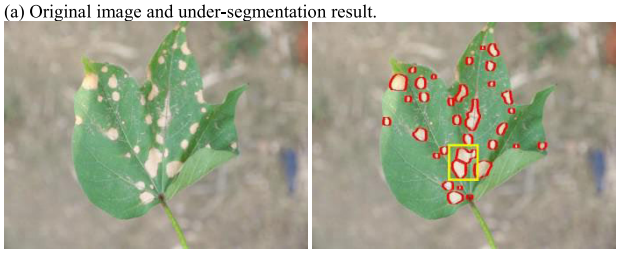
TABLE 13. Comparison of segmentation results.

Leaf image	IHI	ABC-FCM	MRCRG	SPEDCCNN
Cotton blight disease	0.8228	0.8537	0.8945	**0.9379**
Cotton leaf spot	0.8579	0.8962	0.9015	**0.9473**
Cotton powdery mildew	0.8677	0.8994	0.9155	**0.9635**
Solanum nigum blight disease	0.8793	0.9014	0.9265	**0.9811**

TABLE 14. Mutual information for segmentation result.

Leaf image	IHI	ABC-FCM	MRCRG	SPEDCCNN
Cotton blight disease	0.815	0.854	0.893	**0.959**
Cotton leaf spot	0.796	0.887	0.901	**0.961**
Cotton powdery mildew	0.805	0.874	0.899	**0.968**
Solanum nigum blight disease	0.811	0.872	0.921	**0.979**

(a) Original image and under-segmentation result.



(b) Original image and over-segmentation result.

FIGURE 16. The Cotton leaf disease segmentation error case.

FIGURE 17. The sample leaf disease images.

We use dice similarity coefficient (DSC) to evaluate the segmentation efficiency.

The calculation of DSC is as follows.

$$\mathrm{DSC} = 2\mathrm{TP}/(2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}) \tag{12}$$

where TP, FP, FN are explained in section 3.3. The results are displayed in table 13.

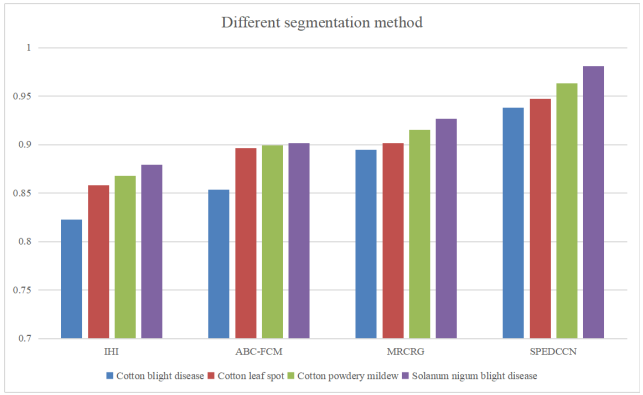


FIGURE 18. Performance of segmentation methods.

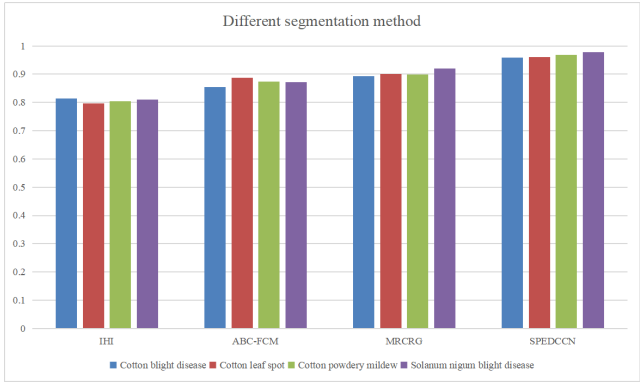


FIGURE 19. Performance of segmentation methods.

TABLE 15. Sensitivity for segmentation result.

	IHI	ABC-FCM	MRCRG	SPEDCCN
Cotton blight disease	0.854	0.925	0.918	**0.957**
Cotton leaf spot	0.905	0.917	0.924	**0.979**
Cotton powdery mildew	0.923	0.931	0.928	**0.972**
Solanum nigum blight disease	0.911	0.922	0.936	**0.983**

It can be seen clearly from above Table 13 that the proposed segmentation method in this paper can effectively segment regions from plant diseases and ensure the maximum similarity rate. This means that the SPEDCCNN method has successfully segmented the disease leaf area. According to the Table 4, the value related graphical representation is shown in figure 18. It can be clearly seen from figure 18 that the SPEDCCNN method ensures that the performance values of different plant leaves, such as cotton fusarium wilt leaves, cotton macula leaves, cotton powdery mildew leaves and cymbidiae fusarium wilt leaves are relatively good. The segmented sections have useful information for analyzing disease-related features and information. Then the mutual information of the segmented regions is shown in Table 14.

TABLE 16. Specificity for segmentation result.

Leaf image	IHI	ABC-FCM	MRCRG	SPEDCCN
Cotton blight disease	0.953	0.961	0.959	**0.979**
Cotton leaf spot	0.962	0.964	0.965	**0.984**
Cotton powdery mildew	0.965	0.972	0.971	**0.979**
Solanum nigum blight disease	0.972	0.976	0.986	**0.993**

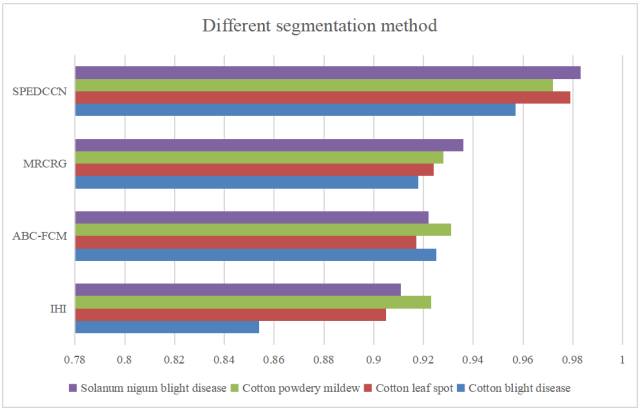


FIGURE 20. Sensitivity performance of segmentation methods.

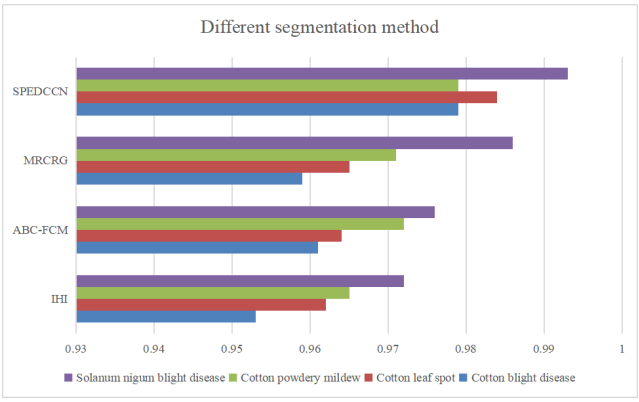


FIGURE 21. Specificity performance of segmentation methods.

It can be clearly seen from above Table 14 that the indicators of the SPEDCCNN method in this paper effectively reflect several information of the disease. It guarantees the maximum value of mutual information, which means that segmented regions based on SPEDCCNN methods have valuable information. According to Table 5, the relative graphs of mutual information values are shown in figure 19.

As can be seen from figure 19, the SPEDCCNN method ensures that different plant leaves, such as cotton fusarium wilt leaves, cotton porphyry leaves, cotton powdery mildew leaves and tarragon fusarium wilt leaves, have high mutual information values. The segmented sections have useful information for analyzing disease-related features and information. We also adopt the sensitivity and specificity metric to examine the accuracy of the segmented region as shown in table 15 and 16.

The above tables 15 and 16, clearly indicates that the SPEDCCNN method has high values (98.3% of sensitivity and 99.3% of specificity) about affected disease with effective manner. The sensitivity and specificity values related graphical representation are shown in figure 20,21.

Thus, from the above analysis, the SPEDCCNN approach successfully recognizes the plant leaf blight disease from cotton and Solanum nigrum images when compared to the other methods.

V. CONCLUSION

Aiming at the traditional crop disease leaf image segmentation problems, this paper proposes a spatial pyramid-oriented encoder-decoder cascade convolution neural network. Firstly, the region localization network (RDL-NET) is used to accurately locate the disease leaf part to reduce the background interference. Then the region segmentation network (RSED-NET) based on encoder-decoder structure is used to accurately segment the disease leaf area. In order to verify the robustness of SPEDCCNN, the SPEDCCNN and other segmentation methods are tested in different scenes. The P, R, F, A, O, U results show that the SPEDCCNN model can complete the disease segmentation task under different scenarios with exceeding 90% values. In terms of segmentation time, SPEDCCNN is superior to other segmentation methods. The segmentation time on a single image is only 0.21s, it indicates that SPEDCCNN can satisfy the requirement of crop disease leaf segmentation. The final experiment results show that the proposed method in this paper has high segmentation accuracy and fast operation speed.

However, this new method can only complete the segmentation task in a small range of scene. We still face two challenges as follows:

1) For a large range of scene or harsh environments, it needs to be further researched.

2) Also some different and complex crop disease types will bring challenges to the precise segmentation.

3) When the adhesion of the disease is particularly similar, and the size of the disease is overlapped, there will be a partial under-segmentation condition.

In the future, we will further compare the potential benefits and limitations of the existing image segmentation methods and explore a robust, fast and accurate disease leaf image segmentation method.

REFERENCES

[1] M. J. Wedger, T. Pusadee, A. Wongtamee, and K. M. Olsen, "Discordant patterns of introgression suggest historical gene flow into Thai weedy rice from domesticated and wild relatives," *J. Heredity*, vol. 110, no. 5, pp. 601–609, Aug. 2019.

[2] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," *Comput. Electron. Agricult.*, vol. 151, pp. 61–69, Aug. 2018.

[3] A. Kaya, A. S. Keceli, C. Catal, H. Y. Yalic, H. Temucin, and B. Tekinerdogan, "Analysis of transfer learning for deep neural network based plant classification models," *Comput. Electron. Agricult.*, vol. 158, pp. 20–29, Mar. 2019.

[4] B. S. Kusumo, A. Heryana, O. Mahendra, and H. F. Pardede, "Machine learning-based for automatic detection of corn-plant diseases using image processing," in *Proc. Int. Conf. Comput., Control, Informat. Appl. (IC3INA)*, Nov. 2018, pp. 93–97.

[5] J. Chen, H. Yin, and D. Zhang, "A self-adaptive classification method for plant disease detection using GMDH-logistic model," *Sustain. Comput., Informat. Syst.*, vol. 28, Dec. 2020, Art. no. 100415.

[6] M. G. Du and S. W. Zhang, "Crop disease leaf image segmentation based on genetic algorithm and maximum entropy," *Appl. Mech. Mater.*, vols. 713–715, pp. 1670–1674, Jan. 2015.

[7] F. Wang, J. W. Li, W. Shi, and G. P. Liao, "Leaf image segmentation method based on multifractal detrended fluctuation analysis," *J. Appl. Phys.*, vol. 114, no. 21, pp. 214905–214909, 2013.

[8] H. Jiang, J. Zhang, Y. Yuan, M. He, and S. Zheng, "Segmentation of cucumber disease leaf image based on MDMP-LSM," *Nongye Gongcheng Xuebao/Trans. Chin. Soc. Agricult. Eng.*, vol. 28, no. 21, pp. 142–148, 2012.

[9] T. Fan and J. Xu, "Image classification of crop diseases and pests based on deep learning and fuzzy system," *Int. J. Data Warehousing Mining*, vol. 16, no. 2, pp. 34–47, Apr. 2020.

[10] P. Sharma, Y. P. S. Berwal, and W. Ghai, "Performance analysis of deep learning CNN models for disease detection in plants using image segmentation," *Inf. Process. Agricult.*, vol. 7, no. 4, pp. 566–574, Dec. 2020, doi: 10.1016/j.inpa.2019.11.001.

[11] E. Goceri, "Challenges and recent solutions for image segmentation in the era of deep learning," in *Proc. 9th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Nov. 2019, pp. 1–6.

[12] S. Yin and H. Li, "Hot region selection based on selective search and modified fuzzy C-Means in remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5862–5871, 2020, doi: 10.1109/JSTARS.2020.3025582.

[13] L. Mauch, C. Wang, and B. Yang, "Subset selection for visualization of relevant image fractions for deep learning based semantic image segmentation," *J. Franklin Inst.*, vol. 355, no. 4, pp. 1931–1944, Mar. 2018.

[14] L. Liu, X. Cheng, and J. Lai, "Segmentation method for cotton canopy image based on improved fully convolutional network model," *Nongye Gongcheng Xuebao/Trans. Chin. Soc. Agricult. Eng.*, vol. 34, no. 12, pp. 193–201, 2018.

[15] Z. Liu, X. Zheng, Z. Xiao, J. Bao, L. Shuang, F. Sheng, L. Jie, T. Yu, and Z. Qiu, "Research on integrated algorithm based on convolutional neural network for rice disease identification," *J. Phys., Conf. Ser.*, vol. 1646, Sep. 2020, Art. no. 012066.

[16] M. Ji, K. Zhang, Q. Wu, and Z. Deng, "Multi-label learning for crop leaf diseases recognition and severity estimation based on convolutional neural networks," *Soft Comput.*, vol. 24, no. 20, pp. 15327–15340, Oct. 2020.

[17] X. Xiong, L. Duan, L. Liu, H. Tu, P. Yang, D. Wu, G. Chen, L. Xiong, W. Yang, and Q. Liu, "Panicle-SEG: A robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization," *Plant Methods*, vol. 13, no. 1, pp. 104–111, Dec. 2017.

[18] J. Li, C. Xu, L. Jiang, Y. Xiao, L. Deng, and Z. Han, "Detection and analysis of behavior trajectory for sea cucumbers based on deep learning," *IEEE Access*, vol. 8, pp. 18832–18840, 2020, doi: 10.1109/ACCESS.2019.2962823.

[19] I. Ahmad, M. Hamid, S. Yousaf, S. T. Shah, and M. O. Ahmad, "Optimizing pretrained convolutional neural networks for tomato leaf disease detection," *Complexity*, vol. 2020, pp. 1–6, Sep. 2020.

[20] M. P. Pound, J. A. Atkinson, A. J. Townsend, M. H. Wilson, M. Griffiths, A. S. Jackson, A. Bulat, G. Tzimiropoulos, D. M. Wells, E. H. Murchie, T. P. Pridmore, and A. P. French, "Deep machine learning provides state-of-the-art performance in image-based plant phenotyping," *GigaScience*, vol. 6, no. 10, pp. 1–10, Oct. 2017.

[21] S. Wen, M. Dong, Y. Yang, P. Zhou, T. Huang, and Y. Chen, "End-to-end detection-segmentation system for face labeling," *IEEE Trans. Emerg. Topics Comput. Intell.*, early access, Nov. 6, 2019, doi: 10.1109/TETCI.2019.2947319.

[22] P. Yin, R. Yuan, Y. Cheng, and Q. Wu, "Deep guidance network for biomedical image segmentation," *IEEE Access*, vol. 8, pp. 116106–116116, 2020, doi: 10.1109/ACCESS.2020.3002835.

[23] A. Van Opbroek, H. C. Achterberg, M. W. Vernooij, and M. D. Bruijne, "Transfer learning for image segmentation by combining image weighting and kernel learning," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 213–224, Jan. 2019, doi: 10.1109/TMI.2018.2859478.

[24] W. Wang, I. Pollak, T.-S. Wong, C. A. Bouman, M. P. Harper, and J. M. Siskind, "Hierarchical stochastic image grammars for classification and segmentation," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3033–3052, Oct. 2006, doi: 10.1109/TIP.2006.877496.
[25] M. Liu, D. Cheng, K. Wang, and Y. Wang, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, vol. 16, nos. 3–4, pp. 295–308, Oct. 2018.
[26] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015, doi: 10.1109/TPAMI.2015.2389824.
[27] W. Zhen, Z. Shanwen, and Z. Baoping, "Crop diseases leaf segmentation method based on cascade convolutional neural network," *Comput. Eng. Appl.*, vol. 56, no. 15, pp. 242–250, 2020.
[28] L. Duan, X. Xiong, Q. Liu, W. Yang, and C. Huang, "Field rice panicle segmentation based on deep full convolutional neural network," *Trans. Chin. Soc. Agricult. Eng.*, vol. 34, no. 12, pp. 202–209, 2018.
[29] N. Zou, Z. Xiang, Y. Chen, S. Chen, and C. Qiao, "Boundary-aware CNN for semantic segmentation," *IEEE Access*, vol. 7, pp. 114520–114528, 2019, doi: 10.1109/ACCESS.2019.2935816.
[30] H. Lu, Z. Cao, Y. Xiao, Z. Fang, Y. Zhu, and K. Xian, "Fine-grained maize tassel trait characterization with multi-view representations," *Comput. Electron. Agricult.*, vol. 118, pp. 143–158, Oct. 2015.
[31] A. Ion, J. Carreira, and C. Sminchisescu, "Probabilistic joint image segmentation and labeling by figure-ground composition," *Int. J. Comput. Vis.*, vol. 107, no. 1, pp. 40–57, Mar. 2014.
[32] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," *Comput. Sci.*, vol. 4, pp. 357–361, Dec. 2014.
[33] X. Wang, S. Yin, K. Sun, H. Li, J. Liu, and S. Karim, "GKFC-CNN: Modified Gaussian kernel fuzzy C-means and convolutional neural network for apple segmentation and recognition," *J. Appl. Sci. Eng.*, vol. 23, no. 3, pp. 555–561, 2020.
[34] W.-J. Niu and Z.-K. Feng, "Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management," *Sustain. Cities Soc.*, vol. 64, Jan. 2021, Art. no. 102562.
[35] S. Kalaivani, S. P. Shantharajah, and T. Padma, "Agricultural leaf blight disease segmentation using indices based histogram intensity segmentation approach," *Multimedia Tools Appl.*, vol. 79, nos. 13–14, pp. 9145–9159, Apr. 2020.
[36] S. K. P. Kumar, M. G. Sumithra, and N. Saranya, "Artificial bee colony-based fuzzy C means (ABC-FCM) segmentation algorithm and dimensionality reduction for leaf disease detection in bioinformatics," *J. Supercomput.*, vol. 75, no. 12, pp. 8293–8311, Dec. 2019.
[37] N. Jothiaruna, K. J. A. Sundar, and M. I. Ahmed, "A disease spot segmentation method using comprehensive color feature with multi-resolution channel and region growing," *Multimedia Tools Appl.*, vol. 80, no. 4, pp. 3327–3335, 2020.

YUXIA YUAN was born in Zhengzhou, Henan, China, in 1982. She graduated from Zhengzhou University. She is currently an Associate Professor with the School of Electrical Engineering, Zhengzhou University of Science and Technology, Zhengzhou. Her research interests include image processing, big data, and cloud computing.

ZENGYONG XU was born in Zhengzhou, Henan, China, in 1982. He graduated from Henan Polytechnic University. He is currently a Lecturer with the School of Automotive Studies, Henan College of Transportation. His research interests include image processing, big data, and cloud computing.

GANG LU was born in Zhengzhou, Henan, China, in 1973. He graduated from Zhengzhou University. He is currently a Lecturer with the School of Electrical Engineering and Automation, Luoyang Institute of Science and Technology, Luoyang, China. His research interests include image processing, big data, and cloud computing.