[1]Jie Yang

[1]Jiya Tian

[1] Jinchao Miao

[1] Yunsheng Chen

[1] Shuping Zhang

[1] YixuanWu

Adaptive Loss-Guided Multi-Stage Residual ASPP for Lesion Segmentation and Disease Detection in Cucumber under Complex Backgrounds

Journal of Electrical Systems

Abstract: - In complex environments, the performance of leaf segmentation models for cucumber diseases tends to decrease due to the overlapping of obstructions such as shadows and leaf debris, as well as the impact of uneven illumination. This, in turn, directly affects the subsequent disease detection tasks. Furthermore, the imbalance in pixel ratio between background and lesion areas can seriously impact the accuracy of lesion extraction. To address these issues, we propose an image segmentation framework based on a two-stage Atrous Spatial Pyramid Pooling (ASPP) with adaptive loss for precise cucumber lesion segmentation under complex scenarios. The model architecture guided by adaptive loss can reduce the loss weight generated by pixels that are easy to classify. Therefore, during the model training process, more attention will be paid to pixels that are challenging to classify, thereby improving the accuracy of lesion segmentation. The two-stage model, which we refer to as the LS-ASPP model, consists of Leaf-ASPP and Spot-ASPP. In the first stage of this model, images are processed through the Leaf-ASPP architecture, which extracts leaf contours from the complex background. In this stage, we incorporate attention modules and residual structures into the ASPP framework, creating an improved residual ASPP network. This can capture multi-scale semantic information of diseased leaves, enhancing the model's edge perception capabilities. In the second stage, the Spot-ASPP model segments the lesion areas from the extracted leaf contours. We adjust the dilation rate of the Atrous Spatial Pyramid Pooling (ASPP) to capture smaller targets, and introduce a Convolutional Attention Block Module (CABM) to highlight important information. Compared to existing deep learning models, this framework can improve semantic segmentation accuracy under complex conditions.

Keywords: Cucumber downy mildew; Lesion semantic segmentation; Convolution channel attention module; Residual structure

1. Introduction

Plant diseases are one of the main causes of crop yield reduction in cucumbers, often leading to significant crop losses and even total crop failure, directly affecting crop quality and yield, and resulting in substantial economic losses [1-2]. Therefore, to improve crop quality and yield, it is crucial to study plant diseases and be able to detect and identify them in order to determine the optimal time for prevention and treatment. When crops are infected by pathogens, most of the symptoms are manifested on the leaves, resulting in various phenomena such as lesions [3] and localized rot and wilting [4]. There are numerous types of crop diseases, and manual diagnosis is complex with a high misdiagnosis rate [5-6]. Simultaneously, the spraying of pesticides is a primary measure for the prevention and treatment of plant diseases. However, pesticide application often overlooks the severity of

[1] [1]* information engineering College, Xinjiang Science and Engineering, Aksu, 843000, Xinjiang, China.Email:42683387@qq.com

Copyright © JES 2024 on-line : journal.esrgroups.org

crop symptoms [7], leading to imprecise pesticide dosages, resulting in soil pollution and overuse of pesticides [8]. Hence, to assist non-specialists in crop production in effectively carrying out their duties, detecting diseases and diagnosing them promptly to avoid further crop losses, techniques in artificial intelligence and digital image processing are typically employed for disease detection. Segmentation and extraction of lesions on cucumber leaves can provide a reliable basis for future plant disease diagnosis and are of significant importance for the prevention and control of plant diseases and pests.

The severity of crop diseases can be assessed using image segmentation methods. Traditional methods assess the severity of crops such as cucumbers by segmenting crop leaves and lesion areas and calculating their areas. The main methods include: (1) Threshold-based segmentation methods [9-10]: These methods, including genetic algorithms and Otsu's method, are relatively simple to implement and have low computational requirements. However, in the real world, scenarios are often complex. Due to the subtle differences in grayscale values of crop leaves and the overlapping grayscale values between multi-scale leaves, data image processing is challenging, causing difficulties for lesion segmentation and detection. (2) Cluster-based segmentation methods [11-12]: Common machine learning clustering segmentation methods include K-means and Fuzzy C-means. These methods are applicable to most samples. However, the segmentation results often depend on the selection of initial parameters, which can lead to local optima and reduce segmentation accuracy. (3) Region-based segmentation methods [13-14]: Common region-based methods include region growing and watershed algorithms. As the region growing method is sensitive to noise, it is not suitable for leaf lesion segmentation and detection under complex scenarios. As can be seen, traditional methods have a high complexity in preprocessing, poor generalizability, and most methods only target a single disease. Their ability to transfer to different types of plant disease segmentation is weak, making it difficult to handle multi-disease and multi-scale lesion segmentation simultaneously, which will severely impact the ability to segment and detect multiple diseases.

In addition to early traditional methods for image segmentation, deep learning technology proposes solutions to enhance the transferability of plant lesion segmentation tasks and to improve segmentation detection accuracy. Differing from early manual feature extraction methods, deep learning segmentation network models avoid cumbersome preprocessing stages, such as Fully Convolutional Networks[15], adopt end-to-end feature extraction methods, and do not require complex preprocessing, like the U-net network[16], which can also achieve more accurate segmentation results, closer to real samples. Additionally, various types of DeepLab network[17] structures, with their higher accuracy and stronger transferability, will encourage more researchers to enter the field of agricultural image processing and make good progress[18-20]. However, due to the complexity of the real environment, ordinary deep learning methods only function in a single environment, and the segmentation effect in complex scenarios is poor, meaning the model lacks universality. For example, Chen[] et al. proposed an improved semantic segmentation network, BLSNet, based on Unet segmentation network introducing attention mechanisms and multi-scale modules, which has high segmentation accuracy and classification accuracy. In recent years, more and more researchers pay more attention to lesion segmentation and disease recognition under complex backgrounds, and explore the importance of lesion extraction for assessing disease severity.

Wang[] et al. proposed a network structure combining DeepLabV3+ and U-Net for lesion segmentation and disease recognition in complex backgrounds, reducing the interference of similar pixel values in complex backgrounds in lesion extraction. However, due to the imbalance of pixel ratio in lesion segmentation, and the difficulty of recognizing leaf edge pixels caused by leaf overlap or debris occlusion, the designers of DUNet did not optimize the network structure for these problems. This paper proposes a novel two-stage LD-ASPP network model guided by adaptive loss, which effectively targets the imbalance of pixel ratio in lesion segmentation, and validates the high precision and accuracy of this method on a cucumber leaf disease dataset. The main contributions of this study are as follows:

(1) To address the imbalance of background pixel and foreground target pixel ratios in lesion segmentation, the proposed adaptive loss algorithm is utilized to enhance the attention to difficult-to-distinguish edge pixels of leaves and to improve the category probability of pixels belonging to the background or foreground. This is mainly achieved through the modulation factor present in adaptive loss to adjust the classification weight. This factor decreases as the pixel classification confidence increases. At this point, during the training process, this modulation factor can reduce the weight of easily classified pixels and increase the weight of difficult-to-classify pixels, prompting the model to effectively focus on the difficult-to-classify pixels. As during the model training process, the classifier needs to classify a large amount of easily classified sample data, which leads to a decrease in the segmentation accuracy of sparse samples.

（2）The first stage, Leaf-ASPP network structure, primarily segments out the complete leaf contour from the complex environment. To enhance the network model's focus on key areas and reduce the impact of overlapping leaves and debris, the Mult Residual ASPP improved module is used in place of the ASPP module to extract multi-scale feature images. An attention mechanism is introduced, combining ordinary convolution with small convolution, to acquire features with stronger discrimination.

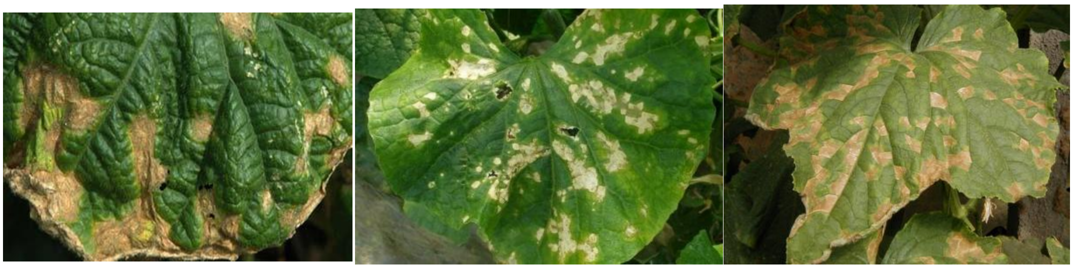
（3）The second stage, Spot-ASPP network structure, extracts lesion areas from the segmented complete leaves. Adjusting the dilation rate of the ASPP module enables the recognition of smaller lesion information, avoids accuracy loss, and aims at lesion segmentation to acquire a more complete lesion area. This stage introduces an enhancement of the network model's focus on key area features and includes a convolutional channel attention block (CABM) to capture attention on important area pixels.

（4）Combining the design of the two-stage model, the LS-ASPP integrated model is constructed for cucumber lesion segmentation and disease detection in complex scenarios. The comprehensive segmentation task will be decomposed into two stages: diseased leaf segmentation and lesion extraction, effectively improving segmentation accuracy and playing a crucial role in other disease classification assessment tasks.

2 Data Sources

2.1 Dataset

The experimental data comes from images of three types of cucumber diseases in the cucumber dataset from the AI laboratory of Xinjiang Institute of Technology, including 48 images of cucumber powdery mildew, 88 images of cucumber angular leaf spot, and 64 images of cucumber downy mildew. Fig. 1 presents examples of diseased samples from the three types of cucumber leaf diseases in the dataset.



a. Cucumber angular leaf spot b. Cucumber downy mildew c. Cucumber powdery mildew

Fig.1 An example of a cucumber leaf with diseased spots

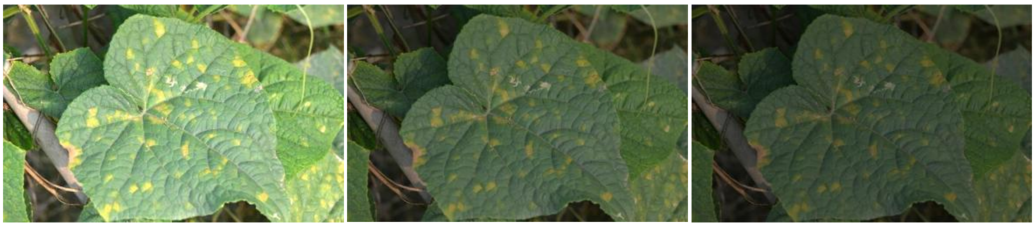
As can be seen from Fig. 1, the identification and segmentation of cucumber leaf diseases mainly have the following four difficulties: 1) Different diseases of cucumber leaves will present different characteristics, and lesion segmentation needs to be conducted according to these disease features; 2) The similarities between the characteristics of different diseases interfere with the recognition task, leading to a low recognition accuracy rate, such as between cucumber downy mildew and cucumber angular leaf spot; 3) Complex backgrounds interfere with leaf segmentation, and shadows from obstructions are misdetected as lesion areas; 4) Due to the irregular shape of cucumber leaf lesion areas, initial small lesions are difficult to discover, which increases the difficulty of segmentation.

2.2 Data Augmentation

2.2.1 Training Data Augmentation

Due to the uneven distribution of samples, the training process may lead to model overfitting. Therefore, data augmentation is timely adopted to improve the generalization performance of the training model. In this study, for each batch of data during the training process, a random augmentation method is chosen. Without increasing the original dataset, the original data features are preserved to better simulate the differences among various samples in a real complex environment. The training set mainly employs the following data augmentation methods: 1) Flipping: The images are manipulated through horizontal flipping, vertical flipping, vertical-then-horizontal flipping, and mirroring. There are four flipping methods in total, simulating the randomness of shooting angles when collecting samples, without changing the shape of lesions or their distribution on the leaf. 2) Color jitter: By adjusting the brightness or saturation of the image, the differences in real-world lighting scenarios are simulated, ensuring no image distortion in the real environment. 3) Adding noise: Noise is added to simulate the noise generated during data collection, preventing the network model from overfitting. The effect of test data augmentation is shown in Fig. 2:

a. Original cucumber image b. Vertical flip c. Clipping

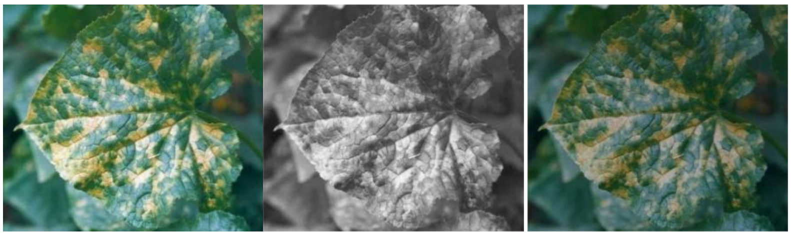


d. Enhanced brightness e. Reduce brightness by 30% f. Reduce brightness by 50%

Fig.2 Illustration of training data sample augmentation

2.2.2 Test Data Augmentation

To address the data fluctuations in real complex environments, the dataset used in this study was shot under laboratory conditions. To simulate the scenarios of insufficient lighting, leaf deviation from the center, and obstructions in natural shooting, three image augmentation methods are employed to enhance the test data: 1) Translation: A part of the image is first cropped, and then the missing pixels are filled in with the border pixel values, simulating situations such as leaf deviation from the lens center and incomplete leaves during shooting. 2) Occlusion: Randomly-sized, randomly-located, and randomly-rotated images of tomato fruit, soil, and green leaves are generated to occlude the leaf area, simulating the target leaf being obscured. 3) Cropping: By cropping a certain proportion of the main area, the diseased part can be effectively highlighted. 4) Reducing brightness: The image brightness is reduced to simulate the condition of insufficient light. The brightness of the diseased leaf images can be sequentially reduced to 50% of the original image and grayscale. The results of the test data augmentation are shown in Fig. 3.

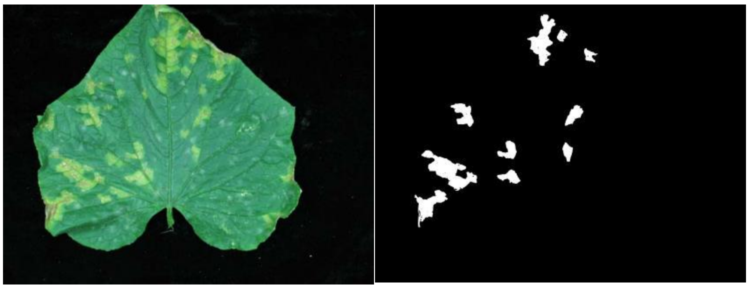


a. Original test image b. Grayscale c. Reduce brightness by 50%

Fig.3 Illustration of test data sample augmentation

2.3 Data Annotation

For the plant disease identification network model inherently equipped with image-level annotations, it is also necessary to process the training disease spot segmentation network model without annotations in the dataset of Xinjiang Institute of Technology laboratory. For the training of disease spot segmentation models, pixel-level annotations are required for disease spots as shown in Fig.4. (Add leaf annotation and disease spot annotation.)



a. Original diseased leaf b. Pixel-level annotation

Fig.4 Pixel-level annotation of diseased leaf spots

Pixel-level annotation requires a substantial amount of human and material resources. Therefore, disease spot annotations are performed on a portion of the samples, selecting a total of 30 diseased non-healthy plant samples from various categories, with the LabelMe image annotation tool used for disease spot area annotation. Among them, 10 disease spot samples are used for training the segmentation model, and 20 disease spot sample images are used for testing the segmentation model.

3 Method

The main issues faced by cucumber disease spot segmentation and disease detection include:

(1) Pixel ratio imbalance: This issue mainly stems from sparse pixels in the target area, leading to an imbalance in the ratio of background area pixels to target pixels, such as in the task of extracting small disease spots. As shown in the figure, the ratio of disease spot area pixels to leaf pixels is small, making disease spot area pixels prone to loss. Additionally, a large number of easily classified pixels in the background area can generate significant losses, resulting in the total loss, when finally computed, far exceeding that of the disease spot area pixels. This directly reduces the efficiency of model training and severely impacts segmentation results.

(2) High number of hard samples: Hard samples mainly originate from complex background images in the natural environment. Due to interference from clutter, leaf overlap, shadow occlusion, uneven lighting, etc., the pixel areas of the interfering parts can be considered hard samples. This leads to incomplete leaf edge segmentation and difficulties in disease spot extraction. These hard-to-distinguish pixels all impact disease spot extraction.

To address the above issues, this study makes improvements from two aspects: loss function and model structure. By adopting an adaptive loss function, issues like low model efficiency due to the sum of the losses of a large number of easy samples during training exceeding the total loss are ameliorated. This can to some extent solve problems such as pixel imbalance and low precision due to hard samples. By carrying out the tasks of leaf and disease spot segmentation in stages, interference from pixels in complex backgrounds is reduced. The first stage uses the Leaf-ASPP network model to segment leaf contours, and the second stage uses the Spot-ASPP network structure to extract disease spot areas.

3.1 Adaptive Loss Function

The introduction of an adaptive loss function mainly aims to resolve issues such as pixel ratio imbalance and an excessive number of hard samples in the task of segmenting diseased cucumber leaves, problems that traditional Cross-Entropy (CE) loss cannot solve. While balanced loss can effectively alleviate class sample imbalance, it overlooks the issue caused by an excessive number of hard samples. Therefore, by improving CE loss and balanced loss, an adaptive loss is generated and its optimization effect is discussed.

As a classic loss function in semantic segmentation in image processing, the Cross-Entropy binary classification loss function is defined as shown in the following equation (1):

$$CE(y,p)=\begin{cases}-log(p) & if\ y=1\\ -log(1-p) & if\ y=0\end{cases} \qquad (1)$$

Where $y\in\{0,1\}$ represents whether the pixel value is true or false, $p\in[0,1]$ represents the probability that the model predicts that this pixel belongs to the class y=1. Specifically, in the context of this paper, during image segmentation, it is determined whether the pixel value belongs to the foreground pixel, otherwise it is a background pixel. In the first stage, y=1 indicates that the pixel belongs to the target leaf area。

For easily classified pixels, such as those with probability values far greater than 0.5, the CE loss function generates a very small loss value. However, due to the vast number of pixels, for example, when the number of easily classified pixels in the data greatly exceeds the loss of hard-to-distinguish pixels, it produces overwhelming results, leading to insufficient training and poor network performance. Therefore, their loss cannot be ignored.

$\alpha-balanced$ CE loss is a common method to solve the problem of class imbalance. By introducing a balancing factor α to the CE loss function, the following equation (2) is formed:

$$\alpha-balancedCE(y,p)=\begin{cases}-\alpha log(p) & if\ y=1\\ -(1-\alpha)log(1-p) & if\ y=0\end{cases} \qquad (2)$$

In practical operation, by setting cross-validation，$\alpha-balanced$ CE loss can increase the weight of smaller categories. Although it assigns weights to samples belonging to the same category, it can reduce the impact of a large proportion of data on the loss []. However, the problem of hard samples existing in the data must be considered, such as how to effectively partition off pixels in areas of leaves covered by shadows, raindrops, dust, or interference pixels overlapping with other leaves in the background, to exclude interference. CE loss, however,$\alpha-balanced$ cannot effectively solve the problem of hard samples. Therefore, a new type of adaptive loss function will be tried to solve the above problems.

The question of whether the model can actively focus on hard-to-classify pixels during training, without the need for human intervention in setting weights, becomes a key link in disease spot segmentation. A modulation factor $[cos(p+\pi/2)+1]$ is introduced into the CE loss function. This term will decay as the pixel classification confidence increases, thereby changing the loss weight of hard-to-classify pixels and sparse category pixels in the overall loss. The adaptive loss function (3) is as follows:

$$\alpha - balancedCE(y,p) = \begin{cases} -[sin(p+\pi)+1]log(p) & if\ y=1 \\ -[sin(1-p+\pi)+1]log(1-p) & if\ y=0 \end{cases} \quad (3)$$

Here, p represents the probability value that the model predicts the pixel belongs to the y=1 class, and the value of the modulation factor is $[sin(p+\pi)+1]$ determined by the probability P. It decays as the probability value p increases, thereby reducing the loss value of easily classified pixels. Equation (3) includes the following content:

（1）When the probability p decays, it indicates that the pixel value is hard to classify, and the size of the modulation factor increases as the probability p decays. When the probability p is 0, the modulation factor is 1. The loss is infinitesimal $[sin(p+\pi)+1]$ and does not affect the overall loss.

（2）When the probability value p increases, indicating that the pixel is easy to classify, the modulation factor decreases as the probability p rises, thus the loss value of easily classified pixels will decrease. When the probability p rises to 1, the modulation factor reaches its minimum value. The pixel loss value will be reduced to a minimum.

The modulation factor can dynamically adjust the weight size according to the difficulty level of the probability values belonging to different categories, thereby adaptively adjusting the loss value. This process reduces the impact of the total loss of easily classified pixels on model performance. The adaptive loss function can reflect dynamic attention to pixels of two categories of different difficulty levels. To a certain extent, it can alleviate the problem of an excessive number of hard samples in leaf segmentation. It can adaptively assign gradually decreasing weight values to easily classified pixels in the background area, improving segmentation accuracy and enhancing network model performance. At the same time, it effectively mitigates the imbalance problem of the pixel ratio in disease spot segmentation. Overall, the adaptive loss function can effectively improve the network model's disease spot segmentation performance.

3.2 Two-Stage LS-ASPP Network Model

As mentioned in Section 1.1, most leaves in complex environments overlap each other, and the background clutter and irrelevant leaves overlap with the target segmentation leaves, affecting the segmentation effect. In addition, there may be diseased leaves in the background image, and areas similar to disease spots can also interfere with the target leaf segmentation. Therefore, a single-stage segmentation task may result in an incomplete disease spot segmentation area, and the low segmentation accuracy can lead to inaccurate disease detection. Therefore, the segmentation task is refined into two stages, from obtaining the disease leaf outline to extracting the disease spot area, optimizing the segmentation process, and improving segmentation precision. This study uses the ASPP as the benchmark network structure, designing a two-stage segmentation model for cucumber disease leaf and disease spot segmentation.

The two-stage LS-ASPP segmentation network model consists of Leaf-ASPP and Spot-ASPP. Both stages use the Atrous Spatial Pyramid Pooling (ASPP) as the benchmark network structure. The proposed model's first-stage network structure uses Leaf-ASPP to extract target leaves from complex scenes. Then, in the second stage, Spot-ASPP is used to segment the more complete disease spot area in the segmented disease leaf. Each stage focuses only on one specific type, reducing the difficulty of segmentation. The framework structure of the proposed method is shown in the figure below:

3.2.1 Leaf-ASPP

In real scenarios, the image background often contains overlapping leaves, which makes it difficult to accurately extract the contours of target leaves, as well as other leaves. More so, uneven illumination, raindrops, and dust can also directly affect segmentation. Therefore, to address these issues and enhance the capability of capturing cucumber disease leaf outlines, we have improved upon the original Atrous Spatial Pyramid Pooling (ASPP), rebranding the optimized network as the Leaf-ASPP network. The main structural optimizations include the replacement of the ASPP module with the Mult Residual ASPP module, enhancing the model's ability to perceive disease leaf outlines in complex backgrounds. The detailed Leaf-ASPP network model is composed of encoder and decoder parts, and its architecture (Fig.6) is shown below:

To improve the disease leaf segmentation performance of the model in complex scenes, we introduced the Mult Residual ASPP module[111], also known as the MRA-Net network, to capture more different multi-scale feature leaf outlines.

Generally, the larger the receptive field, the better the network's ability to perceive and judge each pixel. However, due to the characteristics of large neural network models, the number of network layers that increase sequentially, and the frequent use of up-sampling and down-sampling modules to process features, can lead to loss of detail information and reduced segmentation accuracy.

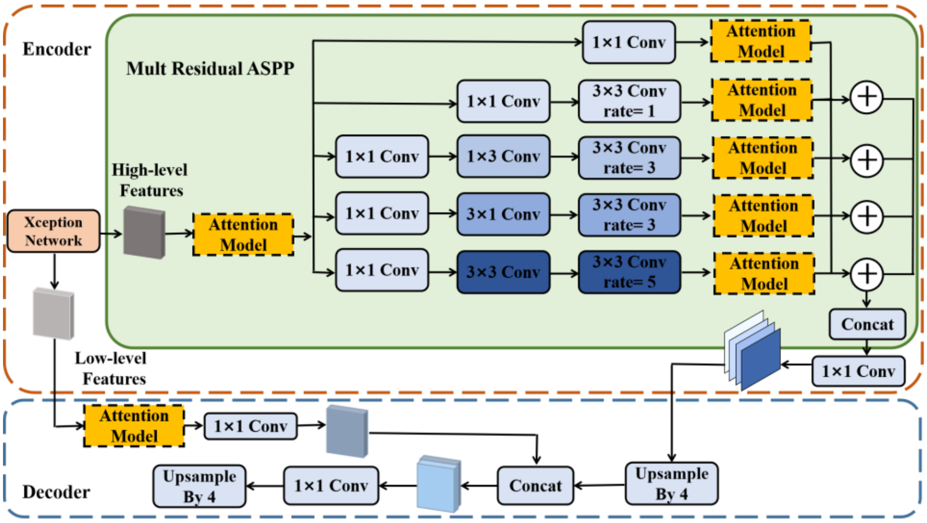


Fig.6 Architecture of MRA-Net

Both the Mult Residual ASPP and ASPP modules use dilated convolutions to enlarge the receptive field to obtain different scale feature maps. However, the original ASPP model mainly consists of three parallel dilated convolutions applied to a feature map, with the basic kernel size being 3×3. As such, in the initial model, the features extracted by the convolution kernel are similar and cannot distinguish difficult pixel features, which leads to an inability to accurately capture disease leaf outlines. Therefore, we have improved upon the original network, and the embedding of the MRA module enhances the model's edge extraction ability.

As shown in Fig.6, each branch of the MRA module consists of ordinary convolution, dilated convolution, and attention modules. The difference from the original spatial pyramid pooling structure lies in the different kernel sizes between the branches of the MRA module. In ordinary convolution, different kernel sizes will capture different receptive fields for each branch. Each branch's basic feature map will effectively capture different information and improve feature distinguishability. Finally, the outputs of each branch are fused to form multi-scale features.

In the encoder, two features are output: low-level features and high-level features. Low-level features are extracted by the Xception backbone network, mainly containing shallow information such as disease spot outlines and shapes. High-level features are processed by the backbone network and residual ASPP, mainly containing deep information such as texture and color features.

The Residual ASPP inputs the original features into three 1 × 1 convolution modules, four extended attention convolution units, and one residual unit. Each extended attention convolution unit consists of an ordinary attribute convolution module, a 3 × 3 convolution module, and an attention module, while the residual unit is composed of a 1 × 1 convolution module and an attention module. Among them, the dilation rates of the four extended convolution attentions are 1, 3, 3, and 5, with a kernel size of 3 × 3. Then, the outputs of each extended attention convolution unit are added to the output of the residual unit to get the four output feature maps of the Residual ASPP. Finally, the four feature maps are concatenated, and the merged result is input into a 1 × 1 convolution module. Through the above operations, the high-level features are finally obtained.

In the decoder, the outputs of low-level and optimized high-level features from the encoder are received. First, the low-level features are input into the attention module and the 1 × 1 convolution layer, yielding a small-scale refined low-level feature map. Then, the up-sampled high-level features are concatenated with the shallow features to obtain a fused feature map. Finally, the fused feature map is input into a 3 × 3 convolution layer for up-sampling processing to get the network's prediction base map.

By improving upon the original ASPP module, the MRA module's extraction of multi-scale features will reduce more irrelevant information, enhancing the model's ability to perceive disease leaf edge pixels. This will significantly improve the original network's segmentation performance.

3.2.2 Spot-ASPP

In the first stage of image segmentation, the information of disease leaf contours has been obtained. The disease leaf image contains only a small amount of sparse disease spot features, and the spot pixels account for a small proportion of the total disease leaf area pixels. This increases the difficulty of disease spot extraction in the second stage, resulting in a lower accuracy of disease spot segmentation. Therefore, the original network's spatial pyramid pooling is optimized again to enhance the model's segmentation performance. The main improvements are: (1) Adjusting the dilation rate in the ASPP module to reduce the loss of detail information. (2) The Convolutional Block Attention Module (CBAM) is introduced to highlight important information features again, capture small area pixels such as disease spots, and suppress irrelevant other disease leaf information to improve disease spot segmentation accuracy. The improved network structure is named Spot-ASPP, and its framework is shown in the figure:

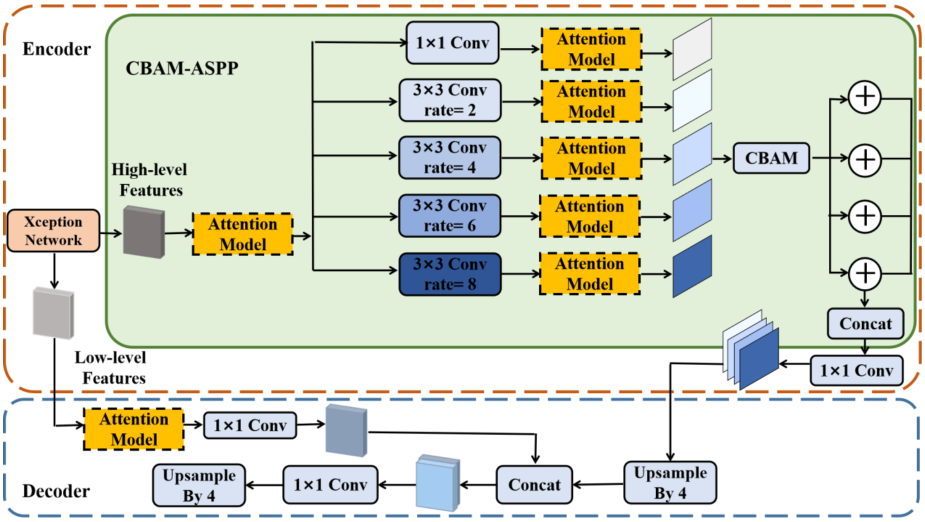


Fig.7 Architecture of CBAM-Net

Firstly, to enhance the segmentation effect of disease spots, smaller size dilated convolutions in the original network are retained, such as those with dilation rates of 2, 4, 6, and 8. The improved network structure is referred to as the CBAM-Net network. The receptive field range is mainly expanded by increasing the dilation rate, but due to the reduced correlation of adjacent local information in the feature map, small target area details will be lost directly. Therefore, the CBAM-Net structure retains smaller dilation rate dilated convolutions, which are more conducive to small disease spot pixel extraction, to achieve a more precise segmentation accuracy.

Secondly, to enhance the robustness of the model's segmentation performance, the Convolutional Block Attention Module (CBAM) is introduced following the optimization of the original ASPP network. In the channel attention module, the feature maps are input into the max pooling layer (Maxpool) and the average pooling layer (Avgpool), generating feature maps that are passed to a Multilayer Perceptron (MLP), thus creating the channel attention map.

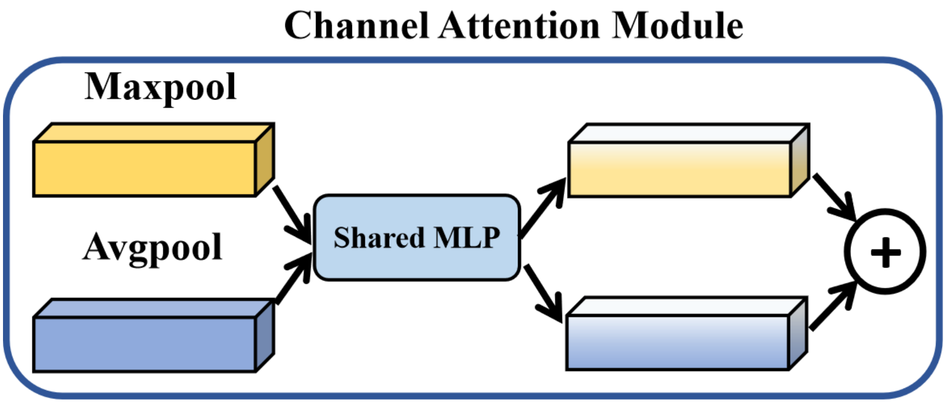


Fig.8 Architecture of Channel Attention Module

The channel attention module uses Avgpool and Maxpool modules to integrate the channel information of the feature maps, outputting two types of spatial information contexts processed by max pooling and average pooling respectively. Following a matrix summation operation, the fused matrix map is multiplied by the input features. This operation will effectively enhance the extraction capability for important features and strengthen the expressive power of the features.

2.3 Model Training

2.3.1 Experiment Configuration

To validate the effectiveness of the U-shaped network model of the LS-ASPP network, the proposed method or model is applied to the task of cucumber leaf disease spot image segmentation, and is compared with other methods or models. The network model training and testing environment are both the Ubuntu 18.04 LTS 64-bit operating system. The proposed method is designed using the Python programming language, with Python version 3.7, and the experiment platform uses PyTorch 1.10.2 as the deep learning open-source framework. The experimental hardware platform environment includes an Intel(R) Core(TM) i9-10900F CPU @ 2.80GHz processor, 32GB of memory, and an NVIDIA GeForce RTX 3080Ti with 26G of video memory. CUDA_11.6.0 and CUDNN_10.2 are used as the library tools for network model training acceleration.

2.3.2 Model Parameters

The U-shaped network convolutional layer of the LS-ASPP network refers to the Unet network model pre-training parameters for initialization, which has now been adopted by PyTorch as the default parameter initialization function. The negative slope of the activation function is 0. Kaiming initialization is mainly designed for deep neural networks using nonlinear activation, which can effectively prevent the explosion or disappearance of activation layer outputs in the forward propagation of deep neural networks, thus accelerating model convergence. The model learning rate is 0.0001, the number of training epochs is 15, the total number of iterations is 360, and the batch size for disease spot segmentation training is 4. The optimizer is Adam[17], with a weight decay of 0.00005.

3. Experimental Results and Analysis

This study selects at least three network models for comparison. Some disease spots are quite similar, and the same type of disease spots show different features at different disease stages, and the feature extraction network pays excessive attention to disease features.

Not using pre-training parameters and having a small proportion of training samples can lead to poor model recognition performance. After preliminary leaf position shifts, due to changes in their background shapes, some required disease spot area features are cropped, resulting in a decline in model recognition and segmentation performance.

The Vit network model pays more attention to the disease spot area. The Unet network model is sensitive to the orientation and location of its disease spots. Changes in the relative position of the background and the disease spot area can lead to poor recognition performance. Global pooling can enhance the relationship between feature maps and categories, showing invariance to spatial changes, and performs well in shift tests.

3.1 Analysis of Disease Segmentation Results

3.2 Segmentation Result Metrics

In order to accurately assess the segmentation precision of the diseased spot area, this study typically employs four conventional evaluation metrics in semantic segmentation: Pixel Accuracy (PA), Mean Pixel Accuracy (MPA), Mean Intersection over Union (MIoU), and Frequency Weighted Intersection over Union (FWIoU).

Pixel Accuracy refers to the ratio of the number of correctly predicted pixels to the total number of pixels, as shown in equation (7):

$$R_{PA} = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij}} \qquad (7)$$

m represents the number of categories， q_{ii} represents the number of correctly predicted positive pixel samples by the model.

$$R_{MPA} = \frac{1}{\mathrm{k}+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \qquad (8)$$

Mean Pixel Accuracy is the proportion of each category of pixels correctly classified, averaged over categories.

$$R_{MIoU} = \frac{1}{\mathrm{k}+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \qquad (9)$$

3.3 Disease Spot Segmentation Precision

The experiment used 84 images with disease spot annotations, processed by data layering, as training samples, and 36 images as test samples. The annotation only divided the images into foreground and disease spot areas, without considering disease category. The experimental results were based on the average of 225 image test results. Table 31 presents a comparison of the segmentation precision for each algorithm. As can be seen from Table 1, the segmentation precision of LS-ASPP is significantly improved compared to FCN, U-Net, and VIT. The segmentation precision of the Unet model is not significantly different from that of other networks trained with FCN, indicating that this method can achieve good segmentation results even when using very few training samples. The structures of U-Net and VIT are superior to the FCN model, because FCN only uses a single feature map from the output of the last three pooling layers during upsampling, without merging low-level semantic features, which leads to poor segmentation results. On the other hand, network models that merge low-level semantic features have optimized the disease spot segmentation effect.

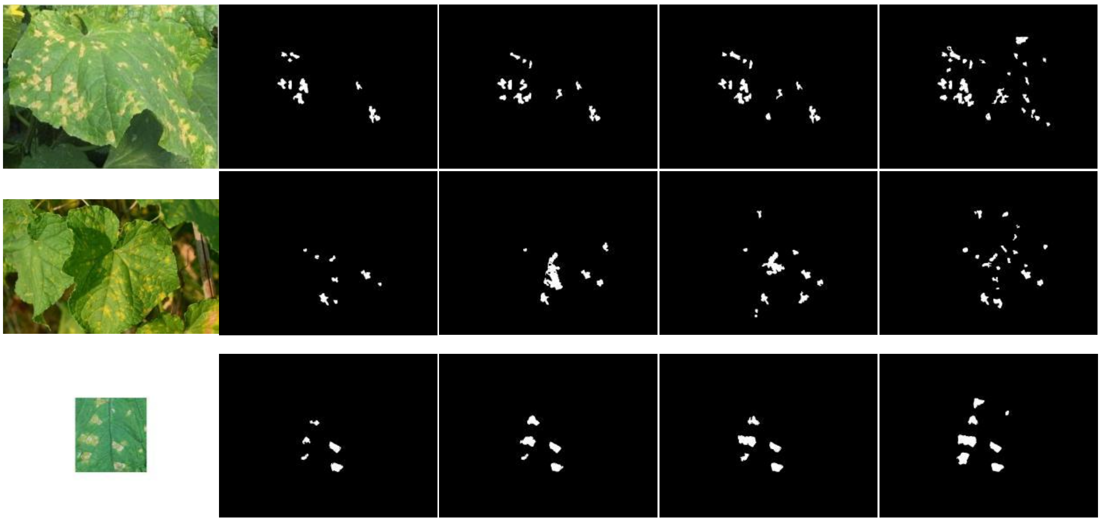


Fig.7 Comparison of segmentation results of various deep learning network models

Fig. 7 shows the segmentation results of cucumber downy mildew leaves by FCN, U-Net, VIT, and the LS-ASPP model with self-attention mechanism used in this experiment. It can be seen that adding skip-connection modules can fully merge low-level semantic feature information. U-Net, VIT, and LS-ASPP models can segment smaller disease spot areas, while the FCN model will lose some disease spot information, with incomplete and relatively vague segmentation boundaries and low acquisition capability for small disease spots at a distance. The use of the U-Net model can lead to uneven edges in the disease spot area, mainly because U-Net only has a convolutional layer in the upsampling process and does not capture all global disease spot information.

Because this experiment uses the LS-ASPP network structure with attention mechanism modules, this structure has the ability to capture all information, accurately capturing global information at a distance, thus more accurately capturing global disease spot semantic information. Therefore, it has stronger segmentation capability for less prominent small disease spot areas and can more accurately capture global leaf spot information.

Table 1 showcases the segmentation results of three diseases on three models: FCN, U-Net, and Vit. It is evident that the addition of skip connections and the fusion of low-level semantic features enable the U-Net and Vit models to segment smaller diseased areas, while the FCN structure tends to lose some details in its segmentation results. The segmentation boundaries are blurry and closely situated diseased areas tend to stick together. The LS-ASPP model produces smoother disease boundaries than the U-Net, whose segmentation edges appear jagged. This is because the U-Net only introduces a convolutional layer after upsampling, while the LS-ASPP, with its convolution operation in the reconstruction layer and attention mechanism, can further smooth edges and gather global segmentation information.

Table 1. Segmentation Accuracy of Each Algorithm

Method	Pixel Accuracy/%	Mean Pixel Accuracy/%	Mean Intersection over Union/%	Frequency Weighted	Run time on

				Intersection over Union/%	Graphic Processing Unit (GPU)/ms
FCN	92.56	81.51	71.11	87.45	22
U-Net	93.41	81.92	72.29	88.49	13
Vit	94.66	83.28	75.68	90.34	12
LS-ASPP	95.67	84.23	76.35	91.45	12

Different diseases present different levels of segmentation difficulty. For instance, the diseased areas of cucumber powdery mildew are light yellow, with early symptoms and indistinct edges, making segmentation challenging. The average segmentation results of 84 test samples for each type of disease spot indicate that the LS-ASPP model exhibits high segmentation accuracy for diseases of the spot type, as shown in Table 2.

The model's input size influences segmentation results: when the input size is increased from 224 × 224 to 384 × 384, while keeping the patch size at 4, the LS-ASPP network's input token sequence becomes larger, thereby improving the model's segmentation performance. Despite a slight improvement in segmentation accuracy, the computational burden of the entire network also increases significantly. To ensure algorithm efficiency, this study uses an input resolution scale of 224 × 224.

The model's scale also influences results: akin to reference [19], we believe deepening the network will impact the model's performance. An increase in model scale does not enhance performance but instead raises the computational cost of the entire network. Balancing precision and speed, we opt for an attention mechanism-based symmetric model for plant image disease segmentation.

Table 2. Segmentation Accuracy for Diseases

Disease classes	Pixel Accuracy/%	Mean Pixel Accuracy/%	Mean Intersection over Union/%	Frequency Weighted Intersection over Union/%
Cucumber Powdery Mildew	93.52	86.24	78.65	88.71
Cucumber Angular Leaf Spot	93.25	85.68	77.94	88.23

Cucumber Downy Mildew	93.78	83.72	74.56	91.37

Conclusion

Plant disease segmentation models are prone to interference from shadows and obstructions, and the extraction of features has an inherent uncertainty. To address these challenges, the LS-ASPP network model was constructed using images from the dataset. The use of the LS-ASPP module in the network model enhances the model's ability to capture global information, thereby improving the segmentation of disease spots.

The model is trained with only a small amount of annotated disease spot samples, significantly reducing the annotation cost. Compared to the U-Net and LS-ASPP models, this model achieves superior segmentation accuracy. It performs well in terms of pixel accuracy, mean intersection over union, and frequency weighted intersection over union, all of which are key segmentation evaluation metrics. This suggests that the model exhibits robustness against shadows and obstructions.

By adding skip connections, the network model can integrate low-level features, and by restoring detailed features, it can retain smaller disease spots and refine segmentation boundaries. The trial demonstrates that the LS-ASPP model, equipped with a self-attention mechanism, exhibits good generalization performance and robustness.

Reference

[1] Liu C, Zhu H, Guo W, et al. EFDet: An efficient detection method for cucumber disease under natural complex environments[J]. Computers and Electronics in Agriculture, 2021, 189: 106378.

[2] Zhang P, Yang L, Li D. EfficientNet-B4-Ranger: A novel method for greenhouse cucumber disease recognition under natural complex environment[J]. Computers and Electronics in Agriculture, 2020, 176: 105652.

[3] Bai X, Li X, Fu Z, et al. A fuzzy clustering segmentation method based on neighborhood grayscale information for defining cucumber leaf spot disease images[J]. Computers and Electronics in Agriculture, 2017, 136: 157-165.

[4] Pixia D, Xiangdong W. Recognition of greenhouse cucumber disease based on image processing technology[J]. Open Journal of Applied Sciences, 2013, 3(01): 27-31.

[5] Luo, Y., Sun, J., Shen, J., Wu, X., Wang, L., Zhu, W., 2021. Apple leaf disease recognition and sub-class categorization based on improved multi-scale feature fusion network. IEEE Access 9, 95517–95527. http://dx.doi.org/10.1109/ACCESS. 2021.3094802.

[6] Jiang, P., Chen, Y., Liu, B., He, D., Liang, C., 2019. Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. IEEE Access 7, 59069–59080.

[7] Mu, H., Wang, K., Yang, X., Xu, W., Liu, X., Ritsema, C.J., Geissen, V., 2022. Pesticide usage practices and the exposure risk to pollinators: A case study in the north China plain. Ecotoxicol. Environ. Saf. 241, 113713.

[8] Pan, D., He, M., Kong, F., 2020. Risk attitude, risk perception, and farmers' pesticide application behavior in China: A moderation and mediation model. J. Clean. Prod. 276, 124241.

[9] Yang P, Song W, Zhao X, et al. An improved Otsu threshold segmentation algorithm[J]. International Journal of Computational Science and Engineering, 2020, 22(1): 146-153.

[10] Pare S, Kumar A, Singh G K, et al. Image segmentation using multilevel thresholding: a research review[J]. Iranian Journal of Science and Technology, Transactions of Electrical Engineering, 2020, 44: 1-29.

[11] Rajabi A, Eskandari M, Ghadi M J, et al. A comparative study of clustering techniques for electrical load pattern segmentation[J]. Renewable and Sustainable Energy Reviews, 2020, 120: 109628.

[12] Chavolla E, Zaldivar D, Cuevas E, et al. Color spaces advantages and disadvantages in image color clustering segmentation[J]. Advances in soft computing and machine learning in image processing, 2018: 3-22.

[13] Abbas Q, Celebi M E, García I F. Breast mass segmentation using region-based and edge-based methods in a 4-stage multiscale system[J]. Biomedical Signal Processing and Control, 2013, 8(2): 204-214.

[14] Ji X, Li Y, Cheng J, et al. Cell image segmentation based on an improved watershed algorithm[C]//2015 8th International Congress on Image and Signal Processing (CISP). IEEE, 2015: 433-437.

[15] Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440.

[16] Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

[17] Liu C, Chen L C, Schroff F, et al. Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 82-92.

[18] Czajkowska J, Badura P, Korzekwa S, et al. Automated segmentation of epidermis in high-frequency ultrasound of pathological skin using a cascade of DeepLab v3+ networks and fuzzy connectedness[J]. Computerized Medical Imaging and Graphics, 2022, 95: 102023.

[19] Yan J, Yan T, Ye W, et al. Cotton leaf segmentation with composite backbone architecture combining convolution and attention[J]. Frontiers in Plant Science, 2023, 14.

[20] Storey G, Meng Q, Li B. Leaf disease segmentation and detection in apple orchards for precise smart spraying in sustainable agriculture[J]. Sustainability, 2022, 14(3): 1458.

FUNDING

Natural Science Foundation of Xinjiang

Uygur Autonomous Region (No.2022D01C461);This work was supported by Sponsored by Xinjiang University of Science and Technology Campus level Research Project on Cucumber Leaf Disease Identification and Spot Segmentation (No. ZQ202201); Natural Science Foundation of Xinjiang

Uygur Autonomous Region (No. 2022D01C349)

ABOUT THE AUTHOR

Jie Yang was born in Dali, Yunnan China, in 1992. She received the master's degree from University of South China.Now, she works in School of Information Engineering, Xinjiang Institute of Technology. Her research interest include Artificial intelligence and image processing.

E-mail: yangjie.xit.cs@gmail.com

Jiya Tian was born in Heze, Shandong.P.R.China, in 1979. He received the master's degree from Jilin University, P.R. China. Now, he works in School of Information Engineering, Xinjiang Institute of Technology. His research interest includes Artificial intelligence and image processing.

E-mail: 42683387@qq.com

Jinchao Miao was born in Xi'an, Shaanxi China, in 1987. She received the master's degree from Xi’an University of Posts & Telecommunications. Her research interest include image processing.

E-mail:miaojinchao@sina.cn

Chen Yunsheng was born in Zhumadian City, Henan Province. Born in 1995.He holds a master's degree from Tarim University. I am currently working at the School of Information Engineering, Xinjiang University of Technology.My main research direction is agricultural informatization.

E-mail: cys13239861903@163.com

Shuping ZHANG graduated from Xinjiang University in 2006 with a Master's degree in Computer Science and Technology.She has presided and participate over 3 provincial-level projects. Published 7 papers. Her research interests include computer vision and deep learning.

E-mail: 442658545@qq.com

Wu Yixuan was born in Shaanxi, China in 2004. Majored in Computer Science and Technology at Xinjiang Institute of Technology.His research interests include computer vision and deep learning.

E-mail: 2992075809@qq.com