



Semantic Segmentation of Rice Disease Images based on DeepLabV3+

Jinghao Li^{1,*}, Zirui Ren² and Letian Zhou³

¹ School of Economics and Management, Tiangong University, 300387, No.399 BinShuiXi Road, XiQing District, Tianjin, China

² College of information Science and Engineering, Hunan University, 410082, Lushan South Road, Yuelu District, Changsha, China

³ School of Computer Science and Technology, Qingdao University, 266075, 308 Ningxia Road, Qingdao, China

*Corresponding author: jinghal@bgsu.edu

Abstract. Disease area segmentation is an important task in the field of smart agriculture, which is of great significance for analyzing the fine-grained information inside disease spots and supporting prevention and control decisions. Early disease area segmentation mostly relied on image processing or manual features, and its accuracy could not meet the practical application requirements in field scenarios. Thanks to the rapid development of pattern recognition technology, semantic segmentation algorithms based on deep learning provide new solutions for accurately and automatically identifying diseased areas. In this paper, we present a semantic segmentation approach for rice leaf disease images using DeepLabV3+. Specifically, we combine the encoder-decoder structure with atrous convolution as well as the spatial pyramid pooling to further improve segmentation accuracy. We constructed a dataset of rice leaf images containing four different types of diseases, and trained and tested models on this dataset. The model performance is evaluated with standard metrics such as mean intersection over union (mIoU) and pixel accuracy. In addition, we design some other sets of corresponding experiments to test the performance in some specific circumstances, including in poor light conditions, on the background of different situations, with low resolution and with noises. All outcomes demonstrate the efficacy and reliability of our methodology. We also discuss the challenges and limitations of the model, as well as possible future directions for improvement.

Keywords: DeepLabV3+, semantic segmentation, rice leaf disease, ASPP, intelligent agriculture.

1 Introduction

Rice leaf disease segmentation is a process of separating the disease regions from the healthy regions in the images of rice leaves, which can help to identify the type and severity of rice leaf disease. Rice leaf disease segmentation plays an important role in disease diagnosis and prevention. First, it is beneficial to reduce the labor intensity

and time consumption of manual disease assessment, which is often subjective and unreliable. Second, rice leaf disease segmentation can help to improve the accuracy and robustness of disease recognition by using image processing and machine learning techniques. Third, rice leaf disease segmentation can help to assess the disease severity and provide early warning for crop protection, so that appropriate control measures can be applied timely, which cost savings as well. To this end, accurate segmenting rice leaf disease attracts more and more research attentions in recent years.

In the field of rice leaf disease recognition, various methods have been proposed to detect and classify different types of diseases using image processing and machine learning techniques. Shrivastava et al. extracted color and texture features from the images of rice leaves and classified different types of diseases using a support vector machine (SVM) [1]. Haque et al. performed annotations on 1500 collected datasets and presented a rice leaf disease classification and detection approach using YOLOv5, achieving a disease recognition accuracy of 94.65% and a detection speed of 50 FPS [2]. Wang et al. introduced an attention-based depth-wise separable neural network with Bayesian optimization (ADSNN-BO) for the detection and classification of rice diseases. They used an attention mechanism to enhance the feature representation and focus on the key regions of interest, which achieved a test accuracy of 94.65% and outperformed the existing models in the literature [3]. Though the aforementioned methods solve the challenging rice leaf disease recognition task, their performances are still limited due to the manual features.

As deep learning has advanced, more sophisticated approaches have emerged that utilize convolutional neural networks (CNNs) to automatically extract features from extensive annotated datasets, resulting in highly accurate disease recognition. The surveyed literature reveals several unresolved challenges [4]: To begin, the pixel representation of the leaf needs to be labeled in a plain and objective manner. The backdrop, the ill parts, and the leaf area should all be labeled. Second, there is a demand for accurate deep learning models that can be used for leaf and lesion segmentation. These segmentation models and highly accurate CNN classifiers will detect infection(s) and determine the severity of an infection. Thirdly, there are not enough datasets for training disease recognition models. As a result, high-performance architectures are desired even though they have this disadvantage. To solve those above problems, we applied DeepLabV3+, for its superb performance in rice leaf disease semantic segmentation compared to other methods. DeepLabV3+ is a state-of-the-art semantic segmentation framework that can be used to identify rice leaf disease by classifying each pixel of the leaf image into different categories. First, DeepLabV3+ can achieve high accuracy and robustness in disease recognition by using deep CNNs to automatically learn features from large-scale annotated data [5]. Second, DeepLabV3+ can accurately segment the disease regions and assess the disease severity by using atrous convolution and ASPP to handle objects of different sizes and shapes [5]. Third, DeepLabV3+ can reduce computational complexity and memory consumption by using a modified Xception network as the backbone and adding a decoder with skip connections to improve the segmentation quality [5].

In summary, our research makes contributions in two aspects. Firstly, we have collected and annotated a large-scale dataset of common rice diseases, which is publicly available. Secondly, we have successfully applied the fixed DeepLabV3+ model in the field of rice leaf disease segmentation.

2 Method

2.1 Model Introduction

The work presented in this paper is based on the DeepLabV3+ model. The DeepLabV3+ model, proposed by Google in 2018, is an improved version of the DeepLabV3 model [6]. It includes an efficient decoder module called Decoder, which aims to achieve more accurate segmentation boundaries. The model in Fig.1 adopts an encoder-decoder architecture [5], which enables it to quickly capture higher-level semantic information during the encoding stage and gradually restore the object boundaries during the decoding stage. This allows the model to maintain high-resolution output while having a larger receptive field. The DeepLabV3+ model also incorporates spatial pyramid pooling technology [5], which fuses multiscale information and achieves good performance as an encoder based on deep network architecture.

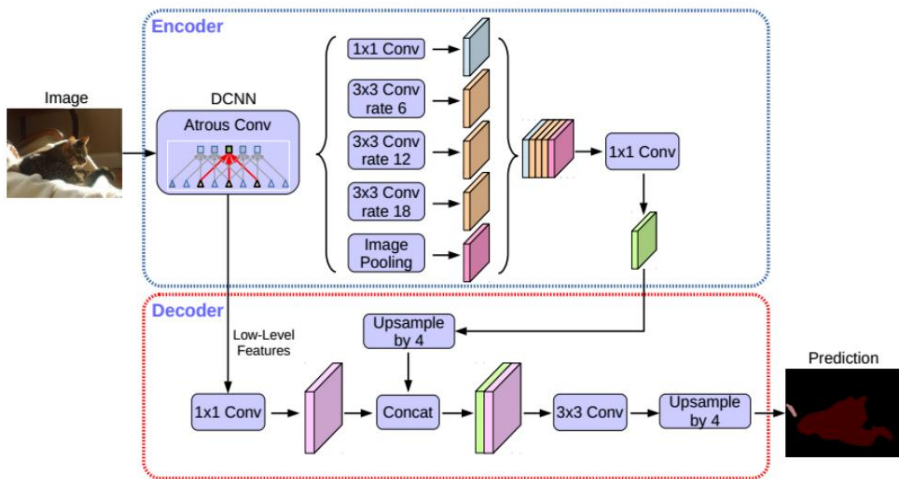


Fig. 1. Structure diagram of the DeepLabV3+ model [5].

2.2 Backbone

Xception (Extreme Inception) is employed as the backbone network for feature extraction [7]. Xception utilizes depthwise separable convolutional techniques, which reduce the number of parameters and improve the computational efficiency of the model.

Depth-wise Separable Convolution with Dilated Convolution. Depth-wise separable convolution consists of two main processes: depth-wise and point-wise [8]. Depth-wise refers to each convolutional kernel processing a single channel independently, without involving convolution between layers. The processed depth-wise feature maps are then stacked. Point-wise involves convolving the stacked feature maps with a 1×1 kernel, allowing for the extraction of inter-channel information. In DeepLabV3+, depth-wise separable convolution is enhanced with dilated convolutions, as shown in Fig.2 (C) [5]. Dilated convolutions expand the receptive field without sacrificing information, allowing the convolutional output to contain a broader range of data.

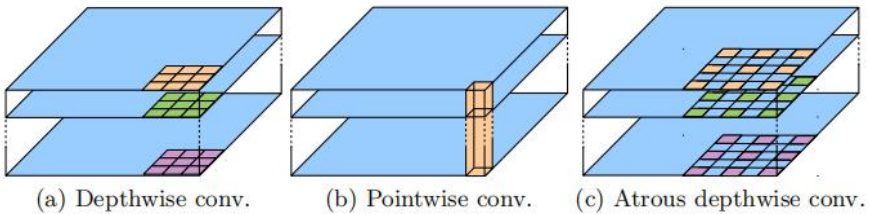


Fig. 2. (a) and (b) illustrate the two processes of depth-wise separable convolution, while (c) introduces the addition of the dilated convolution process based on (a) [5].

Xception. Inception V3 has a three-layer convolutional structure from Base to Concat [9]. To simplify the convolutional structure of Inception, starting from the Input, the first layer is a 1×1 convolution, and the second layer is a 3×3 convolution. Since a 1×1 convolution is applied after the Input, the 1×1 convolution module is extracted so that each subsequent 3×3 convolution shares the same feature maps produced by the 1×1 convolution. Furthermore, each 3×3 convolution processes different channels of the feature maps obtained after the 1×1 convolution, instead of processing the entire feature map. The channels do not overlap [7]. The extreme case of each 3×3 convolution kernel processing a single channel is the structure of the Xception network.

2.3 ASPP Structure

The ASPP structure is a key component used in DeepLabV3+ to enlarge the receptive field. The ASPP structure is designed for enhancing feature extraction. The entire process is based on multi-scale atrous convolutions, which generate feature maps with different sampling rates. These feature maps are then stacked and integrated using 1×1 convolutions, allowing the feature maps to have a larger receptive field [5]. Moreover, the ASPP structure can capture object information at different scales, thereby improving the accuracy of semantic segmentation [5].

2.4 Decoder Structure

The process of the Decoder module is as follows: First, the low-resolution feature maps extracted from the backbone network are processed with 1×1 convolutions. Then, the output from the ASPP module is upsampled by a factor of 4. These two

parts are then stacked together and passed through a combination of 3×3 convolutions and upsampling layers to obtain the semantic segmentation result [5]. The Decoder module is tasked with restoring the resolution of feature maps and producing segmentation results that match the dimensions of the input image. The upsampling is performed using bilinear interpolation. Additionally, the Decoder module introduces an additional 1×1 convolutional layer to reduce the dimensionality of low-level features. This helps prevent the degradation of high-level features obtained from the encoder and enhances the expressive power of the features [5].

2.5 Loss function

The loss function used in this study consists of two parts: Cross Entropy Loss and Dice Loss [10]. Cross Entropy Loss is a widely employed loss function in image segmentation, which examines individual pixels independently and compares the class predictions to the target vectors. Dice Loss utilizes the Dice coefficient as a loss function for semantic segmentation. The Dice coefficient is commonly employed for quantifying the similarity between two samples, with a scale ranging from 0 to 1. A higher value indicates a stronger overlap between the predicted and ground truth results, making a higher Dice coefficient preferable. However, since loss functions are generally minimized, Dice loss is defined as 1 minus the Dice coefficient. This allows the Dice loss to be used as a loss function for semantic segmentation.

3 Experimental Results

3.1 Datasets

The dataset we chose was called Rice Leaf Disease Image Samples [11], which includes more than 5000 pictures shot on-site covering four kinds of rice leaf disease including Rice Blast, Bacterial Blight, Brown Spot and Rice Tungro Spherical Virus Disease. All the images of this dataset have already been converted to the size of [300, 300], and each of them only has different numbers of disease spots for one known disease.

However, the images given were not labelled, and a large proportion of our workload is to get all images labelled using Labelme 5.2.0.post4. Finally, we picked a test set of 395 images and 4366 labelled images from the data set, dividing them into a train set of 3492 images and a validation set of 874 images (8:2).A Subsection Sample.

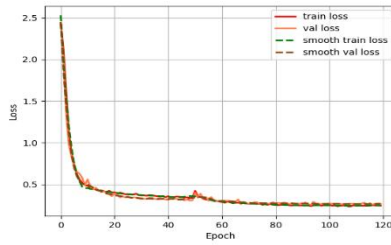
3.2 Training

Training configuration and platform. In the experiment, we made semantic segmentation on four common diseases of rice based on the DeepLabV3+ with the train set and the validation set, both of which have been resized into [512, 512]. We ran this model on the Google Colab platform with an NVIDIA A100-SXM 40G GPU and did a 150-epoch train with a breakpoint training of 35 epochs that continued from

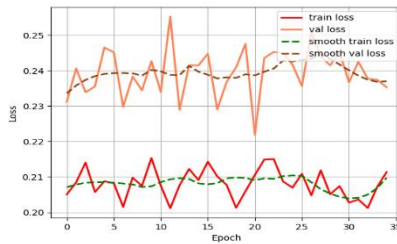
the 115th epoch. During the training, the model made a 50-epoch freeze train with a batch size of 16 and a 105-epoch unfreeze train with a batch size of 8, which balanced both stability and generalization ability. We chose an optimizer named "Adam" with a learning rate of "5e-4" and a "cosine" decay type of learning rate.

Training Process. Our training result based on DeepLabV3+ is relatively great with the performance listed as follows. As is shown in Fig.3, both the train loss and validation loss started from about 2.435 at the beginning, and as the training proceeded, they converge to about 0.24 at the breakpoint (115th epoch). Then, as we continued the training from the breakpoint, it is shown that the validation loss reached a dynamic stability at a mean value of 0.23948803414607484, and the training loss at a mean value of 0.20811066252015134. Another evaluation index like the best value of Mean Intersection over Union(mIoU) was 78.79%, f_score reached 0.889, Mean Pixel Accuracy (mPA) hit the level of 89.78%, and Accuracy achieved a level of 98.59%. We picked up the weight document with the least validation loss as our best weight file to move on to the process of the prediction.

(a)Convergence of Loss
(before breakpoint)



(b)Dynamic Stability of the Loss (after breakpoint)



(c)Change in MIoU(before breakpoint)

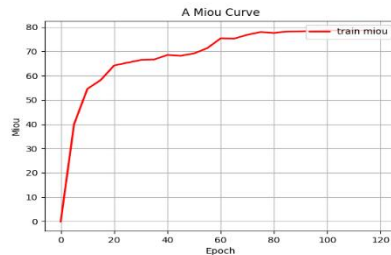


Fig. 3. Experimentation Performance.

3.3 Prediction

Original Test Set. To test the stability, robustness, and accuracy of our model, we set five followed different scenarios. In the first scenario, we test the model based on the original test set. As is shown in Fig.4, the edge of every segmentation was clear, although some of the disease spots were not segmented due to either too small or hidden in the shade of the foreground leaves, which was rare in the test set as a whole. The generalization ability and stability of the model have been well taken into account.

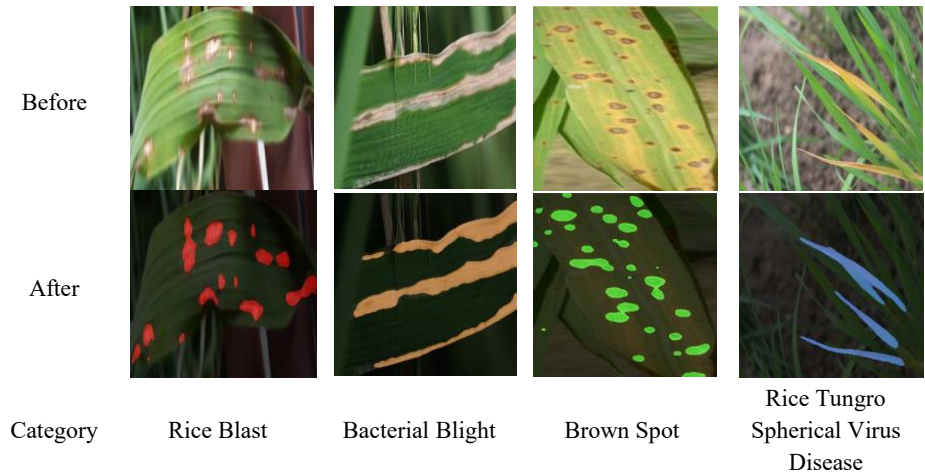


Fig. 4. Segmentation results on original test set.

Illumination Test. In the second scenario, we changed the brightness of the images in our test set into three different exposure levels to test in what way the overall exposure level can affect our model’s performance. To control the exposure level, we defined a variable called “exposure_factor” and multiplied it by the RGB of every pixel of each image. In this test, we found that the performance of the model was not so affected by the exposure level. But it was also noticed that the overall exposure level might cause the loss of detailed features of the disease spots and thus reduced the performance. However, if the brightness didn’t make that loss happen, the model could still segment the disease spots well as is shown in Fig.5.

Background Test. In the third scenario, we selected images with pure white background from other data sets, cut them into 300 pixels wide and 300 pixels high, and did a comparison to the original test set. The results were shown in the Fig.6. And through the analysis of the entire test set, we haven’t found enough evidence to prove that the background clearly affects the accuracy of the segmentation.

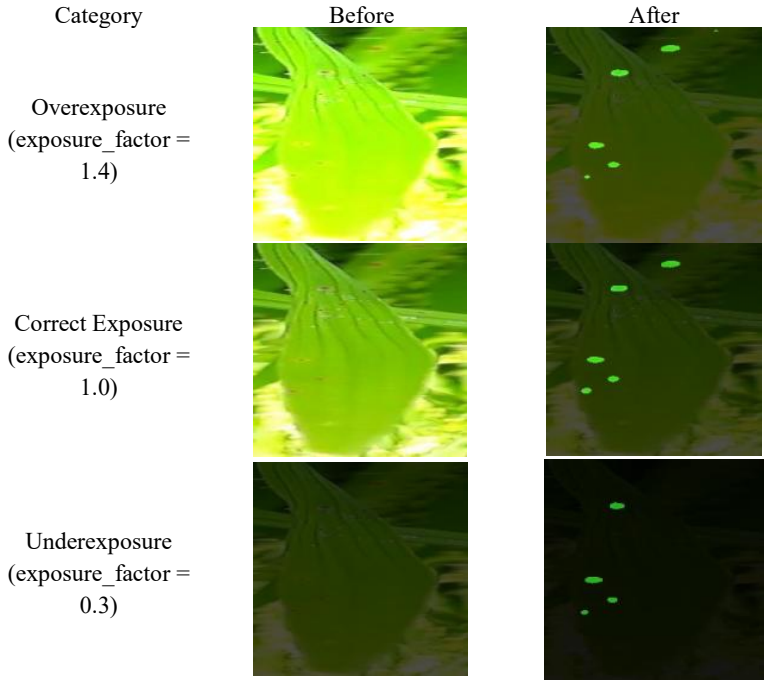


Fig. 5. Segmentation results for various illuminations.

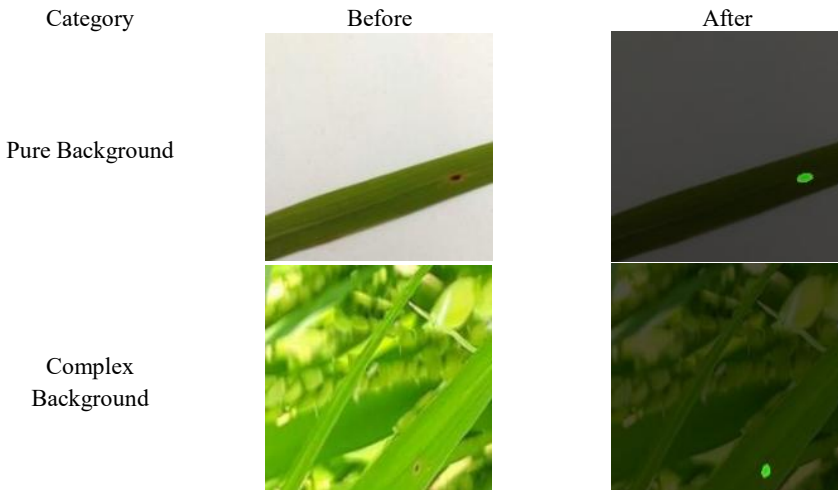


Fig. 6. Segmentation results under various backgrounds.

Test 4: Resolution Test

In the fourth scenario, we selected high-resolution images from another data set (all with resolutions above [5400, 3600]) which includes three different disease, namely Rice Blast, Bacterial Blight, and Brown Spot. The results were as follows in Fig.7. It can also be seen that the resolution of the images doesn't affect the accuracy and stability of our model, which implies that our model has relatively great robustness and generalization ability.

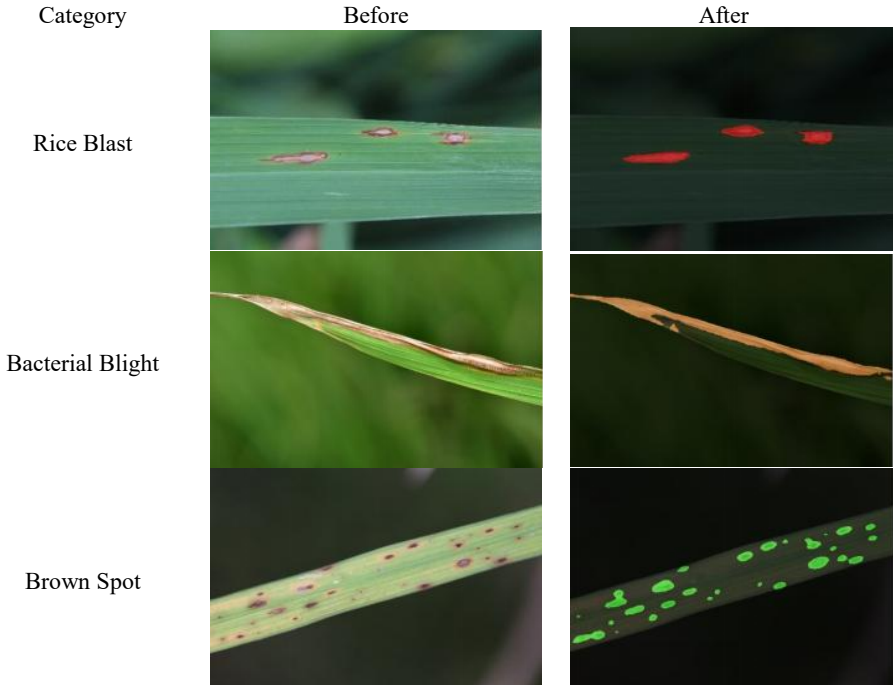


Fig. 7. Segmentation results for various high-resolution images.

Random Color Spots Test. In the fifth scenario, for the original test set, we added several random color blocks of one-pixel size to the original images in the size of [300, 300], to give our model a final test of stability, robustness, and accuracy. And the results are shown in Fig.8. We have gone through the whole test result and found that the accuracy of segmentation was affected by the partial occlusion of the disease spot. And it was true that there were cases where it couldn't be recognized and segmented, and there were cases of miscategorization, which implies the ultimate limitations of the robustness of our model.

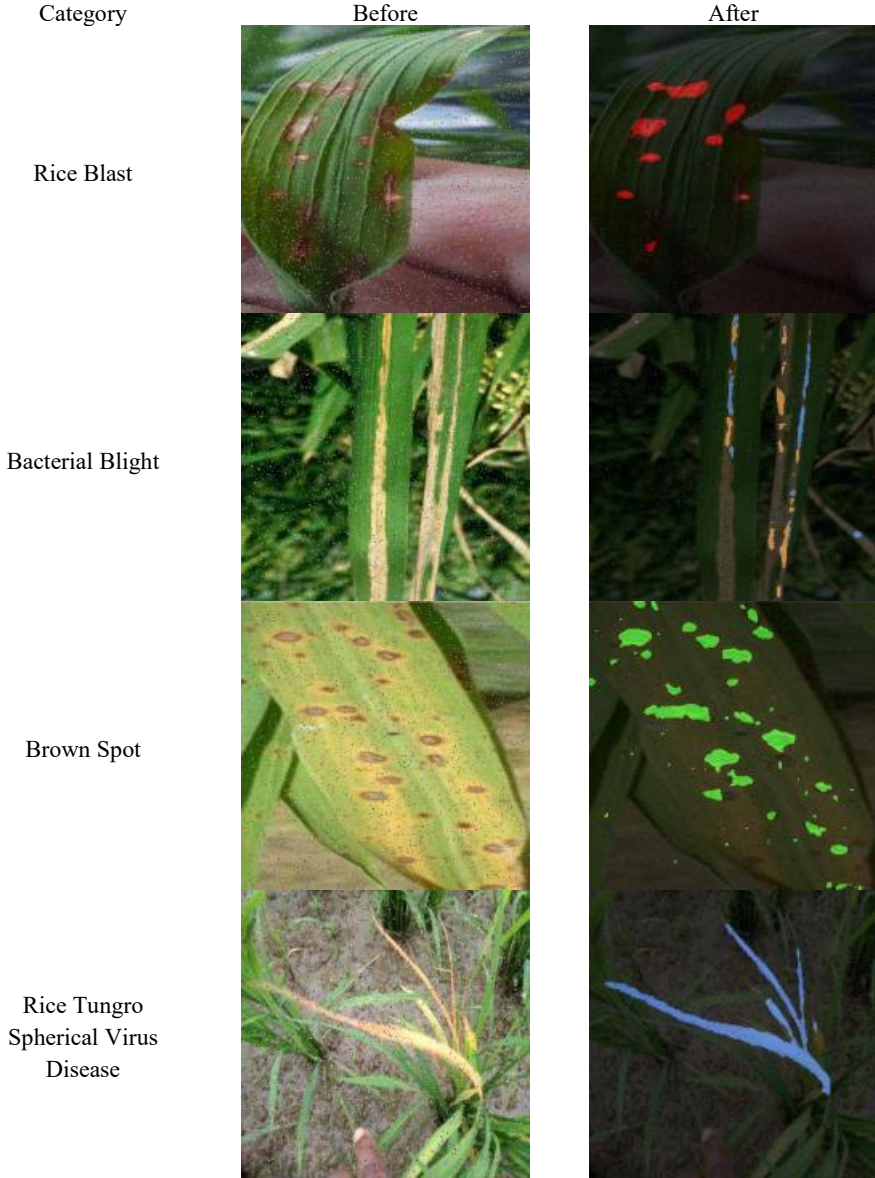


Fig. 8. Segmentation results for random color spots.

To sum up, we can clearly notice that for the original test set, the model performed great segmentation with high accuracy by Test 1. And it is found by Test 3 and Test 4 that both the background and the resolution of the images don't clearly affect the performance of our model and the accuracy of the result, which implies that it has great stability and generalization ability when dealing with images no matter they are with complex or pure background, or in low or high resolution. It is also found by

analyzing Test 2 and Test 5 is that the model is of good robustness, and even when the features of disease spots suffer loss to an extent due to the incorrect exposure or block, the model can well recognize the disease spot. All in all, through the five prediction cases above, we hold a firm belief that the stability, robustness, and accuracy of our model make it acceptable to be put into actual agricultural production activities to help farmers with little knowledge of rice disease recognize and deal with them.

4 Discussion

The task of rice leaf disease segmentation is accomplished in this paper using an advanced semantic segmentation model, DeepLabV3+. Firstly, the paper performs selection and annotation on a dataset of rice disease images, evaluates the performance of DeepLabV3+ on this dataset, and compares it with several baseline methods.

Nevertheless, there are certain limitations in this study that should be addressed in future research. Firstly, the dataset used in this paper only cover four types of diseases, which limits the diversity and representativeness of the data. There is a significant room for expansion in terms of the variety of diseases and the sampling approach. Secondly, the experimental results in this paper only demonstrate the segmentation performance on single-disease plants, and the segmentation accuracy on plants with multiple diseases remains to be tested. Thirdly, the modifications made to the network structure in this paper are relatively simple, and the model's performance could be further optimized.

On the other hand, this paper also demonstrates certain potential and value. Firstly, the dataset was carefully selected and annotated by the authors, ensuring the quality and reliability of the data. Secondly, the model proposed in this paper has been experimentally validated and achieved good segmentation results in complex natural environments and other datasets, showcasing its strong generalization ability. These work provides a solid foundation for further research on rice disease analysis.

Smart agriculture is an important field of artificial intelligence application. With the advancement of computer image processing technology, the authors anticipate that future solutions will achieve:

- (1) Higher accuracy and robustness
- (2) Real-time and automation
- (3) Application of multimodal data
- (4) Data sharing through networks

5 Conclusion

In this paper, we propose a DeepLabV3+ based method for semantic segmentation of rice leaf disease images. To improve feature quality and segmentation accuracy, we will combine the encoder-decoder structure with atrous convolution and spatial pyramid pooling. In addition, we also constructed a rice leaf image dataset containing

four different types of diseases. To verify the model accuracy and generalization ability, we tested the model results in low-light conditions, in the background of different situations, in low-resolution and noisy conditions. Our methodology has been shown to be both effective and reliable by all of the findings. We also discuss the challenges and limitations of the model, as well as possible directions for future improvements.

Acknowledgment

Jinghao Li is responsible for Experimentation and Result. Zirui Ren is responsible for Method, Discussion and References. Letian Zhou is responsible for Abstract, Keywords, Introduction and Conclusion.

All the authors contributed equally and their names were listed in alphabetical order.

References

1. Shrivastava, V.K., & Pradhan, M.K. (2020). Rice plant disease classification using color features: a machine learning paradigm. *Journal of Plant Pathology*, 103, 17 - 26.
2. Haque, M.E., Rahman, A., Junaid, I., Hoque, S.U., & Paul, M. (2022). Rice Leaf Disease Classification and Detection Using YOLOv5. *ArXiv*, abs/2209.01579.
3. Wang, Y., Wang, H., & Peng, Z. (2021). Rice diseases detection and classification using attention based neural network and bayesian optimization. *Expert Syst. Appl.*, 178, 114770.
4. Udayananda, G.K., Shyalika, C., & Kumara, P. (2022). Rice plant disease diagnosing using machine learning techniques: a comprehensive review. *SN Applied Sciences*, 4.
5. Chen, L., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *European Conference on Computer Vision*.
6. Chen, L., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking Atrous Convolution for Semantic Image Segmentation. *ArXiv*, abs/1706.05587.
7. Chollet, F. (2016). Xception: Deep Learning with Depthwise Separable Convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1800-1807.
8. Sifre, L., & Mallat, S. (2014). Rigid-Motion Scattering for Texture Classification. *ArXiv*, abs/1403.1687.
9. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2818-2826.
10. Li, X., Sun, X., Meng, Y., Liang, J., Wu, F., & Li, J. (2019). Dice Loss for Data-imbalanced NLP Tasks. *ArXiv*, abs/1911.02855.
11. Sethy, D.K. (2020). Rice Leaf Disease Image Samples.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

