

# AOP: An Anti-overfitting Pretreatment for Practical Image-based Plant Diagnosis

Takumi Saikawa\*, Quan Huu Cap\*, Satoshi Kagiwada<sup>†</sup>, Hiroyuki Uga<sup>‡</sup>, Hitoshi Iyatomi\*

Email: takumi.saikawa.5x@stu.hosei.ac.jp huu.quan.cap.78@stu.hosei.ac.jp

kagiwada@hosei.ac.jp uga.hiroyuki@pref.saitama.lg.jp iyatomi@hosei.ac.jp

\*Applied Informatics, Graduate School of Science and Engineering, Hosei University, Tokyo, Japan

<sup>†</sup>Clinical Plant Science, Faculty of Bioscience and Applied Chemistry, Hosei University, Tokyo, Japan

<sup>‡</sup>Saitama Agricultural Technology Research Center, Saitama, Japan

**Abstract**—In image-based plant diagnosis, clues related to diagnosis are often unclear, and the other factors such as image backgrounds often have a significant impact on the final decision. As a result, overfitting due to latent similarities in the dataset often occurs, and the diagnostic performance on real unseen data (e.g. images from other farms) is usually dropped significantly. However, this problem has not been sufficiently explored, since many systems have shown excellent diagnostic performance due to the bias caused by the similarities in the dataset. In this study, we investigate this problem with experiments using more than 50,000 images of cucumber leaves, and propose an anti-overfitting pretreatment (AOP) for realizing practical image-based plant diagnosis systems. The AOP detects the area of interest (leaf, fruit etc.) and performs brightness calibration as a preprocessing step. The experimental results demonstrate that our AOP can improve the accuracy of diagnosis for unknown test images from different farms by 12.2% in a practical setting.

**Index Terms**—automated plant diagnosis, preprocessing, segmentation

## I. INTRODUCTION

Plant health is important to ensure a balance between global food demand and agricultural productivity. Protecting plants against disease plays a crucial role in the agricultural domain, helping to improve crop yields and increase food security. In general, the detection of plant diseases is time-consuming, as it requires a great deal of work related to evaluation and treatment by plant pathologists, especially on large farms. The detection of plant disease via automatic techniques is therefore highly desirable.

Plant leaves are the most common part used to detect disease. The automated detection of plant diseases from leaf images has been widely studied in the scientific community, and many proposals have achieved promising results using powerful approaches based on deep convolutional neural networks (CNNs) [1] and related models. As a pioneer study, Kawasaki et al. [2] developed a three-layer CNN to classify three classes of cucumber disease (two diseases and healthy specimens) using leaf images from a real farm. Their model achieved an average accuracy of 94.9% for a wide variety of photographic conditions and complex backgrounds. Fujita et al. [3] extended their study by including early-stage and ill-conditioned images in order to consider more practical situations. They achieved an average classification performance of 82.3% for an eight-class diagnosis scheme for cucumbers.

Wang et al. [4] classified black rot apple diseases using the PlantVillage open dataset [5] and obtained an average accuracy of 90.4%. Diagnosis studies on tomatoes [6], [7] and several other plant species [8] have also demonstrated good classification performance.

More practical diagnosis systems using a wider range of images, including several types of object, have been also proposed. Fuentes et al. [9] used Faster R-CNN [10] and SSD [11] in an object detection and recognition strategy that was state-of-the-art at the time. They investigated a total of 5,000 annotated tomato field leaf images and achieved a maximum mean average precision of 83.6%. Lu et al. [12] designed a similar framework using a fully convolutional network, and achieved a mean recognition accuracy of 98.0% on a wheat disease dataset; they claim that their system can also be deployed in mobile applications.

Although the above-mentioned studies have shown good results for disease diagnosis thanks to breakthroughs in deep learning models, one of the remaining issues is the overfitting problem, which significantly reduces the test diagnosis performance, particularly when the test images are collected from a different environment than the training images. Although this is a very serious problem, only a limited number of studies have highlighted this aspect.

Mohanty et al. [13] used a total of 54,306 plant leaf images consisting of 38 classes of crop-disease pairs (14 crop species) from the PlantVillage dataset, and achieved a classification accuracy of 99.3%. However, they also noted that the accuracy dropped to around 31% when diagnosis was carried out with different test settings. In [9], a pre-trained CNN attained a success rate of 99.5% when identifying 58 classes (plant types and diseases) based on a dataset drawn from both the laboratory and cultivated fields. However, this rate dropped to about 33% when the model was trained solely with laboratory-condition images and tested with cultivation-condition images. This problem also arises when the diagnostic system is trained and tested on different in-farm leaf datasets. Cap et al. developed a practical plant diagnosis system that is capable of using practical wide-angle images taken by surveillance cameras [14]. In the images used, several dozen leaves were present, and were heavily overlapped. Their diagnosis model was composed of a VGG-based network fine-tuned

TABLE I  
DISEASE DATASET USED IN THIS STUDY

Diseases		Disease classification dataset		
		Training	Validation	Test (Different farm)
Viral diseases	MYSV	7,448	827	2,596
	ZYMV	9,189	1,021	3,364
	CMV	4,109	456	563
	WMV	2,511	279	306
Fungal diseases	Brown Spot	2,014	224	654
	Downy Mildew	1,311	146	380
	Powdery Mildew	2,204	245	114
Healthy		6,908	768	1,138
Total		35,694	3,966	9,115

with more than 30,000 cucumber leaf images and achieved a mean accuracy of 97.4% on a test dataset from the same farm; however, this dropped to 65.8% when they tested it on images from a different farm.

In plant diagnosis tasks, the appearance of symptoms is not always clear, and thus the influence of other factors such as variance in the background and image condition is much larger than in general object recognition tasks. Thus, overfitting becomes important. Fujita et al. [15] visualized the key regions of diagnostic evidence for their cucumber diagnosis system using GradCAM [16]. Although their model achieved an average accuracy of 93.6%, they claimed that it sometimes responded to the background areas instead of the target leaf regions, due to the overfitting problem.

These results indicate that some sort of standardization process for images is required. In particular, when plants are grown in a controlled environment to ensure the quality of the training labels, the background diversity tends to be limited. For instance, a set of images infected with a certain type of disease may all have a specific background. For this reason, we believe that the introduction of an efficient preprocessing step to segment out only the region of interest (RoI) is needed to reduce the abovementioned overfitting problem, which is particularly prominent in plant diagnosis.

Segmentation itself has a long research history, and the robust extraction of a RoI against complex backgrounds is generally difficult. Recently, several robust methods based on deep learning have been proposed and have been shown to have outstanding capabilities. For instance, U-net [17] and its improved version called pix2pix [18] do not require any knowledge of the target domain, and have obtained very reasonable segmentation performance just by learning pairs of training images and their segmentation areas. In the past, segmentation of an RoI as part of preprocessing was an essential step in most image recognition or computer vision tasks; however, in the era of deep learning, the need for this step has decreased, and it is generally omitted except in some medical tasks. For the reasons described above, a sophisticated region extraction process is now required in practical image-based plant diagnosis tasks.

In this study, in order to address the serious overfitting

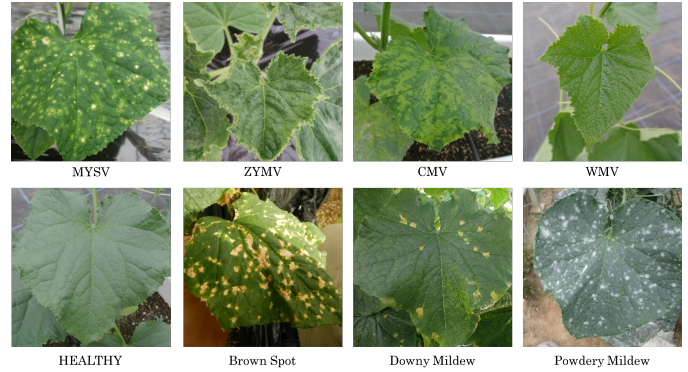


Fig. 1. Typical examples of seven diseases and a healthy leaf. Each image is labeled as one of four types of viral disease (MYSV, ZYMV, CMV, and WMV), three types of fungal disease (Brown Spot, Downy Mildew, and Powdery Mildew) or a healthy leaf.

problem, we propose an anti-overfitting pretreatment (AOP) to detect the area of interest (i.e. leaf, fruit, etc.) as a preprocessing step for image-based plant diagnosis systems.

To the best of our knowledge, this is the first study to conduct segmentation of the area of interest for a plant diagnosis task and to confirm the validity using real farm images.

## II. MATERIALS AND METHODS

### A. Dataset

In this study, we use two types of cucumber leaf datasets. The first is the segmentation dataset, which consists of a total of 8,000 cucumber leaf images and their corresponding leaf region pixel-level annotated ground-truths. This dataset is used to train the core network of our proposed AOP. Note that we collected only images in good condition for our segmentation dataset (i.e. we removed images that were too bright or too dark, and other ill-conditioned images, in advance). Because this data set consists of images with various aspect ratios, we apply a square center crop with the short side of the image as the side length. We then resize them to  $316 \times 316$  pixels.

The second is a classification dataset consisting of 48,775 single-leaf cucumber images, as summarized in Table I. These

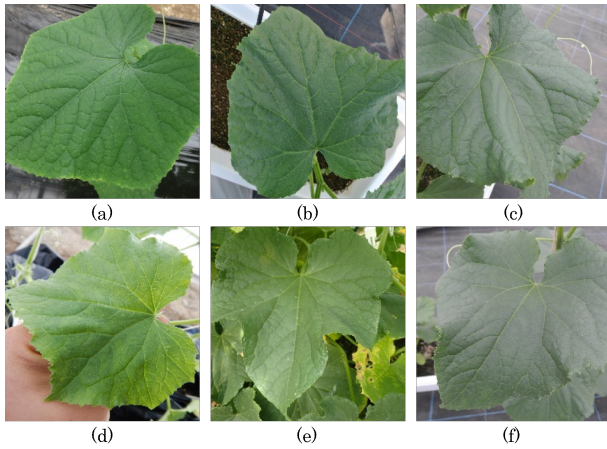


Fig. 2. Early symptoms of (a) MYSV; (b) CMV; (c) WMV; (d) Brown Spot; (e) Powdery Mildew; and (f) a healthy leaf. Early symptoms are almost indistinguishable from healthy leaves and the early stages of other diseases.

images were labeled by experts from agricultural technology research centers in six prefectures of Japan, and each image contains a single cucumber leaf in the middle. Each image in this dataset is of one of four types of viral disease (MYSV, ZYMV, CMV, and WMV)<sup>1</sup>, three fungal diseases (Brown Spot, Downy Mildew, and Powdery Mildew) or a healthy leaf. This classification dataset has three subsets, i.e. the  $disease_{training}$ , the  $disease_{validation}$ , and the  $disease_{test}$ , which are used to train/validate and test the final disease diagnosis networks (details of this are given later). The  $disease_{training}$  set contains 35,694 images, the  $disease_{validation}$  set 3,966 images, and the  $disease_{test}$  set 9,115 images. All the images are resized to  $224 \times 224$  pixels. Note that the training/validation sets were taken from the same farm, but that the testing set was taken from completely different farms. Fig. 1 shows typical example images for each category.

It should also be mentioned that the condition of some of the images in this classification dataset is not good compared to the segmentation dataset. Moreover, this dataset contains many images of early-stage disease, i.e. MYSV, CMV, WMV, Brown Spot and Powdery Mildew. These early-stage images are almost indistinguishable, both from each other and from healthy images. Powdery Mildew is the class containing most of the early-stage images (early-stage images account for more than 25% of the total images for Powdery Mildew disease). Fig. 2 also shows examples of these early cases.

### B. The proposed anti-overfitting pretreatment (AOP)

The key features of our proposed anti-overfitting pretreatment (AOP) are a robust segmentation process of the RoI (i.e. the leaf shape) and image brightness calibration. These functions reduce the serious negative effects of the background in terms of causing the overfitting problem. In our experiment,

<sup>1</sup>MYSV: Melon yellow spot virus, ZYMV: Zucchini yellow mosaic virus, CMV: Cucumber mosaic virus, WMV: Watermelon mosaic virus

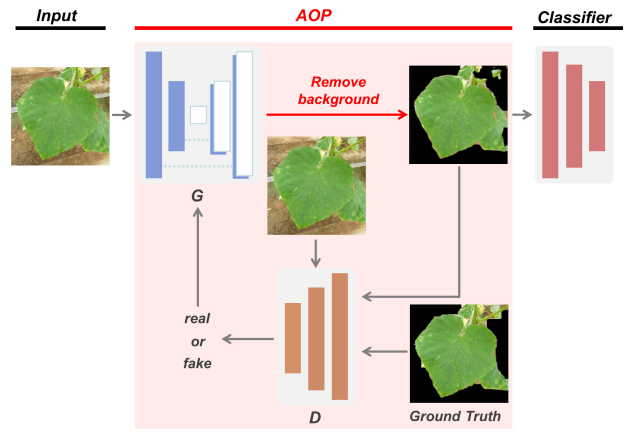


Fig. 3. AOP network architecture. Our AOP network is based on the pix2pix architecture, and is composed of a generator and a discriminator.

we examine the effect of the AOP on diagnosis of cucumber disease based on leaf pictures taken from real farms.

Our AOP network is based on the pix2pix architecture [18], and is composed of a generator and a discriminator. Fig. 3 shows our plant diagnosis schema with the proposed AOP. The generator is based on the U-net architecture [17], and the encoder and decoder are basically the same as in the original pix2pix model. Specifically, the encoder consists of eight convolution and eight deconvolution layers, while the decoder has five convolutional layers. Since the vanilla U-net often generates checkerboard artifacts [19] arising from the deconvolution process in the image generator, we added an unpooling layer after each convolution layer to avoid this.

Pix2pix has been shown to have excellent segmentation and image translation capabilities. Since pix2pix is used for image translation, we believe that it can be used to improve the condition of bad images, such as those with poor lighting conditions. Here, our generator carries out image segmentation and also improves the image condition. Our images are taken under different photographic conditions on different farms, meaning that poorly conditioned images are included. This can lead to incorrect segmentation and lower disease diagnostic performance, since important parts of some images are missing. In order to deal with this problem, we apply brightness calibration to the input images using gamma correction to train our AOP models. We assume that this brightness calibration allows our model to generate segmented images with better photographic conditions.

To achieve this, we apply gamma correction after performing several online data augmentation techniques to train the AOP, i.e. image rotation with a step size of  $90^\circ$ , image mirroring, and random cropping to a size of  $256 \times 256$  pixels. Gamma correction is used, with gamma values randomly selected from [0.8, 1.0, 1.5, 2.0, 2.2]. In addition to the mean absolute error (MAE) content loss used in pix2pix, we also introduce structural similarity (SSIM) [20] to our models, since improving the target structure is expected to improve the

diagnostic performance of the final disease diagnosis system. In our experiment, we compare the difference in diagnosis performance through several AOP models.

### C. Disease classification networks

The final aim of this study is to prove the effectiveness of our proposed AOP against the overfitting problem when testing the diagnostic systems on unknown datasets. We therefore design different disease classification networks and test them on the unknown  $\text{disease}_{\text{test}}$  dataset. These networks are based on the VGG-16 network [21] and are pre-trained with the ImageNet dataset [22]; we fine-tune them on the  $\text{disease}_{\text{training}}$  dataset.

We use small numbers of neurons (1024 and 32) in the FC layers before the eight-class output in order to avoid model overfitting. The details of the training disease classification networks will be described in the next section.

### D. Details of training the networks

1) *Training AOP models:* In the training of the AOP models, the data augmentation techniques described earlier are performed on the fly. We randomly select 80% of the segmentation dataset for training and the remaining 20% is used for testing the AOP models. In this work, we train three AOP models with different loss functions for comparison purposes: the  $\text{AOP}_{\text{MAE}_{\text{prob}}}$ , which generates a segmentation mask with the original MAE loss from pix2pix as its content loss (i.e. the output is in a probabilistic form and the thresholding forms the mask); the  $\text{AOP}_{\text{MAE}}$ , with a MAE content loss that directly generates pixel values; and the  $\text{AOP}_{\text{SSIM}}$ , which uses the structural similarity as its loss function.

We apply an Adam optimizer [23] with  $\alpha = 0.0002$  and  $\beta = 0.5$  for both the generator and discriminator of the AOP models. The minibatch was set to 16, and the training process finished after 100 epochs.

2) *Training disease classification networks:* In this experiment, we train four disease classification networks in order to compare the final diagnostic performance. The first classifier is trained, validated, and tested with the original versions of the  $\text{disease}_{\text{training}}$ ,  $\text{disease}_{\text{validation}}$ , and the  $\text{disease}_{\text{test}}$  datasets, respectively. The other three classifiers are trained, validated, and tested on the segmented versions of these datasets produced by the three AOP models, as described in the previous subsection.

To train these networks, the online data augmentation techniques described above were applied, but with a rotation angle step size of  $10^\circ$  and randomly selected gamma values of [0.5, 1.0, 1.5]. Our disease classifiers were optimized using the momentum stochastic gradient descent [24] with the minibatch set to 32, and the initial learning rate was 0.001. The training process was also stopped after 100 epochs.

## III. EXPERIMENTAL RESULTS

### A. Segmentation performance of the AOP networks

Table II compares the segmentation performance for the test dataset (20% of the segmentation dataset), and Fig. 4

TABLE II  
SEGMENTATION PERFORMANCE WITH AOP NETWORKS

	Precision [%]	Recall [%]	F1-score [%]
$\text{AOP}_{\text{MAE}_{\text{prob}}}$	100.0	95.8	95.8
$\text{AOP}_{\text{MAE}}$	98.4	97.1	97.8
<b>(Proposed) <math>\text{AOP}_{\text{SSIM}}</math></b>	98.6	97.5	<b>98.1</b>

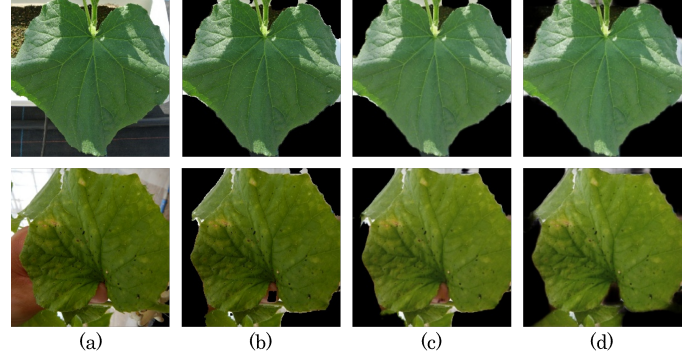


Fig. 4. Examples of (a) an input image; (b) an image of the extracted leaf area  $\text{AOP}_{\text{MAE}_{\text{prob}}}$ ; (c) an image generated by  $\text{AOP}_{\text{MAE}}$ ; and (d) an image generated by  $\text{AOP}_{\text{SSIM}}$ . It can be seen that the leaf region is extracted properly.

shows examples of results of segmented leaves. We use the precision, recall, and F1-score to evaluate the segmentation performance for the three AOP models. The segmentation results of  $\text{AOP}_{\text{MAE}_{\text{prob}}}$  and  $\text{AOP}_{\text{MAE}}$  show that the leaf region is extracted properly.

Note that we choose a threshold value of 0.5 to generate the mask from the  $\text{AOP}_{\text{MAE}_{\text{prob}}}$  model. All of the AOP models achieve high segmentation performance, and the best F1-score of 98.1% is obtained from the proposed  $\text{AOP}_{\text{SSIM}}$ . The results confirm that these segmentation models are strong enough for background removal and are expected to improve the disease diagnostic performance of the subsequent diagnosis systems.

### B. Disease classification performance

The disease classification performance was evaluated and compared, both with and without the AOP models. Table III summarizes the performance in terms of average accuracy. While there are only small differences in the classification accuracy between the  $\text{disease}_{\text{training}}$  and  $\text{disease}_{\text{validation}}$  sets, there are significant undesirable differences between the  $\text{disease}_{\text{validation}}$  and  $\text{disease}_{\text{test}}$  sets. The biggest drop in performance is shown by the classification model without AOP. The accuracy of this model drops maximally from 97.5% on  $\text{disease}_{\text{validation}}$  to only 40.3% on  $\text{disease}_{\text{test}}$  dataset. In contrast, the classification network with our  $\text{AOP}_{\text{SSIM}}$  significantly improves the accuracy on the unknown  $\text{disease}_{\text{test}}$  dataset by 12.2%, reaching the best performance of 52.5% classification accuracy.

Fig. 5 compares the confusion matrices for the disease classification networks (a) without AOP and (b) with  $\text{AOP}_{\text{SSIM}}$ . Although there are several misclassifications for Powdery



TABLE III  
DIAGNOSIS PERFORMANCE WITH AND WITHOUT AOP

	Average Accuracy [%] (8 Classes)		
	Disease <sub>training</sub>	Disease <sub>validation</sub>	Disease <sub>test</sub>
w/o AOP	98.2	97.5	40.3
AOP <sub>MAE<sub>prob</sub></sub>	98.2	97.2	48.8
AOP <sub>MAE</sub>	98.0	97.1	49.7
<b>(Proposed) AOP<sub>SSIM</sub></b>	98.0	97.4	<b>52.5</b>

Mildew in both cases, we can confirm that the AOP<sub>SSIM</sub> markedly improves the diagnostic performance. In addition, we also find that our AOP<sub>SSIM</sub> helps the classification network to focus on the leaf area instead of on its background. Fig. 6 shows the diagnostic evidence heatmap of the classification network (a) without AOP and (b) with the proposed AOP<sub>SSIM</sub>, using the Grad-CAM [16] technique. The darker the heatmap, the stronger the classifier considers that areas. The results with AOP<sub>SSIM</sub> show that the classification network responds to the leaf area, while the model without AOP responds to the background areas.

#### IV. DISCUSSION

These experiments indicate that the diagnosis performance for the test dataset is much lower than for the training/validation dataset in all cases (see Table III). The primary reason for this low performance is the misdiagnosis for Powdery Mildew, as shown in the confusion matrix in Fig. 5. We can see that there are numerous false positives for Powdery Mildew compared to other diseases and healthy data. WMV is also mostly misclassified as healthy or Powdery Mildew.

Since there are 2,204 images of Powdery Mildew in the training set, but almost 25% of these show early symptoms, our model learns to mistakenly classify the early symptoms of the other diseases of WMV, MYSV, CMV, Brown Spot and healthy images as Powdery Mildew, as they are very similar to each other (see Fig. 2). This problem with early symptoms needs to be addressed, and we leave this for future work.

We confirm that the introduction of AOP draws the classifier's attention to the leaf, rather than being confused by the background, as shown in Fig. 6, and this significantly improves the classification performance for a completely different dataset. As shown in Table III, the accuracy of the disease classifier without AOP for the test data was 40.3%, while the classifier with AOP<sub>SSIM</sub> achieved an accuracy of 52.5%, thus improving the overall accuracy by 12.2%. This result also supports our hypothesis that SSIM, which captures structural features better, is more effective for the final diagnosis in situations where leaf regions have been properly extracted and where a background that causes disturbance has been removed.

We can also see from the results in Table III that the AOP<sub>MAE</sub>, which acts as a brightness calibrator, improves the diagnosis performance by 0.9% compared to the AOP<sub>MAE<sub>prob</sub></sub>. However, this improvement was smaller than expected. This is because the brightness gamma correction

(a) Without AOP

MYSV	854	40	112	42	61	32	2	1453
ZYMV	688	2105	53	139	21	2	1	355
CMV	45	6	128	7	39	0	2	336
WMV	3	3	20	4	40	0	0	236
HEALTHY	6	1	66	2	217	0	0	846
Brown Spot	21	0	18	0	18	154	36	407
Downy mildew	43	0	0	0	50	7	109	171
Powdery mildew	0	0	0	0	5	0	1	108

(b) With AOP<sub>SSIM</sub>

MYSV	1046	107	226	28	79	40	6	1064
ZYMV	516	2404	132	129	59	2	3	119
CMV	80	6	261	9	77	0	1	129
WMV	7	3	36	6	118	0	0	136
HEALTHY	15	2	133	2	687	0	1	298
Brown Spot	147	10	35	0	14	153	15	280
Downy mildew	70	1	10	0	66	11	131	91
Powdery mildew	1	1	0	0	11	0	1	100

Fig. 5. Confusion matrices for disease classification networks with and without AOP. Although there are a significant number of misclassifications for Powdery Mildew in both cases, the AOP<sub>SSIM</sub> markedly improves the diagnostic performance.

data augmentation in the diagnosis classifier also contributes to leveraging the final diagnostic performance.

In summary, introducing the AOP eliminates the influence of the background and significantly improves the diagnosis performance. Although the overall performance of 52.5% for the disease<sub>test</sub> set is not sufficient for practical systems, mainly due to the limited performance in regard to early-stage symptoms, we have confidence that these results reflect a real system accuracy based on tens of thousands of training images, including a large number of early-stage cases. We will continue to investigate other crucial factors to achieve a practical diagnosis performance.

#### V. CONCLUSION

In this paper, we confirm the problem caused by latent similarities in image-based plant diagnosis studies. We propose an AOP that provides appropriate segmentation of the RoI,

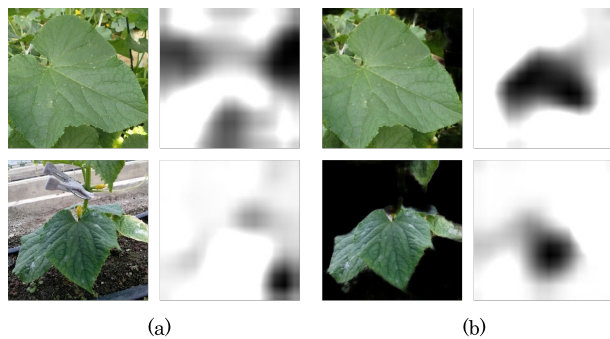


Fig. 6. Comparison of estimated evidence of diagnosis (a) without AOP; and (b) with the proposed AOP<sub>SSIM</sub>. In the results with AOP, the highlighted areas overlap the leaf area, while without AOP they do not, as reported in the literature.

in order to realize robust image-based plant diagnosis that reduces serious overfitting problems. With the introduction of the AOP, the final diagnosis performance for cucumber leaves improved by 12.2% on average for an eight-class classification task.

#### ACKNOWLEDGMENT

This research was supported in part by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Fundamental Research (C), 17K08033, 2017–2020.

#### REFERENCES

- [1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [2] Yusuke Kawasaki, Hiroyuki Uga, Satoshi Kagiwada, and Hitoshi Iyatomi. Basic study of automated diagnosis of viral plant diseases using convolutional neural networks. In *International Symposium on Visual Computing*, pages 638–645. Springer, 2015.
- [3] Erika Fujita, Yusuke Kawasaki, Hiroyuki Uga, Satoshi Kagiwada, and Hitoshi Iyatomi. Basic investigation on a robust and practical plant diagnostic system. In *15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 989–992. IEEE, 2016.
- [4] Guan Wang, Yu Sun, and Jianxin Wang. Automatic image-based plant disease severity estimation using deep learning. *Computational intelligence and neuroscience*, 2017, 2017.
- [5] David Hughes, Marcel Salathé, et al. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *CoRR preprint arXiv:1511.08060*, 2015.
- [6] Halil Durmuş, Ece Olcay Güneş, and Mürvet Kırıcı. Disease detection on the leaves of the tomato plants by using deep learning. In *6th International Conference on Agro-Geoinformatics*, pages 1–5. IEEE, 2017.
- [7] Habibollah Agh Atabay. Deep residual learning for tomato plant leaf disease identification. *Journal of Theoretical & Applied Information Technology*, 95(24), 2017.
- [8] Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk, and Darko Stefanovic. Deep neural networks based recognition of plant diseases by leaf image classification. *Computational intelligence and neuroscience*, 2016, 2016.
- [9] Alvaro Fuentes, Sook Yoon, Sang Kim, and Dong Park. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors*, 17(9):2022, 2017.
- [10] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems (NIPS 2015)*, pages 91–99, 2015.
- [11] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision (ECCV 2016)*, pages 21–37. Springer, 2016.
- [12] Jiang Lu, Jie Hu, Guannan Zhao, Fenghua Mei, and Changshui Zhang. An in-field automatic wheat disease diagnosis system. *Computers and electronics in agriculture*, 142:369–379, 2017.
- [13] Sharada P Mohanty, David P Hughes, and Marcel Salathé. Using deep learning for image-based plant disease detection. *Frontiers in plant science*, 7:1419, 2016.
- [14] QH Cap, K Suwa, E Fujita, H Uga, S Kagiwada, and H Iyatomi. An end-to-end practical plant disease diagnosis system for wide-angle cucumber images. *International Journal of Engineering & Technology*, 7(4.11):106–111, 2018.
- [15] E Fujita, H Uga, S Kagiwada, and H Iyatomi. A practical plant diagnosis system for field leaf images and feature visualization. *International Journal of Engineering & Technology*, 7(4.11):49–54, 2018.
- [16] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proc. of the IEEE International Conference on Computer Vision*, pages 618–626, 2017.
- [17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [18] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proc. of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
- [19] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and checkerboard artifacts. *Distill*, 1(10):e3, 2016.
- [20] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
- [21] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR preprint arXiv:1409.1556*, 2014.
- [22] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [23] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR, abs/1412.6980*, 2014.
- [24] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1):145–151, 1999.