IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 22, 2019, accepted November 15, 2019, date of publication November 21, 2019, date of current version December 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2954845

Strawberry Verticillium Wilt Detection Network Based on Multi-Task Learning and Attention

XUAN NIE[1], LUYAO WANG[1], HAOXUAN DING[2], AND MIN XU[3]
[1]School of Software, Northwestern Polytechnical University, Xi'an 710072, China
[2]School of Power and Energy, Northwestern Polytechnical University, Xi'an 710072, China
[3]School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia

Corresponding author: Min Xu (min.xu@uts.edu.au)

This work was supported by the 2018 Science and Technology Plan of Shaanxi Province under Project 2018ZDXM-GY-108.

ABSTRACT Plant disease detection has an inestimable effect on plant cultivation. Accurate detection of plant disease can control the spread of disease early and prevent unnecessary loss. Strawberry verticillium wilt is a soil-borne, multi-symptomatic disease. To detect strawberry verticillium wilt accurately, we first propose a disease detection network based on Faster R-CNN and multi-task learning to detect strawberry verticillium wilt. Then, the strawberry verticillium wilt detection network (SVWDN), which uses attention mechanisms in the feature extraction of the disease detection network, is proposed. SVWDN detects verticillium wilt according to the symptoms of detected plant components (i.e.,young leaves and petioles). Compared with other existing methods for detecting disease from the whole plant appearance, the SVWDN automatically classifies the petioles and young leaves while determining whether the strawberry has verticillium wilt. To provide a dataset for evaluating and testing our method, we construct a large dataset that contains 3, 531 images with 4 categories (Healthy_leaf, Healthy_petiole, Verticillium_leaf and Verticillium_petiole). Each image also has a label to indicate whether the strawberry is suffering from verticillium wilt. With the proposed strawberry verticillium wilt detection network, we achieved a mAP of 77.54% on object detection of 4 categories and 99.95% accuracy for strawberry verticillium wilt detection.

INDEX TERMS Computer vision, image classification, multitasking, object detection.

I. INTRODUCTION

Early detection of plant disease helps to control disease spread and reduce economic loss. Most plant diseases cause abnormalities in leaves and petioles, which makes early plant disease detection possible by analyzing the appearance of leaves and petioles [1]–[3]. Verticillium wilt is a common soil-borne disease in strawberry plants that invades the root and shows symptoms on the ground. When the strawberries are infected with verticillium wilt, the young leaves turn yellow-green and twist into a navicular shape. The surface of the infected leaves loses luster, the bilateral lobules on the compound leaves are asymmetrical, one or two of the three lobules often become deformed, the petiole becomes tawny, and the leaf starts to dry from the leaf edge until the whole plant dies. Since strawberry is a perennial plant, it is planted in the same field for many years, resulting in a large accumulation of pathogenic bacteria in the soil. In addition, the disease resistance of the strawberry is reduced, and the field management is poor. The incidence of strawberry verticillium wilt is increasing year by year, and verticillium wilt has become an important disease in strawberry production, which has a very serious effect on strawberry yield [3]. As a soil-borne disease, verticillium wilt is very easy to spread. The infected plant must be pulled out and taken out of the field and burned as soon as possible. The symptoms of infected strawberry are complex and are often identified and diagnosed only by professional personnel. In recent years, due to the insufficient recognition of verticillium wilt by strawberry growers, the economic loss of strawberry yields has been serious. Therefore, we propose a detection network to assist strawberry growers, which will improve the accuracy of verticillium wilt identification.

Plant disease detection has always been a highly regarded research field. Existing research on plant disease detection has mainly focused on the biological characteristics of diseases [1]. For instance, studies on potato [4] and tomato [5] plants identify, understand and manage diseases by focusing

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

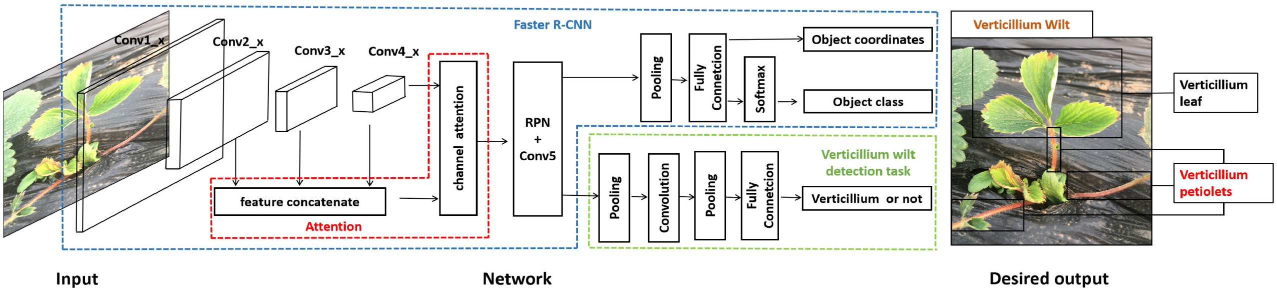


FIGURE 1. The strawberry verticillium wilt detection network is based on Faster R-CNN and multi-task learning, which extracts shared features using attention mechanisms and determines whether the strawberry has verticillium wilt according to the detected symptoms of young leaves and petioles.

on the pathogens and viruses. Recently, researchers have started applying machine learning methods for plant disease detection [6]–[8]. In [6], Ramesh *et al.* extracted the histogram of an oriented gradient [7] (HOG) feature of the crop image and classified by a random forest, they only classified the plant images by the handcrafted extracting feature, without detection. Fuentes *et al.* [8] detected tomato diseases and recognized pests by deep learning. In these methods, the disease detection of plants is only based on single components, such as leaves, to determine whether the plant is suffering from disease.

For verticillium wilt detecting, we comprehensively consider the polymorphic characteristics of verticillium wilt in petioles and young leaves. In this paper, the strawberry verticillium wilt detection network (SVWDN) (see Figure 1) based on Faster R-CNN and multitasking is proposed to detect and recognize verticillium wilt in strawberry images. In SVWDN, in addition to detecting and classifying the petioles and young leaves in the images, the detected components are also used to determine whether the whole plant is infected with verticillium wilt. The first four stacked blocks of ResNet-50 [9] are used as the backbone for features extraction. In the red dotted frame, we concatenate multi-level feature maps to generate a channel-wise weight attention vector and further conduct the feature extraction for region proposal network (RPN). After RPN and the fifth stacked block of ResNet-50, the object detection task of Faster R-CNN that locates proposal region (i.e.,regresses region coordinates) and decides the class label of each proposal region, is performed as shown in the blue dotted frame. Moreover, the strawberry verticillium wilt detection task that classifies whether the strawberry has verticillium wilt is performed using features extracted from proposal regions, as shown in the green dotted frame.

The main contributions of this research are introduced as follows:

- A new dataset of strawberry verticillium wilt is constructed, which contains 3, 531 images collected from a strawberry growing greenhouse, among which 789 are healthy samples, and 2, 742 are verticillium wilt samples. To improve the sample diversity, we took photos from different angles during the collection process.

FIGURE 2. Examples of our verticillium wilt dataset. The red box is a sample of verticillium wilt, and the blue box is a healthy sample.

- Multi-task learning is adapted to introduce a new disease detection task, which judges whether the plant has verticillium wilt from the detected young leaves and petioles. This task achieves an accuracy of 99.85% on the collected dataset. We address Faster R-CNN with the disease detection task as a disease detection network (DDN).
- A novel attention mechanism for verticillium wilt detection is proposed. We use the multi-level features to generate the channel-wise attention weight vector for the high-level feature maps. Compared with the disease detection network (DDN), the object detection performance of the network with attention is improved and the accuracy of the disease detection task is 99.95% on the collected dataset. We address DDN with the attention mechanism as strawberry verticillium wilt detection network (SVWDN).

II. RELATED WORKS

In this section, we briefly introduce related works in four aspects. Initially, several representative object detection methods are reviewed, and applications of object detection are discussed. Second, research on multi-task learning is briefly introduced. Third, the application of the attention mechanism is reviewed. Finally, existing methods of plant disease detection are compared.

A. OBJECT DETECTION

Detection of strawberry verticillium wilt requires the detection of petioles and young leaves first and then determines whether the plant has verticillium wilt by symptoms of the young leaves and petioles. We need to detect and classify young leaves and petioles by object detection methods. There are many CNN methods for object detection. R-CNN [10] input the region proposals found by selective search into the CNN for classification and bounding-box regression, which greatly improves the detection performance compared with OverFeat [11]. A region proposal network (RPN) was used in Faster R-CNN [12] to generate the region proposals, including their class (i.e.,object or not object) and box coordinates. Then, the region of interest (RoI) pooling layer extracted the feature of each region proposal for classification and bounding-box regression. Both YOLO [13] and SSD [14] are end-to-end object detection methods. The detection speed of YOLO is fast, but object detection effect on small object is poor. SSD detects objects from feature maps of different scales. Thus, SSD can detect objects of various sizes contained in the images with high speed.

Object detection is used in many fields. Mao *et al.* [15] considered aggregating different kinds of extra features into the Fast R-CNN based pedestrian detection framework. Dan *et al.* [16] proposed a novel method that relies on a cross-modality learning framework for detecting pedestrians under adverse illumination conditions. Zhang *et al.* [17] combined multi-layer features based on Faster R-CNN and reversed weight for these features to detect road markings.

B. MULTI-TASK LEARNING

Multi-task learning [18], [19] is popular and effective in computer vision. Ranjan *et al.* [20] simultaneously detected the face, localized landmarks, estimated the pose, and recognized the gender in CNN for an image. In Faster R-CNN [12], Ren *et al.* shared the same feature maps to perform the classification and bounding-box regression simultaneously. Dan *et al.* [21] predicted a set of intermediate auxiliary tasks for final depth estimation and scene parsing. Hariharan *et al.* [22] simultaneously learned object detection and semantic segmentation in a network based on R-CNN.

C. ATTENTION MECHANISM

The attention mechanism is widely used in computer vision. Fei *et al.* [23] generated attention-aware features by stacking attention CNN and considerably improved image classification. Long *et al.* [24] achieved good performance on image captioning by using attention mechanisms on channels and feature maps. For scene parsing, Zhao *et al.* [25] connected each position on the feature maps with all the other positions through a learning attention mask. Zhang *et al.* [26] used the spatial and channel-wise attention guided recurrent network for salient object detection. Xiao *et al.* [27] adopted stacked hourglass networks [28] to generate context attention maps for human pose estimation.

D. DISEASE DETECTION IN PLANTS

Plant disease detection is a vital topic that has been studied in recent years and is motivated by the need to reduce loss during plant growth. Before deep learning became popular in computer vision, most studies on plant disease detection were only classification without detection, the handcrafted feature extraction method was widely used in plant disease classification. In [6], [29]–[31], Hu's moment, color histogram, histogram of an oriented gradient (HOG) and gray-level cooccurrence matrix (GLCM) were used as the distinctive attributes for classification by SVM, random forest, naive Bayes and ANN. Dalal and Triggs extracted the histogram of an oriented gradient [7] (HOG) feature of the cropped image for disease classified in plants. Rothe and Kshirsagar [29] extracted Hu's moments from cotton images as a distinctive feature and classified by using an ANN. Deep learning had good performance in image classification in [32], [33], where CNN was used for disease detection by classifying leaf images. After a considerable improvement in object detection, Fuentes et al. [8] considered Faster R-CNN, R-FCN and SSD as detectors to detect tomato plant diseases and recognize pests.

In contrast to the aforementioned plant disease methods, since verticillium wilt has to be diagnosed by two components (i.e.,petioles and young leaves), SVWDN based on Faster R-CNN is proposed. Multitasking is adapted in the novel SVWDN to introduce a disease detection task that classifies whether the plant has verticillium wilt by the detected components. Feature concatenate and channel-wise attention is used in SVWDN to improve the performance.

III. METHOD

A. BASELINE: FASTER R-CNN

Faster R-CNN has performed well in object detection, especially in improving the speed and accuracy in detecting small objects compared with R-CNN. In our verticillium wilt dataset, there are some small objects, and we expect our network to have a high detection accuracy. Therefore, our baseline detector is an implementation of Faster R-CNN [12], initialized with ResNet-50 weights and pretrained on ImageNet [34]. Faster R-CNN is a two-stage detector that first extracts the shared features from the input image and then generates the region proposals through RPN. Finally, objects are detected from the region proposals that were cropped from the RoI pooling layer.

B. AN EXTRA TASK FOR DIAGNOSIS OF VERTICILLIUM WILT IN FASTER R-CNN BASED ON MULTITASKING

To determine whether the strawberry has verticillium wilt according to the detected symptoms of components, a new disease detection task is proposed in our baseline detector Faster R-CNN. As shown in Figure 1, $Conva\_x$ represents the output of the x-th block in the a-th blocks stacked in ResNet-50, after RPN and $Conv5$, the new disease detection task is performed in the green dotted frame of Figure 1.

In disease detection task, features extracted from proposal regions is used to classify whether the plant is suffering from verticillium wilt. In the following, we describe the

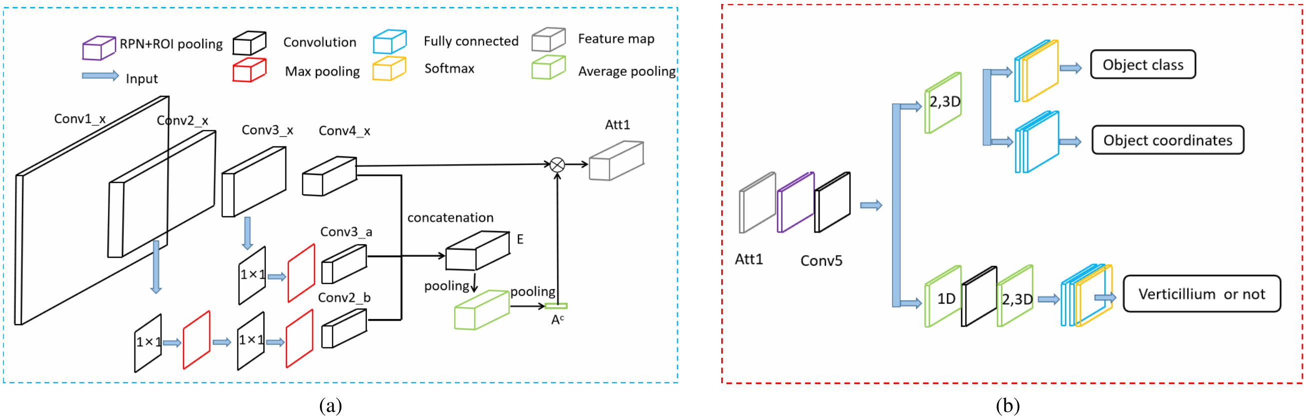


FIGURE 3. **(a) is the CNN (ResNet-50) that generates a channel-wise attention weight vector using multi-level features to conduct the high-level features extraction. (b) is consisted of object detection task and disease detection task.**

disease detection task in detail. As shown in Figure 3(b), after the *Conv*5, the branch on the top demonstrates the task of object detection using Faster R-CNN, while the branch on the bottom is the strawberry verticillium wilt detection task. F is denoted as the input feature, which is the output of *Conv*5 and has the shape of $N \times W \times H \times C$. Here, N is the number of RoIs, W and H are the width and height of feature maps, and C is the number of channels. We unfold F as $F = [F_1, F_2, \ldots, F_N]$, where F_i is the i-th RoI. In the strawberry verticillium wilt detection task, average pooling is used to concatenate features extracted from N RoIs.

We denote G^c as the average pooling of features extracted the N RoIs:

$$G^c = \frac{1}{N}\sum_{i=1}^{N} F_i^c \tag{1}$$

The shape of G^c is $1 \times W \times H \times C$. After G^c was convolved by a convolutional layer with kernel size of 3×3 and channel number of 2, 048, the average pooling for each channel of G^c in spatial dimensions is performed. G_{ij}^c denotes the position in the i-th row and j-th column of the c-th channel and V^c denotes the average pooling result for each channel.

$$V^c = \frac{1}{W \times H}\sum_{i=1}^{W}\sum_{j=1}^{H} G_{ij}^c \tag{2}$$

Then, V^c is input into two fully connected layers and a softmax layer to classify whether the strawberry has verticillium wilt.

The object detection task in the top branch (shown in Figure 3(b)) first performs the average pooling in the spatial dimensions of F. Then, the object is classified through a fully connected layer and a softmax layer and the object coordinates are regressed through a fully connected layer. Since the disease detection task is added in Faster R-CNN, the loss function in SVWDN is defined as

$$Loss = L_{RPN}(\{p_i\}, \{f_i\}) + L_{ODN}(w, u, t^u, v) + L_{NC}(m, n) \tag{3}$$

We denote $Loss$ as the total loss that is composed of $L_{RPN}(\{p_i\}, \{f_i\})$ as the loss of the RPN, $L_{ODN}(w, u, t^u, v)$ as the loss of the (objects detection network) ODB and $L_{NC}(m, n)$ as the loss of the new classification task which detects whether the strawberry has verticillium wilt.

$$\begin{aligned} L_{RPN}(\{p_i\}, \{f_i\}) &= \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) \\ &\quad + \lambda_1 \frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*) \end{aligned} \tag{4}$$

$$\begin{aligned} L_{ODN}(w, u, t^u, v) &= L_{cls}(w, u) \\ &\quad + \lambda_2 [u \geq 1] L_{reg}(t^u, v) \end{aligned} \tag{5}$$

Here, $L_{RPN}(\{p_i\}, \{f_i\})$ and $L_{ODN}(w, u, t^u, v)$ are the same as in the Faster R-CNN. i represents the anchor index in a mini-batch, p_i represents the predicted probability of being an object for anchor i, p_i^* represents the ground truth label, t_i represents the coordinates of the predicted bounding box for anchor i, and t_i^* represents the coordinates of the ground truth box. N_{cls} ($N_{cls} = 256$) [12]is the number of anchors in mini-batch and N_{reg} ($N_{reg} \sim 2, 400$) [12] is the number of locations. L_{cls} is the cross-entropy loss, and L_{reg} is the smooth L1 loss in the Faster R-CNN. w is the predicted class score, u is the true class score, t^u is the true box coordinate and v is the predicted coordinate.

$$L_{NC}(m, n) = \lambda_3 \sum_i m_i \log \frac{1}{n_i} \tag{6}$$

$L_{NC}(m, n)$ is the cross-entropy loss of the disease detection task, where m is the true class score, and n is the predicted class score. These loss terms are normalized with balancing parameters λ_1, λ_2, and λ_3.

C. GENERATE CHANNEL-WISE ATTENTION VECTOR USING MULTI-LEVEL FEATURES

In the attention mechanism, the research attempts to focus on the effective regions or features to achieve better results. Additionally, in our method, we consider each channel of the feature maps to be unequally important for subsequent regression and classification tasks. In the task of leaves and petioles detection, according to human experience, color information is generally useful to distinguish healthy and unhealthy petioles. However, color information might not be properly represented by high-level feature maps (i.e.,$Conv4\_6$) [35]. Therefore, we apply multi-level feature maps to set different weights to various channels of $Conv4\_6$ features in order to emphasise the channels which remain color information.

Faster R-CNN was initialized with ResNet-50 and used $Conv4\_6$ as the input to the RPN. $Conv2\_3$, $Conv3\_4$, and $Conv4\_6$ are used to generate a channel-wise attention weight vector A^c as the weight of different channels for $Conv4\_6$ layer. The shape of $Conv2\_3$, $Conv3\_4$, and $Conv4\_6$ feature maps is $1 \times 4W \times 4H \times 256$, $1 \times 2W \times 2H \times 512$ and $1 \times W \times H \times 1024$.

In the following, we describe channel-wise attention weight vector generation using multi-level features in detail. Figure 3(a) shows the multi-level features concatenation process. Firstly, $Conv3\_4$ is convolved by a convolutional layer with kernel size of 1×1 and channel number of 1. Then, a max pooling is performed with a stride size of 2. After applying a convolutional layer with kernel size of 3×3 and channel number of 512, we obtained $Conv3\_a$. Since the scale of $Conv2\_3$ is twice that of $Conv3\_4$, convolution and max pooling is performed twice for $Conv2\_3$. Then, $Conv2\_b$ is obtained through a convolutional layer with kernel size of 3×3 and channel number of 256. Both $Conv2\_b$ and $Conv3\_a$ have the same scale as $Conv4\_6$. Finally, we concatenated $Conv3\_a$ and $Conv2\_b$ with $Conv4\_6$. We define the concatenation as $E = [E^1, E^2, \ldots, E^{1792}]$, which has the shape of $1 \times W \times H \times 1792$. W and H is the width and height of feature maps, and 1792 is the number of channels. E is convolved by a convolutional layer with kernel size of 3×3 and channel number of 1024 before generating the attention vector for $Conv4\_6$. We use C to represent the channel number 1024. The channel-wise attention weight vector A^c is obtained as follows:

$$M^c = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} E_{ij}^c \tag{7}$$

$$A^c = softmax(M^c) \tag{8}$$

M^c is the average pooling of E in the spatial dimensions. The length of M^c is the same as the number of $Conv4\_6$ channels 1024. The channel-wise attention weight vector A^c is generated after M^c passes the softmax layer. The feature maps $Att1$ as the input of RPN are generated as follows:

$$Att1 = Conv4\_6 \times (A^c + 1) \tag{9}$$

where $Conv4\_6$ represents the feature maps of the $Conv4\_6$ layer.

IV. EXPERIMENTS AND RESULTS

Before showing the results, we first describe our dataset and implementation details.

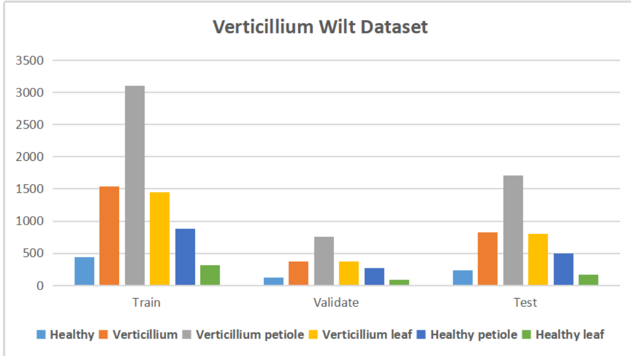


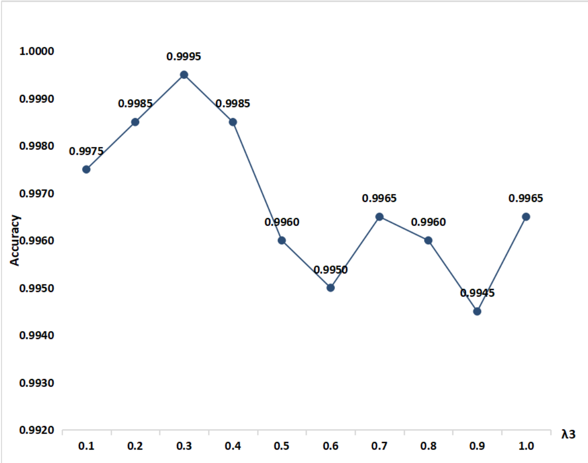
FIGURE 4. **The number of each category in verticillium wilt dataset.**

A. VERTICILLIUM WILT DATASETS

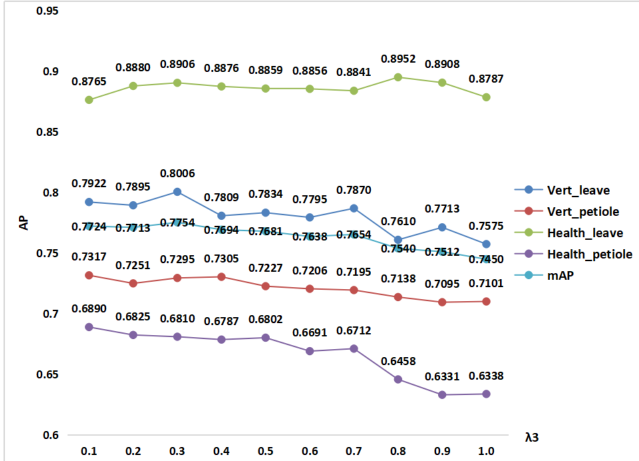
To our knowledge, there are no public strawberry datasets with manual annotations for verticillium wilt. Therefor, we trained and evaluated our proposed network on verticillium wilt datasets that we constructed. Our verticillium wilt dataset consists of 3, 531 images. Using smartphones, images were collected in different strawberry growing greenhouses. To increase the diversity of data, we took photos from different angles for one plant. According to the demand, the four categories Healthy_leaf, Healthy_petiole, Verticillium_leaf and Verticillium_petiole were manually annotated, each image also was labeled as either healthy or verticillium wilt. The image examples are shown in Figure 2. The verticillium wilt dataset contains a total of 3, 531 images with 789 images of healthy samples and 2, 742 images of verticillium wilt samples. The resolution of all processed images is $1,024 \times 768$ or $768 \times 1,024$. The main advantage of our dataset is that we collected one plant samples from three different angles. Unlike other datasets, we take photos for the whole plant in different backgrounds rather than for the individual leaf in the same background. The main challenge of our dataset is the imbalance of samples in each category, as shown in Figure 4. The strawberry verticillium wilt dataset is available at *URL : https : //github.com/WanlgLuYao/Strawberry_wilt_dataset*

B. IMPLEMENTATION DETAILS

We used ResNet-50 [9] as the backbone network in our experiments and set the shortest side pixel size of the input image to 768. Our models based on Faster R-CNN were end-to-end trained on a NVIDIA GTX1080(8G) GPU for 60, 000 iterations with an initial learning rate of 0.001, which was then divided by 10 at 10, 000, 30, 000 and 50, 000 iterations. Following Faster-RCNN [12] parameter setting, three aspects ratios 1 : 2, 1 : 1, 2 : 1 [12] and three anchor scales 8, 16, 32 [12] were set to cover objects of different shapes, pooling size was set to 7 [12]. Same as [12], we set the RPN batch size as 256, the IoU(intersection over union) threshold for NMS as 0.7, the RPN positive and negative IoU(intersection over union) threshold as 0.7 and 0.3, weight decay and momentum as 0.0001 and 0.9 respectively. Object detection



(a)



(b)

FIGURE 5. (a) is the verticillium wilt detection task accuracy of SVWDN using different values λ_3. (b) is the object detection AP of different category and mAP of the SVWDN using different values λ_3.

TABLE 1. All set parameters for SVWDN.

GPU	NVIDIA GTX1080(8G)	Initial learning rate	0.001
Batch size	1	Learning rate	0.0001(1w), 0.00001(3w), 0.000001(5w)
RPN batch size	256	Training step	60,000
Momentum	0.9	Shortest side pixel size	768
Weight decay	0.0001	Loss weighting factors λ_1, λ_2, λ_3	1.0, 1.0, 0.3
RPN positive IOU	0.7	Anchor ratios	1:2, 1:1, 2:1
RPN negative IOU	0.3	Anchor scales	8, 16, 32
RPN NMS threshold	0.7	Pooling size	7

mAPs of 77.93%, 78.05% and 77.68% were achieved when we set loss weighing factors λ_1 [12] as 0.1, 1.0, 10 respectively. Therefore, we decide to apply 1.0 for λ_1 on SVWDN. According to [12], the loss weighting factors λ_2 was set to 1.0. In our SVMDN, a new loss weighting factor λ_3 was introduced for the verticillium wilt detection task. We set λ_3 as 0.3 experimentally. The results of applying different values of λ_3 are shown in Figure 5. In Figure 5(a), the verticillium wilt detection task achieves maximum accuracy (99.95%) when we set λ_3 as 0.3. When λ_3 was set as 0.3, the leaves and petioles detection task also has the best mAP (77.54%), as shown in Figure 5(b). To increase the training samples, the verticillium wilt detection was trained with image flipping. The parameters used in our model are shown in Table 1. Our model was evaluated on the verticillium wilt dataset.

C. EVALUATION METHOD

We performed experiments in two groups. In each group, our dataset was randomly divided into the 56% training set, 14% validation set, and 30% testing set. The training set, validation set and testing set were 1,976, 495 and 1,060. Then, the average of the two results was treated as the final result. The validation set was used to minimize overfitting and was a typical way to stop the network from learning. The training and validation sets were used to perform the training process and parameter selection, respectively. The testing set was used to evaluate the results on unknown data. We first trained Faster R-CNN with ResNet-50 as the baseline on our verticillium wilt dataset. Then, we evaluated our proposed network in two parts, leaves and petioles detection and sample classification (i.e whether the strawberry has verticillium wilt). We used mean average precision (mAP) with an IoU threshold of 0.7 and average precision to evaluate the leaves and petioles detection performance and sample classification performance, respectively. After we added the disease detection task in Faster R-CNN, the DDN performed well on the sample classification task with an accuracy of 99.85%, but the mAP of the DDN object detection decreased compared with the original. After adding the channel-wise attention model using multi-level features on the DDN, both the leaves and petioles detection task and the disease detection task performed well on SWVDN.

D. RESULTS DISCUSSION

Our results are shown in Table 2 and the detection examples are shown in Figure 6. The mAP of Faster R-CNN object detection on our strawberry verticillium wilt dataset was 78.05%. The AP of petioles detection was lower than that of leaves. Compared with leaf, the petiole is a slender object. When performing convolution, petioles only occupy a thin

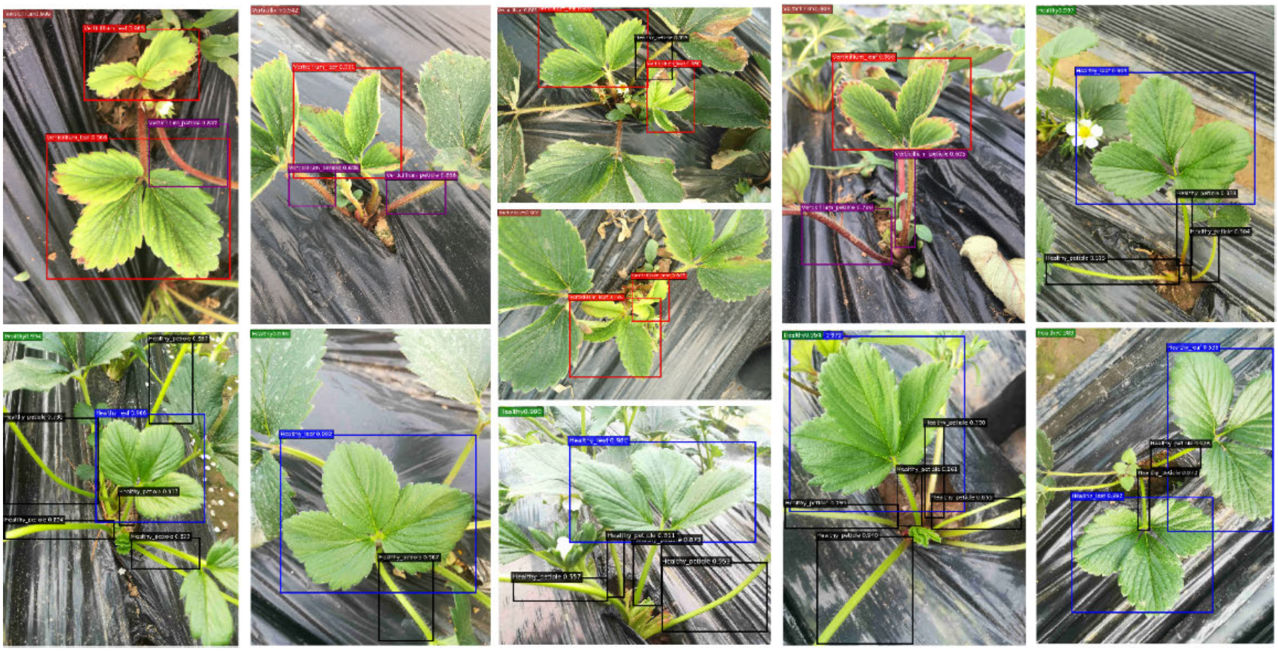


FIGURE 6. **Detection examples on the verticillium wilt dataset with SVWDN. Each color corresponds to an object category. We use the brown mark for verticillium wilt in the upper left corner, and the green mark is healthy. The blue box denotes healthy leaves, the black box denotes healthy petioles, the red box denotes verticillium wilt leaves, and the purple box denotes verticillium wilt petioles.**

TABLE 2. **The detection result on our dataset.**

Method	Verticillium petiole	Verticillium leaf	Healthy petiole	Healthy leaf	mAP	Accuracy
Faster R-CNN	0.7273	0.7978	0.6971	0.8996	0.7805	
DDN	0.7307	0.7987	0.6704	0.8871	0.7717	0.9985
SVWDN	0.7295	0.8006	0.6810	0.8906	0.7754	0.9995

strip in a kernel with size of 3×3. This might make that the feature map can not represent the characteristics of a petiole. After adding the disease detection task in Faster R-CNN, the DDN achieved 99.85% accuracy in the task of classifying whether the strawberry is suffering from verticillium wilt. However, the mAP of DDN object detection decreased to 77.17% and decreased by 0.88%. The AP of object detection on 3 classes Healthy_leaf, Healthy_petiole, Verticillium_leaf decreased. When adding a disease detection task to Faster R-CNN, training process needs to consider both tasks and tune the network parameters towards an overall satisfactory. This might make the network parameters of leaves and petioles detection part are not as good as before. Therefore, the mAP of object detection was decreased. Adding channel-wise attention of the using multi-level features to the DDN, the accuracy of SVWDN was 99.95% on the task of classifying whether the strawberry has verticillium wilt, and the mAP of SVWDN on the object detection task increased to 77.54%. Compared with the DDN, the object detection AP of 3 classes Healthy_petiole, Healthy_petiole, Verticillium_leaf increased, and mAP increased by 0.37%. The SVWDN achieved 99.95% accuracy for determining whether the strawberry is suffering from verticillium wilt without reducing the performance too much in the object detection task. The possible reason that generating channel-wise attention weight vector using multi-level features improved the detection performance is that the attention based method benefits the detection and classification task through setting a larger value for the favorable feature maps.

V. CONCLUSION

SVWDN based on Faster-RCNN was proposed for strawberry verticillium wilt detection, which detects whether a strawberry has verticillium wilt according to the detected components (i.e.,leaves and petioles) and achieved 99.95% accuracy for verticillium wilt detection. More importantly, we constructed a large dataset for strawberry verticillium wilt. To our knowledge, we are the first to use the detection network to determine whether strawberries have verticillium wilt through the multiple detected components. We will attempt to add other disease samples to our dataset and welcome researchers to use our dataset for plant disease detection.

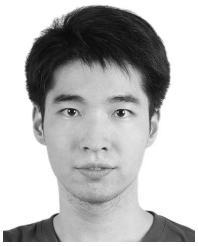
REFERENCES

[1] D. C. Harris, "Control of verticillium wilt and other soil-borne diseases of strawberry in britain by chemical soil disinfestation," *J. Horticultural Sci.*, vol. 65, no. 4, pp. 401–408, 1990.

[2] R. T. Mckinley and P. W. Talboys, "Effects of *Pratylenchus penetrans* on development of strawberry wilt caused by *Verticillium dahliae*," *Ann. Appl. Biol.*, vol. 92, no. 3, pp. 347–357, 2010.
[3] Y. Ebihara, S. Uematsu, and S. Nomiya, "Control of *Verticillium dahliae* at a strawberry nursery by paddy-upland rotation," *J. Gen. Plant Pathol.*, vol. 76, no. 1, pp. 7–20, 2010.
[4] J. E. Munyaneza, J. M. Crosslin, J. L. Buchman, and V. G. Sengoda, "Susceptibility of different potato plant growth stages to purple top disease," *Amer. J. Potato Res.*, vol. 87, no. 1, pp. 60–66, 2010.
[5] O. Batuman and R. L. Gilbertson, "Emerging viral and other diseases of processing tomatoes: Biology, diagnosis and management," *Acta Horticulturae*, vol. 971, no. 6, pp. 35–48, 2013.
[6] S. R. Maniyath, Vinod P V, Niveditha M, Pooja R, Prasad Bhat N, Shashank N, and R. Hebbar, "Plant disease detection using machine learning," in *Proc. Int. Conf. Design Innov. 3Cs Comput. Commun. Control (ICDI3C)*, Apr. 2018, pp. 41–45.
[7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2005, pp. 886–893.
[8] F. Alvaro, Y. Sook, K. Sang, and P. Dong, "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition," *Sensors*, vol. 17, no. 9, p. 2022, 2017.
[9] K. He, X. Zhang, S. Ren, and S. Jian, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
[11] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Lecun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," 2013, *arXiv:1312.6229*. [Online]. Available: https://arxiv.org/abs/1312.6229
[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," 2015, *arXiv:1506.01497*. [Online]. Available: https://arxiv.org/abs/1506.01497
[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2015, *arXiv:1506.02640*. [Online]. Available: https://arxiv.org/abs/1506.02640
[14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
[15] J. Mao, T. Xiao, Y. Jiang, and Z. Cao, "What can help pedestrian detection?" in *Proc. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 6034–6043.
[16] D. Xu, W. Ouyang, E. Ricci, X. Wang, and N. Sebe, "Learning cross-modal deep representations for robust pedestrian detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5363–5371.
[17] X. Zhang, Y. Yuan, and Q. Wang, "ROI-wise reverse reweighting network for road marking detection," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*. Newcastle, U.K.: Northumbria Univ., Sep. 2018, p. 219.
[18] S. Ruder, J. Bingel, I. Augenstein, and A. Søgaard, "Sluice networks: Learning what to share between loosely related tasks," *ArXiv*, vol. abs/1705.08142, 2017.
[19] S. Ruder, "An overview of multi-task learning in deep neural networks," 2017, *arXiv:1706.05098*. [Online]. Available: https://arxiv.org/abs/1706.05098
[20] R. Ranjan, V. M. Patel, and R. Chellappa, "HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 121–135, Jan. 2016.
[21] D. Xu, W. Ouyang, X. Wang, and N. Sebe, "PAD-Net: Multi-tasks guided prediction-and-distillation network for simultaneous depth estimation and scene parsing," 2018, *arXiv:1805.04409*. [Online]. Available: https://arxiv.org/abs/1805.04409
[22] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Simultaneous detection and segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 297–312.
[23] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," 2017, *arXiv:1704.06904*. [Online]. Available: https://arxiv.org/abs/1704.06904
[24] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," 2016, *arXiv:1611.05594*. [Online]. Available: https://arxiv.org/abs/1611.05594
[25] H. Zhao, Y. Zhang, S. Liu, J. Shi, C. C. Loy, D. Lin, and J. Jia, "PSANet: Point-wise spatial attention network for scene parsing," in *Proc. 15th Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, Sep. 2018, pp. 267–283.
[26] X. Zhang, T. Wang, J. Qi, H. Lu, and G. Wang, "Progressive attention guided recurrent network for salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, Jun. 2018, pp. 714–722.
[27] X. Chu, W. Yang, W. Ouyang, C. Ma, A. L. Yuille, and X. Wang, "Multi-context attention for human pose estimation," in *Proc. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1831–1840.
[28] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," 2016, *arXiv:1603.06937*. [Online]. Available: https://arxiv.org/abs/1603.06937
[29] P. R. Rothe and R. V. Kshirsagar, "Cotton leaf disease identification using pattern recognition techniques," in *Proc. Int. Conf. Pervas. Comput.*, Jan. 2015, pp. 1–6.
[30] Y. Tian, C. Zhao, S. Lu, and X. Guo, "SVM-based multiple classifier system for recognition of wheat leaf diseases," in *Proc. World Automat. Congr.*, 2012, pp. 189–193.
[31] A. Rastogi, R. Arora, and S. Sharma, "Leaf disease detection and grading using computer vision technology & fuzzy logic," in *Proc. Int. Conf. Signal Process. Integr. Netw.*, 2015, pp. 500–505.
[32] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Comput. Electron. Agricult.*, vol. 145, pp. 311–318, Feb. 2018.
[33] A. Dhakal and S. Shakya, "Image-based plant disease detection with deep learning," *Int. J. Comput. Trends Technol.*, vol. 61, pp. 2231–2803, Jul. 2018.
[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[35] T. Y. Lin, P. Dollár, R. Girshick, K. He, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2016, pp. 2117–2125.

XUAN NIE was born in 1976. He received the B.S. degree in automatic control, the M.S. degree in pattern recognition, and the Ph.D. degree in computer application technology from Northwestern Polytechnical University, Xi'an, China, in 1998, 2001, and 2005, respectively. He joined the School of Software, Northwestern Polytechnical University, in 2006, as a Lecturer, where he has been an Associate Professor, since 2010. He was a Visiting Professor with The Hong Kong Polytechnical University, in 2010 and with the University of Michigan, USA, in 2013. He has authored and coauthored over 30 journal and conference papers, three monographs and co-invented patents. His main research interests include machine learning, computer vision, image processing, and artificial intelligence. He is the honor of the Science and Technology Achievement Award of Xi'an 2015. He was a Reviewer for the IEEE Internet of Things Journal.

LUYAO WANG was born in 1996. He received the B.S. degree in software engineering from the Xi'an University of Architecture and Technology, China, in 2018. He is currently pursuing the M.S. degree with Northwestern Polytechnical University, Xi'an, China, since 2018. His current research interests include deep learning, and object detection and their applications in plant disease detection.



HAOXUAN DING was born in 1995. He received the B.S. degree in flight vehicle propulsion engineering from Northwestern Polytechnical University, Xi'an, China, in 2018, where he is currently pursuing the M.S. degree. His current research interests include the generative adversarial networks, object detection, and industrial applications.

MIN XU received the B.E. degree from the University of Science and Technology of China, the M.S. degree from the National University of Singapore, and the Ph.D. degree from the University of Newcastle, Australia. She is currently an Associate Professor with the School of Electrical and Data Engineering, Faculty of Engineering and IT, University of Technology Sydney. She has published over 100 research articles in prestigious international journals and conferences. Her research interests include multimedia data analytics and computer vision. She is an Associate Editor of the *Journal of Neurocomputing* (Elsevier). She has been invited to be a member of the program committee for many international conferences.