



International Symposium on Imaging, Sensing, and Optical Memory (ISOM '20),
Takamatsu, Japan

Development of health monitoring method for pecan nut trees using side video data and computer vision

Ryota Nomura¹ · Kazuo Oki^{1,2}

Received: 31 March 2021 / Accepted: 16 August 2021 / Published online: 27 August 2021
© The Author(s) 2021

Abstract

Increasing efficiency and productivity in the field of agriculture is important to provide sufficient food to the world's increasing population. It is important to monitor crops using image processing in order to realize these increases in efficiency and productivity. In order to monitor crops with high quality and accuracy, high resolution images are needed. In this research, a crop monitoring method for pecan nut trees was developed using high-resolution video taken from the side of a vehicle. First, trees were extracted by applying an object detection model to the video data. Second, the extracted trees were divided into canopy and trunk areas. Finally, using labels made by experts and the canopy image as input, the convolutional neural network (CNN) model was trained to classify unhealthy and healthy trees. The model achieved an area under the curve for classification over 0.95. Gradient-weighted Class Activation Mapping (Grad-CAM) was also applied to the model for the purpose of evaluation, and it clarified that the model is focusing on the hollow features of the canopy when performing its classification.

Keywords Disease detection · Computer vision · Object detection · Video processing

1 Introduction

Along with recent progress in the field of machine learning and computer vision, there have been many attempts to also apply these technologies in the field of agriculture [1]. One of the applications is disease detection [2]. Detecting disease at an early stage is important for financial reasons and for food security. Application of these technologies to disease detection can increase efficiency and reduce human labor, especially for wide farmlands where labor costs are high.

Unmanned aerial vehicle (UAV) is one of the methods used for crop monitoring, and it is able to cover a moderate range with high resolution. Previous research used UAV images to detect disease in citrus trees [3]. However, higher quality monitoring can be achieved using images with higher

resolution. Some researchers used multi-spectral imagery from UAV to monitor forest health [4]. Other researchers used hyperspectral and thermal imagery to detect disease in olive trees [5]. However, these multi-spectral and hyperspectral cameras are expensive and difficult to actually apply in practice.

Some previous studies used RGB video data for disease detection. Previous research used RGB video for maize detection [6]. Other researchers also used RGB video and its color features to distinguish cauliflower from weeds [7]. Both of these methods used video data from above, which can be difficult to use for fruits or nuts since they are usually covered by leaves. Also, disease recognition using video data has room for improvement.

Other previous researches used IoT devices to collect sensor data in higher resolution, including images. One research group collected sensor data such as temperature, image, humidity, and integrated to classify disease using Raspberry PI [8]. Other research group used segmentation and Support Vector Machine to detect disease affected leaves from high resolution images collected by IoT devices [9]. High resolution images which leads to sophisticated applications can be

✉ Ryota Nomura
ryota15765@gmail.com

¹ Institute of Industrial Science, The University of Tokyo,
4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, Japan

² Kyoto University of Advanced Science, Kameoka, Japan

collected using IoT devices, but the cost for application will be expensive to cover a wide area.

In this research, a crop monitoring method was developed using RGB video data taken from the side of the crops from a vehicle. This data have four advantages.

One is that it has higher resolution compared with UAV images since the distance between the camera and target is smaller. This is because UAV needs to keep constant distance from the canopy top to fly safely. Also, considering that the tree height differs by trees, the distance between UAV and canopy top becomes large for most of the trees. On the other hand, lateral view from vehicles has less safe distance and does not crash even if the vehicle hits the branch, and the difference of canopy thickness between trees are smaller compared to the difference of canopy height, which results in higher resolution image.

The second advantage is especially for plants such as fruits. When leaves grow over the plant to maximize photosynthesis, detecting features other than leaves such as the fruit size or color is difficult using images taken from above, while the lateral video data used in this research can provide such information. Also, detection of canopy is difficult using images taken from above, since the color of background is usually similar to that of the canopy (see Fig. 1).

The third is the affordability of the equipment. This data can be collected using smartphones or any RGB camera equipment, which is available at a lower cost compared to UAVs. Also, the collection is easier to actually implement.

The fourth advantage is to make use of video data instead of single image data. Segmentation of videos has higher accuracy compared to segmentation to images, since features of near frames, such as, objects will be in similar location with similar shapes between adjacent frames, can be used. Disease classification can also benefit from using videos. Using single image, the prediction is made once, but using videos, frames that take the same tree from different angles can be used as prediction, and ensembling the predictions can result in higher accuracy for the classification. Also,

using video data can enhance easiness and sureness for data acquisition. Selective acquisition for images using location data can also be a choice, but setting this location hyperparameters can be a cost for application and also can lead to miss-acquisition by the reliability of hyperparameters. Using these lateral RGB video data, a new method for disease detection is proposed.

2 Data and location

In this research, the proposed method was applied to pecan nut trees. Pecan nuts are grown for use as food and contain abundant nutrition. Xylella is one of the diseases that affects pecan trees. The leaves of trees affected by xylella become small and drop, and the canopy becomes hollow as shown in Fig. 2. However, the progression of xylella is slow, and it can be treated successfully with early detection and appropriate care.

The data used in this research was collected in White-tail Creek Orchard in Arizona, USA on 2019/10/12 and 2019/10/17. Images of the orchard are shown in Fig. 2. Pecan nut trees were planted in this area in 74 columns. Each column had approximately 120 trees. In this orchard, for the early detection of xylella, pink marks are put on the trunks by experienced farmers to identify unhealthy trees. These trees are candidates to develop disease, and a chemical inspection is conducted on them. However, this label data can be improved in future researches using the results of chemical inspection on trees, conducted at a near date from the date of video acquisition. Video was collected with a GoPro7 at a resolution of 1,920*1,440. The video was taken at 30 fps. For analysis, frames were chosen after being sampled at 3 fps (1/10).



Fig. 1 Comparison between examples of (left) image taken from above and (right) lateral image

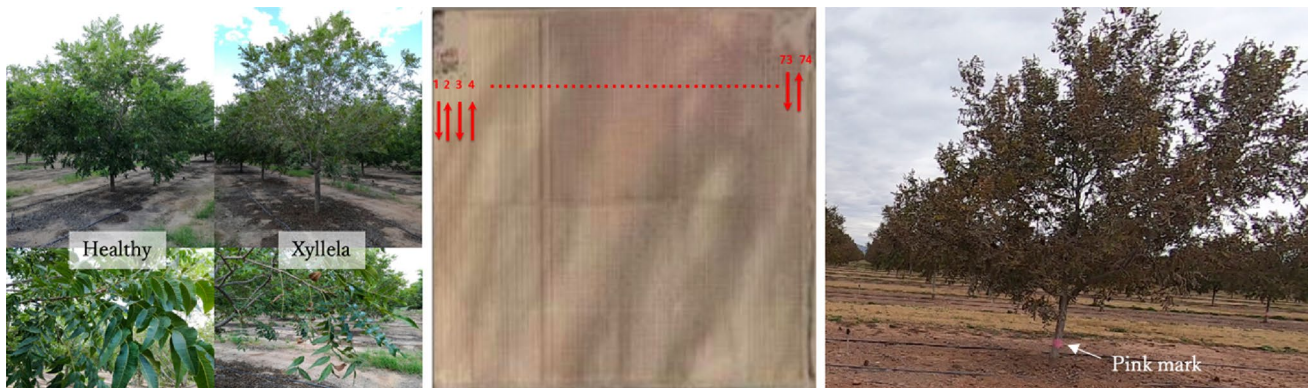
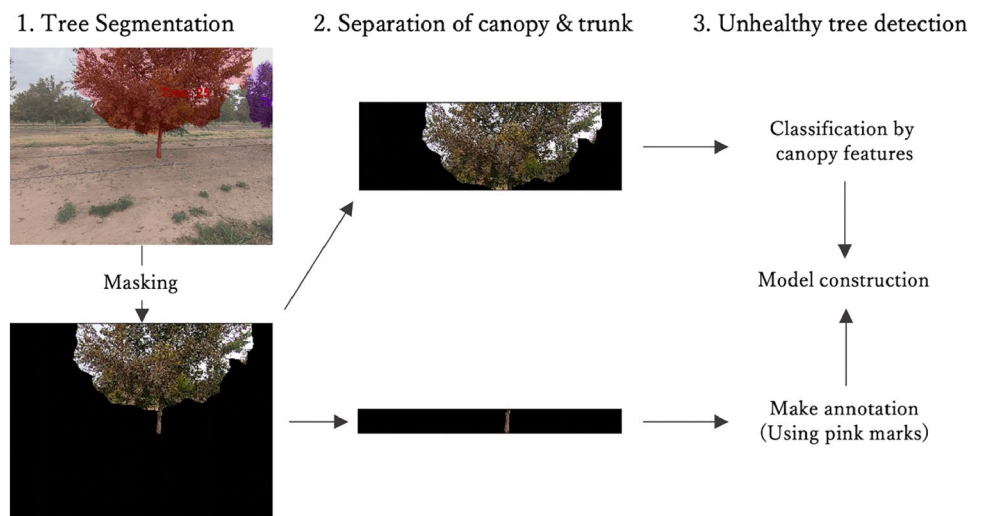


Fig. 2 (left) Comparison of healthy trees and trees with xylella. (middle) Ortho image of the orchard from 150 m above. Light colored black dots represent the trees. (right) Pink marks on planted tree

Fig. 3 Flowchart of unhealthy tree detection method



3 Method

Figure 3 show a flowchart of the method developed in this research.

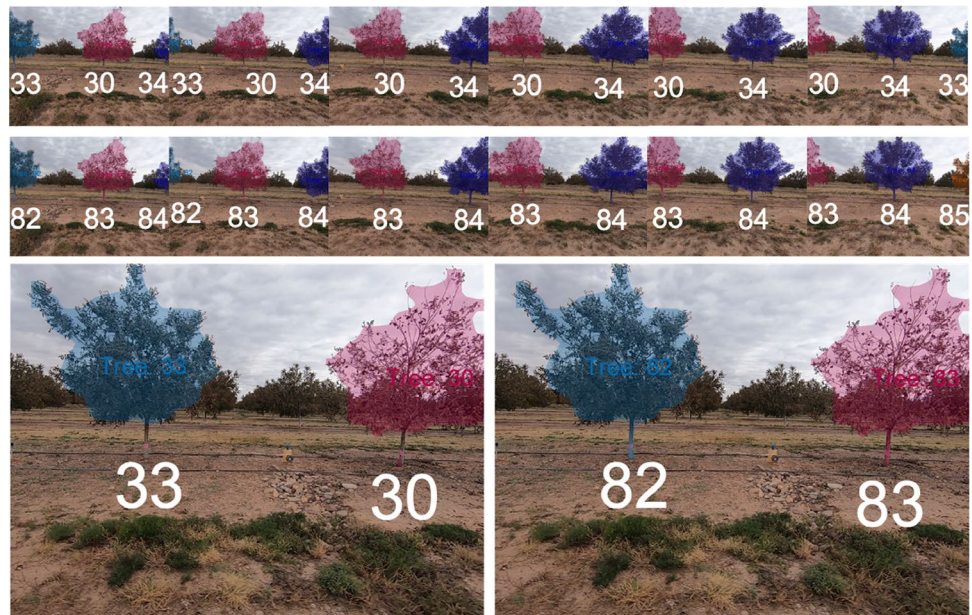
3.1 Detection of trees using TrackR-CNN

TrackR-CNN [10] is a state-of-the-art of object detection and tracking from video data. This is a supervised training model that detects objects by instance segmentation. This model is able to track objects between frames and distinguish them by detecting with different IDs.

This model was applied in several steps. First, the model was pre-trained using the KITTI MOTs dataset [11]. The KITTI MOTs dataset is a video dataset which consists of 5027 frames of training data and 2,981 frames of test data. This is a dataset for autonomous driving and includes annotation for pedestrian and car masks. This dataset consists of images taken outside with various light conditions, which

matches the application for pecan images. The model was pre-trained for 5 epochs using this dataset. Next, this model was fine-tuned, and tree masks were detected by applying this model. Only the structure of the output layer was modified for the fine-tuning. Four out of the 74 columns were used for training. These columns consisted of 7033 frame images. Of these, 210 frame images from 4 sequences were sampled and used for the training. However, the output of this model had two problems. The first was ID repeating. When trees fade out from one side of frame, the next tree appearing at the other side of the frame was recognized as the same tree. To overcome this problem, the rule “Trees only move in one direction” was set. The second problem was misdetection of the trunk part of the tree. Since the trunk area is smaller compared with the canopy, the penalty for the model for mis-detecting trunks was small. To overcome this, another TrackR-CNN model for detecting only trunks was trained, and the outputs were merged as the final output. The detection results obtained before and after this modification are shown in Fig. 4. The trunk area

Fig. 4 (top) Series of detection before and after modification. White labels shows the id of the detected trees. (bottom-left) First output of TrackR-CNN. (bottom-right) Output after modification



was detected more clearly after the modification. Also, miss-detection of IDs can be seen from detection data shown in series. Trees with same ID (shown in same colors) appear multiple times.

The performance of this model was evaluated using the first column of trees (82 trees). For training, 146 frames (15 trees) were used, and 608 frames (67 trees) were used for the test. The detection accuracy was 98 in F number (shown in formula below), and the ID accuracy, which is the ratio of tree IDs tracked correctly for all frames, was 89%.

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Fnumber = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

3.2 Separation of canopy and trunk

For detected tree masks, the trunk and canopy were separated. The intention of this step was to extract the canopy for to input the classification model since the features of unhealthy trees are mainly seen in the canopy. The algorithm for the separation of the trunk and canopy area is as follows:

- For the tree mask, the rows of mask were iteratively looked up from the bottom.
- If the pixels in the row split into multiple groups or if the range of pixel in the row becomes wider than a hyper-

parameter (set to 60 pixels in this research), the upper area is considered to be canopy.

This pixel-based method relies on an experimental hyperparameter which can differ by image resolution or the distance between camera and trees. The method in this section can be replaced by learning based approach to increase generality.

3.3 Detection of unhealthy trees

Using the canopy image extracted in the previous step as the input and the pink marks as correct label, the CNN model was trained to classify healthy and unhealthy trees.

Videos for five different rows recorded under different weather and light conditions were used for this step, shown in Fig. 5. Four columns were used for training and one column was used for evaluation to prevent data leakage. The dataset information is shown in Table 1. The dataset included 712 different trees. Since multiple canopy images of each tree were taken from different angles to provide the video data, 9225 canopy images were extracted in total. Thirteen canopy images were extracted per tree on average. Predictions were made for each canopy image, and the average value was used (ensembled) to make the final prediction.

This dataset is unbalanced since unhealthy trees make up a minority of the total. The loss due to mis-predicting unhealthy trees was set to be three times greater to prevent the model from only making predictions of healthiness. Setting the loss balance to an integer less than 3 resulted in predicting all of the trees to be healthy. Figure 6 shows



Fig. 5 Snapshot from each video. Videos with blue frames show train data and those with red frames show test data

Table 1 Information regarding created dataset

Index	Value
Tree	712
Canopy image	9225
Average image per tree	13.0
Healthy tree	655 (92%)
Unhealthy tree	57 (8%)

the basic information of the model. Activation by rectified linear unit (ReLU) was applied after every convolutional layer and the first dense layer after convolutional network (dense_1 in Fig. 6). The dropout rate was set to 0.7 for dropout_1 and dropout_2 and to 0.8 for dropout_3 to prevent overfitting.

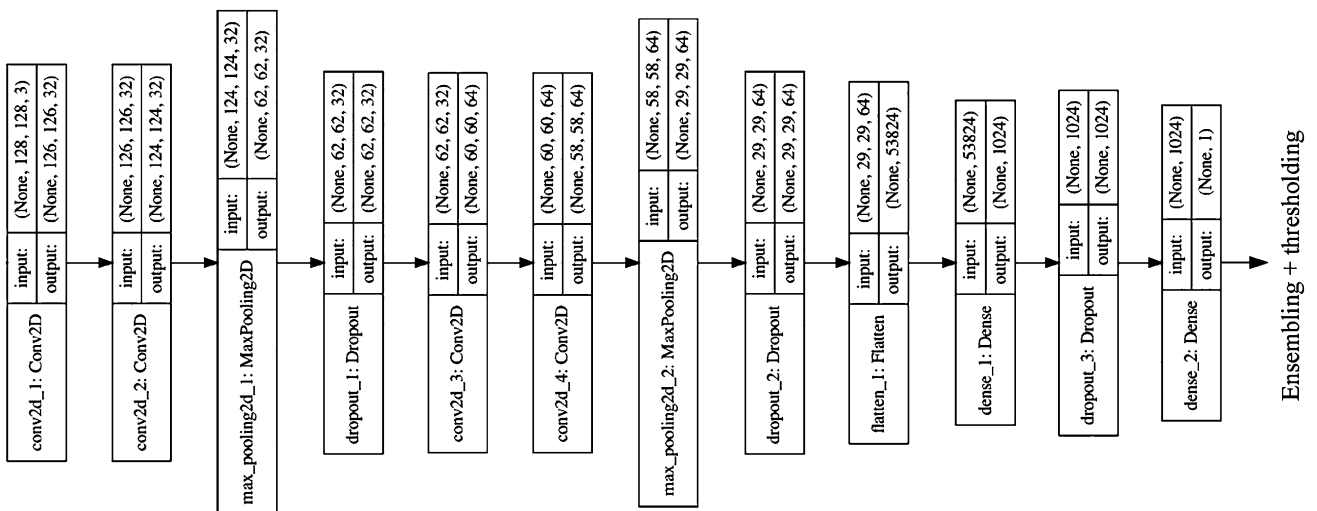


Fig. 6 Model architecture for classification

3.4 Model evaluation by Grad-CAM

Grad-CAM: Gradient-weighted Class Activation Mapping [12] is widely used to visualize explanations for model predictions and for evaluation. This method can visualize which part of the input image is contributing to the classification by visualizing the distribution of the activation in convolutional layers. This was applied to the bottom layer of the model trained in step 3.

4 Result and discussion

4.1 Detection of unhealthy trees

Figure 7 shows the results of the classification model. The area under the curve (AUC) was 0.95 for binary classification. From the histogram for prediction, the threshold could be set around 0.2. However, this threshold should be determined by considering the type of disease and the cost of a chemical investigation. Devastating disease should be detected by small false negative while extracting more trees leads to a more costly investigation.

Images in Fig. 7 shows mis-predicted trees. Mis-predicted pink-marked trees seem to have more leaves and to be healthy while mis-predicted healthy trees have fewer leaves and seem to be unhealthy. The pink marks were put on the trees a few months before this video was made might be the reason for this result. Also, this result suggests the possibility that the accuracy might be increased using statistical methods rather than human eyes, as the latter relies on experience and intuition and can be inconsistent.

4.2 Evaluation by Grad-CAM

Figure 8 shows a visualization from Grad-CAM [12]. By examining the bottom layer of the convolutional model (conv2d_4 in Fig. 6), the focus of the model can be visualized by this method. The middle mask represents the distribution of the activated weights on the target layer when the left image was input to the model. The area displayed in red in the right combined image represents the main focus of the model. The model is likely to be focusing on the hollow features of trees correctly. By making use of this method for different images combined with label data such as the chemical investigation results, this method can also lead to a

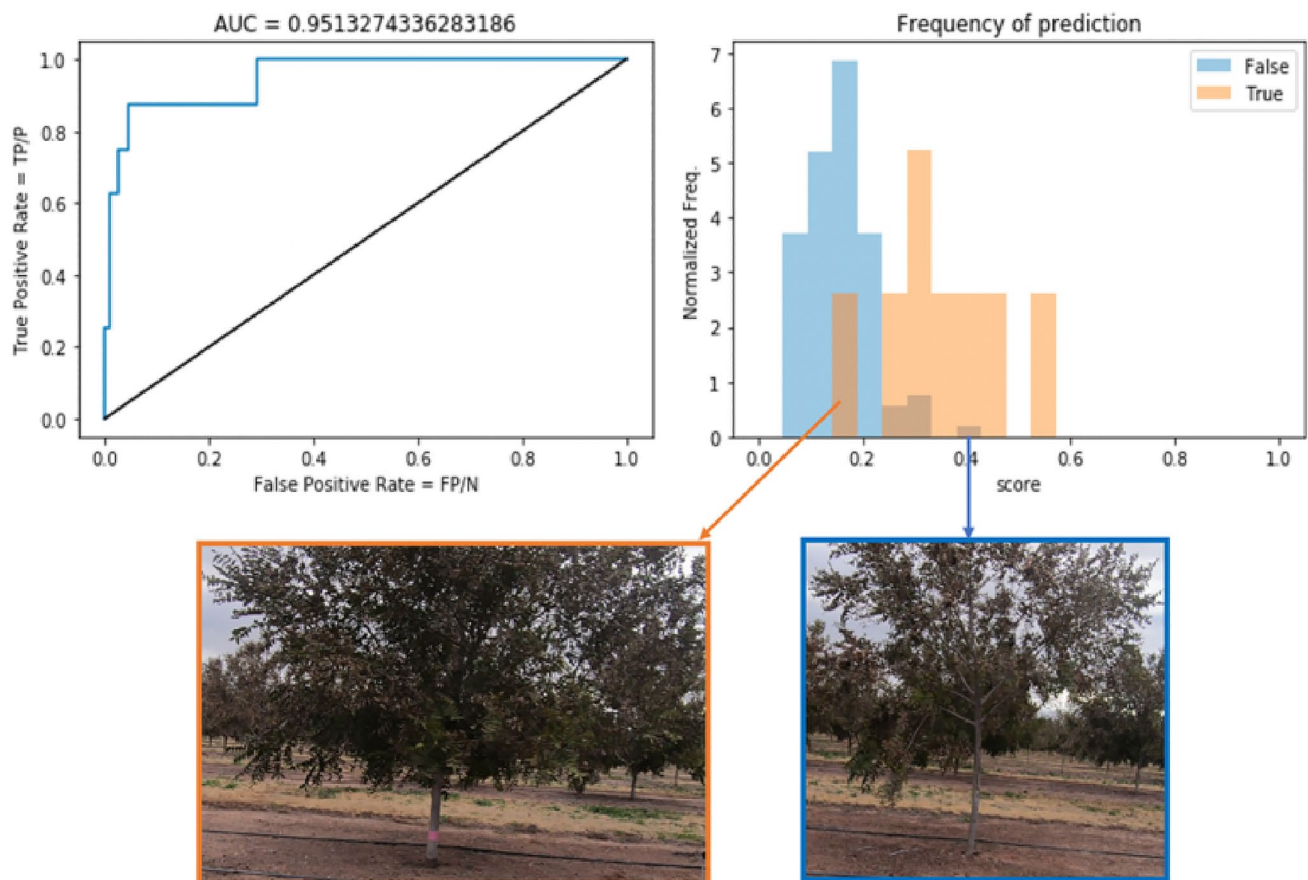


Fig. 7 (above) Classification result in AUC and histogram. (bottom) Sample image for miss-classified tree

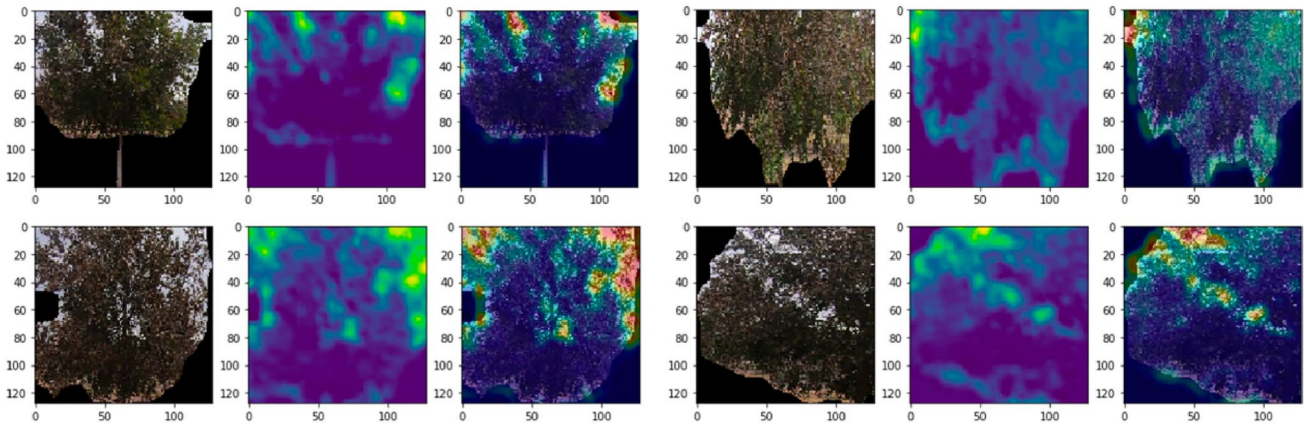


Fig. 8 Results for Grad-CAM evaluation. (left) Original image. (middle) Output from Grad-CAM. (right) Combined image

new index or focus point for detecting diseased or unhealthy trees.

The proposed method consists of two models, instance segmentation model and classification model, for the disease detection. However, given that the latter CNN model focus on the hollow features of the canopy images, it is difficult to quantitatively define the effect of the former model performance to the latter model since same loss to canopy area and trunk area can have different effect. However, performance of the former model can be more important since the output of this model will be the base for the latter model, and linkage of the trees to the trees in the actual field will also rely on the performance of the former model.

5 Conclusion

Early detection of disease is important for food security and to protect farmers' profits. Using computer vision for monitoring is essential to increase productivity and efficiency. However, establishing a method for detecting disease in crops such as fruits or nuts is challenging since they are difficult to see from above. Also, methods using multi-spectral cameras are high in cost and difficult to apply in practice.

In this research, a method of detecting unhealthy trees using video data obtained from the side rather than from above was developed. Side video data has a high resolution compared to UAV images, and features such as fruits, branches, and canopy can be extracted from the side. Also, the device to record RGB video is used in this method and is low in cost. First, instance segmentation model was applied to video data for tree detection. Detection accuracy was 98 in F number. Next, canopy area was extracted using predicted instance segmentation masks. Finally, a CNN classification model was trained using canopy images as the input and the

classification of healthy and unhealthy trees as the output. The model achieved an AUC of 0.95. Also, from an evaluation made using Grad-CAM, the trained CNN model seems to be focusing correctly on hollow features of the canopy. This method can also be applied to other trees and crops. Also, by changing the frame sampling rate from the videos, this method can be applied to video taken at various speeds.

In future research, diseases other than xylella can also be classified by providing annotation data for training. Also, using visualizing technologies such as Grad-CAM, currently unknown features that can distinguish healthy from disease-affected trees can be discovered. Also, using deeper networks, disease classification can be made more accurate. Through cost–benefit analysis, the benefits of introducing this disease detection model can be clarified.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. D. Patrício and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for graincrops: A systematic review," *Computers and Electronics in Agriculture*, 2018.
2. M. Sharif, M. Khan, Z. Iqbal, M. Azam, . M. Lali and M. javed, "Detection and classification of citrus diseases in agriculture

- based on optimized weighted segmentation and feature selection," *Computers and Electronics in Agriculture*, 2018.
3. B. Garza, V. Ancona, J. Enciso, H. Perotto-Baldivieso, M. Kunta and C. Simpson, "Quantifying Citrus Tree Health Using True Color UAV Images," *Remote Sensing for Plant Pathology*, 2019.
 4. J. Dash, M. Watt, G. Pearse, M. Heapy and H. Dungey, "Assessing very high resolution UAV imagery for monitoring forest health during a simulated disease outbreak," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2017.
 5. R. Calderon, J. Navas-Cortés, C. Lucena and P. Zarco-Tejada, "High-resolution airborne hyperspectral and thermal imagery for early detection of Verticillium wilt of olive using fluorescence, temperature and narrow-band spectral indices," *Remote Sensing of Environment*, 2013.
 6. D. Shrestha, B. Steward and S. Birrell, "Video Processing for Early Stage Maize Plant Detection," *Biosystems Engineering*, 2004.
 7. E. Hamuda, B. McGinley, M. Glavin and E. Jones, "Automatic crop detection under field conditions using the HSV colour space and morphological operations," *Computers and Electronics in Agriculture*, 2017.
 8. A. Thorat, S. Kumari and N. D. Valakunde, "An IoT Based Smart Solution for Leaf Disease Detection," *2017 International Conference on Big Data, IoT and Data Science (BIG-IoT)*, 2017.
 9. S. A. Nandhini, R. Hemalatha, S. Radha and K. Indumathi, "Web Enabled Plant Disease Detection System for Agricultural Applications Using WMSN," *Wireless Personal Communications*, 2018.
 10. P. Voitaender, M. Krause, A. Osep, J. Luiten, B. Sekar, A. Geiger and B. Leibe, "MOTS: Multi-Object Tracking and Segmentation," *Computer Vision and Pattern Recognition*, 2019.
 11. A. Geiger, P. Lenz and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," *Computer Vision and Pattern Recognition*, 2012.
 12. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *Computer Vision and Pattern Recognition*, 2016.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.