Water Sprouts Detection of Cacao Tree Using Mask Region-based Convolutional Neural Network

Nurul Maulidyah
Postgraduate Student of Electrical Engineering Department
Hasanuddin University
Makassar, Indonesia
lidyah@unhas.ac.id

Indrabayu
Informatics Engineering Department
Hasanuddin University
Makassar, indonesia
indrabayu@unhas.ac.id

Intan Sari Areni
Electrical Engineering Department
Hasanuddin University
Makassar, indonesia
intan@unhas.ac.id

Abstract—Water sprouts are one of the causes of cherelle wilt in cacao plants, thus water sprouts must be pruned regularly. On the other hand, the number of cacao workers is increasingly reduced because young people prefer to work in urban areas. Therefore, an automation system to prune water sprouts on cacao plants is needed. Machine vision technology plays an important role in detecting water sprouts in the automatic pruning system in the plantation area. In this paper, the Mask R-CNN (Region-based Convolutional Neural Network) method is used to detect and segment water sprouts from images taken at a cacao plantation. The obtained data consist of 150 images which are for a training dataset of 120 images and a testing dataset of 30 images. To determine the system performance, the threshold parameter in the detection step was tested from 0.1 to 0.9. The best results are obtained at a threshold of 0.6 with an F1score of 0.907. This result shows that the Mask R-CNN method is able to detect water sprout properly.

Keywords—water sprouts, cacao plants, automation, Mask R-CNN

I. Introduction

Water sprouts are fast-growing branches with long segments that mostly appear on the main stem and branches of the cacao tree [1], [2]. The growth of cherelle or young cacao pod will be disrupted by the presence of water buds because the absorption of cherelle nutrients is not optimal. For this reason, water sprouts must be pruned regularly. However, pruning requires a lot of work and can cost around 20% of the annual pre-harvest production cost [3]. In addition, the number of cocoa farmers has decreased significantly due to the young generation today prefer to work in urban areas [4]. Therefore, an automation system for pruning water sprouts on cacao is needed.

Research on automation in agriculture has been carried out for the last few decades. However, there are many problems in its implementation such as the complexity of field operations and the inconsistency of crop systems that hinder the execution of automation technology in plantations [5]. Lately, most of the research on automation solutions for fruit production has focused on the harvesting process [3], [6], [7], pruning [8]–[12] and early detection of pest and disease attacks on fruit [13], where machine vision is one of the important technologies used to detect and localize fruits for yield estimation [14] and harvest, reconstruction of branches for pruning [15] and detection of diseases in plants.

Water sprouts can be detected from the unique color of the leaves. There are many studies that have explored leaf segmentation using various types of methods including edge detection [16], [17], color segmentation [18], [19] and deep learning [20]. Leaf segmentation is mostly performed for the classification of plant species, monitoring of plant growth status, the introduction of plant diseases and insect pests [16].

In 2019, Singh implemented segmentation and classification to detect various diseases on images of sunflower leaves. Segmentation was carried out using the Particle Swarm Optimization (PSO) algorithm. PSO is applied to obtain segmentation output which is used to classify leaf diseases. This method gives promising results with an average accuracy of 98.0% [18]. Another study tried to improve K-means algorithm by calculating the Davies-Bouldin index. This was organized to determine the value of clustering. Then the initial clustering center was given to prevent the clustering calculation from entering the optimal locality. This approach has proven effective in segmenting leaves in tomato plants more accurate and more efficient than the traditional K-means algorithm [19].

Wang et al. in 2018 detected edges based on the Chan-Vese model and the Sobel operator. This approach is used to improve the accuracy of segmentation on overlapping plant leaves. The improved Chan-Vese model and Sobel operator were implemented on the leaf area to extract contours and detect edges [16].

Deep learning is one of machine learning techniques that is widely used in various fields to perform object detection and classification [21]–[23]. This is because deep learning has high accuracy in detecting, classifying and segmenting objects which is important in developing machine vision systems for agricultural applications, especially in complex environments such as plantations. In 2018, Xu et al. use the Mask R-CNN to extract the form and number of leaves in tobacco and Arabidopsis images. This study produces an accuracy of 89.9% for segmentation. This showed that the method was quite promising in leaf segmentation [20].

Color segmentation becomes a difficult task when images contain many complex objects with less specific colors [24]. Previous research on leaf detection was carried out on green leaves, but water sprouts on cacao plants had leaves of various colors ranging from pale green, pink, to light brown. For that reason, as an initial step in the research on automation of water

978-1-7281-6587-5/20/$31.00 ©2020 IEEE

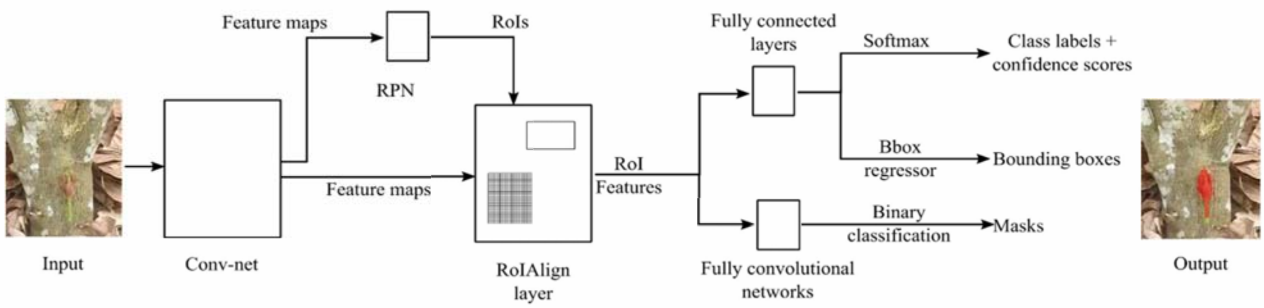


Fig. 1. Mask R-CNN architecture for instance segmentation [26]

sprouts pruning system in cacao plants is detecting water sprouts using the Mask R-CNN method as it is more effective and efficient in object detection and segmentation.

II. Literature Study

A. Mask R-CNN

Mask R-CNN is a deep learning framework for instance segmentation so that each object can be recognized by classes, bounding box and pixel values. According to He et al., there are two stages in Mask R-CNN procedure, i.e. the Region Proposal Network (RPN) and the classification and box offset [25].

However, Shi et al. divided Mask R-CNN into three parts: (i) feature extraction network, (ii) region-proposal network, and (iii) instance detection and segmentation networks as shown in Fig. 1 [26]. The feature-extraction network uses convolutional backbone architecture (Resnet and Feature Pyramid Network) to extract high-level features from images to obtain better accuracy and speed [25]. Convolutional networks use various filters to perform operations such as edge detection, bluring, sharpening, and Gaussian blur. RPN as a region proposer scans every pixel on feature maps generated from convolutional networks using multiple bounding boxes with different sizes and ratios which is centered on each pixel. These bounding boxes are called anchors and act as sliding windows. RPN has a classifier and regressor. The classifiers determine whether the proposal area is a target or background, while the regressor returns the coordinate value of the proposal area [27]. Region proposals that are classified as objects are called Region of Interest (RoI). RoIAlign is applied to align the features extracted from each RoI with its input feature maps without quantizing RoI boundaries and bins. Instead, RoIAlign applies bilinear interpolation to bend the RoI into a fix-sized feature map that is useful for mask prediction [25], [28]. Instance detection and segmentation network consists of fully connected layers that detect class labels, confidence scores and bounding boxes and fully convolutional networks that predict pixel masks inside bounding boxes [26].

B. Pretrained Model

Pre-trained models are the result of the training from a huge of datasets and classes. Basically, CNN training aims to find suitable values for each filter so that certain neurons from the last layer can be activated to predict the correct class. These values are in the pre-trained model that is used to find the features combination in the image and to train the dataset that will be used in this study [29], [30].

A pre-trained model of COCO (Common Objects in Context) dataset is used in this paper, which is a large-scale dataset for object detection, segmentation, and captioning. This dataset has 1.5 million object instances, 80 object categories, 91 stuff categories with 5 captions per image. Although the number is smaller than other datasets, it has more examples for each category. This makes COCO suitable to study object features in more detail and localize more precisely [31].

III. Research Methodology

The Mask R-CNN Framework is used to create a model for water sprouts detection on cacao plants. There are three stages carried out in this study, i.e. image acquisition & image pre-processing, training, and testing as shown in Fig.2.

A. Data

1) Image Acquisition

Data were collected in a cacao farm located in Wajo Regency, South Sulawesi, Indonesia. The images are taken using the Nikon Coolpix P610 camera from a distance of 100-200 cm with a height of 100 cm. The distance and height are chosen with the consideration that water sprouts are found on the main stem of the cacao plant and the spacing between trees is 200-400 cm, which is the standard planting distance in cacao cultivation practices. Some examples of water sprouts on cacao plants are shown in the yellow box in Fig. 3.

2) Image Pre-processing

The data consist of 150 images and divided into 120 images for the training dataset and 30 images for the testing dataset. In preprocessing, data in the form of images are labeled manually using the VIA (VGG Image Annotation) application. Unlike other R-CNN methods that use square or circular bounding boxes, this method uses bounding boxes with polygon shape

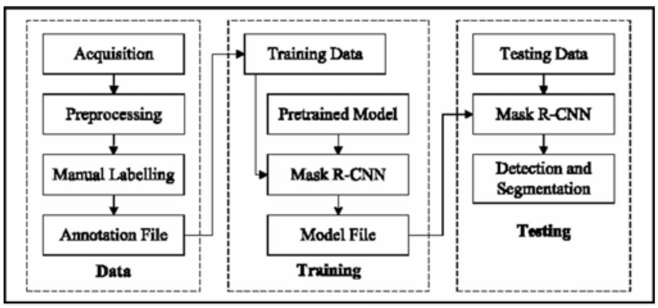


Fig. 2. System Framework

Fig. 3. Examples of water sprout data on cacao plants

Fig. 4. Example of an annotated water sprout

that resemble the shape of the object as shown in Fig. 4. This is useful when the instance segmentation process is performed. The output of the annotation process is a .json file containing the bounding box coordinates and the label of each water sprouts object contained in the image.

B. Training

The training process was executed in Google Colab with the Tesla K80 GPU and 12 GB GDDR5 VRAM. The training process begins by loading the training dataset, which is images with RGB color channels and various image size to the feature extraction network in this case ResNet. This feature extraction network extracts high-level features from input images at the last convolutional layer to produce a high dimensional feature map. In the region proposal network (RPN), anchor boxes with scales (32, 64, 128, 256, and 512) and ratio (0.5, 1, 2) scans the feature map. An intersection-over-union (IoU) anchor with a ground-truth box> 0.7 is considered an RoI (see Fig. 5).

Non-maximum suppression (NMS) is applied to avoid multiple bounding boxes on the same object with a threshold of 0.7. The RoIAlign layer is then applied to map RoI coordinates accurately on fixed-size feature maps. The results are then inputted to a network consisting of two branches to do instance detection and segmentation. The first branch is a fully connected network that detects bounding boxes, classes and confidence scores. The other branch is a fully convolutional network for object masking [26].

Training is carried out using the Stochastic Gradient Descent (SGD) optimizer; hence the RPN and R-CNN can share the convolutional layer. The model was trained with learning momentum 0.9, learning rate 0.0002, weight decay 0.0001 and COCO pre-trained models for initialization on convolutional networks [32].

C. Testing

In the testing stage, an experiment was carried out using the minimum confidence threshold of 0.1 to 0.9. It aims to obtain the threshold value that produces the best system performance. Minimum confidence used to measure the certainty or reliability related with each discovered pattern. To determine the performance of the model based on threshold values, an accuracy calculation using a confusion matrix with two classes, namely water sprouts and not water sprouts, was performed.

D. Evaluation Metrics

Evaluation of the Mask R-CNN model was carried out using a confusion matrix. In this study, F1score system performance measurement is used with a range between 0 (bad) and 1 (very good). *F1score* calculated the average value of *precision* and *recall* as shown in Eq. (1). Precision indicates ratio of the correct detection of water sprouts to all the results of the detection of water sprouts produced by the system, while recall indicates the ratio of the correct detection of water sprouts detected by the system to all water sprouts in the image.

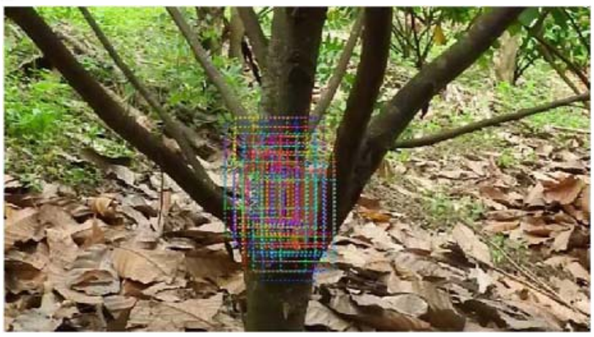
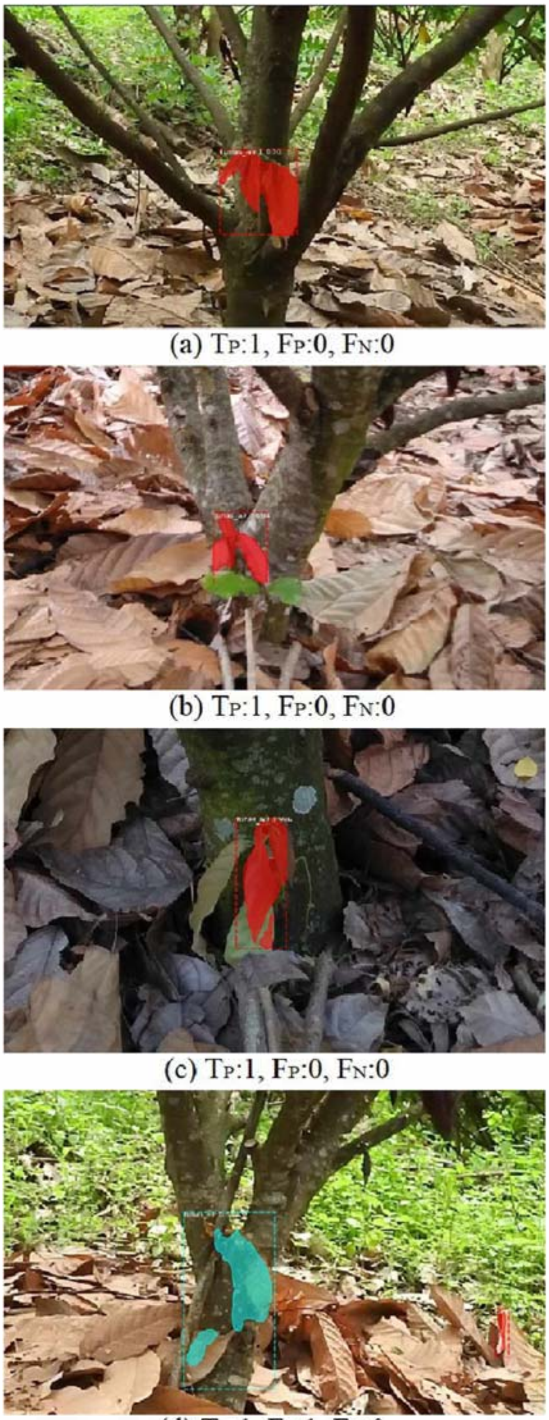


Fig. 5. Example of anchors that are considered as RoIs in the RPN process



(a) TP:1, FP:0, FN:0

(b) TP:1, FP:0, FN:0

(c) TP:1, FP:0, FN:0

(d) TP:1, FP:1, FN:0

Fig. 6. Example of detection and segmentation of water sprouts using Mask R-CNN

$$F1score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{1}$$

The recall and precision can be obtained using (2) and (3)

$$recall = \frac{T_P}{T_P + F_N} \tag{2}$$

$$precision = \frac{T_P}{T_P + F_P} \tag{3}$$

where T_P is true positive, F_N is false negative and F_P is false positive of all images in the testing dataset.

IV. Results and Discussion

Some result examples from instance segmentation in the test dataset can be seen in Fig. 6. In Fig. 6 (a), a water sprout can be detected and segmented very well. However, there are some objects detected but not segmented properly as shown in figures Fig. 6 (b) and Fig. 6 (c). This is because in Fig. 6 (b), the detected water sprout has green leaves, while the trained water sprouts data have a slightly brownish color. Moreover, Fig. 6 (c) has poor lighting so the object is not detected properly. Finally, Fig. 6 (d) shows the result of false positive detection, where dry leaves are detected as water sprouts. These unwanted results can be reduced by adding a variety of training data.

Water sprouts detection system is done by changing the minimum confidence threshold value from 0.1 to 0.9 so that optimal performance is obtained. Evaluation is done at the object level by counting the number of water buds detected in the image. The results obtained show that at the threshold 0.1 to 0.5, the value of F1score is constant at 0.818. This is due to the lowest confidence score value of all objects detected at the threshold 0.1-0.5 is 0.513 (see Fig. 6 (d)). However, at the 0.6 threshold, this object is no longer detected by the system and causes changes to f1score. At the 0.7 to 0.9 threshold, object detection results begin to decrease so that the F1score value also decreases. The best results are shown at the threshold of 0.6 with F1score 0.907 as shown in Fig.7.

V. Conclusion

Water sprouts detection system on cacao trees has been carried out in this study with the Mask R-CNN method. The stages of the study were divided into the training process using COCO pre-trained models with 120 data and the testing process using 20 data. System performance is evaluated based on F1score values obtained by changing the threshold value from 0.1 to 0.9. The best result are obtained when the threshold is 0.6 with an F1score of 0.907. For future work, separation of leaf and stem classes in water buds will be carried out so that each part can be specifically identified.

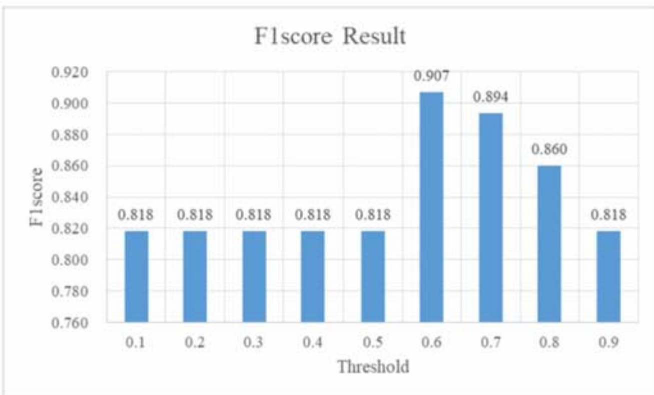


Fig. 7. F1score results for threshold 0.1 to 0.9

References

[1] T. J. Harris, "CACAO GROWING," *J. R. Soc. Arts*, vol. 62, no. 3226, pp. 910–912, 1914.

[2] A. A. Prawoto *et al.*, Technical Guidelines for Good Cocoa Cultivation. Jakarta: Directorate General of Plantation, Ministry of Agriculture, 2014. [in Indonesia]

[3] S. Amatya, M. Karkee, A. Gongal, Q. Zhang, and M. D. Whiting, "Detection of cherry tree branches with full foliage in planar architecture for automated sweet-cherry harvesting," *Biosyst. Eng.*, vol. 146, pp. 3–15, Jun. 2016, doi: 10.1016/j.biosystemseng.2015.10.003

[4] W. Chandra, "Looking for Solutions to Improve Cocoa Plantation Ecosystems in South Sulawesi" *Mongabay Environmental News*, 11-Nov-2018. [Online]. Available: https://www.mongabay.co.id/2018/11/11/mencari-solusi-perbaikan-ekosistem-pertanaman-kakao-di-sulsel/. [Accessed: 06-Mar-2019]. [in Indonesia]

[5] T. Baugher and L. He, "Orchard Automation - Engineering Opportunities for Fruit Growers," *Penn State Extension*, 19-Jun-2018. [Online]. Available: https://extension.psu.edu/orchard-automation-engineering-opportunities-for-fruit-growers. [Accessed: 24-Mar-2019].

[6] H. Altaheri, M. Alsulaiman, G. Muhammad, S. U. Amin, M. Bencherif, and M. Mekhtiche, "Date fruit dataset for intelligent harvesting - ScienceDirect," *Data Brief*, vol. 26, Oct. 2019.

[7] H. A. M. Williams *et al.*, "Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms," *Biosyst. Eng.*, vol. 181, pp. 140–156, May 2019, doi: 10.1016/j.biosystemseng.2019.03.007.

[8] S. A. Akbar, N. M. Elfiky, and A. Kak, "A novel framework for modeling dormant apple trees using single depth image for robotic pruning application," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 5136–5142, doi: 10.1109/ICRA.2016.7487718.

[9] F. J. Castillo-Ruiz, R. R. Sola-Guirado, S. Castro-Garcia, E. J. Gonzalez-Sanchez, J. T. Colmenero-Martinez, and G. L. Blanco-Roldán, "Pruning systems to adapt traditional olive orchards to new integral harvesters," *Sci. Hortic.*, vol. 220, pp. 122–129, Jun. 2017, doi: 10.1016/j.scienta.2017.03.043.

[10] S. Chattopadhyay, S. A. Akbar, N. M. Elfiky, H. Medeiros, and A. Kak, "Measuring and modeling apple trees using time-of-flight data for automation of dormant pruning applications," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1–9, doi: 10.1109/WACV.2016.7477596.

[11] N. M. Elfiky, S. A. Akbar, J. Sun, J. Park, and A. Kak, "Automation of dormant pruning in specialty crop production: An adaptive framework for automatic reconstruction and modeling of apple trees," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015, pp. 65–73, doi: 10.1109/CVPRW.2015.7301298.

[12] M. Karkee, B. Adhikari, S. Amatya, and Q. Zhang, "Identification of pruning branches in tall spindle apple trees for automated pruning," *Comput. Electron. Agric.*, vol. 103, pp. 127–135, Apr. 2014, doi: 10.1016/j.compag.2014.02.013.

[13] Basri, Harli, Indrabayu, I. S. Areni, and R. Tamin, "Image Processing System for Early Detection of Cocoa Fruit Pest Attack," *J. Phys. Conf. Ser.*, vol. 1244, p. 012003, Jun. 2019, doi: 10.1088/1742-6596/1244/1/012003.

[14] Y. Ekawaty, Indrabayu, and I. S. Areni, "Automatic Cacao Pod Detection Under Outdoor Condition Using Computer Vision - IEEE Conference Publication," in *2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, Yogyakarta, Indonesia, Nov. 2019, doi: 10.1109/ICITISEE48480.2019.9003733.

[15] L. He and J. Schupp, "Sensing and Automation in pruning of Apple Trees: A Review," *Agron. 2018*, 2018.

[16] Z. Wang, K. Wang, F. Yang, S. Pan, and Y. Han, "Image segmentation of overlapping leaves based on Chan–Vese model and Sobel operator - ScienceDirect," *Inf. Process. Agric.*, vol. 5, no. 1, pp. 1–10, Mar. 2018.

[17] J. Zhang, F. Kong, J. Wu, S. Han, and Z. Zhai, "Automatic image segmentation method for cotton leaves with disease under natural environment," *J. Integr. Agric.*, vol. 17, no. 8, pp. 1800–1814, Aug. 2018, doi: 10.1016/S2095-3119(18)61915-X.

[18] V. Singh, "Sunflower leaf diseases detection using image segmentation based on particle swarm optimization," *Artif. Intell. Agric.*, Nov. 2019, doi: 10.1016/j.aiia.2019.09.002.

[19] K. Tian, J. Li, J. Zeng, A. Evans, and L. Zhang, "Segmentation of tomato leaf images based on adaptive clustering number of K-means algorithm," *Comput. Electron. Agric.*, vol. 165, p. 104962, Oct. 2019, doi: 10.1016/j.compag.2019.104962.

[20] L. Xu, Y. Li, Y. Sun, L. Song, and S. Jin, "Leaf Instance Segmentation and Counting Based on Deep Object Detection and Segmentation Networks," in *2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*, 2018, pp. 180–185, doi: 10.1109/SCIS-ISIS.2018.00038.

[21] Indrabayu, N. L. Jamaluddin, and I. S. Areni, "Detection and Distance Estimation against Motorcycles as Navigation Aids for Visually-impaired People," in *2019 12th International Conference on Information Communication Technology and System (ICTS)*, 2019, pp. 224–228, doi: 10.1109/ICTS.2019.8850936.

[22] L. Lu, Y. Yi, F. Huang, K. Wang, and Q. Wang, "Integrating Local CNN and Global CNN for Script Identification in Natural Scene Images," *IEEE Access*, vol. 7, pp. 52669–52679, 2019, doi: 10.1109/ACCESS.2019.2911964.

[23] Y. Majeed *et al.*, "Apple Tree Trunk and Branch Segmentation for Automatic Trellis Training Using Convolutional Neural Network Based Semantic Segmentation," *IFAC-Pap.*, vol. 51, no. 17, pp. 75–80, Jan. 2018, doi: 10.1016/j.ifacol.2018.08.064.

[24] S. J. Sangwine and R. E. N. Horne, *The Colour Image Processing Handbook*. Springer US, 2012.

[25] [24] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *ArXiv170306870 Cs*, Jan. 2018.

[26] W. Shi, R. van de Zedde, H. Jiang, and G. Kootstra, "Plant-part segmentation using deep learning and multi-view vision," *Biosyst. Eng.*, vol. 187, pp. 81–95, Nov. 2019, doi: 10.1016/j.biosystemseng.2019.08.014.

[27] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *ArXiv150601497 Cs*, Jan. 2016.

[28] X. Zhang, G. An, and Y. Liu, "Mask R-CNN with Feature Pyramid Attention for Instance Segmentation," in *2018 14th IEEE International Conference on Signal Processing (ICSP)*, Beijing, China, 2018, pp. 1194–1197, doi: 10.1109/ICSP.2018.8652371.

[29] M. Huh, P. Agrawal, and A. A. Efros, "What makes ImageNet good for transfer learning?," *ArXiv160808614 Cs*, Dec. 2016.

[30] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," Nov. 2014.

[31] T.-Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context," *ArXiv14050312 Cs*, Feb. 2015.

[32] W. Abdulla, "Mask_RCNN: Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow," *Github*, 2017.