JTEP Jurnal Keteknikan Pertanian,
Vol. 12 No. 3, p 327-339, 2024
P-ISSN 2407-0475 E-ISSN 2338-8439

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.3.327-339

The Development Mask R-CNN Model for Identification of Melon Plant Leaves and Branches

Meia Noer Muslimah[1], Sri Wahjuni[1*], Toto Haryanto[1]

[1]Department of Mechanical and Biosystem Engineering, Faculty of Agricultural Engineering and Technology, IPB University, Lingkar Akademik Street, IPB Dramaga Campus, Bogor, East Java 16680, Indonesia
*Corresponding author, email: my_juni04@apps.ipb.ac.id

Article Info

Submitted: 21 July 2024
Revised: 16 September 2024
Accepted: 18 October 2024
Available online: 28 October 2024
Published: Desember 2024

Keywords:
hyperparameter tuning; Mask R-CNN; melon; pruning

How to cite:
Muslimah, M. N., Wahjuni, S., Haryanto, T. (2024). The Development Mask R-CNN Model for Identification of Melon Plant Leaves and Branches. Jurnal Keteknikan Pertanian, 12(3): 327-339. https://doi.org/10.19028/jtep.012.3.327-339.

Abstract

The quality of melons can be enhanced and optimized by pruning melon plants. Pruning is a removal process that is carried out on specific parts of a plant. Currently, melon plants are still manually pruned by farmers, but this method has many drawbacks. In this research, pruning was conducted on the branches and leaves of melon plants. Pruning can be facilitated with the assistance of a robot capable of recognizing leaves and branches. In this study, a Mask Region-based Convolutional Neural Network (Mask R-CNN) was used to detect branches and leaves. The hyperparameter tuning technique was employed to obtain the best parameter values, including the learning rate, weight decay, and learning momentum. Two scenarios were considered in this study: one with 10 epochs and the other with 30 epochs. The Average Precision (AP) values obtained at 10 epochs were 32.2% for leaf objects and 0% for branches. At 30 epochs, the AP values were 56.8% for the leaf objects and 4.1% for the branches. The mean Average Precision (mAP) is 16.1% for 10 epochs and 28.4% for 30 epochs.

Doi: https://doi.org/10.19028/jtep.012.3.327-339

1. Introduction

Melon (Cucumis melo L.) is a horticultural plant that is in high demand by communities. The taste of melon fruit is sweet, and the color of its flesh varies (Khumaero et al. 2015). The texture of the flesh is soft and the aroma of the fruit is fragrant, making it a special attraction for people in all circles (Maryanto and Daryono 2011). From a marketing perspective, melons have economic value and are promising prospects for farmers (Sudiyarto, 2011).

The need for melons continues to increase due to public awareness of health (Sukmaningtyas and Hartoyo 2011), and melon production has also increased annually. According to Badan Pusat Statistik (BPS 2022), melon production has increased over the past four years. From 2017 to 2020, melon production in Indonesia was 92,434 tons, 118,708 tons, 122,105 tons, and 138,177 tons, respectively.

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.3.327-339

Increasing melon production is expected to result in melon fruit yields of high quality so as not to harm farmers and consumers.

Melon quality can be improved and optimized in various ways, one of which is by pruning melon plants. Pruning is a disposal process performed on certain parts of a plant (Koentjoro, 2012). According to Dinas Pertanian Tanaman Pangan and Holtikultura Jawa Tengah (DPTPH 2009), the purpose of pruning melon plants is to ensure plant growth so that the production process takes place optimally, and can reduce moisture in the plant crown to reduce the risk of pest and disease attacks. Pruning affects the number of leaves and fruits, fruit length, fruit diameter, and fruit weight (Sofyadi et al. 2021).

The pruning of melon plants is still performed manually by the farmers. This has several disadvantages, including only being done on narrow land, requiring high costs to pay for labor, and requiring a long trimming time (Daryono et al. 2014). Pruning on large areas of land can be performed with the help of machines (pruning automation robots) and can be used as an alternative for labor efficiency (Primilestari and Purnama 2019). Pruning automated robots can reduce the maintenance costs of melon plants, which are high (Wijaya et al. 2021).

The process of pruning Melon plants were pruned on their leaves and branches. Leaves and branches are distinguished to help optimize crop yields, and farmers can obtain information about the location/position and condition of leaves and branches, which can affect the quality and quantity of melons produced. In addition, the separate identification of leaves and branches can aid in pest and disease control (Wahjuni et al. 2023). Research to identify the leaves and branches of melon plants has been carried out by (He et al. 2017) by utilizing the Faster Region Convolutional Neural Network (Faster R-CNN) deep learning approach. The study produced an average precision (AP) of 49.72% for the leaf class, and the branch class was not detected. Furthermore, Faster R-CNN is expanded by adding a binary mask to the output layer, known as the Mask Region-based Convolutional Neural Network (Mask R-CNN). The Mask R-CNN approach allows algorithms to be more precise in object recognition, faster in computational processes, and designed for pixel-to-pixel alignment of the input and output layers (He et al. 2017). Related research using Mask R-CNN for identification of branches and leaves in melon plants has been conducted by (Muslimah 2021). The study obtained better AP results than the Faster R-CNN branch class of 0.1796 and leaf class AP of 0.5751. This study used the default parameters of proprietary research from (He et al. 2017). However, branches were not detected in the present study. The testing used similar data to develop and evaluate the model.

According to previous research, it is necessary to improve the results by hyperparameter tuning using the Mask R-CNN method. Mask R-CNN was selected in this study because it is useful in the object segmentation process. This makes it possible to detect precise object constraints and requires detailed analysis of the objects.

2. Materials and Methods

2.1 Stage of Research

This research consists of data preprocessing, data sharing, hyperparameter tuning, model building, testing, best model optimization, and result analysis. A diagram of the research stages is shown in Figure 1.

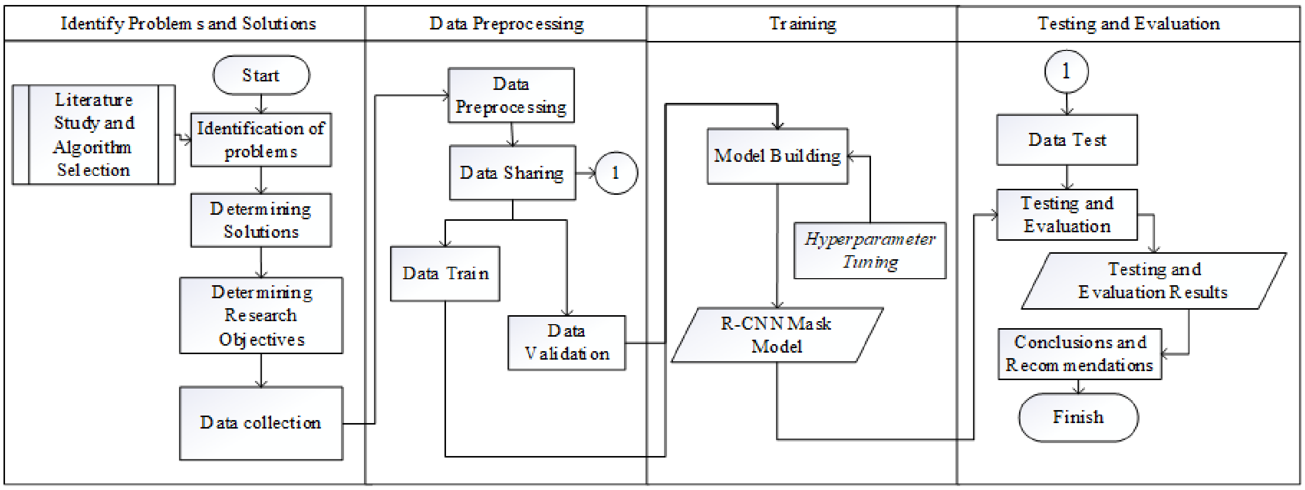


Figure 1. Stage of research

2.2 Research Data

This study used secondary data from a previous study (WahjunData Imaging). Data and forms of images. The images used were as many as 180 original images and 90 augmented images from previous studies. An example of a melon plant is shown in Figure 2.

Figure 2. Image of melon plant

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.3.327-339

The images were captured manually using an Intel Real Sense D435i RGB-D camera. The type of plant in the image is the golden melon found in the Agribusiness and Technology Park (ATP) IPB.

2.3 Preprocess Data

The data preprocessing stage was divided into two stages: data annotation and sharing. Annotation was performed by labelling the classes on the images. Labelling can be performed by coding or by using an annotation tool. Labelling was performed to obtain the ground-truth information for each object labelled on each image (Krig, 2014). The ground truth is a collection of information regarding the labels, images, and labelling features in each image. This study carried out a labelling stage to obtain balanced ground truth information of branches and leaves because in previous studies, the ground truth obtained was not balanced. The annotation tool used in this study was the visual geometry group (VGG) image annotator, which labels the leaves and branches in the image. The stages for labelling on the VGG image annotator (VIA) are preparing all images to be labelled (accepted formats in JPEG or PNG), opening a web browser to access the VIA, uploading all images to be labelled, choosing what shape will be labeled (in this study, polygon shape), pressing and pulling the mouse to make marks around objects, and labeling objects that have already been marked. The same is performed for other objects that will be labelled for all the image data that have been prepared, and then the labelling results are saved or exported according to the required format. The next stage is data sharing. The data used in this study were the ground-truth data generated from the labelling or annotation of the images. The data were divided into three types: training, validation, and testing. Data sharing was randomly conducted. The proportion of data sharing was 80% for the training data, 10% for the validation data, and 10% for the test data.

2.4 Model Development

At this stage, the branches and leaves were detected using the Mask R-CNN method. This method can detect objects in an image and produce an output in the form of segmentation masks and objects detected with a high quality. The construction of the R-CNN Mask model uses code derived from the repository belonging to the inventor of the Mask R-CNN, Matterport (MIT, USA). The code can be accessed on https://github.com/matterport/Mask_RCNN. The architecture of Mask R-CNN is shown in Figure 3.

JTEP Jurnal Keteknikan Pertanian,
Vol. 12 No. 3, p 327-339, 2024
P-ISSN 2407-0475 E-ISSN 2338-8439

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.3.327-339

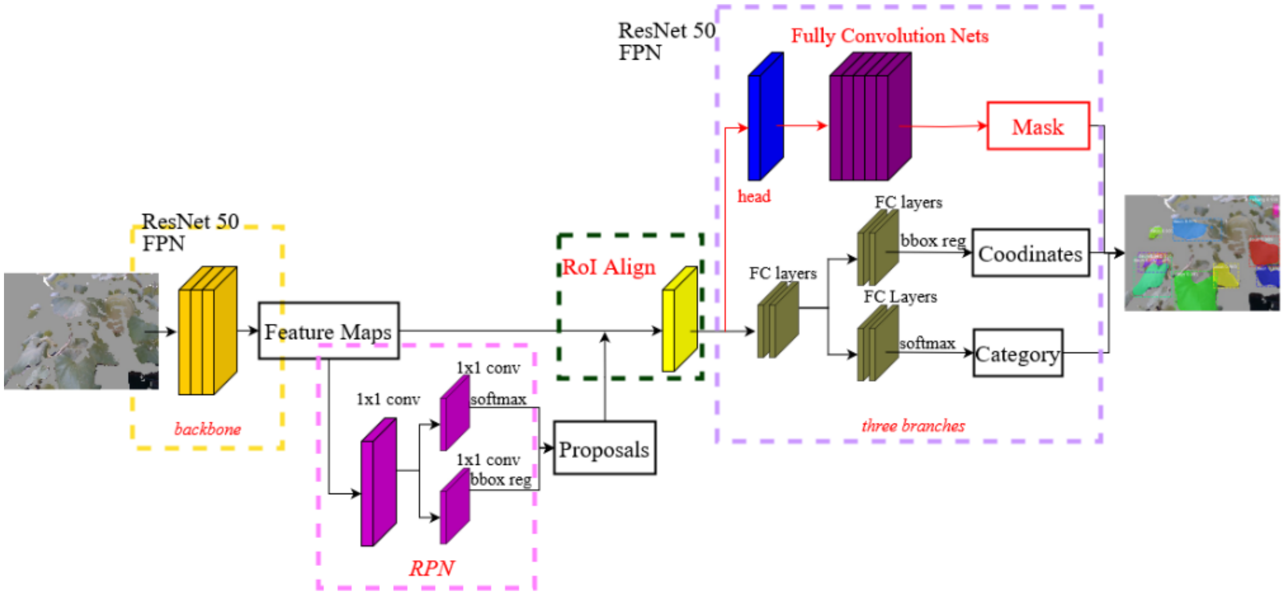


Figure 3. The architecture of Mask R-CNN

2.5 Hyperparameter Tuning

Hyperparameter tuning aims to configure certain parameters to be trained in the model. The hyperparameter configurations are listed in Table 1. The reference values of the parameter combinations were obtained from previous CNN studies. The total number of combinations of parameters used in this study was 18.

Table 1. The hyperparameter configuration

Parameter type		
Learning rate	**Learning momentum**	**Weight decay**
0.001	0.7	0.0001
0.001	0.7	0.0005
0.001	0.7	0.005
0.001	0.9	0.0001
0.001	0.9	0.0005
0.001	0.9	0.005
0.01	0.7	0.0001
0.01	0.7	0.0005
0.01	0.7	0.005
0.001	0.7	0.0001

Continue

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.3.327-339

Continue

Parameter type		
Learning rate	**Learning rate**	**Learning rate**
0.01	0.9	0.0001
0.01	0.9	0.0005
0.01	0.9	0.005
0.02	0.7	0.0001
0.02	0.7	0.0005
0.02	0.7	0.005
0.02	0.9	0.0001
0.02	0.9	0.0005
0.02	0.9	0.005

2.6 Testing and Evaluation

The model testing process in this study uses 10% of the annotated data that has been shared in the previous data sharing process and has never been trained by the model. This testing process was performed to determine the performance of the Mask R-CNN algorithm in predicting or detecting objects in the branches and leaves of melon plants. Testing and evaluation were performed by calculating the precision, recall, average precision (AP), and mean average precision (mAP).

Several steps were implemented to test the model obtained in the previous stage. Model testing was performed by obtaining the mAP values. To obtain the mAP value, it is necessary to define the intersection over union (IoU) value. IoU is an actual evaluation metric for positional accuracy that measures how true image boundaries match the prediction limits of a dataset, in this case, image data (Powers, 2011). The IoU formula is given by Equation (1), and it is necessary to define the threshold value used to classify the data.

$$IoU = \frac{area(true\ bounding\ box \cap predicted\ bounding\ box)}{area(true\ bounding\ box \cup predicted\ bounding\ box)} \tag{1}$$

Before calculating the precision value, true positive (TP) and false positive (FP) values were calculated. TP is the positive class data, is the positive class predicted by the prediction model, and FP is the negative class data and positive class predicted by the prediction model. The threshold value used was 0.5; the prediction could be classified as TP if the IoU value was > 0.5, and the prediction could be classified as FP if the IoU value was < 0.5.

The precision and recall formulas are shown in Equations (2) and (3), respectively, where precision is the proportion of cases predicted by the positive class data, and it is true that the data are classified as

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.3.327-339

positive class data. The recall represents the proportion of cases of class data that are actually positive and correctly predicted as positive class data.

Next, the precision and recall values are used to represent the maximum precision values at each recall level. This is commonly known as interpolated precision (Pinterp). The formula for obtaining the Pinterp value is given in Equation 4 (Padilla et al. 2021). Furthermore, precision recall and Pinterp values were visualized on the curve to facilitate representation of the AP value. AP is an evaluation metric used in object detection. The AP is a measure of performance in a particular class for detecting objects and evaluating the performance of the datasets in the model. The formula for AP is given in Equation (5), and the calculation of the AP value is based on the calculation of the area under the recall precision curve. The average value of the AP in multiple classes is mAP, as shown in Equation (6) (Padilla et al. 2021).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} = \frac{\text{TP}}{\text{all detections}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} = \frac{\text{TP}}{\text{all ground truths}} \quad (3)$$

$$Pinterp(r) = maxP(r), r: r \geq \text{r} \quad (4)$$

$$AP = \sum_{i=1} (\text{Recall}_i - \text{Recall}_{i-1}) \,.\, (\text{Pinterp}_i\,), \quad (5)$$

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i; \quad (6)$$

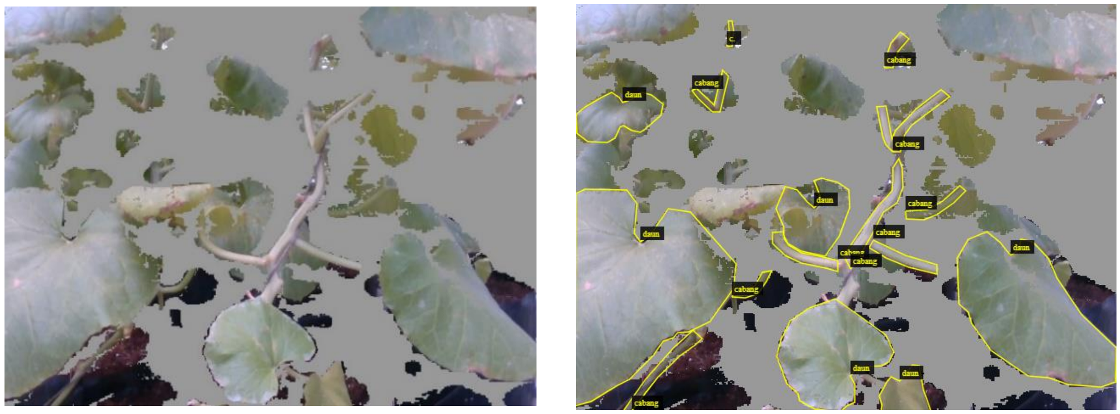
where r is the recall, maxP is the maximum value of the precision at each recall level, and C is the number of classes.

3 Result and Discussion

3.1 Data Prepocessing

At this stage, data annotation was performed using the VIA annotation tool. An example of the data annotation results is shown in Figure 4. Data annotation was performed on 270 images. Each image was labeled according to the selected object. The naming of objects can be customized as desired; in this case, the leaves and branches. As many as 1,693 bounding boxes and branches were obtained, and as many as 1,674 bounding boxes. The resulting extension of the annotation of the entire image was. csv, .json, or coco format.

jTEP Jurnal Keteknikan Pertanian,
Vol. 12 No. 3, p 327-339, 2024
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.3.327-339



(a) (b)

Figure 4. Data (a) original image (b) image after annotation using VIA

3.2 Training and Data Validaton

The process stage of Training and validation of leaf and branch objects on melon plants using the Mask R-CNN algorithm with hyperparameter tuning. In this process, two training scenarios were applied using epochs 10 and 30. The treatment of two training scenarios with epoch numbers of 10 and 30 was one of the approaches used in this study to evaluate model performance. Both epochs were trained by using a combination of 18 predetermined parameters. This was performed to analyze the training results against the performance of the model on the validation data, training time, and test results.

Training with an epoch count of 10 took 134.87 minutes of training time and obtained as many as 180 models. Training with epochs 30 took longer at 403.71 minutes and obtained as many as 540 models. The model extension for this process was. h5. The best model was selected on the basis of the smallest validation loss value for each epoch.

The validation loss value of Mask R-CNN is an important metric in the model training process. Validation loss is a measure of how well the model works on the validation data during the training process. The purpose of the validation loss is to measure the extent to which the model can generalize from the training data to the never-before-seen data (validation data). The smaller the validation loss value obtained by the model, the better the model for predicting the validation data. The best combinations of the parameters obtained during the training and validation processes at each epoch are listed in Table 2.

The validation loss values obtained at epochs 10 and 30 have the same regularity; namely, six combinations of parameters do not appear in the chart. This is because the experiment on the combination of these parameters contains a validation loss value of NaN (Not-a-Number) or is undefined. The NaN value indicates a problem that causes the calculation results to become invalid or undefined during the training process. The NaN value can also be attributed to the complexity of the parameter combinations. In addition, the NaN value in the validation loss is caused by overfitting. Overfitting can occur when the model memorizes excessive training data and cannot be generalized well.

Table 2. The best model for training and validation

Epoch	Parameter		Best epochs	Smallest validation loss
	Name	Value		
10	Learning rate Learning momentum Weight decay	0,001 0,7 0,005	3	1,51
30	Learning rate Learning momentum Weight decay	0,001 0,7 0,0005	27	1,32

3.3 Model Testing and Evaluation

Model testing on the Mask R-CNN algorithm was performed to determine the performance of the model produced using the mAP. The mAP value was obtained by calculating the precision, recall, and average precision (AP) values.

Subsequently, the precision and recall of each object in each class were calculated. High-precision values result in highly accurate detections but miss small objects, whereas high recall or sensitivity values result in more object detections but more misidentified objects (FNs). This research is more concerned with high-precision values because of their accuracy in recognizing an object. Therefore, this study used the AP value as a model evaluation. Next, the precision interpolation value was calculated to estimate the value between precision and recall. This precision interpolation was also visualized on the curve to make it easier to evaluate AP values. The precision and recall (PR) curves and interpolation results are shown in Figure 5 and 6, respectively. The PR curve aims to visualize the AP value from the results of the precision interpolation and the actual precision value. The x-axis of the PR curve represents recall, which is also known as sensitivity. The y-axis on the curve represents precision. The x- and y-axes are ratios or proportions, respectively; therefore, they have no units.

Furthermore, the accuracy calculations for both scenarios were obtained from the AP values of each object. AP at epoch 10 leaf objects by 32.2%, and branch objects by 0.0% (no branch objects were detected). AP at epoch 30 leaf objects by 56.8% and branch objects by 4.1%. The AP value of each object

JTEP Jurnal Keteknikan Pertanian,
Vol. 12 No. 3, p 327-339, 2024
P-ISSN 2407-0475 E-ISSN 2338-8439

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.3.327-339

was used to calculate the mAP values. The mAP values for epochs 10 and 30 were 16.1% and 28.4%, respectively. The AP value obtained in this study can still be said to be small; however, the resulting model can detect leaf and branch objects. The AP value is small because the branch object is too small; therefore, the model does not recognize the characteristics of the object well. The epoch used can still be added so that the model can learn as much as possible to recognize the object. The loss validation value is still above one, which means that the resulting model is not optimal, and the image data used are still less focused on the object.

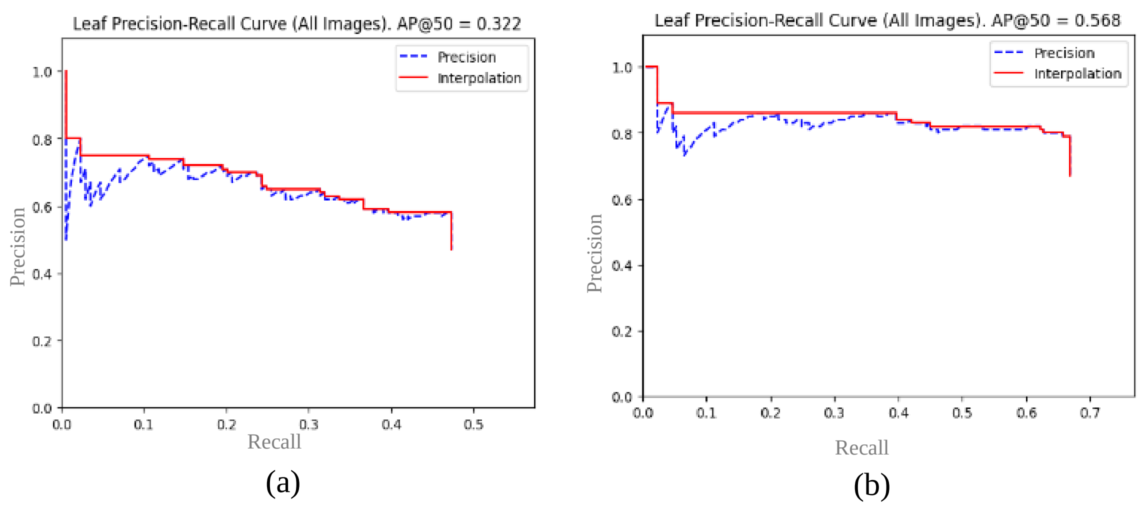


Figure 5. PR curve and leaf interpolation (a) epoch 10 (b) epoch 30

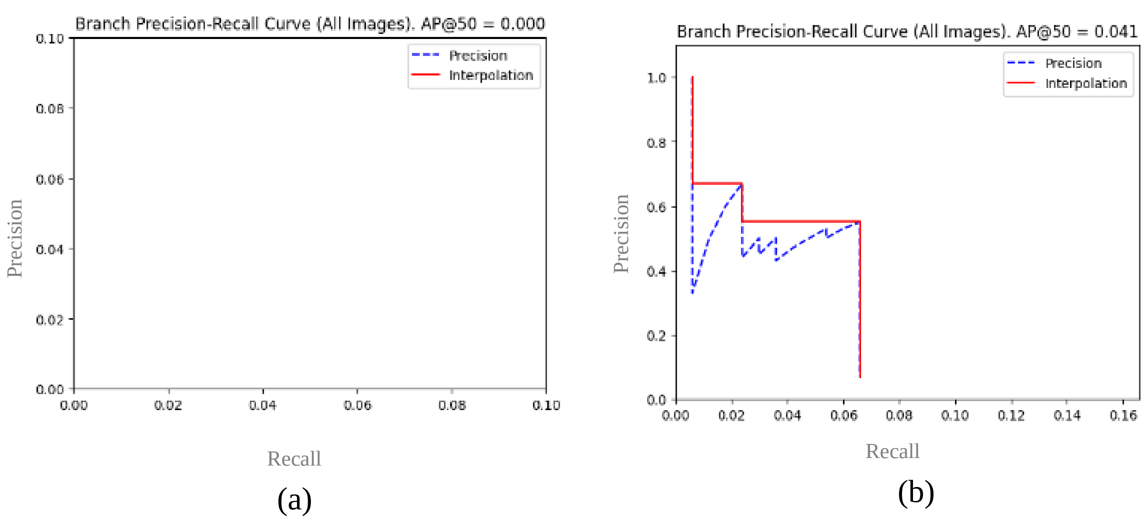


Figure 6. PR curve and branch interpolation (a) epoch 10 (b) epoch 30

JTEP Jurnal Keteknikan Pertanian,
Vol. 12 No. 3, p 327-339, 2024
P-ISSN 2407-0475 E-ISSN 2338-8439

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.3.327-339

The selection of the best model can be observed from the validation loss values and the test results (mAP). First, if viewed based on the obtained validation loss value, the best model used 30 epochs. At epoch 10, the smallest validation loss value is higher than epoch 30, which is also reinforced by the AP value generated by each epoch where epoch 10 (Figure 6 (a)) is 0, which means that it does not detect any branches. Finally, when viewed from the test results with a higher mAP value, the best model chosen was one trained using epoch 30, even though the mAP value was only 12.3% higher. At epoch 10, branches could not be detected, whereas at epoch 30, branches could be detected even though it was only 1. This is because the branch object is too small and the color is almost the same as the background color. In this study, the best model was one that could identify objects using epoch 30.

(a) (b)

Figure 7. Objects that are successfully detected by the best model (a) leaf (b) branch of melon plant

The visualization of the object detection using the best model is shown in Figure 7. From the displayed image, it appears that a plant branch was successfully detected by the object detection model with a confidence level of 0.91, which shows that the model is highly confident in its detection results. The detected branches had certain characteristics that made them stand out from the rest of the plant. One of the factors influencing this detection is the optimal level of lighting and contrast in the area, which allows the algorithm to better recognize visual features such as the shape, color, and texture of the branches. In addition, the Intersection over Union (IoU) value of 0.56 indicates that there is a moderate level of overlap between the main detection box (bounding box), which is colored green,

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.3.327-339

and the comparison box, which is colored red. Although this value indicates that there is room for improvement in accuracy, it provides sufficient information that the detected area has significant relevance to the target object. The physical characteristics of the detected branches included a uniform green color with the surrounding leaves, a smooth surface texture, and the position of the branch, which may be the main focus of the camera. This branch also has a clear and defined shape, making it easier for the algorithm to distinguish it from more complex and diverse leaf backgrounds. The object detection algorithm used appears to have been trained with a dataset that includes similar examples; therefore, it can recognize and detect these on all visible melon branches with a high level of confidence.

4. Conclusion

This research succeeded in detecting leaves and branches using test data. The selection of the best model is seen from the lowest value based on the length of time required during the training and validation process, validation loss value, and test results (mAP). The best model was obtained using epoch 30 with an AP value for leaf objects of 56.8% and branches of 4.1%. It can be concluded that the mAP obtained was 28.4%.

5. References

[BPS] Badan Pusat Statistik. (2022). Produksi Tanaman Buah-Buahan. [accessed 2022 Jan. 1]. https://www.bps.go.id/indicator/55/62/5/produksi-tanaman-buah-buahan.html.

[DPTPH] Dinas Pertanian Tanaman Pangan & Holtikultura Jawa Tengah. 2009. SOP Melon Kabupaten Sragen 2009.

Daryono, B. S., Joko, T., & Purnomo. (2014). Budidaya Melon Ramah Lingkungan di Area Sekitar Gumuk Pasir. Bioedukasi. 7(2011):56–59. https://jurnal.uns.ac.id/bioedukasi/article/view/2842

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. Proc IEEE Int Conf Comput Vis., forthcoming.

Khumaero, W. W., Efendi, D., Suwarno, W. B., & Sobir. (2015). Evaluasi Karakteristik Hortikultura Empat Genotipe Melon (Cucumis melo L.) Pusat Kajian Hortikultura Tropika IPB. J Hortik Indones. 5(1):56. doi: https://doi.org/10.29244/jhi.5.1.56-63

Koentjoro, Y. (2012). Efektifitas model pemangkasan dan pemberian pupuk majemuk terhadap tanaman melon (Cucumis melo L.). Berk Ilm Agroteknologi Plumula. 1(1):9–17. https://www.academia.edu/7492633/Efektifitas_Model_Pemangkasan_Melon_Yonny_Plumula_Vol_1_no_1_

Krig, S. (2014). Computer Vision Metrics: Survey, Taxonomy, and Analysis, Barkeley. CA: Apress.

Maryanto, S. D., & Daryono, B. S. (2011). The comparison of melon (Cucumis melo l.) phenotypic characters among melodi gama 1, gama melon basket, and commersial cultivars using multilocation and multiseason test in indonesia. In: Pasific Science Congress XXII. Kuala Lumpur.

Muslimah, M. N. (2021). Identifikasi cabang dan daun tanaman melon pada citra RGB-D menggunakan metode Mask R-CNN. Bogor: Institut Pertanian Bogor.

Padilla, R., Passos, W. L., Dias, T. L. B., Netto, S. L., & Da Silva, E. A. B. (2021). A comparative analysis of object detection metrics with a companion open-source toolkit. Electron. 10(3):1–28. doi: https://doi.org/10.3390/electronics10030279

Powers, D. M. W. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation J Mach Learn Technol. 2(1). https://doi.org/10.48550/arXiv.2010.16061

Primilestari, S., & Purnama, H. (2019). Teknologi budidaya jeruk di lahan gambut untuk meningkatkan produktivitas dan pendapatan petani di Kabupaten Tanjung Jabung Barat. Pros Semin Nas Lahan Suboptimal. September:978–979. https://conference.unsri.ac.id/index.php/lahan suboptimal/article/view/1467

Sofyadi, E., Lestariningsih, S. N. W., & Gustyanto, E. (2021). Pengaruh Pemangkasan terhadap Pertumbuhan dan Hasil Mentimun Jepang (Cucumis sativus L.) 'Roberto'. Jurnal Unsur Agroscience. 11(1):14–28.doi: https://doi.org/10.35194/agsci.v11i1.1572

Sudiyarto. (2011). Strategi Pemasaran Buah Lokal Jawa Timur. Jurnal Sosial Ekonomi Pertanian. 5(1):65–73. https://jurnal.unej.ac.id/index.php/JSEP/article/view/383

Sukmaningtyas, A., & Hartoyo. (2011). Pengaruh nilai dan gaya hidup terhadap preferensi dan perilaku pembelian buah-buahan impor. Jurnal Ilmu Keluarga dan Konsumen. 6(1):39–48. https://journal.ipb.ac.id/index.php/jikk/article/view/7701

Wahjuni, S., Wulandari., & Nurarifah, H. (2023). Faster RCNN based leaf segmentation using stereo images. J Agric Food Res. 11 January:100514. doi:https://doi.org/10.1016/j.jafr.2023.100514

Wijaya, A. A., Cupriadi, E. D. I., & Fadel, I. (2021). Pengaruh Pemangkasan Buah Terhadap Hasil Semangka Poliploid (Citrullus vulgaris Schard L.). Jurnal Ilmu Pertanian dan Peternakan. 9:37–43. doi:https://doi.org/10.31949/agrivet.v9i1.1182