Research on Melon Fruit Selection Based on Rank with YOLOv4 Algorithm

Nur Azizah Eka Budiarti*, Sri Wahjuni1, Willy Bayuardi Suwarno2, and Wulandari1
1 Department of Computer Science, IPB University, Bogor, Indonesia
2 Department of Agronomy and Horticulture, IPB University, Bogor, Indonesia

*Email: 97_nurazizah@apps.ipb.ac.id

Abstract. Melon is one of the most popular fruits that is exceptionally favoured in Indonesia because it can be consumed directly as fresh fruit or processed as juice or salad. To meet the national market demand, several technologies are used to increase production, one of which is fruit selection. Plants need to be pruned based on fruit size so that fruit quality is maintained. One of the new approaches to detect plant fruits is using deep convolutional neural networks. The goal is to build a melon fruit detection system based on fruit size ranking for selection reliability. Recent work in deep neural networks has developed an excellent object detector, namely the one-stage You Only Look Once (YOLO) algorithm. We used the YOLOv4 model, the fourth generation of YOLO with speed acceleration and detection accuracy better than the previous versions. In addition, eight model schemes were tested with three different hyper-parameters: batch size, iterations, and learning rate. It was found that Scheme G using batch size 64, iterations 2000, and learning rate 0.001 obtained the highest score for both F1-score and mAP with values of 84.47% and 87.68%, respectively. It can be said that the F1-score value is directly proportional to the mAP value.

1. Introduction
Melon is one of Indonesia's foremost fruit products, adding up to the production of 122,150 tons in 2019[1]. The demand for this fruit is relatively high because it can be consumed as fresh fruit, juice, salad, even used as primary ingredients for industry and cosmetics [2,3]. To meet the national market demand, the main attention is focused on stability and increasing melon production. This can be maintained if handled intensively through specific treatments. One form of special treatment is thinning or pruning fruit. Pruning is a particular treatment intended for cell enlargement, reducing competition between fruits, and improving plant position to maximize fruit size and quality[4,5]. Besides, the decline in the number of farmers of productive age and dominated by old age also exceptionally affects agricultural productivity[6].

One solution is to use agricultural robots with object detection methods; besides helping and facilitating the work of farmers, they can also detect and select pruned melons using deep learning methods in real-time compared to traditional ones[7]. Two sorts of deep learning-based object detection algorithm models are developed, namely one-stage and two-stage detectors. One-stage has the advantage that its computation time is speedy compared to the two stages because of the absence of the proposed region [8].

The current eminent one-stage detection algorithm, the YOLO (You Only Look Once) series algorithm, has high detection exactness and quick detection and is broadly utilized in different target location assignments [9]. YOLOv4 is the fourth series of YOLO that puts more accentuation on the
speed and precision of object detection seen from the increment in AP (Average Precision) and FPS (Frame Per Second) by 10% and 12% compared to YOLOv3[10]. Considering that there is still a lack of research on object detection in melons accompanied by ranking, especially for pruning, this study aims to detect and select with rank the pruned melons using the YOLOv4 algorithm.

2. Research theory and method
2.1. Fruit detection
In recent years, object detection has been widely developed in the agricultural sector, specifically in detecting fruit to be harvested, called fruit detection. Whereas object detection is a procedure for estimating the approximate location of an object and the class to which it belongs by yielding the bounding box around the object[11]. Research related to deep learning-based fruit detection, such as works in [12–14].

2.2. You only look once version 4 (YOLOv4)
YOLOv4 is a one-stage algorithm that is currently excellent, has increased speed and performance is represents continued improvements to the previous generation of YOLO[10]. YOLO principally detects objects by dividing the input image into an S × S grid can be seen in Figure 1. Then the regional proposal and classification process is carried out in one network to detect so that it is called one-stage [15]. Examples of YOLO applications are presented in[16,17].

![Figure 1. YOLO algorithm detection principle](image1)

2.3. Data acquisition, augmentation features, and labeling

![Figure 2. Image acquisition schematic](image2)

The experimental data were collected from Agribusiness and Technology Park (ATP) Cikarawang Bogor, West Java, using an Intel Real Sense camera. The image-taking process is carried out by photographing the part of the melon plant that contains young melons with three viewing angles, 0°,
45°, and 90° positions, and four height levels, namely a height of 20 cm, 30 cm, 40 cm, and 50 cm from the planting medium, as shown in Figure 2. A total of 900 original images were obtained from data acquisition. Referring to Queiroga et al. [18], melon pruning keeps the fruit set between nodes 5 and 9, because it is located not far from the roots so that the optimal distribution of photoassimilate makes melons higher dry mass, flesh thickness, and taste is sweet.

The acquired data is then augmented to overcome the problem of limited data quantities to reduce overfitting. The three augmentation techniques used are crop, horizontal flip, and rotation. Then, the image is labeled. Labeling is a process of giving the same label to a set of pixels forming objects that are close together in an image [19]. The labeling process is carried out utilizing labelImg software to find out the coordinates and bounding boxes of the image. The block diagram of this research method is shown in Figure 3.

**Figure 3.** Block diagram of research method

### 2.4. Model building and training

The experimental environment is to build a connected neural network utilizing the deep learning framework YOLOv4 running on HPC of LIPI. This infrastructure was accessed using a PC with IntelCore i3-4005U processor, integrated graphics Nvidia Tesla V100-SXM2, and Linux operating system. YOLOV4 provides Darknet-53 as the backbone of the pre-trained models. This experiment uses the data set proportion of 80% training data and 20% test data. Meanwhile, there are 3 hyperparameters used, namely batch size, epoch, and learning rate. The hyperparameter configuration scheme is shown in Table 1.

| Scheme | Batch Size | Epoch | Learning Rate |
|--------|------------|-------|---------------|
| A      | 32         | 1000  | 0.001         |
| B      | 32         | 1000  | 0.0001        |
| C      | 32         | 2000  | 0.001         |
| D      | 32         | 2000  | 0.0001        |
| E      | 64         | 1000  | 0.001         |
| F      | 64         | 1000  | 0.0001        |
| G      | 64         | 2000  | 0.001         |
| H      | 64         | 2000  | 0.0001        |

### 2.5. Rank technique

The rank technique is carried out on the detected bounding box by taking the horizontal line bounding box used as the diameter or width of the fruit. Then the diameters of the fruit obtained were classified from the largest to the smallest and ranked. The bigger the size, the smaller the ranking and vice versa.
Based on expert knowledge at ATP Cikarawang and Silva et al. [5], no more than two fruits per plant are raised because it will make a better fruit quality and provide a high production value. The pseudocode of ranking and selection on melons can be seen in Figure 4.

```
Algorithm: Ranking melon for pruning selection based on width bounding box

Input: Width (Horizontal line) bbox value
Output: Ranked melon and decision pruning
1 image_h, image_w ← image.shape; out_boxes, num_boxes ← bbox
2 new_width : / /
3 for i ← range(num_boxes[0]) do
4 coor ← out_boxes[0][i]
5 coor[0] ← int(coor[0] * image_w)
6 coor[1] ← int(coor[1] * image_h)
7 coor[2] ← int(coor[2] * image_w)
8 coor[3] ← int(coor[3] * image_h)
9 x1 ← (coor[1], coor[0])
10 x3 ← (coor[3], coor[0])
11 width ← euclidean(x1, x3)
12 iwidth ← int(width)
13 new_width.extend(iwidth)
14 s ← sorted(new_width, reverse=True)
15 rank ← ranking_melon
16 for rank ← enumerate(s, 1) do
17 | print (rank)
18 if rank > 2
19 | print ("pruned")
```

**Figure 4.** Pseudocode for ranking and decision pruned

### 2.6. Evaluation

Evaluation is carried out to determine the performance of the detection algorithm. According to the assessment of the neural network model, this paper uses Recall (R), Precision (P), F1-score, and mean Average Precision (mAP) as assessment indexes. Precision measures the fraction of correctly detected items, while recall is the part of items correctly detected things among all ground truth samples [20,21]. The F1-score is a weighted comparison of the average recall and precision [22]. The True Positives (TP) demonstrates the number of melons that the algorithm has detected successfully. The False Positives (FP) demonstrates the number of non-melon objects that were wrong recognized as melons. The False Negative (FN) shows the number of melon objects that are not recognized as melons. In representation, if the F1-score has a good score, it indicates that the classified model has good precision and recall go.

The calculation Recall (R), Precision (P), and F1-score calculations are shown in Equations 1, 2, and 3.

\[
R = \frac{TP}{TP + FN} \quad (1)
\]

\[
P = \frac{TP}{TP + FP} \quad (2)
\]

\[
F1 - score = 2 \times R \times P / (P + R) \quad (3)
\]

The calculation formula for Average Precision (AP) is shown in Equation 4, where \( r \) represents the integral variable used to determine the integral precision (P) \( \times \) recall (R) between 0 and 1. Meanwhile, the mean Average Precision (mAP) is the average of several, shown in Equation 5.

\[
AP = \int_0^1 P \times R \, dr \quad (4)
\]

\[
mAP = \bar{x} \, AP \quad (5)
\]
3. Results & discussion

3.1. Results

The experimental result of the comparison of eight models is shown in Table 2.

| Scheme | TP   | FP   | FN   | Recall(R) | Precision(P) | F1-score(%) | mAP(%) |
|--------|------|------|------|-----------|--------------|-------------|--------|
| A      | 2229 | 46   | 887  | 0.72      | 0.83         | 77.1        | 78.5   |
| B      | 1776 | 678  | 1330 | 0.57      | 0.72         | 63.6        | 59.9   |
| C      | 2864 | 607  | 1154 | 0.71      | 0.83         | 76.5        | 78.3   |
| D      | 2120 | 395  | 986  | 0.68      | 0.84         | 75.1        | 76.3   |
| E      | 2356 | 557  | 750  | 0.76      | 0.81         | 78.4        | 79.6   |
| F      | 1869 | 677  | 1237 | 0.60      | 0.73         | 65.8        | 62.23  |
| G      | 2591 | 413  | 515  | 0.85      | 0.86         | 85.4        | 87.6   |
| H      | 2193 | 480  | 913  | 0.71      | 0.82         | 76.1        | 77.45  |

Table 2. Validation set experiment results.

Figure 5. Result detection, rank and decision pruning on models (a) scheme A; (b) scheme B; (c) scheme C; (d) scheme D; (e) scheme E; (f) scheme F; (g) scheme G; (h) scheme H
Based on Table 2., the average F1-score is 70%. Scheme G gives the highest F1-score, 85.4%. Meanwhile, the lowest value is obtained in the model that uses scheme B of 63.6%. The results of the F1-score in batch size 64 are higher than batch size 32. In addition, the Recall, Precision, and F1-score values are also higher at a learning rate of 0.001 than 0.0001. A significant decrease in FN and a significant increase in TP are the main reason to upgrade correctly identified positive samples, demonstrating their effectiveness of algorithm improvements.

Likewise, for the mAP value scheme G achieved the highest score of 87.6%, and scheme B gave the lowest value which did not reach 60%. For comparison, the F1-score and mAP values are not much different. So it can be said that the higher the F1-score, the higher the mAP. This shows that the better the F1-score and mAP values, the better the model detects and ranks the objects based on ground truth.

In line with the results of the F1-score and mAP, Figure 5. shows that the scheme G provides the best detection results and can provide an appropriate ranking, which predicts 5 out of 7 objects. Meanwhile, the model with the least number of detections was the schema B. model, which could only detect and rank 3 of the 7 actual objects. An overlap detected object occurred in the Scheme because only using 1000 iterations during training makes the loss value not stable. That condition needs to be avoided because it can be challenging to decide the fruit that needs to be pruned. Moreover, this model succeeds in making decisions for ranking values greater than 2 then there is a pruned sign.

3.2. Discussion
Comparing the visual result in Figure 5. Gotten utilizing model scheme G has considerably higher precision and recall values. Besides, in Figure 5. that scheme G has markedly higher values of F1-score and mAP. Thus, model scheme G for YOLOv4 can handle detection and rank better in terms of the number of detected objects compared to other models. Using YOLOv4 for detecting and ranking melons in our research surpasses the past work. The F1-score values for model scheme G on YOLOv4 from Table 2. exceed the values of F1-score 0.832 on rockmelon reported in Sa et al. [23]. In line with the results of research from Li et al. [24] and Srivastava et al.[25] who said the YOLO algorithm could be generally complete, fast, and accurate compared to faster R-CNN.

For decision pruning, ranking techniques based on fruit size are very helpful in selecting fruit. In line with research, Barzegar et al. [26] and Ferreira et al. [27] Ferreira et al. said the less fruit that is maintained at the time of pruning or thinning, the less distribution of photoassimilate to all parts of the plant. So that the fruit produced is of the maximum size, mass thickness and tasted sweet. There is a challenging problem to be able to robustly detect and rank the melon fruit only based on small and imbalanced data.

4. Conclusion
At present, deep learning detection using YOLOv4 is successfully applied to detecting, ranking, and making decision pruning for melon. We achieved 85.49% and 87.68% for the F1-score and mAP using the G scheme with 64 batch size, 2000 epoch, and 0.001 learning rate. Experimental results show that this scheme G works well on YOLOv4 to detect melons in study of using ranking for fruit selection, which is important for applications such as pruning fruits.

In further research, one could consider other tuning hyperparameters and different algorithms. In addition, including the depth data of the object as an extra input layer to YOLOv4 may be a conceivable way to get a much better result.

Acknowledgements
The authors would like to thank the Department of Computer Science, IPB University, for facilitating the work and LIPI facilitating the model building.
References

[1] BPS. Produksi Tanaman dan Buah-buahan 2019 [Internet]. (2019) Available from: https://www.bps.go.id/indicator/55/62/1/produksi-tanaman-buah-buahan.html

[2] Handajaningsih M, Hasanudin, Saputra HE, Marwanto, Yuningtyas AP. (2019) Modification of growing medium for container melon (Cucumis melo L.) production using goat manure and dolomite. Int J Adv Sci Eng Inf Technol. vol. 9(2):441–7.

[3] Abidzar Tawakal H, Prayoga A. (2019) The Development of Methods for Detecting Melon Maturity Level Based on Fruit Skin Texture Using the Histogram of Oriented Gradients and the Support Vector Machine. In: 2019 Fourth International Conference on Informatics and Computing (ICIC). Institute of Electrical and Electronics Engineers Inc. p. 1–4.

[4] Ferreira RMDA, Aroucha EMM, De Medeiros JF, Do Nascimento IB, De Paiva CA. (2018) Effect of main stem pruning and fruit thinning on the postharvest conservation of melon. Rev Bras Eng Agric e Ambient. vol. 22(5):355–9.

[5] Silva GL da, Queiroga RCF, Pereira FHF, Sousa FF de, Silva ZL da, Ferreira RP, et al. (2019) Effects of Fruit Thinning and Main Stem Pruning in Melon Crops. J Exp Agric Int [Internet]. vol. 39(3):1–10. Available from: https://journaljeai.com/index.php/JEAI/article/view/30333

[6] Prasetyo B. Indonesia Darurat Petani Muda [Internet]. (2016) Tempo.com. Available from: https://m.tempo.co/read/news/2016/08/11/090795077/indonesia-darurat-petani-muda

[7] Yadav SM, Chaware SM. (2020) Video Object Detection through Traditional and Deep Learning Methods. Int J Eng Adv Technol. vol. 9(4):1822–6.

[8] He Y, Zeng H, Fan Y, Ji S, Wu J. (2019) Application of Deep Learning in Integrated Pest Management: A Real-Time System for Detection and Diagnosis of Oilseed Rape Pests. Mob Inf Syst [Internet]. Available from: https://doi.org/10.1155/2019/4570808

[9] Hu X, Liu Y, Zhao Z, Liu J, Yang X, Sun C, et al. (2021) Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO-V4 network. Comput Electron Agric. vol. 185:106135.

[10] Bochkovskiy A, Wang C-Y, Liao H-YM. (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv [Internet].p. 1–17. Available from: https://arxiv.org/abs/2004.10934v1

[11] Pathak AR, Pandey M, Rautaray S. (2018) Application of Deep Learning for Object Detection. In: Procedia Computer Science. Elsevier B.V. p. 1706–17.

[12] Tian Y, Yang G, Wang Z, Li E, Liang Z. (2019) Detection of apple lesions in orchards based on deep learning methods of cyclegan and YoloV3-dense. J Sensors. vol. 2019.

[13] Kang H, Chen C. (2019) Fruit detection and segmentation for apple harvesting using visual sensor in orchards. Sensors (Switzerland). vol. 19(20).

[14] Stein M, Bargoti S, Underwood J. (2016) Image based mango fruit detection, localisation and yield estimation using multiple view geometry. Sensors (Switzerland). vol. 16(11).

[15] Redmon J, Divvala S, Girshick R, Farhadi A. (2016) You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society. p. 779–88.

[16] Liu G, Nouaize JC, Mboumbe PLT, Kim JH. (2020) YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3. Sensors (Switzerland) [Internet]. vol. 20(7):2145. Available from: www.mdpi.com/journal/sensors

[17] Shi R, Li T, Yamaguchi Y. (2020) An attribution-based pruning method for real-time mango detection with YOLO network. Comput Electron Agric. vol. 169:105214.

[18] Queiroga RCF de, Puiatti M, Fontes PCR, Cecon PR. (2009) Características de frutos do meloeiro variando número e posição de frutos na planta. Hortic Bras [Internet]. vol. 27(1):23–9. Available from: http://www.scielo.br/j/hb/a/Xprw9DSNyGQ8ybyv3wLRgMb/?lang=pt
[19] Mozef E. (2004) Algoritma Labeling Citra Biner dengan Performansi Optimal Processor-Ttime [Internet]. Jurnal Informatika. p. 67–77. Available from: https://www.researchgate.net/publication/43329220_ALGORITMA_LABELING_CITRA_BINER_DENGAN_PERFORMANSI_OPTIMAL_PROCESSOR-TIME

[20] Godil A, Bostelman R, Tsai WS, Shneier HM. (2014) Performance Metrics for Evaluating Object and Human Detection and Tracking Systems [Internet]. National Institute of Standards and Technology. Available from: http://dx.doi.org/10.6028/NIST.IR.7972

[21] Hui J. (2018) mAP (mean Average Precision) for object detection [Internet]. Available from: https://medium.com/@jonathan_hui/map-mean-average-precision-for-object-detection-45c121a31173.

[22] Goutte C, Gaussier E. (2005) A Probabilistic Interpretation of Precision, Recall and F-score, with implication for evaluation. In: Advances in Information Retrieval, 27th European Conference on IR Research [Internet]. p. 345–59. Available from: https://www.researchgate.net/publication/221397399_A_Probabilistic_Interpretation_of_Precision_Recall_and_F-Score

[23] Sa I, Ge Z, Dayoub F, Upcroft B, Perez T, McCool C. (2016) Deepfruits: A fruit detection system using deep neural networks. Sensors. vol. 16(8):1222.

[24] Li M, Zhang Z, Lei L, Wang X, Guo X. (2020) Agricultural Greenhouses Detection in High-Resolution Satellite Images Based on Convolutional Neural Networks: Comparison of Faster R-CNN, YOLO v3 and SSD. Sensors [Internet]. vol. 20(17):4938. Available from: https://www.mdpi.com/1424-8220/20/17/4938

[25] Srivastava S, Divekar AV, Anilkumar C, Naik I, Kulkarni V, Pattabiraman V. (2021) Comparative analysis of deep learning image detection algorithms. J Big Data [Internet]. vol. 8(66):1–27. Available from: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00434-w

[26] Barzegar T, Badeck FW, Delshad M, Kashi AK, Berveiller D, Ghashghaie J. (2013) 13C-labelling of leaf phoassimilates to study the source-sink relationship in two Iranian melon cultivars. Vol. 151, Scientia Horticulturae. p. 157–164.

[27] Ferreira RM de A, Arruda EMM, Paiva CA de, Medeiros JF de, Barreto FP. (2016) Influence of the main stem pruning and fruit thinning on quality of melon. Rev Ceres [Internet]. vol. 63(6):789–95. Available from: http://www.scielo.br/j/rceres/a/wQ4HgS4RLCpvyk7TbHYwqC/?lang=en