The Model for The Classification of The Ripeness Stage of Pomegranate Fruits In Orchards Using

Nguyen Ha Huy Cuong (nhhcuong@sdc.udn.vn)
The University of Danang https://orcid.org/0000-0003-3223-2909

Research Article

Keywords: Agriculture, Deep Learning, Deep Convolutional Networks, Ripeness Estimation, Image, Grapefruit Segmentation, Classifier

Posted Date: August 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-834262/v1

THE MODEL FOR THE CLASSIFICATION OF THE RIPENESS STAGE OF POMEGRANATE FRUITS IN ORCHARDS USING DEEP LEARNING

Nguyen Ha Huy Cuong[1], Trinh Trung Hai[2], Trinh Cong Duy[3]

[1] The University of Danang - Software Development Centre - Vietnam
[2] The University of Danang - Vietnam - Korea University of Information and Communication Technology, Da Nang City - Vietnam
[3] The University of Danang - Software Development Centre - Vietnam
[4]

nhhcuong@vku.udn.vn,,tthai@vku.udn.vn, tcduy@sdc.udn.vn

Abstract. In agriculture, a timely and accurate estimate of ripeness in the orchard improves the post-harvest process. Choosing fruits based on their maturity stages can reduce storage costs and increase market results. In addition, the estimation of the ripeness of the fruit based on the detection of input and output indicators has brought about practical effects in the harvesting process, as well as determining the amount of water needed for irrigation. pepper, the amount of fertilizer for the end of the season appropriate. In this paper, propose a technical solution for a model to detect persimmon green grapefruit fruit at agricultural farms, Vietnam. Aggregation model and transfer learning method are used. The proposed model contains two object detection sub models and the decision model is the pre-processed model, the transfer model and the corresponding aggregation model. Improving the YOLO algorithm is trained with more than one hundred object types, the total proposed processing is 500,000 images, from the COCO image data set used as a preprocessing model. Aggregation model and transfer learning method are also used as an initial step to train the model transferred by the transfer learning technique. Only images are used for transfer model training. Finally, the aggregation model with the techniques used to make decisions selects the best results from the pre-trained model and the transfer model. Using our proposed model, it has improved and reduced the time when analyzing the maximum number of training data sets and training time. The accuracy of model union is 98.20 %. The test results of the classifier are proposed through a data set of 10000 images of each layer for sensitivity of 98.2 %, specificity 97.2 % with accuracy of 96.5 % and 0, 98 in training for all grades.

Keywords: Agriculture, Deep Learning, Deep Convolutional Networks, Ripeness Estimation, Image, Grapefruit Segmentation, Classifier

1 Introduction

During the 4.0 revolution, the access to high technology helped the Agriculture sector change its fragmented, small-scale and self-sufficient production organization. In Vietnam, there are many models of the high-tech applications being implemented such as: "The best rice cultivation" model for farmers, in cooperation with technology companies, providing cultivating varieties. Smart farming (slow fertilization and one-time spraying of probiotics, using solar sensors to regulate water levels) has helped yield 7 tons of rice/ha, while reducing the seed from 20 kg/work to 6-8 kg, reduces pests, fertilizers and the number of sprays from 5 times to 3 times, and save labour. A complete application model of smart devices such as dairy farming in "TH True Milk", Its 2,000 hectares of grasslands apply a variety of automated solutions, advanced techniques from soil preparation, seeding, watering, to automatic harvesting ... with a productivity of 800 people. There are also models of smart vegetable farming monitoring and control systems that monitor and control temperature, humidity, light, ventilation and watering the plants, helping them grow better, safer, higher productivity, higher economic efficiency. Other agricultural access activities are encouraging, such as the application of automatic and semiautomatic technologies in rice, maize, fruit, vegetable production, dairy cows, breeding pigs and aquatic products. Innovation in Vietnamese agriculture does not only stop learning from international technology and techniques but also the self-exploration and creativity of Vietnamese farmers. Solutions that bring high economic efficiency, increase productivity, and output are always recommended. The solutions tend to create state-invested processes, the research teams create unique product lines, in which each new product goes through many trials until it reaches the optimum to bring to market, and then continue to be renovated and improved to meet the constantly changing needs of the market. However, many models still lack smart application solution for examples, recognizing ripe fruit at harvest time and identification of microalgae species along with the long coastlines of Vietnam. Automation in agriculture and ancillary industries are thriving today globally. The four pillars of fast-changing 4.0 technologies are Cloud Computing, Big Data, Internet of Things and Artificial Intelligence. In the rapid development of the Internet of Things (IoT) and deep learning methods, IoT is being widely applied in agriculture, automated solutions are deployed to help automate agricultural tasks such as monitoring soil moisture, pH, dispersal, automatic watering for plants, etc. [12]. The delicate combination of deep learning and artificial intelligence sciences has helped develop solutions to real-world problems that are difficult in nature. In agriculture, especially in Vietnam and Southeast Asia, important issues need to be identified, such as pests causing plant diseases from leaf images, warning the causes weather causes a decrease in productivity, it is necessary to have a weather forecast to cultivate suitable crops based on the weather of the common seasons in Vietnam, the solution to check the plowed soils and do well before plant breeding grows successfully. A meticulous awareness of near-real-time fruit detection (combined with ripe or unripe sorting) will help in improving the quality and yield of the fruits during harvest, yield estimates and corresponding

harvest times. Automated robotic harvesting systems can use this localization system [1] [3] [6].

Previous research methods considered the ripening of fruit to be an object detection problem in YOLO-v3 [3] [6], [21] [19] [15] [16] which significantly increased the training fee.

In our research approach, our solution is divided into five parts of the image classification of fruits: fruit detection and fruit ripening state, fruit ripening state [13]. Scope of this article:

- Collecting image data of grapefruit from different farms, different grapefruit varieties.
- Caption image data with YOLO-v3 tool in [6].
- Training YOLO-v3 on data with darknet53 weight as the backbone.
- Training a set of images to classify pomelos with quality grapefruit and grapefruit varieties.
- Training the image classifier for detecting the ripening stages of the grapefruit fruit.

For the purpose of this paper, developed methods that are beneficial in reducing the cost of time spent on training and the flexibility in connecting different classification tasks such as detecting diseases on the harvested grapefruit.

In this paper, develop a collective modelling method to detect objects, the solution will not change and there is no need to retrain any learning steps even if we increase bulk input data. As large data increases, this newly increased dataset only needs training as an additional model. The proposed model has three parts including: the data input to be processed, the model transferred to detect additional objects, and the aggregation model to determine the final decision

The model to be transferred is very simple, the model of detecting augmented objects using the transferred deep learning method. This has reduced the time and number of images needed during the model training. Aggregation model is made using the synthesis technique of bootstrap component (Bagging).

The rest of the paper is structured as follows. In section 2, we present related work. We describe data accumulation, annotation tool and its usage and literature survey is discussed in section 3. In section 4, we present our proposed approach. Experimental results are detailed in Section 5. Finally, we conclude with future work directions and limitations in section 6.

2 Related Works

Our problem is to provide a technical solution, with the goal of detecting ripening objects from the identification of sets of objects from trained object classes, or in other words, object class identification. from the pictures. The output can simply locate each detected object in the image for a limited amount of time, based on the information, along with the name of the object class. The object detection input can be an image or a video with one or more objects.

You Only Look Once (YOLO) and R-CNN family include (R-CNN, Fast R-CNN, Faster R-CNN, and Mask R-CNN)... are modle popular object detection.

Ross Girshick, Shaoqing and Joseph Redmon and colleagues provided the research models here, we found that there are differences in applied algorithms, calculation time and the performance of solutions are different. [21] [19] [15] [16]

Two research groups of Zhong-Qui Zhao, Ross Girshick, proposed model (R-CNN) [5] [21] This is a basic technique that uses Neural networks to detect objects, but this technical solution requires a lot of time. processing time. The input to the technique using the R-CNN model is an image that is extracted into small dimensions called area suggestions for ease of handling. The R-CNN model uses a selective search method to extract reference ranges, areas that can be divided into groups of three objects or a set of objects. Selective search provides a range of candidate suggestions. Next, all of these regional proposals are packaged and sent to the cumulative Neural Network (CNN) [18]. Then, use the Support Vector Machine (SVM) to classify the presence of the object [14].

In recent years, researchers have improved the CNN model that processes each image of one input one after another, announcing the R-CNN improvement model. The R-CNN model has a faster computation time, the research method uses a region of aggregated group of the same size to create a fixed-size vector of regional proposals. Quick R-CNN provides all images based on CNN training techniques to create cumulative feature maps. The Softmax function is used in the FC class to classify objects. [19] [8] [5]

Recently, the Fast R-CNN technical solution has shown to be more effective than the R-CNN method, the input images are included in CNN to create feature maps. Regional suggestion network (RPN) is used to create regional recommendations instead of selective search method in R-CNN and Fast R-CNN. RPN method helps reduce the time in the process of making regional proposals. Subsequently, regional proposals were presented for the lumped class to create a fixed size [5] [15].

Mask R-CNN use the same basic structure as Faster R-CNN. The method Mask R-CNN able to align pixel-to-pixel and has a better instance segmentation while the pooling layer in Faster R-CNN unable. [5] [9] [16].

Recently when detecting subjects researchers use YOLO algorithm. The YOLO algorithm has effectively brought the fact that it is faster, stronger and more efficient than all previous R-CNN families. For the R-CNN family, multiple regional proposals need to be created and included in CNN to predict and locate objects. The YOLO algorithm used does not need to generate regional suggestions, it uses a single convolutional network and only sees the algorithm of the image once in the image [5] [9] [16].

The image is divided into grid cells then creating bounding boxes. For each bound box, the model gives a layer probability to position the object in the image. Deep learning method is used a lot and performance is increasingly improved. However, if models detect new objects, the model needs to retrain the models to recognize both old and new objects. Usually, training a deep learning model requires a lot of training data due to the large number of model parame-

ters. The detection of objects by deep learning is no exception. It is a process that consumes time and calculation. Therefore, this study aims to propose a model of increasing object detection by quickly detecting and avoiding retraining of [5] [16].

Suchet Bargoti et al. [11] used method detect fruit in orchards by implementing a Faster-RCNN to localize fruits (mangoes, almonds, star apple, and apples), but method of Suchet Bargoti et al. [11] not to detect ripening stages.

A. Koirala developing an architecture MangoYolo based on YOLO-v2, YOLO-v3 to detect mangoes in orchards with an average precision of 0.983% [**?**].

Tian et al. using YOLO-v3 and detecting different growth stages of apples in orchards. The ripening detection problem in requests to train different stages of apple growth as discrete objects and therefore, so it would induce immense training costs [3]. S. Kim et al, proposed detection stage added to the model given by . [7].

3 Data collection and annotations

In this paper, used image-net open-source data and images of green grapefruit are used on agricultural farms. Advanced search on agricultural farms throughout the territory of Vietnam, allowing to collect delicate data sets of grapefruit images from many orchards. The total number of images is 5000. Here are five sample images from the cumulative dataset.

Fig. 1. Sample images from the data set

Total 5000 Images were annotated using YOLO-v3 tool in. To detect and avoid overfitting in classification, in this data-set apply method an augmentation for generating more training data from collection image data. In the test scenario, employed also various image transformations such as zoom, lighting, flip, rotate and warp. [6]

In this paper, the model proposed consists of two model detection and decision model which are a pre-trained model, a transferred-model and an ensemble model as shown in Fig 2.

Figure 3 and Figure 4 shows example images from an initial one. The model a detection created four artificial randomly transformed images for one actual image.

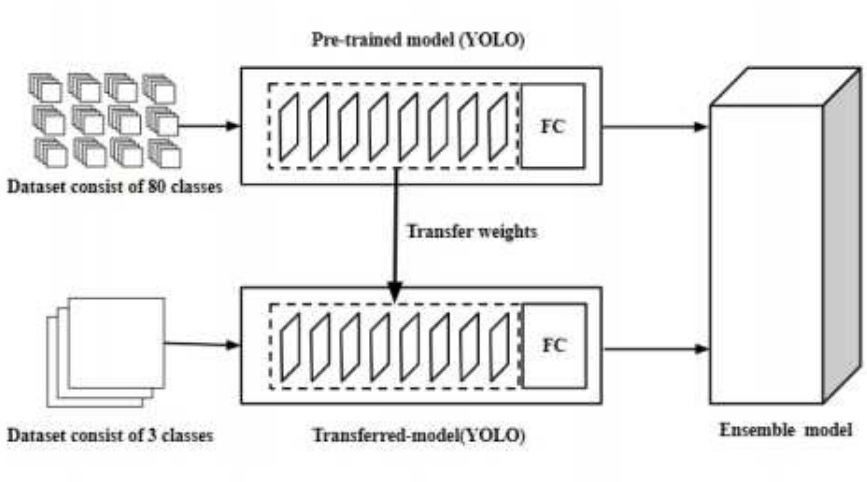

Fig. 2. Proposed Model for Incremental Object Detection with YOLO as a pre-trained model

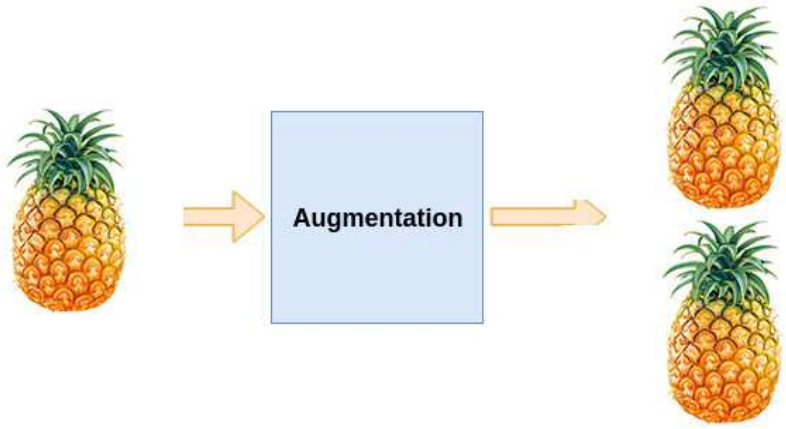

Fig. 3. Augmentation and transformations for grapefruit fruits images

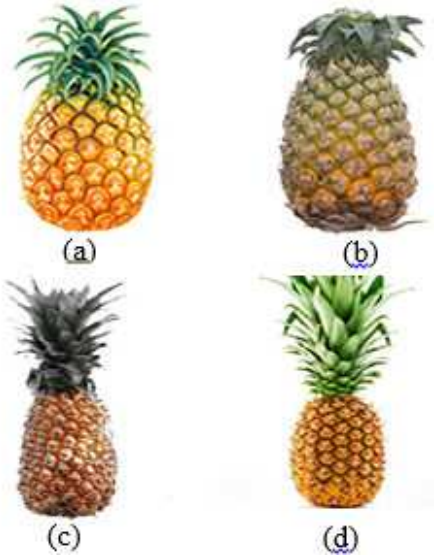

Fig. 4. (a, b) Ripe grapefruit (c, d) Unripe grapefruit

4 Proposed Method

In the session, the introduction, the objective of the article is to provide technical solutions for detecting faulty grapefruit for specific classification after the remaining grapefruit is free to detect ripe fruits. The proposed method is implemented by the improved YOLO-V3 toolkit [16].

4.1 YOLOv3

J. Redmon, A. Farhadi et al. [16] announced YOLO V3, is an improvement of YOLO-v2 and YOLO [6]. The technique of YOLO algorithms to detect objects is unlike the method of Faster RCNN, Family CNN was previously published. YOLO V3 uses a combination of regression techniques, bounding boxes processing, and class probability calculation through regression. This innovative solution has increased the computational speed, thus yielding quite high test results.

The image of input data collected is divided into a matrix grid (S x S) by YOLO V3. If an object with a center is predicted to fall into a certain grid, then that grid is responsible for detecting the object. Check by browsing by column and row of the grid matrix. If no object falls into the grid, the test object cannot be detected, resulting in a zero-trust score. YOLO V3 proposed method in the paper, using the row-and-column calculation to delete the checked objects. This technique will gradually and gradually eliminate the object that has been marked. As a result, the solution selects the best limit boxes if more than one limit box detects the same object. The technique also solves the object

layers in individual grids, it also calculates the number of bounding boxes and the probability of detecting the object classes tested [4] [8] .

4.2 Image Classifier

Previous studies have shown that the pre-trained model has yielded satisfactory results in test scenarios such as VGG etc. In the paper, Resnet architecture was used as the backbone for classifying images from images collected by the data warehouse. Resnet technology is assessed quite quickly in training the given sample images, reducing lost signals, increasing the appearance effect and having quite high accuracy. The test scenario has progressed with the Resnet100 architecture for the task of categorizing the training images obtained [17].

Figure 5 illustrates the object detection process by the proposed model of the paper. After calculation, by using the mesh matrix browsing method (SxS), the objects are detected and the probability that the class is detected. The model detected an object of a ripe grapefruit when the skin color turned to a glossy yellow with a probability of 0.9868. The following results were tested and continued after that, with positive results. The corporate model should choose the highest probability class as the most accurate final decision [20].

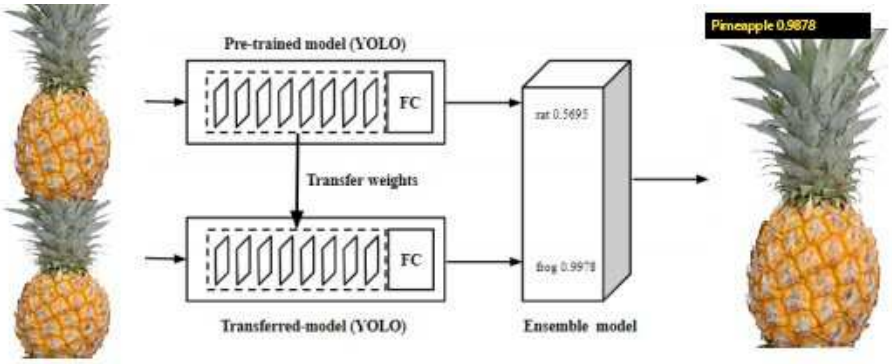

Fig. 5. grapefruit regions detected by YOLOV3 R: Ripe UR: Unripe

Real-time detection of ripe grapefruit is an extremely difficult problem. Therefore, in this paper, technical solutions are divided into 5 stages. First, processing input data, sorting, and training data sets are images of grapefruit by toolkit YOLO V3 improved.

Secondly, in the step presented to build a degree classification procedure based on CNN resnet100 architecture. In this step, you can classify ripe or unripe fruits.

In the real-world scenario, when conducting experiments on agricultural farms in Vietnam, when the input data set is an image of grapefruit from agri-

cultural farms sent to the improved YOLO-v3 network. Experiments can detect ripe, unripe grapefruit based on the graded frames trained. Through practice, it can be segmented into useful tasks that help warn decisions to assist farmers, in the classification process, as well as diagnose pests and diseases that affect the standard fruits. be planned. Helping farmers to stop growing spraying and preventing pests and diseases, etc.

5 Results and Discussion

To train YOLO-v3, the following are initial parameters [23] .

Table 1. Parameters for Experimental analysis

	Image Size	Batch Size	Decay	Momentum	Training steps
1	1366 x 768	64	0.9	0.00005	5000
2	1360 x 768	64	0.9	0.00005	10000
3	1280 x 720	64	0.9	0.00005	20000
4	1280 x 600	64	0.9	0.00005	25000
1	1366 x 768	64	0.9	0.00005	25000
2	1360 x 768	64	0.9	0.00005	30000
3	1280 x 720	64	0.9	0.00005	35000
4	1280 x 600	64	0.9	0.00005	45000

Improved YOLO-v3 model was trained with 5000 training steps on a HP k80 GPU, achieving the mean accuracy of 89%.

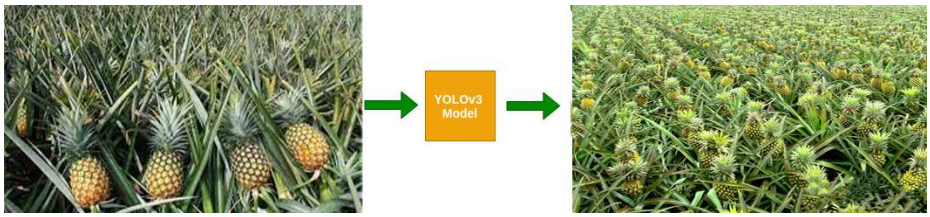

Fig. 6. Input frame (to the Yolo model) B. Output detections

YOLO-v3 improved model detected only 90 out of 100 grapefruits, present in the input frame in Fig 6 achieving the mean accuracy of 90%.

5.1 Classifier implementation

In the data set in Table 5.2 explains the causes of the losses, the losses are considered valid and the error rate in 5 training cycles with the learning rate of

0.0005. The accuracy we achieved was 97%. The proposed classification process was based on PyTorch's deep learning library. We have trained the Resnet100 model with all the fruits in include [23] .

Table 2. Loss and Accuracy in Classification

Image Size	Epoch	Training loss	Validation loss	Accuracy
1366 x 768	1	0.81	0.16	0.99
1360 x 768	2	0.91	0.11	0.98
1280 x 768	3	0.89	0.08	0.97
1280 x 720	4	0.87	0.06	0.95
1280 x 600	5	0.85	0.03	0.98
1024 x 768	6	0.71	0.16	0.93
800 x 600	7	0.83	0.11	0.96
1280 x 600	8	0.85	0.03	0.98
1024 x 768	9	0.90	0.16	0.93
800 x 600	10	0.94	0.11	0.98

Below is the confusion matrix for the trained baseline model. A confusion matrix helps evaluate the performance of the trained model on a data-set with known target classes. The model was evaluated by calculating precision and recall, using True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN) in the confusion grid matrix.

Example 7. Confusion matrix with ripe and unripe accuracy (baseline model)

Predictive values given by Precision = TP / TP + FP. Sensitivity according to Recall = TP / TP + FN. When assessing the results of the test scenarios, the results are predictive values and the sensitivity is equal to 0.967. Figure 8 confirms the classes processed during the base model training.

Example 8. Resulted performance for Training vs Validation phases (baseline model)

Finally, to conduct evaluation and make decisions to confirm ripe grapefruit with unripe fruit. In the paper, the refining and release techniques were applied to all layers of the input image data dataset trained from the improved YOLO V3 model. In order to maintain the calculation results through the trained steps, the set of typical classes of the trained base model. Experiments conducted to extract some of the selected data a subset compared to the initially divided classes. Experiment still holds the learning rate is 0.00005 for the initial classes and 0.000055 for the final grades. After 20 minutes of training, the accuracy achieved is 98.33 %.

The results obtained from the collection and calculation based on the probability of accurate evaluation and the calculation results from the based on confusion matrix for improved model are 0.97 and 1. respectively. The full observation of the empirical cases the problem of training loss is very low and the technique

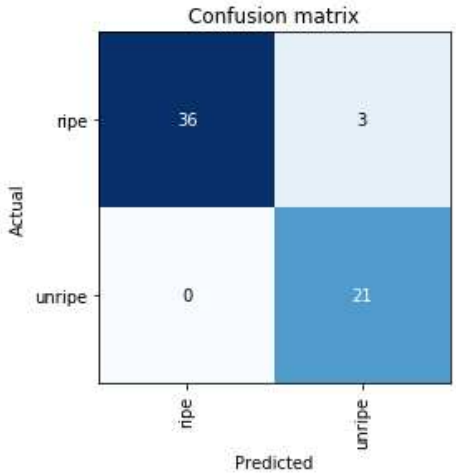

Fig. 7. Confusion matrix with ripe and unripe accuracy (baseline model)

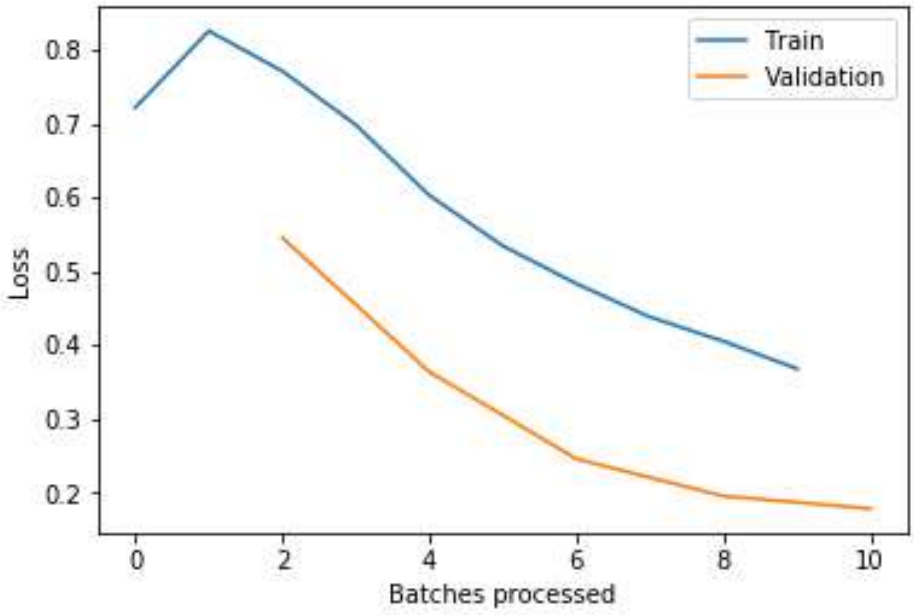

Fig. 8. Resulted performance for Training vs Validation phases (baseline model)

Table 3. Loss and Accuracy in DNN

Image Size	Epoch	Training loss	Validation loss	Accuracy
1366 x 768	1	0.11	0.16	0.93
1360 x 768	2	0.11	0.11	0.96
1280 x 768	3	0.09	0.08	0.96
1280 x 720	4	0.07	0.06	0.95
1280 x 600	5	0.05	0.03	0.98
1024 x 768	6	0.11	0.16	0.93
800 x 600	7	0.13	0.11	0.96
1280 x 600	8	0.15	0.03	0.98
1024 x 768	9	0.10	0.16	0.93
800 x 600	10	0.14	0.11	0.96

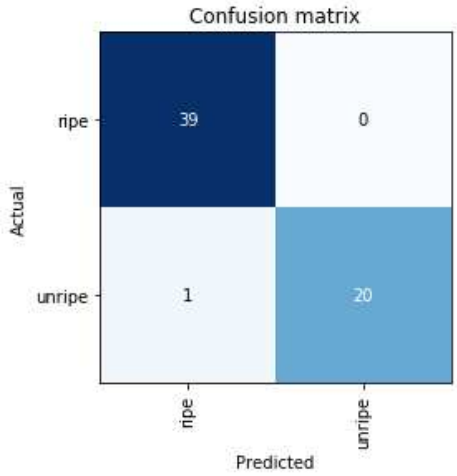

Fig. 9. Confusion matrix with ripe and unripe accuracy (fine-tuned model)

meeting the results of the high validation is quite high. Therefore, the model is evaluated quite effectively to bring economic benefits to society which is quite high.

5.2 Fruit classification

In Figure 11, ripe and unripe grapefruit, discovered from improved YOLO-v3, has been improved to provide trained image classifiers. These results allow farmers to cockroach ripe or unripe fruit. The results of this research are very practical and have brought a lot of benefits to farmers, when conducting the supply to supermarkets and exports to other countries around the world.

Example 11. Results of the proposed model

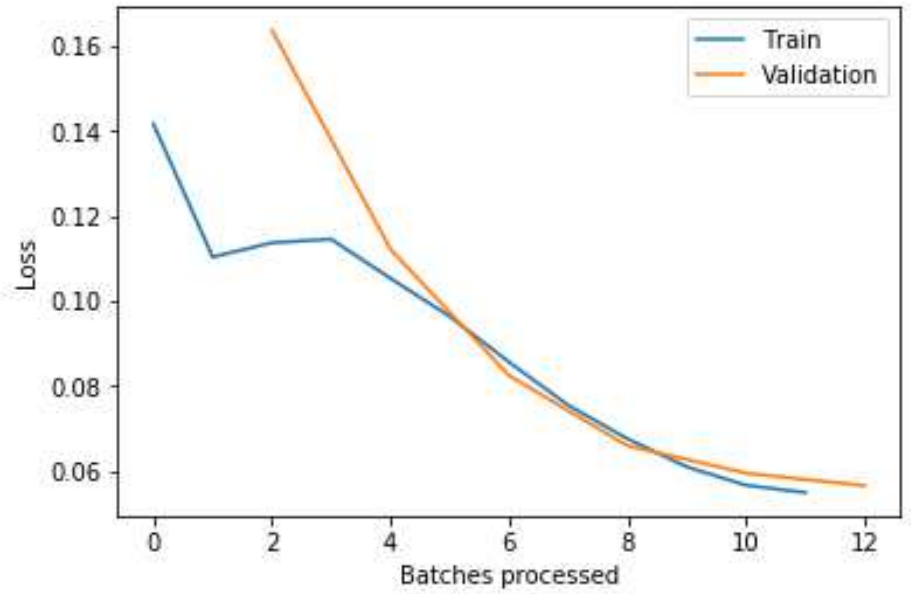

Fig. 10. Resulted performance for Training vs Validation phases (fine-tuned model)

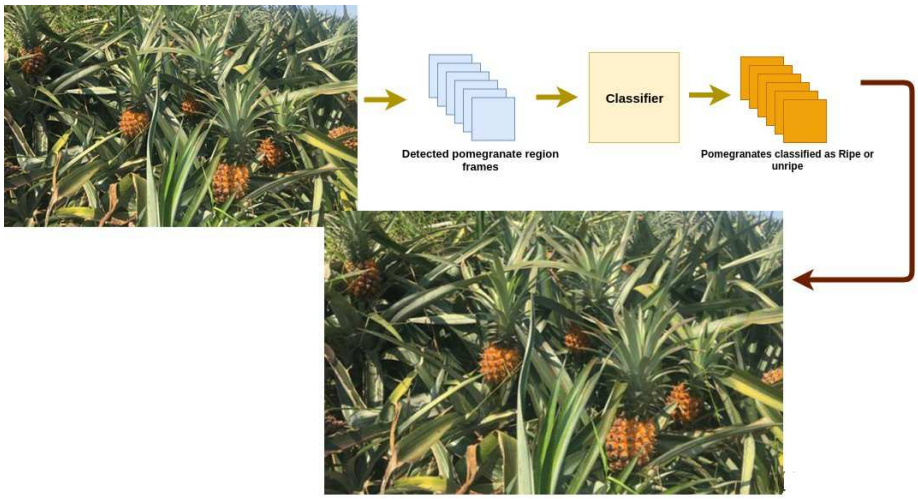

Fig. 11. Results of the proposed model

6 Conclusion

In this paper, we published effective technical solutions to detect grapefruit on agricultural farms in Vietnam and classify them as ripe or unripe. The purpose of the article is to develop a new algorithm that detects the middle nine stages. A Deep Neural Network classification process (DNN) is used to predict ninth and unripe grades. Therefore, the solution has been effective by reducing the cost of heavy training, instead of using the YOLO detection model to detect and classify grapefruit simultaneously. Proposing an aggregation model aimed at increasing the discovering of new objects, reducing time and calculating for model training. The results of the proposed research technique are: Firstly, in the experiments, the first sub-model is a pre-trained model trained with eighty objects, a total of 500,000 images, from the COCO image data set. Second, the transfer model using transfer learning techniques only trains new images containing three additional objects. There are about 5,000 total images, one thousand images per layer. It only takes a few hours to train the model. Finally, the aggregation model using the bagging technique is used as a decision model. The accuracy of our proposed model is as high as 98.3%. Our proposed method is flexible because various classification tasks (such as disease detection in grapefruit) are easily added to the model.

Acknowledgment

The article was conducted with the support from the scientific research project University of Danang.

References

1. Lakshmanaprabu S.K., Sachi Nandan Mohanty, Shankar K., Arunkumar N.: Optimal deep learning model for classification of lung cancer on CT images, Future Generation Computer Systems, 19(1), 374-382 (2019).
2. P. Rajeshwari, P. Abhishek, P. Srikanth, T. Vinod.: Object Detection: An Overview. International Journal of Trend in Scientific Research and Development (IJTSRD), 3(1), 1663-1665 (2019)
3. Tian, Y., Yang, G., Wang, Z., et al.: Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J]. Computers and Electronics in Agriculture 157, 417–426 (2019).
4. Xiongwei Wu, Doyen Sahoo, Steven C.H. Hoi.: Recent Advances in Deep Learning for Object Detection; arXiv:1908.03673v1 [cs.CV], Aug. (2019)
5. Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, Xindong Wu.: Object Detection with Deep Learning; arXiv:1807.05511 [cs.CV], Apr. (2019)
6. A. B. Alexey: Apple detection during different growth stages in orchards using the improved YOLO-V3 model. https:
github.com/AlexeyAB/Yolo (2018)
7. S. Kim, Y. Ji and K. Lee.: An Effective Sign Language Learning with Object Detection Based ROI Segmentation, second IEEE International Conference on Robotic Computing (IRC), Laguna Hills, CA, 330-333 (2018).
8. Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, Stephen Marshall.: Activation Functions: Comparison of trends in Practice and Research for Deep Learning; arXiv:1811.03378 [cs.LG], Nov. (2018).
9. Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick.: Mask R-CNN; arXiv:1703.06870 [cs.CV], Jan. (2018)

10. Munera, S., Amigo, J. M., Blasco, J.; Cubero, S.; Talens, P., Alexios, N.: Ripeness monitoring of two cultivars of nectarine using VIS-NIR hyperspectral reflectance imaging. Journal of Food Engineering, 214(3), 29-39 (2017).
11. S. Bargoti and J. Underwood.: Deep fruit detection in orchards; IEEE International Conference on Robotics and Automation (ICRA), Singapore, 3626-3633. doi: 10.1109/ICRA.2017.7989417 (2017).
12. Yuting Zhang, Kihyuk Sohn, Ruben Villegas, Gang Pan, Honglak Lee.: Improving Object Detection with Deep Convolutional Networks via Bayesian Optimization and Structured Prediction; arXiv:1504.03293 [cs.CV], Jan. (2016).
13. Yongxi Lu, Tara Javidi, Svetlana Lazebnik.: Adaptive Object Detection Using Adjacency and Zoom Prediction; arXiv:1512.07711 [cs.CV], Apr. (2016).
14. Santagapita, P.R., Tylewicz, U. Panarese, V., Rocculi, P.,Dalla, Rosa, M.: Non-destructive assessment of kiwifruit physic-chemical parameters to optimize the osmotic dehydration process, A study on FT-NIR spectroscopy. Journal of Biosyst. Eng. 142(2), 101-129 (2016).
15. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks; arXiv:1506.01497v3 [cs.CV], Jan.(2016)
16. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi.: You Only Look Once: Unified, Real-Time Object Detection; arXiv:1506.02640 [cs.CV], May (2016).
17. K. Simonyan and A. Zisserman.: Very deep convolutional networks for large-scale image recognition. In ICLR, (2015).
18. Keiron O'Shea, Ryan Nash.: An Introduction to Convolutional Neural Networks; arXiv:1511.08458 [cs.NE], Dec. (2015).
19. Ross Girshick.: Fast R-CNN; arXiv:1504.08083 [cs.CV], Sep. (2015).
20. Dumitru Erhan, Christian Szegedy, Alexander Toshev, Dragomir Anguelov.: Scalable Object Detection using Deep Neural Networks. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2147-2154 (2014).
21. Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik.: Rich feature hierarchies for accurate object detection and semantic segmentation; arXiv:1311.2524 [cs.CV], Oct. (2014).
22. Jia, K.,Wang, X., Tang, X.: Image transformation based on learning dictionaries across image spaces. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(2), 367–380 (2013).
23. Common Objects in Context. [online]. Available: http://cocodataset.org/
24. Open Images Dataset V5. [online]. Available: https://storage.googleapis.com/openimages/web/index.html.