SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V10I7P103

Volume 10 Issue 7, 31-40, July 2023

Original Article

CNN-YOLOv8 - Based Tomato Quality Inspection System - A Case Study in Vietnam

Thi-Mai-Phuong Dao[1*], Ngoc-Khoat Nguyen[2], Van-Kien Nguyen[1]

[1]*Faculty of Electrical Engineering, Hanoi University of Industry, Hanoi, Vietnam*
[2]*Faculty of Control and Automation, Electric Power University, Hanoi, Vietnam*

[*]*Corresponding Author : daophuong@haui.edu.vn*

Received: 23 April 2023 Revised: 21 June 2023 Accepted: 09 July 2023 Published: 31 July 2023

Abstract *- Quality classification is one of the final stages of the process of consuming agricultural products not only in Vietnam but also in other countries having agriculture sector. It determines the quality and directly affects the price of agricultural products in the market. With the significant development of science and technology, many advanced techniques have been used, in which computer vision and artificial neural networks have been widely applied with undeniable achievements. This has helped increase the quality of agricultural products, improve sorting efficiency, and reduce operating costs. In the present study, the YOLOv8- based deep learning network model using a convolutional neural network is proposed to solve the problem of detecting several surface diseases on the tomatoes considered significant crops in tropical countries, e.g., Vietnam. The results of training the YOLOv8 network model with a dataset of 500 product images, including both good and bad features, mean Average Precision (mAP) value up to 99.5%, a precision of 96.3%, and recall of 96.1% demonstrate a statement: the YOLOv8 algorithm can be effectively applied in agricultural product quality inspection systems.*

Keywords *- Agricultural inspection, Computer vision, YOLOv8, Convolutional Neural Network (CNN), Tomato quality.*

1. Introduction

Vietnam is a developing country, and the economy heavily depends on agriculture. As a tropical country, agricultural products still contribute a significant share of exports in Vietnam. Therefore, the quality of agricultural products is increasingly required to meet strict export standards for fastidious markets such as Europe, America, and Japan. Even in Vietnam, introducing Vietnamese Good Agricultural Practices (VietGAP) standards is a breakthrough, requiring serious efforts and development of agriculture, especially in the export domain of agricultural products [1].

To achieve the VietGAP and export standards, the problem of identification and classification of agricultural products has become highly important. This problem has appeared for a long time, and many scientific studies and reports have been published to solve it. One of the earliest methods is *image processing*, which focuses on developing algorithms to extract information from input images, such as colour, shape, and size, for fruit identification [2-6]. However, the results achieved by these methods are not highly sufficient, and the application scope for many different agricultural products is also limited, owing to the complex variation in the characteristics of colour, shape, size, etc.

With the tremendous development of computer science, deep learning-based models have emerged as the most advanced techniques for image classification. They hold great promise in challenging fields, e.g., agriculture, where they can handle large variabilities of data better than classical computer vision methods. Deep Learning (DL) is one of the most popular Machine Learning (ML) based methods. An essential feature of DL is its ability to automatically learn and analyse patterns contained in images [7-10]. With the rapid development of deep learning techniques, convolutional neural networks (CNNs) have been successfully applied in agricultural research to solve the limitations of machine learning.

A CNN model usually consists of two leading operators: convolution and composite layers. The convolution layer can automatically extract an image's more critical and complex features. Due to the high computational cost of the convolutional network, the aggregation layer reduces the number of parameters of the data. Most current studies investigating the classification and identification of agricultural products are based on the CNN model. Since 2012, when Krizhevsky [11] won the ImageNet competition (ILSVRC), CNN has become a popular method for image classification in many fields because it has obtained effective results.

In the agricultural field, CNN-based methods have been applied to classify agricultural products [12-16] and detect them [17-20]. These methods have proven the feasibility and efficiency of fruit identification and quality assessment in the agricultural industry. In the CNN architecture, the You Only Look Once (YOLO) algorithm stands out for its remarkable balance between speed and accuracy, allowing fast object recognition and reliability [27]. Many articles and scientific studies have applied the YOLO algorithm to classify agricultural products [21-26] and achieved remarkable success. In [21], the authors used YOLOv3 to identify and classify chilli quality; in the best condition, the classification accuracy reached 99.4%, but in the worst case (the peppers are on top of each other), the correct result was only 75.6%. J. Yao et al. [23] proposed an advanced YOLOv5 network to identify defects in kiwifruit, showing that the error detection results reached 94.7% and meaning that it was 9% higher than the original algorithm. In [24], the authors simulated two networks, YOLOv5 and Mask R-CNN, to recognize a variety of fruits and their quality using a dataset of 10545 images. The results for YOLOv5 were superior to those of the Mask R-CNN method when real-time object detection was required. J. L. De Moraes and colleagues [26] used the YOLOv7 algorithm to identify and classify nine different diseases on papaya fruit, achieving an average mAP of 86.2%, even at types with high internal variation, such as “mechanical damage”.

This study applies the YOLOv8 algorithm, the newest version of YOLO software, to classify the tomato quality. The primary goal of this paper is to design a classification model, including hardware and software, effective in assessing the quality of tomatoes. In addition, such a detection and classification system is integrated with a control and supervisory part to build up a successfully applicable prototype. Promising practical results obtained with such a tomato quality inspection model verify the dominant applicability of this work.

2. Quality Inspection System for Agricultural Products

2.1. Schematic Design of the Inspection System

The schematic diagram of the classification system is illustrated in Figure 1. Some explanations are as follows:

PC/Laptop (Personal computer/Laptop computer) block is set with classification software to receive product images through the classification camera. Besides, this PC sends the control signal to the PLC and observes the operation of the whole system. PLC (Programmable Logic Controller) is used to receive the control and classification signals from the PC via Modbus TCP/IP protocol in combination with signals from sensors; thereby, it can regulate an actuator employed for product classification. Sensors and push buttons receive the control signals and determine the positions of agricultural products on the convey. The actuator controls the convey and the classification part. POWER is to provide a 220VAC source for the PC and a 24VDC source for the PLC, sensors, and actuators.

2.2. Overview of YOLOv8

The YOLO is a fast and accurate object detection model in computer vision. With the YOLO method, object detection and positioning occur only once on the entire image instead of dividing the image into small areas for processing as in other traditional methods. The YOLO uses a convolutional neural network (CNN) architecture to extract features from images and then applies convolutional layers to predict feature locations and types. This helps the YOLO achieve fast processing speed and accurate object detection. The YOLO model was improved over versions such as YOLOv2, YOLOv3 and YOLOv4, with increased accuracy and performance. The YOLO has been widely applied in practical applications, such as object recognition, object tracking, auto-driving, and security monitoring.

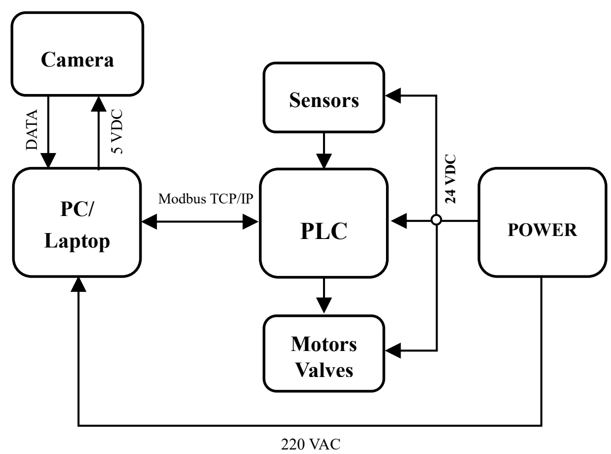

Fig. 1 The schematic diagram of the proposed system

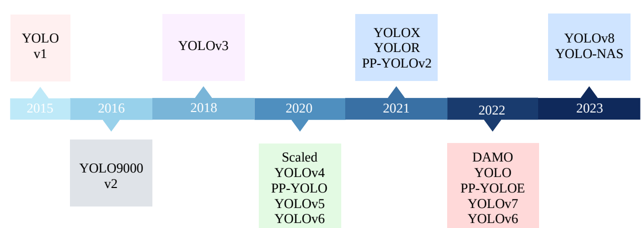

Fig. 2 The developing timeline of different versions of YOLO [27]

Until now, the YOLOv8 is the latest version of YOLO for ultralytics. As a modern cutting-edge model, YOLOv8 is buillt on the success of previous versions by introducing new features and enhancements to enhance performance, flexibility, and efficiency. The YOLOv8 fully supports vision AI tasks, including detection, segmentation, posture estimation, tracking, and classification. This flexibility allows users to leverage YOLOv8 capabilities across various applications.

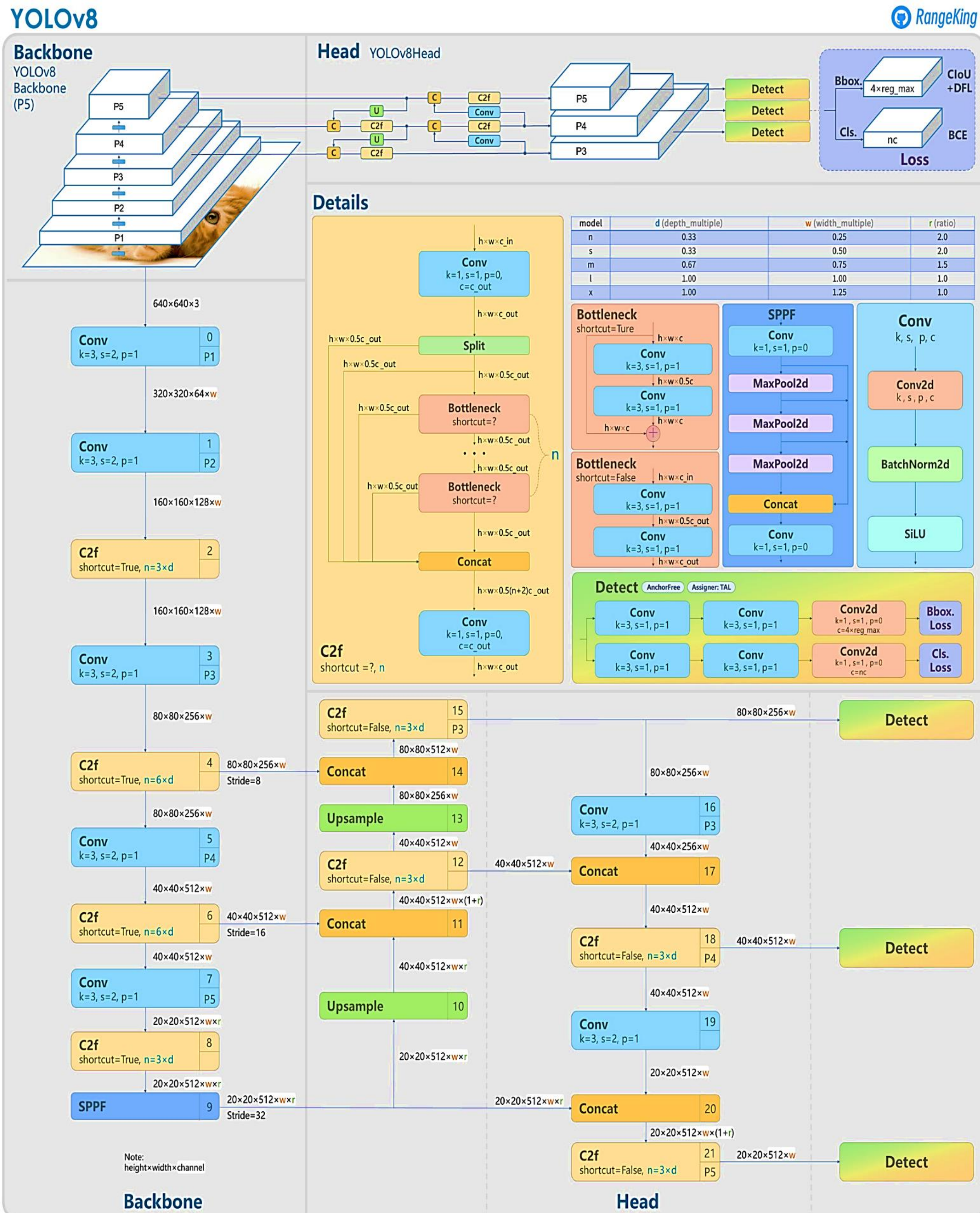

Fig. 3 The architecture of YOLOv8 created by the user: GitHub Range King

Figure 3 shows the detailed structure of the YOLOv8. The YOLOv8 network architecture consists of two main parts: the backbones and the head. The YOLOv8 uses a convolutional neural network (CNN) as the backbone to extract features from the YOLOv5 analog input images. This convolutional neural network is typically built on top of *EfficientNet* or *CSPDarknet*53 architectures, which are powerful and efficient architectures for feature extraction. The *C*2*f* module replaces the CSPLayer used in YOLOv5 [27]. In the YOLOv8 architecture, the head part is the model component placed at the end of the processing. It generates the final predictions regarding object detection, classification, and regression. The head of YOLOv8 includes two components: *Detection Layer* (responsible for predicting the bounding box and objectness score) and *Activation Functions* (objectness scores in YOLOv8 are usually activated by the sigmoid function). These functions show the probability that a bounding box contains an object. The softmax function usually activates classification probabilities showing an object belonging to different classes.)

In addition, YOLOv8 applies remarkable loss functions to train the model efficiently. Typically, the loss function *CioU* (Complete Intersection over Union) and *DFL* (Distribution Focal Loss) are used [27] to ensure that the location prediction of the bounding boxes and the feature classification are performed accurately and stably. Combining these loss functions, the YOLOv8 achieved good results in object recognition and detection in images and videos.

2.3. Procedure to Design the Fruit Classification System

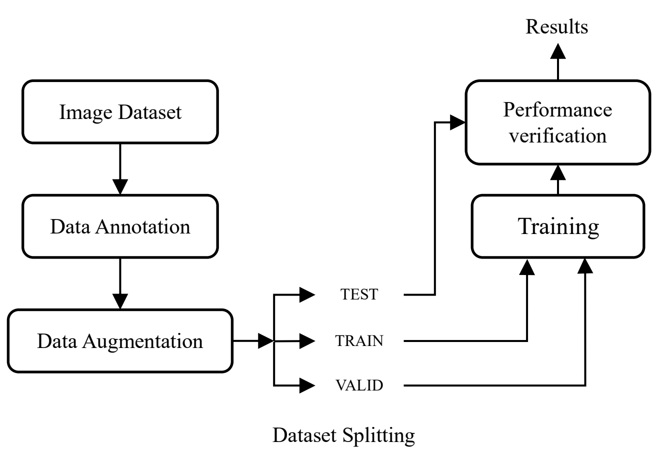

Fig. 4 A procedure to design the classification system

This work aims to identify the surface defects of tomato fruits using Deep Learning as the system's main component. An overview of the steps involved in building the classification model is presented in Figure 4.

2.3.1. Design of Sample Dataset and Its Labelling

Building a sample set for the fault identification problem on tomatoes using the YOLOv8 network is crucial to ensure the model can learn the product features and detect possible errors.

The original dataset consisted of two images: good and bad tomatoes, including 370 images with full HD resolution. Good tomato images are images of good quality in appearance that are uniform in shape, free from distortion, warping, or cracking; have a uniform bright red colour over the entire fruit; and are free from black spots or blemishes, beautiful colour, and smooth. Tomatoes should be medium, not too small, or too large for the standard. In contrast, the image of defective tomatoes contains errors such as cracked, rotten, crushed, corroded, and infected with pests. In addition, poor-quality tomatoes often have signs such as non-uniform colour, uneven fruit, distorted or bent, with cracks or dark spots.

Labelling the dataset is essential in classifying error-free and error-free tomatoes using YOLOv8. By adequately labelling each image, we can determine the location and type of the tomatoes in the image. This allows the YOLOv8 deep learning model to recognize and classify defective and non-faulty tomatoes accurately.

Labelling involved assigning a corresponding label to each tomato in the image. Usually, we use special software or graphics tools to draw bounding boxes around each tomato and specify the corresponding label, such as "OK" or "NG.” This process requires care and precision, ensuring the labels are affixed to the correct position and type for each tomato in the dataset.

2.3.2. Data Annotation

Deep-learning models usually provide better results when there is a large amount of data. However, collecting large amounts of data is complex, and the available data sources are limited. This led to missing data in the analysis. In addition, the lack of data causes overfitting in the training process. Data augmentation has been proposed to solve these problems.

In this study, several techniques have been used to increase the number of images in the dataset. These techniques include geometry transformations (rotating and flipping the image), saturation transformations (from -23% to 23%), and exposure transformations (from -25% to +25%). This helps expand and diversify the training dataset and improve the model's generalizability. After the data enhancement step, we obtained a tomato classification dataset of 500 images. Figure 5 shows an example of increasing the size of the dataset.

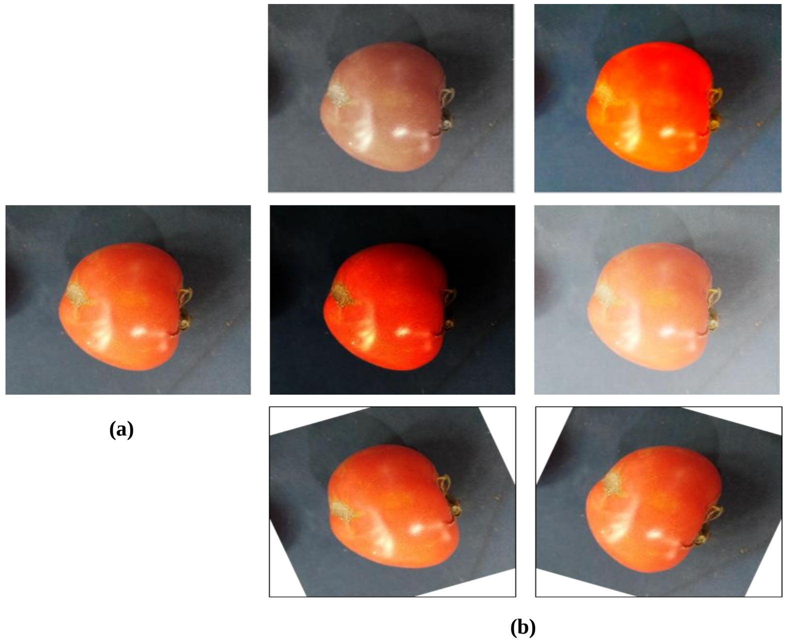

Fig. 5 (a) Original images; (b) Annotation images: change of saturation and light exposure rotary

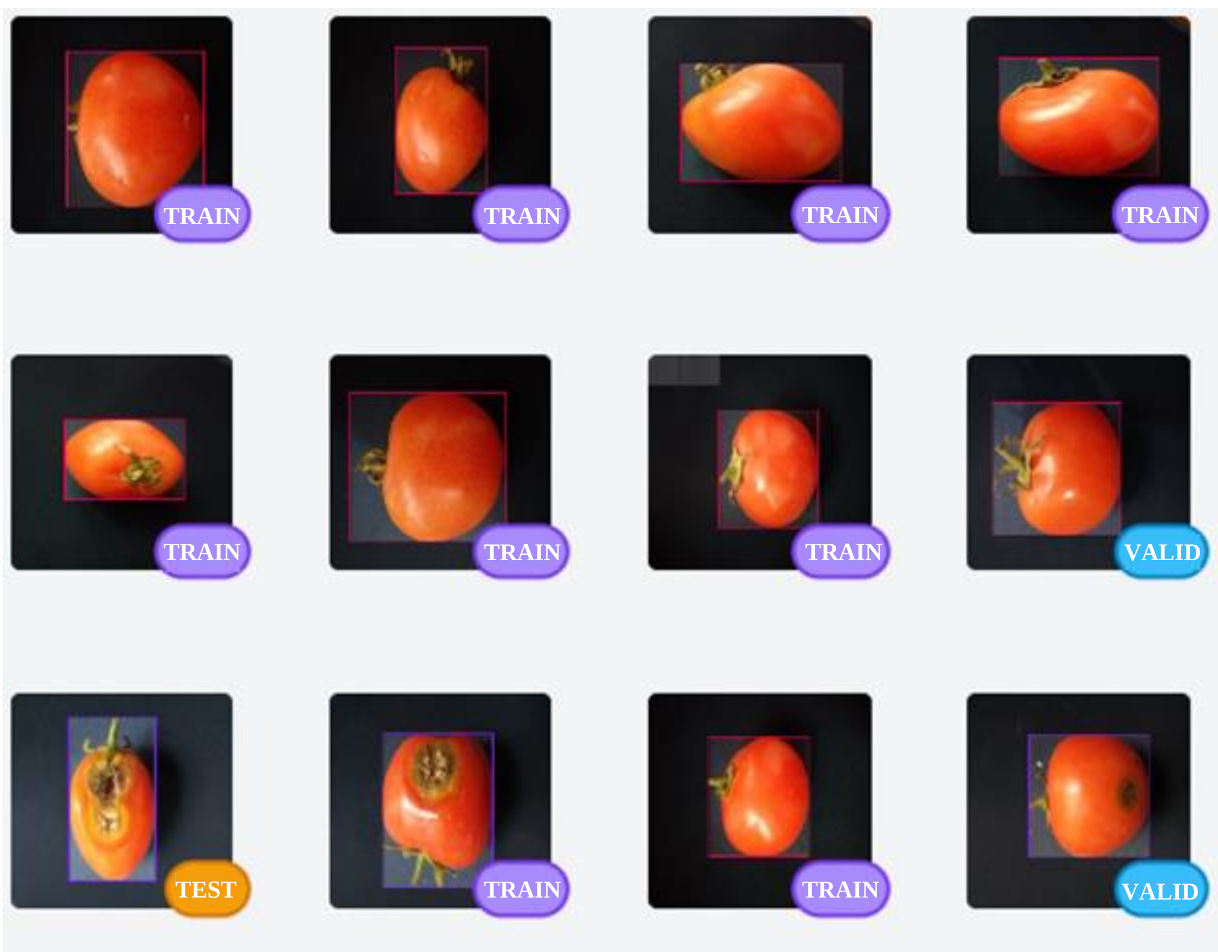

Fig. 6 Dataset with labelling classified into three subsets: TRAIN (70%), VALID (15%), và TEST (15%)

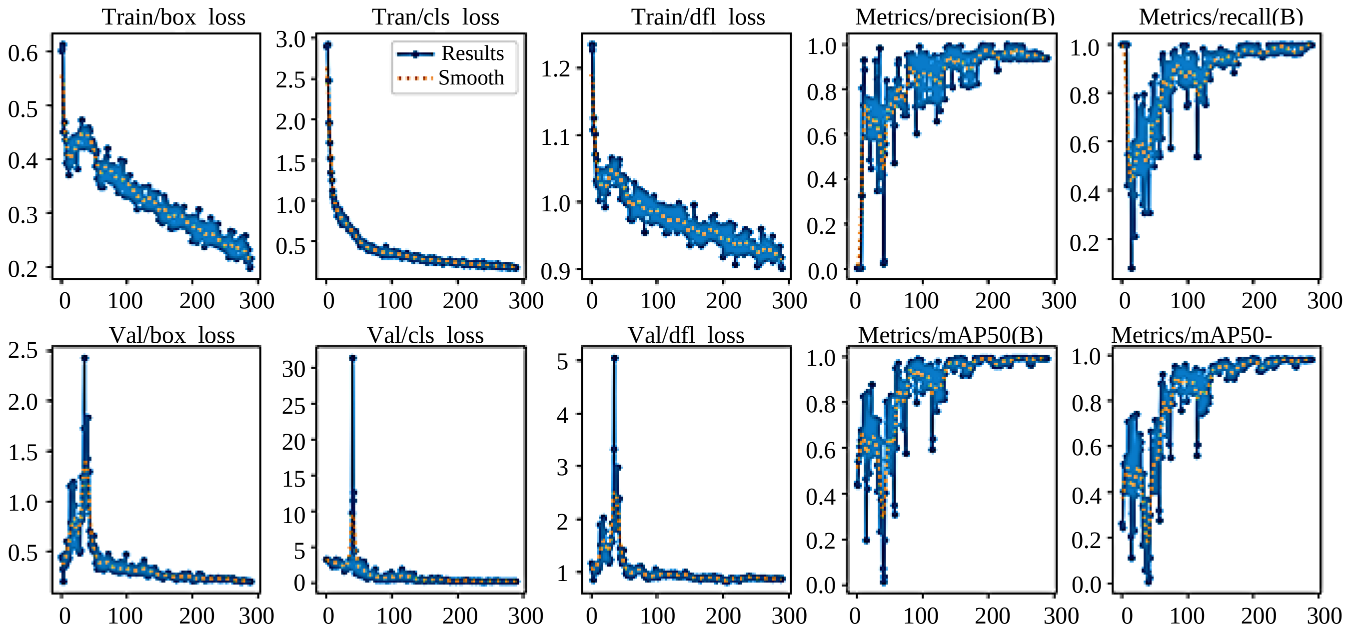

Fig. 7 Training results of YOLOv8 on Google Colab

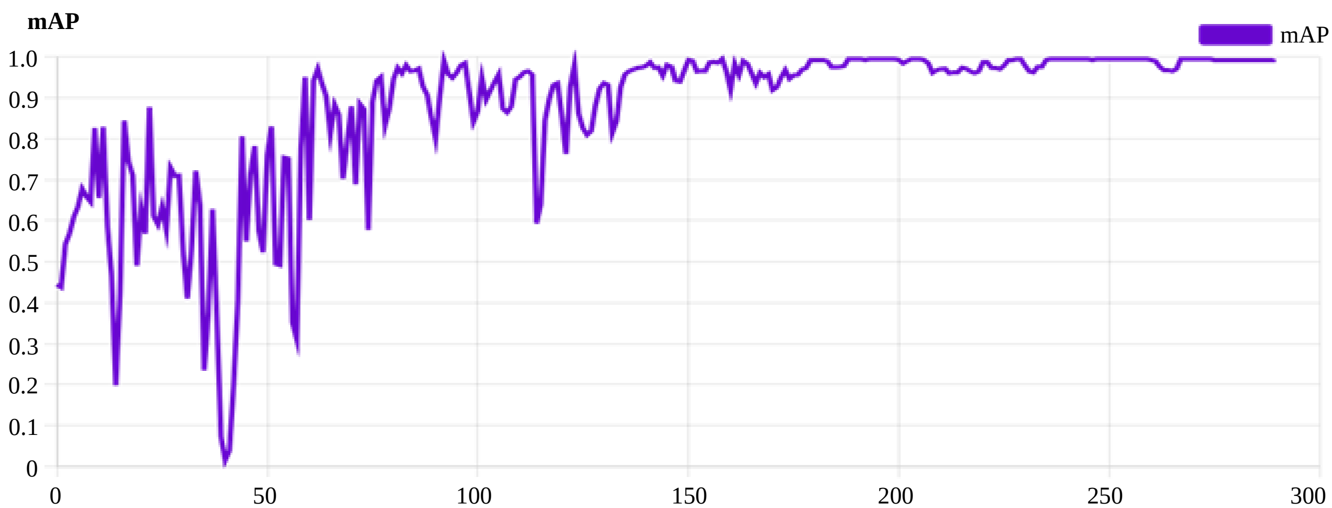

Fig. 8 Testify the efficiency of YOLO - the based model through an mAP diagram

After labelling and enhancing the dataset, the labelled dataset was divided into three subsets: Train, valid, and test, at a ratio of 70:15:15, respectively (see Figure 6). The train set was used to train the model to adjust the weights and parameters of the YOLO network to optimize the object detection performance.

The valid set evaluated the model's performance during training using evaluation parameters such as accuracy, sensitivity (recall), and positive accuracy (precision). After completing the training, the test set was employed to evaluate the model's performance fully. This helps to determine whether the model can correctly and accurately predict new data and allows an overview of the model's ability to detect objects in practice.

2.3.3. Network Training

In this study, the advantage of the Tesla T4 GPU power on Google Colab is taken to train the model. Details of the training results after 289 epochs are shown in Figure 7. As plotted in Figure 8, training the YOLO network on Google Colab yielded impressive results. The model achieved a mAP (mean Average Precision) of up to 99.5%, with a precision of 96.3% and a recall of 96.1%. These results demonstrate the high performance and accuracy of the object detection of the model.

Fig. 9 The practical tomato quality classification model

Fig. 10 Results of the inspection for good tomatoes

Fig. 11 Results of the inspection for rotten tomatoes with several defects

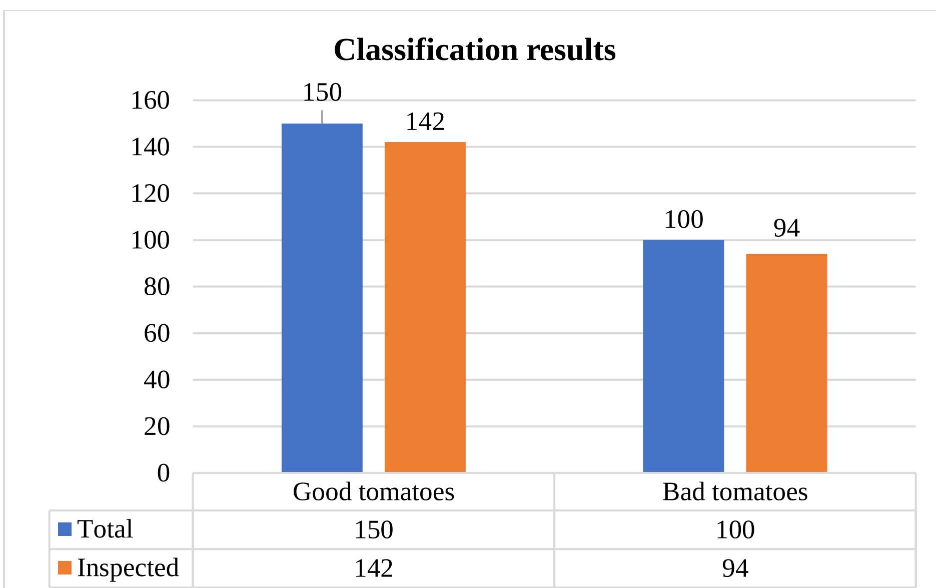

Fig. 12 Results of the classification for both good and bad tomatoes

Table 1. Results of the classification

	Good Tomatoes	**Bad Tomatoes**
Total	150	100
Inspected	142	94
Accuracy	94,67%	94%

3. Results and Discussions

This section presents and analyses some significant experiment results for inspecting tomatoes. As shown in Figure 9, a practical prototype with complete hardware and software can completely classify tomatoes for experiments. This model consists of three grading areas: area 1 (good tomatoes), area 2 (bad tomatoes) and area 3 (tomatoes with failure of classification) (see Figure 9).

Implementing the experiments, several results are illustrated in Figure 10, 11 and 12 and Table 1. Figure 10 shows the classification of good tomatoes with no surface defects. Figure 11 plots four sub-images describing bad tomatoes with defects. These tomatoes need to be removed from a list of exported or supermarket-placed fruits. From Table 1 or Figure 12, the classification percentages are high (more than 94%), demonstrating the promising applicability of the practical inspection system proposed in this study.

4. Conclusion and Future Work

In this study, a CNN-based YOLOv8 deep learning network model was proposed to detect defects on tomato surfaces. A dataset of 500 RGB images with good and bad tomatoes has been used. The training results have shown that the model's Average Precision (mAP) reached 99.5%, showing superior object detection and positioning capabilities. This results from model optimization and advanced techniques during the training process. The precision and coverage (recall) reached 96.3% and 96.1%, respectively. These results demonstrated that the model could accurately identify and cover objects in the images. This achievement contributes to a significant improvement in the model's performance for object detection tasks. The classification experimental results on the real model reached over 94%, indicating that the YOLOv8 algorithm has great potential for agricultural product quality classification applications. Hopefully, the proposed system has made e a significant contribution to the field of agricultural research. Future studies should focus on improving the current results and extending different fruits to be inspected.

Funding Statement

This work was supported by the Project granted number [DTKHCN.18/2022], Electric Power University, Hanoi, Vietnam.

References

[1] [Online]. Available: https://quacert.gov.vn/en/good-agriculture-practice.nd185/vietgap-standard.i88.html

[2] Longsheng Fu et al., "Classification of Kiwifruit Grades Based on Fruit Shape using a Single Camera," *Sensors*, vol. 16, no. 7, pp. 1-14, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[3] Hassan Sadrnia et al., "Classification and Analysis of Fruit Shapes in Long Type Watermelon using Image Processing," *International Journal of Agriculture & Biology*, vol. 9, no. 1, pp. 68-70, 2007. [Google Scholar] [Publisher Link]

[4] Chanki Pandey et al., "Quality Evaluation of Pomegranate Fruit using Image Processing Techniques," *International Conference on Communication and Signal Processing*, India, pp. 38-40, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] D. Surya Prabha, and J. Satheesh Kumar, "Assessment of Banana Fruit Maturity by Image Processing Technique," *Journal of Food Science and Technology*, vol. 52, pp. 1316-1327, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[6] Dameshwari Sahu, and Ravindra Manohar Potdar, "Defect Identification and Maturity Detection of Mango Fruits using Image Analysis," *American Journal of Artificial Intelligence*, vol. 1, no. 1, pp. 5-14, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[7] C. Murugesh, and S. Murugan, "Moth Search Optimizer with Deep Learning Enabled Intrusion Detection System in Wireless Sensor Networks," *SSRG International Journal of Electrical and Electronics Engineerin*g, vol. 10, no. 4, pp. 77-90, 2023. [CrossRef] [Publisher Link]

[8] S. Thirumal, and R. Latha, "Teaching and Learning Based Optimization with Deep Learning Model for Rice Crop Yield Prediction," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 10, no. 4, pp. 105-114, 2023. [CrossRef] [Publisher Link]

[9] Gargi Sharma, and Gourav Shrivastava, "Crop Disease Prediction using Deep Learning Techniques-A Review," *SSRG International Journal of Computer Science and Engineering*, vol. 9, no. 4, pp. 23-28, 2022. [CrossRef] [Publisher Link]

[10] G. P. Dimf, P. Kumar, and K. Paul Joshua, "CNN with BI-LSTM Electricity Theft Detection Based on Modified Cheetah Optimization Algorithm in Deep Learning," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 10, no. 2, pp. 35-43, 2023. [CrossRef] [Publisher Link]

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Image Net Classification with Deep Convolutional Neural Networks," *In Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012. [Google Scholar] [Publisher Link]

[12] Yuzhen Lu, "Food Image Recognition by using Convolutional Neural Networks (CNNs)," *Computer Vision and Pattern Recognition*, pp. 1-6, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[13] Yu-Dong Zhang et al., "Image based Fruit Category Classification by 13-Layer Deep Convolutional Neural Network and Data Augmentation," *Multimedia Tools and Applications*, vol. 78, pp. 3613-3632, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[14] Jan Steinbrener, Konstantin Posch, and Raimund Leitner, “Hyperspectral Fruit and Vegetable Classification using Convolutional Neural Networks,” *Computers and Electronics in Agriculture*, vol. 162, pp. 364-372, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[15] Hafiz Muhammad Rizwan Iqbal, and Ayesha Hakim, “Classification and Grading of Harvested Mangoes using Convolutional Neural Network,” *International Journal of Fruit Science*, vol. 22, no. 1, pp. 95-109, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Laila Marifatul Azizah et al., “Deep Learning Implementation using Convolutional Neural Network in Mangosteen Surface Defect Detection,” *7th IEEE International Conference on Control System, Computing and Engineering*, Malaysia, pp. 242-246, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[17] Peichao Cong et al., Research on Instance Segmentation Algorithm of Greenhouse Sweet Pepper Detection Based on Improved Mask RCNN, *Agronomy*, vol. 13, no. 1, pp. 1-24, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Abdul H. Halimi, and Ashebir H. Tefera, “Application of Cropwat Model for Estimation of Irrigation Scheduling of Tomato in Changing Climate of Eastern Europe: the Case Study of Godollo, Hungary,” *SSRG International Journal of Agriculture & Environmental Science*, vol. 6, no. 1, pp. 1-11, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Longsheng Fu et al., “Kiwifruit Detection in Field Images using Faster R-CNN with ZF Net,” *IFAC-Papers Online*, vol. 51, no. 17, pp. 45-50, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[20] C. Sudha, K. JaganMohan, and M. Arulaalan, “Real Time Riped Fruit Detection using Faster R-CNN Deep Neural Network Models,” *International Conference on Smart Technologies and Systems for Next Generation Computing*, India, pp. 1-4, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Sudianto et al., “Chilli Quality Classification using Deep Learning,” *International Conference on Computer Science and Its Application in Agriculture*, Indonesia, pp. 1-5, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] Abdul Saboor Dawlatzai, R. Jayanthi, and Saidajan Atiq Abdiani, “Efficacy of Graded Doses of Pusa Hydrogel on Growth and Quality of Coleus (Coleus blumeiL.) under Polyhouse Condition,” *SSRG International Journal of Agriculture & Environmental Science*, vol. 4, no. 4, pp. 32-36, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[23] Jia Yao et al., “A Real-Time Detection Algorithm for Kiwifruit Defects Based on YOLOv5,” *Electronics*, vol. 10, no. 14, pp. 1-13, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[24] Kashish Goyal, Parteek Kumar, and Karun Verma, “AI-based Fruit Identification and Quality Detection System,” *Multimedia Tools and Applications*, vol. 82, pp. 24573-24604, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] K. Vasumathi, S. Selvakani, and P. Rajesh, “Deep Learning for Analysing Rice Quality,” *International Journal of Computer and Organization Trends*, vol. 13, no. 1, pp. 16-22, 2023. [CrossRef] [Publisher Link]

[26] Jairo Lucas de Moraes et al., “Yolo-Papaya: A Papaya Fruit Disease Detector and Classifier using CNNs and Convolutional Block Attention Modules,” *Electronics*, vol. 12, no. 10, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[27] J. Terven, and D. Margarita Cordova-Esparza, “A Comprehensive Review of Yolo: from Yolov1 and Beyond,” *Computer Vision and Pattern Recognition*, pp. 1-33, 2023. [CrossRef] [Google Scholar] [Publisher Link]