



Implementation of YOLOv5 for Real-Time Maturity Detection and Identification of Pineapples

Trung Hai Trinh¹, Ha Huy Cuong Nguyen^{2*}

¹ Faculty of Computer Science, Vietnam – Korea University of Information and Communication Technology, The University of Danang, Danang City 550000, Vietnam

² Software Development Centre, The University of Danang, Danang City 550000, Vietnam

Corresponding Author Email: nhhcuong@sdc.udn.vn

<https://doi.org/10.18280/ts.400413>

ABSTRACT

Received: 16 February 2023

Revised: 27 June 2023

Accepted: 11 July 2023

Available online: 31 August 2023

Keywords:

detect fruit ripening, ripeness estimation, segmentation, pineapples classification, YOLOv5

The assurance of fruit freshness during harvest is an enduring challenge that is faced by farmers and suppliers. Traditional methodologies, currently utilized for evaluating fruit freshness, have been characterized by their time-consuming nature, high costs, labor-intensity, and susceptibility to inaccuracies. To address these issues, machine-based detection and sorting systems have been proposed, offering increased efficiency by leveraging technological advancements to analyze fruit attributes such as color, texture, physical appearance, size, and shape. These attributes are critical determinants of fruit quality and value, making accurate fruit evaluation a necessity in the agricultural and food industries. This is particularly true for products such as organic juices and jams, which have significant implications for human health. Providing unfit fruits not only affects the economy adversely but also contributes to augmented carbon dioxide emissions. This work presents a novel technical solution for the detection of pineapple freshness on Vietnamese farms, employing the Fast R-CNN and YOLOv5 techniques. The YOLO model was trained on a diverse dataset, comprised of over a hundred object categories and 50,000 preprocessed images. An aggregated model, combining the outputs of the pre-trained and transfer models, demonstrated improved performance while reducing training time, owing to the extensive training dataset. The classifier displayed an impressive accuracy sensitivity of 94.5% when tested on 50,000 images. Experimental results validate the superior performance of the trained YOLOv5s model, which attains a ripe pineapple recognition accuracy of 98%, outperforming Faster R-CNN by 9.27% and trailing behind YOLOv5x by a mere 0.22%. Additionally, the YOLOv5s model exhibits an impressive detection speed, requiring only 9.2ms to detect a single image—67.88% faster than Faster R-CNN and only 34.06% slower than YOLOv5x. These findings confirm that the YOLOv5s target recognition model meets the requirements for accurate recognition and high-speed processing of ripe pineapples. Its compatibility with agricultural-embedded mobile devices makes it a prime candidate for supporting precision operations in ripe pineapple harvesting machines.

1. INTRODUCTION

Agricultural automation is currently undergoing a global upsurge, owing to its substantial contributions to improving productivity, efficiency, and output in the agricultural sector. This trend transcends geographical boundaries, significantly impacting Vietnamese agriculture, where a fusion of international technology and techniques, augmented by the innovation and adaptability of local farmers, drives progress. The escalating demand for clean and safe food, coupled with the imperative of population health, necessitates an emphasis on agricultural automation.

The advent of intelligent features and technologies, including artificial intelligence (AI), cloud computing, big data, and the Internet of Things (IoT), has instigated a revolution across multiple sectors, including agriculture. These innovations, which leverage advanced analytics and digital technology, offer enhanced processing speeds, adaptability, and economic value. Despite the transformative potential of these technologies, their integration into the

agricultural sector, particularly precision agriculture, has been markedly slow during the ongoing digital revolution.

In the specific context of Da Nang City, food safety and traceability have emerged as prominent concerns for the authorities in charge of healthcare, social security, and food safety. The lack of comprehensive traceability labels and the inconsistency in labeling practices across various production regions in Vietnam underscores the urgent need for improved agricultural automation.

Spurred by these challenges, this study endeavors to bridge the research gap, focusing on the development of a ripe fruit database, with a particular emphasis on the classification of pineapples in Vietnam. The primary objective is to generate a reliable and accessible database for key agricultural products, thereby facilitating effective management and traceability throughout the supply chain. By reducing the training time for classification and disease detection in harvested pineapples, the research aims to streamline agricultural processes, boosting productivity and ensuring the delivery of safe, high-quality produce.

Significant strides have been made in the realm of pineapple classification using object detection techniques. Evidence from this study demonstrates that the amalgamation of a collective method for object detection and transfer learning techniques significantly enhances the efficiency and accuracy of pineapple classification. Comprehensive experiments and analyses have yielded a classifier with a sensitivity accuracy of 94.5% on a test set of 50,000 images. These results underscore the efficacy of the proposed approach in minimizing training time while preserving high recognition accuracy. The YOLOv5s target recognition model, being suitable for deployment on agricultural embedded mobile devices, offers valuable technical support for accurate operations in ripe pineapple harvesting machines.

The structure of this paper is as follows: Section 2 reviews relevant literature in the field, elucidating existing research and identifying knowledge gaps. Section 3 details the process of accumulating data, annotating tools, and conducting literature surveys, thereby delineating the methodology employed in the study. The proposed approach is introduced in Section 4, detailing the techniques used in the development of the ripe fruit database and the implementation of the collective method for object detection. Section 5 presents and analyzes the experimental results, providing insights into the performance and effectiveness of the developed model. Finally, Section 6 concludes the paper by summarizing the findings, discussing future research directions, and acknowledging the study's limitations.

2. LITERATURE REVIEW

The technology of barcodes plays a crucial role in locating fruits in supermarkets and tracking product information such as origin. However, shopkeepers face the challenge of managing barcodes for each fruit category. Machine learning algorithms, particularly in the field of object detection and recognition, have received significant attention [1]. In supermarkets and fruit shops, fruits and vegetables are typically packed in small boxes and priced using barcodes. Despite the availability of prepackaged fruit, many customers prefer to handpick their fruits. To address this, a fruit recognition system that combines a joint scale with an imaging dataset can be employed.

Visual image recognition holds immense importance in control systems, information processing systems, and automated decision-making systems. Convolutional neural networks (CNNs) are widely used to develop control systems for object recognition. The researches utilized CNNs and parameter optimization to accomplish fruit detection and recognition tasks [2]. CNNs are capable of recognizing images and classifying objects by processing and classifying the received image feeds. Deep learning, particularly using CNNs, is a powerful technique for image processing.

In the realm of fruit recognition, various studies have explored methods based on deep learning. The researches employed computer vision to rapidly, automatically, and accurately detect and register apples from sequential input images, even under artificial lighting conditions [3]. Previous studies have also addressed the challenges and difficulties associated with fruit recognition systems [4, 5]. Transfer learning has been found effective in reducing parameters and

computation costs for fruit recognition [6, 7]. The similarity between the target dataset and the training dataset in terms of size and breadth is worth mentioning. Notably, studies have presented different methods, computation times, and performance outcomes [8, 9].

The Faster R-CNN model, which combines a region proposal network (RPN) with a CNN model, has shown superior performance compared to its predecessor, Fast R-CNN. However, identifying ripening phases in various fruits such as mangoes, almonds, star apples, and apples remains a challenge. The researchers developed the MangoYolo architecture based on YOLOv5, achieving an impressive 0.983% accuracy in locating mangoes in orchards [8].

In the domain of fruit ripeness assessment, which is a significant research area in agricultural harvesting, deep learning has received considerable attention. The researches proposed a CNN-based technique to recognize different ripe pineapple types [10, 11]. They built a mask R-CNN model to identify the ripeness of pineapple fruits in a greenhouse setting. Liu developed an enhanced YOLO-v4 identification model based on YOLOv3, specifically for accurately detecting yellow pineapples under slightly obstructed conditions. Chung and colleagues suggested an approach that combines YOLOv4 and HSV to identify mature pineapples in their natural environments [12]. Deep learning techniques eliminate the need for laborious manual feature extraction and offer greater precision and reliability compared to conventional methods. However, achieving both model recognition speed and size is challenging for current deep-learning methods used in greenhouse-picking devices. The YOLOv5 series target detection algorithm, a representative of the "one-stage" model of the target detection algorithm, can enhance pineapple maturity recognition accuracy, stability, and practicality, particularly when dealing with pineapples in different environments and overlapping conditions. It is crucial to differentiate pineapples based on their maturity stages, but further work is required to address these challenges.

Pineapple fruit recognition in greenhouse picking devices can benefit from the YOLOv5s model, which offers rapid running speed, high recognition accuracy, tiny model size, and data augmentation to increase target detection accuracy and precision. The model's real-time performance and accuracy are confirmed by comparing it with other target detection models, including Faster R-CNN, YOLOv5m, YOLOv5l, and YOLOv5x. Additionally, the combination of computer vision and CNNs has been proposed in a collaborative research study [13, 14]. They utilized TensorFlow to implement their mathematical model.

The integration of new technologies in agriculture enhances production, efficiency, and product quality, leading to long-term sustainability. The identification and classification of crops based on their similarities are common approaches for locating them. For instance, a study used photographs of tomatoes and other plants on dark backdrops to analyze variance in fruit shape and extract information on the sizes of cut tomatoes. Vision-based algorithms were employed to detect and localize fruit pickers, enabling the identification and categorization of post-harvest fruits using multilayer neural networks [15]. CNN recognition models have been widely utilized in numerous studies, and ongoing efforts continue to improve these models for digital image identification and detection [16, 17].

3. METHODOLOGY

We use open-source image net data and images of pineapple ripening stages and searched for pineapple images from many agricultural farms in Quang Nam - Da Nang province Vietnam, providing us with delicate data sets of Pineapple images. The images total led over 50,000. See some sample images from the cumulative dataset as shown in Figure 1.



Figure 1. Sample images from the dataset

Annotating YOLOv5 tool was used to label 50000 images. A technique for generating additional training data from collecting picture data has been used to this dataset as a means of identifying and avoiding overfitting in classification. During the course of the test scenario, a number of other picture modifications, such as zooming, lighting, flipping, rotating, and warping, were used. According to Figures 2, 3, and 4, the suggested model consists of two model detections and choices. Additionally, the pre-trained model, the transferred-model, and the ensemble model are all included. The model produced four randomly generated pictures that were created artificially based on one genuine image.

Using the features extracted from the input image, ConvNet produces feature maps as output. A 16×16 filter and three convolutional layers are proposed. Convolutional kernels of 9×9 and 25×25 outperformed each other. We developed feature maps based on these evaluations. A single layer of the feature maps is shown in Figure 2. This figure illustrates the architecture or structure of the proposed model. It showcases the two main components: the pre-trained model and the detection and decision modules. The pre-trained model refers to a pre-existing model that has been trained on a large dataset for general object recognition tasks. It serves as a foundational model that has learned a wide range of features and patterns from the training data. The detection and decision modules, which are not explicitly depicted in the figure, are responsible for further processing the outputs of the pre-trained model and making decisions based on the detected objects.

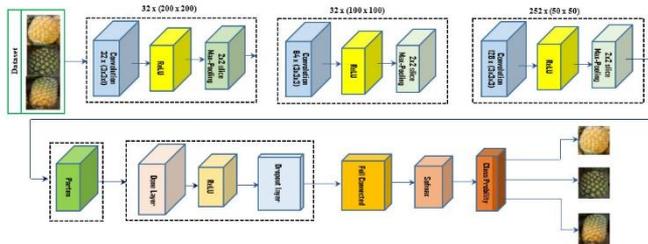


Figure 2. Framework architecture of pineapples classification

As part of our proposal for deep learning ReLU is used as an activation function and Softmax function to augment the hidden layers, increasing the training data in the CNN network with the YOLOv5 model. In addition to introducing nonlinearity, ReLU as an activation allows models to learn faster and perform better [18], shown in Figure 3. This figure

likely presents the process flow or workflow of the proposed model. It demonstrates the sequence of steps involved in the fruit recognition system. The details of the individual steps and their connections should be provided in the accompanying text or figure legend. Please refer to the specific description in the text or legend to gain a better understanding of the processes depicted in Figure 3.

In the proposed network, a convolutional module generates feature vectors with 128 dimensions. Convolutional layers connect the layers in the proposed network. On the fully connected and Softmax layers of the network output, linear classification is built. Dropout layers are employed at a rate of 0.3 in our proposed method. Overfitting results when there are too many data points matched to a small function, which we avoid by using dropout layers, shown in Figure 4(a), (b).

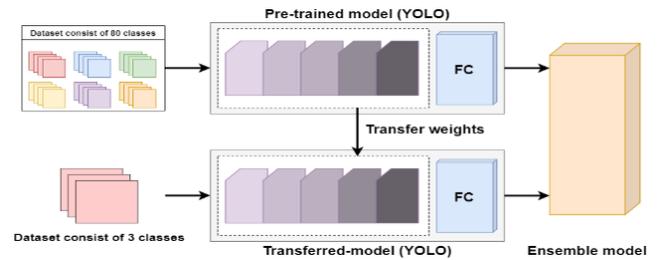


Figure 3. Proposed model for incremental object detection with YOLO

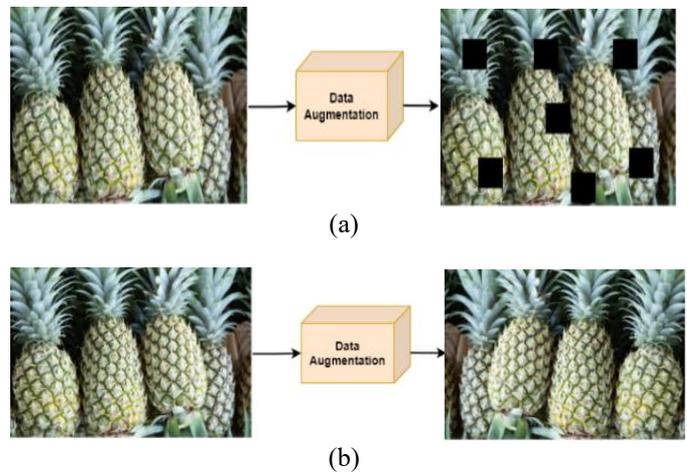


Figure 4. Augmentation and transformations for Pineapple fruits images

Figure 4 illustrates the process of augmentation and transformations applied to the images of pineapple fruits. Augmentation techniques are commonly used in deep learning to increase the size and diversity of the training dataset, which can help improve the model's performance and generalization capabilities. Transformations refer to various operations applied to the images to create variations in their appearance.

The specific augmentation and transformation techniques used in Figure 4 might include:

Rotation: The images of pineapple fruits are rotated by a certain angle, which introduces variability in the orientation of the fruits.

Scaling: The images are resized to different scales, either larger or smaller, to simulate variations in the size of the fruits.

Flipping: The images are horizontally flipped, creating mirror images, which can help the model learn from different perspectives.

Brightness/Contrast adjustment: The brightness and contrast levels of the images are modified, creating different lighting conditions that the model should be able to handle.

Noise addition: Random noise is added to the images, mimicking real-world imperfections or variations in image quality.

These augmentation and transformation techniques aim to enhance the robustness and generalization of the model by exposing it to a wider range of variations in the training data. By applying these operations to the pineapple fruit images, the model becomes more capable of recognizing and accurately classifying pineapple fruits under different conditions.

For further details on the specific augmentation and transformation techniques employed in Figure 4, please refer to the corresponding explanation provided in the text or figure caption.

4. PROPOSED METHOD

The objective of this paper is to provide several technological techniques for the identification of objects of pineapples using the YOLOv5 toolkit [19]. We proposed method implement for object detect and classification the pineapple. The first detect for a pineapple, then classification of unripe pineapple, Semi-Ripe Pineapple and Ripe Pineapple with YOLOv5 algorithm to identify pineapples for good predictive result [18], shown in Figure 5(a), (b), (c).

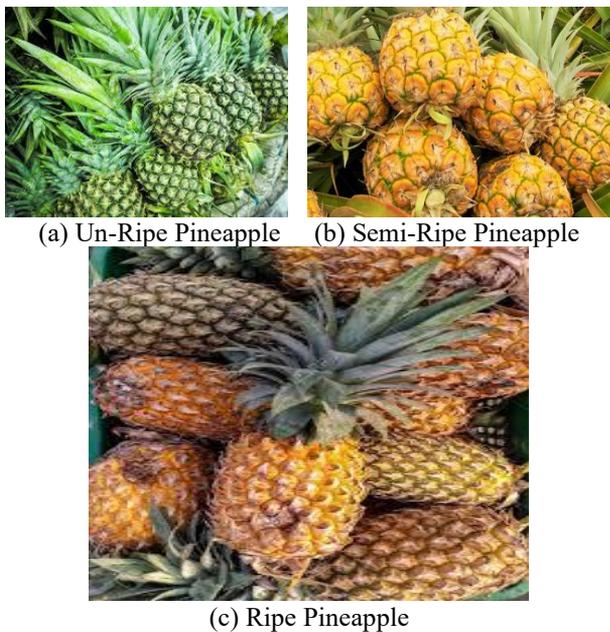


Figure 5. Fruits images: Un-Ripe Pineapple, Semi-Ripe Pineapple, and Ripe pineapple

As can be seen in Figure 6, the YOLOv5 network design makes use of not one but two different CSP structures. CSP1 and CSP2 respectively. The CSP1 structure is used by backbone networks, while the CSP2 structure is utilised by neck networks. CSP structural design creates a split in the feature maps into two distinct portions. They are integrated with the feature maps that were previously convolutionally worked on in the first portion in order to gain additional in-depth information about the characteristics being sought for. On the other hand, a portion of them is used to carry on the operation of convolution in order to get more in-depth

information about the characteristics. In addition to enhancing the capabilities of the network to learn new information, cross-stage designing can also cut down on the number of parameters used by the network. This allows the network to keep its high level of accuracy even as the number of parameters is decreased, and it can also boost inference speed while simultaneously cutting down on the amount of memory needed [20]. In the article, research based on deep convolutional neural network technique, the researcher has developed and proposed an algorithm to classify and detect fruits, in an experimental case study deployed on the basis of real-time data. ripening period of pineapple. This study applied different types of data expansion techniques to other fruits. The study has improved by reducing the convolutional layers in the neural network, and further improving the performance in each convolutional layer. Research has developed using three convolutional layers and two connected layers. You don't need to do any preprocessing in order to discover the most useful features from huge picture collections. The proposed network has undergone rigorous testing in a variety of actual-world use cases. The results of the experiments indicate that it has a high level of accuracy. The data is divided using the Focus module, which is located at the first layer of the backbone network. In the YOLOv5s algorithm, a regular image with dimensions of 3'608'608 is imported into the network, converted to a feature map with dimensions of 12'304'304 following the clipping operation, and then converted to a feature map with dimensions of 32'304'304 following convolution with 32'304'304 worth of convolution kernels (Figure 7). The fundamental objective of this algorithm is to get a higher level of speed [21].

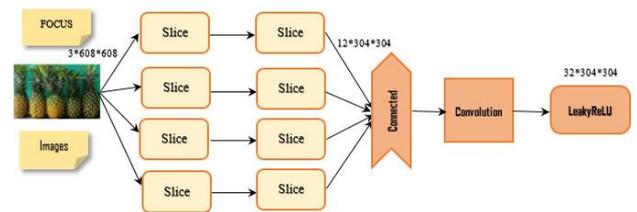


Figure 6. Diagram of the focus module

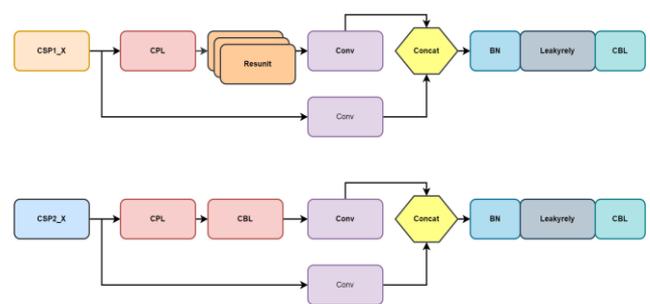


Figure 7. Schematic diagram of the CSP module

4.1 YOLOv5 improved

Select and get rid of each individual thing that has been tagged. In the event that many limit boxes identify the same item, the solution will choose the limit box that is most suited to the situation. As well as the bounding boxes for each layer, we calculate the probability of finding different types of objects. New versions of YOLOv5 have been announced by the authors. In contrast to YOLO algorithms, Faster R-CNN

uses a different detection method. As part of YOLOv5, regression techniques, bounding box processing, and regression are used to determine class probability. We increased the computation speed using this innovative solution, resulting in very high-test results [22, 23].

Loss function: This function is used binary cross entropy with log loss A matrix grid (S×S) is used in YOLOv5 to collect data. Centres of objects are displayed by a grid matrix and automatically detected. A confidence score of 0 applies to subjects with no scores. By calculating row and column values, the selected objects will be deleted in YOLOv5.

Backbone: In order to extract features from pictures that are made up of cross-stage partial networks, it uses CSPDarknet as the framework. Including:

YOLOv5 Neck: It employs PANet to construct a feature pyramid network in order to do aggregation on the features, and then it sends that information to the Head so that it may be predicted.

Layers in YOLOv5 Head that are responsible for object identification and provide predictions based on the anchor boxes.

In addition to this, YOLOv5 employs the following options for training—both Activation and Optimisation: Leaky ReLU and sigmoid activation are used in YOLOv5, and the optimizer choices SGD and ADAM are available.

4.2 Data set description

When the skin colour changed from a dull green to a glossy yellow, the model had a chance of 0.9868 of detecting an item that was a ripe pineapple. The subsequent findings were tested, and the process was maintained after that, with successful outcomes. The ultimate option made by the business model need to be the one that corresponds to the category with the greatest likelihood, as shown in (Figure 8).

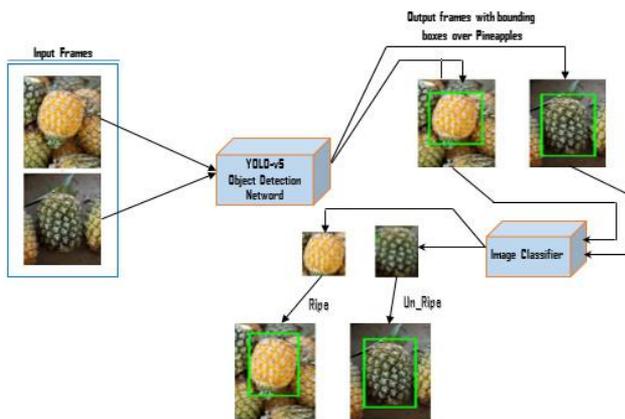


Figure 8. Pineapple regions detected by YOLOv5 (R: Ripe UR: Unripe)

It is extremely difficult to detect ripe Pineapple in real time. Therefore, technical solutions are presented in this paper in five stages. By using toolkit YOLOv5 improved, you can process input data, sort, and train data sets. After conducting a review of the test scenarios in the lab. Research and pilot implementation at pineapple farms in the central region of Vietnam, the input data set is images of pineapple growing areas, images taken with mobile devices and images from Camera pineapples from mature farms, information tested via

App/Web installed on Android Smart Phone platform. With the improved YOLOv5, we recommend. The experiment has the ability to differentiate between ripe and unripe pineapples thanks to the taught grading frame. With enough experience, it may be simplified into actionable steps that provide farmers with the information they need to make educated choices as shown in (Figure 9).

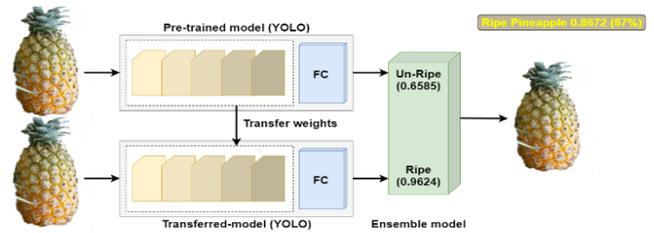


Figure 9. Detected by YOLOv5

5. RESULTS AND DISCUSSION

5.1 Data set description

We analyzed 50,000 images of pineapple fruits from a database. We captured each image using a 200×200 pixel HD Logitech webcam. Besides light, darkness, sunlight, pose variation, changes in lighting, and camera artifacts, this database collection presented a variety of challenges. A split-and-merge operation is used to separate images from their backgrounds. Our model must be able to cope with lighting variations, artifacts captured, specular reflection shading, and shadows in real-world scenarios in supermarkets and fruit shops. We tested all our models for robustness and they all performed well. Each channel has 8 bits and is stored in RGB (Red, Green, Blue) color space. On the same day, we took different times for the same category of images.

This results in the dataset having a greater degree of uncertainty, and the scenario having a greater degree of realism. The photographs all had varying degrees of quality and brightness. We did our data collection on fruits in an environment that imposed little restrictions. There are photographs that were obtained by repositioning the weight machine so that it was in close proximity to the windows and then drawing and shutting the drapes in order to snap the photographs. The following is the disarray grid for the prepared benchmark model. A disarray grid assesses the presentation of the prepared model on an informational index with realized objective classes. The model was evaluated by figuring out the accuracy and review, using True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) in the chaotic framework lattice.

Values that are predictive as determined by the formula $Precision = TP / TP + FP$. Recall's formula for sensitivity is as follows: TP divided by TP plus FN.

When analyzing the outcomes of the test cases, we get predictive value findings, and the sensitivity comes out to be 0.967. The classes that were processed during the base model training are shown in Figure 10 and Figure 11.

To carry out an analysis and settle on a decision on whether or not to confirm ripe pineapple with unripe fruit. In the research study, the refining and releasing approaches were applied to all of the layers of the input picture data dataset that was trained using an upgraded version of the YOLOv5 model. In order to keep the computation outputs consistent all the way

through the training stages, the set of typical classes of the trained base model was used. We carried out tests in order to extract part of the chosen data as a subset in comparison to the classes into which it was first classified. The results of the experiment have shown that the learning rate is 0.00005 for the preliminary courses and 0.00046 for the culminating grades. The accuracy that may be reached after just 20 minutes of training is 94.2 percent.

The outcomes obtained from the collection and estimate based on the likelihood of precise assessment and the computation results from the dependent on confusion framework for the further evolved model are 0.93 and 1.0, respectively. Both sets of outcomes pertain to the further developed model. In the comprehensive view of the specific circumstances, the likelihood of preparing for disaster is really low, however the likelihood of successfully confronting the aftereffects of widespread approbation is exceptionally high.

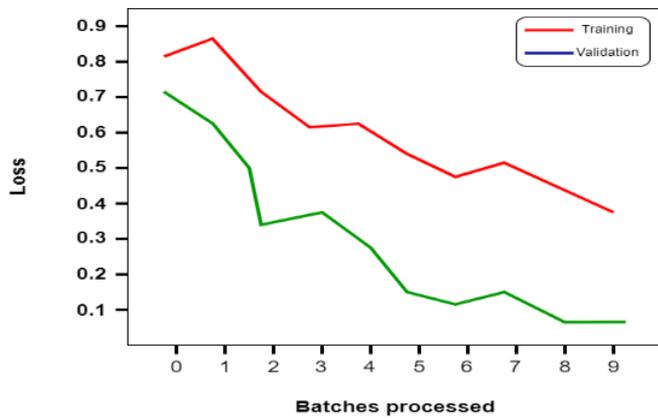


Figure 10. Resulted training vs validation phases (baseline model)

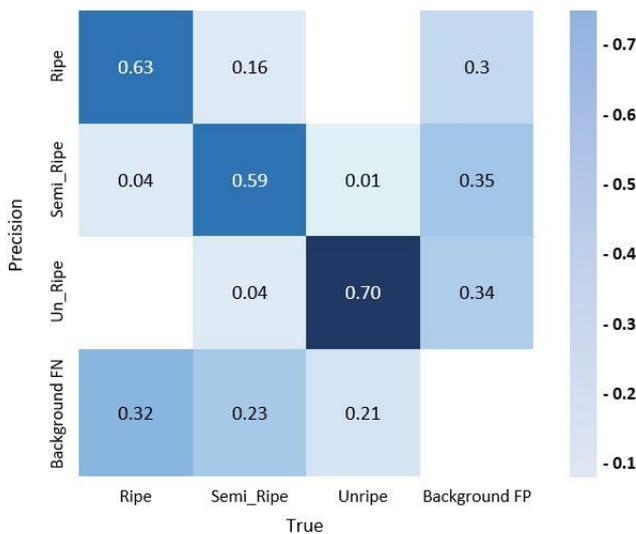


Figure 11. Confusion matrix with ripe and unripe accuracy (fine-tuned model)

5.2 Comparative analysis of other state-of-the-art methods

In Figure 12, Figure 13 and Figure 14, semi - pineapple and unripe pineapple, found from further developed YOLOv5, has been improved to give prepared picture classifiers. These outcomes permit ranchers to cockroach ready or unripe organic products. The consequences of this exploration are

extremely useful and have carried a lot of advantages to ranchers when leading the stock to stores and products to different nations throughout the planet.



Figure 12. The recognition affects the comparison of models detecting objects with ripe pineapples

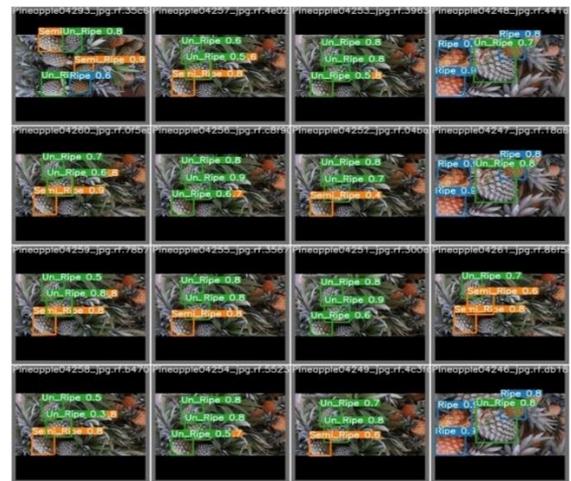


Figure 13. The recognition affects the comparison of models detecting objects with Semi-ripe pineapples



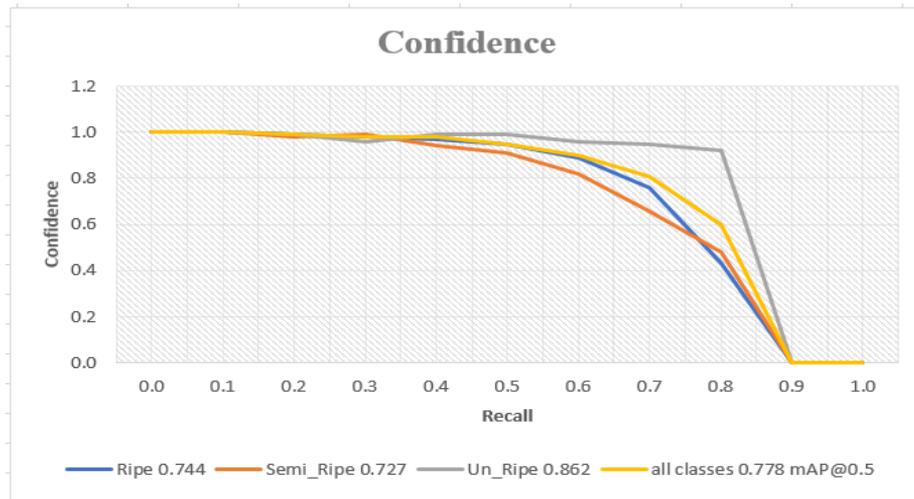
Figure 14. The recognition models with unripe pineapples

5.3 Simulation results

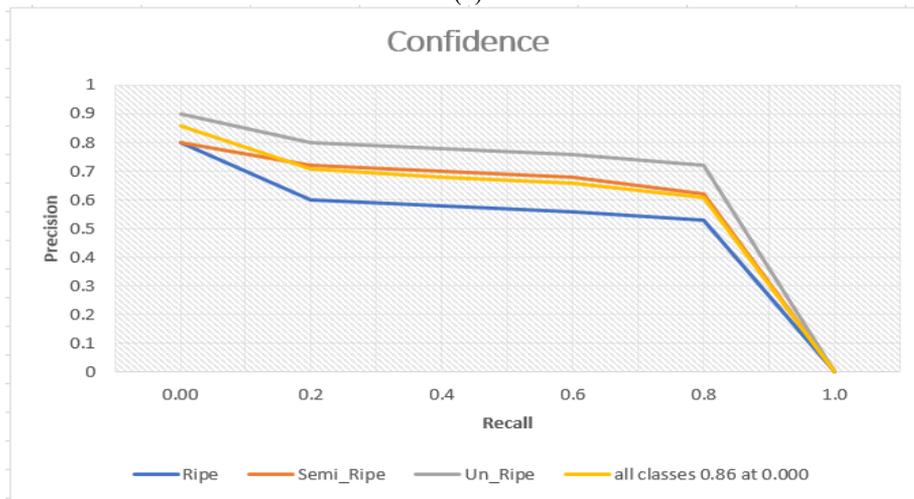
It may be difficult to ascertain the number of layers, the different kinds of layers, and the number of neurons that are contained inside each layer of the brain. In this part, we investigate various different network topologies to determine which one is the most suitable for educational purposes.

During the training of the network, a total of twenty epochs were used. The training loss curves flatten out and become less steep as the number of training epochs rises. In overfitting strategies, dropout layers, augmented data, generalized architectures, and adjusting hyper parameters have all proven to be advantageous. The first test, based on Figures 15 and 16, was 96% accurate with 20 epochs, indicating that overfitting was eliminated.

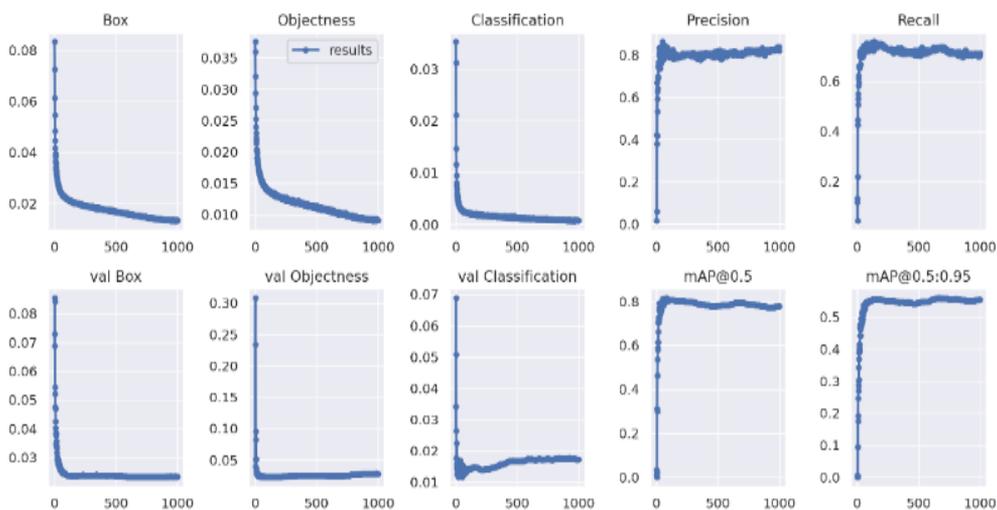
We investigated a number of different situations, including as lightning and stances, to establish how effectively the suggested network would function under various conditions as shown in Figure 17. When dealing with real-time object identification, selecting a suitable architecture for a faster R-CNN might be difficult.



(a)



(b)



(c)

Figure 15. The models detecting objects with unripe pineapples

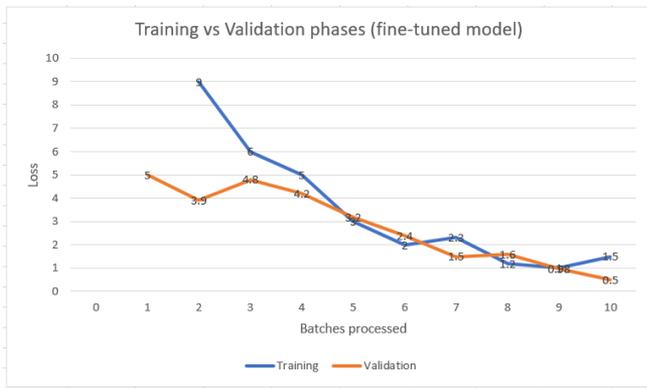


Figure 16. Training vs Validation phases (fine-tuned model)

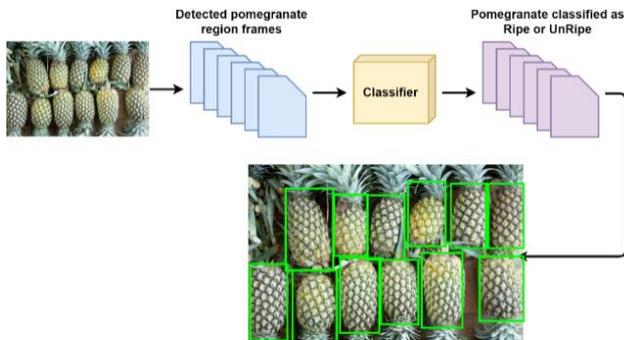


Figure 17. Results of the proposed model



Figure 18. Testing loss curves in a pineapple scenario for training

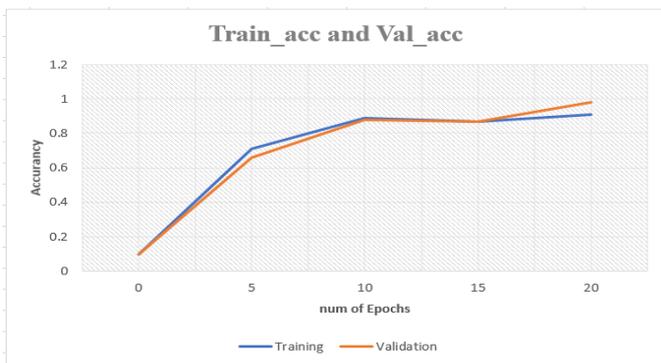


Figure 19. Test scenario with pineapple: Valid training and accuracy curves

In Figure 18, the training and validation loss function curves show that as the quantity of repetitions increases, the loss value of the network gradually decreases and stabilizes. Model loss

values decrease rapidly. In the first 15 iterations and taper off in the last 15 iterations. The model reaches optimality after 20 iterations when the loss value is stable near 0.15.

In Figure 19, the training and validation accuracy function curves show that as the quantity of repetitions increases, the accuracy value of the network gradually decreases and stabilizes. Model loss values decrease rapidly. In the first 15 iterations and taper off in the last 20 iterations. The model reaches optimality after 20 iterations when the accuracy value is stable near 0.8.

In the course of this research, a mod base for the identification of ripe pineapples based on YOLOv5s was suggested. When compared to other research projects of a similar kind that are currently being conducted on the identification of ripe pineapples, the model that has been provided has a number of benefits, including a quicker recognition speed and a higher mAP value. When the model size is reduced by almost around 80%, the missed identification rate of tomato fruits under occlusion and small target conditions is still less than 2%. This is the case even though the model size has been reduced. Therefore, based on the findings of the comparison, the conclusion that can be drawn, which is displayed as in (Table 1), is that the YOLOv5s has considerable advantages for the deployment of embedded devices for the harvesting of pineapples in an orchard. Table 2 demonstrates that once the models were subjected to rigorous testing, the four network architectures of YOLO-v4 improved, and YOLOv5 developed experimentally using pineapples obtained great performance. Every one of their accuracies is more than 95%, every one of their recalls is close to 90%, and every one of their mean average accuracies is greater than 96%. Even though it has the lowest mAP, faster RCNN does not have the best detection effect. The detection performance of YOLOv3 is noticeably superior than that of Faster RCNN, YOLOv4 improved, and YOLOv5 enhanced, while it is only marginally inferior to that of YOLOv6 improved. In terms of accuracy, recall, and average accuracy, the results of YOLOv5 improved are, respectively, 0.11%, 0.33%, and 0.09% lower than the results of YOLOv5 improved. These differences may be seen when comparing the two versions of the test. best detection. When compared to YOLOv5 improved, the model size of YOLOv5 improved has only an 8.88% and 34.06% improvement in each image's detection rate, respectively. According to the findings of the comparison, YOLOv5 enhanced is capable of achieving greater identification accuracy, quicker detection speed, and a smaller model size than the other five models. The weight and file size of the existing deep learning model seem to be excessive, and the real-time detection rate appears to be sluggish. The recognition accuracy does not seem to be very great when dealing with little targets or a congested pineapple. To put this into perspective, the improved YOLOv4 better maturity identification model that was developed in this research and is based on the improved YOLOv5 is able to differentiate between four distinct stages of pineapple ripeness. In this experiment, in contrast to other research of a similar kind, the identification of ripe orchards of pineapples was separated into four categories of maturity. This was done to facilitate the harvesting of pineapples meeting varying criteria for levels of ripeness during real production. When compared to the recognition models used in the laboratory, the model that was suggested for this investigation had the qualities of having a higher recognition speed, a smaller model size, and comparatively less requirements for the setup of the hardware.

The suggested model, when compared to other recognition models in the greenhouse, has the features of greater recognition accuracy and lower model size, with the proposed model having around an 80% reduction in model size when

compared to other models in the greenhouse. The effectiveness of the suggested model's pineapple detection on both congested areas and tiny targets is superior to that of current methods in this regard.

Table 1. Ripe pineapple recognition results comparison of some models

Models	Precision P/%	Recall R/%	Mean Average Precision/%	Test Time/ms	Memory Size/M
YOLOv5s	96.40	91.45	96.8	10	30MB
YOLOv5m	97.34	92	97.9	14	80MB
YOLOv5l	98.9	93.76	98.2	28.8	146MB
YOLOv5x	99.68	95.67	98.8	40.28	289MB
Faster R-CNN	88.7	81.6	83.45	16.8	150MB

Table 2. The proposed method is compared with existing methods

Source	Method	Environment	Maturity Level	mAP	Speed
Chung and Van Tai [12]	HSV	Laboratory	Ripe pineapples	90.70%	68.7 ms
Girshick [3]	SR-SVM	Laboratory	Ripe pineapples	98.30%	72.5 ms
Bargoti and Underwood [9]	Improved Mask R-CNN	Greenhouse	Ripe pineapples	95.45%	65.8 ms
Liu et al. [23]	YOLOv3	Greenhouse	Ripe pineapples	96.40%	54 ms
Gai et al. [22]	Tiny-YOLOv3	Greenhouse	Ripe pineapples	87.70%	227.1 ms
Cuong et al. [14]	Tiny YOLO-v4 Improved	Orchard	Ripe pineapples	98.26%	22.18 ms
Proposed	YOLOv5 improved	Orchard	Ripe pineapples	98.76%	9.2 ms

Following the completion of the comparison study that is provided in Table 1, the YOLOv5s-pineapples model has gained benefits for embedded deployment as well as for the mobile usage of pineapple picking in the frame.

In addition to presenting the simulation results, it is crucial to discuss the managerial implications derived from our findings. The fruit recognition system based on deep learning techniques has several practical implications for different stakeholders in the fruit industry.

Inventory and supply chain management: The accurate and automated fruit detection and tracking provided by the system can significantly improve inventory management practices. Supermarkets and fruit shops can utilize the system to optimize their stock levels, reduce waste, and streamline the supply chain. By having real-time information on fruit availability and quality, managers can make data-driven decisions to ensure sufficient stock and minimize losses.

Customer satisfaction and experience: The fruit recognition system can enhance the overall shopping experience for customers. By enabling quick and precise identification of fruits, customers can conveniently choose their preferred items. This reduces the time spent searching for specific fruits and promotes customer satisfaction. Additionally, the system can assist in providing accurate information about fruit characteristics, helping customers make informed purchasing decisions.

Quality control and grading: The system's ability to classify fruits based on various attributes such as ripeness, size, and color offers significant benefits for quality control and grading purposes. It enables fruit producers and distributors to implement consistent quality standards and sort fruits into different grades. This ensures that customers receive fruits of desired quality and helps maintain product reputation.

Product traceability and transparency: By utilizing barcode technology and tracking information, the fruit recognition system enhances traceability and transparency in the supply chain. Retailers and consumers can access detailed information about the origin, production methods, and

transportation history of the fruits. This promotes trust, supports sustainable sourcing practices, and allows customers to make informed choices based on their preferences and values.

Operational efficiency and cost savings: Implementing a fruit recognition system can lead to improved operational efficiency and cost savings. By automating the fruit detection process, manual labor requirements can be reduced, freeing up resources for other value-added tasks. Moreover, the system's ability to accurately track fruits can minimize losses due to spoilage, optimize storage conditions, and reduce unnecessary handling and transportation costs.

These managerial implications demonstrate how the fruit recognition system based on deep learning techniques can bring practical benefits to the fruit industry. By addressing these implications, stakeholders can make informed decisions, enhance customer satisfaction, improve operational efficiency, and promote sustainable practices throughout the supply chain.

6. CONCLUSIONS

In conclusion, this study focused on developing specialized solutions for recognizing and classifying pineapples on agricultural farms in Vietnam. The main objectives were to identify semi-ripe and unripe pineapples, as well as to assess the quality of pineapples based on their condition, whether good or rotten. The motivation behind this research was to introduce a new algorithm capable of recognizing the nine core stages of pineapple development.

The proposed approach successfully achieved its goals by reducing the computational cost associated with heavy training. Instead of using the YOLOv5 detection model for simultaneous pineapple detection and classification, our model focused on streamlining model training and increasing the number of detectable items. The key components of the research strategy are as follows:

Pre-trained Model: The main sub-model in the analysis is a pre-trained model constructed using 80 classes and a total of 30,000 ripe pineapple images from the dataset as shown in (Figure 17).

Transfer Learning: The transfer model utilizes transfer learning techniques to train on new images, including 10,000 images of semi-ripe pineapples and 10,000 images of unripe pineapples. The training process is efficient, taking only a few hours.

Decision Model: The complete model, created using the transfer learning approach, serves as the foundation for the decision-making process. The suggested model achieves an impressive accuracy of approximately 94.2% and can easily adapt to new classification tasks, such as identifying the condition of a sick patient pineapple.

The classification accuracy was further improved by incorporating various image enhancement methods and applying filters, leading to overall enhanced validation accuracy. However, it is worth noting that there is a lack of reliable and generative systems for automated sorting, counting, detecting rotten fruits, and grading across different types of fruits. This limitation arises from the fact that existing studies often rely on datasets containing only one fruit of a given kind in the photos.

In the future, an improved deep learning system could be developed to predict the remaining shelf life of pineapples, enabling timely sales before spoilage. Such a system would benefit stakeholders involved in the fruit selling sector, consumers, and the overall economy of the nation.

In summary, this research has made significant strides in developing a specialized pineapple recognition and classification system. The findings contribute to the advancement of autonomous pineapple grading and quality prediction, particularly in the smart agricultural business sector. Further research efforts should focus on constructing a faster R-CNN model and expanding the capabilities of the system to encompass a broader range of fruits. With continuous improvement and innovation in deep learning techniques, we anticipate a promising future for the fruit industry, marked by enhanced productivity, reduced waste, and improved profitability.

ACKNOWLEDGMENT

Training for supporting this research project as part of the Ministerial Program of Science and Technology CTB.2021.DNA (Grant No.: B2021.DNA.09).

REFERENCES

[1] Khaing, Z.M., Naung, Y., Htut, P.H. (2018). Development of control system for fruit classification based on convolutional neural network. 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIconrus), Moscow and St. Petersburg, Russia. <https://doi.org/10.1109/EICONRUS.2018.8317456>

[2] Nguyen, H.H.C., Luong, A.T., Trinh, T.H., Ho, P.H., Meesad, P., Nguyen, T.T. (2021). Intelligent fruit recognition system using deep learning. In: Meesad, P., Sodsee, D.S., Jitsakul, W., Tangwannawit, S. (eds) Recent Advances in Information and Communication

Technology 2021. IC2IT 2021. Lecture Notes in Networks and Systems, vol 251. Springer, Cham. https://doi.org/10.1007/978-3-030-79757-7_2

[3] Girshick, R. (2015). Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, pp. 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>

[4] Ren, S.Q., He, K.M., Girshick, R., Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. arXiv:1506.01497V5. <https://doi.org/10.48550/arXiv.1506.01497>

[5] Kim, J.H., Kim, N., Park, Y.W., Won, C.S. (2022). Object detection and classification based on YOLOv5 with improved maritime dataset. Journal of Marine Science and Engineering, 10(3): 377. <https://doi.org/10.3390/jmse10030377>

[6] Kuznetsova, A., Maleva, T., Soloviev, V. (2020). Detecting apples in orchards using YOLOv3 and YOLOv5 in general and close-up images. In: Han, M., Qin, S., Zhang, N. (eds) Advances in Neural Networks – ISNN 2020. ISNN 2020. Lecture Notes in Computer Science(), vol 12557. Springer, Cham. http://dx.doi.org/10.1007/978-3-030-64221-1_20

[7] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2021). You only look once: unified, real-time object detection. arXiv:1506.02640. <https://doi.org/10.48550/arXiv.1506.02640>

[8] Widayawati, W., Febriani, R. (2021). Real-time detection of fruit ripeness using the YOLOv4 algorithm. Teknika: Jurnal Sains Dan Teknologi, 17(2): 205-210. <http://dx.doi.org/10.36055/tjst.v17i2.12254>

[9] Bargouti, S., Underwood, J. (2017). Deep fruit detection in orchards. IEEE International Conference on Robotics and Automation (ICRA), Singapore, pp. 3626-3633. <https://doi.org/10.1109/ICRA.2017.7989417>

[10] Muresan, H., Oltean, M. (2018). Fruit recognition from images using deep learning. Acta Universitatis Sapientiae, 10: 26-42. <https://doi.org/10.2478/ausi-2018-0002>

[11] Zhang, Y.T., Sohn, K., Villegas, R., Pan, G., Lee, H. (2016). Improving object detection with deep convolutional networks via bayesian optimization and structured prediction. arXiv:1504.03293. <https://doi.org/10.1109/CVPR.2015.7298621>

[12] Chung, D.T.P., Van Tai, D. (2019). A fruits recognition system based on a modern deep learning technique. Journal of Physics: Conference Series, 1327: 012050. <https://doi.org/10.1088/1742-6596/1327/1/012050>

[13] Nguyen, H.H.C., Nguyen, D.H., Nguyen, V.L., Nguyen, T.T. (2020). Smart solution to detect images in limited visibility conditions based convolutional neural networks. In: Advances in Computational Collective Intelligence. ICCCI 2020. Communications in Computer and Information Science, 1287: 641-650. https://doi.org/10.1007/978-3-030-63119-2_52

[14] Cuong, N.H.H., Trinh, T.H., Meesad, P., Nguyen, T.T. (2022). Improved YOLO object detection algorithm to detect ripe pineapple phase. Journal of Intelligent & Fuzzy Systems, 43(1): 1365-1381. <https://doi.org/10.3233/JIFS-213251>

[15] Valentino, F., Cenggoro, T.W., Pardamean, B. (2021). A design of deep learning experimentation for fruit freshness detection. IOP Conference Series: Earth and Environmental Science, 794: 012110.

<https://doi.org/10.1088/1755-1315/794/1/012110>

[16] Kumar, Y., Dubey, A.K., Arora, R.R., Rocha, A. (2020). Multiclass classification of nutrients deficiency of apple using deep neural network. *Neural Computing and Applications*, 34: 8411-8422. <https://doi.org/10.1007/s00521-020-05310-x>

[17] Karakaya, D., Ulucan, O., Turkan, M. (2019). A comparative analysis on fruit freshness classification. *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Izmir, Turkey, pp. 1-4. <https://doi.org/10.1109/ASYU48272.2019.8946385>

[18] Agarap, A.F. (2019). Deep learning using rectified linear units (RELU). *Neural and Evolutionary Computing*. <https://doi.org/10.48550/arXiv.1803.08375>

[19] Feng, J., Zeng, L.H., He, L. (2019). Apple fruit recognition algorithm based on multi-spectral dynamic image analysis. *Sensors*, 19(4): 949. <https://doi.org/10.3390/s19040949>

[20] Afaq, S., Rao, S. (2020). Significance of epochs on training a neural network. *International Journal of Scientific and Technology Research*, 19(6): 485-488.

[21] Ren, S.Q., He, K.M., Girshick, R., Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Proceedings of the Advances in Neural Information Processing Systems*, pp. 91-99. <https://doi.org/10.1109/TPAMI.2016.2577031>

[22] Gai, W., Liu, Y., Zhang, J., Jing, G. (2021). An improved Tiny YOLOv3 for real-time object detection. *Systems Science and Control Engineering*, 9(1): 314321. <https://doi.org/10.1080/21642583.2021.1901156>

[23] Liu, G., Nouaze, J.C., Touko Mbouembe, P.L., Kim, J.H.

(2020). YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3. *Sensors*, 20(7): 2145. <https://doi.org/10.3390/s20072145>

NOMENCLATURE

AI	Artificial Intelligence
CNN	Convolutional neural networks
IoT	Internet of Things
Faster R-CNN	Fast Region-based Convolutional Network
YOLO	You Only Live Once
TP	True Positives
TN	True Negatives
FP	False Positives
FN	False Negatives
V5	Version 5

APPENDIX

- 1) <https://universe.roboflow.com/gan-nhan/pineapple-data/dataset/1/images>; (Link dataset image).
- 2) <https://drive.google.com/file/d/1u3LkA7baaPp4cEAO9wEGb6YB9rAWd3jH/view?usp=sharing>; (Link dataset).
- 3) <https://traiduachin.surge.sh/>; (Link Website).
- 4) <https://play.google.com/store/apps/details?id=vn.pineapple.app>; (Link App Play).
- 5) <https://apps.apple.com/us/app/tr%C3%A1i-d%E1%BB%A9a-ch%C3%ADn/id6449583263>; (Link AppStore).