JIAISE

Journal of Integrated Artificial Intelligence Science and Engineering

Vol. 1, No. 1, March 2025, pp. 1-4, DOI: 10.59190/jiaise.v1i1.302

PUBLISHED BY

LSPRAR

Utilizing computer vision for the detection, classification, and mapping of coffee cherries during the harvest: A review

M Fajar Syahputra

Department of Physics, Universitas Riau, Pekanbaru 28293, Indonesia

ABSTRACT

Computer vision is a technology that integrates image processing and pattern recognition to extract information from digital images. This research uses the you only look once (YOLO) method to detect, classify, and map the maturity stage of coffee fruit during harvest. The YOLOv3-tiny model was applied to classify coffee fruits into three categories including unripe, ripe, and overripe. Results showed an average precision of 86.0%, 85.2%, and 80.0%, respectively. The system enables more efficient harvest management by utilizing spatial and temporal information for coffee fruit quality mapping. This mapping can help farmers reduce operational costs and improve efficiency through the application of precision farming techniques. This technology proves great potential in improving the quality and quantity of coffee yields with a fast and accurate computer-based approach.

ARTICLE INFO

Article history:

Received Jan 5, 2025
Revised Feb 8, 2025
Accepted Mar 11, 2025

Keywords:

Coffee Fruit Ripeness
Computer Vision
Object Detection
Spatial Mapping
YOLO

* **Corresponding Author**
E-mail address: m.fajar5044@student.unri.ac.id

1. INTRODUCTION

The majority of farmers still harvest coffee fruits using traditional methods, which yield uneven flowering throughout time. These methods harvest coffee fruits in different maturity processes (raw, ripe, and overripe). Market planning, labour resource management, and agronomic care can all be incorporated into the science of coffee plant ripening. Coffee fruit samples' colour is used to assess the ripening phase. The coffee ripening process is not well displayed spatially due to the low sampling density. Computer vision turned out to be a promising method in this instance, particularly when it came to object detection and offering a thorough pixel-by-pixel description of the object's colour uniformity. Relatively little work has been done to identify coffee fruits and categorise their ripening phase, despite recent advancements in computer vision techniques.

Few studies have been conducted to identify coffee fruits and classify their ripening stage despite recent advancements in computer vision techniques. These studies include developing a machine vision system for mobile devices that can identify and classify coffee fruits on branches regardless of environmental conditions, classifying the vegetative structure of coffee branches based on the acquisition of 2D and 3D features from field videos, and processing images taken during the coffee harvest There is still a knowledge gap, nonetheless, regarding high spatial resolution data on coffee plants. While assessing the crop's ripening stage has advantages, this method is ineffective for assessing coffee ripening uniformity widely at the field level.

2. THEORITICAL REVIEW

2.1. Computer Vision

Pattern recognition and image processing are combined to form computer vision. One area of study in computer science and artificial intelligence is computer vision. Unlike human vision, computer vision requires extensive, intricate logic operations to develop. As a result, it has been used for decades in image recognition, classification, detection, and segmentation. The use of computer vision has expanded to include anything from raw data recording to information extraction and interpretation. incorporates ideas, methods, and concepts from computer graphics, artificial intelligence, pattern recognition, and digital image processing. The majority of computer vision tasks include gathering details about events or descriptions from digital picture input scenes and outside elements. The application domain and the type of data being examined determine the approaches taken to address computer vision issues.

2.2. You Only Look Once (YOLO)

One technique for real-time object detection is the YOLO method. A network with fewer convolutional layers and a simplified feature extractor, YOLOv3-tiny is an adaptation of the YOLOv3 version. Though just at two different scales, YOLOv3-tiny makes predictions at many scales, just like YOLOv3. Even though it is a less resilient network, it should perform well enough to train and detect more quickly for smaller tasks.

3. RESEARCH METHODS

3.1. Identification of Coffee Fruit Maturity

Coffee quality is determined by the stages of coffee fruit harvesting. Harvesting the coffee fruit from the tree is the first step. The harvesting process involves only taking extremely mature coffee fruit because combining it with raw coffee fruit can degrade its quality and flavour. Farmers typically harvest coffee fruit by hand during this season. Due to the physical and hurried nature of coffee harvesting, farmers frequently select coffee cherries that are still considered immature. The process of selecting coffee fruit is followed by soaking the faulty fruit in a tub of water to sort it. It is necessary to differentiate healthy coffee fruit from faulty coffee fruit that floats in water. The coffee fruit then moves on to the drying step.

3.2. YOLO Method

This strategy is used to classify objects and determine what they are. A number of grid cells will be created from the input image. One grid cell must be used to forecast each object in the picture. Every grid cell will forecast class likelihood C and bounding box B. Five elements make up the bounding box: x, y, w, h, and confidence level.

The original YOLO clarifying network versions (YOLOv2 and YOLOv3) underwent a series of adjustments to attain state-of-the-art performance comparable to other reference methods while retaining its high processing speed. The system displays the bounding box (or surrounding rectangle) prediction of the object for detection and categorisation. Each network prediction sensor cell undergoes bounding box prediction, where three bounding boxes are anticipated using three "anchor cities" as the base. Although the bounding box dimensions can be altered, the "anchor boxes" have default beginning dimensions set up to aid in network learning. By using the clustering technique on the training dataset's annotated bounding boxes, the "anchor box" settings were produced. There are five forecasts for every bounding box. The "objectivity" prediction, or the likelihood that an object will be in the anticipated bounding box, is one of these predictions, whereas the other four are connected to the bounding box's coordinates.

Three distinct scales are used to make predictions throughout the YOLOv3 feature extractor. In other words, the initial prediction is created at a size 32 times smaller than the input image following a sequence of convolutional layers. The second prediction is then made by combining the improved features from the current layer with the feature map from the prior layer, and then adding more convolutional layers. A sensor twice as large as the previous one is used to make the second prediction.

For forecasts at the third and final scale, same procedure is repeated. This makes it possible for the prediction to profit from both the initial feature extractor computations and the improved features.

4. RESULTS AND DISCUSSIONS

4.1. Performance of Object Detection

Figure 1 illustrates the three kinds of fruit ripening phases (unripe, ripe, and overripe) identified by the YOLOv3-tiny model in its performance analysis. The model classifies the validation cells as unripe, ripe, and overripe, with an AP of roughly 86.0%, 85.0%, and 80%. Because overripe and ripe fruits had similar colours, there was considerable uncertainty during classification as a result of the poorer precision for overripe fruit classification. One could argue that the model repeatedly produced false-positive results by failing to correctly categorise overripe fruits. As the resolution of the input photos improved, so did the model's AP for identifying and categorising coffee fruits in the validation set. When an input resolution of 800 x 800 pixels was utilised, the performance similarly peaked. The fact that Yolo's convolutional and pooling layers progressively reduce the spatial dimension as network depth increases can be utilised to characterise this improvement in results with increased image resolution. Small or hazy fruits, for example, may be challenging to discern at lesser resolutions.

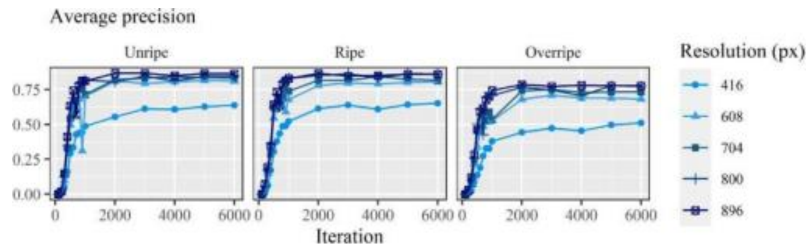

Figure 1. For the validation set, the YOLOv3-tiny model's average precision using Fruit grade coffee.

Figure 2. Two random video frames collected during coffee harvesting: (a) and (c) are the original frame with fruits at different stages of ripening; (b) and (d) are the detection performed by the proposed model.

Depending on the cultivar and growth stage, there may be differences in the colour of the fruits, leaves, and the quantity of contaminants left over after the harvester's pre-cleaning procedure. As a result, fruit recognition and classification outcomes may differ when using algorithms that solely use colour to identify items. The trained model YOLOv3-tiny-800's predictions for the test set's photos were visually evaluated (Figure 2). While the majority of the coffee fruits in these pictures have been accurately recognised and categorised, some have not been found or categorised. The first arbitrary frame (Figure 2 (a)) had a mAP of 85.6%, while the second arbitrary frame (Figure 2 (c)) had a mAP of 72.1%. The higher fruit density in the second frame is the reason for its lower mAP. Increased fruit density could lead to object of interest overlap and make detection more difficult. Additionally, groupings of fruits of the same colour may become blurry if the image is downscaled to 800 × 800 pixels during detection.

4.2. Quality mapping: Coffee maturation phases' spatial variability

A threshold of 40 fruits was chosen from this analysis as the maximum number of detected fruits that may be used, including for coffee fruit picking quality. Figure 3 (a) displays the total number of coffee fruits found in each frame of the random movie; Figure 3 (b) displays the total number of fruits found in each class; and Figure 3 (c) displays the percentage of recognised classes per frame. The percentage of ripe fruits is fairly constant along with the frame, or harvest line, with a slight increase in immature fruits between frames 4500 and 6500. In comparison to ripe fruits (21.0% – 59.6%) and immature fruits (0% – 25.6%), the majority of overripe coffee fruits (31.4% – 75.3%) were found. The experimental area's delayed harvesting start was the cause of the overall greater percentage of overripe fruits. This region was chosen especially for the experiment and was contingent upon labour availability during the busiest harvesting season.

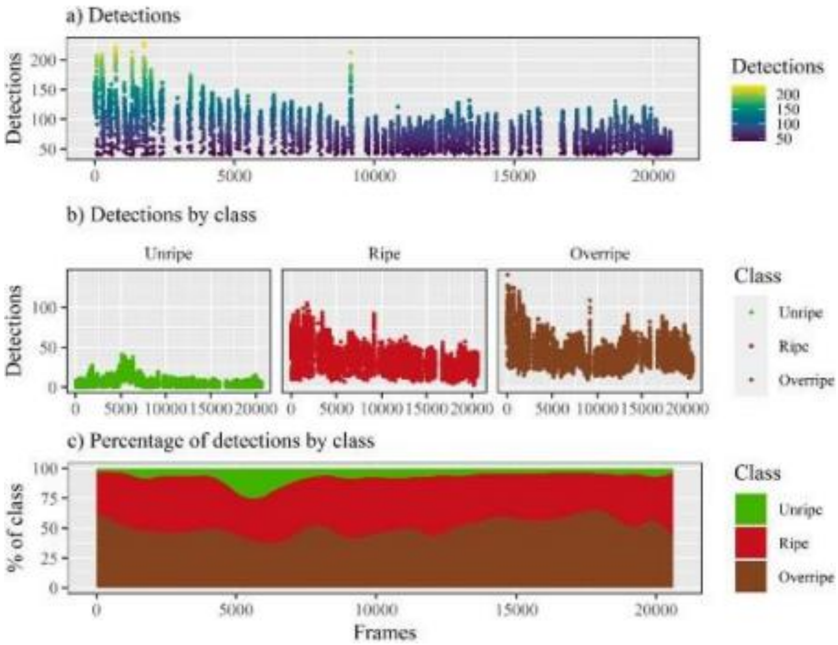

Figure 3.3 shows the fraction of detected classes, the total number of fruits detected, and the total number of detections by ripening class for random films.

5. CONCLUSION

In spite of various contextual factors, including variations in background lighting conditions, contrast between the fruit and its surroundings, and vibrations brought on by various harvesting angles, the deep learning model created in this study shows the ability to identify coffee fruits and categorise their ripening stages. With 86.0% accuracy for immature fruits, 85.2% accuracy for ripe fruits, and 80.0% accuracy for overripe fruits, the model demonstrated remarkable average precision rates. These outcomes demonstrate how well the model can identify and categorise coffee fruit maturation stages.

REFERENCES

[1] Bazame, H. C., Molin, J. P., Althoff, D., & Martello, M. (2021). Detection, classification, and mapping of coffee fruits during harvest with computer vision. *Computers and Electronics in Agriculture*, **183**, 106066.

[2] Munawar, Z., Hasnawi, M., Beno, I. S., Darma, I. W. A. S., Sastradipraja, C. K., Sutramiani, N. P., Amir, A., Manurung, H. E., Kharisma, I. L., Mandowen, S. A., Pomalingo, S., Indra, D., Irmawati, Setyawan, A. A., Mansyur, S. H., & Koibur, M. E. (2023). *Visi komputer: Konsep, metode, dan aplikasi*. Kaizen Media Publishing.

[3] Hidayatulloh, M., Saputro, A. K., Alfita, R., Ulum, M., & Hardiwansyah, M. (2023). Rancang bangun alat sortir kualitas buah kopi Wonosalam berbasis image processing. *Multitek Indonesia*, **17**(1), 79–93.