Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

OBJECT DETECTION USING CONVOLUTIONAL NEURAL NETWORK YOLOV7 TO DETECT BANANA RIPENESS

DICKY ANTONY [1], AHMAD NURUL FAJAR [2]

[1,2] Information System Management Department, BINUS Graduate Program – Master of Information Systems Management, Bina Nusantara University, Jakarta 11480, Indonesia

E-mail: [1]dicky.antony@binus.ac.id, [2]afajar@binus.edu

ABSTRACT

This study addresses the challenge of ensuring fruit quality in Indonesia, the 8th largest fruit producing country globally. Despite favorable environmental conditions, many harvested fruits fail to meet quality standards due to various factors such as inadequate water content and soil conditions. To tackle this issue, Convolutional Neural Network (CNN) modeling is employed to assess the quality of golden bananas. This study utilizes the YoloV7 model to detect bananas based on skin color, distinguishing between grade A and grade B bananas. The model achieves a mean Average Precision (mAP) of 78.1%, with grade A achieving 99.5% and grade B achieving 56.7% in Average Precision (AP). These findings contribute to enhancing fruit quality assessment methods and offer a potential solution to improve the quality of harvested fruits.
Keywords: *Object Detection, CNN, YoloV7. Deep Learning.*

1. INTRODUCTION

Bananas, cherished for their delightful taste and nutritional benefits, have entrenched themselves as a beloved fruit among the populace of Indonesia. Their widespread popularity is evident in the consumption patterns of Indonesian individuals, which exhibit a consistent upward trajectory over the years. While occasional fluctuations, such as the declines observed in 2019 and 2020, may occur, they do little to detract from the enduring appeal of bananas within the Indonesian culinary landscape. This enduring affection for bananas underscores their status as a staple food item and highlights their significance as a cultural icon in Indonesia. As a versatile fruit enjoyed in various forms, from fresh consumption to culinary preparations, bananas play an integral role in the dietary habits and cultural fabric of Indonesian society [1].

Table 1: Average Per Capita Banana Consumption a Week in Indonesia

Tahun	Rata-rata Konsumsi Perkapita Seminggu di Indonesia
2018	0,169
2019	0,158
2020	0,153
2021	0,159

There are myriad factors contributing to the widespread popularity of bananas among the Indonesian populace. Foremost among these is their delectable taste and nutritional richness. Renowned for their naturally sweet flavor and pleasingly soft texture, bananas serve as a delectable snack option cherished by individuals of all ages across Indonesia. Beyond their taste appeal, bananas boast an impressive nutritional profile, brimming with essential vitamins, minerals, and dietary fiber. Rich in potassium, vitamin C, and vitamin B6, bananas offer a nourishing boost to one's diet, promoting overall health and well-being. Additionally, their convenient packaging—encased in a naturally protective peel—renders them a convenient on-the-go snack, perfect for busy individuals seeking a quick and satisfying source of sustenance. It is this irresistible combination of taste, convenience, and nutritional goodness that has cemented bananas as a perennial favorite among the Indonesian populace [2].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Number of Banana Plant Production in Indonesia

Tahun	Jumlah Produksi Tanaman Buah Pisang di Indonesia (Ton)
2018	7.264.383
2019	7.280.658
2020	8.182.756
2021	8.741.147

Moreover, the abundant availability of bananas in Indonesia further contributes to their widespread popularity. Endowed with a tropical climate conducive to banana cultivation, Indonesia boasts an ideal environment for banana growth. Bananas flourish in various regions across the archipelago, thriving both as commercial crops and in the verdant gardens of households. The ubiquity of banana cultivation is evidenced by the continuous upward trend in banana production throughout Indonesia. This bountiful supply ensures a consistent availability of bananas to meet the ever-high demand within the Indonesian market. From bustling urban centers to remote rural villages, bananas remain readily accessible to individuals from all walks of life, serving as a staple component of the Indonesian diet. This seamless integration into the fabric of Indonesian agriculture and culinary culture underscores the enduring significance of bananas as a beloved fruit cherished by millions across the archipelago [3].

In banana production there are stages in the form of sorting or sorting which are still done manually. Because the sorting that is carried out still uses manual methods, it can result in problems both in terms of the quality of the fruit being sorted being not good, the quantity of fruit being reduced, or disrupting the economy of banana farmers. Technological developments can now be applied to the agricultural sector so that the sorting process can be carried out automatically. When compared to other detection models like DPM and R-CNN, research on You Only Look Once (YOLO) models has shown higher object identification speed and performance. The most recent YOLO models, including YoloV7 [4], have undergone constant improvement and feature improved object identification and faster training. Although YoloV7 has been applied to agricultural contexts in a number of studies, this model has not yet been used to evaluate the quality of curly red banana. Consequently, using a camera-interface interface, this work uses YoloV7 for the identification and categorization of banana pepper quality [5]. Our goal is to improve sorting process accuracy and efficiency for farmers, neighborhood collectors, and market vendors. Additionally, tangentially supporting the stability of banana pepper prices in the market.

2. LITERATURE REVIEW

2.1 Convolutional Neural Network

CNN (Convolutional Neural Network) is one method that can be applied from Deep Learning. CNN itself has the meaning of a development method from MLP (Multi Layer Perception) which is intended to process two-dimensional data and has a high network depth and is widely used in image classification [6].

Convolutional Layer is a layer that calculates the output of neurons connected to a small region or receptive field at the input. The receptive field is a small area in the input volume that is connected to weights or filters that can be changed in the convolutional layer. After the convolutional layer, it is usually followed by an activation layer (such as ReLU) and a pooling layer, which will reduce the dimensions of the output convolutional layer [7].

Pooling Layer is a layer that can speed up computation by reducing the dimensions of the feature map because by doing this it can overcome the overfitting problem. Fully connected layer is a layer in a neural network where every node in the output layer is connected to every node in the previous layer. This layer is responsible for performing classification tasks based on previously extracted features.

2.2 YoloV7

The most recent iteration of YOLO, known as YoloV7, was developed by C.-Y. Wang et al. In the 5–120 FPS range, this model can outperform all current state-of-the-art real-time object detectors in terms of accuracy and speed [8]. A complex object detection model is created by combining the best characteristics of the YOLO framework with the YoloV7 architecture. It begins by analyzing an input picture and extracting information at different sizes using a backbone network. Contextual information can be further enhanced with an optional neck component [9]. Detection heads then use feature maps from the neck or backbone to estimate bounding boxes, class probabilities, and object-specific attributes. The most reliable bounding

Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

boxes with their matching class labels and confidence scores are retained, and redundant detections are eliminated through the use of post-processing techniques like non-maximum suppression. To enable accurate and speedy item detection, YoloV7 blends the core YOLO concepts with practical design decisions [10].

3. RESEARCH METHODS

3.1 Data Gathering

Through this research, bananas were used as objects in the research and data collection consisted of 1000 image samples divided into 2 different grades, namely Grade A and Grade B with each grade containing 500 images. In this research, golden banana image data was collected using a Xiaomi Pocophone F1 smartphone camera with a camera resolution of 12 megapixels.

Table 3: Banana Variable Data

Image of Banana	Variable	Variable Description
	Grade A	The image of a grade A banana is a banana that is green or dark in color, has a shape that is very hard, has a bland fruit taste, also a banana that has just grown and is not yet ripe.
	Grade B	The image of a grade B banana is a banana that is green or light in color, has a shape that is still very hard, has a bland fruit taste, and is not yet ripe.

3.2 Data Pre – Processing

The next stage after data processing will be carried out in several phases, namely as follows:

a. Augmentation involves enhancing the dataset to improve model performance. Initially, 200 photos were selected for processing using the Jupiter Python program. The selected photos were cropped to focus on the chili peppers, and each photo was resized to a standard resolution of 640 x 640 pixels. RGB color channels were extracted from the photos to capture the color information of the chilies effectively. Finally, the dataset was organized and prepared for further analysis, particularly for training deep learning models.

b. Annotation is the process of labeling objects of interest in images. In this case, the dataset underwent annotation using the Roboflow website. Each of the 1000 photos in the dataset was segmented into frames for annotation purposes. Bounding boxes were manually placed around the chili peppers in each image to indicate their locations. Additionally, depth labels were assigned to provide information about the distance of the chilies from the camera. These annotations serve as ground truth data, which is crucial for training accurate object detection models.

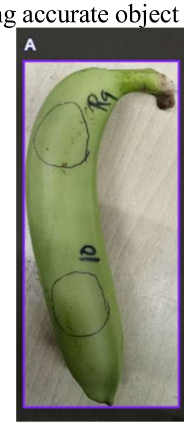

Figure 1: Annotation Image 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

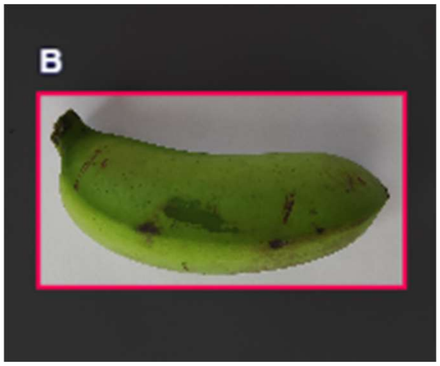

Figure 2: Annotation Image 2

c. Dataset Split To facilitate model training and evaluation, the dataset was divided into three subsets: training, testing, and validation. The training set, comprising 80% of the sample photos, was used to train the model on identifying chili peppers. The validation set, consisting of 10% of the sample photos, was used to tune hyperparameters and monitor the model's performance during training. The testing set, also containing 10% of the sample photos, was used to assess the final performance of the trained model.

3.3 Model Configuration and Training Details

In the model configuration phase, several parameters are set to tailor the model's architecture and behavior. Notably, the `nc` parameter defines the number of object classes to detect, crucial for identifying various types of chili peppers. Additionally, parameters such as `depth_multiple` and `width_multiple` offer control over the depth and width of the model, allowing for customization of its size and complexity. Anchors play a pivotal role in determining the size of bounding boxes at different resolution levels within the model, facilitating accurate object localization. The backbone of the YoloV7 model comprises convolutional layers applied across various resolution levels, extracting features from input images essential for subsequent object detection tasks. This feature extraction process is complemented by the head part of the model, responsible for processing extracted features and generating predictions regarding the presence and location of chili peppers in the image. Iterative convolution, known as RepConv, enhances the model's representation capacity by iteratively applying convolution operations on multiple layers. Finally, the IDetect operation in the last layer of the model produces object detections based on a predefined number of classes and bounding boxes, effectively completing the model's configuration for object detection tasks.

3.4 Testing and Evaluation

In the testing and evaluation phase, Google Colab and the PyTorch package are utilized to assess the performance of the study model. This phase involves the comprehensive utilization of the training, testing, and validation datasets to evaluate the model's accuracy and reliability. Evaluation metrics such as Precision-Recall (mAP), Precision, Recall, and the F1-score are employed to gauge the model's effectiveness in detecting chili peppers in images. Particularly emphasized is the F1-score metric, which offers detailed insights into the model's performance across different classes or categories, providing a more nuanced assessment than overall accuracy. Through rigorous evaluation, the correctness and dependability of the model are ensured, providing valuable insights into its real-world applicability and effectiveness in chili pepper detection tasks [11].

4. RESULTS AND DISCUSSION

Following the meticulously crafted research methodology, the subsequent pivotal step entails the meticulous training of the meticulously collected data. Leveraging the robust architecture of YoloV7, revered for its prowess in feature extraction, the model embarks on a journey through the vast expanse of data, guided by a multitude of meticulously chosen parameters, including epoch, batch size, and learning rate. With the veil of training lifted, the fruits of this laborious endeavor are subjected to meticulous scrutiny and intricate analysis. From this arduous odyssey emerges a plethora of paramount revelations, vividly depicted in the intricate tapestry of curves. The F1-Score, a beacon of holistic evaluation, casts its illuminating gaze upon the model's performance, revealing an average value of 0.80. However, amidst this commendable achievement lies a tempestuous conundrum, as the threshold of 0.178 occasionally casts shadows of doubt, leading to the emergence of false positives (FP). This precipitates an urgent call for recalibration, urging a recalibration of confidence levels and the quest for a loftier threshold to navigate the treacherous waters of detection. In the realm of precision-recall curves, Grade A stands as a paragon of virtuous precision, boasting a resplendent level of 99.5%

true positives (TP). In contrast, Grade B, though valiant in its efforts, achieves a more modest precision level of 56.7% TP, with the specter of false positives haunting its domain. The mean Average Precision (mAP), a cornerstone of model evaluation, ascends to a commendable pinnacle of 0.781, showcasing the YoloV7 model's mettle at a threshold of 0.5. This resounding achievement resounds with the symphony of high precision, echoing throughout the hallowed halls of research. The precision curve, a testament to the model's prowess, unfurls its majestic banner, unfurling a triumphant TP level of 0.738 without the slightest whisper of FP. Yet, amidst this euphoric crescendo lies a shadow of discord, as the recall curve scales dizzying heights, reaching a vertiginous apex of 0.99 at the lowest threshold. While this lofty ascent heralds the promise of exhaustive recall, it also unfurls the banner of caution, beckoning forth the imperative of precision enhancement to stave off the specter of erroneous detections. Furthermore, the model's discerning gaze alights upon Grade B bananas, adorned with their lustrous mantle of dark green skin. This keen-eyed discernment facilitates the demarcation between the ripened hues of maturity and the verdant blush of youth, obviating the need for reliance solely upon the fallible faculties of visual inspection. In the effulgent glow of these revelatory insights, the model's performance is cast into stark relief, underscoring the pivotal importance of precision enhancement in the unending quest for more accurate object detection and the boundless horizons of knowledge creation.

.

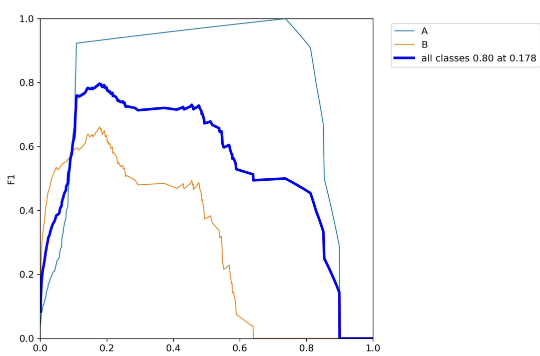

Figure 3: F1-Score Result YoloV7

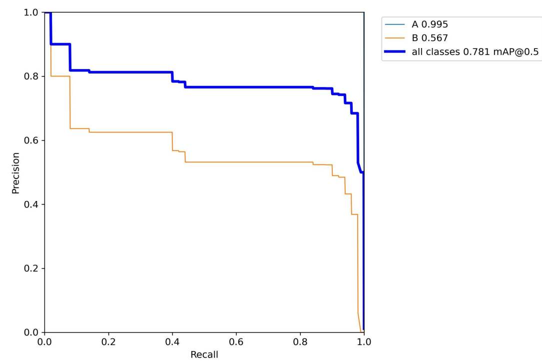

Figure 4: Precision Recall Curve Result YoloV7

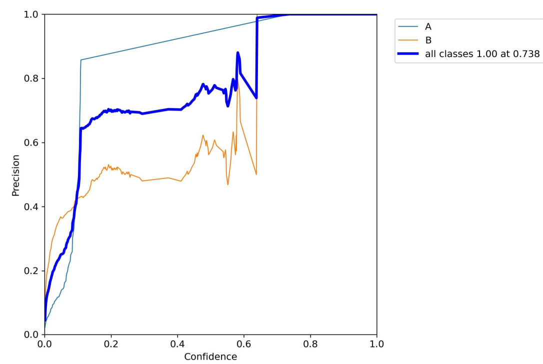

Figure 5: Precision Curve Result YoloV7

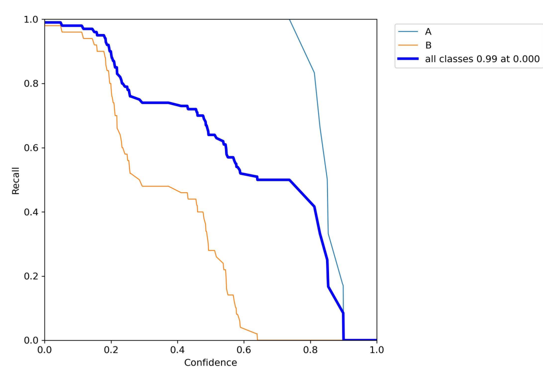

Figure 6: Recall Curve Result YoloV7

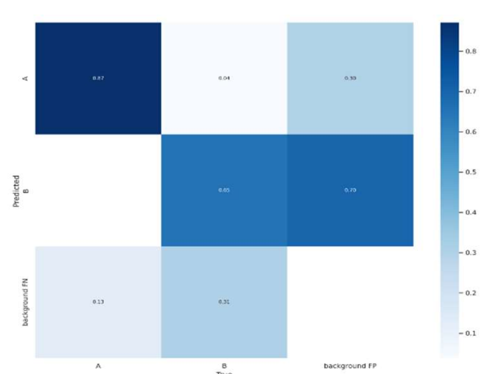

Figure 7: Confusion Matrix

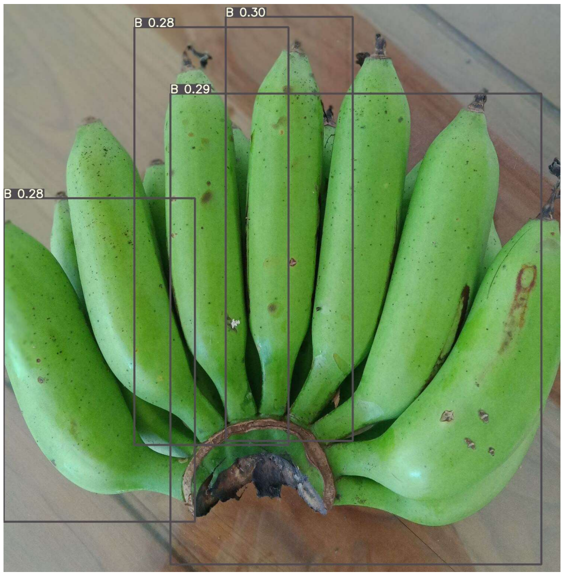

Figure 8: Test Result Banana YoloV7

5. CONCLUSION

In conclusion, this research represents a significant stride towards enhancing the economic prospects of banana collectors through improved sorting methodologies facilitated by object detection techniques. By harnessing the power of advanced technology and leveraging the insights gleaned from our research endeavors, banana collectors stand to benefit from more efficient and accurate banana sorting processes, ultimately leading to enhanced economic outcomes.

The findings of our model training, conducted with a batch size of 16 and over 50 epochs, yielded commendable results. With a mean Average Precision (mAP) of 78.1%, and Average Precision (AP) scores of 99.5% for Grade A bananas and 56.7% for Grade B bananas, the model demonstrated proficiency in accurately detecting and classifying bananas according to predetermined grades. These outcomes underscore the efficacy of our research methodology and the potential for practical implementation in real-world scenarios.

Looking ahead, there is ample opportunity for future research to build upon our findings and propel this technology to even greater heights. One promising avenue for exploration involves the development of practical applications, such as the creation of machinery capable of implementing object detection algorithms in real-time sorting processes. By bridging the gap between research and application, future endeavors have the potential to revolutionize banana sorting practices, further optimizing efficiency and economic viability for stakeholders across the industry.

In essence, our research serves as a catalyst for innovation and progress in the field of banana sorting, offering tangible solutions to real-world challenges. As we continue to push the boundaries of technological advancement, we remain steadfast in our commitment to driving positive change and empowering communities through the transformative power of research and innovation.

REFERENCES:

[1] Badan Pusat Statistika, "Rata-rata Konsumsi Perkapita Seminggu Menurut Kelompok Buah-Buahan Per Kabupaten/kota (Satuan Komoditas), 2021." Accessed: May 18, 2023. [Online]. Available: https://www.bps.go.id/indicator/5/2102/2/rata-rata-konsumsi-perkapita-seminggu-menurut-kelompok-buah-buahan-per-kabupaten-kota.html

[2] Badan Pusat Statistika, "Produksi Tanaman Buah-buahan 2021." Accessed: May 18, 2023. [Online]. Available: https://www.bps.go.id/indicator/55/62/1/produksi-tanaman-buah-buahan.html

[3] Deddy Prayama, Amelia Yolanda, and Andi Wellyno Pratama, "Rancang Bangun Alat Pengontrol Penyiram Tanaman Otomatis Menggunakan Sensor Kelembaban Tanah Di Area Pertanian," *JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 2, pp. 807–812, Dec. 2018.

[4] X. Meng, C. Li, J. Li, X. Li, F. Guo, and Z. Xiao, "YOLOv7-MA: Improved YOLOv7-Based Wheat Head Detection and Counting," *Remote Sens (Basel)*, vol. 15, Nov. 2023, doi: 10.3390/rs15153770.

[5] S. Wang, D. Wu, and X. Zheng, "TBC-YOLOv7: a refined YOLOv7-based algorithm for tea bud grading detection," *Front Plant Sci*, vol. 14, p. 1223410, Nov. 2023, doi: 10.3389/fpls.2023.1223410.

[6] F. Paraijun *et al.*, "Implementasi Algoritma Convolutional Neural Network Dalam Mengklasifikasi Kesegaran Buah Berdasarkan Citra Buah," vol. 11, no. 1, 2022, doi: 10.33322/kilat.v11i1.1458.

[7] Muh. Ibnu Choldun R and Kridanto Surendro, "KLASIFIKASI PENELITIAN DALAM DEEP LEARNING," pp. 25–31.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[8] Y. Sun, Y. Li, S. Li, Z. Duan, H. Ning, and Y. Zhang, "PBA-YOLOv7: An Object Detection Method Based on an Improved YOLOv7 Network," *Applied Sciences*, vol. 13, p. 10436, Nov. 2023, doi: 10.3390/app131810436.

[9] X. Chen, Y. Xue, Q. Hou, Y. Fu, and Y. Zhu, "RepVGG-YOLOv7: A Modified YOLOv7 for Fire Smoke Detection," *Fire*, vol. 6, p. 383, Nov. 2023, doi: 10.3390/fire6100383.

[10] L. Zhao and M. Zhu, "MS-YOLOv7:YOLOv7 Based on Multi-Scale for Object Detection on UAV Aerial Photography," *Drones*, vol. 7, p. 188, Nov. 2023, doi: 10.3390/drones7030188.

[11] S. Sharma and A. Singhal, "A Comprehensive Investigation of Machine Learning Algorithms with SMOTE Integration to Maximize F1 Score," 2023, pp. 187–199. doi: 10.1007/978-981-99-2100-3_16