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 OF CHILI USING CONVOLUTIONAL NEURAL NETWORK YOLOV7

RICHARD SALIM [1], AHMAD NURUL FAJAR [2]

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

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

ABSTRACT

In Indonesia, the production of red curly chili faces challenges in stabilizing market prices, leading to a growing dependence on chili imports to maintain stability. Import figures surged by 237.07% in early 2023, rising from 1.24 million kilograms in January 2022 to 4.18 million kilograms. This reliance on imports is primarily due to the rigid distribution system of chili peppers, which closely follows farmers' harvest schedules. Consequently, inconsistent chili availability and uncertain quality result from prolonged regional distribution times, impacting market prices. The sorting process is crucial in determining prices for all participants within the Indonesian chili supply chain. Unfortunately, the current manual sorting process is plagued with shortcomings, negatively affecting the efficiency of the entire chili supply chain. Therefore, it is crucial to develop innovative strategies to aid supply chain participants in chili cultivation and boost chili sales by automating the sorting process. In this research initiative, our team proposes a solution involving the development of the YOLOv7 model for automatic detection and classification of high-quality red curly chili. Our approach included collecting image data, rigorous data preprocessing, and hyperparameter optimization. The YOLOv7 model demonstrated commendable performance, achieving an impressive overall grade Mean Average Precision (mAP) of 0.977. Additionally, it exhibited noteworthy average precisions (AP) with scores of 0.996 for grade A, 0.947 for grade B, 0.951 for grade C, and 0.996 for grades D and E.

Keywords: *Deep Learning, Object Detection, Object Classification, YOLOv7.*

1. INTRODUCTION

The Indonesian economy can significantly benefit from horticultural goods due to their high economic value and the country's ideal equatorial climate for cultivation. Among these goods, chili peppers hold particular importance as they are integral to Indonesian cuisine, consumed raw or processed into various products such as sambal, chili powder, pickled chili, and more. Their rich vitamin C content further enhances the economic value of chili peppers. Notably, curly red chili stands out as Indonesia's most consumed type of chili pepper, making it a focal point in this study due to its pivotal role in the horticultural commodity market. According to data from the National Socio-Economic Survey (2020-2023), the total consumption of curly red chili in 2021 amounted to 490.83 thousand tons, marking a 9.94% increase from the previous year and representing the highest consumption level in the past five years [1]. Production of curly red chili has also shown a steady increase in the past two years, as indicated in Table 1 by the Central Statistics Agency [2].

Table 1: Curly Red Chili Production in Indonesia

Curly Red Chili Production in 2021	Curly Red Chili Production in 2022
8.601.851	10.173.818

Despite the increasing production of curly red chili, Indonesia continues to import significant quantities to stabilize market prices. Data from CNBC Indonesia indicates a 237.07% surge in chili pepper imports in early 2023, reaching 4.18 million kilograms, compared to 1.24 million kilograms in January 2022. The Ministry of Trade reports that nearly 65% of Indonesia's essential food resources are still imported [3]. This dependency on imports is primarily attributed to the rigid distribution system of chili peppers, which aligns closely with farmers' harvest schedules. This leads to inconsistent chili availability and uncertain quality due to

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

prolonged distribution times between regions. This unpredictability can harm market prices.

Building on previous research [4], the sorting process's significance is evident, occurring three times in the chili pepper supply chain before progressing to the subsequent stages. This holds particular importance for market traders, as interviews with local market traders have revealed that sorting is instrumental in maintaining their selling prices; failure to sort typically results in price reductions. However, manual hand sorting remains prevalent despite inherent drawbacks, including labor intensiveness, time consumption, conflict susceptibility, and human subjectivity that may jeopardize pre-established agreements between parties [5]. To address the limitations of manual sorting, this study investigates the application of computer vision, specifically a convolutional neural network (CNN), known for its effectiveness in object detection and agricultural product classification. Object detection, a challenging aspect of computer vision, involves identifying instances of semantic objects of a specific class and finding applications in various domains such as autonomous driving, security monitoring, and more [6]. Recent advancements in deep learning algorithms have significantly enhanced object detection capabilities, offering improved accuracy, speed, non-destructiveness, and real-time functionality, making it an appealing solution.

Research on You Only Look Once (YOLO) models has consistently demonstrated superior object recognition speed and performance compared to other detection models like DPM and R-CNN [7]. The continuous development of YOLO models, with the latest iteration being YOLOv7, has further accelerated training and improved object detection capabilities. While various studies have applied YOLOv7 in agricultural contexts, there is a notable gap in research utilizing this model to assess the quality of curly red chili. This study seeks to fill this gap by employing YOLOv7 to detect and classify chili pepper quality through a camera interface. The primary objective is to enhance the efficiency and accuracy of the sorting process, benefiting farmers, local collectors, and market traders. The goal is to contribute indirectly to stabilizing chili prices in the market.

2. LITERATURE REVIEW

Computer vision has emerged as a transformative and crucial tool in food and agriculture, addressing real-world challenges through automated sorting, grading, classification, and object recognition [8]. This technology has efficiently replaced manual procedures, providing a robust, accurate, non-destructive, and cost-effective approach to analysis. Among the various methods employed in computer vision, deep learning, a subset of machine learning, has garnered significant attention [9]. Its ability to independently extract complex features from diverse data sources makes it a standout in the research landscape, particularly in image processing applications. Deep learning architectures for image categorization, predominantly based on convolutional neural networks [10], have gained prominence in the food and agriculture sectors, contributing to the advancement of computer vision applications.

Previous research in object detection within agriculture includes the study "Non-destructive thermal imaging for object detection via advanced deep learning for robotic inspection and harvesting of chili peppers" by Steven C. Hespeler, Hamidreza Nemati, and Ehsan Dehghan-Niri [11]. This study focuses on the challenges posed by environmental factors such as debris, pepper overlap, and changing lighting conditions. It compares the performance of two advanced deep learning algorithms, Mask-RCNN and YOLOv3, regarding object detection accuracy and computational efficiency using a chili pepper dataset for training. YOLOv3 stands out by achieving an exceptional mean average precision (mAP) value of 1.0 for overall training, demonstrating superior performance on the chili dataset compared to Mask-RCNN and offering faster computing speed.

Another notable research study is "Drone-based apple detection: Finding the depth of apples using YOLOv7 architecture with multi-head attention mechanism" by Praveen Kumar S and Naveen Kumar K [12]. This research underscores the application of artificial intelligence (AI) and YOLOv7-based apple recognition algorithms in agricultural drones for orchard management. Addressing challenges in accurately identifying apples due to occlusions and other variables, the study proposes using a deep learning model to correct mistakes in real-time drone field operations. Incorporating a multi-head attention mechanism enhances the accuracy of Apple recognition in challenging environments. The research concludes that using drones, YOLOv7 can effectively detect apples in various conditions.

In the context of the current research, which only conducts research on object detection, these studies will serve as valuable references for developing an object detection and image classification model utilising the YOLOv7 framework. This research aims to enhance the accuracy and effectiveness of image classification for curly chilli peppers.

2.1 YOLOv7

The YOLOv7, developed by Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao [13], represents the latest iteration of the YOLO framework. This model surpasses all state-of-the-art real-time object detectors in speed and accuracy, operating within the impressive range of 5 FPS to 120 FPS. The YOLOv7 architecture amalgamates the optimal features of the YOLO framework, culminating in a sophisticated object detection model [14].

The model initiates the process by utilizing a backbone network to analyze an input image and extract information at various scales. An optional neck component is incorporated to enhance contextual information further. Subsequently, detection heads estimate bounding boxes, class probabilities, and object-specific properties using feature maps from the neck or backbone. The final output is derived through post-processing techniques, such as non-maximum suppression, to eliminate redundant detections. The result retains the most confident bounding boxes, corresponding class labels, and confidence scores. YOLOv7 integrates core YOLO principles with thoughtful design choices, delivering precise and rapid object identification [15].

2.2 mAP (Mean Average Precision)

The algorithm performance assessment holds paramount importance in computer vision and object identification, with mean average precision (mAP) emerging as a widely employed metric [16]. Nevertheless, the inefficiency and challenging integration of existing mAP computation approaches into training regimens have faced longstanding criticism. This challenge hinders the seamless evaluation of model progress after each training session. Recognizing the imperative to address these constraints, this study acknowledges the need for a novel mAP computing method that aligns with the requirements of contemporary machine learning processes, facilitating parallel execution.

$$\mathrm{mAP@t} = \frac{1}{k}\sum_{0}^{k-1} \mathrm{AP@t} \quad (1)$$

Equation 1 presents the mAP at a specific intersection over union (IoU) threshold 't'. Figure 1 illustrates constructing a precision-recall (PR) curve, essential for computing each class's average precision (AP). To ensure the resulting PR curve is independent of the order in which the detector analyzes images, the detected bounding boxes (DTBBs) are initially sorted by their confidence ratings in decreasing order [17].

$$q_i = \frac{1}{z}\sum_{0}^{i} 1_{\mathrm{TP}} \quad (2)$$

Equation 2 can calculate recall at a specific position on the PR curve. In this formula, 'qi' represents the recall value, and 'z' denotes the total number of "easy" ground truth bounding boxes (GTBBs) for the class in question. It is important to note that "difficult" GTBBs—such as occluded or severely truncated—do not incur consequences if they are missed during assessment. Conversely, precision assesses the accuracy of a prediction.

$$p_i = \sum_{0}^{i} \frac{1_{\mathrm{TP}}}{1_{\mathrm{TP}} + 1_{\mathrm{FP}}} \quad (3)$$

Equation 3 utilizes the value of 'pi' to represent the precision value, describing precision at a specific point on the PR curve. Notably, there might be more DTBBs than points on the PR curve since some DTBBs may match "difficult" GTBBs or fail to meet the IoU threshold requirements. When DTBBs match GTBBs that higher-confidence DTBBs have already matched, they are also counted as false positives (FPs) if they do not meet the IoU requirement. Furthermore, DTBBs that match "difficult" GTBBs are disregarded, regardless of whether others have matched these GTBBs. DTBBs are considered true positives (TP) if they do not fall into the FP or ignored category. While some benchmarks use multiple IoU thresholds and calculate an average mAP, this research follows the Pascal VOC technique. It emphasizes that the procedure can be extended to support multiple IoU thresholds while employing a single threshold for clarity. Overall, this technique provides a comprehensive and consistent method for determining mAP, facilitating the thorough evaluation of object detection algorithms [18].

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

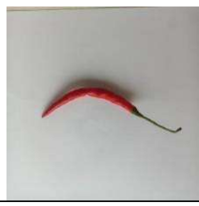
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3. RESEARCH METHODS

3.1 Data Gathering

In this study, the dataset focuses on curly red chili as the primary object of interest. The dataset comprises 200 sample images distributed across five distinct categories, each containing 40 images. These sample images were captured using a Xiaomi Redmi Note 8 Pro camera with an 8 MP resolution. Data collection took place during the afternoon and evening hours (from 13:00 to 15:00) with natural sunlight, and white paper was utilized as the background for the sample images. In addition, the researchers collected a dataset comprising 500 mixed images, each featuring chilies with various grades. This diverse dataset was crucial for enabling the model to learn and adapt to the information present in the image data across different conditions. The research uses five variables denoted as grades A, B, C, D, and E, as detailed in Table 2.

Table 2: Dataset Variable

Image	Variable	Description
	Grade A	Grade A curly red chili peppers, or red curly chili peppers grade A, are fully matured chilies with outstanding quality and exceptional flavor. These chilies are vibrant and fresh, boasting a bright red color.
	Grade B	Grade B curly red chili peppers, or red curly chili peppers grade B, are chilies in the final stage before reaching full maturity. These chilies are predominantly orange or orangish-green, with the orange hue being more dominant than the green in the peppers.
	Grade C	Grade C curly red chili peppers, or red curly chili peppers grade C, are chilies during ripening. These peppers exhibit a green color intermingled with orange hues, with the green color predominating in the peppers.
	Grade D	Grade D curly red chili peppers, or red curly chili peppers grade D, are chilies that are not yet fully mature and display a green color.
	Grade E	Grade E curly red chili peppers, or red curly chili peppers grade E, have deteriorated and exhibit a dark red color and a shriveled appearance compared to other peppers.

3.2 Data Pre – Processing

After the dataset is created, the researchers will move on to the next step, processing the data. In this phase, researchers will do several steps, there are:

a. Augmentation: Processing the initial 700 images commenced with cropping using the

Jupiter Python program. Subsequently, the images were renamed, and their size was reduced to a resolution of 640 × 640 pixels. The pixel values in each chili image represent the red, green, and blue color channels. The author then prepared the dataset for deep learning processing.

b. Annotation: The annotation process will employ the Roboflow website to annotate the sample dataset. 700 images were divided into multiple frames, and annotations were conducted by labeling the chili peppers in each frame using bounding boxes. These bounding boxes were drawn around the chilies, and depth labels were assigned to indicate the relative distance from the camera. The annotations offer crucial ground truth data for training and assessing the model's accuracy.

c. Dataset Split: The entire dataset is categorized into three folders: training, testing, and validation. The training set comprises 80% of all the sample images to evaluate the capability of the trained model. The validation and testing set consists of 10% of all the sample images.

Figure 1: Dataset Sample

3.3 Model Configuration and Training Details

This section provides a detailed description of the YOLOv7 (You Only Look Once version 7) architecture, a well-regarded object detection model known for its real-time capabilities and high accuracy, forming our study's foundation. The YOLOv7 model's configuration includes several crucial parameters, with 'nc' (Number of Classes) being a defining factor, specifying the number of distinct object categories the model can effectively detect. In our research, 'nc' is thoughtfully set to 5 to align with the specific object classes of interest. Furthermore, the 'depth multiple' and 'width multiple' parameters play a pivotal role in customizing the model's depth (the number of layers) and width (the number of channels). In our study, we intentionally maintain these parameters at 1.0, adhering to the default architectural configuration without any alterations to depth or width. Another critical feature of YOLOv7 is the utilization of anchor boxes, which provide the model with prior knowledge about the sizes and shapes of objects to be detected. In our research, we configure anchor boxes as follows:

- P3/8 Anchors: [12, 16, 19, 36, 40, 28]
- P4/16 Anchors: [36, 75, 76, 55, 72, 146]
- P5/32 Anchors: [142, 110, 192, 243, 459, 401]

The backbone of the YOLOv7 model is tasked with feature extraction from input images. This critical component comprises a sequence of layers and operations, including convolutional layers, max-pooling, and concatenation operations organized into blocks designed to capture features at various scales. The architecture of the backbone is thoughtfully designed to extract features at the P3, P4, and P5 scales, enabling effective object detection across a range of object sizes.

The head of the YOLOv7 model plays a pivotal role in the final stages of object detection and classification. It begins with Spatial Pyramid Pooling (SPP) and Concatenated Spatial Pyramid Pooling (CSP) operations, which are crucial for capturing contextual information. The head also integrates convolutional layers and upsampling operations to align feature map resolutions. Feature maps from different scales are concatenated, and additional convolutional layers are applied to make predictions regarding object detection. These predictions include essential information such as bounding box coordinates, object scores, and class probabilities for the detected objects. The YOLOv7 model is meticulously architected to predict objects across diverse scales and consolidate these results into the final detection output.

In model development, researchers rely on hyperparameters to enhance accuracy. These hyperparameters include 'epochs,' 'batch size,' and 'learning rate.' For our research, we set the number of epochs at 65, indicating that the training process iterates 65 times across all datasets, with a batch size 16.

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

3.4 Testing and Evaluation

The research will utilize the PyTorch package and Google Colab to test our study model. Datasets, including training, testing, and validation data, will be integral to the evaluation process. An extensive assessment of the model's performance will be carried out to verify its reliability and accuracy. As explained in the previous chapter, the resulting performance will be evaluated using several metrics such as Precision-Recall (mAP), precision, and recall. Additionally, the researcher will use the F1-score metric to measure the model's performance. The F1-score, an alternative machine learning assessment statistic, elaborates on a model's performance within a class instead of evaluating the model's overall performance based on accuracy [19]. The F1 score is widely used in recent research since it combines two conflicting metrics: a model's precision and recall scores. The F1-score is calculated using Equation 4, where 'P' is the precision score, and 'R' is the recall score.

$$F1 - Score = \frac{2 \times P \times R}{P + R} \tag{4}$$

After obtaining all the metric values, researchers will visualize the YOLOv7 model in two ways: by inputting random images into the model and using a webcam to detect objects in real time.

4. RESULTS AND DISCUSSION

The models are trained on the prepared dataset using the previously provided explanation of research methodologies. The conventional architectures of the YOLOv7 models are employed to extract features. Various categorization techniques of the YOLOv7 are compared based on the architecture. Hyperparameters such as the number of epochs, batch size, and learning rate are adjusted to maximize model accuracy and evaluate the performance of each model.

The results of the YOLOv7 model experiments indicate that the Mean Average Precision (AP) is 0.996 for grade A, 0.947 for grade B, 0.951 for grade C, 0.996 for grade D, and 0.996 for grade E. The overall Mean Average Precision (mAP) is recorded as 0.977. Figure 2 illustrates the relationship between precision along the Y-axis and recall along the X-axis.

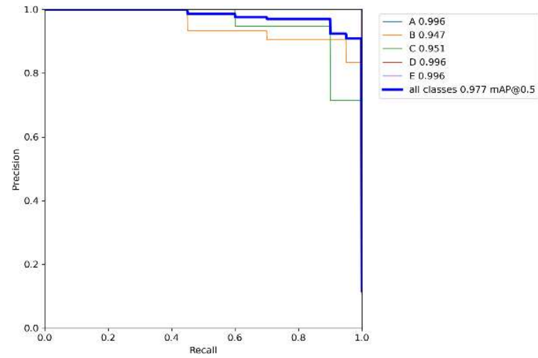


Figure 2: mAP of YOLOv7 Model

In Figure 3, the recall scores are displayed. This graph illustrates how successfully a YOLOv7 model can identify all genuine positive samples in the dataset for a given confidence level. The model can identify almost 100% of all positive samples at a confidence level of 1.00 without producing false optimistic predictions, as indicated by the "all classes" line with a confidence level of 1.00 and a recall of 0.000.

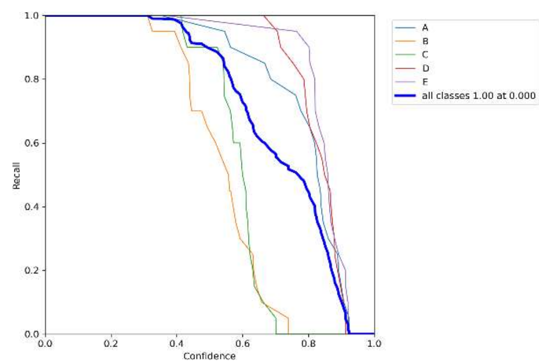


Figure 3: Recall of YOLOv7 Model

In Figure 4, the precision scores are displayed. This graph illustrates the accuracy with which a model predicts positive samples at a given confidence level. The "all classes" line at a confidence level of 0.664 and a precision of 1.00 indicates that the model achieves perfect precision (1.00) at a confidence level of 0.664, suggesting that all its optimistic predictions are accurate. However, the model only generates optimistic predictions when highly confident, which may cause it to miss some accurate positive samples.

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

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

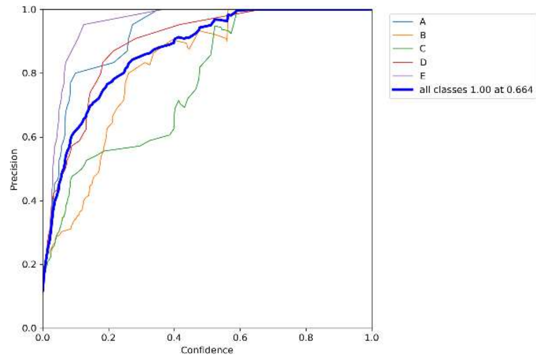


Figure 4: Precision of YOLOv7 Model

In Figure 5, the F1-score is displayed, indicating a high overall accuracy level with a value of 0.94. This score signifies the model's strong performance in correctly classifying data. The reference to "all classes" implies that this F1-score accounts for the model's performance across all distinct classes or categories it aims to predict, providing a comprehensive evaluation of its effectiveness in handling diverse data classifications.

However, "0.412" following the F1-score figure requires further clarification, as it might denote a specific threshold or decision boundary utilized in the classification process. The precise interpretation of this value depends on the context and purpose of the analysis. An F1-score of 0.94 for all classes underscores the model's robust performance in classification tasks across multiple categories. While "0.412" suggests the involvement of a specific decision threshold, its specific meaning necessitates a more detailed explanation based on the specific research or analysis in question.

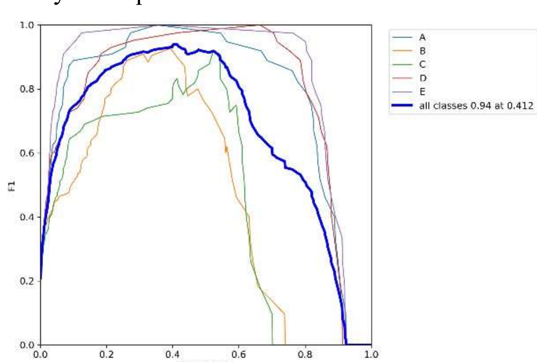


Figure 5: F1-Score of YOLOv7 Model

In conclusion, the YOLOv7 model shows promise for identifying and categorizing red curly chili, benefiting farmers and supply chain participants. However, the accuracy of these assumptions awaits confirmation through concrete experimental data. The researcher will assess the outcomes of model trials generated by testing their ability to recognize and distinguish items through random photo insertion and real-time webcam identification (Figure 6 and Figure 7).

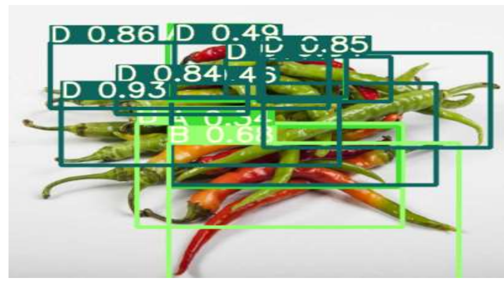


Figure 6: Testing Result Model YOLOv7 with Image

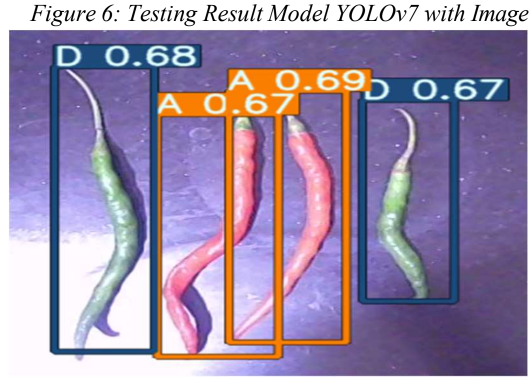


Figure 7: Testing Result Model YOLOv7 with Webcam

The images presented in Figure 6 and Figure 7 showcase the YOLOv7 model's effective identification and classification of items, demonstrating its proficiency in static images and real-time webcam recordings. Figure 6 illustrates the model's accurate recognition of all grade A chilies, assigning the correct class label "A", "B", and "D" to each chili. In Figure 7, the model adeptly identifies grade E chilies in the webcam recording, appropriately labeling them with the class label "A" and "D".

Upon critical assessment of the achieved answers, several strengths and limitations emerge. The model exhibits high accuracy in classifying different chili grades, particularly in real-time scenarios, emphasizing its practical applicability in sorting and classifying chilies within the supply chain. However, the evaluation's focus on specific chili grades (A to E) raises concerns about the model's generalizability to a broader range of chili types. Additionally, the observed high-confidence threshold in precision graphs suggests a conservative approach that might overlook some accurate positive samples, necessitating careful consideration for different applications.

Open issues for further exploration include diversifying the dataset to encompass a broader range of chili types, conditions, and environmental factors to enhance the model's generalization capabilities. Exploring optimization strategies for dynamically adjusting the model's confidence threshold based on specific application requirements is crucial for improving its adaptability to different use cases.

5. CONCLUSION

In conclusion, this study aimed to develop an effective model for detecting and classifying the quality of curly red chili, providing valuable support to supply chain stakeholders, particularly market traders, in addressing challenges related to automatic sorting, grading, and object recognition. The model was trained for 65 epochs with a batch size of 16 through a comprehensive series of tests to identify optimal hyperparameters. The results demonstrated a commendable overall Mean Average Precision (mAP) of 0.977, showcasing the model's robust performance. Specific average precisions (AP) for different chili grades were also noteworthy, with scores of 0.996 for grade A, 0.947 for grade B, 0.951 for grade C, 0.996 for grade D, and 0.996 for grade E.

While applying these hyperparameters consistently yielded positive outcomes, it is essential to critically analyze the paper's objectives and achievements and acknowledge its limitations. The study has made significant strides in developing an object identification model for curly red chili, contributing to the automation of quality assessment in the supply chain. However, the focus on specific chili grades raises questions about the model's adaptability to a broader range of chili types, and the conservative approach observed in precision graphs may impact the model's sensitivity to accurate positive samples. Future research endeavors could explore diversifying the dataset and optimizing the model's confidence threshold to enhance its applicability across various scenarios.

REFERENCES:

[1] Viva Budy Kusnandar, “Konsumsi Cabai Merah Meningkat 9,94% pada 2021,” databoks.
[2] “Badan Pusat Statistik.” Accessed: May 24, 2023.[Online].Available: https://www.bps.go.id/indicator/55/62/1/produksi-tanaman-buah-buahan.html
[3] “RI Borong Cabai China & Malaysia, Impor Pangan Auto Naik!” Accessed: May 15, 2023. [Online].Available: https://www.cnbcindonesia.com/news/20230216095312-4-414242/ri-borong-cabai- china-malaysia-impor-pangan-auto-naik
[4] Richard Salim and Ahmad Nurul Fajar, "Image Classification of The Fertility Level of Chili Using Convolutional Neural Network," *ICIMTech 2023 - International Conference on Information Management and Technology.*, 2023.
[5] MUHAMMAD RAFLI ARFIAN, “ANALISIS KINERJA RANTAI PASOK CABAI RAWIT DI KECAMATAN DUKUN KABUPATEN MAGELANG,” UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN,” Yogyakarta, 2022.
[6] L. Jiao *et al.*, "A Survey of Deep Learning-Based Object Detection," *IEEE Access*, vol. 7, pp. 128837–128868, 2019, [Online].Available: https://api.semanticscholar.org/CorpusID:198147317
[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", Accessed: Nov. 05, 2023. [Online]. Available: http://pjreddie.com/yolo/
[8] N. Begum and M. K. Hazarika, "Maturity detection of tomatoes using transfer learning," *Measurement: Food*, vol. 7, p. 100038,Sep.2022,doi:10.1016/J.MEAFOO.2022.100038.
[9] K. K. Chakraborty, R. Mukherjee, C. Chakroborty, and K. Bora, "Automated recognition of optical image based potato leaf blight diseases using deep learning," *Physiol Mol Plant Pathol*, vol. 117, p. 101781,Jan.2022,doi: 10.1016/J.PMPP.2021.101781.
[10] M. Kang and S. Elbel, "Novel regenerator design for caloric cycles using artificial neural network — Genetic algorithm method and additive manufacturing," *Energy Reports*, vol. 9, pp. 4257–4274, Dec. 2023, doi: 10.1016/J.EGYR.2023.03.031.
[11] S. C. Hespeler, H. Nemati, and E. Dehghan-Niri, "Non-destructive thermal imaging for object detection via advanced deep learning for robotic inspection and harvesting of chili peppers," *Artificial Intelligence in*

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

Agriculture, vol. 5, pp. 102–117, Jan. 2021, doi: 10.1016/J.AIIA.2021.05.003.

[12] P. K. S and N. K. K, "Drone-based apple detection: Finding the depth of apples using YOLOv7 architecture with multi-head attention mechanism," *Smart Agricultural Technology*, vol. 5, p. 100311, Oct. 2023, doi: 10.1016/J.ATECH.2023.100311.

[13] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real time object detectors".

[14] I. Saira Gillani *et al.*, "YOLOV5, YOLO-X, YOLO-R, YOLOV7 PERFORMANCE COMPARISON: A SURVEY," pp. 17–28, 2022, doi: 10.5121/csit.2022.121602.

[15] H. Shi, W. Yang, D. Chen, and M. Wang, "CPA-YOLOv7: Contextual and pyramid attention-based improvement of YOLOv7 for drones scene target detection," *J Vis Commun Image Represent*, vol. 97, p.103965,Dec.2023,doi:10.1016/J.JVCIR.2023.103965.

[16] S. Aliyi, K. Dese, and H. Raj, "Detection of gastrointestinal tract disorders using deep learning methods from colonoscopy images and videos," *Sci Afr*, vol. 20, p. e01628, Jul. 2023, doi: 10.1016/J.SCIAF.2023.E01628.

[17] B.Wang, "A PARALLEL IMPLEMENTATION OF COMPUTING MEAN AVERAGE PRECISION",Accessed: Nov. 05, 2023. [Online]. Available: https://github.com/bwangca/fast- map.

[18] A. Thakuria and C. Erkinbaev, "Improving the network architecture of YOLOv7 to achieve real-time grading of canola based on kernel health," *Smart Agricultural Technology*, vol. 5, p. 100300, Oct. 2023, doi: 10.1016/J.ATECH.2023.100300.

[19] Z. DeVries *et al.*, "Using a national surgical database to predict complications following posterior lumbar surgery and comparing the area under the curve and F1-score for the assessment of prognostic capability," *The Spine Journal*, vol. 21, no. 7, pp. 1135–1142,Jul.2021, doi: 10.1016/J.SPINEE.2021.02.007.