Intelligent Fruit Inspection System: Developing a YOLO-based Model for Identifying Defects on Plums Surface

Sain Rigobert Che
African Institute for Mathematical Sciences Cameroon
P.O. Box 608 Limbe, Southwest Region
Cameroon, Limbe
sain.che@aims-cameroon.org

Arnaud Nguembang Fadja*
Department of Engineering, University of Ferrara
Via Savonarola, 944121 Ferrara FE Italy
arnaud.nguembafadja@unife.it

Abstract

Agriculture is the backbone of Africa's economy, with over 60% of the labor force of over 1.2 billion people largely depending on it for their livelihoods. However, the gap between agriculture and technology in Africa continues to widen, creating a need for innovative solutions to improve productivity and access to high-quality agricultural products. One such problem is manually assessing the quality of agricultural products, which is especially time-consuming and tedious when done on a large scale. To address this challenge, we developed an artificial intelligence solution using YOLOv5 and YOLOv8 algorithms to assess the quality of African pears. We collected, from three regions in Cameroon, a dataset of 2892 damaged and good African pear surfaces. Our YOLOv5 and YOLOv8 models achieved mean average precision scores of 88.2% and 89.9% respectively. The proposed YOLOv8 solution has been deployed and runs on a web application. To the best of our knowledge, this is the first intelligent system for inspecting African plum quality.

1 Introduction

The African plum, known as Dacryodes edulis (1), is a widely cultivated fruit tree in tropical Africa, cherished for its nutritious and edible fruit (2). Despite its importance as a source of income and nutrition for rural communities, post-harvest losses are significant due to inadequate handling and marketing. While recent research has focused on exploring its properties and assessing quality non-destructively (3), there is limited work on automating fruit quality assessment. This study presents an intelligent fruit inspection system for African plums, utilizing a dataset of over 2892 manually collected and annotated images from Cameroon. Computer vision (4; 5) and object detection (6; 7) have gained significant research interest for agricultural product quality assessment and defect detection. Convolutional neural networks (CNNs) have shown promise for detecting defects and grading the quality of fruits, vegetables, and grains (8; 9; 10). However, most previous research has focused on major crops in America, Europe, and Asia. There is limited investigation of real-world challenges like shape, size, color, and imaging variations when applying these models in African

*http://www.arnaud-fadja.it/

5th Deep Learning Indaba Conference (DLI 2023).

settings. To address this gap, we trained and evaluated YOLOv5 and YOLOv8, state-of-the-art object detection models to detect surface defects. The YOLOv8 version was deployed in a web application. Our key contributions include: • Collection and annotation of a novel dataset of African pear images • Investigation of two state-of-the-art object detection models - YOLOv5 and YOLOv8 • Training customized models from scratch and optimizing hyperparameters • Conducting comprehensive experiments to evaluate model accuracy • Deploying the best performing model in a web application for real-time defect detection A running instance is available at https://shorturl.at/hmrzF. The paper is organized as follows: Section 2 discusses related work on fruit defect detection. Section 3 presents the data collection. Section 4 describes YOLOv5 and YOLOv8 architectures. Section 5 presents experimental results and finally, Section 6 concludes and outlines future work.

2 Related Works

Computer vision (4) and object detection (6) have gained significant research interest for agricultural product quality assessment and defect detection in recent years. Convolutional neural networks (CNNs) have been widely adopted for these tasks due to their ability to learn discriminative visual features directly from images. Numerous studies have applied CNNs to detect defects and grade the quality of popular fruits such as apples (11; 8), strawberries (13; 14), mangoes (12; 9) and citrus fruits (10). For example, Bargoti and Underwood (8) trained a deep CNN on over 5000 apple images and achieved 94% accuracy in identifying four common defects. Li et al. (9) systematically compared popular CNN architectures like ResNet (15) and DenseNet (16) for three mango defects, finding ResNet-50 performed best with over 90% accuracy. While most prior work focused on major agricultural products from America, Europe and Asia, a few studies have explored African crops. For example, Lauguico et al. (17) detected grape diseases using Transfer Learning and CNNs, reporting up to 96% accuracy.. However, plums have received little attention despite their economic importance in many African countries. Additionally, few studies have developed full end-to-end vision systems with deployment capabilities. Overall, deep CNNs and advanced object detection models show promise for automating agricultural product quality inspection. Nevertheless, further research is still needed to address diverse tropical crops through effective computer vision solutions applicable across different geographical contexts and production chains in Africa. This study aims to fill these gaps by developing an intelligent vision system for plums, an important African fruit.

3 Data Collection

In this study, a dataset of African pear images was gathered from various regions in Cameroon. This dataset includes 2892 images of both good and defective African pears, captured using an Android phone. The images were collected from three distinct agro-ecological regions, each with its climate: Littoral (coastal tropical), North West (highland tropical), and North (Sudano-Sahelian). This diverse dataset captures variations in pear characteristics, including size, shape, color, and defects. Images were taken at two distinct orchards within each of the mentioned regions, spanning a three-month period during the peak harvesting season. To enhance the model's resilience, photos were captured against diverse backgrounds, including soil, white paper boards, and shed walls. Furthermore, the pears were photographed from various angles, adding depth and variety to the dataset. To address fluctuations in lighting conditions, images were taken at different times of the day, covering early morning, noon, afternoon, and dusk. This encompassed both shaded and direct sunlight scenarios. Additionally, the dataset intentionally included a variety of defective pears, such as bruised, cracked, rotten, spotted, and undamaged good pears. To ensure high-quality and high-resolution images, each pear was photographed multiple times, contributing to the overall dataset. The annotation process involved manual labeling of all images using the LabelImg tool within Roboflow. Two distinct classes were established: "good pears" and "bad pears" (representing defective pears). In order to enhance annotation consistency, a single individual was responsible for labeling the entire dataset. Subsequently, a rigorous data cleaning procedure was executed, culminating in a final curated dataset comprising 2892 images.

4 Model Selection

We chose to employ YOLO (You Only Look Once) as our one-stage object detection model, considering two YOLO variants (19; 20): YOLOv5 and YOLOv8. Our aim was to evaluate their performance within our specific context. Both YOLOv5 and YOLOv8 share a common architecture, comprising a backbone feature extractor network, a neck, and a prediction head. The backbone is responsible for extracting spatial feature maps from the input image, while the neck enhances these features. The prediction head is responsible for estimating bounding box coordinates and class probabilities for each grid cell in the feature maps. The primary distinction between YOLOv5 and YOLOv8 lies in their backbones. YOLOv5 utilizes CSPDarknet (21) as its backbone, whereas YOLOv8 employs CSPResNet (22). In addition, YOLOv8's prediction head incorporates SAM (Spatial Attention Module) (23) with PAN (Path Aggregation Network) (24). Furthermore, YOLOv8's neck introduces a new FSA (Feature Selective Aggregation) module and employs relative encoding for bounding boxes. Both YOLOv5 and YOLOv8 models were optimized using the mean squared error (MSE) loss function during training. YOLOv5 is $L_{v5} = L_{coord} + L_{obj} + L_{cls}$ where $L_{coord} = \lambda_{coord} \sum[(x-x')^2 + (y-y')^2]$ is the localization loss, $L_{obj} = \sum[(C - P_{obj})^2]$ the objectness Loss, $L_{cls} = \lambda_{cls} \sum[(one{-}hot(y) - P_{class})^2]$ the classification loss and $L_{box} = L_{obj} + L_{cls}$ the weighted binary cross-entropy loss. YOLOv8 is $L_{v8} = L_{coord} + L_{obj} + L_{cls} + L_{center}$ where $L_{center} = \lambda_{center} \sum[(x - x')^2 + (y - y')^2]$ is the center loss and $L_{box} = L_{obj} + L_{cls} + L_{center}$ the weighted binary cross-entropy loss

Where λ_{coord}, λ_{cls}, and λ_{center} are coefficients to balance the losses; C is the objectness score; P_{obj} is the predicted objectness score; *one-hot*(y) is the ground truth class vector; P_{class} is the predicted class probabilities; (x, y) are predicted box center coordinates; (x', y') are ground truth box center coordinates.

5 Experimental Results

5.1 Data Preprocessing

The 2892 images were manually annotated using the Roboflow platform to identify good and defective regions on each pear 1. Online data augmentation was applied during training to increase diversity, including rotations, flips, zooms, and hue/saturation shifts. The data was split into 70% training, 20% validation, and 10% test sets in a stratified manner. Additionally, all images were resized to 640 x 640 pixels for uniformity.

(a) Bad fruit labeling

(b) Good fruit labeling

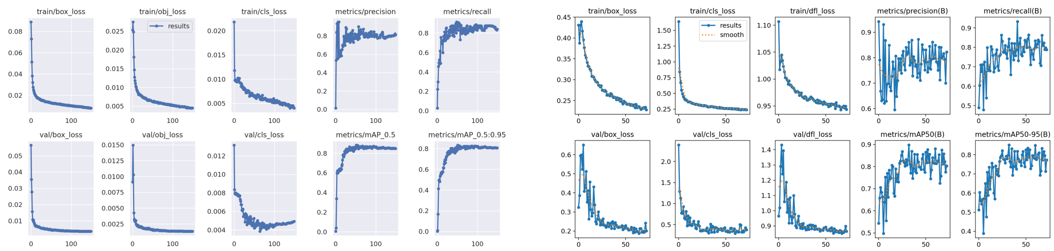
Figure 1: Image labeling.

5.2 Model Training

The Google Colab framework was utilized to train the YOLOv5 and YOLOv8 models from scratch using the African pear dataset. The essential training hyper-parameters are outlined in Table 1. Both models were configured with the Adam optimizer and a batch size of 16. We adhered to original input resolutions i.e 416x416 and 640x640 for YOLOv5 for YOLOv8 respectively. We set the learning rate at 0.001, which ensured stable training without issues like divergence. The learning rate was adjusted schedule to decay every 30 epochs. Training was performed and no pre-trained weights were used. The models were trained end-to-end to optimize the mAP loss, as shown in Figure 2.

Model	Input Resolution	Batch Size	Optimizer	Training Epochs
YOLOv5	416 x 416	16	Adam	150
YOLOv8	640 x 640	16	Adam	80

Table 1: Training details for the YOLOv5 and YOLOv8 models on the African pear dataset.



(a) YOLOv5 training (b) YOLOv8 training

Figure 2: Comparison of YOLOv5 and YOLOv8 training.

5.3 Results and discussion

The trained models were evaluated on the unseen test set of 293 images. The models demonstrate

Model	Precision (%)	Recall (%)	F1 score(%)	mAP (%)
YOLOv5	81.6	86.1	88.8	88.2
YOLOv8	86.0	84.3	85.1	89.8

Table 2: Performance metrics for the YOLOv5 and YOLOv8 models.

exceptional precision and recall values, affirming their capability to accurately detect the majority of defects on the plum surface while minimizing false detections. Notably, both models achieved a mean average precision exceeding 88%. This mAP metric evaluates accuracy across all classes, confirming their competence in distinguishing between defective and normal plum regions. YOLOv8, with a slightly higher mAP of 89.8%, suggests its improved architecture facilitates more robust feature extraction across the image. The strong mAP demonstrates high accuracy for both defective and good classes. Some classifications on the test set images are illustrated in Figure 3. YOLOv8 was deployed in a web application available at https://shorturl.at/cuQV9.

(a) v5 detection (b) v8 detection (c) v5 detection (d) v8 detection

Figure 3: YOLOv5 and YOLOv8: Bad and good plum detection on the test set images.

6 Conclusion

This work showcased the potential of machine learning for automating African plum quality inspection, using YOLOv5 and YOLOv8 models with high accuracy in detecting surface defects. YOLOv8 demonstrated 87% precision and 90% recall, indicating promise for efficient plum sorting and grading in domestic and export markets. Future work includes enhancing the models' optimization, speed, and user interfaces for seamless deployment in African farms and markets. Field tests and farmer evaluations will be conducted to validate the models' real-world performance and usability.

Acknowledgments and Disclosure of Funding

A special to Armel Fameni for performing data collection and to our families and friends.

References

[1] KK Ajibesin and others. *Dacryodes edulis (G. Don) HJ Lam: a review on its medicinal, phytochemical and economical properties. Research Journal of Medicinal Plant*, vol. 5, no. 1, pp. 32–41, 2011. Academic Journals Inc.

[2] Ayuk, E. T., Duguma, B., Franzel, S., Kengue, J., Mollet, M., Tiki-Manga, T., & Zekeng, P. (1999). Uses, management, and economic potential of Dacryodes edulis (Burseraceae) in the humid lowlands of Cameroon. *Economic Botany*, 292–301.

[3] Jagtap, S. T., Phasinam, K., Kassanuk, T., Jha, S. S., Ghosh, T., Thakar, C. M. (2022). Towards application of various machine learning techniques in agriculture. *Materials Today: Proceedings*, 51, 793–797.

[4] George Stockman and Linda G. Shapiro. *Computer Vision.* Prentice Hall PTR, 2001.

[5] Zhou, L., Zhang, L., Konz, N. (2022). Computer vision techniques in manufacturing. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(1), 105–117.

[6] Farhana Sultana, Abu Sufian, and Paramartha Dutta. *A Review of Object Detection Models Based on Convolutional Neural Network. Intelligent Computing: Image Processing Based Applications*, pages 1–16, 2020. Springer.

[7] Diwan, T., Anirudh, G., Tembhurne, J. V. (2023). Object detection using YOLO: Challenges, architectural successors, datasets and applications. *Multimedia Tools and Applications*, 82(6), 9243–9275.

[8] Suchet Bargoti and James Underwood. *Deep fruit detection in orchards*. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages=3626–3633, 2017, IEEE.

[9] J. Li, Y. Liu, Y. Zhang, Z. Zhang, and F. Cheng. *Comparison of CNN architectures for mango defect detection*. *IEEE Access*, 9:66282-66289, 2021, IEEE.

[10] Asad Khattak, Muhammad Usama Asghar, Ulfat Batool, Muhammad Zubair Asghar, Hayat Ullah, Mabrook Al-Rakhami, and Abdu Gumaei. *Automatic detection of citrus fruit and leaves diseases using deep neural network model*. *IEEE Access*, 9:112942–112954, 2021, IEEE.

[11] Nur Alam, MD, Ullah, I., & Al-Absi, A. A. (2021). Deep learning-based apple defect detection with residual squeezenet. In *Proceedings of International Conference on Smart Computing and Cyber Security: Strategic Foresight, Security Challenges and Innovation (SMARTCYBER 2020)* (pp. 127–134). Springer

[12] R. Nithya, B. Santhi, R. Manikandan, Masoumeh Rahimi, and Amir H. Gandomi. *Computer vision system for mango fruit defect detection using deep convolutional neural network. Foods*, 11(21):3483, 2022, MDPI.
.

[13] Ping Lin and Yongming Chen. *Detection of strawberry flowers in outdoor field by deep neural network*. In *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pages=482–486, 2018, IEEE.

[14] Islam, M. T., Tusher, A. N., Nanda, P., Verma, V. K., Srivastava, S., Gupta, R. K., Mazumdar, A. P. (2022). Automatic Detection of Grape, Potato, and Strawberry Leaf Diseases Using CNN and Image Processing. *Data Engineering for Smart Systems. Lecture Notes in Networks and Systems*, 238, Springer.

[15] Brett Koonce and Brett Koonce. *ResNet 50.* In *Convolutional Neural Networks with Swift for TensorFlow: Image Recognition and Dataset Categorization*, pages 63–72, 2021. Springer.

[16] Yi Zhu and Shawn Newsam. *Densenet for Dense Flow.* In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 790–794, 2017. IEEE.

[17] Sandy Lauguico, Ronnie Concepcion, Rogelio Ruzcko Tobias, Argel Bandala, Ryan Rhay Vicerra, and Elmer Dadios. *Grape leaf multi-disease detection with confidence value using transfer learning integrated to regions with convolutional neural networks*. In *2020 IEEE REGION 10 CONFERENCE (TENCON)*, pages 767–772, 2020.

[18] Sasha Targ, Diogo Almeida, and Kevin Lyman. *Resnet in Resnet: Generalizing Residual Architectures. arXiv preprint arXiv:1603.08029*, 2016.

[19] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. *You Only Look Once: Unified, Real-Time Object Detection.* In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788, 2016.

[20] Peiyuan Jiang, Daji Ergu, Fangyao Liu, Ying Cai, and Bo Ma. *A Review of YOLO Algorithm Developments. Procedia Computer Science*, vol. 199, pp. 1066–1073, 2022. Elsevier.

[21] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. *Yolov4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934*, 2020.

[22] Chien-Yao Wang, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, and I-Hau Yeh. *CSPNet: A New Backbone that Can Enhance Learning Capability of CNN.* In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 390–391, 2020.

[23] Bingling Chen, Yan Huang, Qiaoqiao Xia, and Qinglin Zhang. *Nonlocal Spatial Attention Module for Image Classification. International Journal of Advanced Robotic Systems*, vol. 17, no. 5, page 1729881420938927, 2020. SAGE Publications Sage UK: London, England.

[24] Jianjun Wu and Shaowen Liao. *Traffic Sign Detection Based on SSD Combined with Receptive Field Module and Path Aggregation Network. Computational Intelligence and Neuroscience*, vol. 2022, 2022. Hindawi.