

Article

Not peer-reviewed version

Intelligent Vision System for Identifying Defects on African Plum Surfaces

[Arnaud Nguembang Fadja](#)*, [Sain Rigobert Che](#), [Marcellin Atemkemg](#)

Posted Date: 14 August 2024

doi: 10.20944/preprints202408.0992.v1

Keywords: agriculture; artificial intelligence; object detection; african plums; YOLOv5; YOLOv8; YOLOv9; fast RCNN; mask RCNN; VGG-16; detectron-121; MobileNet; ResNet



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Intelligent Vision System for Identifying Defects on African Plum Surfaces

Arnaud Nguembang Fadja ^{1,*}, Sain Rigobert Che ^{2,†} and Marcellin Atemkeng ^{3,†}

¹ Department of Engineering, University of Ferrara, Italy

² African Institute for Mathematical Sciences Cameroon

³ Department of Mathematics, Rhodes University, PO Box 94, Makhanda, 6140, South Africa

* Correspondence: arnaud.nguembafadja@unife.it

† These authors contributed equally to this work.

Abstract: Agriculture stands as the cornerstone of Africa's economy, supporting over 60% of the continent's labor force. Despite its significance, the quality assessment of agricultural products remains a challenging task, particularly at large scale, consuming valuable time and resources. The African plum is an agricultural fruit that is widely consumed across West and Central Africa but remains underrepresented in AI research. In this paper, we collected a dataset of 2,892 African plum samples from fields in Cameroon representing the first dataset of its kind for training AI models. The dataset contains images of plums annotated with quality grades. We then trained and evaluated various state-of-the-art object detection and image classification models, including YOLOv5, YOLOv8, YOLOv9, Fast R-CNN, Mask R-CNN, VGG16, Detectron-121, MobileNet and ResNet, on this African plum dataset. Our experimentation resulted in mean average precision scores ranging from 88.2% to 89.9% and accuracies between 86% and 91% for the object detection models and the classification models respectively. We then performed model pruning to reduce model sizes while preserving performance, achieving up to 93.6% mean average precision and 99.09% accuracy after pruning YOLOv5, YOLOv8 and ResNet by 10-30%. We deployed the high-performing YOLOv8 system in a web application, offering an accessible AI-based quality assessment tool tailored for African plums. To the best of our knowledge, this represents the first such solution for assessing this underrepresented fruit, empowering farmers with efficient tools. Our approach integrates agriculture and AI to fill a key gap.

Keywords: agriculture; artificial intelligence; object detection; african plums; YOLOv5; YOLOv8; YOLOv9; fast RCNN; mask RCNN; VGG-16; detectron-121; MobileNet; ResNet

1. Introduction

Accurate assessment of fruit quality is essential for ensuring food security and optimizing agricultural production. In recent years, there has been a growing need for innovative solutions to address the challenges faced in fruit quality evaluation. Computer vision and artificial intelligence (AI) techniques are increasingly being explored for applications such as fruit detection and grading. Plums are an important fruit crop worldwide, with several varieties commonly grown across regions for their nutritional and economic value. AI and machine learning have been applied to tasks like plum detection, sorting, and quality assessment for plums like the European plum. However, one variety that remains understudied is the African plum.

The African plum, also known as *Dacryodes edulis* [1], is a significant crop grown by smallholders in Africa. As a biodiverse species, it contributes significantly to food security and rural livelihoods in over 20 countries. Commonly found in home gardens and smallholdings, African plum is estimated to support millions of farm households through food, nutrition and income generation [2]. It bears fruit year-round, providing a reliable staple high in vitamins C and A [3]. Leaves are also collected as food seasoning or fodder. Its multi-purpose uses enhance resilience for subsistence farmers. Culturally, African plum plays dietary and medicinal roles, with all plant parts consumed or utilized. It forms part of the region's cultural heritage and traditional ecological knowledge systems [4]. However, lack of improved production practices and limited access to markets have prevented scale-up of its commercial

potential [5]. Better quality assessment and grading methods specific to African plum could help address this challenge. Our study aims to develop such techniques using artificial intelligence as a means to support livelihoods dependent on this vital traditional crop.

This research aims to explore the application of machine learning [6,7] and computer vision algorithms [8,9] to address this issue and enable industrial-scale quality control. This study focuses on developing a computer vision system for contactless quality assessment of African plums, a widely cultivated crop in Africa.

In recent years, the field of computer vision, particularly Deep Learning (DL) algorithms, has emerged as a promising solution for fruit assessment [10,11]. Deep Convolutional Neural Networks (CNNs) have revolutionized image classification and identification [12], enabling accurate and reliable analysis of fruit quality. Recent research studies have demonstrated the effectiveness of object detection models in fruit assessment. For instance, a study published in 2023 utilized the YOLOv5 model to assess apple quality, achieving an accuracy of 90.6% [13]. Similarly, a research paper from 2023 showcased the successful application of the Mask RCNN model in identifying and localizing defects in citrus fruits, achieving an F1 score of 85.6% [14].

Pruning techniques have emerged as a valuable approach to further optimize the performance of object detection models in fruit assessment. Pruning involves selectively removing redundant or less informative parameters, connections, or structures from a trained model, leading to more efficient and computationally lightweight models. By applying pruning techniques, researchers have successfully enhanced the efficiency and effectiveness of fruit assessment models. For instance, in a study published in 2022, pruning techniques were employed to optimize a YOLOv5 model used for apple quality assessment, resulting in a more compact and computationally efficient model while maintaining high accuracy [15].

Recent research highlights the remarkable capabilities of object detection models, such as YOLO, Mask RCNN, Fast RCNN, VGG-16, Detectron-121, MobileNet, and ResNet, in accurately identifying and localizing external quality attributes of various fruits. The successful application of these models demonstrates their potential for automating fruit quality assessment and enhancing grading processes, contributing to overall improvements in fruit assessment and quality control practices.

This study focuses on developing models based on a range of architectures, including YOLOv5 [16], YOLOv8 [17], YOLOv9 [18], Fast RCNN [19], Mask RCNN [20], VGG-16 [21], DenseNet121 [22], MobileNet [23], and ResNet [24], specifically tailored for the external quality assessment of African plums.

To assess the quality of African plums, we collected a comprehensive dataset of over 2,800 labeled images, divided into training, validation, and testing sets. Various models, including YOLOv9, YOLOv8, YOLOv5, Fast RCNN, Mask RCNN, VGG-16, Detectron-121, MobileNet, and ResNet, were trained, fine-tuned, and validated. Among these, YOLOv8 demonstrated the highest accuracy in detecting surface defects. We integrated YOLOv8 into a prototype inspection system to evaluate its effectiveness in contactless quality sorting at an industrial scale. This integration aims to enhance the efficiency of sorting processes, improving productivity and quality assurance in the agricultural industry. The main contributions of this work are:

- Developed models based on YOLOv9, YOLOv8, YOLOv5, Fast RCNN, Mask RCNN, VGG-16, Detectron-121, MobileNet, and ResNet for African plum quality assessment.
- Implemented pruning techniques to optimize YOLOv9, YOLOv8, YOLOv5, MobileNet, and ResNet models, resulting in more efficient, computationally lightweight models.
- Collected a new labeled dataset of over 2,800 African plum samples, the first of its kind for this fruit crop.
- Deployed the best-performing model (YOLOv8) in a web application for real-time defect detection, with an instance available at <https://shorturl.at/hmrzF>.

Such data-driven solutions could enhance African agriculture by addressing production challenges for underutilized native crops.

2. Related Works on Plum

The application of computer vision and deep learning for agricultural product quality assessment and defect detection has gained significant attention in recent years. Several studies have explored the utilization of convolutional neural networks (CNNs) for detecting defects and classifying the quality grades of various fruits, vegetables, and grains.

In the context of fruit defect detection, CNNs have been employed to detect defects on apples [25], oranges [26], strawberries [27], and mangoes [28], among other fruits. For instance, Khan et al. [29] developed a deep learning-based apple disease detection system. They constructed a dataset of 9000 high-quality apple leaf images covering various diseases. The system uses a two-stage approach, with a lightweight classification model followed by disease detection and localization. The results showed promising classification accuracy of 88% and a maximum average precision of 42%. Similarly, Faster R-CNN models have been employed for accurate defect detection in citrus fruits, achieving comparable accuracies [30]. Additionally, Kusurini et al. [31] compared deep learning models for mango pest and disease recognition using a labeled tropical mango dataset. VGG16 achieved the highest accuracy at 89% in validation and 90% in testing, with a testing time of 2 seconds for 130 images.

The application of object detection models such as Faster R-CNN, SSD, YOLO, Detectron, and Mask R-CNN has also been prevalent in agricultural applications [32]. Notably, YOLO models trained on mango images have demonstrated high accuracy in detecting anthracnose disease [33]. YOLOv3 has shown superior performance compared to other models in detecting apple leaf diseases [29]. Recent studies have also utilized Detectron-121 and VGG-16 models for fruit defect detection and quality assessment [34], further highlighting their effectiveness in this domain. Additionally, MobileNet and ResNet architectures have been extensively studied and applied in various computer vision tasks. The MobileNet architecture, introduced by Szegedy et al. [35], utilizes inception modules to achieve high accuracy while maintaining computational efficiency.

While these studies have made significant contributions to fruit quality assessment and defect detection, the incorporation of pruning techniques to optimize model performance and streamline computational complexity has been gaining attention. In our work, we extended the existing research by applying pruning techniques to five models: YOLOv9, YOLOv8, YOLOv5, MobileNet, and ResNet. By selectively removing redundant parameters and connections, pruning helped improve the efficiency and speed of these models without compromising their accuracy.

In summary, deep CNNs and object detection models have demonstrated remarkable capabilities in assessing agricultural product quality and detecting defects. Our work contributes to this field by targeting the African plum and demonstrating the feasibility of an intelligent vision system for automating post-harvest processing. The comprehensive evaluation of our pruned models on the African pear dataset showcased improved efficiency without sacrificing accuracy. However, further research is needed to address real-world complexities, such as variations in shape, size, color, and imaging conditions, when deploying these models practically in African settings. The inclusion of Detectron-121 and VGG16 models in our study further expands the range of models used for fruit defect detection, enhancing the comprehensiveness of our research in this area.

3. Data Collection and Dataset Description

The acquisition of a robust and comprehensive dataset is essential for the development of an effective deep learning model. In this section, we present the details of our data collection process and describe the African pear dataset collected from major pear-growing regions in Cameroon.

We utilized an android phone to capture a total of 2892 images, encompassing both good and defective African pears. Our data collection strategy involved acquiring images from three distinct agro-ecological regions: Littoral (coastal tropical climate), North West (highland tropical climate), and North (Sudano-Sahelian climate). By capturing images across diverse regions, we ensured the inclusion of variations in pear size, shape, color, and defects, making our dataset representative of real-world scenarios.

The image capture process took place at two different orchards within each region over a three-month period, coinciding with the peak harvesting season. To enhance the robustness of our dataset, we took images against varying backgrounds, such as soil, white paper boards, shed walls, etc. This approach aimed to expose the deep learning model to different visual contexts, ultimately improving its performance. Furthermore, We captured images of the pears from multiple angles to portray their comprehensive appearance.

Considering the impact of lighting conditions on image quality, we acquired images at different times of the day, including early morning, noon, afternoon, and dusk. We specifically accounted for both shade and direct sunlight conditions during image capture. To ensure the dataset covered a wide range of defects, we included various types of defective pears, such as bruised, cracked, rotten, spotted, and unaffected good pears. Each pear was imaged multiple times, resulting in high-resolution images that provide detailed information for analysis.

To facilitate subsequent analysis and model training, all images were annotated using the Labeling tool in Roboflow. Our annotation process involved marking bounding boxes around each pear in the images. We defined three distinct classes: good pears, bad pears (defective pears), and a background class that indicates the absence of fruit in the image. To maintain annotation consistency, a single annotator was responsible for labeling the entire dataset. Additionally, another annotator validated the efficiency of the annotations. Also, extensive data cleaning procedures were employed to ensure data quality and integrity. As a result, we obtained a final curated dataset consisting of 2892 annotated images.

Figure 1 showcases sample images depicting plum fruits on the fruit tree, providing a visual representation of the African pear.



Figure 1. Sample images showcasing plum fruits on the fruit tree [36].

Our diverse and comprehensive African pear dataset captures the inherent complexities of the real-world agricultural setting. The dataset incorporates variations in growing conditions, pear quality, and imaging environments, making it an invaluable resource for training robust deep learning models for automated defective pear detection. The availability of such a dataset will contribute to advancements in agricultural technology and pave the way for more efficient and accurate fruit assessment processes.

4. Model Architecture

In this section, we present an evaluation of different state-of-the-art convolutional neural network (CNN) architectures for classification and object detection. The objective is to determine the most suitable approach for our specific application of object detection in the context of the African pear dataset.

4.1. Model and Technique Descriptions

We consider the following eight CNN architectures for evaluation:

1. You Only Look Once (YOLO): YOLO frames object detection as a single-stage regression problem, directly predicting bounding boxes and class probabilities in one pass [37]. We experiment with YOLOv5, YOLOv8, and YOLOv9 which build upon smaller, more efficient backbone networks like CSPDarknet53 compared to earlier YOLO variants. These models divide the image into a grid and associate each grid cell with bounding box data. YOLOv8 and YOLOv9 improve accuracy through an optimized neck architecture that enhances the flow of contextual information between the backbone and prediction heads [17].

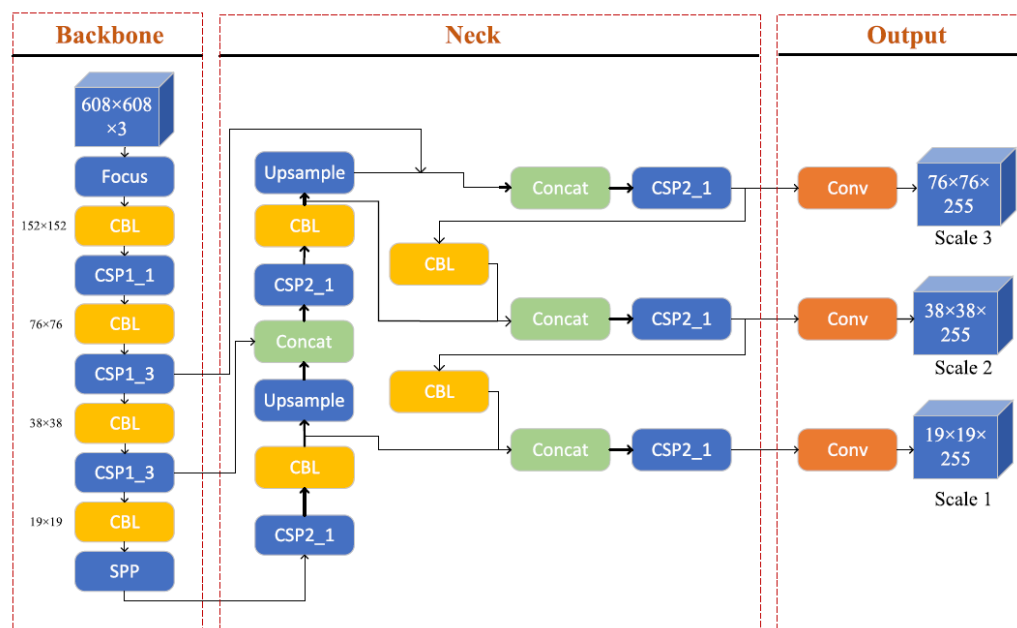


Figure 2. The YOLOv5 is structured into three primary segments: the backbone, neck, and output [41].

2. Faster R-CNN: Faster R-CNN is a two-stage detector that utilizes a Region Proposal Network (RPN) to propose regions of interest (RoIs), followed by classification and refinement of the detected objects in each RoI [38]. It employs a Region-of-Interest Pooling (RoIPool) layer to extract fixed-sized feature maps from the backbone network's feature maps for each candidate box.
3. Mask R-CNN: Building on Faster R-CNN, Mask R-CNN introduces a parallel branch for predicting segmentation masks on each RoI, in addition to bounding boxes and class probabilities [20]. It utilizes a mask prediction branch with a Fully Convolutional Network (FCN) to predict

a binary mask for each RoI. This per-pixel segmentation ability enables instance segmentation tasks alongside object detection.

4. Detectron-121: Detectron-121 is a powerful object detection model based on the Mask R-CNN architecture [39]. It has achieved state-of-the-art performance on various benchmarks and datasets, showcasing its effectiveness in complex object detection scenarios.
5. VGG16: VGG16 is a widely adopted CNN architecture that has shown strong performance in object detection tasks [21]. Its deep network structure and large receptive field contribute to its ability to capture and represent complex visual patterns.
6. MobileNet: MobileNet, also known as Inception-v1, is another popular CNN architecture introduced by Szegedy et al. [23]. It employs the concept of inception modules, which are designed to capture multi-scale features by using filters of different sizes within the same layer. GoogleNet's architecture enables efficient training and inference with a reduced number of parameters.
7. ResNet: ResNet, proposed by He et al. [24], addresses the degradation problem in deep neural networks by introducing residual connections. These skip connections allow the gradients to flow more easily during training, enabling the training of very deep networks. ResNet has achieved state-of-the-art performance in various computer vision tasks, including image classification and object detection.

4.2. Key Features

Each of the evaluated architectures brings unique features and innovations that contribute to their overall performance and effectiveness:

- YOLO: YOLO models provide real-time object detection capabilities due to their single-stage regression approach and optimized architecture.
- Faster R-CNN: The two-stage design of Faster R-CNN, with the RPN and RoIPool layer, enables accurate localization and classification of objects in images.
- Mask R-CNN: In addition to bounding boxes and class probabilities, Mask R-CNN introduces per-pixel segmentation to enable instance-level object detection and segmentation.
- Detectron-121: Based on the Mask R-CNN architecture, Detectron-121 is known for achieving state-of-the-art performance in classification tasks, making it a strong contender for our evaluation.
- VGG16: With its deep network structure and large receptive field, VGG16 has demonstrated strong performance in previous image classification studies.
- MobileNet: MobileNet's inception modules allow it to capture multi-scale features efficiently, leading to good performance with fewer parameters.
- ResNet: ResNet's residual connections address the degradation problem in deep networks, enabling the training of very deep architectures and achieving state-of-the-art performance in various computer vision tasks.

4.3. Supporting Evidence

We support our evaluation of these architectures by referring to relevant papers that describe their architectures and demonstrate their performance in object detection tasks. Please refer to the following citations for more detailed information: YOLO: [17,37], Faster R-CNN: [38], Mask R-CNN: [20], Detectron-121: [39], VGG16: [21], GoogleNet: [23], ResNet: [24]

4.4. Framework and Dataset

To facilitate the evaluation process, we leverage the Roboflow framework [40] for data labeling, augmentation, and model deployment. Roboflow is a comprehensive computer vision platform that streamlines the development lifecycle from data preparation to model deployment and monitoring; it provides tools for dataset creation, annotation, augmentation, and model training using popular frameworks, allowing users to package and deploy models as APIs or embedded solutions, with

hosted deployment, inference, and performance tracking capabilities, as well as support for team collaboration and versioning, enabling efficient development of computer vision applications.

4.5. Application Relevance

Comparing these model architectures is crucial for our specific object detection task in the context of the African plum dataset. By evaluating the performance of these architectures, we aim to identify the most suitable approach that can accurately and efficiently detect objects in our dataset. This determination will help us make informed decisions regarding the choice of model for our application, potentially improving the efficiency and accuracy of object detection in African pear images. Additionally, understanding the strengths and weaknesses of each architecture will provide valuable insights for future research and development in the field of object detection.

5. Experimental Results and Analysis

This section presents the experimental results and analysis of the African pear defect detection system. It covers the data preprocessing steps, model training, evaluation results, and a detailed discussion.

Figure 3 provides an overview of the key steps in our implementation. The process starts with data collection, followed by data preprocessing. We then evaluated various object detection models, including YOLOv5, YOLOv8, YOLOv9, Fast R-CNN, Mask R-CNN, VGG16, Detectron-121, MobileNet, and ResNet, the models were then trained and validated, and finally evaluated on the test set to assess its performance.

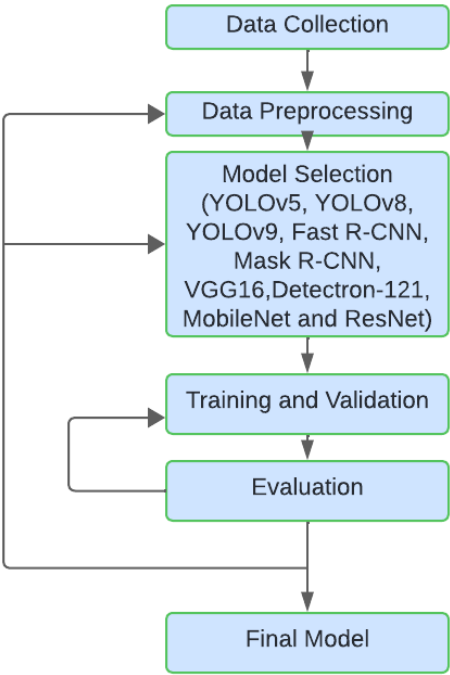


Figure 3. Overview of the key steps in our implementation.

The subsequent subsections will delve into the details of each step, providing further insights into our experimental approach and findings.

5.1. Data Preprocessing and Augmentation

The raw African pear image dataset underwent several preprocessing steps to prepare it for model training and evaluation. These steps are described below:

- **Labeling for object detection models:** The dataset of 2892 images was manually annotated using the Roboflow platform. Each image was labeled to delineate the regions corresponding to good and defective pears. Additionally, a background class was used to indicate areas where no fruit was present in the image (see Figure 4).
- **Labeling for classification Models:** For the classification models, a simplified labeling approach was used. Two separate annotation files were created, one for good pears and one for defective pears. The images were labeled with their respective class, without the inclusion of a background class. This approach was suitable for the classification task performed by these models.
- **Augmentation:** To increase the diversity of the dataset and improve the model's generalization ability, online data augmentation techniques were applied during training. These techniques included rotations, flips, zooms, and hue/saturation shifts. By augmenting the data, we introduced additional variations and enhanced the model's ability to handle different scenarios.
- **Data Splitting:** The dataset was split into three subsets: a training set comprising 70% of the data, a validation set comprising 20%, and a test set comprising the remaining 10%. The splitting was performed in a stratified manner to ensure a balanced distribution of good and defective pears in each subset.
- **Image Resizing:** The image resolutions used for the various models were selected based on the specific requirements and constraints of each model. The YOLOv5, YOLOv8, and YOLOv9 models, which are designed for real-time object detection, used higher input resolutions (416 x 416, 800 x 800, and 640 x 640 respectively) to capture more detailed visual information and improve the model's ability to detect smaller objects. The Mask R-CNN and Fast R-CNN models, used for instance segmentation, required higher-resolution inputs (640 x 640) to accurately delineate object boundaries and capture fine-grained details. In contrast, the VG16, Detectron-121, MobileNet, and ResNet models, which are classification-based and were trained on the ImageNet dataset, used a standard input size of 224 x 224 pixels, as this lower resolution is sufficient for image classification tasks, which focus on recognizing high-level visual features rather than detailed object detection or segmentation.



Figure 4. Sample images showcasing the labeling of good and defective pears with and without the background category.

5.2. Model Training

The training process involved training the Yolov9, YOLOv5, YOLOv8, Mask R-CNN, Fast R-CNN, VG16, and Detectron-121 models using the Google Colab framework. The key details of the model training are summarized in Table 1.

Table 1. Training details for the YOLOv5, YOLOv8, YOLOv9, Mask R-CNN, Fast R-CNN, VG16, Detectron-121, MobileNet, and ResNet models.

Model	Input Resolution	Batch Size	Optimizer	Training Epochs
YOLOv5	416 x 416	16	Adam	150
YOLOv8	800 x 800	16	Adam	80
YOLOv9	640 x 640	16	Adam	30
Mask R-CNN	640 x 640	8	SGD	10,000 iterations
Fast R-CNN	640 x 640	64	SGD	1,500 iterations
VG16	224 x 224	32	Adam	15
Detectron-121	224 x 224	32	SGD	50
MobileNet	224 x 224	32	Adam	40
ResNet	224 x 224	32	Adam	16

The number of training epochs for each model was varied based on the complexity of the task and the dataset, with the Mask R-CNN model requiring the most training iterations (10,000) due to

the more complex instance segmentation task, and the YOLOv9 model requiring the fewest training epochs (30) due to the simpler object detection task and the use of a pre-trained backbone.

The YOLOv5 model was trained with an input resolution of 416×416 pixels, using a batch size of 16. The Adam optimizer was employed with a learning rate determined through hyperparameter tuning. The model was trained for 150 epochs, iterating over the training dataset multiple times to optimize the model's parameters.

Similarly, the YOLOv8 model was trained with an input resolution of 640×640 pixels and a batch size of 16. The Adam optimizer was used, and the model was trained for 80 epochs. The training process involved updating the model's parameters to minimize the detection loss and enhance its ability to accurately detect and classify good and defective pears.

The Mask R-CNN and Fast R-CNN models were also trained on the African pear dataset. These models were trained with an input resolution of 640×640 pixels. The Mask R-CNN model utilized the stochastic gradient descent (SGD) optimizer with a batch size of 8 and was trained for 10,000 iterations. On the other hand, the Fast R-CNN model employed the SGD optimizer with a batch size of 64 and was trained for 1,500 iterations.

The VG16 and Detectron-121 models were trained for the classification task using an input resolution of 224×224 pixels. A batch size of 32 was used for both models. The Adam optimizer was employed, and the models were trained for 15 epochs. The MobileNet and ResNet models were also trained on the African pear dataset. Both models were trained with an input resolution of 224×224 pixels and a batch size of 32. The Adam optimizer was used, and both were trained for 40 epochs and 16 epochs respectively.

The training process involved feeding the models with the annotated dataset, allowing them to learn the features and patterns associated with good and defective pears. The models' parameters were adjusted iteratively during training to minimize the detection and classification error, optimizing their performance for the African pear defect detection task. The models' training details, including the input resolution, batch size, optimizer, and training epochs, were carefully selected to achieve the best possible performance.

We applied pruning techniques to optimize the YOLOv9, YOLOv8, YOLOv5, ResNet, and MobileNet models. Specifically, we utilized a technique called Magnitude-based Pruning, which identifies and removes the least significant weights or filters based on their absolute values. This method involves ranking all weights or filters in the model by their magnitude and setting a pruning threshold to discard those below this threshold. By removing these less important components, we effectively reduced the number of parameters in the models. This pruning process not only decreased the model size and computational requirements but also aimed to maintain the overall performance of the models.

6. Evaluation and Results

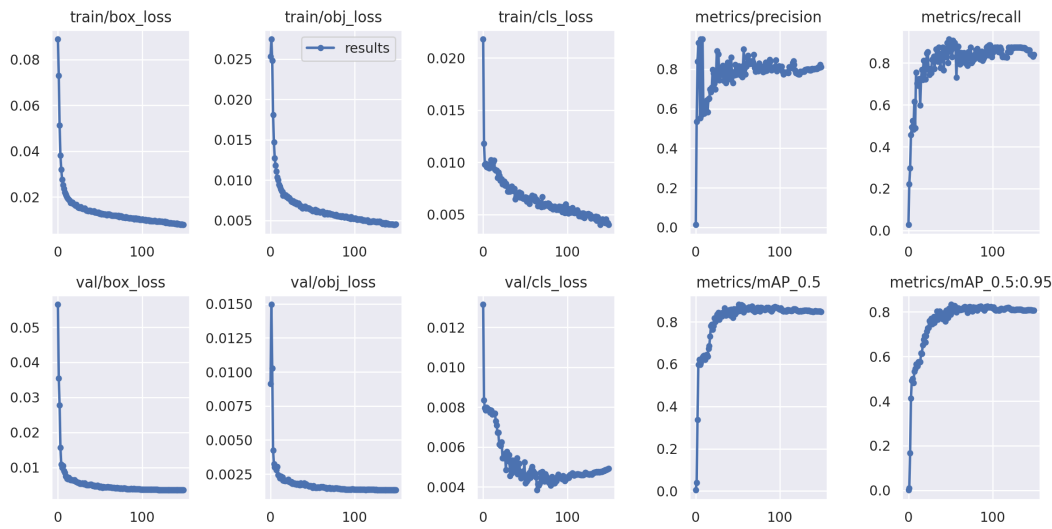


Figure 5. YOLOv5 Training Performance. This figure shows the training curves for the YOLOv5 object detection model. The top plot displays the loss function during the training process, which includes components for bounding box regression, object classification, and objectness prediction. The bottom plot displays the model’s mAP50 and mAP50-95 metrics on the validation dataset, which are key indicators of the model’s ability to accurately detect and classify objects.

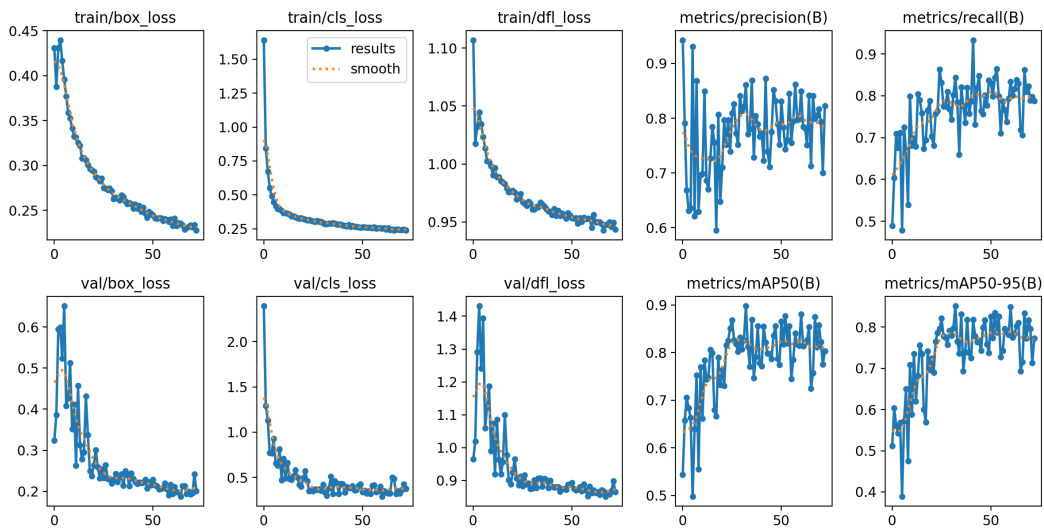


Figure 6. YOLOv8 Training and Evaluation. This figure presents the performance metrics for the YOLOv8 object detection model during the training and evaluation phases. The top plot shows the training loss, which is composed of components for bounding box regression, object classification, and objectness prediction. The bottom plot displays the model’s mAP50 and mAP50-95 metrics on the validation dataset, which are key indicators of the model’s ability to accurately detect and classify objects.

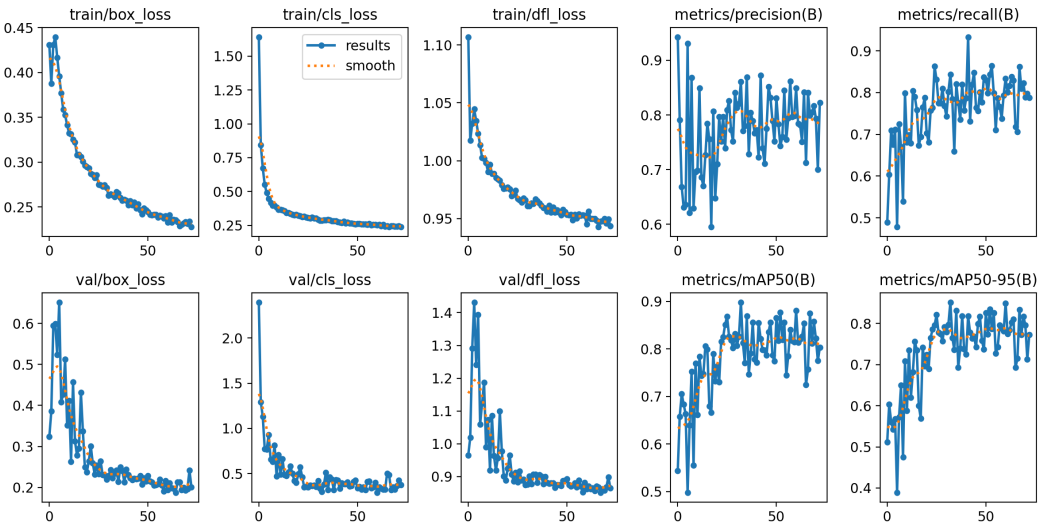


Figure 7. YOLOv9 Training and Evaluation. This figure presents the performance metrics for the YOLOv9 object detection model during the training and evaluation phases. The top plot shows the training loss, which is composed of components for bounding box regression, object classification, and objectness prediction. The bottom plot displays the model’s mAP50 and mAP50-95 metrics on the validation dataset, which are key indicators of the model’s ability to accurately detect and classify objects.



Figure 8. Fast R-CNN Training and Evaluation Metrics. This figure shows the training and validation metrics for the Fast R-CNN object detection model. The blue line represents the overall training loss, which includes components for bounding box regression, object classification, and region proposal classification. The orange and green lines show the validation metrics for the Classification loss and the Regression loss, respectively. These metrics indicate the model’s performance in generating accurate region proposals and classifying/localizing detected objects.

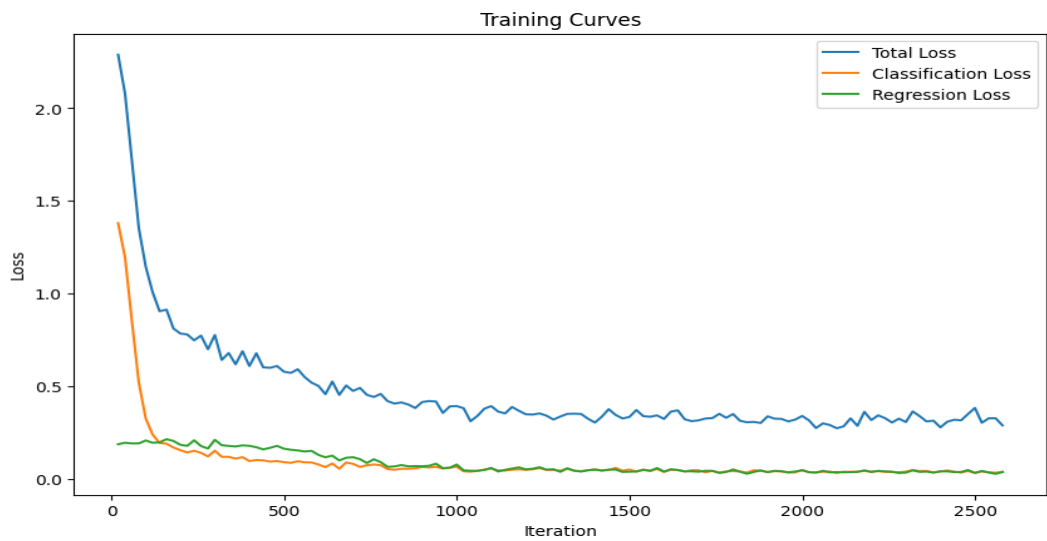


Figure 9. Mask R-CNN Training and Evaluation Metrics. This figure presents the training and validation performance metrics for the Mask R-CNN instance segmentation model. The blue line represents the overall training loss, which includes components for bounding box regression, object classification, and region proposal classification. The orange and green lines show the validation metrics for the Classification loss and the Regression loss, respectively. These metrics indicate the model’s performance in generating accurate region proposals and classifying/localizing detected objects.

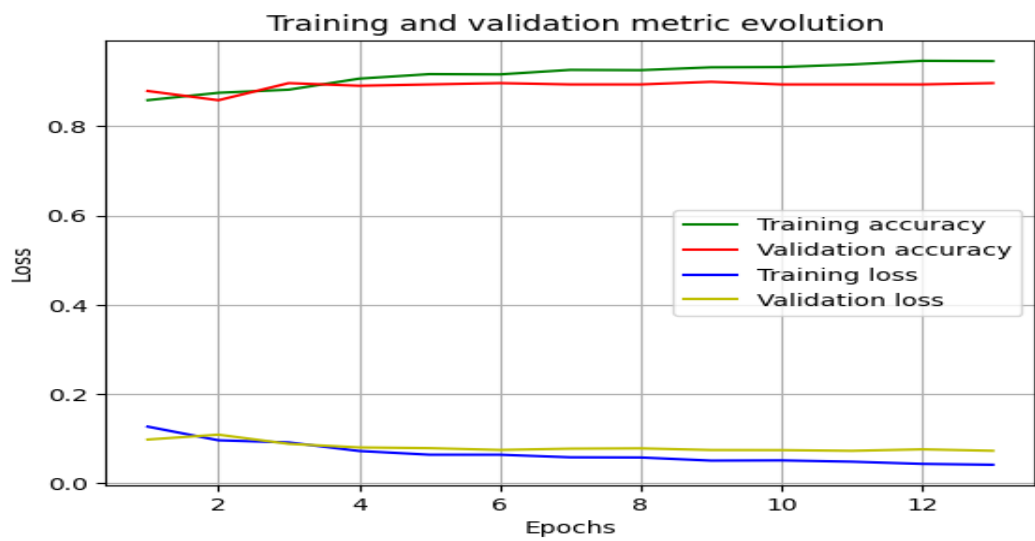


Figure 10. detectron121 training.

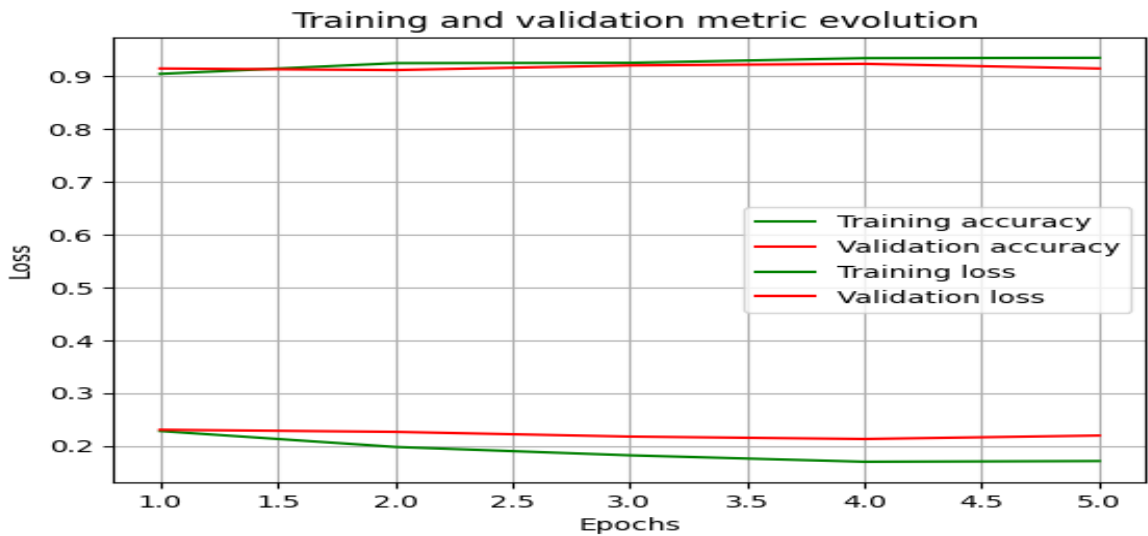


Figure 11. vg16 training.

The results presented in the tables provide valuable insights into the performance of the models for African plum quality assessment. As seen in Table 2, YOLOv8 achieved the highest mAP of 93.6%, precision of 87%, and recall of 90%, indicating it can most accurately detect both defects and good plums. YOLOv5 also performed strongly with 89.5% mAP. This highlights the effectiveness of the YOLO architecture for this object detection task.

Table 3 shows that among the classification models, ResNet achieved the best performance with 94% F1-score and 90% accuracy, demonstrating highly accurate classification of plum quality. MobileNet attained competitive results with 92% F1-score and 86% accuracy.

Table 4 reveals that pruning YOLOv5, YOLOv8 and YOLOv9 up to 20% led to only minor mAP reductions compared to their unpruned versions, showing promise for more compact model deployment with limited accuracy trade-offs.

Likewise, in Table 5, pruning ResNet and MobileNet by up to 20% maintained high performance across metrics such as accuracy, precision, and recall. This aligns with findings by Han et al. (2015), which demonstrated that neural network pruning not only reduces model size but also can enhance overall performance by removing redundant parameters and improving generalization capabilities [42]. In addition to the quantitative results reported in the tables, we also provide visual examples of the model outputs to give readers a better sense of the performance. Figure 12 shows the detection results from the YOLOv5, YOLOv8, and YOLOv9 models on sample plum images. The bounding boxes and class labels demonstrate the ability of the YOLO models to accurately identify both defective and good plums.

Table 2. Evaluation results for the YOLOv5, YOLOv8, Fast R-CNN and Mask R-CNN models.

Model	Precision (%)	Recall (%)	F1-Score (%)	mAP (%)
YOLOv5	80	85	82.5	89.5
YOLOv8	87	90	89	93.6
YOLOv9	85.9	90	87.9	93.1
Fast R-CNN	84.8	86.4	85.6	84.8
Mask R-CNN	61.3	68.2	64.6	61.3

Table 3. Evaluation results for the VG16, Detectron-121 models, MobileNet and ResNet .

Model	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
VG16	78	80	79	91
Detectron-121	80	82	81	86
MobileNet	87	97	92	86
ResNet	91	98	94	90

Table 4. Results of pruned YOLOv5, YOLOv8, and Yolo9 models.

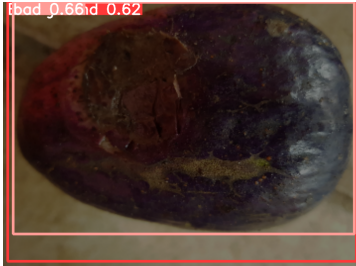
Model	Pruning (%)	mAP
YOLOv5	-	0.895
	30	0.875
	20	0.891
	10	0.898
YOLOv8	-	0.936
	30	0.597
	20	0.81
	10	0.902
Yolov9	-	0.931
	30	0.719
	20	0.924
	10	0.93

Table 5. Results of pruned ResNet and MobileNet models

Model	Pruning (%)	Accuracy	Precision	Recall	F1 Score
ResNet	0	0.9086	0.9132	0.9861	0.9482
	10	0.8614	0.9433	0.8956	0.9188
	20	0.7965	0.7934	1.0000	0.8848
	30	0.6549	0.9649	0.5978	0.7383
MobileNet	0	0.8614	0.8734	0.9753	0.9215
	10	0.8909	0.8968	0.9823	0.9376
	20	0.8820	0.8935	0.9754	0.9327
	30	0.8614	0.8944	0.9373	0.9153



(a) YOLOv5 Good Fruit prediction.



(b) YOLOv5 Bad Fruit prediction.



(c) YOLOv8 Good Fruit Prediction.



(d) YOLOv8 Bad Fruit Prediction.



(e) YOLOv9 Good Fruit Prediction.



(f) YOLOv9 Bad Fruit Prediction.

Figure 12. Model Predictions with Background Class

Overall, the results validate YOLOv8 as the premier model, while also indicating classification architectures like ResNet perform well. Pruning experiments demonstrate potential to compress models without severely impacting quality assessment capability

7. Conclusion

In this research, we have demonstrated the feasibility of utilizing deep learning techniques for automated external quality evaluation of African pears. A comprehensive dataset comprising over 2800 African pear images was curated, encompassing variations in shape, size, color, defects, and imaging conditions across major pear-growing regions in Cameroon. Through the utilization of this dataset, we trained and evaluated different deep learning architectures including YOLOv5, YOLOv8, YOLOv9, Faster R-CNN, Mask R-CNN, VG16, Detectron-121, MobileNet, and ResNet for the detection and quantification of defects as well as classification tasks on the pear surface.

Among the object detection models, YOLOv8 and YOLOv9 exhibited strong performances, with mean average precision (mAP) values of 89% and 93.1%, respectively, and F1-scores of 89% and 87.9%. These results indicate the suitability of the YOLO framework for accurate and efficient defect detection in African pears. Furthermore, both models were successfully integrated into a functional web application, enabling real-time surface inspection of African pears.

For the classification task, the results from VG16, Detectron-121, MobileNet, and ResNet models were evaluated based on precision, recall, F1-score, and accuracy. However, the pruning results for these models were not provided in the available information, making it challenging to discuss the impact of pruning on their performance.

The evaluation of pruning techniques on the YOLOv8 and YOLOv9 models revealed their sensitivity to pruning levels. While moderate pruning (10% and 20%) maintained relatively high mAP values, heavy pruning (30%) significantly degraded their performance. These results highlight the importance of considering pruning strategies carefully to strike a balance between model size reduction and performance preservation.

Moving forward, there are several avenues for expanding upon this research. Firstly, the collection of more annotated data encompassing additional pear varieties, growing seasons, and defect types would further enhance the robustness and generalization capability of the models. Additionally, exploring solutions for internal quality assessment using non-destructive techniques such as hyperspectral imaging would be an important direction for future research.

Furthermore, extending the application of intelligent inspection to other African crops by creating datasets and models specific to those commodities would broaden the impact of this technology. Evaluating model performance under challenging real-world conditions and incorporating active learning techniques for online improvement are also crucial for ensuring the practical applicability and effectiveness of these AI-based systems.

In summery, this pioneering work demonstrates the potential of AI and advanced sensing technologies in transforming African agriculture. By addressing the limitations identified in this study through future research, we can unlock the full capabilities of intelligent technology to enhance food systems and contribute to the sustainable development of African agriculture.

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. K. Schreckenberger, A. Degrande, C. Mboosso, Z. Boli Baboulé, C. Boyd, L. Enyong, J. Kanmegne, C. Ngong. *The social and economic importance of Dacryodes edulis* (G. Don) HJ Lam in Southern Cameroon. *Forests, Trees and Livelihoods*, vol. 12, no. 1-2, pp. 15-40, 2002.
3. Aurore Rimlinger, Stéphanie M Carrière, Marie-Louise Avana, Aurélien Nguengang, Jérôme Duminil. *The influence of farmers' strategies on local practices, knowledge, and varietal diversity of the safou tree (Dacryodes edulis) in Western Cameroon*. *Economic Botany*, vol. 73, pp. 249-264, 2019.

4. Leseho Swana, Bienvenu Tsakem, Jacqueline V Tembu, Ramy B Teponno, Joy T Folahan, Jarmo-Charles Kalinski, Alexandros Polyzois, Guy Kamatou, Louis P Sandjo, Jean Christopher Chamcheu, et al. *The Genus Dacryodes Vahl.: Ethnobotany, Phytochemistry and Biological Activities*. Pharmaceuticals, vol. 16, no. 5, pp. 775, 2023.
5. Roger RB Leakey, Marie-Louise Tientcheu Avana, Nyong Princely Awazi, Achille E Assogbadjo, Tafadzwanashe Mabhaudhi, Prasad S Hendre, Ann Degrande, Sithabile Hlahla, Leonard Manda. *The future of food: Domestication and commercialization of indigenous food crops in Africa over the third decade (2012–2021)*. Sustainability, vol. 14, no. 4, pp. 2355, 2022.
6. Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2(3), 160.
7. Zhou, Z. H. (2021). *Machine learning*. Springer Nature.
8. Szeliski, R. (2022). *Computer vision: algorithms and applications*. Springer Nature.
9. Szeliski, R. (2014). *Concise Computer Vision. An Introduction into Theory and Algorithms*. Springer-Verlag.
10. Apostolopoulos, I. D., Tzani, M., and Aznaouridis, S. I. (2023). A General Machine Learning Model for Assessing Fruit Quality Using Deep Image Features. *AI*, 4(4), 812–830.
11. Xiao, F., Wang, H., Xu, Y., and Zhang, R. (2023). Fruit Detection and Recognition Based on Deep Learning for Automatic Harvesting: An Overview and Review. *Agronomy*, 13(6), 1625.
12. Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
13. Xu, B., Cui, X., Ji, W., Yuan, H., and Wang, J. (2023). Apple grading method design and implementation for automatic grader based on improved YOLOv5. *Agriculture*, 13(1), 124.
14. Rahat, S. M. S. S., Al Pitom, M. H., Mahzabun, M., and Shamsuzzaman, M. (2023). Lemon Fruit Detection and Instance Segmentation in an Orchard Environment Using Mask R-CNN and YOLOv5. In *Computer Vision and Image Analysis for Industry 4.0* (pp. 28–40). CRC Press.
15. Mao, D., Zhang, D., Sun, H., Wu, J., & Chen, J. (2024). Using filter pruning-based deep learning algorithm for the real-time fruit freshness detection with edge processors. *Journal of Food Measurement and Characterization*, 18(2), 1574–1591.
16. Solawetz, J. (2022). What is Yolov5? A guide for beginners. *Roboflow Blog*.
17. Hussain, M. (2023). YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines*, 11(7), 677.
18. Chien-Yao Wang, I-Hau Yeh, and Hong-Yuan Mark Liao. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv preprint arXiv:2402.13616*, 2024.
19. Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1440–1448).
20. He, K., Gkioxari, G., Doll'ar, P., and Girshick, R. (2017). Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2961–2969).
21. Mascarenhas, S., and Agarwal, M. (2021). A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. In *2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)* (Vol. 1, pp. 96–99). IEEE.
22. Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4700–4708).
23. Yanyu Li, Geng Yuan, Yang Wen, Ju Hu, Georgios Evangelidis, Sergey Tulyakov, Yanzhi Wang, and Jian Ren. Efficientformer: Vision transformers at mobilenet speed. *Advances in Neural Information Processing Systems*, volume 35, pages 12934–12949, 2022.
24. Ross Wightman, Hugo Touvron, and Hervé Jégou. Resnet strikes back: An improved training procedure in timm. *arXiv preprint arXiv:2110.00476*, 2021.
25. Chu, P., Li, Z., Lammers, K., Lu, R., and Liu, X. (2021). Deep learning-based apple detection using a suppression mask R-CNN. *Pattern Recognition Letters*, 147, 206–211.
26. Asriny, D. M., Rani, S., and Hidayatullah, A. F. (2020). Orange fruit images classification using convolutional neural networks. In *IOP Conference Series: Materials Science and Engineering* (Vol. 803, No. 1, pp. 012020). IOP Publishing.
27. Lamb, N., and Chuah, M. C. (2018). A strawberry detection system using convolutional neural networks. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 2515–2520). IEEE.

28. Nithya, R., Santhi, B., Manikandan, R., Rahimi, M., and Gandomi, A. H. (2022). Computer vision system for mango fruit defect detection using deep convolutional neural network. *Foods*, 11(21), 3483.
29. Asif Iqbal Khan, SMK Quadri, Saba Banday, and Junaid Latief Shah. *Deep diagnosis: A real-time apple leaf disease detection system based on deep learning*. *Computers and Electronics in Agriculture*, vol. 198, p. 107093, 2022. Publisher: Elsevier.
30. Xiaoyu Liu, Guo Li, Wenkang Chen, Binghao Liu, Ming Chen, and Shenglian Lu. *Detection of dense Citrus fruits by combining coordinated attention and cross-scale connection with weighted feature fusion*. *Applied Sciences*, volume 12, number 13, page 6600, 2022. Publisher: MDPI.
31. Kusrini, Kusrini, Suputa, Suputa, Setyanto, Arief, Agastya, I Made Artha, Priantoro, Herlambang, and Pariyasto, Sofyan. *A comparative study of mango fruit pest and disease recognition*. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 20, no. 6, pp. 1264–1275, 2022.
32. Stevan Čakić, Tomo Popović, Srdjan Krčo, Daliborka Nedić, and Dejan Babić. *Developing object detection models for camera applications in smart poultry farms*. In *2022 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*, pages 1–5, 2022. Organization: IEEE.
33. Analyn N. Yumang, Christian Joseph N. Samilin, and John Christian P. Sinlao. *Detection of Anthracnose on Mango Tree Leaf Using Convolutional Neural Network*. In *2023 15th International Conference on Computer and Automation Engineering (ICCAE)*, pages 220–224, 2023. Organization: IEEE.
34. Sai Sudha Sonali Palakodati, Venkata Rami Reddy Chirra, Dasari Yakobu, and Suneetha Bulla. *Fresh and Rotten Fruits Classification Using CNN and Transfer Learning*. *Rev. d'Intelligence Artif.*, vol. 34, no. 5, pp. 617–622, 2020.
35. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. *Rethinking the inception architecture for computer vision*. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.
36. G.Don. *Pachylobus edulis*. *Genera Historiae*, 2: 89, 1832. Available at: <https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:128214-1>.
37. 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.
38. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. *Faster r-cnn: Towards real-time object detection with region proposal networks*. *Advances in neural information processing systems*, vol. 28, 2015.
39. Ankur Chaturvedi and Vikram Rajpoot. *An Optimized Deep Vision Framework*. *Solid State Technology*, vol. 63, no. 6, pp. 561–569, 2020.
40. Qinjie Lin, Guo Ye, Jiayi Wang, and Han Liu. *RoboFlow: a data-centric workflow management system for developing AI-enhanced Robots*. In *Conference on Robot Learning*, pages 1789–1794, 2022. PMLR.
41. Zhu, Linlin, Xun Geng, Zheng Li, and Chun Liu. "Improving YOLOv5 with attention mechanism for detecting boulders from planetary images." *Remote Sensing* 13, no. 18 (2021): 3776.
42. Han, S., Pool, J., Tran, J., & Dally, W. (2015). Learning both weights and connections for efficient neural network. In *Advances in Neural Information Processing Systems* (pp. 1135–1143).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.