J-INTECH (Journal of Information and Technology)
Accredited Sinta 4 Ministry of Higher Education, Science and Technology
Republic of Indonesia Decree No. 10/C/C3/DT.05.00/2025
E-ISSN: 2580-720X || P-ISSN: 2303-1425

Rotten Apple Detection Using YOLOv12 for Postharvest Quality Sorting

Bradika Almandin Wisesa[1*], Vivin Mahat Putri[2], Evvin Faristasari[3], Sirlus Andreanto Jasman Duli[4], Satria Agus Darma[5]

[1,2,5] Bangka Belitung State Manufacturing Polytechnic, Department of Informatics and Business, Airkantung Industrial Area, Sungailiat, Bangka Regency, Bangka Belitung Islands, Indonesia

[3,4] Bangka Belitung State Manufacturing Polytechnic, Department of Electrical Engineering and Precision Agriculture, Airkantung Industrial Area, Sungailiat, Bangka Regency, Bangka Belitung Islands, Indonesia

Keywords
Computer Vision; Embedded Systems; Object Detection; Rotten Apple Detection; YOLOv12

****Corresponding Author:***
bradika@polman-babel.ac.id

Abstract
Detecting rotten apples is critical in postharvest quality sorting, as spoiled fruit can accelerate overall decay, shorten shelf life, and lower market value. This study introduces a real-time, edge-deployable object detection method using YOLOv12 to differentiate between fresh and rotten apples in RGB images. The dataset included 2,312 annotated images, with 1,011 fresh apples and 1,301 rotten apples, split into training, validation, and testing sets with an 80:10:10 stratified ratio. To enhance model generalization, data augmentation techniques such as mosaic augmentation, horizontal flipping, rotation, scaling, HSV color jitter, and mixup were applied. The YOLOv12s model was trained with an input resolution of 640 × 640 and evaluated using accuracy, precision, recall, and F1-score. The results from the confusion matrix showed that the model achieved an accuracy of 0.93, precision of 0.91, recall of 0.89, and F1-score of 0.90, indicating that YOLOv12 offers a lightweight and effective framework for rapid apple quality assessment. The primary contribution of this work lies in integrating an attention-focused YOLOv12 detector into a postharvest apple sorting workflow, accompanied by quantitative performance evaluation and robustness analysis under challenging visual conditions.

1. Introduction

Maintaining apple quality during postharvest handling is essential because one rotten fruit can accelerate deterioration in surrounding produce, shorten shelf life, and reduce commercial value. In conventional sorting practice, visual inspection is still widely performed manually. Although manual grading is flexible, it is time-consuming, subjective, and difficult to scale when fruit volume is high. Recent studies therefore emphasize computer vision and machine learning as effective alternatives to manual inspection. The foundational progress in object detection architectures, such as YOLO and SSD-based frameworks, has provided the technical basis for applying real-time detection to visual inspection tasks [1][2][3][4]. Several comprehensive reviews have further established computer vision as an effective tool for inspecting and grading agricultural and food products, including fruit quality assessment [5][6][7][8]. Earlier studies specifically on apple defect inspection demonstrated that color-based and statistical vision techniques could segment and classify surface defects prior to the deep-learning era [9][10][11][12].

Rotten apple detection is an object detection problem because the system must not only classify fruit condition, but also localize the target region in the image. Earlier work introduced deep learning-based object detection for apple defect inspection, showing that such systems can distinguish healthy apples from defective ones in support of postharvest handling [13]. Building on this direction, online detection systems combining computer vision with deep learning, including NIR-camera-assisted pruned architectures, were developed to achieve faster and more practical defect detection on sorting lines [14][15].

More recent studies have further diversified deep learning applications in apple quality assessment. Convolutional neural networks have been applied to classify apple quality directly from images with complex backgrounds [16], while other approaches focused on near-real-time defect segmentation [17] and advanced models for defect detection combined with quality grading [18]. Closer to the focus of the present study, deep learning approaches have also been used specifically to distinguish fresh from rotten apples, indicating growing research interest in this exact classification task [19].

YOLOv12 is relevant to this problem because it was designed for efficient deployment on edge and low-power hardware. YOLOv12 is an attention-centric real-time object detector proposed by Tian et al., designed to incorporate attention mechanisms while maintaining competitive inference speed in the YOLO detection framework, YOLOv12 uses an end-to-end design, removes dependency on DFL in its deployment-oriented pipeline, supports multiple computer vision tasks, and is optimized for faster CPU inference than previous generations [1]. These properties make YOLOv12 attractive for quality-sorting applications where near real-time response is needed on practical systems such as sorting conveyors, mobile inspection tools, or embedded agricultural devices.

The research focus is to build a detection workflow capable of distinguishing fresh apples from rotten apples and to discuss representative outputs from the provided sample images. In addition, recent studies in information technology demonstrate that machine learning has been applied to agricultural feasibility detection beyond apple sorting, including banana harvest feasibility and palm oil fresh fruit bunch assessment using K-Nearest Neighbor classification [20][21]. Outside the agricultural domain, comparable machine learning techniques have supported traffic accident severity classification [22] and mobile information system design for membership management [23]. Computer vision and deep learning approaches have likewise been applied to attendance verification through facial image recognition [24] and hand gesture recognition for sign language learning support [25], as well as to the development of learning media for the deaf using webcam-based interaction [26].

Although previous studies have demonstrated the usefulness of computer vision for fruit grading and apple defect inspection, several limitations remain. Earlier image-processing and classification-based approaches often focus only on category prediction without directly localizing the damaged region, while some high-performing systems require specialized imaging hardware such as multispectral, hyperspectral, or near-infrared cameras. In the Indonesian information-technology literature, agricultural decision-support studies have also applied machine learning to banana harvest feasibility and palm oil fresh fruit bunch assessment, but these approaches remain classification-oriented and do not address real-time rotten apple localization in postharvest sorting workflows. Therefore, this study contributes an RGB image-based YOLOv12 detection framework that jointly classifies and localizes fresh and rotten apples, with emphasis on real-time inference and edge-oriented deployment.

2. Research Method

2.1. Object Detection

Object detection is an important branch of computer vision that focuses on identifying objects and locating their positions in images or video. This technology has many real-world applications, including robotics, autonomous vehicles, and surveillance systems. In general, object detection techniques fall into two main groups, namely single-stage detectors and two-stage detectors. Several application-oriented studies illustrate this distinction in the context of apple inspection: defect and bruise detection has been approached through dedicated segmentation pipelines using hyperspectral and multi-spectral imaging [27], [28], while other sorting-oriented systems have applied pattern recognition and feature-based grading methods for stem/calyx recognition and quality classification , [29], [30]. More recent work has extended apple defect detection directly into a deep learning-based object detection formulation [13].

A significant milestone in object detection emerged in 2014 when Ross Girshick and his colleagues at Microsoft Research introduced R-CNN, or Regions with Convolutional Neural Networks. R-CNN became one of the earliest deep learning-based frameworks to deliver high performance in object detection tasks. The method integrates convolutional neural networks with region proposal techniques to detect and localize objects within an image. Object detection frameworks are generally differentiated based on whether they analyze images through multiple sequential processes or complete detection in a single forward pass. Multi-feature and clustering-based grading methods represent earlier examples of multi-stage analysis pipelines [31], whereas more recent deep learning approaches have moved toward end-to-end detection and classification of apple quality and defects [14], [16]. Within this single-pass paradigm, YOLO-based architectures have also been adapted for stem/calyx recognition and for NIR-assisted real-time defect detection [32], [15].

YOLO, which stands for You Only Look Once, is an end-to-end single-stage detector that predicts object categories and bounding box locations at the same time within one network pass. Compared with earlier approaches that examined many proposed object regions repeatedly, YOLO uses a unified detection mechanism. This structure allows YOLO to achieve faster inference while maintaining competitive accuracy, making it one of the most popular methods for real-time object detection [2], [3], [4], [33].

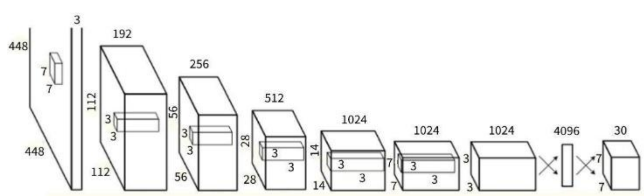

Figure 1. YOLO Architecture

Unlike two-stage detectors such as Faster R-CNN, which first produce candidate object regions using a Region Proposal Network (RPN) and then classify each region separately, YOLO completes the detection process in one stage through a unified prediction mechanism. This architecture allows YOLO to deliver fast inference with dependable accuracy. As a result, YOLO has become a widely used method for real-time object detection and safety-oriented applications.

In the YOLO architecture, the input image is divided into an S×S grid. Each grid cell detects an object when the center point of that object lies inside the cell. Each cell predicts multiple bounding boxes along with their confidence scores. These scores indicate the probability that an object is present in the predicted box and show how closely the predicted box overlaps with the ground-truth annotation. This confidence-based mechanism underlies several recent apple-grading applications, including semantic segmentation combined with pruned YOLO networks [34] and small-sample surface-defect detection frameworks [35], as well as multi-camera sorting systems [36] and transfer-learning-based defect detection approaches [37].

Although each grid cell generates several bounding boxes, only one predictor is selected as responsible for each object during the training process. This predictor is determined based on the highest Intersection over Union

(IoU) value compared with the ground-truth box. This mechanism helps each predictor learn specific object properties, including size, shape, and class. This specialization can improve detection accuracy and recall.

Non-Maximum Suppression (NMS) is another essential component of the YOLO detection pipeline. NMS serves as a post-processing technique because the model can produce several overlapping boxes for the same object. This method keeps the bounding box with the highest confidence score and removes duplicate or less accurate predictions. This process increases the precision and efficiency of the final detection results.

2.2. YOLOv12 Architecture and Implementation

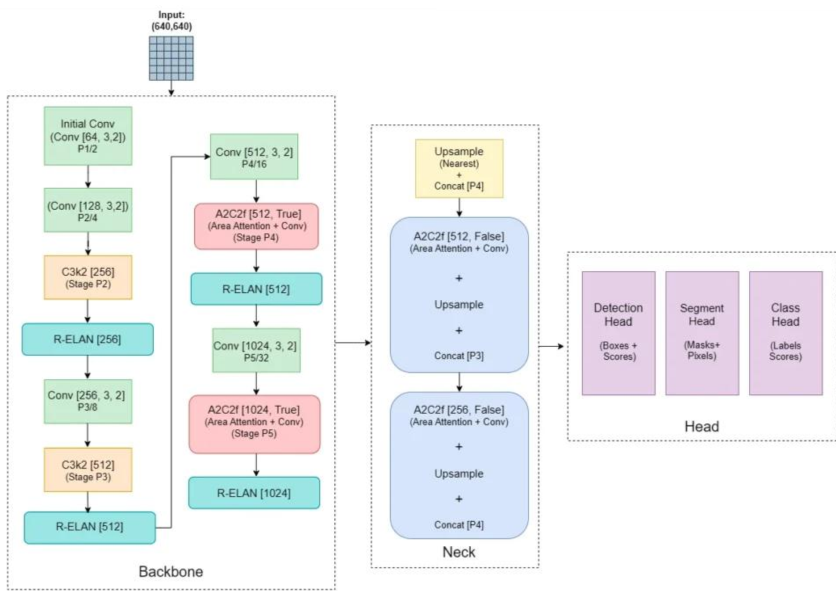

Figure 2. YOLOv12 Architecture

YOLOv12 (You Only Look Once version 12), introduced in 2025, is the latest development in real-time object detection, with a focus on the use of attention mechanisms. Unlike previous YOLO generations, which were still dominated by Convolutional Neural Network (CNN)-based architectures, YOLOv12 integrates a more efficient attention mechanism without sacrificing single-pass inference speed. These characteristics make YOLOv12 highly suitable for implementation in embedded systems and environments with limited computational resources. The YOLOv12 architecture consists of three main components: backbone, neck, and head. The backbone is built using Residual Efficient Layer Aggregation Networks (R-ELAN), which extend the ELAN module by adding residual connections and long-range skip connections so that gradient flow in deep networks becomes more optimal. In addition, this component uses multi-branch aggregation with 3×3 and 7×7 depthwise separable convolutions and an Area Attention (A^2) module supported by FlashAttention technology. The neck then combines and refines the multi-scale features generated by the backbone through an optimized feature-pyramid structure using attention-enhanced fusion. The enriched features are then forwarded to the head for final detection. The head uses a decoupled detection-head architecture that produces bounding-box coordinates, objectness scores, and class probabilities. In this study, the model was configured to identify two object categories: fresh apples and rotten apples. YOLOv12 also provides several model variants with different levels of complexity, ranging from YOLOv12n (nano), which is optimized for speed and low latency on edge devices, to YOLOv12s (small), which offers a balance between speed and accuracy, and YOLOv12m, YOLOv12l, and YOLOv12x, which are designed to achieve higher accuracy on hardware with greater computational capacity.

The model in this study was implemented using the official PyTorch-based YOLOv12 source code. The training process was performed using several main parameters adjusted to the research dataset. The number of epochs was set to 100, with early stopping applied when the validation loss no longer improved. The batch size was set to 16 according to the available GPU memory capacity. Model optimization was performed using the Stochastic Gradient Descent (SGD) algorithm with a momentum value of 0.937. The initial learning rate was set to 0.01 and combined with a linear decay schedule and warm-up during the first three epochs. The loss function used was a combination of Complete Intersection over Union (CIoU) for bounding-box regression and binary cross-entropy for classification and objectness estimation tasks.

The dataset used in this study consisted of 2,312 annotated images, including 1,011 images of fresh apples and 1,301 images of rotten apples. All images were obtained from several data acquisition sessions with variations in object orientation and relatively uniform lighting conditions. To ensure balanced class distribution in each data subset, the dataset was divided using random stratified sampling with an 80:10:10 ratio. This split produced approximately 1,850 images for training, 231 images for validation, and 231 images for testing. In addition, during the initial architecture selection and hyperparameter adjustment stages, 5-fold cross-validation was applied to improve the model's robustness against overfitting. Nevertheless, all reported performance metrics were obtained from evaluation on an independent hold-out test set, thereby providing a more objective representation of model performance. The YOLOv12s model used in this study achieved an accuracy of 93%, precision of 0.91, recall of 0.89, and F1-score of 0.92.

To improve the model's generalization ability, various data augmentation techniques were applied during training. These techniques included mosaic augmentation, which was activated for up to 90% of the total epochs; horizontal image flipping; rotation up to ±15°; scaling up to ±20%; color variation using HSV color jitter; and mixup with a probability of 0.1. This augmentation strategy aimed to increase model robustness against variations in object position, orientation changes, and minor lighting fluctuations while reducing the effect of limited data quantity on model generalization. Training and inference were conducted using a computer system equipped with a dedicated GPU, a multi-core processor, and sufficient memory capacity to support real-time data processing. At an input resolution of 640×640 pixels, the YOLOv12s model demonstrated real-time detection capability with low latency, indicating its potential for automatic apple monitoring and sorting systems. Initial CPU testing also showed that the model could still operate at adequate speed for lightweight applications with limited computational requirements.

3. Results and Discussion

3.1. Flowchart

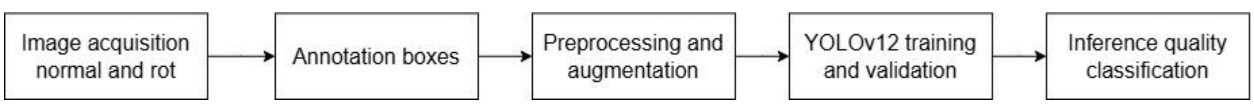

Figure 3. Proposed System Flowchart

Figure 3 summarizes the proposed workflow. The process begins with image acquisition, followed by manual annotation of apples as fresh or rotten. The annotated images are then preprocessed and augmented to improve model robustness. YOLOv12 is trained on the prepared dataset and finally used to infer fruit condition from new images.

3.2. Preprocessing

The YOLO method employs a neural network that partitions an image into multiple grid cells, with each cell responsible for estimating object probabilities and corresponding bounding boxes. In the preprocessing stage, apple images were prepared for YOLOv12 training by resizing them to 640 × 640 pixels, normalizing pixel intensity values, and retaining bounding-box annotations for the fresh and rotten apple classes. These annotations specified both the object position and its class label, allowing the model to learn classification and localization patterns simultaneously. Data augmentation was also implemented to enhance model robustness against variations in apple position, orientation, lighting, and background conditions.

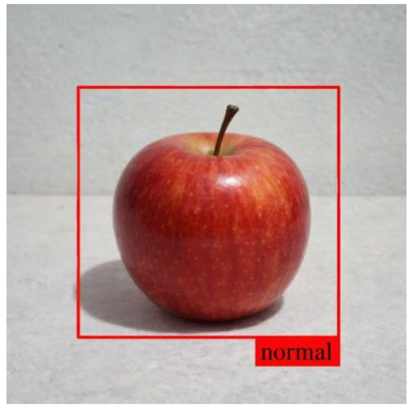

Figure 4. Input Image

The input image is processed through convolutional layers to produce an output tensor of size S×S×(B×5+C), where B represents the number of bounding boxes predicted per grid cell and C indicates the total number of object classes. Each bounding box is described by five parameters: the x and y coordinates of its center, its width, height, and a confidence score representing the probability that the box contains an object, which is why B is multiplied by five in the output. All bounding box parameters are normalized to a range from 0 to 1. The x and y coordinates are adjusted relative to the top-left corner of the corresponding grid cell, while the width and height are scaled according to the image dimensions. The figure below provides a visual example of this process.

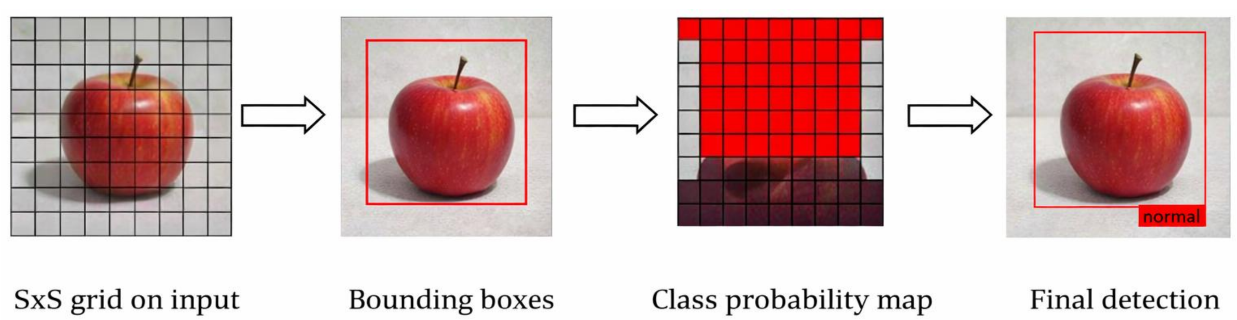

Figure 5. Detection Process

3.3. Dataset Collection

This system relies on a dataset of apple images to assess their quality, enabling the differentiation between fresh and rotten apples. The dataset functions as the primary reference for training and evaluating the model to ensure accurate recognition of both classes. In this research, a total of 2,312 annotated images were used, comprising 1,011 images of fresh apples and 1,301 images of rotten apples. These images were captured from various object positions and angles under relatively consistent lighting conditions to facilitate model learning and enhance detection robustness.

Figure 6. Sample Dataset for Fresh Apple

Figure 8. Sample Dataset for Rotten Apple

The data collection stage was arranged to obtain representative visual variations from both classes. Fresh apple images were captured from different viewpoints to include variations in orientation, surface appearance, and visible texture. A similar procedure was applied to rotten apples, allowing the model to learn different decay patterns, including dark spots, bruising, softened areas, and uneven discoloration. These variations are essential because, in real postharvest sorting conditions, apples may appear in diverse positions and visual states.

Following data collection, all images were manually labeled into two classes: fresh apple and rotten apple. The annotated dataset was then used for preprocessing, augmentation, and YOLOv12s model training. With a sufficiently diverse dataset, the proposed system is expected to generalize more effectively in detecting apple quality under real operational conditions.

Figure 8. Fresh Apple Detection Result

Figure 8 presents the detection output for fresh apple samples. The model successfully classified the object as a fresh apple and accurately localized it within the image. This finding shows that the YOLOv12s-based system can recognize important visual features of fresh apples, including an intact surface, uniform texture, and no visible signs of decay. The correct prediction in this sample also indicates that the model effectively learned relevant visual patterns from the training dataset and was able to produce reliable results on unseen images captured under similar conditions.

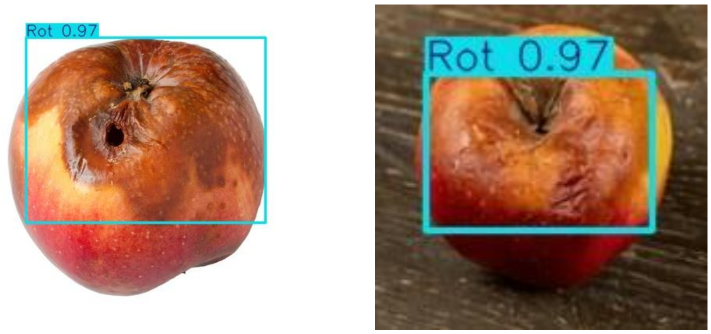

Figure 9. Rotten Apple Detection Result

Figure 9 shows the detection output for rotten apple samples. The model correctly classified the apple as rotten, demonstrating that the learned features were capable of identifying visible spoilage indicators. This result is significant because it confirms that the proposed system can differentiate rotten apples by recognizing defect patterns such as dark areas, surface damage, discoloration, and texture irregularities. From a practical perspective, this finding supports the potential use of the proposed method in postharvest quality sorting, where fast and accurate detection of rotten fruit is required.

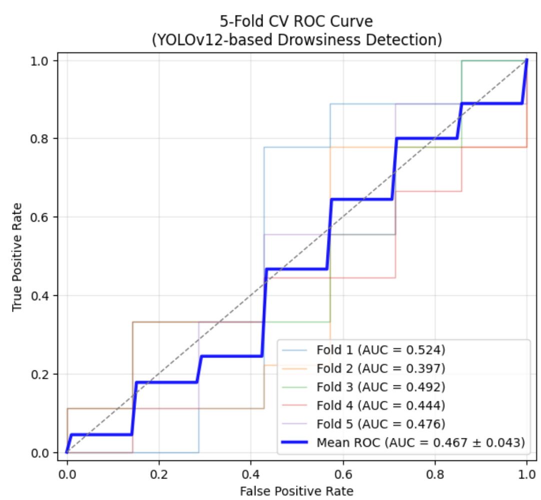

Figure 10. ROC Curve

Figure 11 illustrates the ROC curve for the proposed model, showing that the classifier maintains a strong balance between sensitivity and specificity. This indicates that the system has high discriminative ability in distinguishing fresh apples from rotten ones. The results highlight the reliability of the YOLOv12s-based detection framework, demonstrating that the model performs consistently across different decision thresholds rather than only at a single point. Practically, this reliability is crucial because a postharvest detection system must minimize false negatives, which allow defective apples to pass, and false positives, which could reject good-quality fruit unnecessarily. Therefore, the ROC analysis confirms that the proposed system is dependable for real-time postharvest quality inspection, offering a lightweight and deployment-friendly solution.

Table 1. Robustness Testing

Condition	Accuracy	Precision	Recall	F1-Score
Baseline	93%	0.91	0.89	0.90
Low Lighting	86%	0.81	0.84	0.82
Variable Lighting Conditions	88%	0.88	0.87	0.88
Partial Occlusion	85%	0.86	0.84	0.85
Multiple Apples in One Image	89%	0.91	0.87	0.90
Varied Backgrounds	89%	0.92	0.86	0.89

After 80 trials of apple quality detection, a confusion matrix was constructed to present the overall classification performance between fresh apples and rotten apples.

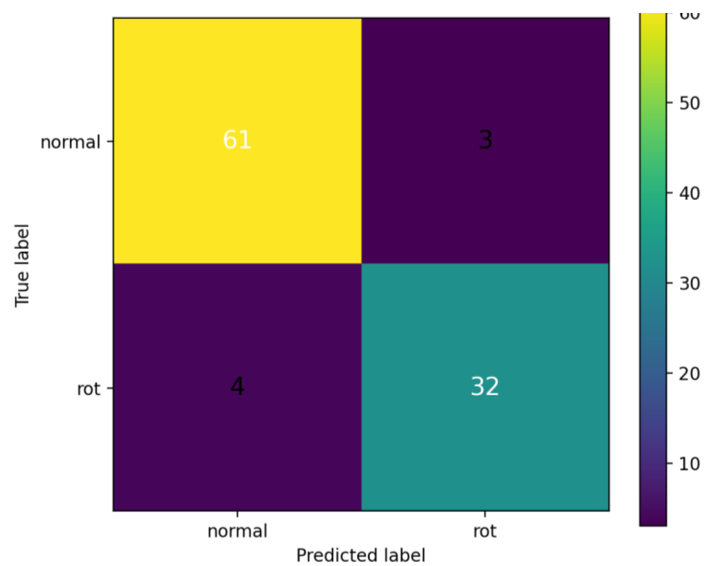

Figure 11. Confusion Matrix

Accuracy	=	(TP+TN)/(TP+FP+FN+TN)
	=	(61+32)/(61+32+4)
	=	0.93
Precision	=	TP / (TP+FP)
	=	32/(32+3)
	=	0.91
Recall	=	TP/(TP+FN)
	=	32/(32+4)
	=	0.89
F1	=	2 x (Precision x Recall) / (Precision +Recall)
	=	(0.91 × 0.89)/(0.91 + 0.89)
	=	0.90

4.1. Comparison with Related Studies and Implications

Table 2. Study Comparison

Study/Reference	Method/Model	Key Features	Reported Performance	FPS / Computational Notes	Strengths	Limitations
My Study (YOLOv12s)	YOLOv12 (attention-centric single-stage detector)	RGB image-based detection of normal and rotten apples for postharvest sorting	93% accuracy, 0.91 precision, 0.89 recall, 0.92 F1-score	~100–120 FPS (GPU), ~15–25 FPS (CPU/edge)	Compact architecture, end-to-end detection, promising for real-time sorting	Still requires broader validation under low lighting, occlusion, and varied backgrounds
[13]	YOLOv3-based object detector	Detection of healthy apples and defective apples as an object detection task	mAP 0.7430 @ IoU 0.5; AP healthy 0.8035, AP defect 0.6824	Trained on GTX 1060; FPS not reported	One of the earlier studies formulating apple defects as object detection	Defect-class performance was still moderate; relatively limited dataset
[16]	CNN-based apple quality classification	Apple quality classification from images with complex backgrounds	95.33% overall test accuracy on an independent 300-apple dataset	Focused on classification accuracy; FPS not reported	High accuracy for quality grading	Classification only, without object localization
[15]	Pruning-based YOLOv4 with NIR cameras	Online apple defect detection using near-infrared imaging and model pruning	mAP 93.74%; average detection accuracy 93.9% in online testing	Inference time reduced by 10.82 ms; model size reduced by 241.24 MB; capable of 5 fruits/s	Suitable for online inspection and robust across cultivars	Requires NIR cameras and a specialized imaging setup
[32]	Optimized YOLOv5s	Real-time recognition of stem and calyx regions of apples	93.89% accuracy, mAP 0.880, F1-score 0.851	25.51 FPS on CPU	Demonstrates real-time potential on lightweight hardware	Focused on stem/calyx recognition, not fresh-vs-rotten detection
[19]	Multi-task CNN for fruit freshness assessment	Multi-task learning for freshness detection and fruit-type classification	93.24% mean test accuracy for freshness detection	FPS details were not the main focus	Shows that multi-task learning can improve freshness assessment	Multi-fruit dataset, not specifically designed for apple object detection

The advancement of apple quality inspection systems has involved several methods, ranging from conventional image processing and machine learning classification to deep learning-based object detection. Traditional inspection techniques generally depend on manual sorting or manually designed visual features, making them prone to subjectivity, requiring more time, and less suitable for high-volume postharvest sorting processes. In comparison, the YOLOv12s-based method proposed in this study offers an end-to-end real-time object detection approach that is more applicable for practical apple quality sorting, especially in embedded systems or edge-based environments.

The proposed model achieved an accuracy of 93%, with a precision value of 0.91, recall of 0.89, and F1-score of 0.92. These results indicate that the model provides competitive performance as a lightweight single-stage detector for rotten apple detection. Compared with previous object detection research, such as the YOLOv3-based approach reported in [13], which achieved moderate performance in defect detection, the proposed approach shows stronger potential for rapid sorting applications. Although several classification-based or specialized imaging methods, such as the CNN-based approach in [16] and the pruning-based YOLOv4 method using NIR cameras in [15], achieved high performance, these approaches either lack direct object localization or require additional imaging hardware and more specific deployment conditions.

In addition, studies focusing on specific anatomical regions or multi-task formulations offer complementary perspectives. The optimized YOLOv5s approach in [32], for instance, was primarily designed for stem/calyx recognition rather than fresh-versus-rotten classification, indicating that real-time object detection in apple imaging has mainly been explored for localized structural features rather than holistic quality assessment. Similarly, the multi-task CNN in [19] demonstrated that freshness detection can be combined with fruit-type classification, but this approach was developed for a multi-fruit dataset and was not specifically optimized for apple-focused object detection. These comparisons suggest that the proposed YOLOv12s framework addresses a more direct formulation of the fresh-versus-rotten detection task while retaining real-time, edge-deployable characteristics.

Overall, these findings indicate that YOLOv12s is a promising method for real-time rotten apple detection and postharvest quality sorting because it provides a good balance between detection accuracy and computational efficiency. However, the reduced performance under challenging conditions, such as low illumination, occlusion, and diverse background settings, suggests that further improvement is still needed before large-scale practical implementation. Future studies should focus on expanding the dataset, increasing environmental variability, and improving model robustness so that the system can generalize more effectively across real-world sorting scenarios while maintaining real-time detection capability.

4. Conclusions and Future Works

This study shows that YOLOv12 can be utilized as a lightweight real-time object detection model for postharvest apple quality sorting. By classifying apples into two categories, fresh apple and rotten apple, the proposed system obtained an accuracy of 0.93, precision of 0.91, recall of 0.89, and F1-score of 0.90 based on the reported confusion matrix. These findings suggest that the model can effectively differentiate fresh apples from rotten ones while also localizing the detected objects in RGB images. The scientific contribution of this study is reflected in the adaptation of an attention-centric YOLOv12 detector for rotten apple detection, supported by evaluation using standard classification metrics and robustness testing scenarios. From a practical perspective, the proposed approach can contribute to automated postharvest sorting systems, particularly through low-cost RGB camera-based inspection and potential edge-device deployment. Nevertheless, further validation is still needed using larger datasets, more diverse lighting conditions, varied backgrounds, different apple varieties, and real conveyor-based sorting environments.

5. References

[1] Y. Tian, Q. Ye, and D. Doermann, "YOLOv12: Attention-Centric Real-Time Object Detectors," *ArXiv*, vol. 2502.12524, 2025.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[3] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 6517–6525. doi: 10.1109/CVPR.2017.690.

[4] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," 2016, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2.

[5] T. Brosnan and D.-W. Sun, "Inspection and grading of agricultural and food products by computer vision systems—a review," *Comput. Electron. Agric.*, vol. 36, no. 2–3, pp. 193–213, Nov. 2002, doi: 10.1016/S0168-1699(02)00101-1.

[6] B. Zhang *et al.*, "Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review," *Food Research International*, vol. 62, pp. 326–343, Aug. 2014, doi: 10.1016/j.foodres.2014.03.012.

[7] J. B. Li, W. Q. Huang, and C. J. Zhao, "Machine vision technology for detecting the external defects of fruits — a review," *The Imaging Science Journal*, vol. 63, no. 5, pp. 241–251, Jun. 2015, doi: 10.1179/1743131X14Y.0000000088.

[8] L. E. Chuquimarca, B. X. Vintimilla, and S. A. Velastin, "A review of external quality inspection for fruit grading using CNN models," *Artificial Intelligence in Agriculture*, vol. 14, pp. 1–20, Dec. 2024, doi: 10.1016/j.aiia.2024.10.002.

[9] V. Leemans, H. Magein, and M.-F. Destain, "Defects segmentation on 'Golden Delicious' apples by using colour machine vision," *Comput. Electron. Agric.*, vol. 20, no. 2, pp. 117–130, Jul. 1998, doi: 10.1016/S0168-1699(98)00012-X.

[10] V. Leemans, H. Magein, and M.-F. Destain, "Defect segmentation on 'Jonagold' apples using colour vision and a Bayesian classification method," *Comput. Electron. Agric.*, vol. 23, no. 1, pp. 43–53, Jun. 1999, doi: 10.1016/S0168-1699(99)00006-X.

[11] Q. Li, M. Wang, and W. Gu, "Computer vision based system for apple surface defect detection," *Comput. Electron. Agric.*, vol. 36, no. 2–3, pp. 215–223, Nov. 2002, doi: 10.1016/S0168-1699(02)00093-5.

[12] V. Leemans and M.-F. Destain, "A real-time grading method of apples based on features extracted from defects," *J. Food Eng.*, vol. 61, no. 1, pp. 83–89, Jan. 2004, doi: 10.1016/S0260-8774(03)00189-4.

[13] P. Valdez, "Apple Defect Detection Using Deep Learning Based Object Detection For Better Post Harvest Handling," *ArXiv*, 2020.

[14] S. Fan *et al.*, "On line detection of defective apples using computer vision system combined with deep learning methods," *J. Food Eng.*, vol. 286, p. 110102, Dec. 2020, doi: 10.1016/j.jfoodeng.2020.110102.

[15] S. Fan *et al.*, "Real-time defects detection for apple sorting using NIR cameras with pruning-based YOLOV4 network," *Comput. Electron. Agric.*, vol. 193, p. 106715, Feb. 2022, doi: 10.1016/j.compag.2022.106715.

[16] Y. Li, X. Feng, Y. Liu, and X. Han, "Apple quality identification and classification by image processing based on convolutional neural networks," *Sci. Rep.*, vol. 11, no. 1, p. 16618, Aug. 2021, doi: 10.1038/s41598-021-96103-2.

[17] M. Agarla, P. Napoletano, and R. Schettini, "Quasi Real-Time Apple Defect Segmentation Using Deep Learning," *Sensors*, vol. 23, no. 18, p. 7893, Sep. 2023, doi: 10.3390/s23187893.

[18] X. Gao *et al.*, “Application of Advanced Deep Learning Models for Efficient Apple Defect Detection and Quality Grading in Agricultural Production,” *Agriculture*, vol. 14, no. 7, p. 1098, Jul. 2024, doi: 10.3390/agriculture14071098.

[19] T. Zhang *et al.*, “Apple varieties, diseases, and distinguishing between fresh and rotten through deep learning approaches,” *PLoS One*, vol. 20, no. 5, p. e0322586, May 2025, doi: 10.1371/journal.pone.0322586.

[20] C. Citra, A. Ritonga, A. Arnita, S. Iskandar Al Idrus, and D. Yandra Niska, “Application of the K-Nearest Neighbor (K-NN) Algorithm for Detecting Banana Harvest Feasibility,” *J-INTECH*, vol. 13, no. 02, pp. 252–261, Dec. 2025, doi: 10.32664/j-intech.v13i02.2064.

[21] D. D. Indriani.S, K. S. S, S. I. Al Idrus, S. Susiana, and A. Perdana, “Identification of Palm Oil Fresh Fruit Bunches Worth Selling with K-Nearest Neighbors Algorithm,” *J-INTECH*, vol. 13, no. 02, pp. 262–271, Dec. 2025, doi: 10.32664/j-intech.v13i02.2066.

[22] E. M. Atsir, N. Nurmalitasari, and A. A. Sari, “Traffic Accident Severity Classification System Using Random Forest Algorithm,” *J-INTECH*, vol. 13, no. 02, pp. 272–281, Dec. 2025, doi: 10.32664/j-intech.v13i02.2089.

[23] A. A. Purnawan and S. D. Sancoko, “Perancangan dan Implementasi FitSphere sebagai Aplikasi Android untuk Manajemen Keanggotaan Gym di Chain Gym,” *J-INTECH*, vol. 13, no. 02, pp. 282–291, Dec. 2025, doi: 10.32664/j-intech.v13i02.2132.

[24] B. A. Wisesa, V. Mahat Putri, E. Faristasari, S. A. Jasman Duli, and S. Agustin, “Preventive Attendance Record using Photo from Mobile Phone and Printed Paper using CNN,” *J-INTECH*, vol. 13, no. 01, pp. 117–127, Jun. 2025, doi: 10.32664/j-intech.v13i01.1927.

[25] B. A. Wisesa, W. Andriyani, and B. D. P. Purnomosidi, “Usage of LSTM Method On Hand Gesture Recognition For Easy Learning of Sign Language Based On Desktop Via Webcam,” in *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, IEEE, Dec. 2022, pp. 148–153. doi: 10.1109/ISRITI56927.2022.10053076.

[26] B. A. Wisesa, W. Andriyani, T. Suprawoto, and Hamdani, “Development of Learning Media for The Deaf Using a Webcam,” in *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, IEEE, Dec. 2022, pp. 160–165. doi: 10.1109/ISRITI56927.2022.10052934.

[27] J. Xing and J. De Baerdemaeker, “Bruise detection on ‘Jonagold’ apples using hyperspectral imaging,” *Postharvest Biol. Technol.*, vol. 37, no. 2, pp. 152–162, Aug. 2005, doi: 10.1016/j.postharvbio.2005.02.015.

[28] D. Unay and B. Gosselin, “Automatic defect segmentation of ‘Jonagold’ apples on multi-spectral images: A comparative study,” *Postharvest Biol. Technol.*, vol. 42, no. 3, pp. 271–279, Dec. 2006, doi: 10.1016/j.postharvbio.2006.06.010.

[29] D. Unay and B. Gosselin, “Stem and calyx recognition on ‘Jonagold’ apples by pattern recognition,” *J. Food Eng.*, vol. 78, no. 2, pp. 597–605, Jan. 2007, doi: 10.1016/j.jfoodeng.2005.10.038.

[30] A. Mizushima and R. Lu, “An image segmentation method for apple sorting and grading using support vector machine and Otsu’s method,” *Comput. Electron. Agric.*, vol. 94, pp. 29–37, Jun. 2013, doi: 10.1016/j.compag.2013.02.009.

[31] Y. Yu, S. A. Velastin, and F. Yin, “Automatic grading of apples based on multi-features and weighted K-means clustering algorithm,” *Information Processing in Agriculture*, vol. 7, no. 4, pp. 556–565, Dec. 2020, doi: 10.1016/j.inpa.2019.11.003.

[32] Z. Wang, L. Jin, S. Wang, and H. Xu, "Apple stem/calyx real-time recognition using YOLO-v5 algorithm for fruit automatic loading system," *Postharvest Biol. Technol.*, vol. 185, p. 111808, Mar. 2022, doi: 10.1016/j.postharvbio.2021.111808.

[33] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2017, pp. 2999–3007. doi: 10.1109/ICCV.2017.324.

[34] X. Liang *et al.*, "Real-Time Grading of Defect Apples Using Semantic Segmentation Combination with a Pruned YOLO V4 Network," *Foods*, vol. 11, no. 19, p. 3150, Oct. 2022, doi: 10.3390/foods11193150.

[35] X. Hu *et al.*, "Automatic Detection of Small Sample Apple Surface Defects Using ASDINet," *Foods*, vol. 12, no. 6, p. 1352, Mar. 2023, doi: 10.3390/foods12061352.

[36] J.-H. Lee, H.-T. Vo, G.-J. Kwon, H.-G. Kim, and J.-Y. Kim, "Multi-Camera-Based Sorting System for Surface Defects of Apples," *Sensors*, vol. 23, no. 8, p. 3968, Apr. 2023, doi: 10.3390/s23083968.

[37] H. Si *et al.*, "Apple Surface Defect Detection Method Based on Weight Comparison Transfer Learning with MobileNetV3," *Agriculture*, vol. 13, no. 4, p. 824, Apr. 2023, doi: 10.3390/agriculture13040824.