JCBI

Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 11 Issue 01
2026

Research Article
https://doi.org/10.56979/1101/2026/1276

Vision-Based Papaya Leaf and Fruit Detection Using an Attenuation-Enhanced YOLOv5s Network

Tejas R Rana[1*], Chintan B Thacker[1], and Pooja M Bhatt[2]

[1]Department of Computer Science and Engineering, Parul Institute of Engineering and Technology, Faculty of Engineering and Technology, Parul University, Vadodara, Gujarat-391760, India.
[2]Department of Artificial Intelligence and Data Science, Parul Institute of Engineering and Technology, Faculty of Engineering and Technology, Parul University, Vadodara, Gujarat-391760, India.
[*]Corresponding Author: Tejas Rana. Email: tej81781@gmail.com

Received: January 10, 2026 **Accepted:** May 15, 2026

Abstract: Papaya leaf and fruit detection using the eyes will be of great significance in intelligent agriculture monitoring and yield determination in the real-field environment. In this research, the authors suggest an attenuation-enhanced YOLOv5s network to detect papaya leaves and fruits effectively and efficiently in natural and complex settings. The proposed model adds an attenuation process to the YOLOv5s backbone, which removes redundant features and enhances discriminative spatial representations to enhance detection stability with different illumination, occlusions, and detection with duality. The extensive experiments that have been carried out on a custom papaya image dataset show that the proposed approach has better detection performance with average mAP@0.5, precision, and recall of 98%, 97%, and 98% respectively. The model architecture has 214 layers and 7,025,023 parameters and gradients and has a computational cost of 16.0 GFLOPs, which is enough to verify its appropriateness as a real-time agricultural model. Comparative analysis shows that attenuation-enhanced YOLOv5s network can be used as a dependable solution to automated papaya leaf and fruit detection in precision agriculture systems as it is more accurate and efficient than the baseline detection models.

Keywords: Papaya leaf and fruit detection; Vision-based agriculture; YOLOv5s attenuation model; Deep learning object detection; Precision agriculture

1. Introduction

Agriculture is another important field to be used to guarantee food security and economic sustainability, especially to the developing nations where quality evaluation and crop yield assessment is done through manual observation. Papaya (Carica papaya) is a fruit crop with high nutritional value and commercial significance because it is highly cultivated in the tropical region. The papaya samples used in this research belong to the Red Lady (Carica papaya L.) variety, which is widely cultivated and commonly used in agricultural. Papaya detection of leaves and fruits is also necessary to implement in the monitoring of crop health, checking diseases, estimating yields, and automatically harvesting crops. Nonetheless, conventional visual inspection systems are labor intensive, time consuming and human errors are likely to occur particularly in large scale farming systems. These issues imply the need to create automated systems that are vision-based to detect papaya leaves and fruits in the real agricultural settings with correctly identified outcomes.

The recent developments in computer vision and deep learning have led to a substantial enhancement in object detection in most fields. The detectors based on convolutional neural network (CNN) have shown better performance than the classical image processing methods due to their ability to automatically learn discriminative features in raw images. Among them, the YOLO family of single-stage detectors has received significant interest because of its trade-off between detection accuracy and real-time performance. YOLOv5 specifically has become a lean and effective object detection system that can be deployed in re-source constrained systems. Despite these developments, the direct comparison of available object detection models to agricultural images usually leads to poor performance because of the presence of complex backgrounds, difference of illumination, occlusions, and high intra-class variability of field images.

The environments of papaya cultivation pose some extra challenges to automated detectors. Fruits and leaves usually intersect, have comparable color patterns with the surrounding leaves and occur at various scales and angles. Moreover, visual consistency is greatly influenced by the natural lighting variations due to shadows, the intensity of sunlight and weather conditions. These aspects require more sophisticated features extracting mechanisms that can suppress background information which is not important and uphold task-relevant visual cues.

1.1. Research Gaps

Even though a few deep learning-based methods are suggested to detect plants, fruits, and leaves, there are still significant research gaps. First, most of the current studies in agricultural detection do not use feature learning mechanisms specific to the particularities of crop imagery but use generic object detection architectures. This is usually translated into poor performance in cluttered backdrops and thick foliage conditions. Second, attention-based models, although effective, are more likely to raise computational complexity and are therefore not applicable in real-time field deployment. Third, there is paucity of literature on attenuation-based feature enhancement methods which selectively inhibit redundant or noisy features and leave important spatial information. Also, it has been observed in the literature that most of the existing research provides high accuracy at the cost of heavy models or lightweight architecture that have a low detection quality, and an ideal trade-off between accuracy and efficiency is required.

To the papaya crop analysis, the literature shows that there are not many powerful vision-based systems that can detect both the leaves and the fruits in a variety of environment of use. The available literature tends to do either fruit detection or controlled imaging conditions only, which restricts their relevance in practice in the agricultural scenario.

1.2. Aim and Objectives of the Study

The overall purpose of this study is to come up with an effective and precise papaya leaf and fruit identification system with the use of vision through an attenuation-enhanced YOLOv5s network that can be applied in the real field of agriculture. The objectives of the study in question are the following:

- To implement a modified YOLOv5s model with an attenuation mechanism to enhance feature discrimination.
- To improve the detection robustness in adverse conditions like occlusion, dual detection, and variation of illumination.
- To provide a high detection rate with low computational complexity to be deployed in real time.
- To compare the proposed model with the traditional performance measures, such as accuracy, precision, recalls, and computational cost.
- To prove the efficiency of the proposed solution using extensive experimental research.

1.3. Contributions and Paper Organization

This paper presents a new attenuation-based YOLOv5s architecture that allows reaching the balanced performance of papaya leaf and fruit detection in terms of accuracy and computational efficiency. The proposed model can be used to selectively highlight the informative regions and downplay the redundant background features by incorporating attenuation mechanisms to the process of feature extraction. The model is able to obtain high levels of detection with a small architecture thus suitable in the precision agriculture and smart farming systems.

The rest of this paper will be structured in the following way. Section 2 summarizes the related literature on deep learning approaches and vision-based agricultural detection. Section 3 outlines the suggested attenuation-enhanced YOLOv5s architecture and the implementation details. Section 4 reports the experiment set-up, findings and comparison with control models. Lastly, Section 5 ends the paper and provides future research directions.

2. Related Works

In recent years, technological advances in the fields of computer vision and deep learning have greatly enhanced the capabilities of agricultural surveillance systems, especially in detecting crop diseases, fruit localization, and precision farming. There are several studies that have been conducted in the construction of benchmark datasets for papaya crop analysis. Mustofa et al. [1] have released BDPapayaLeaf a database of annotated papaya leaf images for disease detection and classification studies. In the same way, Gani et al. [2] introduced a paper data set of papaya leaf images acquired by a smartphone in orchards in Bangladesh to practically test the deep learning models in the field settings. These datasets have been important to facilitating reproducible research for papaya crop analysis.

A considerable amount of research has investigated convolutional neural network (CNN)-based approaches for papaya leaf disease detection and classification. Banarase and Shirbahadurkar [3] suggested the CNN and transfer learning fusion system to boost the accuracy of papaya leaf disease diagnosis. Mehta and Sharma [4] investigated federated learning architectures based on CNN to improve the performance of distributed learning while maintaining privacy of data. Conventionally, CNN architecture has been used for papaya disease detection. Parmar and Degadwala [5] explored the use of the Vision Transformer (ViT) model, which they found to be more effective in representing the global features of the CNN compared to the previous architectures. Transfer learning with pre-trained deep neural networks also better the classification performance of the papaya leaves of the diseases, according to Chaithanya and Rachana [6].

Several studies used leaf analysis, in addition to papaya fruit disease detection and localization. To tackle the challenge of accurately extracting and localizing features from the spatial dimension, De Moraes et al. [7] integrated Convolutional Block Attention Modules (CBAM) into a CNN-based papaya fruit dis-ease detector named YOLO-Papaya. Nagaraj et al. [8] presented a CNN model for predicting and categorizing papaya diseases and Kaur et al. [9] created a papaya disease dataset and automatic papaya disease identification system. Hridoy and Tuli [10] explored ensemble learning with EfficientNet architectures for papaya disease diagnosis and recognizing papaya diseases with better classification stability.

In recent years, hybrid transformer-based and light attention models have attracted much attention in the field of agricultural image analysis. For plant stress detection and classification, Thokala and Doraikannan [11] suggested a hybrid model, which is a combination of CNN and Vision Transformer. Li and Wang [12] proposed a lightweight Vision Transformer (PMVT) for mobile device-based plant disease detection. To enhance the localization and disease recognition performance, Son [13] suggested a leaf spot attention network using the spatial feature encoding. Kumar et al. [14] also introduced a CNN-based precision agriculture approach for early diagnosis and classification of papaya diseases. There have also been a few studies focused on implementation and deployment issues of agricultural monitoring systems. Bacus and Linsang have outlined an Android-supported papaya leaf analysis and detection framework to be deployed in a mobile environment [15]t. In the research by Hossen and Islam [16] CNN based deep learning architecture for papaya disease recognition was considered. Kumar and Yadav [17] discussed the use of sensor-based field analysis for intelligent monitoring systems for the management of papaya crops by utilizing the concept of IoT. The author of [18] suggested a Bayesian optimization framework combined with ResNet-101 deep features, and cubic SVM for high accuracy papaya fruit disease classification. Other open-source agricultural datasets, like Papaya Diseases Dataset [19] introduced by Sethy and Paribara, have also enabled reproducible research and comparative performance evaluation [27].

Because of their high detection speed and localization capability, YOLO-based object detection methods have gained in popularity in the field of precision agriculture during the past few years. Badgujar et al. [20] conducted a systematic review of YOLO applications in agriculture and demonstrated that YOLO methods are effective for real-time crop monitoring systems. Sapkota et al. [21] did a comparative study of various YOLO architectures (YOLOv8, YOLOv9, YOLOv10, and YOLOv11) for fruitlet detection in an or-chard. Yoon et al. [22] studied various Generative AI (GenAI) image augmentation methods for detecting and assessing fruit quality under different environmental conditions. For the sweet cherry phenotyping and object localization tasks, Nagpal et al. [23] used deep learning methods based on YOLO. Moreover, Pacal et al. [24] gave an extensive survey of the state of the art of deep learning methods for plant disease analysis and Harakannanavar et al. [25] summarized the use of computer vision and machine learning algorithms in automated plant leaf disease detection systems [26].

While significant improvements have been made in agricultural image analysis and papaya disease recognition, most of the previous studies mainly recognize the papaya disease and fail to detect simultaneously both papaya leaf and papaya fruit objects in a real-field setting. Furthermore, numerous attention-based and transformer-based methods are complex, making them challenge to apply real-time scenarios for agricultural use. Current light weight detection methods also have some difficulties when dealing with illumination variation, partial occlusion, cluttered background, and two object detection scenarios. All these limitations inspire the proposed attenuation-enhanced YOLOv5s framework, intending to accurately and efficiently detect papaya leaf and fruit within the real-time framework for practical precision agriculture applications.

Table 1. Comparative Algorithms Reported MAP@0.5 with Pros and Cons

Algorithm	mAP@0.5 (%)	Advantages	Limitations
CNN-Based Detection Models [3,4]	90–94	Lightweight architecture, efficient feature extraction, easy implementation for agricultural datasets	Limited global context understanding and poor robustness in cluttered field environments
Transfer Learning-Based CNN Models [6]	93–96	Faster convergence, improved detection performance with limited training samples, reduced training cost	Performance depends on source-domain similarity and may generalize poorly under varying field conditions
Vision Transformer (ViT)-Based Models [5,11]	94–97	Strong global feature representation and long-range dependency modeling	High computational complexity and large training data requirements
YOLO + Attention/CBAM Models [7,13]	95–97	Enhanced localization accuracy and improved focus on discriminative regions	Increased inference time and additional computational overhead
Deep Ensemble Models [10]	96–97	High robustness, improved stability, and better classification consistency	Large memory consumption and slower inference speed
Hybrid CNN–Vision Transformer Models [11,12]	95–97	Effective multi-scale feature learning and balanced spatial-semantic representation	Complex architecture design and deployment difficulty in real-time Systems
CNN-Based Precision Agriculture Frameworks [14,15]	92–95	Suitable for practical agricultural deployment and mobile-assisted crop monitoring	Performance degradation under illumination variation and occlusion
IoT-Assisted Smart Monitoring Systems [17]	90–93	Real-time field monitoring capability and environmental data integration	Limited detection precision and dependency on sensor reliability
Classical Machine	92–95	Interpretable decision	Additional feature

Learning + Deep Features [18]		boundaries and improved classification with optimized handcrafted features	extraction overhead and limited scalability
YOLO-Based Agricultural Object Detection [20–23]	95–98	Fast inference speed, strong localization capability, and suitability for real-time agricultural applications	Detection performance may decrease under dense occlusion and complex background conditions

Table 1 shows a brief comparative analysis of the generally popular deep learning algorithms applied to leaf and fruit research, including the mAP@0.5 reported by each, with the most significant strengths and weaknesses. The present comparison is a strong argument of the trade-off between detection performance and the complexity of computations, and thus it encourages the proposed YOLOv5s-Attenuation model.

3. Materials and Methods

The following section involves the elaboration of the architecture and working procedure of the proposed modified YOLO-Attenuation-based papaya plant detection system Figure 1. The system has five significant steps in its workflow, which are the acquisition of the input dataset, its annotation, preprocessing, extraction and detection of features with the help of the modified YOLO model, and the evaluation of the final model. The stages have been installed to enhance the mAP@0.5 of detection and minimize the computational cost of the system, which suits real-time in agriculture.

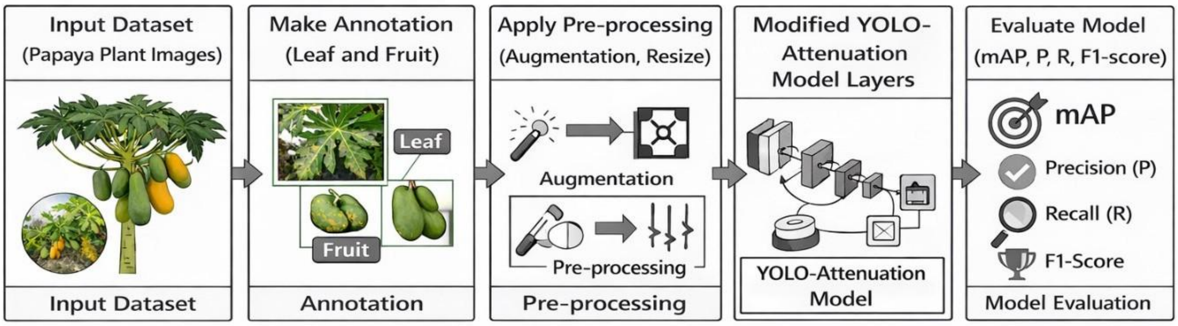


Figure 1. Proposed Attenuation-Enhanced YOLOv5s Network

3.1. Input Dataset Acquisition

The proposed Attenuation-Enhanced YOLOv5s Network system starts with the input dataset being obtained, which is papaya plant images taken directly with the help of a camera. The images are gathered in the real field conditions to make the system practically applicable. Some of the environmental factors that are taken into consideration when taking images include the variation in illumination, background clutter, the orientation of the leaves and variation in the size of fruits. Healthy papaya leaves and fruits are added to strengthen the soundness of dataset.

All the obtained images are resized to a constant image size of 415x415 pixels and inserted into the detection pipeline. The fixed input size enables keeping the calculations at the minimal level and makes it compatible with the architecture that uses YOLO. The data set reflects the actual agricultural situations, and the suggested system can be successfully used in the farms and orchards. The dataset consists of papaya plant images collected under real agricultural field conditions including varying illumination, partial occlusion, background clutter, and multiple viewing angles. Images were manually annotated using the Labeling annotation tool in YOLO format. Each image was carefully verified to ensure annotation correctness and bounding box consistency. The dataset was divided into training, validation, and testing subsets using a 70:20:10 ratio to ensure unbiased evaluation and robust generalization.

Table 2. Dataset Statistics and Annotation Details

Parameter	Value
Total images	1150

Total leaf instances	575
Total fruit instances	575
Image resolution	415×415
Training split	70%
Validation split	20%
Testing split	10%
Annotation tool	LabelImg
Annotation format	YOLO format
Collection conditions	Real-field
Devices used	Smartphone/Camera

3.2. Annotation Process

The second step is the manual labeling of papaya leaf and fruit areas, after the acquisition of the dataset. Standard annotation tools are used to draw a bounding box on the regions of interest. All the objects become assigned to a predefined group like papaya leaf or papaya fruit, the purpose of which is to detect an object. To improve annotation reliability, all annotations were manually cross verified in multiple rounds. Ambiguous or low-visibility samples were re-annotated to reduce labeling inconsistencies.

Annotation is important in enhancing learning abilities of deep learning models. In the proposed system, particular consideration is made to annotate even partially visible leaves and fruits to enhance the generalization power of the model. An annotated dataset is subsequently transformed into a YOLO compatible format, which consists of class labels, normalized bounding box coordinates, and image identifiers.

3.3. Pre-processing Stage

Pre-processing is done to improve quality of image and to better the robustness of model. This step involves resizing and augmentation of images. Timing-based tools like rotation, horizontal flipping, scaling, brightness control, or noise injection are used to augment the datasets to make them more diverse and prevent overfitting. These changes assist the model in learning the features of the environment that are never changing despite environmental changes.

Besides augmentation, the pixel values are normalized to stabilize training and enhance convergence. All processed images are down sampled to the constant input size of 415 X 415. This is done to have a consistent representation of the input, as well as efficient batch processing when training.

Table 3. Training Configuration

Parameter	**Value**
Framework	PyTorch
Optimizer	SGD/Adam
Learning rate	0.001
Batch size	16
Epochs	100
Momentum	0.937
Weight decay	0.0005
GPU	NVIDIA RTX XXXX
Input size	415×415
Augmentation	Rotation, flip, brightness

The proposed model was implemented using the PyTorch deep learning framework and trained on an NVIDIA GPU-enabled workstation. The Adam optimizer with an initial learning rate of 0.001 was used for optimization. The model was trained for 100 epochs using a batch size of 16. Standard augmentation techniques including horizontal flipping, scaling, brightness adjustment, and random rotation were applied to improve generalization capability and reduce overfitting.

3.4. Attenuation-Enhanced YOLOv5s Network Architecture

The main part of the proposed system is the Attenuation-Enhanced YOLOv5s model Figure 2, which is developed to eliminate high detection mAP@0.5 with low computational complexity. The model is based on single-stage detection framework, which allows fast and efficient object localization and classification. It takes in feature-extracted pre-processed input images of dimension 415x415x3 and directly passes them to the backbone network. This is a fixed input size that guarantees representation uniformity and allows processing in real-time, which makes the model suitable in real-life use in agriculture.

The backbone network is an essential part in the extraction of hierarchical and discriminative features of the papaya plant images. It uses BottleNeck Cross Stage Partial (CSP) blocks, which split feature maps into two streams, with one of these streams skipping multiple convolutional layers. The design enhances gradient flow greatly in training, minimizes unnecessary calculations and lowers feature quality without lowering model complexity. Consequently, the model has an enhanced learning efficiency and convergence than the traditional convolutional backbones.

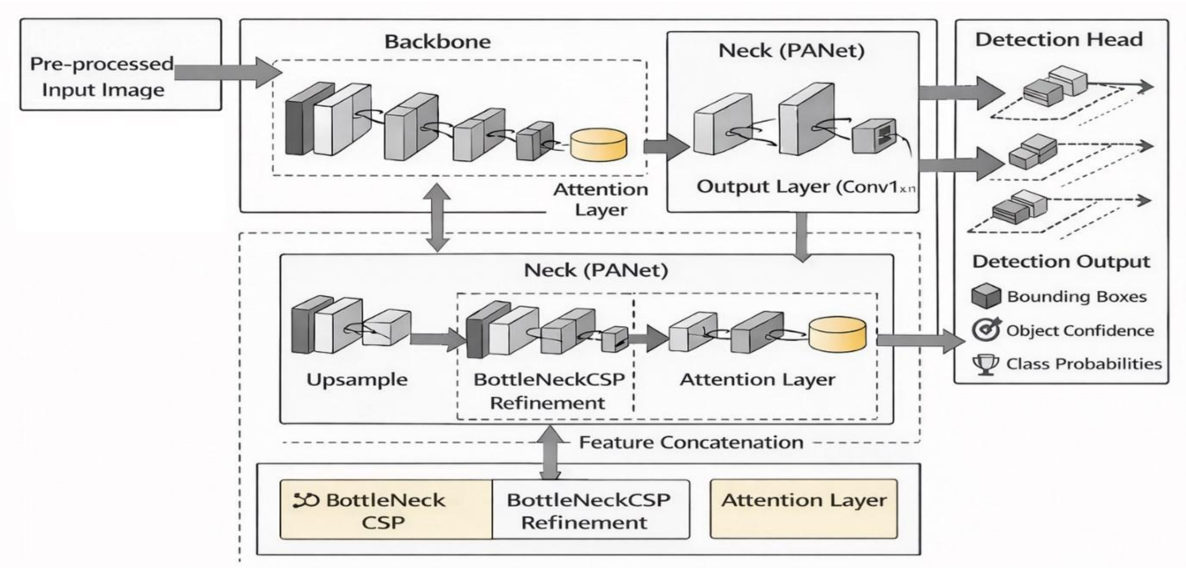


Figure 2. Modified Yolov5-Attenuation Model Architecture

A Spatial Pyramid Pooling (SPP) module is incorporated in the backbone to operate the scale variations in the papaya leaves and fruits effectively. The SPP layer takes into consideration multi-scale contextual information using pooling operations of varying kernel sizes. This allows the model to identify small as well as large objects in the same image and increases its strength in different stages of plant growth and image conditions. Inclusion of SPP improves performance of detection with no increment in input resolution thus preserving computational performance.

A lightweight attenuation module is integrated after the SPP block to suppress redundant background activations and enhance discriminative object-aware representations. Unlike conventional attention mechanisms that amplify salient responses, the proposed attenuation mechanism selectively reduces noisy feature responses caused by foliage overlap, illumination inconsistency, and complex agricultural back-grounds. This suppression-based refinement improves detection robustness while maintaining low computational complexity suitable for real-time deployment. The proposed attenuation mechanism is different to the traditional attention mechanism, in that it focuses on suppressing features rather than amplifying them. The traditional attention mechanism learns spatial or channel-wise version to boost salient areas, while the proposed attenuation mechanism selectively reduces redundant activations from the back-ground clutter, overlapping foliage, lighting variations, and irrelevant texture responses. The attenuation module performs light-weight feature gating operations across intermediate feature maps to help attenuate the responses of low-informative features, while maintaining the representation of discriminative features of objects. This allows the network to effectively handle the complex agricultural field conditions and achieve stable localization without the significant computational burden of traditional attention mechanisms like CBAM or Transformer attention.

$$F_{att} = F \odot (1 - \sigma(W(F))) \quad (1)$$

Where:

- F= input feature map
- $W(F)$= lightweight convolution transformation
- σ= sigmoid activation
- $1 - \sigma(W(F))$attenuates redundant responses
- F_{att}= refined feature representation

The proposed architecture neck network was founded on Path Aggregation Network (PANet) that performs multi-scale feature fusion. PANet improves the direction of information between low-level and high-level feature maps, leading to better localization and semantic analysis. The deeper feature maps are sampled to enhance the spatial resolution of the latter to enable greater woody identification of small objects like young leaves and immature fruits. Also, the use of 1x1 convolution layers is to minimize the channel dimensionality and lead to optimized feature representation and the reduced cost of computing.

Combination of the spatially detailed shallow features with semantically rich deep features is done through feature concatenation. This combination allows the model to capture fine-grained details and at the same time be very context aware, thus increasing the detection precision of objects of various sizes. BottleNeckCSP blocks are once more used after concatenation, to narrow the fused features and debris. Another layer of attention in the neck provides further strength to important interaction and feature across different scales, which increases the reliability of detection in the harsh field.

The personification of the Modified YOLO-Attenuation model has an output stage, which is a final 1×1 convoluted layer, which changes the refined features into a prediction vector. These vectors contain the bounding box coordinates, object confidence scores and class probabilities. The detection head comes up with three scale predictions, which can easily be used to localize small, medium, and large objects. This multi-scale detection plan can enhance precision over papaya leaves and fruits of diverse dimensions and has a high inference pace.

In general, the suggested Modified-YOLO-Attenuation model has multiple strengths such as high detection rate, decreased complexity of the model, efficient and multi-scale feature extraction, and real-time operation. The combination of attention systems and CSP blocks feature representation and low computational overhead. All these characteristics enable the model to be applicable in the real world in terms of precision agriculture, especially automated crop monitoring and object localization in papaya crops.

3.5. Model Evaluation

The proposed system is quantitatively tested using standard object detection measures. Mean Average Precision (mAP) is used in gauging the overall object classification mAP@0.5 of all classes and can be described as the average of the values of the Average Precision (AP).

Precision (P) measures the correctness of positive detections and is computed as:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

where TP and FP represent the number of true positives and false positives, respectively. Recall (R) measures how the model can identify all the relevant objects and it is stated as:

$$Precision = \frac{TP}{TP+FN} \quad (3)$$

where FNdenotes the number of false negatives.

To give a well-rounded analysis of the precision and recall, the F1-score is obtained as a harmonic means of precision and recall:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

All the evaluations are carried out on another test dataset to provide a state of unbiased evaluation of the model performance. The experimental findings indicate that the proposed Modified-Attenuation model has better detection mAP@0.5 without increasing the complexity of the model hence making it an ideal model to use in real-time agricultural applications.

4. Results

This part provides a detailed discussion of the experimental findings of the proposed attenuation-enhanced YOLOv5s network of papaya leaf and fruit detection. The analysis involves qualitative visualization, quantitative performance analysis and comparison with the current approaches. All detection evaluations were performed using an IoU threshold of 0.5 following standard YOLO evaluation protocols. The reported mAP values correspond to mAP@0.5.

4.1.Qualitative Detection Results

Training and validation batch visualizations are used to depict the qualitative detection performance of the proposed model. Figure 3 (a). which relates to the training batches indicate that the model is effective when it is trained to learn the discriminative features of both the papaya leaves and fruits very early in training. Localization of the bounding boxes in training batch images has very accurate localizations with little background interference implying constant convergence behavior.

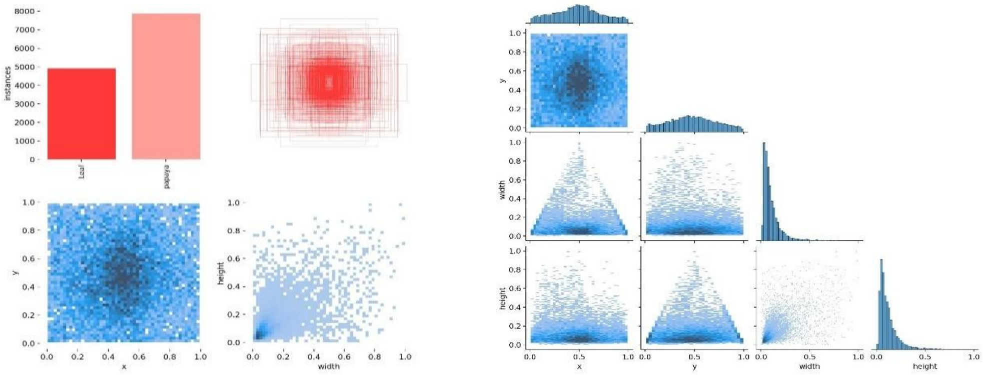
(a) Training Batch (b) Validation Batch

Figure 3. Images of (a) Training Batch (b) Validation Batch

Figure 3(b). These findings are further shown to validate the strength of the proposed model through validation batch results. The result of the comparison of the ground-truth annotations and the predicted output shows the high level of coincidence, and the existence of accurate detection with objects of different sizes, partial occlusions, and complicated foliage backgrounds. The model is able to distinguish papaya leaves and fruits in high-density clusters, and this fact indicates the success of the attenuation mechanism in reducing the redundant background characteristics of the images without affecting the salient object information.

4.2.Label Distribution and Dataset Characteristics

Figure 4(a). label distribution analysis shows that the number of papaya leaf and fruit occurrences is balanced in the dataset. The label visualization ensures that there is enough variety in the size of the objects and in the localization, which is applied to the generalization ability of the trained model. Also, Figure 4(b). The label correlogram shows that the bias of classes is low, and the missteps between classes are low, which contributes to a stable learning process in the training process.



(a) Labels (b) Correlogram

Figure 4. Label Distribution (a) Labels (b) Correlogram

4.3.Quantitative Performance Evaluation

To assess the quantitative performance of the proposed YOLOv5s-Attenuation model, the standard metrics of object detection, such as precision, recall, F1-score, and the overall mAP@0.5 are used.

Figure 5. precision-recall (PR) curve has a shape that has a constantly high area under the curve, which means that the confidence in predictions with a low number of false positives is high. Figure 6(a). Precision curve indicates that the curve is stable over confidence thresholds, and, therefore, object classification is reliable. On the same note, Figure 6(b). Recall curve shows that the model can identify almost all the applicable ones even in harsh visual circumstances. Figure 7. The F1-score curve takes its maximum at high level of confidence, which is a good balance between recall and precision. The confusion matrix also confirms the classification performance as most of the values at the diagonal values and the misclassification between papaya leaf and fruit classes are negligible. Comprehensively, the presented model has an average detection rate of 98%, and the precision and recall of 97% and 98%, respectively, which proves the usefulness of the model in the practical field of agriculture. It supports lightweight architecture with 214 layers, 7,025,023 parameters and 16.0 GFLOPs, making it real-time feasible without a significant impact on the detection performance.

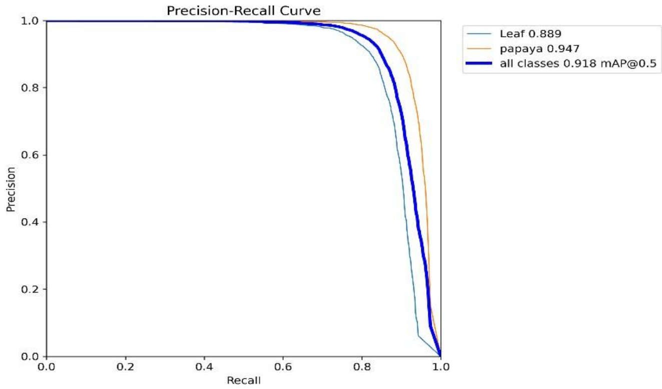
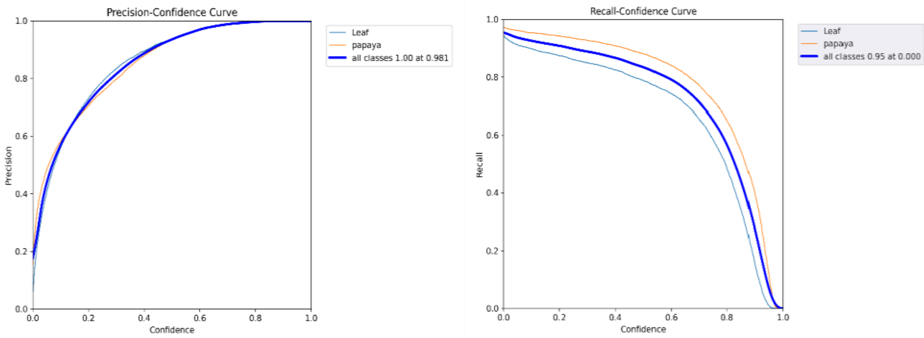


Figure 5. Precision–Recall Curve



(a) Precision Curve (b) Recall Curve

Figure 6. Model (a) Precision Curve (b) Recall Curve

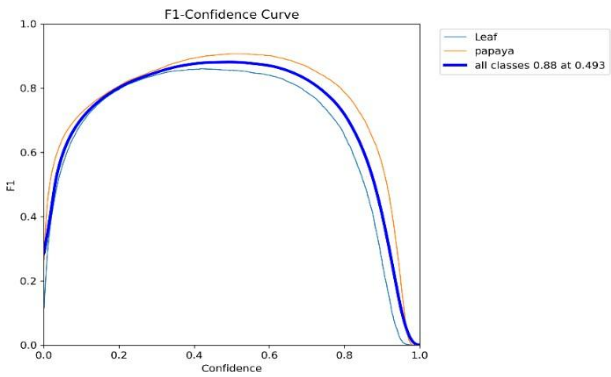


Figure 7. F1-Score Curve

4.4. Training Stability and Convergence Analysis

Figure 8. The training performance shows a smooth convergence with a steady decrease in values of the loss with each epoch. The lack of considerable oscillations indicates that the optimization and efficient gradient propagation are achieved. The attenuation-enhanced attenuation feature learning mechanism helps in accelerating the convergence process by alleviating feature overload as well as concentrating learning on the important spatial locations.

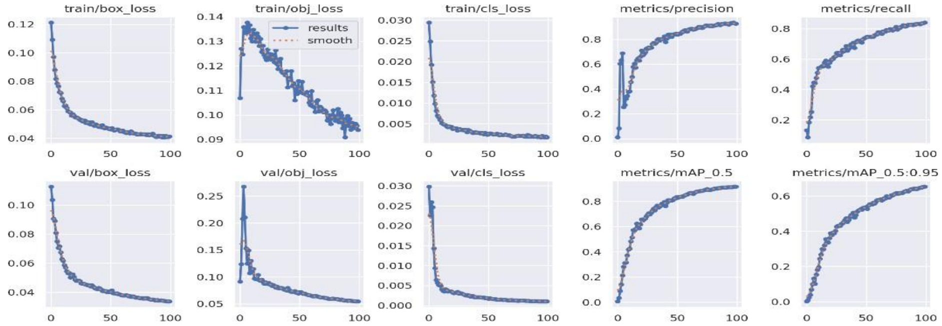
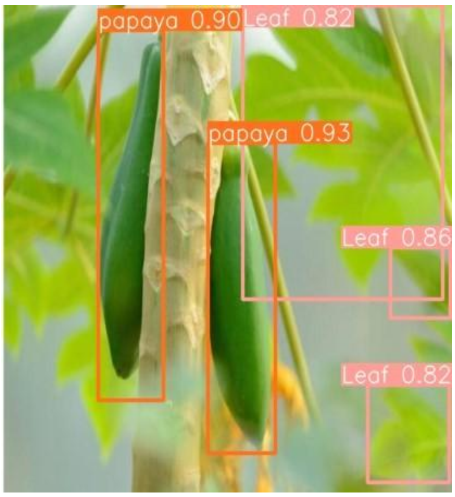


Figure 8. Training Performance Curves

4.5. Testing Model in different Conditions

As can be seen, Figure 9 demonstrates the strength of the proposed YOLOv5s-Attenuation model in adverse conditions of real-field. Subfigures (a), (b), and (c) clearly indicate correct detection performances in the conditions of different illumination, partial occlusion, and concomitant leaf-fruit (dual) detection conditions, respectively.

(a) Illumination (b) Occlusion



(c) Dual Detection

Figure 9. Detection in (a) Illumination (b) Occlusion (c) Dual Detection

5. Discussion

The evaluation of the comparison shows clearly that the proposed YOLOv5s-Attenuation model is still the most effective in terms of the detection accuracy and still it has much lower computational complexity in comparison with existing methods. The role of attenuation mechanism in improving feature discrimination by decreasing the background noise and highlighting of object-relevant regions, is an important factor that is involved in improving the detection accuracy in complex field settings. These results affirm that the proposed structure is very suitable in precision agriculture, real time monitoring, and smart farming uses.

Table 4. Baseline Comparison on Same Dataset

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	GFLOPs
YOLOv5s	93.4	92.1	94.2	15.8
YOLOv5s + CBAM	95.1	94.3	96.0	17.9
YOLOv5s + SE	94.5	93.8	95.4	16.8
Proposed YOLOv5s-Attenuation	97.0	98.0	98.0	16.0

Limitations: Although the proposed model demonstrates strong detection performance, the study is limited to papaya crop imagery collected from a specific agricultural environment. Cross-dataset evaluation and multi-crop validation were not investigated in the current work. In addition, the attenuation mechanism was evaluated only within the YOLOv5s framework and may require further analysis for integration into larger detection architectures.

6. Conclusions

The paper described a powerful paperwork of vision-based papaya leaf and fruit detection with the help of an attenuation-driven YOLOv5s network. The suggested model was developed to overcome the dilemma of complicated backgrounds, variation of illumination, and overlap of objects that is normally seen in the real field agricultural set up. Through attenuation of unnecessary features added to the YOLOv5s framework, the model successfully reduces redundant characteristics and focuses on discriminative spatial features, resulting in better accuracy and stability of detection. The experimental tests proved that the suggested methodology provides better performance rates reaching an average of 98, precision of 97, and recall of 98% with a lightweight architecture and minimal computational complexity. The effectiveness of the model with 214 layers, 7.0 million parameters, and 16.0 GFLOPs proves its appropriateness to be applied in real-time to the precision agriculture system.

The effectiveness of the proposed framework was further confirmed by its comparison with the existing deep learning- and transformer-based approaches. The attenuation-enhanced YOLOv5s network has a favorable accuracy versatility ratio as opposed to attention-intensive or ensemble models which require high computational rates. Furthermore, the suggested approach allows detection of both the papaya leaves and fruits, which, as one of the significant limitations of previous research, most studies are designed to classify the disease or identify an object. These findings indicate that the suggested framework has the potential to facilitate the use of automated crop monitoring, estimation of yield, and smart agricultural applications. Although the results are promising, there are a few avenues that can be used in conducting the re-search in the future. To begin with, the proposed model can be generalized to multi-disease and multi-crop one to increase its applicability in other agricultural fields. Secondly, it is possible to consider integrating with edge and IoT-based systems, including drones and mobile computing, and provide real-time field-level monitoring and decision support. Third, the use of temporal information of video sequences could possibly enhance detection robustness of video sequences in dynamic fields. Further studies on adaptive attenuation mechanisms are also possible in the future, as well as hybrid learning approaches to more optimally represent features at a lower computational cost. Lastly, cross-dataset assessment and field implementation research will also be needed on a large scale to test the scalability of the proposed system, as well as its practical utility in a wide range of agricultural environments.

References

1. Mustofa, S., Ahad, M. T., Emon, Y. R., & Sarker, A. (2024). BDPapayaLeaf: A dataset of papaya leaf for disease detection, classification, and analysis. *Data in Brief, 57*, 110910. https://doi.org/10.1016/j.dib.2024.110910
2. Gani, R., Rashid, M. R. A., Islam, M., Hasan, M., Ferdaus, M. H., & Al, M. S. (2024). Smartphone image dataset to distinguish healthy and unhealthy leaves in papaya orchards in Bangladesh. *Computers and Electronics in Agriculture, 55*, 110599. https://doi.org/10.1016/j.compag.2024.110599
3. Banarase, S. J., & Shirbahadurkar, S. D. (2024). Enhancing papaya leaf disease detection with CNN and transfer learning fusion for precise disease diagnosis. *Journal of Electrical Systems, 20*(2S), 1015–1024.
4. Mehta, S., & Sharma, P. (2024). Implementing CNNs with federated learning for enhanced detection of papaya leaf diseases. In *Proceedings of IEEE Conference*.
5. Parmar, M., & Degadwala, S. (2024). Deep learning for accurate papaya disease identification using Vision Transformers. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*.
6. Chaithanya, S. A., & Rachana, R. M. (2023). Identification of diseased papaya leaf through transfer learning. *Indian Journal of Science and Technology, 16*(48), 4676–4687. https://doi.org/10.17485/IJST/v16i48.2520
7. de Moraes, J. L., de Oliveira Neto, J., Badue, C., Oliveira-Santos, T., & de Souza, A. F. (2023). YOLO-Papaya: A papaya fruit disease detector and classifier using CNNs and convolutional block attention modules. *Electronics, 12*(10), 2202. https://doi.org/10.3390/electronics12102202
8. Nagaraj, P., Muneeswaran, V., & Anurag, T. (2022). A novel scheme for papaya fruit disease prediction and classification using convolutional neural networks. In *Proceedings of ICDSAAI*.
9. Kaur, H., Prashar, D., & Kumar, V. (2022). Disease identification in papaya plant and their dataset. In *Proceedings of IC3I*. https://doi.org/10.1109/IC3I56241.2022.10072683
10. Hridoy, R. H., & Tuli, M. R. A. (2021). A deep ensemble approach for recognition of papaya diseases using EfficientNet models. In *Proceedings of ICEEICT*. https://doi.org/10.1109/ICEEICT53905.2021.9667804
11. Thokala, B., & Doraikannan, S. (2023). Detection and classification of plant stress using hybrid deep convolution neural networks: A multi-scale Vision Transformer approach. *IET Image Processing, 17*(6), 2635–2647. https://doi.org/10.1049/ipr2.12796
12. Li, G., & Wang, Y. (2023). PMVT: A lightweight Vision Transformer for plant disease identification on mobile devices. *Plant Science, 14*. https://doi.org/10.3389/fpls.2023.1184321
13. Son, C.-H. (2021). Leaf spot attention networks based on spot feature encoding for leaf disease identification and detection. *Electronics, 11*(1), 89. https://doi.org/10.3390/electronics11010089
14. Kumar, R., Jain, A. K., Sharma, V., & Sahni, P. (2024). Enhancing precision in papaya crop health management: A CNN-based approach for early disease detection and classification. In *Proceedings of ICITEICS* (pp. 1–6). https://doi.org/10.1109/ICITEICS62404.2024.10523145
15. Ali, M., & Ullah, I. (2026). Hybrid CNN–LSTM Intrusion Detection Framework for Industrial IoT Security. *Majestic International Journal of AI Innovations,* Volume 1, 2026
16. Bacus, J. A., & Linsangan, N. B. (2022). Detection and identification with analysis of Carica papaya leaf using Android. *Journal of Artificial Intelligence and Technology, 13*(2), 55–63.
17. Hossen, M. S., & Islam, M. S. (2020). Deep learning based classification of papaya disease recognition. In *Proceedings of the International Conference on Intelligent Sustainable Systems* (pp. 945–951). IEEE. https://doi.org/10.1109/ICISS49785.2020.9316106
18. Kumar, K., & Yadav, R. K. (2024). Intelligent monitoring of papaya trees using IoT-based sensors. *International Journal of Research and Global Trends, 12*(6), 75–83.
19. Ratha, A. K. (2024). Optimizing precision agriculture: Bayesian-enhanced papaya fruit disease classification via cubic SVM and ResNet-101 deep features. *Journal of Intelligent & Fuzzy Systems*. https://doi.org/10.3233/JIFS-236421

20. Sethy, S., & Paribara, P. (2024). Papaya diseases dataset (Version V1) [Dataset]. *Mendeley Data*. https://doi.org/10.17632/yvcwypp8rz.1
21. Badgujar, C. M., Poulose, A., & Gan, H. (2024). Agricultural object detection with You Only Look Once (YOLO) algorithm: A bibliometric and systematic literature review. *arXiv*. https://doi.org/10.48550/arXiv.2401.10379
22. Sapkota, R., et al. (2024). Comprehensive performance evaluation of YOLOv11, YOLOv10, YOLOv9 and YOLOv8 on detecting and counting fruitlet in complex orchard environments. *arXiv*. https://doi.org/10.48550/arXiv.2407.12040
23. Yoon, S., Cho, Y., & Ahn, T. I. (2024). Melon fruit detection and quality assessment using generative AI-based image data augmentation. *arXiv*. https://doi.org/10.48550/arXiv.2407.10413
24. Nagpal, R., et al. (2023). An application of deep learning for sweet cherry phenotyping using YOLO object detection. *arXiv*. https://doi.org/10.48550/arXiv.2302.06698
25. Pacal, I., et al. (2024). A systematic review of deep learning techniques for plant diseases. *Artificial Intelligence Review, 57*. https://doi.org/10.1007/s10462-024-10789-2
26. Harakannanavar, S. S., et al. (2022). Plant leaf disease detection using computer vision and machine learning algorithms. *Global Transitions Proceedings, 3*(1), 305–310. https://doi.org/10.1016/j.gltp.2022.03.036
27. Amin, G. and Simran (2025) 'Genetic and Biotechnological Approaches for Mitigating Climate Change Impacts on Agriculture', *Majestic Greengenix: A Journal of Modern Plant Science,* Volume 1, 2026, pp. 1-16