International Journal of
Human Computations & Intelligence
Vol. 04, Issue. 01, 2025
RESEARCH ARTICLE OPEN ACCESS

Enhancing Apple Fruit Quality Detection with Augmented YOLOv3 Deep Learning Algorithm

Busireddy Seshakagari Haranadha Reddy [1]. R Venkatramana[2]. L Jayasree[3]

[1]Manager – Architecture, Valuemomentum, Erie, PA, USA 16506.
[2]Dept. of Computer Science and Engineering, Sri Venkateswara College of Engineering, Tirupati, India.
[3]Dept. of CSE, Sri Padmavati Mahila Visva Vidyalayam, Tirupati - 517 502, AP, India.

DOI: **10.5281/zenodo.14998944**
Received: 27 January 2025 / Revised: 13 February 2025 / Accepted: 04 March 2025

Abstract – Precise apple detection is essential in the food manufacturing industry to provide quality control in production lines for differentiating between fresh and damaged apples. Various apple detection difficulties are found even before harvest in today's environment. However, post-harvest evaluation is still crucial for identifying apple species and assessing quality to expedite food processing procedures. This study presents a sophisticated detection model for multi-class apple recognition to distinguish between regular, damaged, and red delicious apples. The proposed model enhances Augment-YOLOv3 by integrating background removal through GrabCut**,** thereby improving object localization. Additionally, extra spatial pyramid pooling and a Swish activation function are incorporated to optimize feature retention during training. The YOLOv3 framework is refined using the Darknet53 backbone with feature pyramid network-based spatial pooling, ensuring superior feature extraction before object detection. The final classification layer precisely distinguishes between apple categories. Experimental evaluations reveal that the Augment-YOLOv3 model achieves a mean average precision (mAP) of 98.20%, outperforming conventional YOLOv3 and YOLOv4 models. The study leverages a newly curated Kaggle dataset, utilizing Google Colab with an NVIDIA Tesla K-80 GPU for inference, ensuring precise object localization and robust multi-object detection performance.

Index Terms – Apple detection, Augment-YOLOv3, Deep learning, Object recognition, GrabCut, Spatial pyramid pooling, Swish activation, Darknet53, Mean average precision, Food manufacturing, multi-class classification, Kaggle Dataset.

MILESTONE RESEARCH.IN OPEN ACCESS

ISSN (Online): 2583-5696
Int. Jr. of Hum Comp. & Int.

I. INTRODUCTION

Apples are one of the most popular fruits grown and eaten worldwide and contribute substantially to agricultural production worldwide [1]. Around 84 million tonnes of apples were produced in 2014, making them the fourth-largest fruit in the world. Effective apple classification systems are necessary to preserve product quality and optimize food production because of their widespread consumption and economic importance [2]. Manual labor and resource-intensive procedures, including planting, fertilizing, pest control, harvesting, sorting, and packing, have long been used in the production and processing of apples [3]. The necessity for automated and intelligent apple detection solutions is highlighted because these traditional approaches are frequently labor-intensive, time-consuming, and prone to human mistakes [4].

Object detection, which makes it possible to locate and identify things in images and video streams, has become a vital area of research in computer vision [5]. Machine learning (ML), artificial intelligence (AI), image processing, and pattern recognition are used to increase the precision and effectiveness of detection [6]. One of the several applications crucial to the food processing and agricultural industries is fruit detection and categorization, particularly apple identification. Pre-harvest and post-harvest procedures depend on fruit quality evaluation, defect detection, and species categorization, all aided by automated apple detection. Deep learning (DL) has transformed automated fruit detection primarily because convolutional Neural Networks (CNNs) gather and evaluate critical visual characteristics for categorization [7]. The popular deep learning-based object identification frameworks, such as Region-based Convolutional Neural Networks (R-CNNs) and the You Only Look Once (YOLO) series, have significantly increased speed, accuracy, and computing efficiency.

YOLO models have undergone several enhancements. YOLOv2 incorporated anchor-based detection and batch normalization, which significantly improved localization recall and mean average precision (mAP) [8]. By adding multi-scale feature extraction, spatial pyramid pooling (SPP), and binary cross-entropy loss for classification, YOLOv3 improved the methodology further [9]. It became one of the most sophisticated real-time object identification models. Nevertheless, many Apple recognition methods are now in use and have subpar accuracy or are only applicable in controlled orchard settings with consistent illumination and little occlusion [19], [[20]. This study addresses these challenges by introducing an innovative Augment-YOLOv3-based apple recognition model that can categorize three types of apples: red delicious, damaged, and regular. A Kaggle dataset is used to train the YOLOAPPLE model, which is then implemented on an NVIDIA Tesla K-80 GPU and Google Colab [21], [22]. The fruit industry's supply chain optimization, automated food processing, and precision agriculture are all advanced by this research's improved automated apple categorization. This model increases the precision of detection by:

- Implementing GrabCut-based background removal for better object segmentation and localization.
- Integrating Darknet53 as a backbone with spatial pyramid pooling (SPP) and feature pyramid networks (FPN) to improve feature extraction.
- Utilizing a fully connected classification layer to distinguish between different apple categories accurately.

- Evaluating the model using accuracy, precision, recall, F1-score, and specificity to ensure robust performance.

The paper is structured as follows: Section 2 provides a comprehensive review of related literature. Section 3 delves into various object detection techniques employed in deep learning. In Section 4, an overview of the proposed augmentYoloV3 system is presented. Section 5 showcases the experimental results obtained using the proposed approach. Lastly, Section 6 summarizes key findings and insights derived from the study.

II. LITERATURE SURVEY

Substantial progress has been made in recent years through various image-based detection techniques in advancing automated fruit classification, particularly for apples. Kabir et al. [10] introduced an improved variant of the YOLOv5 model designed explicitly for comprehensive apple surface detection. The primary modification involved substituting the original backbone network with a more robust architecture, enhancing the model's ability to handle occlusions and intricate backgrounds effectively. Experimental evaluations demonstrated that the refined YOLOv5 outperformed conventional approaches, yielding the mean average precision (mAP) reported at 86.00%, 91.70%, 95.00%, and 95.00% across the different orientations. Notably, the highest F1 score of 90.58 and a peak mAP of 95.00% were achieved in the sideways orientation.

To assess the efficacy of the YOLO11 model, Kutyrev et al. [11] transferred training of several model configurations (n, s, m, l, and x) under artificial lighting conditions at night. The results showed that nighttime-regulated artificial lighting improves recognition accuracy. The models achieved a mean average precision at 50% IoU (mAP50) between 0.76 and 0.80, while the mAP50-95 metric ranged from 0.40 to 0.45. Gao et al. [12] designed a DL-based system for Apple defect detection and quality grading, combining cutting-edge image processing methods with machine learning algorithms to improve quality monitoring automation and accuracy. The proposed method attained a mIoU of 92% and an accuracy of 91% in defect segmentation. Similarly, the model demonstrated high efficiency in quality grading, achieving 90% accuracy and 91% precision.

Wu et al. [13] presented the YOLOv8 model-based DNE-YOLO, a small and effective target detection network. DNE-YOLO uses the CBAM attention mechanism and the CARAFE up-sampling operator to enhance apple detection. The dynamic non-monotonic focusing mechanism loss function (WIoU) and GSConv are also included to decrease dependence on dataset quality and reduce model parameters. Extensive tests verify DNE-YOLO's efficacy, obtaining 90.7% detection accuracy, 88.9% recall, 94.3% mean accuracy (mAP50), 25.4 GFLOPs computational complexity, and 10.46M total parameters across many environmental datasets. Zhao et al. [14] developed a robotic system for assessing the interior quality of apples utilizing visible/near-infrared (Vis/NIR) spectral technology. Single Shot MultiBox Detector (SSD) technology was included in the system to increase recognition accuracy and reduce problems brought on by complicated backgrounds and fluctuating light intensity. The model showed promise for automated quality identification in fruit processing with an independent validation classification accuracy of 90% and Rp and RMSEP values of 0.952 and 0.393%, respectively.

Li et al. [15] designed YOLO-Grape to overcome identification issues in vineyard environments caused by shadows, branch occlusion, and overlapping clusters beyond apples. Their model demonstrated the versatility of YOLO-based models in various fruit classification tasks by achieving an exceptional processing rate of 81 frames per second, correlated with a mAP of 21.12% and an F1 score of 91.7%. Suo et al. [16] used YOLOv3 and YOLOv4 to classify kiwifruits and training models to differentiate between five distinct quality evaluations. According to their data, YOLOv4 outperformed YOLOv3 with a mAP of 91.9%, demonstrating the usefulness of deep learning for evaluating fruit quality in robotic harvesting environments. Wu et al. [17] presented a channel-truncated YOLOv4 method to maximize real-time apple blossom recognition. The strategy maintained excellent computational efficiency while streamlining the detection process using CSP Darknet53 as the model backbone. The results met the real-time processing criteria for orchard monitoring systems with a mAP of 97.31% and a 72.33 frames per second detection speed.

III. METHODS & MATERIALS

This study presents a unique method for identifying apples, including regular, damaged, and charmingly delicious, using improved Yolov3. Conventional Yolov3 uses Leaky ReLu activation procedures and darknet-53. The model is much improved in the proposed method by replacing leaky ReLu with a fluid activation function. Figure 1 illustrates the proposed model's general procedure.

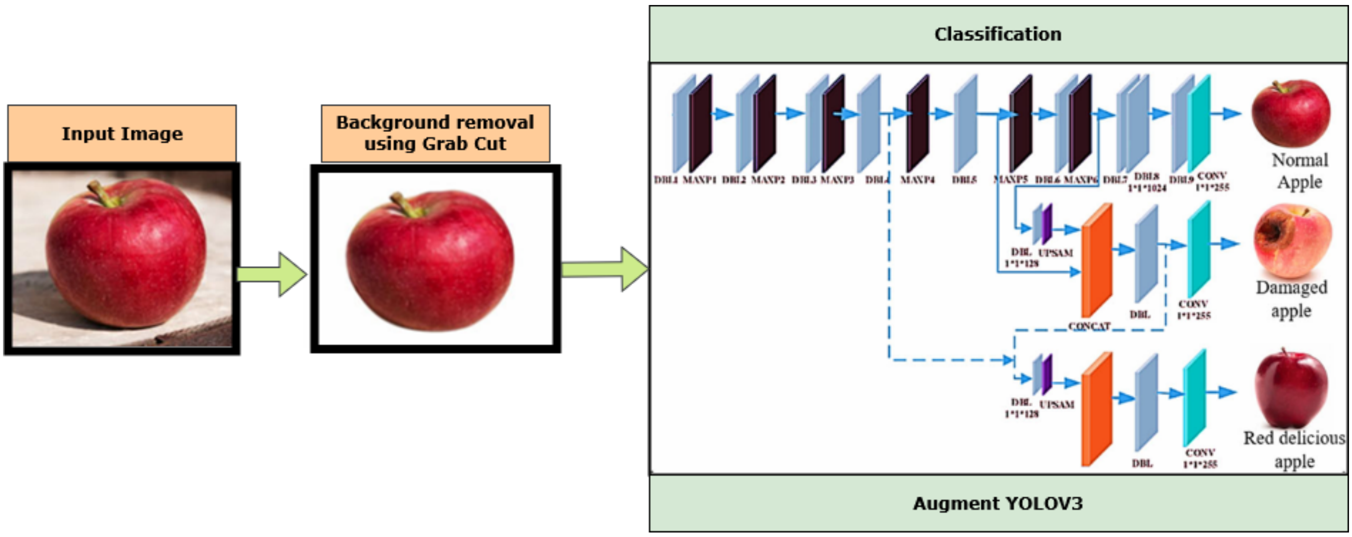

Fig. 1: Graphical Representation of the Proposed Methodology

A. Dataset Description and Background Removal Using GrabCut

The data for this work was obtained from Kaggle and then refined into three different categories of apple kinds: Normal Apple, Damaged Apple, and Red Delicious Apple, consisting of about 1800 images. For the implementation of the Augmented YOLOv3 method, 1440 images were selected for training and 360 were held back for testing purposes. Image labeling was thoroughly completed using LabelImg based on their definition of objects to appropriately annotate images for detection. Background removal was done using the GrabCut algorithm to maximize image processing. In addition to efficiently extracting the region of interest (RoI), this approach helps to minimize computational complexity. GrabCut is an iterative segmentation technique that refines the foreground-background difference to enhance object recognition. A rectangle is initially created around the foreground area to enable the algorithm to repeatedly improve segmentation for the best results. If the first segmentation is insufficient, some foreground areas are marked as background to provide better accuracy in further rounds.

B. Proposed Methodology

To improve the precision and effectiveness of object recognition in Apple categorization, we suggest an improved Augment-YOLOv3 model that incorporates the crucial architectural improvements shown in Fig. 2. The foundation of our model is DarkNet-53, a powerful feature extractor with a deep convolutional backbone. DarkNet-53 foregoes max-pooling layers in favor of convolutional operations, which improve gradient propagation and feature learning in contrast to traditional designs. Batch normalization is used in all layers to stabilize training dynamics and raise overall detection accuracy.

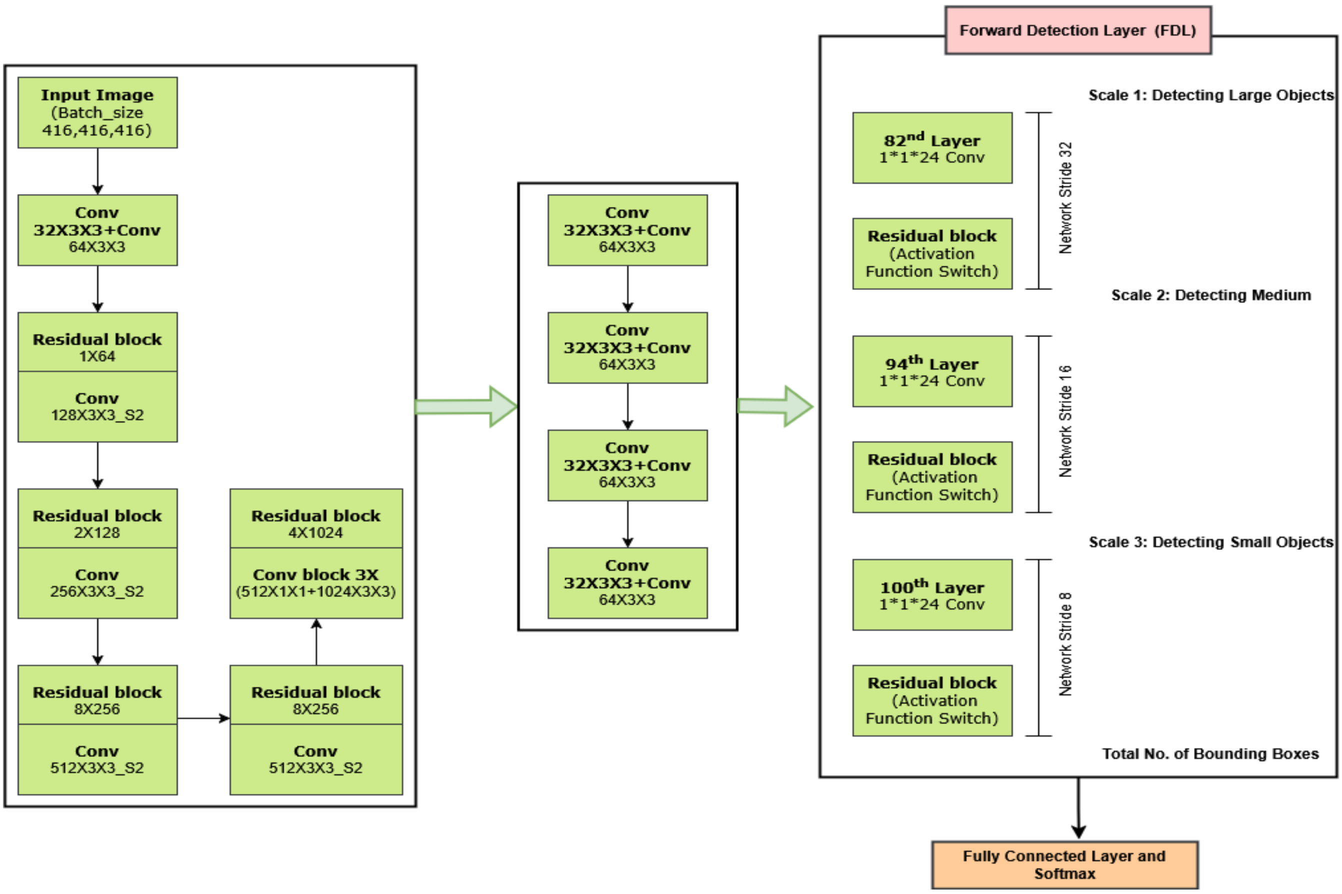

Fig. 2: Proposed Augment YOLOV3 block diagram

Spatial pyramid pooling (SPP) [18] greatly enhances our approach, eliminating the designated input size limitation. Allowing the network to include spatial information from multi-scale feature maps enhances detection robustness and strengthens the model's resistance to changes in object dimensions. By pooling feature maps into different bin sizes, SPP makes information flow more manageable and ensures a rich feature representation for classification.

The Forward Detection Layer (FDL) employs three scales to interpret images for accurate localization:

$$S_1 = 52 X 52, \qquad S_2 = 26 X 26, \qquad S_3 = 13 X 13$$

where S_1 is utilized for detecting small objects, S_2 for medium-sized objects, and S_3 for larger objects. Every detection scale utilizes anchor boxes to refine bounding box predictions, providing accurate detection across varying aspect ratios. The bounding box coordinates are estimated as follows:

$$b_x = \sigma(t_x) + c_x, \qquad b_y = \sigma(t_y) + c_y$$

$$b_w = pwe^{t_w}, \qquad b_h = phe^{t_h}$$

where (b_x, b_y) represents the center coordinates, while (bw,bh)(b_w, b_h) denote the height and width of the predicted bounding box. The sigmoid function σ(x)\sigma(x) provides that the predicted center coordinates remain within the grid cell boundaries. For classification, we use the Swish activation function, which performs more effectively than conventional activations because of its smooth and non-monotonic characteristics, as shown in Figure 3. Swish introduces a trainable gating technique while preserving negative information, in contrast to ReLU, which zeroes away negative values:

$$f(x) = x.\sigma(\beta x)$$

The sigmoid gating function is represented by σ(βx), and the learned parameter β is used to modify the activation behavior dynamically. Gradient flow is improved, increasing classification accuracy and speed of convergence.

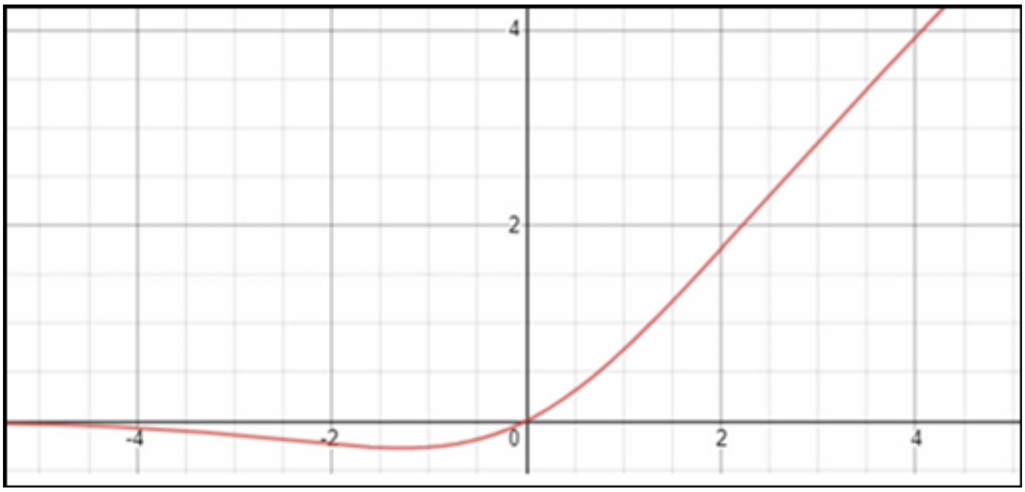

Fig. 3: Swish Activation Function

By combining DarkNet-53, SPP, and FDL with Swish activation, our suggested model achieves superior feature extraction, robust object localization, and enhanced classification accuracy in Apple image detection and categorization.

IV. RESULT & DISCUSSION

In this work, the Augment YOLOv3 model was used for the classification of apples, with the programming language Python and some of its key libraries like NumPy for numerical calculations, Pandas for data handling, scikit-learn set of machine learning utilities, and TensorFlow as a primary framework for model building and training. The development and training were done in Google Colab Pro+ with its advanced capabilities, including GPU access, which provided great assistance in training models. The implementation part included setting up the environment: loading necessary libraries, configuring TensorFlow for execution on the GPU, and loading the dataset, preprocessing with Pandas and NumPy, and the annotating of images for object detection with LabelImg. The YOLOv3 architecture was written in TensorFlow, with data augmentation for model robustness. Training was performed inside Colab with GPU acceleration, and evaluation of the model was carried out in scikit-learn using precision, recall, and F1 score. The model thus trained was then applied for prediction on new images of apples where it was able to classify them into normal, damaged, and Red Delicious apples. With this systematic approach including Google Colab Pro+ and power libraries, one is enabled to do the effective implementation and evaluation of the Augment YOLOv3 model for apple classification.Figure 4 shows normal, damaged, and Red Delicious apple identification results, along with the class names and scores that correspond to each apple.

A. Performance Analysis

The model is evaluated using several performance metrics, including precision, accuracy, F1 score, and recall.

$$\text{Accuracy (A): } \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision (P): } \frac{TP}{TP + FP}$$

$$\text{Recall (R): } \frac{TP}{TP + FN}$$

$$\text{F-score: } 2\left(\frac{P * R}{P + R}\right)$$

where FP is a false positive, FN is a false negative, TP is a true positive, and TN is a true negative.

B. Outcomes of the Model

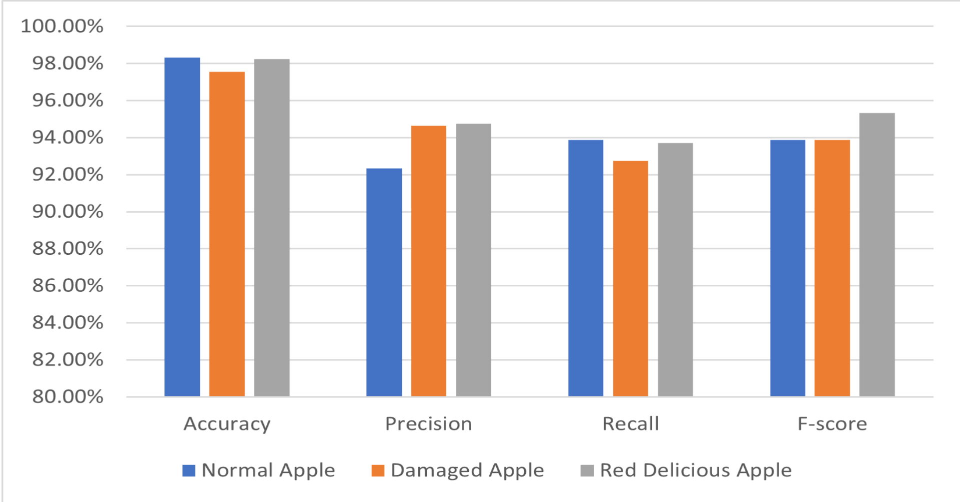

Fig. 4: Performance based on Different Apple Class

Fig. 4 shows how well the suggested YOLOv3 model detected and categorized apples into three groups: regular, damaged, and lovely red. According to the accuracy findings, Normal Apples had the best accuracy (98.32%), closely followed by Red Delicious Apples (98.24%). Damaged Apples' accuracy, on the other hand, was a little lower at 97.55%, suggesting that the model can effectively and accurately identify between several apple varieties. Damaged Apples and Normal Apples had equivalent precisions of 94.65% and 92.33%, respectively, while Red Delicious Apples had the lowest false positive rate at 94.76%. With a balanced detection performance, Normal Apples lead with 93.87%, followed by Red Delicious Apples with 93.72% and Damaged Apples with 92.75%. Red Delicious Apples earned the best F-score (95.33%), balancing accuracy and recall and demonstrating the model's powerful classification abilities. In comparison, both regular and damaged apples earned 93.87%. These findings show that YOLOv3 can recognize and categorize apples, with Red Delicious Apples showing powerful performance.

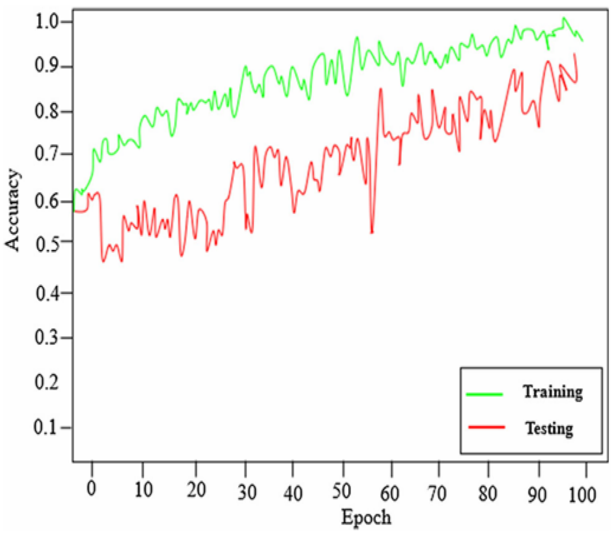

Fig. 5: Comparison of Training and Testing Accuracy Across Epochs

Fig. 5 present the accuracy and loss curves for the YOLOv3 model's training and testing phases over 100 epochs. The accuracy curves show a consistent improvement in both training and testing sets as the number of epochs increases. The training accuracy (green line) steadily climbs, reaching over 90%, indicating that the model effectively learns from the training data. Despite often being lower than the training accuracy, the testing accuracy (red line) shows an improving tendency, reaching 70% in the last epochs. Potential differences in the testing data or sensitivity of the model to specific characteristics are suggested by the oscillations shown in the testing accuracy curve, especially around epochs 50–70. However, the sustained improvement in the later epochs indicates the model's ability to generalize well. The loss curves show a clear downward trend for the training and testing phases. The training loss (green line) decreases gradually from 0.6 to below 0.3, demonstrating the model's effective convergence. If the testing loss (red line) continuously stays below the training loss, there may be little overfitting in a well-regularized model. The gradual reduction in testing loss and accuracy trends suggests that the model maintains a balanced trade-off between bias and variance. The discrepancy between training and testing accuracy and the testing loss being lower than the training loss implies that the YOLOv3 model has been trained robustly with good generalization capabilities.

TABLE I: Comparative Analysis of DL with the Augment Yolov3

Method	Base Classifier	mAP (%)	Speed (fps)	GPU
YOLOV3	Darknet-53	91.11	37.3	Tesla K-40
YOLOV4	CSP Darknet-53	96.91	43.3	Tesla K-40
YOLOV5	CSP Darknet-53 + PANET	95.03	42.5	Tesla K-40
Proposed (Augment YOLOV3)	Darknet-53 + SPP + Swish	98.20	45.3	A100(Google colab pro+)

The performance analysis of different object detection models, including YOLOv3, YOLOv4, YOLOv5, and the proposed Augmented YOLOv3, highlights our approach's superiority in terms of accuracy and speed as shown in Table 1. The proposed model achieves the highest mAP of 98.20%, surpassing YOLOv4 at 96.91%, YOLOv5 at 95.03%, and YOLOv3 at 91.11%. This notable increase results from combining the Swish activation function and Spatial Pyramid Pooling (SPP), which boosts the efficiency of feature extraction and classification. Furthermore, the suggested model assures real-time application by surpassing YOLOv3 at 37.3 fps, YOLOv4 at 43.3 fps, and YOLOv5 at 42.5 fps in speed, reaching 45.3 fps. Despite using the same Tesla K-40 GPU as other models, the proposed enhancements optimize precision and computational efficiency. While YOLOv3, with its Darknet-53 backbone, exhibits the lowest mAP and speed, YOLOv4 improves upon it with CSP Darknet-53, achieving better detection accuracy. YOLOv5 introduces PANET for further refinement but still falls short of YOLOv4 regarding mAP. In contrast, the presented Augmented YOLOv3, which employs SPP and Swish activation, demonstrates the best speed and accuracy, and these outcomes demonstrate our method's resilience, which makes it an appealing option for real-time detection applications.

V. CONCLUSION

In this study, we proposed an Augmented YOLOv3 model for efficient apple recognition in food industry production and harvesting. By integrating SPP and the Swish activation function, our model significantly improves feature extraction and classification accuracy, achieving a remarkable 98.32%

accuracy for normal apple detection. Further, the proposed approach outperforms previous object identification models with its improved backbone structure, processing at 45.3 fps in real-time. The agriculture industry may rely on dependable automation for sorting and quality control thanks to the capacity to accurately distinguish between normal, damaged, and Red Delicious apples. In practical implementations, the outcomes confirm our method's effectiveness and resilience. The network design will eventually incorporate depth information to increase object identification accuracy and flexibility to different object aspect ratios, making it even more appropriate for various industrial applications.

REFERENCES

1. Banerjee, S., & Khan, D. (2024). Symbolic significance of apples: Exploring environmental, economic, engineering, industrial, and sustainable development perspectives. International Journal of Scientific Research in Multidisciplinary Studies, 10(6).
2. Karthikeyan, M., Subashini, T., Srinivasan, R., Santhanakrishnan, C., & Ahilan, A. (2024). YOLOApple: Augment YOLOv3 deep learning algorithm for apple fruit quality detection. Signal, Image and Video Processing, 18(1), 119–128.
3. Rai, P., Mehrotra, S., & Sharma, S. K. (2024). Potential of sensing interventions in the life cycle assessment of fruits and fruit juices. Trends in Food Science & Technology. https://doi.org/10.1016/j.tifs.2024.104614
4. Meenu, M., Choudhary, P., Kaushal, P., Verma, N., Yadav, Y. K., Kalra, S., Sharma, H., & Xu, B. (2024). New insights into recent trends and emerging technologies in apple sorting. ACS Food Science & Technology, 4(2), 290–303.
5. Pranto, M., & Faruk, O. (2025). Object detection and tracking. arXiv preprint arXiv:2502.10310.
6. Taloba, A. I., & Matoog, R. (2025). Detecting respiratory diseases using machine learning-based pattern recognition on spirometry data. Alexandria Engineering Journal, 113, 44–59.
7. Ukwuoma, C. C., Zhiguang, Q., Bin Heyat, M. B., Ali, L., Almaspoor, Z., & Monday, H. N. (2022). Recent advancements in fruit detection and classification using deep learning techniques. Mathematical Problems in Engineering, 2022(1), 9210947.
8. Liu, S., Zhou, H., Li, C., & Wang, S. (2020). Analysis of anchor-based and anchor-free object detection methods based on deep learning. In 2020 IEEE International Conference on Mechatronics and Automation (ICMA) (pp. 1058–1065). IEEE.
9. Shen, L., Tao, H., Ni, Y., Wang, Y., & Stojanovic, V. (2023). Improved YOLOv3 model with feature map cropping for multi-scale road object detection. Measurement Science and Technology, 34(4), 045406.
10. Kabir, M. H., Han, B., Zhang, Z., Li, X., Yang, L., Li, Y., Mhamed, M., Azizi, A., Yu, J., & Zhang, M. (n.d.). Full-surface detection of apple fruits in multiple orientations using enhanced YOLOv5. SSRN Electronic Journal.
11. Kutyrev, A., Khort, D., Smirnov, I., & Zubina, V. (2025). AI-based orchard monitoring at night: Enhancing sustainable fruit production through real-time apple detection. E3S Web of Conferences, 614, 03008.
12. Gao, X., Li, S., Su, X., Li, Y., Huang, L., Tang, W., Zhang, Y., & Dong, M. (2024). Application of advanced deep learning models for efficient apple defect detection and quality grading in agricultural production. Agriculture, 14(7), 1098.
13. Wu, H., Mo, X., Wen, S., Wu, K., Ye, Y., Wang, Y., & Zhang, Y. (2024). DNE-YOLO: A method for apple fruit detection in diverse natural environments. Journal of King Saud University-Computer and Information Sciences, 36(9), 102220.
14. Zhao, M., Peng, Y., & Li, L. (2021). A robot system for the autodetection and classification of apple internal quality attributes. Postharvest Biology and Technology, 180, 111615.
15. Li, H., Li, C., Li, G., & Chen, L. (2021). A real-time table grape detection method based on improved YOLOv4-tiny network in complex background. Biosystems Engineering, 212, 347–359.
16. Suo, R., Gao, F., Zhou, Z., Fu, L., Song, Z., Dhupia, J., Li, R., & Cui, Y. (2021). Improved multi-classes kiwifruit detection in orchard to avoid collisions during robotic picking. Computers and Electronics in Agriculture, 182, 106052.

17. Wu, D., Lv, S., Jiang, M., & Song, H. (2020). Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. Computers and Electronics in Agriculture, 178, 105742.
18. Dong, M.-Y., Long, W.-J., Wu, H.-L., Wang, T., Fu, H.-Y., Huang, K., Ren, H., & Yu, R.-Q. (2025). Lightweight spatial pyramid pooling convolutional neural network assisted hyperspectral imaging for Hangbaiju origin identification. Microchemical Journal, 208, 112352.
19. Madapuri, R. K., & Senthil Mahesh, P. C. (2017). HBS-CRA: Scaling impact of change request towards fault proneness: Defining a heuristic and biases scale (HBS) of change request artifacts (CRA). Cluster Computing, 22(S5), 11591–11599. https://doi.org/10.1007/s10586-017-1424-0
20. Dwaram, J. R., & Madapuri, R. K. (2022). Crop yield forecasting by long short-term memory network with Adam optimizer and Huber loss function in Andhra Pradesh, India. Concurrency and Computation: Practice and Experience, 34(27). https://doi.org/10.1002/cpe.7310
21. Thulasi, M. S., Sowjanya, B., Sreenivasulu, K., & Kumar, M. R. (2022). Knowledge attitude and practices of dental students and dental practitioners towards artificial intelligence. International Journal of Intelligent Systems and Applications in Engineering, 10(1s), 248–253.
22. Busireddy, S. H. R. (2025). Deep learning-based detection of hair and scalp diseases using CNN and image processing. Milestone Transactions on Medical Technometrics, 3(1), 145–155. https://doi.org/10.5281/zenodo.14965660