Fruit Recognition Using Deep Learning: A Comparative Study Between Custom CNN and Transfer Learning Approaches

Aashna Kushwaha
Department of Computer Engineering and Application
GLA University
Mathura, India
aashnakushwaha0709@gmail.com

Rajesh Kumar Tripathi
Department of Computer Engineering and Application
GLA Univeristy
Mathura, India
rajesh.tripathi@gla.ac.in

***Abstract*—The automated identification of fruits is of significant importance in the domains of agriculture, retail, and the food industry. This paper explores two deep learning-based methods for fruit classification: a custom CNN method and a transfer learning method using MobileNetV2. Both these models were trained and evaluated on a comprehensive fruit dataset containing 36 classes. The first approach utilized a simple CNN architecture with moderate accuracy, while the second employed a pre-trained MobileNetV2 model with advanced data augmentation techniques. Experimental results show that the MobileNetV2 model significantly outperforms the custom CNN, achieving a test accuracy of 94.71% compared to 89% accuracy from the custom CNN. Thus, Transfer Learning proves to be a more effective method for fruit recognition tasks.**

***Index Terms*—Deep Learning, MobileNetV2, Transfer Learning, Data Augmentation**

I. INTRODUCTION

The precise classification of fruits has emerged as a critical endeavor within the fields of computer vision and artificial intelligence, encompassing diverse applications in agriculture, supply chain management, food retail, and automated inspection systems [1]. Conventional approaches to fruit classification, which predominantly depend on manual inspection, are not only labor-intensive but also prone to human error, inconsistency, and inefficiency when applied to large datasets or industrial processes. As global agricultural production intensifies to meet increasing food demands, there is a pressing need for automated, rapid, and reliable classification systems capable of operating with minimal human involvement. [9]

In recent years, deep learning (DL), particularly Convolutional Neural Networks (CNNs), has transformed the field of image recognition by removing the necessity for manual feature extraction. CNNs have exhibited exceptional performance by autonomously learning hierarchical features directly from raw images, rendering them particularly suitable for tasks that involve intricate patterns and variations, such as differentiating between various types of fruits that frequently possess similar color, shape, and texture characteristics.

Building a Convolutional Neural Network (CNN) model from the ground up requires substantial quantities of labeled data, significant computational resources, and meticulous model tuning to attain satisfactory outcomes without the risk of overfitting. In contrast, Transfer Learning—where a model pre-trained on datasets such as ImageNet is modified for new tasks—presents a viable alternative. This approach enables researchers and practitioners to utilize learned visual features from large datasets, thereby considerably minimizing training time, data demands, and computational expenses, while frequently enhancing the performance of the final model.

However, training a CNN from scratch has some inherent challenges:

- Data and Computational Requirements: Training deep models typically requires large labeled datasets and high computational resources, which may not be readily available in some practical applications.
- Overfitting Risk: Custom CNNs built from scratch may have a higher risk of overfitting, especially when the dataset is relatively small or imbalanced.
- Training Time: Training deep models from scratch can be time-consuming, requiring days or even weeks of computation for optimal performance.

A. Problem Statement

In the domain of fruit classification, the aforementioned challenges impede the advancement of efficient and accurate automated systems. The primary issues encountered include elevated computational costs, limited data availability, and the potential for overfitting when training Convolutional Neural Network (CNN) models from the ground up. Additionally, the manual labeling of extensive datasets consisting of fruit images for model training is both financially burdensome and time-consuming. Therefore, there exists an urgent necessity for a solution that is more efficient, scalable, and accurate in the domain of fruit recognition.

B. Our Contribution

To address these challenges, we propose a dual-model approach that evaluates two distinct strategies:

- Custom CNN Model: We design a custom CNN architecture to serve as a baseline model. This model is trained from scratch and optimized for fruit classification, with a relatively shallow architecture, suitable for applications requiring moderate computational power.
- MobileNetV2 Transfer Learning Model: We leverage Transfer Learning using the pre-trained MobileNetV2 model. This approach allows us to fine-tune a model that has already learned general image features from large datasets like ImageNet, significantly reducing the training time and computational requirements. Data augmentation techniques are also employed to further enhance the model's generalization capabilities.

The objective of this study is to evaluate the efficacy of two methodologies based on the following criteria:

- Classification accuracy
- Training efficiency
- Generalization performance
- Practical applicability in real-world systems

Through the implementation of a transfer learning approach utilizing MobileNetV2, this research seeks to address the limitations associated with conventional convolutional neural networks (CNNs) while preserving high classification accuracy and robustness, even when applied to a comparatively smaller fruit dataset.

II. Literature Review

In recent years, deep learning techniques have transformed the domain of image classification, markedly surpassing traditional machine learning methods that were predominantly dependent on manually crafted features. Convolutional Neural Networks (CNNs) have emerged as the foundational element of contemporary computer vision, demonstrating the ability to learn hierarchical feature representations directly from pixel data without the necessity for manual intervention.

A seminal contribution by Krizhevsky et al. (2012) introduced AlexNet [11], showcasing the potential of deep neural networks to achieve exceptional performance on extensive datasets such as ImageNet. This advancement catalyzed the creation of deeper and more intricate architectures, including VGGNet (Simonyan and Zisserman, 2014), ResNet (He et al., 2015), and MobileNet (Howard et al., 2017), each designed to improve accuracy and efficiency.

Transfer Learning has emerged as an effective strategy to tackle the challenges associated with training deep networks from the ground up, particularly in situations where there is a scarcity of labeled data. Models that have been pre-trained on extensive datasets such as ImageNet can be adapted for domain-specific tasks, leading to a substantial reduction in both computational expenses and data requirements, while frequently enhancing overall performance.

A. *Fruit Recognition Specific Studies*

In the field of fruit recognition, traditional Convolutional Neural Networks (CNNs) and transfer learning methodologies have been extensively investigated:

- Zhang et al. (2017) established a fruit grading system utilizing deep CNNs, which attained significant accuracy while highlighting the necessity for extensive labeled datasets.
  M
- ujtaba et al. (2021) illustrated the effectiveness of MobileNetV2 in conjunction with data augmentation techniques applied to agricultural datasets, leading to enhanced accuracy even in scenarios with limited data availability.
- Rahman et al. (2020) underscored the importance of integrating data augmentation with deeper CNN architectures to enhance the robustness of fruit classification systems.
- Pereira et al. (2022) proposed a lightweight model based on EfficientNetB0 for fruit classification, achieving high accuracy with minimal computational requirements. This research underscores the trend toward the utilization of more efficient models suitable for deployment on mobile and Internet of Things (IoT) devices.
- Chowdhury et al. (2023) introduced a hybrid model that integrates Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) for the classification of fruits and vegetables, achieving state-of-the-art performance across multiple datasets. Their findings indicate that attention mechanisms enhance the model's focus on subtle features within images.
- Alam et al. (2023) investigated the application of self-supervised learning techniques to pre-train models using unlabeled fruit images, thereby significantly reducing reliance on labeled datasets while preserving high classification performance.
- Gupta and Singh (2024) conducted a comprehensive analysis comparing traditional CNNs, various MobileNet variants, and emerging lightweight models such as MobileNetV3 and EfficientNetV2. Their study revealed that MobileNetV3-small exhibited the most favorable balance between speed and accuracy for practical deployment in real-world scenarios.

The current body of literature underscores a notable progression towards the development of lightweight, efficient models and hybrid architectures that enhance classification accuracy while reducing computational expenses. Furthermore, the incorporation of self-supervised learning methodologies effectively mitigates the significant challenge posed by the scarcity of labeled data. These emerging trends reinforce the notion that Transfer Learning, data augmentation, and the utilization of efficient pre-trained models are essential strategies for the establishment of robust and scalable fruit recognition systems.

In consideration of these findings, this research aims to compare a custom-built Convolutional Neural Network (CNN) with an architecture based on MobileNetV2. Gupta and Singh (2024) conducted a comprehensive analysis comparing traditional CNNs, various MobileNet variants, and emerging lightweight models such as MobileNetV3 and EfficientNetV2. Their study revealed that MobileNetV3-small exhibited the most favorable balance between speed and accuracy for practical deployment in real-world scenarios.

III. Methodology

This section presents the models used for fruit classification: a custom Convolutional Neural Network (CNN) designed from scratch as a **Baseline Model**, and a MobileNetV2-based transfer learning model as the **Proposed Method**. Both models were trained, validated, and compared using the same dataset under similar conditions .

1) Baseline Method: Custom Convolutional Neural Network (CNN): The baseline model is a custom-designed Convolutional Neural Network (CNN) developed from the ground up to establish a comparative benchmark. The architecture adheres to a conventional design that incorporates multiple convolutional and pooling layers, succeeded by fully connected layers.

2) Architecture Details:

- **Input**: Images resized to $64 \times 64 \times 3$ (RGB).
- **Convolutional Layers**:
  - Conv2D (64 filters, 3×3, ReLU activation)
  - MaxPooling2D (pool size 2×2)
  - Conv2D (128 filters, 3×3, ReLU activation)
  - MaxPooling2D (pool size 2×2)
  - Conv2D (256 filters, 3×3, ReLU activation)
  - MaxPooling2D (pool size 2×2)
  - Conv2D (512 filters, 3×3, ReLU activation)
  - MaxPooling2D (pool size 2×2)
- **Flatten Layer**: Transforms feature maps into a singular vector.
- **Fully Connected Layers**:
  - Dense (512 units, ReLU activation)
  - Dropout (rate = 0.5) to prevent overfitting
  - Dense (36 units, Softmax activation) for multi-class classification

3) Training Details:

- Optimizer: Adam optimizer
- Loss Function: Categorical Crossentropy
- Metrics: Accuracy
- Batch Size: 32
- Epochs: 30

4) Limitations: Although the CNN achieved a training accuracy of 89% and a validation accuracy of 93%, it showed minor signs of overfitting and slower convergence due to the relatively shallow architecture and limited use of regularization and augmentation.

A. Proposed Method: MobileNetV2-based Transfer Learning Approach

The proposed methodology utilizes MobileNetV2, which is a lightweight and efficient convolutional neural network that has been pre-trained on the ImageNet dataset. The model was modified for the purpose of fruit classification by substituting its upper classification layers and subsequently fine-tuning the network.

1) Motivation: MobileNetV2 employs depthwise separable convolutions and inverted residuals with linear bottlenecks to significantly decrease computational demands while preserving high accuracy, thereby rendering it suitable for real-world deployment scenarios that necessitate speed and efficiency.

2) Preprocessing Steps:

- **Input Resizing**: Images resized to $224 \times 224 \times 3$.
- **Normalization**: Pixel values scaled to $[0, 1]$:

$$x_{\text{norm}} = \frac{x}{255}$$

- **Data Augmentation**:
  - Random horizontal and vertical flips
  - Random rotations ($\pm 20^\circ$)
  - Random zoom ($\pm 20\%$)

3) Preprocessing Steps: **Input Resizing**: Images resized to $224 \times 224 \times 3$. **Normalization**: Pixel values scaled to $[0, 1]$:

$$x_{\text{norm}} = \frac{x}{255}$$

Data Augmentation:

- * Random horizontal and vertical flips
  * Random rotations ($\pm 20^\circ$)
  * Random zoom ($\pm 20\%$)

4) Model Architecture: The final architecture includes:

1) Data Augmentation Layers
2) Rescaling Layer for pixel normalization
3) MobileNetV2 Base:
   - Pre-trained on ImageNet
   - Top layers removed (include_top=False)
   - Initially frozen during early training stages
4) Global Average Pooling 2D:

$$GAP(F) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{i,j}$$

5) Dense Layer with 512 units and ReLU activation:

$$h = \text{ReLU}(W_1 \times GAP(F) + b_1)$$

6) Dropout Layer (rate = 0.5) to prevent overfitting:

$$h' = \text{Dropout}(h)$$

7) Output Layer (Dense, 36 units, Softmax activation):

$$\hat{y} = \text{Softmax}(W_2 \times h' + b_2)$$

5) Training Strategy:

- Optimizer: Adam optimizer
- Loss Function: Categorical Crossentropy:

$$\mathcal{L}_{CCE} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

- Evaluation Metrics: Accuracy
- Batch Size: 32
- Epochs: 30

– Callbacks:
  * EarlyStopping (patience = 5)
  * ModelCheckpoint (saving the best model)

Initially, MobileNetV2 layers were frozen. Top layers can be unfrozen for fine-tuning once the classification head stabilizes, using a smaller learning rate.

6) Advantages over Baseline:

– Faster convergence due to pre-trained weights
– Better generalization from ImageNet features
– Higher validation and test accuracy
– Reduced overfitting compared to the baseline CNN

IV. Experimental Results

A. Dataset Description

The dataset employed for training and validation comprises images of fruits that are categorized into 36 distinct classes. Each image is resized to uniform dimensions ($64 \times 64 \times 3$ for the baseline model and $224 \times 224 \times 3$ for the proposed model). The dataset was categorized into training and validation sets as follows:

– **Training set**: 70% of the data
– **Validation set**: 30% of the data

Images were systematically labeled and categorized into their respective class folders, and TensorFlow's utility was employed to facilitate the efficient loading of datasets.

B. Performance Metrics

The following metrics were used for evaluating the models:

– **Accuracy**: Proportion of correctly classified samples.
– **Loss**: Categorical Crossentropy loss between predictions and ground truth.
– **Confusion Matrix**: Evaluates detailed class-wise performance.

C. Comparison of Models

Table I presents the comparison between the baseline CNN and the MobileNetV2-based model.

TABLE I
Comparison of Baseline CNN and Proposed MobileNetV2 Model

Metrics	Baseline CNN	Proposed MobileNetV2
Training Accuracy	89%	91%
Validation Accuracy	93%	94%
Test Accuracy	88%	**94.71%**

D. Results and discussion

1) Accuracy and Loss Curves: This section provides a comprehensive evaluation of the performance of the proposed MobileNetV2 model, which includes an analysis of the confusion matrix as well as the accuracy and loss curves. These visual representations facilitate a deeper understanding of the model's performance and yield valuable insights into the classification outcomes.

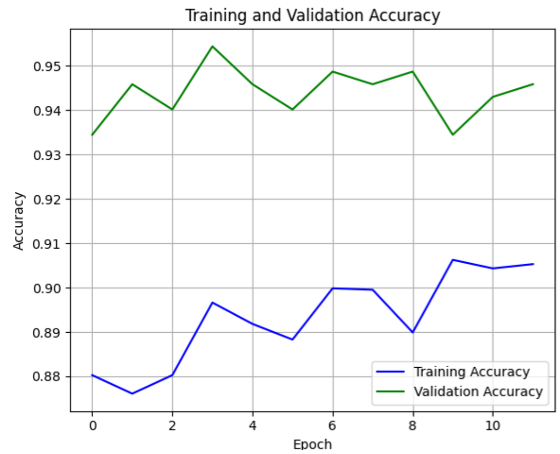

Fig. 1. Training and Validation Accuracy for Proposed MobileNetV2

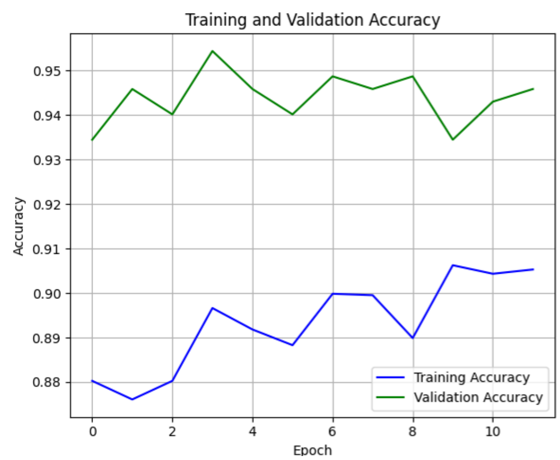

Fig. 2. Loss curves for Proposed MobileNetV2

2) Confusion Matrix: The confusion matrix illustrated in Figure 3 provides a thorough evaluation of the model's performance across all 36 classes. The elements along the diagonal indicate correct classifications, whereas the off-diagonal elements reflect instances of misclassification. This visualization is essential for evaluating the model's efficacy in relation to each individual class. The analysis of the confusion matrix reveals that the model exhibits commendable performance across various fruit classes, as evidenced by a significant number of accurate predictions reflected along the diagonal. Nevertheless, instances of misclassification have been identified, particularly among fruits with similar visual characteristics, highlighting potential areas for enhancement in the model's accuracy. Specifically, certain classes that possess comparable color or texture attributes may pose challenges in differentiation.

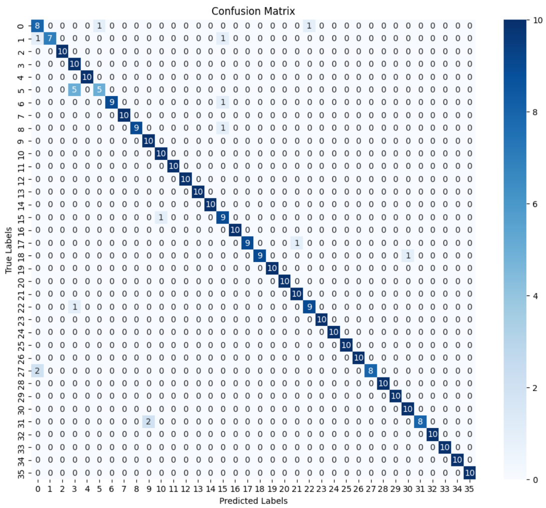

Fig. 3. Confusion Matrix for Proposed MobileNetV2 Model

– The experimental findings underscore the notable advantages of employing Transfer Learning with MobileNetV2 in comparison to the custom Convolutional Neural Network (CNN) model. Although the custom CNN attained a commendable test accuracy of 88%, it necessitated substantial tuning and computational resources to mitigate overfitting, particularly when trained on a relatively limited dataset. Furthermore, the model's test accuracy was constrained, highlighting the inherent challenges associated with training deep CNNs from the ground up, which require a larger dataset and prolonged training duration to achieve optimal performance.
– Conversely, the MobileNetV2 model, which utilizes pre-trained weights from ImageNet, achieved a test accuracy of 94.71%, demonstrating enhanced generalization and reduced training time. The application of transfer learning facilitated the MobileNetV2 model's efficacy with fewer labeled samples, thereby significantly augmenting both efficiency and performance. This approach enables more rapid convergence during training, as the model benefits from the extensive features acquired from millions of images, rendering it more resilient to variations in input data. In summary, the MobileNetV2-based transfer learning methodology surpassed the custom CNN in terms of both accuracy and computational efficiency.

V. Conclusion

This research illustrated the comparative effectiveness of two deep learning methodologies: one employing a custom Convolutional Neural Network (CNN) and the other utilizing Transfer Learning with MobileNetV2 for the purpose of fruit recognition. Through comprehensive experimentation, it was found that although the custom CNN model achieved a validation accuracy of 93% and a test accuracy of 88%, it necessitated substantial effort in terms of architectural design, hyperparameter optimization, and training duration. Furthermore, its propensity to overfit highlighted inherent limitations in scalability and generalization, particularly when applied to relatively small or imbalanced datasets.

In contrast, the MobileNetV2 model, which incorporates data augmentation techniques and dropout regularization, demonstrated a significant improvement over the baseline CNN model. It achieved a test accuracy of 94.71%, reflecting enhanced generalization and robustness to previously unseen data. The utilization of pre-trained weights from ImageNet enabled the model to capitalize on transferable low- and high-level image features, thereby considerably diminishing the reliance on extensive labeled datasets and extended training periods. Furthermore, the lightweight architecture of MobileNetV2 renders it particularly advantageous for deployment in real-time and resource-constrained environments, such as mobile or embedded systems within agricultural applications.

The findings substantiate that Transfer Learning, when meticulously fine-tuned, presents a viable and robust alternative to conventional deep Convolutional Neural Networks (CNNs), particularly for domain-specific applications such as fruit classification. Future research may aim to enhance accuracy by fine-tuning the deeper layers of MobileNetV2, integrating self-supervised learning methodologies, or investigating next-generation efficient architectures such as EfficientNetV2 and MobileNetV3. Furthermore, broadening the dataset to encompass a greater diversity of fruit types and environmental conditions could significantly improve the model's adaptability in practical, real-world scenarios.

References

[1] W. Min, Z. Wang, J. Yang, C. Liu, and S. Jiang, "Vision-Based Fruit Recognition Via MCNNi-Scale Attention Cnn," SSRN Electronic Journal, 2022, doi: 10.2139/ssrn.4162668.

[2] S. Hamid and M. Walia, "Convolution Neural Network Based Image Recognition," International Journal of Science and Research (IJSR), vol. 10, no. 2, pp. 1673–1677, Feb. 2021, doi: 10.21275/sr21225214136

[3]]L. Hou, Q. Wu, Q. Sun, H. Yang, and P. Li, "Fruit recognition based on convolution neural network," 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), pp. 18–22, Aug. 2016, doi: 10.1109/fskd.2016.7603144.

[4] Y. Zhan, Y. Chen, X. Li, M. Chen, and Z. Luo, "Fruit Recognition Based on Convolution Neural Network," Journal of Physics: Conference Series, vol. 1651, no. 1, p. 012176, Nov. 2020, doi: 10.1088/1742-6596/1651/1/012176

[5] "Fruit and Vegetable Recognition and Nutrition Analysis Based on Multilayer Perceptron and Neural Network," International Journal of Neural Network, vol. 1, no. 4, Dec. 2020, doi: 10.38007/nn.2020.010406.

[6] M. B. Sardar, "Fruit Recognition and its Calorie Measurement: An Image Processing Approach," International Journal Of Engineering And Computer Science, Oct. 2016, doi: 10.18535/ijecs/v5i10.56.

[7] D. M. Bongulwar, "Fruit Recognition Using Deep Learning Approach," Research Developments in Science and Technology Vol. 6, pp. 46–52, May 2022, doi: 10.9734/bpi/rdst/v6/15821d.

[8] A. Marin and E. Radoi, “Image-based Fruit Recognition and Classification,” 2022 21st RoEduNet Conference: Networking in Education and Research (RoEduNet), Sep. 2022, doi: 10.1109/roedunet57163.2022.9921050.
[9] Q. He, K. Xia, and H. Pan, “Fruit Recognition Using Color Statistics,” 2022 International Conference on Automation, Robotics and Computer Engineering (ICARCE), pp. 1–4, Dec. 2022, doi: 10.1109/icarce55724.2022.10046437.
[10] Y. Ayman, N. ElShabasy, and N. Nazih, “A Deep Learning Approach for Fruit Recognition and Nutrition Detection,” 2024 International Telecommunications Conference (ITC-Egypt), pp. 1–6, Jul. 2024, doi: 10.1109/itc-egypt61547.2024.10620582.
[11] Q. He, K. Xia, and H. Pan, “Fruit Recognition Using Color Statistics,” 2022 International Conference on Automation, Robotics and Computer Engineering (ICARCE), pp. 1–4, Dec. 2022, doi: 10.1109/icarce55724.2022.10046437.
[12] S. Jana, S. Basak, and R. Parekh, “Automatic fruit recognition from natural images using color and texture features,” 2017 Devices for Integrated Circuit (DevIC), pp. 620–624, Mar. 2017, doi: 10.1109/devic.2017.8074025.
[13] H. Mureşan and M. Oltean, “Fruit recognition from images using deep learning,” Acta Universitatis Sapientiae, Informatica, vol. 10, no. 1, pp. 26–42, Aug. 2018, doi: 10.2478/ausi-2018-0002.
[14] H. S. Gill, G. Murugesan, B. S. Khehra, G. S. Sajja, G. Gupta, and A. Bhatt, “Fruit recognition from images using deep learning applications,” Multimedia Tools and Applications, vol. 81, no. 23, pp. 33269–33290, Apr. 2022, doi: 10.1007/s11042-022-12868-2.
[15] D. Thi Phuong Chung and D. Van Tai, “A fruits recognition system based on a modern deep learning technique,” Journal of Physics: Conference Series, vol. 1327, no. 1, p. 012050, Oct. 2019, doi: 10.1088/1742-6596/1327/1/012050.
[16] Swapnil Srivastava, “A Fruit Recognition System based on Modern Deep Learning Technique,” International Journal of Engineering Research and, vol. V9, no. 07, Jul. 2020, doi: 10.17577/ijertv9is070352.
[17] S. C, “Robust Deep Learning Based Fruit Recognition System for Autonomous Harvesting System in Complex Cashew Orchards,” 2025, doi: 10.2139/ssrn.5197538.
[18] Z. Xu, J. Chen, X. Wu, and C. Tian, “Intelligent Recognition of Packed Fruits Based on Deep Learning,” 2024 9th International Conference on Image, Vision and Computing (ICIVC), pp. 67–73, Jul. 2024, doi: 10.1109/icivc61627.2024.10837532
[19] J. Ji, X. Zhu, H. Ma, H. Wang, X. Jin, and K. Zhao, “Apple Fruit Recognition Based on a Deep Learning Algorithm Using an Improved Lightweight Network,” Applied Engineering in Agriculture, vol. 37, no. 1, pp. 123–134, 2021, doi: 10.13031/aea.14041.
[20] P. Zhou, L. Zheng, M. Wang, and M. Li, “A Lightweight Apple Fruit Instance Segmentation Network: YOLO-AppleSeg,” Proceedings of the International Conference on Computer Vision and Deep Learning, pp. 1–7, Jan. 2024, doi: 10.1145/3653781.3653825.
[21] Z. Zhao, J. Wang, and H. Zhao, “Research on Apple Recognition Algorithm in Complex Orchard Environment Based on Deep Learning,” Sensors, vol. 23, no. 12, p. 5425, Jun. 2023, doi: 10.3390/s23125425.
[22] Y. Yang et al., “A New Kiwi Fruit Detection Algorithm Based on an Improved Lightweight Network,” Agriculture, vol. 14, no. 10, p. 1823, Oct. 2024, doi: 10.3390/agriculture14101823.