Automatic Fruit and Vegetable Classification Using Deep Learning and Computer Vision: An Approach for Intelligent Weighing Systems

Daniel Lloyd, Thiago Borsoni and Bernardo Pinto

Abstract— **Automatic fruit and vegetable recognition is an essential component of intelligent weighing systems in supermarkets and self-service stores. These systems must operate under strict latency and hardware constraints while maintaining high classification accuracy under varying lighting, backgrounds, and occlusions. This work investigates compact deep-learning classifiers using the *Fruits and Vegetables Detection for YOLOv4* dataset. We build a YOLOv8s-based pipeline trained on high-resolution RGB images labeled by product type and the presence of plastic wrapping (bagged vs. non-bagged). First, we merge bag information into eight semantic classes (*chilli, lemon, banana, apple, tomato, grapes, raspberry, blackberries*), achieving 100% accuracy. We then refine the taxonomy to fourteen classes by splitting six categories into with-bag and without-bag variants, while keeping raspberries and blackberries only in the non-bag condition; the model again reaches 100% accuracy. We report a mean latency of $193\,\mathrm{ms}$ per image and a model size of $9.81\,\mathrm{MB}$, and show that predictions remain stable under strong brightness changes and partial occlusions. Finally, we discuss overfitting risks and outline cross-dataset validation and embedded deployment as next steps.**

Keywords— **Computer Vision, Deep Learning, Fruits, Vegetables, Classification, YOLOv8s.**

I. INTRODUCTION

Automatic classification of fruits and vegetables plays a key role in intelligent weighing systems and retail automation. In self-service scenarios, visual identification removes the need for manual barcode entry, reducing errors and customer waiting times. Real deployments, however, must cope with variations in lighting, reflections from plastic bags, overlapping items, and inconsistent camera perspectives, which can degrade the accuracy of traditional machine-learning approaches and call for lightweight deep-learning models that balance precision and efficiency [1].

Compact convolutional neural networks (CNNs) such as MobileNetV2 and EfficientNet have shown strong performance when fine-tuned on domain-specific datasets [2], and integrating color information via histograms or clustering can improve robustness under translucency and occlusion [3]. YOLO-based architectures further combine detection and recognition for real-time inference on embedded hardware, yet challenges such as dataset imbalance, label inconsistency, and limited cross-domain generalization remain [4], [5]. Recent advances, including more challenging datasets [6] and optimized CNN pipelines [7], [8], reinforce the need for adaptable, high-performance models capable of operating in dynamic retail and agricultural environments.

In this context, we propose an end-to-end pipeline for automatic fruit and vegetable classification tailored for edge deployment. Using the Fruits and Vegetable Detection for YOLOv4 dataset [9], we train a YOLOv8s classifier to distinguish fourteen categories under both bagged and non-bagged conditions.

II. SYSTEM DESCRIPTION AND PROBLEM DEFINITION

Intelligent weighing systems rely on visual input to identify produce items placed on the scale. In many existing setups, users must manually select the product type from a list—introducing human error and slowing service throughput. The system developed in this research aims to eliminate this manual step by deploying a lightweight neural classifier capable of automatically identifying the item in real time. The problem can be formulated as a multi-class image classification task, where each input frame from the weighing camera must be assigned to one of the known product categories. Our target environment imposes three constraints:

Low latency – inference must complete within milliseconds to maintain the user experience. In our experiments, the YOLOv8s classifier processed a single image in a mean time of $193.24\,\mathrm{ms}$ over 30 runs (best $158.62\,\mathrm{ms}$, worst $388.46\,\mathrm{ms}$), which is acceptable for interactive kiosk scenarios, although further optimization may be required for very low-power edge devices or high-throughput setups.

Limited memory footprint – the model must be compatible with embedded GPUs or mobile SoCs. The final checkpoint (`best.pt`) occupies only $9.81\,\mathrm{MB}$, indicating that the classifier is lightweight enough to fit comfortably alongside other application components on typical embedded platforms.

Visual variability – robustness to plastic bag translucency, occlusions, and lighting changes. Stress tests on held-out images showed that the classifier consistently predicted the correct class (*grapes_with_bag*) with confidence 1.00 under strong brightening, darkening, and the insertion of a synthetic occlusion block, suggesting promising robustness to illumination shifts and partial masking in this controlled setting.

To address these constraints, a compact YOLOv8s architecture was selected for training and evaluation, as it balances accuracy and computational efficiency. In parallel, we annotated a binary "Bag" attribute (bagged/unbagged) using filename patterns to assess the potential influence of packaging on classification accuracy.

Fig. 1. Grapes in bag with different visual variability

III. Methodology

A. Dataset Preparation

The experiments were carried out using the Fruits and Vegetable Detection for YOLOv4 dataset [1], which contains thousands of labeled images representing a variety of fruit and vegetable categories. Class labels were extracted from filenames with a regular-expression parser, and each image was additionally annotated with a binary attribute indicating whether the product was bagged (`wb`) or unbagged (`wob`). After cleaning and preprocessing, we adopted a fine-grained taxonomy with fourteen classes: for *chilli, lemon, banana, apple, tomato* and *grapes* we explicitly distinguished *with_bag* and *without_bag* variants, while for *raspberry* and *blackberries* only the without-bag condition was retained due to data availability. This decision allows the model to learn the visual impact of plastic wrapping without discarding the limited categories that appear exclusively without a bag. The final dataset comprises 4 692 images, with class frequencies ranging from 140 to 728 samples, split into training and validation sets using stratified sampling (70% for training and 30% for validation) to preserve label balance across splits.

B. Model and Training Setup

A pretrained YOLOv8s classifier (yolov8s-cls.pt) was fine-tuned over 50 epochs to adapt it to the task. The training was conducted with a batch size of 16 and an input image resolution of 224×224 pixels, using the Adam optimizer. The process took place in a GPU-enabled environment with CUDA support, ensuring efficient computation. Throughout training, progress was tracked using the Ultralytics logging framework. Upon completion, the best-performing model checkpoint (best.pt) was selected for validation.

C. Evaluation Metrics

To evaluate the model's performance, several metrics were employed, including overall accuracy, as well as precision, recall, and F1-score calculated for each class.

D. Methodology Workflow

In the proposed workflow, **Dataset Preparation** gathers images from the original Kaggle folders, extracts class labels via regular expressions, derives the `wb/wob` (bagged/unbagged) attribute, and maps everything to fourteen classes (six products with both with-bag and without-bag variants, plus raspberries and blackberries only without bags). After cleaning, the 4 692 images are stratified into 70% training and 30% validation and copied into class-specific folders for YOLO classification. **Model and Training Setup** fine-tunes a pretrained YOLOv8s classifier (`yolov8s-cls.pt`) on 224×224 RGB crops for 10 epochs (batch size 16) on a CUDA-enabled GPU, using the Ultralytics loop to track accuracy and loss and saving the best checkpoint as `best.pt`. **Evaluation Metrics** compute accuracy, precision, recall, F1-score. Finally, **Final model evaluated and results visualized** combines these quantitative results with qualitative examples under different lighting and occlusion patterns, providing interpretable evidence to support deployment decisions.

IV. Results

The model achieved remarkable performance, demonstrating near-perfect convergence within the first few epochs.

A. Qualitative Analysis

Qualitative inspection of predictions classified both bagged and non-bagged fruits. Even in the presence of transparent, blue or black bags, as well as overlapping objects, detections remained robust and visually precise.

B. Quantitative Results

TABLE I. YOLOv8 Classification Model Performance Metrics for 8 classes

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000
F1-score	1.000
Validation images	1378

TABLE II. YOLOv8 classification performance for 14 fine-grained classes (with/without-bag variants).

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000
F1-score	1.000
Validation images	1378

V. Threats to Validity

Despite the encouraging results, several factors may limit the model's ability to generalize. The dataset is relatively small and visually homogeneous (similar lighting, distance, and angle), so achieving 100% accuracy in both the 8-class and 14-class settings may reflect overfitting to specific

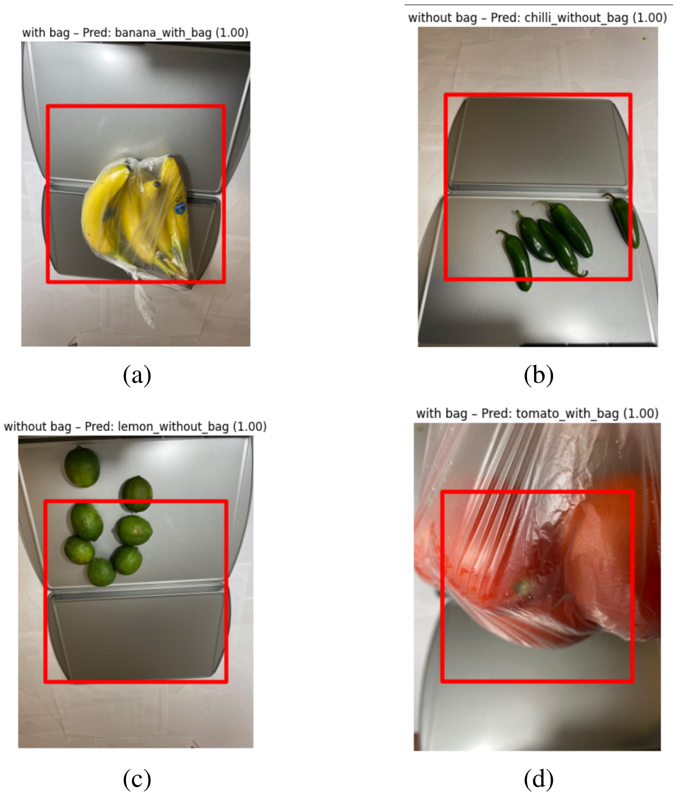

Fig. 2. Representative validation images with and without plastic bags, overlaid with a fixed central bounding box used for visualization. From left to right and top to bottom: *banana_with_bag*, *chilli_without_bag*, *lemon_without_bag*, and *tomato_with_bag*; in all cases the YOLOv8s classifier predicts the correct label with confidence 1.00.

acquisition conditions rather than true robustness. The class distribution is also imbalanced: categories such as *raspberry* and *blackberries* are underrepresented and appear only in the without-bag condition, making it difficult to assess how well the with/without-bag split scales to rarer products. In addition, bag information is derived heuristically from filename patterns (e.g., `wb` vs. `wob`), which may not always match the actual visual presence of plastic wrapping, introducing potential label noise. Finally, the model has only been evaluated on a GPU-enabled environment and on a single dataset, so performance on embedded hardware and under different camera characteristics remains uncertain, motivating further work on data scaling, domain adaptation, and hardware-aware optimization.

VI. Conclusions

The proposed YOLOv8-based classifier achieved **100% accuracy** on the validation set for both an 8-class configuration (ignoring bag information) and a 14-class setting that distinguishes with-bag and without-bag variants for six products, while keeping raspberries and blackberries only in the non-bag condition. The final checkpoint is lightweight ($9.81\,\mathrm{MB}$), with single-image inference around $193\,\mathrm{ms}$ on a GPU, and stress tests under strong brightness changes and synthetic occlusions showed stable predictions, suggesting promising robustness to realistic visual variability.

Nonetheless, the combination of very high scores and a limited, relatively uniform dataset raises concerns about overfitting and potential label noise from filename-based bag annotations. As future work, we plan to expand the dataset to multiple stores and camera setups, perform cross-dataset evaluations to measure domain shift, benchmark inference on embedded hardware, and explore architectures that can handle scenes with multiple items, moving toward a robust, deployable solution for real-world supermarket weighing systems.

References

[1] K. Patel, "Fruits & vegetable detection for yolov4," https://www.kaggle.com/datasets/kvnpatel/fruits-vegetable-detection-for-yolov4, 2021.

[2] J. L. Rojas-Aranda, J. I. Nunez-Varela, J. C. Cuevas-Tello, and G. Rangel-Ramirez, "Fruit classification for retail stores using deep learning," in *Mexican Conference on Pattern Recognition*. Springer, 2020, pp. 3–13. [Online]. Available: https://doi.org/10.1007/978-3-030-49076-8_1

[3] H. S. Gill, G. Murugesan, A. Mehbodniya, G. S. Sajja, G. Gupta, and A. Bhatt, "Fruit type classification using deep learning and feature fusion," *Computers and Electronics in Agriculture*, vol. 211, p. 107990, 2023. [Online]. Available: https://doi.org/10.1016/j.compag.2023.107990

[4] M. S. Hossain, M. Al-Hammadi, and G. Muhammad, "Automatic fruit classification using deep learning for industrial applications," *IEEE transactions on industrial informatics*, vol. 15, no. 2, pp. 1027–1034, 2018. [Online]. Available: https://doi.org/10.1109/TII.2018.2875149

[5] Y. Gulzar, Z. Ünal, S. Ayoub, F. A. Reegu, and A. Altulihan, "Adaptability of deep learning: datasets and strategies in fruit classification," in *BIO Web of Conferences*, vol. 85. EDP Sciences, 2024, p. 01020. [Online]. Available: https://doi.org/10.1051/bioconf/20248501020

[6] N. Mamat, M. F. Othman, R. Abdulghafor, A. A. Alwan, and Y. Gulzar, "Enhancing image annotation technique of fruit classification using a deep learning approach," *Sustainability*, vol. 15, no. 2, p. 901, 2023. [Online]. Available: https://doi.org/10.3390/su15020901

[7] J. L. Joseph, V. A. Kumar, and S. P. Mathew, "Fruit classification using deep learning," in *Innovations in Electrical and Electronic Engineering: Proceedings of ICEEE 2021*. Springer, 2021, pp. 807–817. [Online]. Available: https://doi.org/10.1007/978-981-16-0749-3_62

[8] S. Khanna, C. Chattopadhyay, and S. Kundu, "Enhancing fruit and vegetable detection in unconstrained environment with a novel dataset," *Scientia Horticulturae*, vol. 338, p. 113580, Dec. 2024. [Online]. Available: http://dx.doi.org/10.1016/j.scienta.2024.113580

[9] P. Kamat, S. Gite, H. Chandekar *et al.*, "Multi-class fruit ripeness detection using yolo and ssd object detection models," *SN Applied Sciences*, vol. 7, 2025. [Online]. Available: https://doi.org/10.1007/s42452-025-07617-7

[10] M. Mukhiddinov, "Improved classification approach for fruits and vegetables," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–10, 2022. [Online]. Available: https://doi.org/10.3390/s22218192