

Minimizing Checkout Time: A Vision-Based Solution for Produce Identification

Rodrigo Ferrett, João Pedro Santana, Pedro Silvestre, Lucas Alves, Sarah Ferrari, and Rigel P. Fernandes 

Abstract—This project focuses on the development of a self-checkout system for produce. The system integrates a camera and a digital scale to automate the process of identifying and pricing fruits and vegetables. The camera captures images of the produce to determine its type using image recognition techniques, while the scale measures its weight. By combining these two sources of data, the system automatically calculates the price of the selected product, streamlines the checkout process, reduces human error, and improves efficiency in retail environments. The proposed system measures and recognizes produce in 4 seconds in average.

Keywords—Image Recognition, Digital Scale Integration, Automated Checkout, Produce Classification, Real-Time Processing, Retail Automation

I. INTRODUCTION

Automated checkout systems are transforming the retail experience by reducing wait times and minimizing human error, particularly in environments where speed and accuracy are critical [1]. Recent advances in computer vision and embedded systems have made it possible to integrate cameras, sensors, and lightweight deep learning models to recognize items in real time, even under hardware constraints [2]. Solutions such as ARC have demonstrated the potential of vision-based checkout systems to eliminate the need for barcodes by leveraging convolutional neural networks for object classification [3].

The implementation of image recognition systems for fresh produce faces unique challenges, such as the high variability in the appearance, color, shape, and ripeness of fruits and vegetables. Furthermore, lighting conditions in the retail environment can significantly affect classification accuracy. To overcome these obstacles, many researchers have turned to deep learning models, especially Convolutional Neural Networks (CNNs), which have proven highly effective at extracting complex features from images for classification tasks [4].

Models with lightweight architectures, such as MobileNetV2, are particularly well-suited for embedded systems as they offer a good balance between accuracy and computational efficiency, enabling real-time processing on low-cost hardware [5]. The effectiveness of these models depends on robust training with large datasets that capture the diversity of produce in various scenarios. Enhanced approaches, using variations of models like YOLO, have also been developed not

only to classify products but also to assess their quality and freshness, adding even more value to the automated checkout process [6].

This paper introduces a Vision-based Self-Checkout for Produce system, an embedded self-checkout solution that combines image recognition and weight measurement to automatically identify and price produce. Our key contributions include: (1) Implementation of a lightweight image recognition model optimized for embedded hardware; and (2) Integration of real-time weight data to ensure accurate and efficient pricing for diverse types of produce.

II. THE PROPOSED METHOD

Figure 1 shows the sequence diagram for the automated process of weighing and identifying a product at a checkout system. The process begins when the Customer places an item on the scale, prompting the Cashier System to request a stable weight from the Digital Scale. Once the weight is received, the Cashier System commands the Camera to capture an image of the product. This image is then sent to the AI Service (Recognition), which analyzes it to identify the product type, such as an "Apple." With the product name in hand, the system queries the Price Database to retrieve the corresponding price per unit or kilogram. Finally, the Cashier System calculates the total price, displays the product's details—including its name, weight, and final price—and awaits the customer's confirmation to add or remove the item from the cart.

A. System architecture

The methodology for this project is based on a hybrid, two-part architecture that leverages the strengths of both a micro-controller for real-time sensor reading and a host computer for intensive computational tasks. The process was structured into four main stages: (1) designing the hybrid system architecture, (2) specifying and assembling the distinct hardware units, (3) implementing the firmware and software for each component, and (4) creating the dataset and training the classification model for deployment on the host system.

Inspired by vision-based automatic checkout systems that aim to eliminate the need for barcodes [1] [3] our system was designed with a modular architecture. The core of the system integrates two real-time data sources: a weighing module to determine the precise mass of the product and a computer vision module for its identification. Both modules operate in parallel and send their data to a central processing unit, which is responsible for consolidating the information, calculating the price, and presenting the final result to the user.

Rodrigo de Sousa Ferrett, João Pedro Santana, Pedro Silvestre, Lucas Alves, Sarah Ferrari, and Rigel P. Fernandes are affiliated with the Tech programs at the IbmeC University Center, Rio de Janeiro, Brazil. Emails: {rodrigoferrett, jpedrossantana, phsmmb, lucasacastela, sarahferrari, rigelfernandes}@gmail.com.

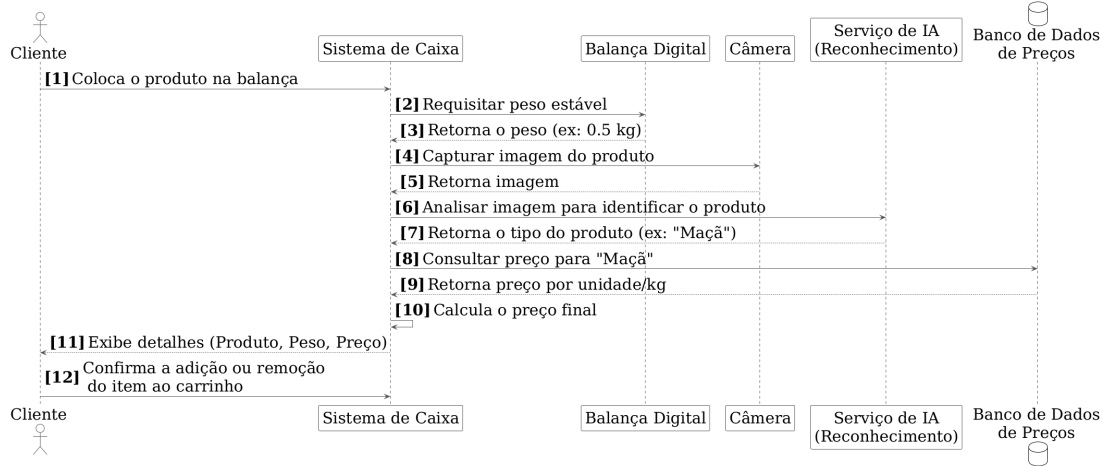


Fig. 1. Diagrama sequencial da balança.

The integration of real-time weight sensing with a dedicated classification model follows the architecture of other successful smart sensor-based measurement systems, even in diverse applications such as advanced environmental monitoring. [7]

- **The Sensor Unit:** powered by an Arduino Mega 2560, is solely responsible for interfacing with the weighing module (load cell and HX711). Its dedicated task is to perform precise weight measurements and stream this data continuously to the host.
- **The Host Processing Unit (a PC or single-board computer):** serves as the main brain of the operation. It receives weight data from the Sensor Unit via a serial connection. Simultaneously, it captures images using an attached camera and runs the deep learning model to perform product classification. The host unit then integrates the two data streams (product identity and weight) to calculate the final price. This distributed architecture ensures reliable real-time sensor readings without compromising the performance needed for computer vision tasks [2].

B. Hardware Components

Figure 2 depicts the hardware components that were selected to create a distinct separation between main processing tasks and real-time sensor data acquisition [2].

- **Arduino Mega 2560:** A dedicated microcontroller for polling sensors and serial communication.
- **Load Cell and HX711 A/D Converter Module:** The load cell (a Wheatstone bridge strain gauge) measures weight, and the HX711 ADC amplifies and digitizes the signal.
- **Physical Structure:** A 3D-printed enclosure houses the Weighing Assembly and the Sensor Control Unit, creating a stable platform.

C. Dataset and Model Training

The performance of the system's identification capability is dependent on the vision model. To enhance results a

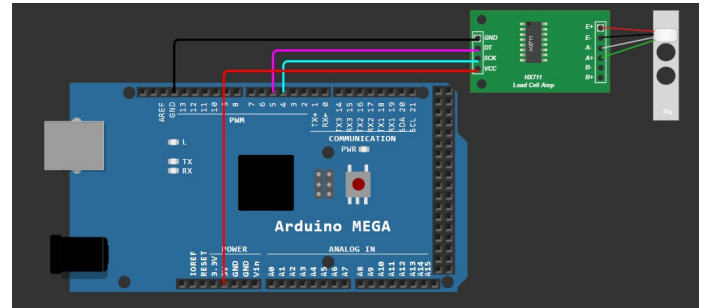


Fig. 2. Arduino Mega 2560.

data collection and augmentation step was performed. A custom dataset was built with [Number of Images] images across [Number of Classes] different types of produce. To mitigate the challenges of appearance variability and lighting conditions [5] [5], [6], data augmentation techniques were applied, including random rotations, zooms, and brightness adjustments.

The transfer learning technique was used, fine-tuning a pre-trained MobileNetV2 model with our specific dataset. This approach is well-suited for vision tasks on custom datasets [4]. The training was conducted in a GPU-enabled environment. The final validated model was then converted to the TensorFlow Lite (.tflite) format for efficient deployment on the Host Processing Unit.

III. RESULTS

The initial experimental phase is designed to validate the transition from the virtual model to a functional prototype, focusing on data integration and AI inference performance under hardware constraints.

The early configuration will prioritize modular testing across three core components: the Sensor Unit, the Host Processing Unit, and the system's database.

1) *Initial Experiment Goals:* The first experiments will be structured to achieve the following: (1) Validate Sensor Real-Time Constraints: Measure the time required for the Sensor Unit (Arduino Mega 2560 and Load Cell) to achieve stable

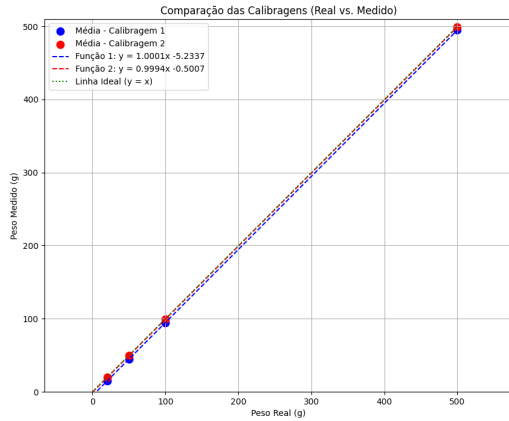


Fig. 3. Graphics of the realized tests

TABLE I. Tests results and error margin, the most distance datas before and after calibration step

Real Weight (g)	Measured Reading 1 (g)	Measured Reading 2 (g)	Average Reading (g)	Absolute error (g)
500.00	494.82	494.85	494.835	5.165
100.00	94.92	94.65	94.785	5.215
50.00	44.86	44.73	44.795	5.205
20.00	14.66	14.83	14.745	5.255

Real Weight (g)	Measured Reading 1 (g)	Measured Reading 2 (g)	Average Reading (g)	Absolute error (g)
500.00	499.09	499.36	499.22	0.775
100.00	99.30	99.46	99.38	0.620
50.00	49.48	49.57	49.525	0.475
20.00	19.39	19.59	19.49	0.510

weight readings and transmit data consistently to the Host. (2) Measure AI Inference Performance: Test the Host Processing Unit's ability to execute the MobileNetV2 classification model efficiently, measuring the latency of the image recognition process. (3) Test End-to-End Integration: Conduct tests that combine weight data, product ID from the AI, and a reference price from the database (which must be created) to successfully calculate the final price.

This section details the results of a two-phase calibrations and tests process using specific weights. with the objective of correcting the initial inaccuracies and confirming the performance of the system.

The test procedure involved measuring a set of certified reference weights (20g, 50g, 100g, 500g). For each reference weight, two separate readings were taken to ensure consistency. This process was executed twice, first to establish parameters and a second time to confirm the data of the first test.

We conducted 10 consecutive purchases and measured the time from when the food was placed on the scale until the purchase was finalized. We obtained the following in seconds: [4.03, 4.55, 3.93, 4.38, 3.41, 4.52, 4.35, 3.06, 4.42, 3.54, 3.46] Resulting in a mean of 4.02 seconds with a standard deviation of 0.44 seconds.

The performance of the system was validated by testing its three core components: weighing accuracy, recognition

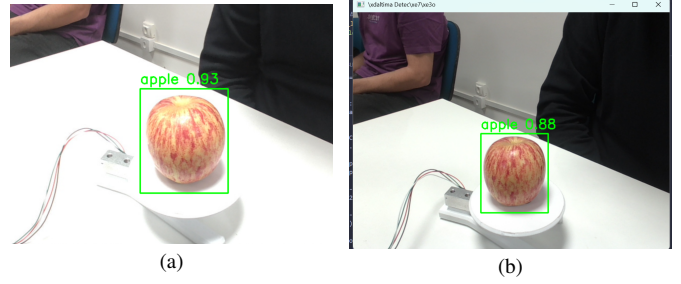


Fig. 4. Apple tests (a) (b)

performance, and end-to-end checkout time.

After a two calibration process using certified weight, the scale measured a high precision, post calibration, the absolute error was minimal, such as -0.620g for a 100g weight and -0.775g for a 500g weight. The calibration graph (Fig. 3) confirms a strong linear relationship between the real and measured weights.

The total checkout time was measured over 10 consecutive purchases. The system recorded a mean processing time of 4.02 seconds from placing the item to the final price display, with a low standard deviation of 0.44 seconds. This confirms the system is fast and consistent.

IV. CONCLUSIONS

This paper presents a vision-based self-checkout system that automates identification by fusing image recognition with precise weight data. The hybrid architecture proved highly effective, with the calibrated weighing module showing minimal error (e.g. -0.62 g on 100g) and the end-to-end system processing purchases in just 4.02 seconds on average. Future work will build a physical prototype, expand the product database, and refine the classification model for more varieties of produce.

REFERENCES

- [1] M. Shoman, A. Aboah, A. Morehead, Y. Duan, A. Daud, and Y. Adu-Gyamfi, "A region-based deep learning approach to automated retail checkout," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2022, pp. 3210–3215. [Online]. Available: <https://doi.org/10.48550/arXiv.2204.08584>
- [2] J. Fang and X. Zhang, "Embedded image recognition system for lightweight convolutional neural networks," ser. AIAM2021. New York, NY, USA: Association for Computing Machinery, 2022, p. 1373–1377. [Online]. Available: <https://doi.org/10.1145/3495018.3495401>
- [3] S. Talha Bukhari, A. Wahab Amin, M. Abdullah Naveed, and M. Rzi Abbas, "ARC: A Vision-based Automatic Retail Checkout System," *arXiv e-prints*, p. arXiv:2104.02832, Apr. 2021. [Online]. Available: <https://doi.org/10.48550/arXiv.2104.02832>
- [4] W. Wang, A. Zhu, H. Wei, and L. Yu, "A novel method for vegetable and fruit classification based on using diffusion maps and machine learning," *Current Research in Food Science*, vol. 8, p. 100737, 2024. [Online]. Available: <https://doi.org/10.1016/j.crfs.2024.100737>
- [5] E. Talha-Mendez, I. A. Cruz-Albarran, S. Tovar-Arriaga, and L. A. Morales-Hernandez, "Deep learning-based method for classification and ripeness assessment of fruits and vegetables," *Applied Sciences*, vol. 13, no. 22, 2023. [Online]. Available: <https://doi.org/10.3390/app132212504>
- [6] M. Mukhiddinov, A. Muminov, and J. Cho, "Improved classification approach for fruits and vegetables freshness based on deep learning," *Sensors*, vol. 22, no. 21, 2022. [Online]. Available: <https://doi.org/10.3390/s22218192>
- [7] S. Cecchi, S. Spinsante, A. Terenzi, and S. Orcioni, "A smart sensor-based measurement system for advanced bee hive monitoring," vol. 20, no. 9, p. 2726, 2020. [Online]. Available: <https://doi.org/10.3390/s20092726>