Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Artificial Intelligence (37^{th}cycle)

A computer vision framework to transform produce waste into value for agricultural sustainability

By

Mirko Agarla

******

Supervisor(s):
Prof. Paolo Napoletano, Supervisor
Prof. Raimondo Schettini, Co-Supervisor

Doctoral Examination Committee:
Prof. Tatiana Tommasi, Politecnico di Torino
Prof. Claudio Cusano, Università di Pavia
Prof. Gabriele Gianini, Università di Milano-Bicocca
Prof. Giuseppe Boccignone, Università di Milano
Prof. Francesco Bianconi, Università di Perugia

Politecnico di Torino
2025

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Mirko Agarla
2025

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Contents

1 Introduction

Fruits and vegetables are sources of nutrients that promote health and well-being. The growing global demand for fresh produce introduces significant challenges. The market is projected to expand at an annual rate of 9.3% between 2023 and 2033 [1], but increased production also leads to higher levels of waste. In the United States, it is estimated that 30% to 40% of fruit is wasted during harvesting and processing [2]. This waste not only reduces valuable food supplies but also consume resources such as water, energy, and human labor [3].

Traditional food quality control methods have relied on manual inspections and chemical tests, which are time-consuming, labor-intensive, and prone to human error [4–6]. Current computer vision approaches are expensive and tailored to specific tasks in controlled environments [7, 8]. The most common methods use expensive hardware such as hyperspectral cameras, limiting their broader applicability [9, 10]. With a declining workforce and rising costs, these methods are becoming increasingly impractical.

This research presents a general computer vision framework designed to reduce waste by transforming it into value. The framework automates the processes of detecting, sorting, and assessing the quality of fruits and vegetables, enabling the recovery of good parts of produce rather than discarding entire items due to minor defects. To demonstrate the practical impact of this research, I developed a prototype system in collaboration with experts from Politecnico di Torino. The system can identify and remove damaged areas from apples, allowing the healthy portions to be used for products such as juice, dried fruit, or desserts. The proposed framework enhances the efficiency of quality inspections and supports sustainable agricultural practices. The overall framework is a low-cost solution that can be easily implemented in real farming conditions, helping large-scale and small-scale producers maintain high-quality standards while reducing environmental impact.

2 Proposed framework

The proposed framework for produce waste reduction can be divided into five main parts, as shown in Figure 1.

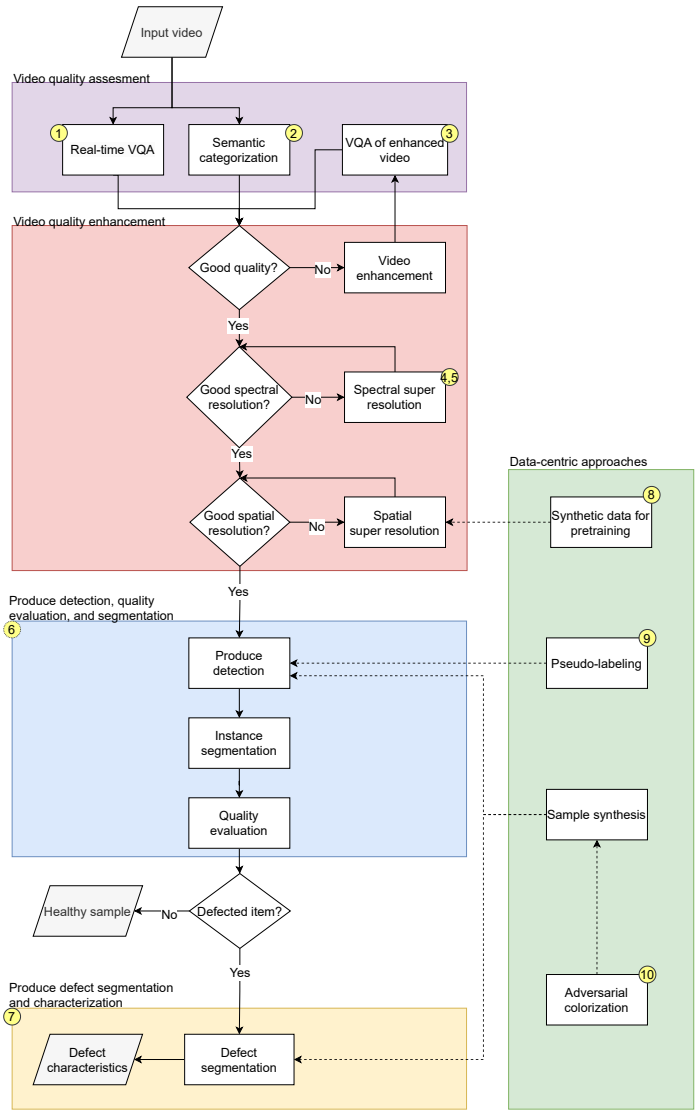

Fig. 1 Overview of the proposed framework for automated food quality inspection. Video quality assessment: (1) real-time video quality assessment [11], (2) semantic categorization [12], (3) video quality assessment of enhanced video [13]. Video quality enhancement: (4) spectral metrics [14], (5) spectral super resolution [15], video quality enhancement, spatial super resolution. (6) Produce detection, segmentation, and quality evaluation. (7) Produce defect segmentation and characterization [16]. Data-centric approaches: (8) synthetic data for hyperspectral pretraining [17], (9) pseudo-labeling [18], (10) adversarial colorization [19].

The process starts with **Video Quality Assessment**:

- **(1) Real-time video quality assessment** [11] evaluates the quality of incoming video frames considering issues introduced by environment like poor lighting or low cost cameras like motion blur, poor colors and shakiness.
- **(2) Semantic categorization** [12] adjusts quality evaluations by considering the context, such as distinguishing between indoor and outdoor scenes.
- **(3) Video quality assessment of enhanced video** [13] ensures that any improvements made to the video do not introduce new artifacts.

Then **Video quality enhancement** exploits the quality assessment results to optimize the video signal:

- **Video enhancement** improves and normalizes the video under varying conditions.
- **(4) Spectral super-resolution** [14, 15] converts standard RGB images into hyperspectral images with up to 31 bands to reveal more detailed information. (5) To evaluate spectral reconstruction, a method understands the correlation of spectral metrics with signal quality [14].
- **Spatial super-resolution** increases the resolution of RGB and hyperspectral images to capture finer details.

With enhanced video data, the framework moves to **Produce detection, segmentation, and quality evaluation**. Here, advanced segmentation techniques are developed to detect and isolate individual produce items, which are then evaluated based on size, shape and color.[1]

In the final part, the system focuses on **Produce defect segmentation and characterization** [16], where, exploiting the isolated produce, it precisely identifies and measures defective areas. This detailed analysis supports decisions on whether parts of the produce can be recovered or if the entire item should be discarded.

To support the whole framework **Data-centric approaches** address the issue of limited training data:

[1] The paper about this work is under construction.

- **Synthetic data for hyperspectral pretraining** [17] generates artificial hyperspectral images from existing RGB data to pre-train hyperspectral super-resolution models.
- **Pseudo-labeling** [18] automatically labels new data samples for classification and detection scenarios, enriching the training set.
- **Adversarial colorization** [19] diversifies the dataset by modifying image colors to create new realistic and challenging samples to improve the accuracy of the network and generalization.
- **Sample synthesis** generates artificial examples of produce and defects [16], allowing the system to learn from a wider range of scenarios.

3 Research contributions and methodologies

This chapter provides a summary of the research contributions and findings for produce waste reduction. The proposed framework for automated produce quality control is designed as a modular system that addresses real-world challenges in video-based inspection within agricultural settings. In the following sections, we summarize the key contributions of each component of the framework and discuss the overall impact on improving accuracy, efficiency, and adaptability in produce quality control applications. While agricultural datasets were used to evaluate components like produce detection, segmentation, and defect characterization, our methodology also leveraged diverse real-world datasets for video quality assessment and enhancement. These broader datasets, which cover a wide range of distortions and variations, allowed us to extensively test the robustness and adaptability of our methods under challenging conditions in the absence of agricultural-specific data.

3.1 Video quality assessment

The first stage of the framework focuses on assessing video quality under real-world conditions. The developed real-time video quality assessment method [11] employs a sampling algorithm that eliminates temporal redundancy by selecting representative frames from a continuous video stream. These frames are then processed by

two lightweight convolutional neural networks (CNNs) that extract quality-related and semantic features. The frame-level features are aggregated into a video-level representation, and a Support Vector Regressor (SVR) maps these features to a final quality score. Experimental evaluations on four large-scale user-generated video databases demonstrate that the method generalizes well and is significantly more efficient than comparable approaches—achieving up to 185 FPS on an NVIDIA X Pascal GPU and 12 FPS on an Intel i7-770 CPU for 1080p videos.

Building upon this foundation, the proposed semantic categorization component [12] refines quality predictions by incorporating the scene context. A dedicated model categorizes video frames according to the 16 basic-level categories of the SUN397 dataset. By grouping videos based on scene content, the system reveals that the performance of No-Reference VQA methods is highly domain-dependent; some scenes prove more challenging than others. This insight has been crucial in enhancing the robustness of the overall quality assessment process, ensuring that the system maintains high accuracy regardless of environmental variations.

In addition, the quality assessment of enhanced videos component [13] validates that video enhancement techniques have improved the quality without introducing additional artifacts. Using a multi-encoder approach, features related to technical aspects, aesthetic quality, and attributes such as sharpness and contrast are extracted and combined. These features are then processed by an SVR to produce a final quality score. Results on the VDPVE dataset and performance in the NTIRE 2023 Quality Assessment of Video Enhancement Challenge (where the method ranked sixth) confirm that this approach outperforms state-of-the-art techniques.

3.2 Video quality enhancement

Video quality enhancement consists of advanced techniques to emphasize spectral and spatial super-resolution. In the spectral domain, the research presents a comprehensive analysis of fourteen commonly used spectral similarity measures using the Munsell atlas [14]. The measures were categorized by the type of error they capture, revealing asymmetries and sensitivities related to color properties such as Chroma and Lightness. This analysis guides the selection of appropriate metrics for accurate spectral reconstruction and has been made available publicly to promote reproducibility.

For spectral super-resolution, the Fast-n-Squeeze method is designed [15]. This hybrid approach first estimates a global RGB-to-spectral linear transformation matrix using low-level image features, then refines the output with a global scaling factor determined by a lightweight CNN. Validated in the NTIRE 2022 Spectral Reconstruction Challenge, Fast-n-Squeeze effectively reconstructs hyperspectral data from standard RGB images while achieving high inference speeds—up to 198.45 FPS on a GPU.

Spatial super resolution techniques enhance image detail beyond the limitations of current imaging hardware. While traditional RGB super-resolution methods reconstruct high-resolution images from low-resolution inputs, applying these techniques to hyperspectral imaging is more challenging due to the inherent trade-off between spectral and spatial resolution. Hyperspectral imaging offers rich spectral information valuable for applications such as agriculture and remote sensing, but high-resolution hyperspectral datasets are scarce and costly to obtain. This scarcity limits the effectiveness of existing super-resolution methods in data-scarce environments. To overcome these challenges, in Section 3.5 we introduce a data-centric approach that generates synthetic datasets to support and improve hyperspectral spatial super resolution.

3.3 Produce detection, segmentation, and quality evaluation

The framework's subsequent stage automates the detection, segmentation, and evaluation of produce items. State-of-the-art deep learning models, including YOLO and DETR, are utilized to accurately identify individual produce items within complex scenes. Once detected, these items are segmented to isolate each produce element, enabling a detailed evaluation of attributes such as size, shape and color. By integrating data-centric strategies like pseudo-labeling [18] and synthetic data generation [19], the system overcomes the common challenge of limited annotated data. Experimental results demonstrate that synthetic data augmentation and advanced architectures enhance detection and segmentation capabilities. DETR R101 trained with synthetic data emerged as the most effective method, achieving the highest F-scores for both detection and segmentation of produces while balancing computational cost and speed. Meanwhile, YOLO-based models trained with synthetic data provided a viable alternative for scenarios requiring lightweight, real-time solutions.

3.4 Produce defect segmentation and characterization

A major contribution of this thesis is the development of a specialized approach for the defect segmentation and characterization, with a focus on apple defects [16]. A U-shaped CNN with strategically placed skip-connections within the noise reduction block is employed to accurately isolate defective regions of each identified produce. To address the scarcity of annotated defect data, an innovative data synthesis technique was designed. This technique generates new training samples by extracting defects from non-healthy produce and inserting them into healthy images, with random rotations and warping transformations to increase variability. Experimental results reveal that the proposed method outperforms traditional hand-crafted techniques and other CNN-based approaches by up to 35% in F-score, while also demonstrating strong performance using both multispectral (RGB+NIR) and conventional RGB inputs. Moreover, the method operates in quasi real-time—achieving around 100 FPS on a GPU and nearly 7–8 FPS on a CPU—highlighting its potential for industrial application.

3.5 Data-centric approaches

Recognizing the limitation of scarce high-quality labeled data in agricultural applications, the framework incorporates several data-centric methods to improve model performance and generalization. One approach involves generating synthetic hyperspectral data for pretraining [17]. By converting high-resolution RGB images into plausible hyperspectral images using advanced spectral reconstruction techniques, a large-scale synthetic dataset is created. This dataset significantly improves the training of hyperspectral super-resolution models, as demonstrated by notable gains in standard quality metrics such as MPSNR.

Another innovative strategy is pseudo-labeling, which automatically generates labels for unlabeled data [18]. Initially applied in cross-lingual speech emotion recognition (SER) with Transformer-based methods, this approach was later extended to produce detection. In the SER experiments, the use of hard pseudo-labels yielded an average accuracy improvement of 40% compared to state-of-the-art methods, underscoring the versatility and effectiveness of pseudo-labeling in enhancing model performance across domains.

Finally, the adversarial colorization method, GLoFool [19], contributes to data augmentation and model robustness. GLoFool is a black-box approach that iteratively applies global enhancement and local color perturbations to generate adversarial images. Two variants—GLoFool-Q, which focuses on maintaining color fidelity and perceptual quality, and GLoFool-T, which is optimized for transferability and robustness. These new samples can be applied in two distinct ways. When targeted for adversarial purposes, it generates challenging samples that test and enhance the model's ability to handle unexpected variations, improving its robustness. When used in a non-adversarial approach, it modifies image colors to create diverse and realistic samples, expanding the training set and improving the model's generalization to different scenarios.

4 Conclusions and industrial integration

This Ph.D. research presented the development of a comprehensive, modular framework for automated produce quality control that combines artificial intelligence, robotics, and data-centric approaches. The framework achieves high accuracy in detection, segmentation, and defect removal. The system is evaluated in environments with limited annotated data and constrained computational resources.

A key strength of the proposed framework is its dynamic adaptability. Its modular architecture allows individual components to be activated or deactivated based on operational needs. The practical applicability of this approach is demonstrated by an apple defect removal prototype, which identifies and removes defects while preserving usable produce. This directly addresses the global challenge of food waste caused by minor imperfections.

The framework addresses the high cost associated with advanced imaging hardware. Traditional methods often depend on high-definition RGB or hyperspectral cameras, which increase operational costs and reduce large-scale deployment. In contrast, our approach leverages RGB-to-hyperspectral conversion and spatial super-resolution techniques to perform cost-effective spectral analysis using conventional cameras. By first applying video quality assessment techniques to verify signal integrity and then employing signal enhancement methods to improve data quality, the system obtains high-fidelity representations of produce that support precise detection and segmentation.

The scalability of the system is achieved through the interconnected design of the components. This design enables integration across multiple processes and supports data analysis throughout the industrial chain. This integrated structure enables effortless adaptation across different environments and industrial applications, making it suitable for a global produce processing market valued at approximately $70 billion.

The initial prototype serves as a proof of concept for practical deployment, and the ongoing patent highlights the commercial potential and innovation of the technology. The next phase of this research will focus on transitioning the system from a controlled laboratory setting to real-world operational processing lines. Strategic partnerships with manufacturers of food processing equipment and agricultural cooperatives will accelerate the deployment and refinement of the technology. The long-term vision is to establish a scalable, adaptable, and practical solution that transforms quality control processes in the produce processing industry. By optimizing resources, reducing waste, and improving operational efficiency, the framework contributes to a more sustainable global food supply chain and aligns with the ONU Sustainable Development Goals.

References

[1] Market.us. Fruit and vegetable market: 10-year projection from 2023 to 2033. Accessed: 2024-11-04, 2023. Market report.

[2] FDA. Food waste and loss statistics, 2017.

[3] OECD. *Improving Energy Efficiency in the Agro-food Chain*. OECD Green Growth Studies. OECD Publishing, Paris, 2017.

[4] Tadhg Brosnan and Da-Wen Sun. Improving quality inspection of food products by computer vision—a review. *Journal of Food Engineering*, 61(1):3–16, 2004. Applications of computer vision in the food industry.

[5] Sergio Cubero, Nuria Aleixos, Enrique Moltó, Juan Gómez-Sanchis, and Jose Blasco. Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. *Food and bioprocess technology*, 4:487–504, 2011.

[6] Yankun Peng and Sagar Dhakal. Optical methods and techniques for meat quality inspection. *Transactions of the ASABE*, 58(5):1371–1386, 2015.

[7] Naoshi Kondo. Automation on fruit and vegetable grading system and food traceability. *Trends in Food Science & Technology*, 21(3):145–152, 2010.

[8] Yuzhen Lu and Renfu Lu. Non-destructive defect detection of apples by spectroscopic and imaging technologies: a review. *Transactions of the ASABE*, 60(5):1765–1790, 2017.

[9] Erick Saldaña, Raúl Siche, Mariano Luján, and Roberto Quevedo. Computer vision applied to the inspection and quality control of fruits and vegetables. *Brazilian Journal of Food Technology*, 16:254–272, 2013.

[10] LA Weston and MM Barth. Preharvest factors affecting postharvest quality of vegetables. *HortScience*, 32(5):812–816, 1997.

[11] Mirko Agarla, Luigi Celona, and Raimondo Schettini. An efficient method for no-reference video quality assessment. *Journal of Imaging*, 7(3):55, 2021.

[12] Mirko Agarla, Luigi Celona, et al. On the semantic dependency of video quality assessment methods. In *Final program and proceedings-London imaging meeting*, volume 2021, pages 49–53, 2021.

[13] Mirko Agarla, Luigi Celona, Claudio Rota, and Raimondo Schettini. Quality assessment of enhanced videos guided by aesthetics and technical quality attributes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1533–1541, 2023.

[14] Mirko Agarla, Simone Bianco, Luigi Celona, Raimondo Schettini, and Mikhail Tchobanou. An analysis of spectral similarity measures. In *Color and Imaging Conference*, volume 2021, pages 300–305. Society for Imaging Science and Technology, 2021.

[15] Mirko Agarla, Simone Bianco, Marco Buzzelli, Luigi Celona, and Raimondo Schettini. Fast-n-squeeze: Towards real-time spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1132–1139, 2022.

[16] Mirko Agarla, Paolo Napoletano, and Raimondo Schettini. Quasi real-time apple defect segmentation using deep learning. *Sensors*, 23(18):7893, 2023.

[17] Emanuele Aiello, Mirko Agarla, Diego Valsesia, Paolo Napoletano, Tiziano Bianchi, Enrico Magli, and Raimondo Schettini. Synthetic data pretraining for hyperspectral image super-resolution. *IEEE Access*, 2024.

[18] Mirko Agarla, Simone Bianco, Luigi Celona, Paolo Napoletano, Alexey Petrovsky, Flavio Piccoli, Raimondo Schettini, and Ivan Shanin. Semi-supervised cross-lingual speech emotion recognition. *Expert Systems with Applications*, 237:121368, 2024.

[19] Mirko Agarla and Andrea Cavallaro. Glofool: Global enhancements and local perturbations to craft adversarial images. In *Computer Vision – ECCV 2024 Workshops*, pages 383–399. Springer Nature Switzerland, 2025.