Smart Vision Technology for Quality Control in E-Commerce

Dhruv Khichi

220150002

Abstract: This project aims to develop a robust smart vision system for quality control in e-commerce, ensuring efficient product assessment through advanced image recognition and analysis. The system focuses on item recognition, packaging inspection, fresh produce assessment, and text extraction via OCR for automated inventory management. The objective is to ensure products meet the required standards and help automate processes for more effective and efficient e-commerce operations.

1 Objective

This project aims to develop a robust smart vision sys- tem for quality control in e-commerce, ensuring efficient product assessment through advanced image recognition and analysis. The system focuses on:

- **Item Recognition:** Identifying unique products and categorizing them.
- **Packaging Inspection:** Ensuring labeling accu- racy and integrity.
- **Fresh Produce Assessment:** Determining qual- ity, freshness, and defects.
- **Text Extraction:** Using OCR for automated in- ventory management and traceability.

2 Methodology

The methodology of this project involves the develop- ment and deployment of various deep learning mod- els for quality control in e-commerce. The following sections detail the steps taken during the implemen- tation of the system, addressing the use of Convolu- tional Neural Networks (CNNs), YOLO (You Only Look Once), EfficientNet-B0, and Optical Character Recog- nition (OCR) for packed items.

2.1 Initial Approach: Convolutional Neu- ral Networks (CNNs)

Implementation: The initial approach used Convolu- tional Neural Networks (CNNs) for basic classification tasks. A dataset with 10 categories was collected and manually annotated to distinguish between fresh and rotten fruits/vegetables.

A CNN was built using Keras, incorporating con- volution, pooling, and dense layers to classify images. The model was designed to classify fruits and vegetables based on their quality.

- Data preprocessing included augmentation, nor- malization, and filtering to improve model gener- alization.
- The model was trained on the dataset with an ap- propriate loss function and optimizer.
- The Convolutional Neural Network (CNN) achieved an accuracy of 60% in predicting whether an object is classified as *good* or *bad*.

Challenges:

- A large dataset was required for accurate predic- tions, which resulted in overfitting when trained on a smaller dataset.
- Time-intensive model tuning was needed to im- prove the classification accuracy for complex multi-class problems.
- The absence of spatial information (object local- ization) limited its application for tasks like defect detection.

2.2 Transition to YOLO (You Only Look Once)

Motivation: The limitations of CNNs, particularly in real-time object detection and classification, prompted a transition to YOLO (You Only Look Once). YOLO enables both object detection and classification in real- time.

Implementation:

- The dataset was expanded by manually annotating 10x more images to improve generalization.
- YOLOv8 was chosen for its real-time detection ca- pabilities, using the Ultralytics library to stream- line implementation.
- A custom model was trained on annotated bound- ing boxes to detect and classify multiple classes (e.g., fresh vs. rotten produce) within a single image.

Advantages:

- YOLOv8 enabled faster processing, achieving real-time detection up to 45 FPS.
- The model provided spatial information through bounding boxes and class labels, useful for localization.

Challenges:

- Manual annotation of bounding boxes was labor-intensive and time-consuming.
- YOLOv8, while suitable for object detection, struggled with high text recognition accuracy due to poor OCR results in certain conditions.

2.3 Integration of EfficientNet-B0

Motivation: EfficientNet-B0, a pre-trained lightweight architecture, was adopted to handle dual-task classifica- tion effectively. This architecture allows for both cate- gory and quality classification, ensuring better general- ization and efficiency.

Implementation:

- EfficientNet-B0 was adapted for dual-task classification, distinguishing between fruit/vegetable types and labeling quality as "Good" or "Bad."
- Transfer learning was applied using pre-trained weights from ImageNet to enhance feature extraction.
- The model's head was customized with two fully connected layers for dual-task predictions.
- A preprocessing pipeline was created to resize and normalize the input images as required by EfficientNet.

Advantages:

- EfficientNet-B0 provided improved classification accuracy over CNN, especially with smaller datasets due to transfer learning.
- The model was highly efficient, enabling the classification of both category and quality in one step.
- The model achieved a maximum train set accuracy of 98.2%, with a validation accuracy of 97.5%, and a test accuracy of 98%.

Challenges:

- Significant data preprocessing was required to ensure the consistency of input images with EfficientNet's input requirements.
- Task-specific tuning was necessary to achieve optimal performance for dual-task classification.

2.4 Summary of Results

- The CNN model provided foundational classification capabilities but lacked localization abilities, limiting its use for defect detection.
- YOLOv8 enabled real-time object detection with bounding boxes but faced challenges due to the large-scale annotation effort required for 90,000 images, which was time-consuming and labor-intensive.
- EfficientNet-B0 offered an efficient and accurate solution for dual-task classification, particularly for small datasets.

3 Results and Insights

3.1 Performance Summary

CNN:

- The CNN model was effective for basic image classification tasks, allowing for the classification of fresh and rotten fruits and vegetables.
- However, the lack of spatial and localization capabilities limited its ability to detect defects or perform tasks requiring object localization, such as identifying specific areas of the produce.

YOLO:

- YOLOv8 demonstrated real-time object detection with bounding box localization, marking a significant advancement over the CNN model.
- Despite its success in detecting objects and providing bounding boxes with class labels, the model faced challenges due to the large-scale an- notation effort required for 90,000 images, mak- ing the annotation process time-consuming and labor-intensive.

EfficientNet-B0:

- EfficientNet-B0 was the best-suited model for quality classification tasks, as it effectively han- dled dual-task classification—distinguishing be- tween fruit/vegetable types and labeling quality as "Good" or "Bad."
- This model's efficiency, particularly with trans- fer learning, allowed for enhanced accuracy, even when trained on smaller datasets, making it a prac- tical solution for real-time quality control tasks.

3.2 Challenges Overcome

- **Increased Dataset Size:** The dataset was ex- panded by manually annotating a larger set of images to improve model generalization and re- duce the risk of overfitting, especially for YOLO and EfficientNet-B0.
- **Modular Approach:** The project adopted a mod- ular approach by integrating specialized models for different tasks (CNN for classification, YOLO for detection, and EfficientNet-B0 for dual-task classification). This allowed for targeted improvements in each area.
- **Refined Preprocessing and Training Pipelines:** Data preprocessing and model training pipelines were refined, ensuring better adaptability of the models to the dataset. For instance, image aug- mentation and normalization techniques were en- hanced for better model performance.

3.3 Conclusion

This project successfully integrated multiple deep learn- ing models—CNN, YOLO, and EfficientNet-B0—to ad- dress distinct challenges in e-commerce quality control. Each model contributed uniquely:

- **CNN:** Served as the foundation for basic classification tasks, providing initial insights into image-based classification.
- **YOLO:** Addressed the need for real-time object detection, significantly improving the ability to detect and localize different types of produce.
- **EfficientNet-B0:** Provided an optimized solution for dual-task classification, enhancing accuracy while maintaining efficiency, making it ideal for quality classification tasks.