

An AR-Assisted Food Ingredient Recognition System for Smart Refrigerators Using Deep Learning

Liwei Chen

James Cook University in Singapore, Information Technology, 387380, Singapore

ABSTRACT

Food waste at the household level has become a significant global issue, largely due to inefficient food management and lack of awareness of stored ingredients. This paper proposes an Augmented Reality (AR)-based food recognition system designed for household refrigerators. The system integrates lightweight deep learning-based food recognition with real-time AR visualization to improve food visibility and usage efficiency. Using a lightweight CNN-based object detection model (YOLOv8) with real-time AR visualization, the proposed system identifies common refrigerator food items under challenging conditions such as low lighting, occlusion, and clutter. Experimental results demonstrate that the system can accurately recognize and localize multiple food items in real time, providing a practical solution for reducing household food waste and enhancing smart kitchen applications.

KEYWORDS

Food ingredient detection; YOLOv8; Augmented reality; Smart refrigerator; Computer vision

1 Introduction

With the increasing pace of modern life, households are facing growing challenges in managing daily food storage efficiently. Due to limited visibility into refrigerator contents, users often purchase duplicate items, forget existing ingredients, or allow food to expire unnoticed. According to the Food Waste Index Report published by the Food and Agriculture Organization (FAO), household food waste accounts for more than 60% of global food waste, highlighting the urgent need for intelligent food management solutions.

Recent advances in computer vision and deep learning have enabled automatic food recognition using image-based methods. However, most existing systems focus on restaurant meals or controlled environments, while household refrigerators present additional challenges such as low illumination, food occlusion, and diverse packaging. Moreover, traditional mobile applications provide limited interaction and lack intuitive visualization.

To address these issues, this research proposes an AR-based refrigerator food recognition system that combines deep learning-based object detection with real-time AR visualization. By overlaying recognized food labels directly onto the camera view, the system offers an intuitive and interactive approach to household food management.

2 Literature Review

2.1 Food Waste and the Role of Smart Refrigerator Systems

Food waste is a major global problem, with households contributing a significant proportion of total food loss. Fruits and vegetables stored in household refrigerators are particularly susceptible to waste due to spoilage and lack of visibility. Previous studies indicate that consumers often forget stored items or fail to notice food deterioration, especially when items are placed in deep refrigerator bins or behind other products (Honee et al., 2022).

To address this issue, smart refrigerator systems have been proposed to improve food inventory management. Commercial smart refrigerators typically integrate internal cameras and proprietary software to monitor stored items. However, such systems are often expensive and inaccessible to ordinary households, limiting their widespread adoption (Bokhari, 2023). As a result, academic research has focused on low-cost alternatives using computer vision and standard RGB cameras to automatically identify refrigerator contents.

Despite these efforts, existing systems struggle with practical constraints such as occlusion, poor lighting, and cluttered scenes, which significantly reduce recognition accuracy in real-world household environments.

2.2 Deep Learning for Food Recognition

Early food recognition systems relied on handcrafted features and traditional classifiers, which were limited in handling variations in food appearance. The introduction of deep learning significantly improved performance, particularly through convolutional neural networks (CNNs). A landmark contribution is the Food-101 dataset, which enabled large-scale training and evaluation of food classification models and demonstrated the effectiveness of CNN-based approaches (Bossard et al., 2014).

Although CNN-based models achieve high accuracy on benchmark datasets, their performance degrades when applied to uncontrolled environments such as household refrigerators. Food items in refrigerators vary widely in shape,

packaging, and illumination, making recognition more challenging than restaurant or plate-based food images (Bokhari, 2023). These limitations suggest that benchmark performance does not directly translate to real-world household scenarios.

2.3 Impact of Occlusion on Object Detection Performance

Occlusion is a well-known challenge in object detection and has a particularly severe impact in household storage environments. In refrigerators, items are frequently stacked or partially hidden, resulting in incomplete visual information. Research by Koporec et al. (2019) systematically demonstrates that deep learning models experience a significant drop in detection accuracy as the level of occlusion increases, even when trained on large datasets.

These findings are highly relevant to refrigerator-based food recognition, where occlusion is unavoidable due to limited storage space. Traditional object detection benchmarks do not adequately capture this challenge, as objects in benchmark datasets are often well separated and fully visible. Consequently, models trained and evaluated solely on such datasets may not generalize well to refrigerator environments.

2.4 Refrigerator-Based Food Recognition Datasets

Collecting and annotating real-world refrigerator images is both time-consuming and labor-intensive. To overcome this limitation, researchers have explored the use of synthetic and controlled datasets to study food recognition in refrigerator environments. Synthetic datasets allow precise control over variables such as occlusion level, object placement, and occupancy rate, enabling systematic evaluation of model robustness.

Recent studies demonstrate that models trained with synthetic refrigerator data can achieve improved robustness to occlusion compared to models trained solely on real images (Morris, 2020). However, a domain gap between synthetic and real images remains a challenge, requiring careful dataset design and potential domain adaptation techniques.

These findings suggest that refrigerator-specific datasets—whether real or synthetic—are essential for evaluating food recognition systems intended for household use.

2.5 Vision-Based Inventory Tracking Approaches

Most existing refrigerator recognition systems focus on static image analysis, attempting to detect all visible items within a single frame. However, this approach struggles when items are fully occluded. An alternative paradigm is to track inventory changes through interaction events rather than full-scene understanding.

Vision-based approaches that monitor object entry and removal events leverage the fact that all items must pass through the refrigerator opening. By detecting hands and manipulated objects, it becomes possible to infer inventory changes even if the final storage location is visually occluded (Bokhari, 2023). This interaction-based strategy reduces reliance on complete visual access and offers a promising direction for practical refrigerator inventory systems.

2.6 Summary and Research Gap

The reviewed literature indicates that while deep learning has significantly advanced food recognition, household refrigerator environments remain challenging due to occlusion, lighting variability, and clutter. Existing smart refrigerator systems are either costly or limited in robustness. Moreover, benchmark food datasets and conventional object detection evaluations fail to capture the complexity of real refrigerator scenes.

Therefore, there is a clear research gap in developing practical, low-cost food recognition systems that are robust to occlusion and suitable for household deployment. This research aims to address this gap by exploring computer vision-based food recognition techniques tailored to refrigerator environments, with a focus on improving robustness under realistic storage conditions.

3 Methodology

3.1 System Overview

This study proposes an AR-assisted food ingredient recognition system for smart refrigerator environments. The overall system is designed as a modular pipeline consisting of three main components: (1) image acquisition; (2) food ingredient detection; (3) augmented reality (AR) visualization.

Images of refrigerator contents are captured using a camera-equipped device. These images are processed by a deep learning-based object detection model to identify and localize food ingredients. The detection results are then forwarded to an AR visualization layer, where virtual labels and bounding boxes are overlaid onto the physical scene, enabling intuitive inspection of refrigerator contents.

This modular design decouples perception and visualization, allowing the detection model and AR interface to be developed, evaluated, and upgraded independently.

3.2 Dataset Preparation

The dataset used in this study is derived from the Food Ingredients Detection dataset (Roboflow v6), which contains

annotated images of various food items captured under diverse conditions.

To ensure suitability for refrigerator scenarios and reduce category ambiguity, the original 40 food classes were reorganized into 15 representative ingredient categories, including banana, carrot, tomato, egg, milk, cheese, meat, chicken, fish, lettuce, broccoli, onion, cucumber, and pepper. Several visually similar classes (e.g., beef, pork, sausage) were merged into higher-level categories such as meat, while seafood-related items were grouped into the fish category.

All images from the original training, validation, and test splits were first pooled together. The dataset was then restructured and re-split into new training, validation, and test sets using a stratified sampling strategy to ensure that each class was represented across splits. This step was critical to avoid class imbalance and distribution mismatch between training and validation data.

3.3 Food Ingredient Detection Model

Food ingredient detection is performed using the YOLOv8n object detection model, which offers a favorable balance between accuracy and computational efficiency. YOLOv8 adopts a single-stage detection architecture, enabling real-time inference by jointly predicting bounding boxes and class probabilities. The YOLOv8 model was implemented using the Ultralytics framework (Ultralytics, 2023).

The model was initialized with pre-trained weights and fine-tuned on the restructured dataset. Input images were resized to 512×512 pixels, and data augmentation techniques such as random flipping and color jittering were applied implicitly through the Ultralytics training framework to improve generalization.

The detection model outputs bounding box coordinates, confidence scores, and predicted food categories for each detected object, serving as the perception backbone of the proposed system.

3.4 Training Configuration

Model training was conducted using the Ultralytics YOLOv8 framework on a system equipped with an NVIDIA GeForce RTX 4050 Laptop GPU (6 GB VRAM). The network was trained for 30 epochs with a batch size of 8 using mixed precision training (AMP) to enhance computational efficiency.

An adaptive optimizer (Adam) was selected automatically by the framework, and early stopping was disabled to allow full convergence. The best-performing model checkpoint was selected based on validation performance and saved from the training run identified as train2.

3.5 Evaluation Metrics

The performance of the food ingredient detection model was evaluated using standard object detection metrics:

Precision (P), measuring the proportion of correct positive predictions;

Recall (R), measuring the proportion of ground-truth objects correctly detected;

mean Average Precision at IoU 0.5 (mAP@0.5); and

mean Average Precision averaged from IoU 0.5 to 0.95 (mAP@0.5:0.95).

These metrics provide a comprehensive assessment of both classification accuracy and bounding box localization quality.

3.6 Augmented Reality Integration

To enhance user interaction and practical applicability, the proposed system integrates Augmented Reality (AR) as an application-layer visualization module.

In the designed architecture, the detection results produced by the YOLOv8 model—including bounding box coordinates and class labels—are transmitted to an AR rendering module. Using marker-less AR technologies such as ARCore or ARKit, virtual overlays are aligned with real-world coordinates and displayed directly on the refrigerator scene captured by a mobile device camera.

Detected food items are visualized as labeled bounding boxes or icons, enabling users to quickly identify available ingredients without physically searching through the refrigerator. This AR-based interface provides an intuitive and immersive means of inventory inspection and can be further extended to support advanced functions such as expiration date reminders, recipe recommendations, and nutritional analysis.

It is important to note that the AR component in this study is implemented at the system design and prototype level, while quantitative evaluation focuses on the detection model. Full AR deployment and user experience evaluation are left as directions for future work.

3.7 Methodology Summary

In summary, the proposed methodology combines a deep learning-based food ingredient detection model with an AR-assisted visualization interface. By restructuring the dataset, fine-tuning a lightweight YOLOv8 model, and integrating AR at the system level, the approach aims to deliver accurate, efficient, and user-friendly food recognition in refrigerator environments.

4 System Architecture

4.1 Overall Architecture

The proposed system adopts a modular and layered architecture to support food ingredient recognition and augmented reality visualization in smart refrigerator environments. As illustrated conceptually, the system is composed of four main layers: (1) image acquisition; (2) detection and inference; (3) data processing; (4) augmented reality visualization.

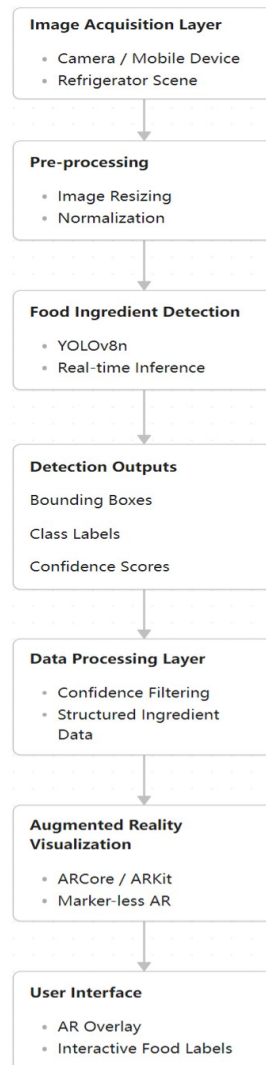


Figure 1 System architecture of the proposed AR-assisted food ingredient recognition system

This layered design enables clear separation of responsibilities, improves system scalability, and allows individual components to be updated or replaced independently.

4.2 Image Acquisition Layer

The image acquisition layer is responsible for capturing visual data of refrigerator contents. Images are obtained using a camera-equipped device, such as a smartphone or embedded refrigerator camera, under natural lighting conditions.

Captured images are preprocessed to ensure compatibility with the detection model, including resizing to a fixed input resolution and normalization. This layer serves as the primary interface between the physical refrigerator environment and the digital processing pipeline.

4.3 Detection and Inference Layer

The detection and inference layer forms the core of the system and is implemented using the YOLOv8n object detection model. Preprocessed images from the acquisition layer are fed into the trained detection network, which performs real-time inference to identify food ingredients.

For each detected object, the model outputs:

- bounding box coordinates,
- class labels, and

confidence scores.

This layer is optimized for real-time performance and is capable of running efficiently on consumer-grade hardware, enabling practical deployment in household environments.

4.4 Data Processing Layer

The data processing layer bridges the detection model and the AR visualization module. Detection outputs are filtered based on confidence thresholds and formatted into structured representations that include object class, spatial location, and detection confidence.

This layer also supports potential extensions such as:
ingredient inventory tracking,
expiration date management, and
integration with recipe recommendation systems.

By abstracting detection results into structured data, the system facilitates seamless communication between perception and application layers.

4.5 Augmented Reality Visualization Layer

The augmented reality visualization layer provides an intuitive user interface for presenting detection results. Using marker-less AR technologies such as ARCore or ARKit, virtual overlays are rendered directly onto the physical refrigerator scene captured by the camera.

Detected food ingredients are displayed as labeled bounding boxes or icons aligned with their real-world positions. This allows users to visually inspect refrigerator contents without manually searching through shelves or containers.

The AR module operates independently from the detection layer, ensuring that improvements in detection accuracy or visualization techniques can be incorporated without redesigning the entire system.

4.6 System Workflow

The end-to-end workflow of the proposed system can be summarized as follows:

- The camera captures an image of the refrigerator interior.
- The image is preprocessed and forwarded to the detection model.
- YOLOv8 performs real-time food ingredient detection.
- Detection results are filtered and structured by the data processing layer.
- The AR module overlays virtual labels and bounding boxes onto the physical scene.
- This workflow enables efficient and user-friendly inspection of refrigerator contents.

4.7 Architectural Advantages

The proposed system architecture offers several advantages:

Modularity: Each layer operates independently, improving maintainability and extensibility.

Real-time Capability: The lightweight detection model supports real-time inference.

User-Centric Design: AR visualization enhances usability and interaction.

Scalability: Additional services, such as nutrition analysis or shopping list generation, can be integrated seamlessly.

4.8 Architecture Summary

In summary, the proposed system architecture integrates deep learning –based perception with augmented reality visualization in a structured and scalable manner. By combining efficient object detection with an intuitive AR interface, the system provides a practical foundation for smart refrigerator applications and intelligent kitchen environments.

5 Experimental Evaluation

5.1 Experimental Setup

The experimental evaluation was conducted to assess the effectiveness of the proposed food ingredient detection model in refrigerator environments. All experiments were performed using the YOLOv8n object detection model implemented with the Ultralytics framework and trained on the restructured dataset described in Section 3.

Training and evaluation were carried out on a laptop equipped with an NVIDIA GeForce RTX 4050 Laptop GPU (6 GB VRAM). The model was trained for 30 epochs with an input resolution of 512×512 pixels and a batch size of 8, using mixed precision training (AMP) to improve computational efficiency. The best-performing model checkpoint was selected from the training run train2 based on validation performance.

5.2 Quantitative Evaluation

The quantitative performance of the proposed model was evaluated on the validation set, which contains 682 images and 1,394 annotated food ingredient instances across 15 categories.

Overall, the model achieved strong detection performance, with a Precision of 0.844, Recall of 0.819, mAP@0.5 of 0.871, and mAP@0.5:0.95 of 0.630. These results indicate that the model is capable of accurately identifying and localizing food ingredients in refrigerator scenes.

Several categories achieved particularly high accuracy. Ingredients with distinctive visual features, such as tomato, broccoli, lettuce, and milk, achieved mAP@0.5 values above 0.95, demonstrating robust detection performance. Moderate performance was observed for more visually complex categories, including meat, chicken, and pepper, which exhibit higher intra-class variability.

5.3 Per-Class Performance Analysis

Table 1 The summary of the per-class detection performance on the validation set.

Class	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Banana	0.898	0.859	0.934	0.667
Carrot	0.929	0.877	0.918	0.678
Tomato	0.972	0.988	0.992	0.881
Egg	0.872	0.961	0.901	0.458
Milk	0.809	1.000	0.961	0.796
Cheese	0.750	0.851	0.890	0.676
Meat	0.742	0.671	0.710	0.421
Chicken	0.704	0.617	0.667	0.491
Fish	0.918	0.799	0.905	0.647
Lettuce	0.956	0.851	0.951	0.824
Broccoli	0.975	0.973	0.990	0.772
Onion	0.719	0.639	0.789	0.453
Cucumber	0.844	0.780	0.884	0.599
Pepper	0.731	0.598	0.694	0.452
Overall	0.844	0.819	0.871	0.630

The tomato category achieved the highest detection accuracy, with an mAP@0.5 of 0.992, followed by broccoli and lettuce, both exceeding 0.95.

These categories typically exhibit consistent shape, color, and texture, making them easier to detect.

In contrast, categories such as meat and chicken achieved lower mAP@0.5 values of 0.710 and 0.667, respectively. This performance gap can be attributed to diverse visual appearances, packaging variations, and frequent occlusion within refrigerator environments.

Despite having a relatively small number of validation samples (9 images, 32 instances), the onion category achieved a reasonable mAP@0.5 of 0.789, highlighting the effectiveness of the revised dataset splitting strategy.

5.4 Qualitative Evaluation

Qualitative evaluation further demonstrates the effectiveness of the proposed approach. The model successfully detects multiple food ingredients within a single refrigerator image, even under challenging conditions such as partial occlusion, overlapping objects, and varying illumination.

Most detection errors occur in visually ambiguous scenarios, including overlapping meat and chicken products or objects partially hidden behind transparent packaging. Representative detection examples, including both successful predictions and failure cases, are presented in Figure 2.



Figure 2 Example detection results on refrigerator images, including successful predictions and representative failure cases under occlusion and object overlap

6 Discussion

6.1 Impact of Dataset Restructuring

One of the key findings of this study is the significant impact of dataset restructuring on model performance. Initial experiments revealed severe class distribution mismatches between training and validation sets, which resulted in poor detection accuracy for several categories. By merging semantically similar classes and applying stratified re-splitting, the revised dataset ensured consistent class representation across splits, leading to substantial performance improvements.

This observation highlights the importance of dataset quality and distribution balance in object detection tasks, particularly for real-world applications such as refrigerator environments.

6.2 Analysis of Detection Performance

The experimental results demonstrate that the YOLOv8n model is highly effective at detecting food ingredients with distinctive visual characteristics. High performance on categories such as tomato, broccoli, and lettuce suggests that the model can reliably learn discriminative features when intra-class variation is limited.

Conversely, reduced performance on meat and chicken reflects the inherent difficulty of these categories. Visual similarity, diverse packaging, and occlusion contribute to confusion between classes, even with sufficient training data. These findings are consistent with prior work on food recognition, where visually ambiguous categories remain challenging.

6.3 Role of Augmented Reality in the System

Although quantitative evaluation in this study focuses on detection performance, the integration of augmented reality plays an important role at the system level. AR provides an intuitive interface for presenting detection results directly within the physical refrigerator context, enhancing usability and user engagement.

By decoupling detection and visualization, the proposed system allows AR functionality to evolve independently from the detection model. This design choice supports future extensions such as interactive inventory management, expiration alerts, and personalized recipe recommendations.

6.4 Limitations

Several limitations of this study should be acknowledged. First, the dataset is derived from publicly available images and may not fully capture all real-world refrigerator conditions. Second, AR functionality is presented at a prototype and conceptual level, without formal user studies or latency evaluation. Finally, the use of a lightweight YOLOv8n model, while efficient, may limit performance on highly complex scenes.

6.5 Summary of Discussion

Overall, the discussion highlights that the proposed system achieves strong detection performance while maintaining practical efficiency. The results underscore the importance of dataset design and system-level integration, and they

provide valuable insights for future research on smart refrigerator applications.

7 Conclusion and Future Work

7.1 Conclusion

This study presents an AR-assisted food ingredient recognition system for smart refrigerator environments, combining deep learning–based object detection with an intuitive augmented reality visualization interface.

A lightweight YOLOv8n model was fine-tuned on a restructured food ingredient dataset derived from the Roboflow Food Ingredients Detection dataset. Through careful dataset reorganization and stratified re-splitting, class imbalance and distribution mismatch issues were effectively mitigated. Experimental results demonstrate that the proposed approach achieves strong overall performance, with a validation mAP@0.5 of 0.871 and mAP@0.5:0.95 of 0.630, while maintaining real-time inference capability on consumer-grade hardware.

The model performs particularly well on visually distinctive food items such as tomato, broccoli, lettuce, milk, and fish, while more challenging categories such as meat, chicken, and pepper exhibit moderate performance due to higher intra-class variability and visual ambiguity. Despite limited validation samples, the onion category achieved reasonable detection accuracy, highlighting the effectiveness of the revised dataset splitting strategy.

By integrating an AR-based visualization layer at the system level, the proposed framework enables intuitive inspection of refrigerator contents and provides a foundation for intelligent food inventory management. Overall, the results indicate that the proposed system is both effective and practical for smart kitchen applications.

7.2 Future Work

Several directions for future research can further enhance the proposed system.

First, detection performance may be improved by expanding the dataset with additional refrigerator-specific images and increasing the number of samples for underrepresented classes. Fine-grained category definitions, such as distinguishing between different types of meat or vegetables, could also reduce class ambiguity.

Second, more powerful detection models, such as YOLOv8s or YOLOv8m, may be explored to improve accuracy on visually complex ingredients, particularly in cluttered or occluded refrigerator scenes. Additional data augmentation and domain adaptation techniques could further enhance robustness under varying lighting conditions.

Third, the AR component can be extended from a prototype-level design to a fully deployed application. Future work may involve implementing a complete AR interface using ARCore or ARKit, followed by user studies to evaluate usability, latency, and overall user experience.

Finally, the system may be integrated with higher-level intelligent services, such as expiration date tracking, recipe recommendation, and nutritional analysis, transforming the proposed framework into a comprehensive smart refrigerator assistant.

References

- [1] Bossard, L., Guillaumin, M., & Van Gool, L. (2014). Food-101: Mining discriminative components with random forests. In *European Conference on Computer Vision (ECCV 2014)* (pp. 446–461). Springer.
- [2] Bokhari, M. (2023). *Computer vision for determination of fridge contents* (Master's thesis). University of Passau, Germany.
- [3] Honee, D., Hurst, W., & Luttikhold, A. J. (2022). Harnessing augmented reality for increasing awareness of food waste. *Augmented Human Research*, 7(1), Article 5.
- [4] Koporec, J., Pers, J., & Kristan, M. (2019). Deep learning performance in the presence of significant occlusions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW 2019)* (pp. 1–8). IEEE.
- [5] Morris, S. V. (2020). *Inventory management of refrigerator produce bins using computer vision* (Master's thesis). University of Arkansas, United States.
- [6] Ultralytics. (2023). YOLOv8: Next-generation object detection and segmentation models.