Multi-Ingredient Food Image Classification and Context-Aware Recipe Recommendation System

Dhanush Biligiri N H, Fnu Sowrabh, Ram Bagaria, Akhilesh Dandavati

***Abstract*—Food recognition and recipe recommendation systems have become essential for smart cooking and personalized nutrition. This project explores the integration of deep learning for ingredient detection and a machine learning-based recommendation system to enhance accuracy and personalization. By leveraging computer vision and hybrid recommendation techniques, the system aims to provide context-aware recipe suggestions while considering ingredient availability and user preferences. This approach contributes to improved food tracking, meal planning, and automated culinary assistance.**

***Index Terms*—Deep Learning, Food Recognition, Recipe Recommendation, Computer Vision, Context-Aware System.**

I. Introduction

The increasing availability of food images and recipes on digital platforms has created a demand for intelligent food recognition and recommendation systems. Traditional methods rely on ingredient-based matching or predefined recipe databases but face several challenges, including:

- **Ingredient Recognition Complexity:** Variations in appearance, lighting, and occlusion make multi-ingredient detection difficult.
- **Lack of Context Awareness:** Many systems fail to consider user preferences, dietary restrictions, and available ingredients.
- **Scalability Issues:** Large-scale recipe databases require efficient filtering techniques for personalized recommendations.

To address these limitations, this project leverages deep learning for food recognition and machine learning-based recommendations to enhance accuracy and personalization. These systems can benefit applications in smart kitchens, automated diet tracking, and health monitoring by helping users track their food intake, monitor macronutrient consumption, and suggest healthier alternatives. Additionally, ingredient detection plays a crucial role in dietary restriction compliance, allowing individuals with allergies or specific nutritional needs to verify food ingredients before consumption. In restaurants and food industries, automated ingredient recognition can assist in quality control and menu personalization.

II. Related Work

Previous studies have explored different aspects of food recognition and recommendation systems:

- **Ingredient Detection:** Deep learning models such as CNNs and Mask R-CNN have been used to classify food ingredients, improving recognition accuracy [1].
- **Recipe Generation:** Research on inverse cooking has explored generating structured recipes from images by predicting ingredients and cooking instructions [2].
- **Context-Aware Recommendations:** Modern approaches incorporate user preferences and dietary constraints to improve personalization [3].

CNNs and Mask R-CNN struggle with real-time inference and overlapping objects. YOLOv8 enables fast, accurate detection and reduces reliance on manually labeled datasets through advanced augmentation. This research builds on these approaches by integrating multi-ingredient detection with a context-aware recommendation system.

III. Methodology

The proposed system consists of four main components: Ingredient Detection, Feature Extraction, Hybrid Recommendation System, and Model Deployment.

A. Data Collection and Preprocessing

A well-labeled and diverse dataset is essential for robust ingredient detection. The following datasets are considered:

- **Public Datasets:** Large-scale datasets provide high-quality images and labels for training:
  - **Food101** – Contains 101 food categories [4].
  - **Recipe1M+** – Pairs food images with recipes for ingredient extraction [5].
  - **VIREO-172** – Includes 172 labeled ingredients [6].
  - **UECFOOD256** – A fine-grained dataset with ingredient labels [7].
- **Custom Dataset Creation:** If additional data is needed:
  - Web scraping food images and ingredient annotations.
  - Manual annotation using tools like LabelImg.
  - Data augmentation techniques such as rotation and brightness adjustments.

B. Ingredient Detection using YOLO

A YOLO-based deep learning model is used for multi-ingredient detection in food images. The training process consists of:

- Converting images into a YOLO-compatible format (bounding boxes and labels).
- Fine-tuning a pre-trained YOLO model to improve recognition accuracy.
- Evaluating detection performance using **mean Average Precision (mAP)**.

YOLOv8 and YOLOv5 were chosen for their improved accuracy, anchor-free detection, and instance segmentation. They balance speed and precision, effectively distinguishing visually similar ingredients. Data augmentation enhances generalization across diverse lighting conditions and food types. Additionally, YOLOv5 offers better pre-trained models and optimized algorithms, making it a strong candidate for ingredient detection.

C. Hybrid Recipe Recommendation System

The recommendation engine combines:

- **Content-Based Filtering:** Suggests recipes based on detected ingredient similarity.
- **Collaborative Filtering:** Utilizes user interaction data to refine suggestions.
- **Hybrid Model:** Integrates both methods to enhance personalization.

Content-Based Filtering uses TF-IDF and ingredient embeddings to map ingredient similarity. Collaborative Filtering applies SVD and Neural Collaborative Filtering (NCF), refining personalization. A hybrid weighting system prevents the cold-start problem

D. Model Deployment

The deployment strategy includes:

- Local inference for initial testing.
- Cloud-based API hosting for scalability.
- Web/Mobile applications for user accessibility.

IV. Expected Results and Outcomes

The system is expected to achieve:

- **Accurate Multi-Ingredient Detection:** High **mAP** score in recognizing multiple ingredients from food images.
- **Efficient Recipe Recommendations:** Improved personalization through hybrid filtering.
- **Scalability:** Seamless integration with structured recipe databases and real-time inference.
- **Future Enhancements:** Potential integration with NLP models for dynamic recipe generation.

Ingredient detection is evaluated using mAP and false positive rates. We aim to assess Recommendation quality using Precision@K, Recall@K and user retention tracking.

V. Discussion and Conclusion

This project integrates deep learning for ingredient detection and a hybrid recommendation system to enhance food recognition and recipe suggestions. Using YOLO-based object detection, the system ensures accurate ingredient identification, while a hybrid recommendation approach improves personalization.

Future improvements will focus on increasing dataset diversity, optimizing model performance, and refining recommendation accuracy. Deployment feasibility in real-world applications will also be explored.

VI. Future Work

Planned enhancements include:

- **Expanding Dataset Diversity:** Incorporating additional real-world images to improve generalization.
- **Enhancing Model Performance:** Exploring advanced deep learning architectures.
- **Real-Time Implementation:** Deploying as a cloud-based API or mobile application.
- **User Feedback Mechanism:** Implementing a learning-based system for better recommendations.

Multi-modal learning integrates images, text, and user preferences. Graph Neural Networks (GNNs) enhance ingredient relations. Explainable AI and reinforcement learning improve personalized recommendations and user engagement. These advancements will improve the system's adaptability and usability.

References

[1] M. K. Morol, M. S. J. Rokon, I. B. Hasan, A. M. Saif, R. H. Khan, and S. S. Das, "Food recipe recommendation based on ingredients detection using deep learning," in *Proceedings of the 2nd International Conference on Computing Advancements*, March 2022, pp. 191–198.

[2] A. Salvador, M. Drozdzal, X. Giro-i Nieto, F. Marques, and A. Moreno, "Inverse cooking: Recipe generation from food images," *arXiv preprint*, 2018. [Online]. Available: https://arxiv.org/abs/1812.06164

[3] P. Chhikara, D. Chaurasia, Y. Jiang, O. Masur, and F. Ilievski, "Fire: Food image to recipe generation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 8184–8194.

[4] L. Bossard, M. Guillaumin, and L. Van Gool, "Food-101 – mining discriminative components with random forests," in *European Conference on Computer Vision (ECCV)*, 2014, pp. 446–461.

[5] J. Marin, A. Salvado, M. Drozdzal *et al.*, "Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 12, pp. 3245–3259, 2019.

[6] J. Chen, C. W. Ngo, and X. Zhu, "Deep understanding of cooking procedure for cross-modal recipe retrieval," in *ACM Multimedia*, 2016, pp. 192–196.

[7] Y. Kawano and K. Yanai, "Food image recognition with deep convolutional features," in *ACM International Conference on Multimedia*, 2014, pp. 589–592.