# Recipix:
# A Multimodal Food Recognition and Recipe Recommendation System Using AI

Ivan Zepeda (Main Developer — research)
ivanzepeda@duck.com
c0883949@mylambton.ca

Camilo Florez (Developer — Research) Erwin Mongui (Architect)
Preet Dhorojiya (Developer)

October 2024

**Abstract**

This paper presents Recipix, a multimodal system designed to optimize food management by leveraging computer vision and natural language processing (NLP). Recipix identifies ingredients in fridge images, applies classification techniques, and uses a recommendation engine to suggest personalized recipes based on user dietary preferences, allergies, and ingredient availability. The system integrates deep learning models, including ResNet-50 for image classification and SAM (Segment Anything Model) for segmentation, delivering real-time food identification and meal suggestions.

## 1. Introduction

The challenges of reducing food waste and improving meal planning have long been significant problems in both household and industrial settings. Globally, food waste accounts for a large percentage of environmental and economic losses. One contributing factor is that consumers often fail to utilize ingredients in their refrigerators effectively. To address this, Recipix offers a novel AI-powered solution that provides food identification from fridge images and suggests recipes tailored to the available ingredients and user preferences.

This project integrates cutting-edge AI models and techniques to enhance the user's experience in meal planning and sustainability. Recipix also serves as an example of how AI and machine learning technologies can be applied to everyday tasks.

## 2. Literature Review

Several AI-powered kitchen assistants have emerged in recent years. Systems like Samsung's Family Hub have introduced similar functionalities, integrating AI meal planning with food inventory management. While these solutions provide a foundation, their reliance on proprietary platforms and incomplete integration of computer vision limits their accessibility and customization.

Other studies have explored convolutional neural networks (CNNs) for food classification and NLP techniques for recipe generation. However, Recipix distinguishes itself by combining the strengths of computer vision and NLP in an end-to-end system, offering real-time interaction via Telegram and a highly personalized recommendation engine.

## 3. Methodology

### 3.1. Data Collection

To train Recipix's food classification model, we utilized a dataset comprising over 41,000 images representing 88 distinct food categories. Images were sourced from public datasets like FVIRD (Fruit and Vegetable Image Recognition Dataset) and Food101, as well as web scraping of food-related websites for additional images and recipes.

### 3.2. Image Preprocessing

Images were preprocessed using techniques like resizing, normalization, and augmentation (e.g., random rotations, flips). These preprocessing steps ensured the models could generalize well across various conditions such as lighting, angles, and partial occlusion of food items.

### 3.3. Model Selection

ResNet-50 was used for food classification due to its proven performance in object detection tasks. The model was fine-tuned on the augmented dataset to improve its accuracy. Segment Anything Model (SAM) was integrated for image segmentation to detect and differentiate between food items in cluttered fridge images. SAM proved particularly effective for recognizing overlapping or partially hidden objects.

### 3.4. Recipe Recommendation System

The NLP component of Recipix was built using models like SpaCy and Word2Vec to match the detected ingredients with a curated recipe database. Recipes were filtered based on user preferences such as allergies, dietary restrictions, and liked ingredients. The system then generated recipe suggestions in PDF format using DALL·E API to provide a visual component.

# 4. Technical Description

## 4.1. System Architecture

The Recipix system follows a client-server architecture where:

- **Client Side:** Users interact with the system via a Telegram bot or a web interface. They upload an image of their fridge or pantry, which is processed by the backend to identify ingredients.
- **Backend:** The backend is built using FastAPI to handle image processing requests, leveraging GPU acceleration (CUDA) for real-time performance. The models are deployed as microservices that handle tasks such as image segmentation, classification, and recipe recommendation.
- **Data Storage:** Recipes, user preferences, and classified ingredients are stored in a PostgreSQL database, with Redis used for caching to ensure quick response times.

## 4.2. Computer Vision Pipeline

**Image Segmentation:** The SAM model segments the fridge image into individual food items, which are then passed through the classification model.

**Food Classification:** The ResNet-50 model classifies each segmented food item, with accuracy reaching 81.05% across unseen test data.

**Mask Generator:** The segmented regions are passed through the SAM Mask Generator, which applies specific configuration for generating detailed masks (high, medium, or low detail, depending on the user's request).

## 4.3. NLP-Based Recommendation Engine

TF-IDF and Word2Vec models are used to analyze available ingredients and match them to recipes in the database. The system factors in user preferences like diets (e.g., vegan, gluten-free) and allergies, adjusting recipe suggestions dynamically. Recipe suggestions are compiled and delivered in an interactive format via Telegram or a web interface, complete with visual representations of the final dish.

## 5. Results and Performance

### 5.1. Image Classification Accuracy

The system achieved an accuracy of 81.05% in classifying food items using the ResNet-50 model. Further optimization, such as increasing dataset diversity and fine-tuning the hyperparameters, improved the classification results.

### 5.2. Segmentation Efficiency

With the integration of the SAM model, segmentation performance was optimized for both speed and accuracy. Mask generation times varied based on configuration:

- Low detail: 30 seconds per image
- Medium detail: 50 seconds per image
- High detail: 80 seconds per image

### 5.3. Recipe Matching and User Satisfaction

The NLP recommendation engine successfully matched recipes based on available ingredients with a user satisfaction rate of 87%, based on preliminary testing. Users reported improved meal variety and a reduction in food waste due to tailored suggestions.

## 6. Future Work

### 6.1. Inventory Tracking

Future iterations of Recipix aim to implement an inventory tracking system, where the system can remember the contents of the fridge and provide real-time updates on what is available. This will integrate with online grocery platforms for automated shopping lists and restocking.

### 6.2. Dietary-Specific Recommendations

Additional development will be focused on expanding support for specific diets such as keto, paleo, and low-carb, as well as incorporating dynamic ingredient substitutions for users with strict dietary needs or allergies.

### 6.3. Enhanced NLP

The recommendation system will evolve by integrating custom transformer-based models (e.g., BERT) or Large Language Models (LLM) to provide deeper contextual understanding of recipes and improve the overall user experience.

## 7. Conclusion

Recipix demonstrates the power of AI to solve real-world problems in food management and sustainability. By integrating state-of-the-art computer vision models like ResNet-50 and SAM with an intelligent NLP-based recommendation engine, Recipix offers a robust solution for reducing food waste and optimizing meal planning. The results indicate high classification accuracy, effective recipe matching, and a significant reduction in food waste for users.

Future work will focus on expanding the system's capabilities, including real-time inventory tracking and deeper integration with smart home ecosystems.

## References

[1] Perez, M. (2023). What is Web Scraping and What is it Used For?. ParseHub.

[2] Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv*.

[3] Gustavsson, J., Cederberg, C., Sonesson, U., van Otterdijk, R., & Meybeck, A. (2011). Global food losses and food waste. Food and Agriculture Organization of the United Nations.