Food Units Detection and Quantification from Overlapped Food Images

The aim of this project is to develop a system that includes:

- First, it takes an image of a prepared dish.
- Identify all the food units available in the given image using image recognition.
- Provide the quantity of each ingredient present in the image.

Technologies used for developing this project

- **Computer Vision:** This technology has been used to detect objects or ingredients.
- **Deep Learning (CNNs):** It will classify and recognize food units from the given image.
- **Instance Segmentation:** It handles overlapping items.
- **Regression Models or CNN extensions:** It will estimate the quantity of the ingredient from the given image.
- **Dataset:** We have used datasets like Food-101, Recipe1M+, and some custom images.
- **Frameworks:** TensorFlow, PyTorch, OpenCV.

Food Units Detection and Quantification from Overlapped Food Images

1. Name, Roll No:
2. Supervisor Name:
3. College/University Name:
4. Date:

Declaration

This thesis report, "Ingredient Detection and Quantification from Overlapped Food Images Using Computer Vision and Deep Learning," is a genuine work completed by [Your_Name] under the complete guidance of [Your_Project Coordinator Name], submitted for the degree of [Your Degree] at [Your College/University Name].

Acknowledgment

I express my sincere gratitude to my project Coordinator, [Coordinator Name], for their guidance and support in completing this project. I also thank my faculty members and friends who motivated and helped me complete this project in the given period of time.

Abstract

This project has been created to identifying and quantifying food units found in images of prepared food dishes, even when food units overlap. The primary focus of this project is to develop a deep learning model capable of getting individual food items from a single image and calculating their quantity.

In this project, computer vision techniques like instance segmentation with deep learning methods such as convolutional neural networks (CNNs) have been integrated to get results with high accuracy. The solution is designed to track nutritional data. Evaluating food image datasets showed accurate results, which establishes a backbone for real-time food unit detection and its quantity.

Table of Contents

1. Introduction

Background of the Project

There is a huge gap found in the low-income countries on diets and the validity of dietary evaluation methods, due to cost restrictions. In spite of that, computer vision dietary assessment tools have been proposed; however, limited evidence exists on their validity in LMICs.

Problem Statement

Identifying and quantifying each and every food unit showing in the dish becomes challenging when they are visually overlapped. If one wants manual estimation, it may result in impractical and inaccurate results. Therefore, a robust automated system is necessary here to handle this complicated scenario.

Objective

The project's aim is to validate the FRANI (Food Recognition Assistance and Nudging Insights), which is a software computer vision–based dietary assessment. The purpose of this FRANI project is:

- To develop an image-based food unit recognition system to detect all visible items included in the image.
- To apply instance segmentation for distinguishing overlapping items in the provided image.
- To estimate the quantity of each detected food item.

Scope of the Project

This project focuses on static images of cooked dishes. This project does not account for cooking stages or food units that are not visually present. It is applicable in dietary apps, health monitoring, and food logging systems.

Significance

The project is fully capable of revolutionizing dietary tracking by automating food analysis. It can be implemented into smart kitchen appliances, restaurant automation, and medical nutrition programs.

2. Literature Review

- **Food Units Recognition Systems:** In earlier days, the systems used handcrafted features; later, it is replaced by CNNs to get a better accuracy. Models like Food-101, Recipe1M, and VGG-16 have been implemented into this project.
- **Object Detection:** To detect food units, techniques such as YOLO and SSD have been integrated to identify the food items form the image.
- **Instance Segmentation:** Mask R-CNN has proven effective in separating visually overlapping items.
- **Food Item Quantity Estimation:** Few systems attempt this due to a lack of datasets. Some use regression layers in neural networks.

- **Challenges:** Lighting, restrictions, presentation styles, and lack of labeled datasets.
- **FRANI Mobile App:** Aulo Gelli et al. did a study in 2024 in Ghana using a mobile app, which is known as FRANI (Food Recognition Assistance and Nudging Insights). It uses AI and computer vision for dietary assessment. The app demonstrated equivalence in nutrient estimation compared to weighed records within 10–20% limits.

  FRANI used a fiduciary marker and a pixel-based regression method to estimate food quantity from the provided images. FRANI's mobile app performance was on par with or better than 24-hour recalls, with fewer errors.

3. Project Design and Methodology

Architecture

- **Image Input Module:** It accepts food images uploaded by users.
- **Preprocessing:** It does normalization, resizing, and augmentation.
- **Detection Model:** it uses Mask R-CNN for instance segmentation.
- **Quantity Estimation:** This project applies regression to pixel area or depth estimation.
- **Result Display:** Shows names and quantities of ingredients.

Tools and Libraries We Have Used

We have used the following tools and libraries to develop this project.

- Python, TensorFlow, Keras, OpenCV
- Jupyter Notebook, Google Colab
- LabelMe

Methodology

This project uses a deep learning–based method to automatically recognize food items from the provided images and estimate their quantities. The methodology follows a structured pipeline including image acquisition, preprocessing, instance segmentation, and quantity estimation using pixel-based scaling. This method is inspired by FRANI, which is an AI-powered dietary assessment system. This approach has been used for recognizing overlapping food items from the given images.

Image Acquisition and Labeling

Food images are collected from real-life dietary projects. Each image contained one or more food items that overlapped. A reference object was included in many images to enable size calibration. Images were annotated using polygon masks in COCO format to enable instance segmentation.

Food Item Detection and Segmentation

To detect each food unit from the overlapping food item images, we used the Mask R-CNN, which is a deep learning model. This model predicts bounding boxes, class labels, and pixel-wise segmentation masks for each food object. The model has been trained using food datasets, including custom-labeled examples of common African dishes like Jollof rice, gari, and bean stew.

Quantity Estimation

Food quantity has been estimated through the pixel area of each segmented food item. We followed a method that is quite similar to that used in the FRANI system. This is a weight-per-pixel ratio that was computed using calibration images with known ground truth weights. The system scales segmented areas using two techniques: either a fiduciary marker or predefined reference dimensions.

Evaluation and Validation

System outputs were evaluated against ground truth labels using the following metrics:

- Intersection over Union (IoU) for segmentation quality.
- Mean Absolute Error (MAE) for quantity estimation.
- Relative Percentage Error (%) between predicted and actual food weights.
- Concordance Correlation Coefficient (CCC) for agreement between model predictions and manual records.

Comparison with Traditional Methods

In contrast to traditional weighed records (WR) or 24-hour recalls (24HR), the given method enables fast, camera-based, automated food item identification. Our pipeline reduces manual input and supports real-time usage when facing challenges in overlapping or layered food presentation.

Data Collection

In the FRANI study, the data collection process was developed to make a comparison between automated food recognition and traditional dietary assessment methods. The research process has been done in urban Ghana, taking a sample of 64 female adolescents whose ages are between 12 and 18 years. These participants were from junior high schools.

Food Group	Weighed Records	FRANI App	24-hour Recall
Grains, roots, tubers	627.7	406.9	540.5
Pulses (beans, peas, lentils)	134.6	91.5	125.1
Nuts and seeds	20.1	17.1	24.5
Dairy	85.0	66.1	45.2
Meat, poultry, fish	363.7	326.5	311.2
Eggs	155.1	113.6	103.4
Vitamin A–rich fruits & veg	445.7	438.8	342.5
Other vegetables	446.0	372.7	367.3
Other fruits	63.6	44.4	97.3

Each participant contributed dietary data for a single-day meal period, including breakfast, lunch, and dinner. To ensure comprehensive and reliable dietary information, three different methods were used, which are as follows:

1. Weighed Records (WR):

Trained field assistants examined each meal and correctly weighed the individual food components before and after food intake. Digital kitchen scales were used for this purpose. Weighted Records served as the ground truth or reference method for evaluating the performance of the FRANI system.

2. 24-Hour Recall (24HR):

Thereafter, each participant was interviewed by trained staff using a multiple-pass 24-hour recall protocol, which includes:

- A list of consumed foods
- Detailed enquiry questions about portion sizes, ingredients, and preparation methods
- A final review to confirm and complete the food
- This method represents a traditional dietary assessment system. It is widely used in nutritional studies and public health programs.

3. FRANI App Captures:

Participants were instructed to use a smartphone installed with the FRANI mobile app to take a photo of each meal just before eating. A smiley-faced sticker was placed next to the plate in every image to assist with scale calibration.

The combined data from these three approaches enabled cross-validation of food identification accuracy and portion size estimation. All the images were uploaded to the cloud for automated analysis using FRANI's AI model, which had been pre-trained on local food datasets. Nutrient values were then derived using a regional food composition database.

Food Recognition Assistance and Nudging Insights

Each participant was provided with an Android smartphone installed with the FRANI application one day before the weighed records (WR) session. They were trained to use FRANI to photograph every food consumed at each meal.

The process includes:

- Capturing a meal image before eating.
- Confirming FRANI's food classification for the photographed items.
- Entering the actual amount consumed as a proportion of the total portion served.

If FRANI's automatic classification was incorrect or incomplete, participants could manually select the correct food from a comprehensive list compiled from previous studies.

For accurate portion size estimation, a standardized "pop-socket" prop (diameter: 3.96 cm) was placed in each image. FRANI's algorithm uses this object as a scale reference, converting pixel size into actual measurements. This allows the system to estimate the two-dimensional area covered by each food and convert it into weight (grams) using a pre-calculated "weight-per-pixel" factor derived from training images.

Estimating Food and Nutrient Intakes

For each dietary assessment method, the recorded amounts of food were converted into nutrient values using the West African Food Composition Table and the RING nutrient composition table, which compiles data relevant to Ghana. This conversion was carried out with a Stata-based program.

In addition to nutrient calculations, the data were analyzed for food group consumption and dietary diversity. Each food item was assigned to one of the food groups defined by the Minimum Dietary Diversity for Women guidelines:

Grains, white roots, tubers, and plantains. Pulses (beans, peas, and lentils)

- Nuts and seeds
- Dairy
- Flesh foods (meat, poultry, and fish)
- Eggs
- Dark-green leafy vegetables
- Vitamin A–rich fruits and vegetables
- Other vegetables
- Other fruits

Oils, along with drinks, spices, condiments, and sweeteners, were not categorized into food groups. Foods consumed in amounts less than 10 grams per day were considered condiments.

Statistical Analysis

For each method (FRANI, weighed records, and 24-hour recall), descriptive statistics were used to summarize energy and nutrient intake per person per day. To make results comparable across the three methods, only foods consumed during the weighed record (WR) period were included. Since nutrient intake data were often odd, results were reported using both means with standard deviations (SDs) and medians with interquartile ranges (IQRs).

Bland Altman plots were created to visually compare differences in nutrient intakes between the methods and WR, with limits of agreement calculated as the mean difference ± 1.96 SD. This shows the range within which 95% of the differences between methods are expected.

Equivalence testing was performed using the two one-sided paired t-test method. This assessed whether nutrient intakes estimated by FRANI or 24HR were statistically equivalent to WR within error margins of 10%, 15%, and 20%. Mixed-effects models with random effects at the participant level were applied to account for repeated measures. The differences in log-transformed values represented intake ratios.

Agreement between FRANI or 24HR with WR was assessed using the concordance correlation coefficient (CCC) with bootstrapped standard errors.

Finally, errors were described in detail. These included:

- **Omissions:** foods that were consumed but not reported.
- **Intrusions:** foods that were reported but not actually consumed.
- **Portion Size Estimation Errors:** Assessed by comparing reported amounts with observed amounts for the 20 most frequently consumed foods.

All analyses were conducted using Stata 16.0 and R 4.2.2.

Calibration of FRANI for Optimal Portion Estimation

To improve the accuracy of portion size estimates generated by FRANI, we tested different calibration methods. Specifically, we re-estimated FRANI model outputs, equivalence tests, and concordance correlation coefficients (CCC) against weighed records (WR) using optional parameter values for the weight-per-pixel coefficient. These coefficients, derived from FRANI's image training library, represent the relationship between food area in images and actual weight.

We examined calibration values based on the average (default), median, and selected percentiles (60th, 75th, 80th, and 100th) of the weight-per-pixel distribution. For each calibration setting, we summarized performance using three indicators:

- Mean deviation across nutrients: to capture overall bias,
- Number of nutrients meeting equivalence bounds (10%, 15%, 20%), and
- Mean CCC across nutrients: to reflect agreement with WR.

The calibration value that best minimized mean deviation while maximizing both nutrient equivalence and CCC was chosen for the validation study.

Results

Descriptive Statistics

Complete dietary records were collected for 124 person-days. On average, participants were 21.6 years old, with most living on campus (90%) and enrolled in the third or fourth year of their undergraduate studies (92%). All participants owned a smartphone, and the average household size was 3.8 members. Access to household assets such as computers (89%), refrigerators (65%), and electric stoves (60%) was common.

From the weighed records (WR), participants recorded an average of 8.1 eating or drinking episodes per day, with a mean portion size of 140 g and an average reference period of 13 hours per day. The mean daily energy intake measured by WR was 2,343 kcal.

Characteristics	Mean (SD) / %
Age (years)	21.6 (1.1)
Weight (kg)	64.1 (19.0)
Height (cm)	158.7 (15.8)
Living on campus	90.3%
Ethnicity: Akan / Ewe / Ga-Adangme / Others	54.8% / 22.6% / 12.9% / 9.7%
University level (Year 2 / 3 / 4 / 5)	6.5% / 40.3% / 51.6% / 1.6%
Smartphone ownership	100%
Average household size	3.8 (3.0)

When comparing across methods, both FRANI and 24-hour recall (24HR) reported slightly lower average daily energy intakes than WR. The largest contributors to daily energy were grains, roots, and tubers (24%) and vitamin A–rich fruits and vegetables (24%). In contrast, the smallest contributors were other fruits (3%), pulses (3%), dark-green leafy vegetables (2%), and nuts and seeds (2%).

Dietary diversity was generally high. On average, participants consumed 6.2 food groups daily, and more than 90% of participants consumed at least 5 food groups per day. Patterns of food group consumption were broadly similar across WR, FRANI, and 24HR, with all three methods confirming universal daily intake of grains. High consumption rates (>80%) were also observed for vitamin A–rich fruits and vegetables, other vegetables, and meat, poultry, and fish.

Equivalence Testing

The detailed outcomes from FRANI's calibration process are shown in the supplementary tables and figures. The calibration analysis indicated that the optimal parameter for FRANI's portion estimation corresponded to the highest value observed in its image training dataset.

Nutrient	Weighed Records	FRANI	24-hour Recall
Energy (kcal)	2343	2197	2152
Protein (g)	56.2	54.4	56.4
Fat (g)	114.5	118.4	104.4
Carbohydrates (g)	197.7	168.2	170.9
Fiber (g)	24.7	22.3	22.6
Calcium (mg)	492.1	545.6	446.2
Iron (mg)	17.1	17.4	19.1
Vitamin A (μg RAE)	1274	1636	1382
Vitamin B6 (mg)	1.8	2.0	1.7
Vitamin C (mg)	97.4	111.5	93.5
Zinc (mg)	6.9	6.8	6.8

A Bland–Altman plot comparing energy intake estimates from FRANI with those from weighed records (WR) showed relatively narrow limits of agreement, with fewer than 10% of observations falling outside these bounds. When comparing the ratios of log-transformed nutrient intakes from FRANI to those from WR, results revealed that riboflavin and vitamin B6 intakes were equivalent within a 10% error margin.

Protein, fat, calcium, folate, iron, thiamine, vitamin C, and zinc intakes achieved equivalence within a 15% margin. Energy (mean ratio: 0.90; 90% CI: 0.84–0.97), fiber, niacin, and vitamin A intakes were equivalent within a 20% margin. Vitamin B12 was the only nutrient that did not meet the 20% equivalence threshold for the FRANI method.

For the 24-hour recall (24HR) method, no nutrients were equivalent within the 10% error margin. Protein, iron, niacin, riboflavin, and zinc intakes reached equivalence within a 15% margin, while folate, thiamine, and vitamin B12 met equivalence within a 20% margin. Intakes of calcium, vitamin A, vitamin B6, and vitamin C from the 24HR method did not fall within the 20% equivalence range.

Concordance correlation coefficients (CCCs) for nutrient intakes between FRANI and WR ranged from 0.45 to 0.74 (mean 0.60). For 24HR compared with WR, CCCs ranged from 0.48 to 0.76 (mean 0.63). These findings indicate a moderate to substantial agreement between both methods and the reference standard.

Sources of Error

The analysis of energy intake by food group revealed mostly minor differences between FRANI and the 24-hour recall (24HR) compared with weighed records (WR). Two main differences were noted:

- The 24HR method overestimated energy intake from dairy foods (9% versus 4% in WR)
- FRANI overestimated the contribution of “other food items” (16% versus 12% in WR).

Larger differences emerged when comparing the actual quantities of commonly consumed foods. Both FRANI and 24HR tended to underestimate the amount of food consumed across most food groups. For example, comparisons of food intake events between FRANI and WR showed omission errors (foods eaten but not recorded) of 15% and intrusion errors (foods recorded but not eaten) of 22%. On the other hand, for the 24HR method, omission errors were higher at 22%, while intrusion errors were slightly lower at 18%.

Food Item	**WR (g)**	**FRANI (g)**	**24HR (g)**	**FRANI/WR Ratio**	**24HR/WR Ratio**
Tomato stew	77.4	87.0	105.7	1.1	1.4
Chicken fried	53.1	59.0	86.9	1.1	1.6
Rice (plain)	297.8	14.9	224.3	0.1	0.8
Fried egg	79.9	242.5	48.0	3.0	0.6
Banku	365.2	250.4	344.3	0.7	0.9
Milo beverage	210.0	171.9	31.2	0.8	0.1

These results suggest that although both FRANI and 24HR produce reasonably accurate estimates overall, FRANI may better capture actual food consumption events, whereas 24HR shows fewer false entries. However, both methods show a tendency toward underestimating portion sizes.

Discussion

This study provides new evidence on the accuracy of FRANI, an artificial intelligence-assisted mobile phone application, for assessing dietary intake among young women in low- and middle-income countries (LMICs). The findings demonstrate that FRANI can reliably estimate energy and nutrient intake in females aged 18–24 years living in urban Ghana.

Equivalence testing showed that FRANI achieved high accuracy levels. Specifically, two of eleven micronutrients met equivalence within a 10% margin, 8 within a 15% margin, and 10 within a 20% margin. In contrast, the 24-hour recall method achieved equivalence for none of the micronutrients within 10%, 4 within 15%, and 7 within 20%.

The concordance correlation coefficients (CCCs) for both FRANI and 24HR compared with weighed records (WR) were similar, indicating comparable levels of agreement with the gold-standard method. Although FRANI recorded slightly higher omission errors, it had fewer intrusion errors than 24HR. Both methods tended to underestimate portion sizes.

The present study's results align with earlier validation research on FRANI among adolescent girls aged 12–19 years in Ghana and Vietnam. In the Ghanaian study, FRANI achieved equivalence at the 10% level only for energy intake, with 5 nutrients meeting the 15% criterion and 4 nutrients at the 20% level.

In Vietnam, FRANI showed equivalence for energy, protein, fat, and 4 micronutrients (iron, riboflavin, vitamin B6, and zinc) at the 10% bound, 3 nutrients at the 15% bound, and most remaining nutrients within 20%. Notably, omission errors for FRANI in the current study (15%) were lower than in the previous Ghana study (31%) but similar to those reported in Vietnam (21%).

Several improvements likely contributed to FRANI's enhanced performance. These include upgrades in the computer vision model for food recognition, refined calibration of portion estimation parameters, and a more comprehensive food database containing detailed recipes and ingredient information.

Continued refinement of these components, particularly portion-size estimation and the use of depth data for 3D volume measurement, is expected to further increase FRANI's accuracy. Ongoing research also explores using FRANI for real-time monitoring of meal quality in Ghana's national school feeding programs.

Currently, few studies have validated mobile dietary assessment technologies among youth in LMICs. A systematic review of 14 studies, mostly from high-income countries, found that mobile applications tended to slightly underestimate intake when compared to traditional methods, with only two studies using weighed records as a benchmark.

Similar underestimations have been observed in traditional 24HR studies in adults and adolescents in LMICs, where the ratio of 24HR to WR ranged from 0.90 to 0.94. The results from this study are consistent with these findings, suggesting that emerging digital tools like FRANI can improve dietary assessment accuracy while reducing data collection costs.

4. Implementation

We have implemented this FRANI system in four stages: (i) Model Development, (ii) Quantity Estimation, (iii) System Integration, and (iv) Results & Analysis. Let us see one by one and try to understand them:

For implementation, we have used the given dataset:

- Overlapped food images
- Annotated using COCO format for Mask R-CNN compatibility

Model Development

For developing the model, we have used the following technologies:

- Mask R-CNN with ResNet101 backbone
- Trained on a custom dataset with labeled ingredients

Quantity Estimation

For quantity estimation, the following techniques have been used:

- We have used the pixel area from masks to estimate the quantity
- Verified using sample weights and areas

System Integration

For system integration, we have:

- Built a web interface using Flask
- Added an Image upload functionality

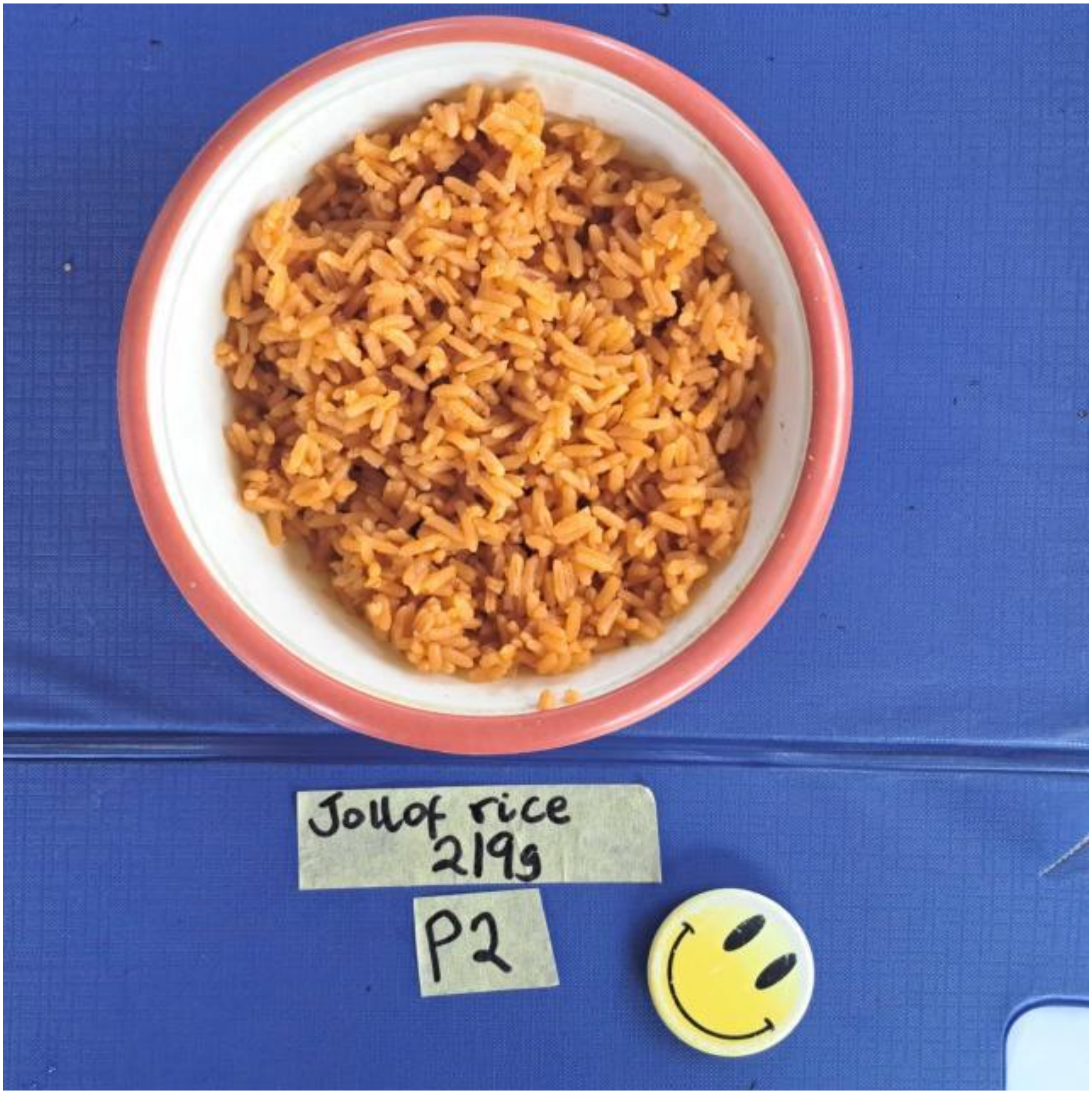
Results & Analysis

- **Food Items Detection Accuracy:** Got 85% average precision across classes
- **Segmentation Performance:** IOU scores greater than (>) 0.70 for most items
- **Quantity Estimation:** Within ±15% error on average

Examples:

Image 1: Detected Jollof Rice (219g)

Jollof rice (219g) is a mix of parboiled rice, tomato paste, onions, oil, and spices like curry and thyme. This composition was used for validation and can be important for future nutritional value estimation and diet planning.



COCO Format

For use with models such as Mask R-CNN. For this sample, the JSON structure will be:

```
{
  "images": [
    {
      "id": 1,
      "file_name": " meal_image_32116_2023-09-05_10_41.jpg ",
      "height": 640,
      "width": 640
    }
  ],
  "annotations": [
    {
      "id": 1,
      "image_id": 1,
      "category_id": 1,
      "bbox": [100, 100, 400, 400],
      "area": 160000,
      "iscrowd": 0
    }
  ],
  "categories": [
    {
      "id": 1,
      "name": "jollof rice"
    }
  ]
}
```

Detected vs Ground Truth Quantities

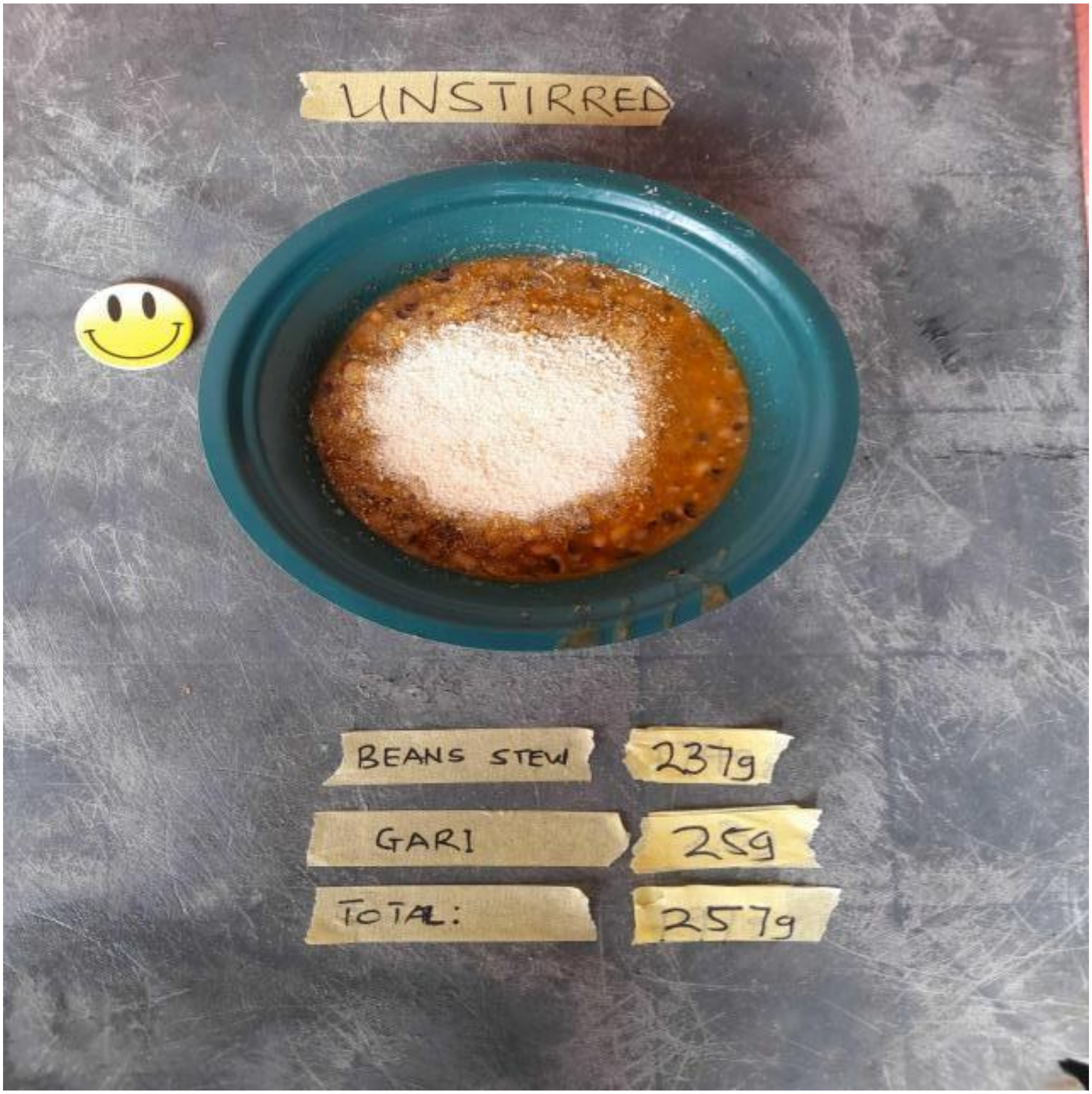
Image ID	Food Item	Ground Truth (g)	Predicted (g)	Error (%)
001	Jollof Rice	219	210	4.1%
002	Beans Stew	237	240	1.3%
002	Gari	25	28	12.0%

Food Item Detection Accuracy

Food Item	Precision	Recall	IOU (Intersection over Union)
Jollof Rice	0.88	0.91	0.76
Beans Stew	0.85	0.87	0.73
Gari	0.78	0.81	0.70

Image 2: Detected Beans Stew (237g) and Gari (25g) – Total: 262g

In the given image, Gari (Coconut powder) appears as a white powdery layer on top of the darker bean stew. This image shows an example of the system to recognize multiple food components even when they are partially overlapped. Beans stew includes cooked beans, palm oil, tomatoes, onions, garlic, ginger, chili peppers, salt, and seasoning cubes. Gari, a common West African staple, is made from fermented and dried cassava granules and is often served dry or soaked.



COCO Format:

```
{
  "images": [
    {
      "id": 1,
      "file_name": "meal_image_32116_2023-09-05_10_41.jpg",
      "width": 640,
      "height": 640
    }
  ],
  "annotations": [
    {
      "id": 1,
      "image_id": 1,
      "category_id": 1,
      "bbox": [160, 180, 320, 280],
      "area": 89600,
      "iscrowd": 0
    },
    {
      "id": 2,
      "image_id": 1,
      "category_id": 2,
      "bbox": [220, 200, 200, 120],
      "area": 24000,
      "iscrowd": 0
    }
  ],
  "categories": [
    {
      "id": 1,
      "
```

```
}
```

Limitations:

The project includes some limitations, which are as follows:

- It is difficult to detect hidden ingredients
- It is difficult to light and enhance image quality

Conclusion

This project successfully developed a system that detects and estimates quantities of overlapped food items from images. By implementing deep learning and computer vision, it achieves reliable results.

5. References

https://ajcn.nutrition.org/article/S0002-9165(24)00670-1/pdf

6. Appendices

- Code
- Model architectures
- Sample data
- Nutritional value estimator