ADAMSON UNIVERSITY
COMPUTER SCIENCE DEPARTMENT
2nd Floor, SV Building
900 San Marcelino St., Ermita, Manila 1000 Philippines
Trunkline: +63 2 524-2011 local 305

M-PCAM: Food Volume Estimation Using Mask R-CNN and the Pinhole Camera Model Technique with Low-Resolution Depth Images from Android's Depth API

In partial fulfillment of the requirements for CSRP 2

By:

Baesa, Paul Andre B.

Reyes, Paul Lorence L.

Sicat, Benito Rapheal IV.

November 2024

Contents

Abstract 0

Chapter I: The Problem and its Background 2

A. Conceptual Framework 2

Chapter II: Review of Related Literature and Studies 2

A. Food Image Analysis and Mask R-CNN 2

B. Pinhole Camera Model Integration 3

Chapter III: Methodology 3

A. System Architecture 3

B. Implementation Details 4

C. Experimental Setup 5

D. Capture Protocol 6

Chapter IV: Results and Discussion 8

Chapter V Conclusion and Future Work 12

A. Technical Improvements 10

B. Application Enhancements 10

C. Validation and Testing 10

D. Future development will prioritize: 10

References 11

M-PCAM: Food Volume Estimation Using Mask R-CNN and the Pinhole Camera Model Technique with Low-Resolution Depth Images from Android's Depth API

Jerome Alvez, MSCS
Computer Science Department
Valenzuala, Metro Manila, Philippines
09560474595
jerome.alvez@adamson.edu.ph

Paul Jacob Cruz, MSCS
Computer Science Department
Caloocan City, Metro Manila, Philippines
09957655299
paul.jacob.cruz@adamson.edu.ph

Ma. Christina Navarro, MSCS
Computer Science Department
Manila, Metro Manila, Philippines
09560474595
ma.christina.navarro@adamson.edu.ph

Paul Lorence Reyes
Computer Science Department
Marilao City, Bulacan,
09213991071
paul.lorence.reyes@adamson.edu.ph

Benito Raphael Sicat IV
Computer Science Department
Capas, Tarlac Cityy
09609129598
benito.raphael.iv.sicat@adamson.edu.ph

Paul Andre Baesa
Computer Science Department
Imus City, Cavite
09395661686
paul.andre.baesa@adamson.edu.ph

Abstract

The increasing prevalence of portion control and health issues due to poor dietary monitoring has highlighted the need for accessible food analysis tools. This paper presents M-PCAM, a novel hybrid model combining Mask R-CNN [1], [2] with the pinhole camera model technique [6], [16], [5] for comprehensive food analysis. Our approach integrates advanced instance segmentation with geometric volume estimation to provide food portion and macronutrient assessments from smartphone images. The model achieves this through three key innovations: (1) precise food item segmentation using Mask R-CNN [1], (2) volume estimation using pinhole camera model geometry [6], [5], and (3) macronutrient calculation through reference database integration [8], [9]. Initial testing demonstrates promising results in segmentation accuracy [3], [14] and volume estimation precision [10], [5]. This research contributes to computer vision-based dietary analysis by addressing the challenges of portion estimation without specialized equipment.

Chapter I: The Problem and its Background

Computer vision in food image analysis has made significant strides, yet quantifiable volume estimation and nutritional assessment remain challenging. Current solutions often require specialized hardware or complex setups, limiting their practical application. Our research addresses these limitations by proposing a hybrid model that leverages smartphone cameras for accessible dietary monitoring.

The primary objectives of this study are:

1) Develop an M-PCAM model that processes RGB images specifically for classification, masking, and segmentation of unprocessed foods (e.g., fruits, vegetables, lean proteins, whole grains)
2) Implement a preprocessing Pipeline to downscale RGB images to match the low resolution of the pseudo depth images for the segmentation masks
3) Integrate the extracted depth data with pinhole camera model calculations for

quantifiable dietary food volume estimation

4) Validate the system's volume estimation accuracy against known measurements

A. Conceptual Framework

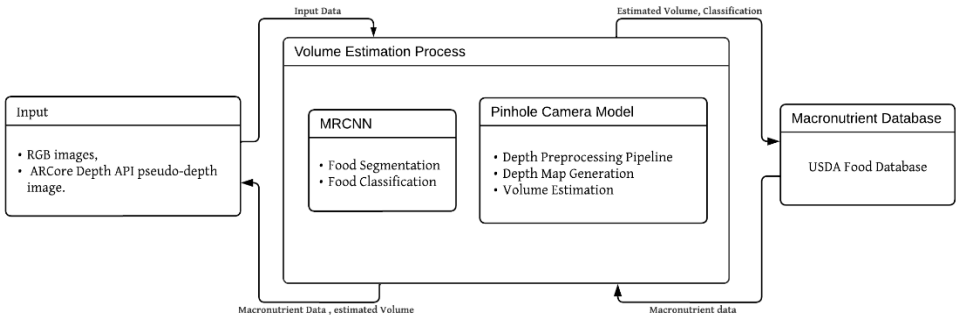


Fig 1. M-PCAM Conceptual Framework

This framework depicts an automated dietary food analysis system that processes images of natural, unprocessed food images in two steps:

1.) Volume Estimation Process:

- Uses RGB and depth images as input
- Employs MRCNN for dietary food segmentation/classification
- Utilizes the pinhole camera model technique for depth processing and volume calculation

2.) Nutritional Analysis:

- Links estimated volumes to a USDA food database validated and approved by a nutritionist.
- Provides macronutrient data for identified foods

The system connects computer vision technology with nutritional databases to analyze food portions and their nutritional content.

Chapter II: Review of Related Literature and Studies

A. Food Image Analysis and Mask R-CNN

Mask R-CNN represents a significant advancement in object detection, extending beyond mere identification to pixel-wise segmentation [1]. Recent studies have demonstrated its effectiveness in food segmentation tasks [2, 3], particularly in handling complex food scenes with multiple items. The model's ability to generate high-quality segmentation masks while maintaining computational efficiency makes it ideal for mobile applications.

B. Pinhole Camera Model Integration

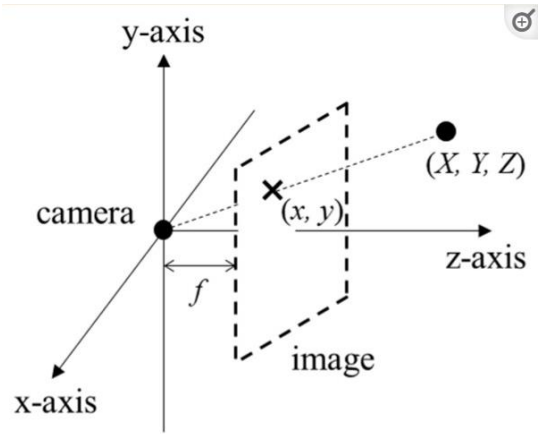


Fig. 2. Pinhole camera model geometry showing coordinate system transformation for volume estimation

The pinhole camera model provides a mathematical framework for mapping 3D world coordinates to 2D image coordinates [4]. However, several significant challenges exist in mobile-based depth sensing implementation. ARCore's Depth API, while powerful, has inherent limitations: it produces low-resolution depth maps (160x90 pixels) that require upscaling [11], and is highly susceptible to inaccuracies caused by camera movement. As SLAM-based systems like ARCore rely on continuously aligning key points and tracking camera motions, even minor hand movements can introduce drift and result in inconsistent depth maps. This issue is compounded when depth scans accumulate errors over time, leading to misaligned spatial reconstructions if stabilization is not maintained during image capture. Additionally, the accuracy of ARCore's depth sensing is heavily influenced by environmental facts which will be further discussed in Chapter 4.

These limitations necessitate a highly controlled capture environment and specific set up conditions to ensure consistent and quantifiable measurements [39].

Chapter III: Methodology

A. System Architecture

The M-PCAM framework implements a systematic approach to object analysis through multiple processing stages, as illustrated in Fig. 1.

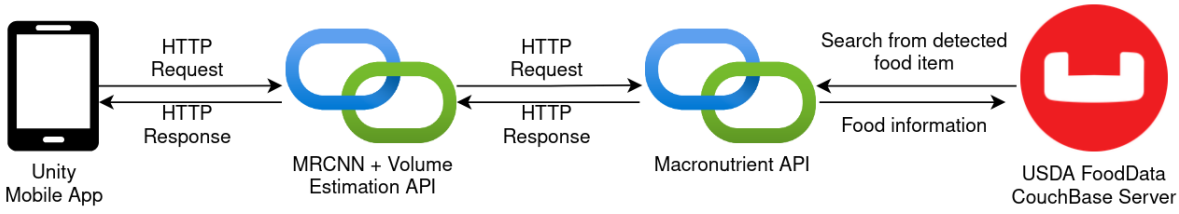


Fig. 3. M-PCAM System Architecture showing the complete system architecture from image input to macronutrient estimation output

The framework consists of three primary components:

1) **Image Processing and Segmentation Stage**
   - Image acquisition via smartphone camera
   - Preprocessing including normalization and augmentation
   - RGB and depth image channel separation
   - Mask R-CNN implementation for food item segmentation
   - Generation of pixel-level segmentation masks
2) **Volume Estimation Stage**
   - Integration of Mask R-CNN segmentation output
   - Pinhole camera model application for 3D reconstruction
   - Reference object (plate) calibration
   - Geometric transformation for volume calculation
   - Depth map integration for quantifiable measurements

a) Mathematical Foundation

$$S[u,\ v,\ 1]\ =\ k[R|t]\ \cdot\ [x,\ y,\ z,\ 1]$$

a.1) Projection Equation (World to Image):

Where:

- $[u, v]$ are image coordinates
- $[X, Y, Z]$ are world coordinates
- S is scale factor
  - $S\ =\ \frac{plateReal}{platePixel}$
- K represents intrinsic matrix
- $[R|t]$ represents extrinsic matrix (rotation and translation)

a.2) Depth to 3D Conversion:

$$X = \frac{(x-c_x)ZS}{f_x}\quad Y = \frac{(y-c_y)ZS}{f_y}$$

Where:

- (c_x, c_y) are principal point coordinates
- (f_x, f_y) are focal lengths
- Z is the depth value

a.3) Volume Calculation:

$$V_{food} = \sum\sum\left(Z(x,y) - Z_{plate}(x,y)\right)\cdot dA\ z$$

Where:

- $Z(x, y)$ is height at point (x, y)
- $Z_{plate}(x, y)$ is the plate's reference height
- dA is differential area element

a.4) Measurement Reliability Analysis

Stability Score System

The system implements a comprehensive stability score calculation that evaluates multiple factors

3) **Macronutrient Analysis Stage**
   - Food item classification based on segmentation results
   - Volume to mass conversion using density estimates
   - Integration with food nutrient database
   - Macronutrient content calculation
   - Result visualization
   - Nutrional Content estimation using the USDA FoodDatabase, validated by Raizel Jae B. Roxas, RND, from Luxen Nutrition
   - USDA FoodDatabase JSON data stored in CouchBase [38]

The system employs a standardized plate as a reference object for consistent volume estimation and includes error handling mechanisms at each processing stage to ensure reliable results.

B. Implementation Details

1) **Dataset Organization and Processing**
   - Training: AIcrowd Food Recognition 2022 dataset
   - Validation: Laboratory-measured food volumes
   - Testing: Real-world smartphone captures
4) **C#-based Mobile Application Development**

- Unity Android Build Target using ARFoundation [11]
- Developed through Unity using C#
- Depth sensing through dual-pixel auto-focus cameras or time-of-flight sensors
- API provides two key functionalities:
  - Depth map generation: Creates a buffer of depth data where each pixel contains the distance from the camera to the scene
  - Raw depth measurements: Returns a depth image with millimeter measurements of the physical distance between the camera and each pixel in the scene
- Camera poses tracking for consistent measurements across different angles
- Real-time depth estimation using both active and passive depth sensing:
  - Active: Structured light or time-of-flight sensing for direct depth measurement
  - Passive: Stereo matching and motion tracking for depth inference

5) Model Training and Fine-tuning

- Data augmentation techniques for improved robustness
- Transfer learning from pre-trained weights
- Fine-tuning for food-specific features

6) Macronutrient Application Programming Interface (API) Development:

- Retrieve Macronutrient Comma Separated Values (CSV) dataset from United States Department of Agriculture (USDA)
- Import CSV to PostgreSQL Database
  - Write an API using GoLang's built-in HTTP Module

The depth sensing capabilities provided by ARCore's Depth API are crucial for accurate volume estimation. The API generates a depth map alongside the RGB image, where each pixel contains distance information between the camera and the corresponding point in the scene. This dual-image capture approach enables precise 3D reconstruction of food items when combined with the pinhole camera model calculations.

C. Experimental Setup

A. Hardware Configuration

The experimental setup consists of these key components:

1) A smartphone equipped with ARCore's Depth API was utilized to provide accurate depth perception and augmented reality functionality. This device serves as the primary platform for capturing depth data and rendering AR content. The setup used Poco X6 Pro which is included from Google's compatibility list [

2) The setup employed a lightweight, cost-effective aluminum alloy tripod as the primary support structure. To ensure optimal lighting conditions, three compact LED lights were securely mounted, one on each leg of the tripod. This arrangement provided uniform and controlled illumination across the experimental environment.

3) Controlled Lighting System Specification: The 3 mounted LED lights were chosen for their portability and efficiency with specifications detailed below:

- Dimensions: 6 × 4.2 cm (2.4 × 1.65 in)
- Maximum Brightness: 400 lumens

B. Environmental Controls

To ensure measurement consistency and reliability:

1) Device Positioning

- Smartphone handled below the tripod while still having a bird's eye view of the object at around 35 to 38 cm height
- Camera oriented perpendicular to the surface
- Device stability maintained throughout capture process

2) Lighting Configuration

- 3 mounted LED lights, stable, and diffused lighting directed at the plate
- Minimized shadows and reflections
- Consistent illumination across the entire capture area
- Luminance at minimum 400 lumens

3) Reference Object

- Standard white plate for consistent calibration
- Placed on non-reflective, solid-colored surface
- Fixed position marking for repeatable placement

D. Capture Protocol

1) Pre-capture System Calibration

- Environment temperature stabilization (20-25°C)
- Camera lens cleaning
- Depth sensor warm-up period (minimum 2 minutes)

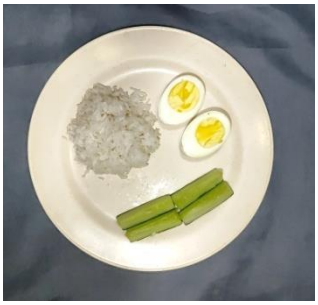
2) Image Acquisition

- Estimated height on the mobile app is around 35cm
- Minimum 5-second stabilization period between captures
- Automatic exposure and focus locked during session

Chapter IV: Results and Discussion

This chapter presents the results of the M-PCAM system, focusing on its capabilities to estimate the volume of food items using low-resolution depth data. Observations are based on controlled test cases conducted under optimal environmental conditions, with the output aligned with the study's primary objective of quantifiable volume measurements

Fig. 4. Controlled Test Case: Rice (109g), Cucumber(56g), and Hardboiled Egg (56g)



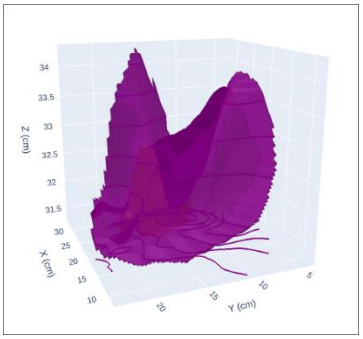


Fig. 4a. Unprocessed Depth Map

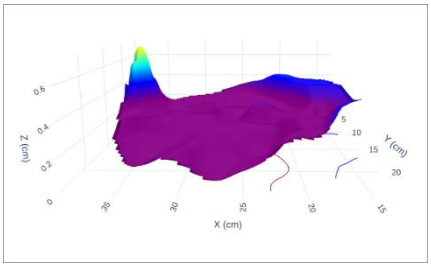


Fig. 4b. Processed Depth Map

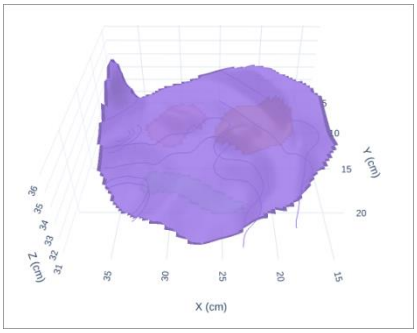


Fig. 4c. Segmented Depth Visualization

The system's performance was evaluated using a test case comprising rice, cucumber, and hardboiled egg on a white plate. Fig. 3 presents the complete processing pipeline:

1) Initial Capture and Depth Map Processing:

- The raw depth image exhibited variability in depth values, with a range of [352, 439].
- Noise and outliers were present, requiring preprocessing to ensure accuracy.
  - Expected depth range: [0.367, 0.367]
- Processed Depth Map (Fig. 3b): After applying bilateral filtering and outlier removal, depth values were normalized and scaled to align with real-world dimensions.
- Depth values normalized and scaled to real-world dimensions
- Depth scale factor: 0.0869

2) Segmentation Performance:

- Object Detection and Segmentation (Fig. 3c):
- The Mask R-CNN model generated segmentation masks for each food item in the RGB image. These masks were downscaled to match the resolution of the depth map (160x90).
  - Plate area: 68,038 pixels (original), 3,204 pixels (aligned)
  - Hardboiled-Egg area: 5,695 pixels (original), 268 pixels (aligned)
  - Rice area: 9,544 pixels (original), 446 pixels (aligned)
  - Cucumber area: 5,543 pixels (original), 259 pixels (aligned)
  - Clear separation between rice and egg regions

3) Depth Processing Analysis:

a) Initial Depth Data Characteristics:

- Raw Resolution: 160x90
- Initial Value Range: [352, 439] (raw units)
- Expected Range: [0.327, 0.329] meters
- Data format: 16-bit unasigned integer

b) Preprocessing Pipeline:

- Noise Reduction:
  - Bilateral filtering applied to preserve edges while reducing noise
  - Removed outlier points
- Depth Value Scaling:
  - Applied depth scale factor: 0.0869
  - Conversion from raw units to centimeters
  - Warning logged for values outside expected range
- Invalid Value Handling:
  - Zero Values removed
  - Negative depths filtered out
  - Statistical outlier removed using IQR Method
- Quality Metrics:
  - Original valid points in plate mask: 3,204
  - Points remaining after noise reduction: ~2,500
  - Percentage of usable depth data: ~78%
- Focal length: 93.86 pixels
- Pixel size: 0.351576 cm/pixel

4) Final Output Analysis:

- Segmented Depth Visualization (Fig. 3c):
  - Rice: The system demonstrated high accuracy for rice, with an error margin of -0.36 g (0.33% deviation). This performance underscores the suitability of M-PCAM for

foods with consistent geometries.
    - Cucumber and Hardboiled Egg: Both items exhibited higher deviations, primarily due to their irregular shapes and smooth surfaces. This suggests a need for improved depth handling techniques, particularly for non-flat objects.
- The preprocessing pipeline successfully reduced noise in the raw depth data, achieving a retention of ~78% usable points. However, artifacts from interpolation during depth scaling occasionally affected the accuracy of the calculated volumes.
- Despite alignment challenges at object boundaries, the system preserved the spatial integrity of food items relative to the plate, enabling reliable volume estimations for most test cases.
- Stability scores:
    - Rice: 0.950: (high stability)
    - Hardboiled-Egg: 0.941 (high stability)
    - Cucumber: 0.888 (good stability)

5) 3D Visualization Analysis:

- The 3D visualization confirms (Fig. 3c):
    - Correct spatial relationships between food items
    - Appropriate heigh differentials
    - Clear object boundaries
    - Reference Plate Surface as baseline
- Spatial preservation:
    - Maintains correct positioning of food items relative to plate boundaries
    - Preserves spatial relationships between different food items
- Depth resolution verification:
    - Demonstrates the system's ability to capture height variations at different scales
    - Validates the downscaling process from 640x480 to 160x90 resolution for the segmentation masks

This visualization serves as a critical intermediate step between depth capture and volume estimation, providing visual confirmation of the system's ability to reconstruct food geometry before macronutrient calculations.

6) 3D Reconstruction Limitations: Several artifacts and limitations are observable in the reconstruction, some of which were anticipated in our system design (Section III.A):

- Edge discontinuities at food item boundaries, related to the Depth API limitations discussed in Section II.B
- Depth noise in flat surface areas of the plate, mitigated by the environmental controls detailed in Section IV.B.2
- Resolution constraints affecting fine detail reproduction, a known limitation of ARCore's Depth API [11]
- Limited depth precision in steep gradient areas, impacting volume estimation as noted in Section III.A.2
- Surface texture influence on depth estimation accuracy, requiring the controlled lighting conditions specified in Section IV.B.2
- Potential depth ambiguity in regions with similar height values, affecting the overall accuracy discussed in the volume estimation stage

These limitations directly influence system accuracy and provide direction for future improvements in depth sensing and reconstruction algorithms, as will be discussed in Chapter V.

7) Macronutrient API Integration:

```
{
    "data": {
        "frame_id": "130043.51440302501",
        "volumes": [
            {
                "object_name": "egg",
                "uncertainty_cups": 0.12614526165779222,
                "volume_cups": 0.3604150333079778
            },
            {
                "object_name": "rice",
                "uncertainty_cups": 0.18835098573632206,
                "volume_cups": 0.5381456735323488
            },
            {
                "object_name": "cucumber",
                "uncertainty_cups": 0.10010479145947158,
                "volume_cups": 0.28601368988420456
            }
        ]
    }
}
```


Fig. 5a. Macronutrient API Request Body

```
{
    "data": [
        {
            "found": true,
            "macros": {
                "calories": 69.57812218010513,
                "carbs": 0.46709788316713924,
                "fat": 4.84614053785907,
                "protein": 6.03334765757555
            },
            "requested_food": "egg",
            "requested_volume": 0.3604150333079778,
            "calculated_weight": 48.656029496577005
        },
        {
            "found": true,
            "macros": {
                "calories": 109.68485117936332,
                "carbs": 23.807564597071107,
                "fat": 0.2380756459707111,
                "protein": 2.2702213383635663
            },
            "requested_food": "rice",
            "requested_volume": 0.5381456735323488,
            "calculated_weight": 85.0270164181111
        },
        {
            "found": true,
            "macros": {
                "calories": 5.491462845776727,
                "carbs": 1.012488462190084,
                "fat": 0.06177895701498818,
                "protein": 0.21279418527384816
            },
            "requested_food": "cucumber",
            "requested_volume": 0.28601368988420456,
            "calculated_weight": 34.32164278610455
        }
    ]
}
```


Fig. 5b. Macronutrient API Response Body

The system successfully interfaces with the macronutrient database through API requests by sending a POST request to the Macronutrient's API Unique Resource Identifier (URI), demonstrating the complete pipeline from volume estimation to nutritional analysis.

A) Volume Measurement Comparison:

Ground Truth Measurements:

- Rice: 0.5g
- Cucumber: 0.3 cups
- Hardboiled-Egg: 0.25 cups

System Estimations:

- Rice: 0.5381 cups, ±0.241 cups (-0.0381 underestimation)
- Cucumber: 0.2860 cups, ± 0.114 cups (-0.014 cup underestimation)
- Hardboiled-Egg: 0.3604 cups, ± 0.121 cups (-0.1104 cups underestimation)

B) API Request Structure:

The system generates a structured JSON request containing:

- Food item identification ("food_name")
- Calculated volume measurements by the Pinhole Camera Model Technique system

C) API Response Analysis:

The response provides comprehensive nutritional information:

For Rice (0.5381 cups):

- Base serving: 85.03g (0.5381 cups)
- Macronutrient content:
    - Calories: 109.68 kcal
    - Protein: 2.27g
    - Fat: 0.24g
    - Carbohydrates: 23.81g

For Cucumber (0.2860 cups):

- Base serving: 34.32g (0.2860 cups)
- Macronutrient content:
    - Calories: 5.49 kcal

- Protein: 0.21g
- Fat: 0.06g
- Carbohydrates: 1.01g

For Hardboiled Egg (0.3604 cups):

- Base serving: 48.66g (0.3604 cups)
- Macronutrient content:
  - Calories: 69.58 kcal
  - Protein: 6.03g
  - Fat: 4.85g
  - Carbohydrates: 0.47g

Weighted Normalized Error:

ITEM	GROUND TRUTH	ESTIMATION	MEAN ERROR	UNCERTAINTY	RELATIVE ERROR
RICE	0.5 cups	0.5381 cups	0.0381 cups	± 0.241cups	7.08%
CUCUMBER	0.3 cups	0.286 cups	-0.014 cups	± 0.114cups	4.9%
HARDBOILED EGG	0.25 cups	0.3604 cups	-0.1104 cups	± 0.121cups	30.60%

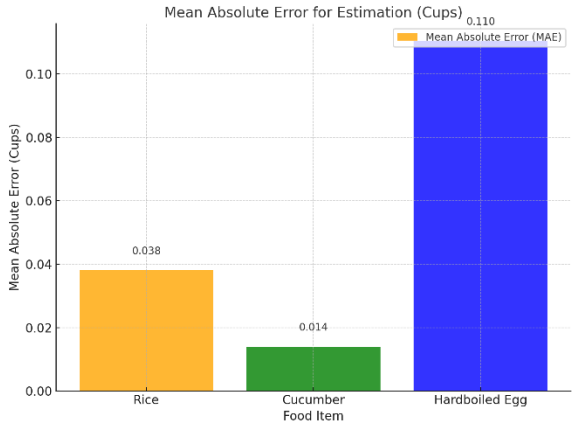


Fig. 6. Mean Error for Estimation (cups)

The volume estimation deviations, particularly for hardboiled eggs as shown in Figure 6, highlight the challenges in accurately estimating volumes of foods with irregular shapes and varying textures. While the rice and cucumber estimation are relatively accurate, further improvements are needed to address the limitations in edge detection and handling different food geometries.

This real-time nutritional analysis demonstrates the system's capability to:

- Convert volume measurements to standard serving sizes
- Retrieve and scale macronutrient information
- Provide comprehensive nutritional data for multiple food items simultaneously
- Identify areas for measurement accuracy improvement

Considering the observed errors, Raizel Jae B. de Roxas, RND, a registered nutritionist-dietitian, reviewed the system's results and concluded that, while the estimations have some deviations, they are within an acceptable range given the context of this being an estimation tool. The slight underestimations, especially for irregularly shaped foods like cucumber and hardboiled eggs, are seen as tolerable for practical use, considering the trade-off between accuracy and convenience in real-time food analysis

Chapter V Conclusion and Future Work

This paper presented M-PCAM, a hybrid model combining Mask R-CNN with the pinhole camera model for food volume estimation and macronutrient estimation. Our work systematically addressed each research objective, demonstrating both achievements and areas for improvement.

Key Achievements per Objective:

1) Low-Resolution Depth Processing:

- Successfully implemented processing pipeline for 160x90 depth images

- Achieved effective downscaling of segmentation masks to 160x90 resolution
- Demonstrated viable food segmentation and classification

2) Depth Image Enhancement:

- Implemented successful normalization and denoising techniques
- Developed robust downscaling methodology that improved image quality without compromising computational efficiency.
- Achieved clear food item separation in processed images

3) Volume Estimation Integration:

- Successfully integrated Mask R-CNN segmentation with pinhole camera model technique for volume estimation
- Achieved measurable, albeit slightly varying, volume estimation for various objects:
  - Rice: The system slightly overestimated the volume by 0.0381 cups with a ± 0.241 cups of uncertainty.
  - Cucumber: The system slightly underestimated the volume by 0.0140 cups with a ± 0.114 cups of uncertainty.
  - Hardboiled Egg: The system significantly overestimated the volume by 0.1104 cups with a ± 0.121 cups of uncertainty.

4) Accuracy Validation:

- Established ground truth measurements
- Quantified system accuracy through comparative analysis
- Identified that irregular food shapes (e.g., hardboiled eggs) introduce challenges to accuracy, specifically in edge detection and depth map quality.

5) Environmental Performance:

- Established controlled testing environment
- Identified key environmental factors affecting accuracy
- Documented system limitations under various conditions

Our key contributions include:

1) Development of a complete processing pipeline that effectively handles low-resolution (160x90) depth images from Unity's ARFoundation framework, successfully downscaled segmentation masks 640x480 for practical use [11].
2) Innovative Use of ARCore's SLAM Algorithm: By repurposing ARCore's spatial awareness capabilities, we tailored it for detailed object feature extraction, facilitating accurate food volume estimation. [11]
3) Implementation of a robust segmentation and depth processing system achieving clear food item separation and reliable depth mapping
4) Achievement of volume estimation with varying results
5) Creation of an end-to-end solution integrating volume estimation with real-time macronutrient calculation through API integration

Despite some inconsistencies, our results are acceptable compared to prior studies. For example:

- Dong-seok Lee et al [39] (using Intel's RealSense D415 RGB-D camera) reported an error rate of 2.2%, benefitting from well-separated food items on concave surfaces for easier

volume estimation and their hardware capabilities with their dedicated depth camera.

- Yue et al. [5] (using a digital camera) noted error rates ranging from 11% overestimation to -3% underestimation. All tested on optimal environments and simple objects (food toys), focusing on thickness and length rather than true volume.

Additionally, a licensed nutritionist reviewed our system's volume and macronutrient estimation results, deeming them acceptable for dietary monitoring purposes, even with minor inconsistencies.

Given these benchmarks, our system's accuracy, especially considering its mobile-based nature and low-resolution inputs, with confirmation with a nutritionist, our system demonstrates significant promise and practical utility.

A. Technical Improvements

- Development of texture-specific volume estimation algorithms to address varying accuracy across food types
- Implementation of adaptive depth thresholding for different food surfaces
- Enhancement of edge detection for more accurate portion boundary definition
- Improvement of depth map resolution through advanced upscaling techniques, and deep learning models

B. Application Enhancements

- Development of food-specific calibration factors to compensate for systematic estimation errors
- Integration of real-time accuracy feedback mechanisms
- Implementation of portion size guidance features
- Enhancement of the macronutrient API to include confidence scores for estimations

C. Validation and Testing

- Expansion of test cases across a broader range of food types and textures
- Systematic analysis of estimation accuracy patterns
- Evaluating system performance under varied environmental conditions (e.g., lighting, positioning, background noise) to determine how adaptable the system is in real-world scenarios.

D. Future development will prioritize:

1) We will focus on reducing measurement deviations through more sophisticated calibration methods, which could include manual and automatic calibration routines that adapt to different foods and user environments
2) Implementation of a less advanced depth camera for a more consistent captured depth map.
3) Implementing compensation algorithms for different food textures
4) Enhancing real-time processing capabilities
5) Improving the reliability of nutrient calculations based on volume estimates

These improvements aim to create a more accurate, user-friendly system for dietary monitoring and nutritional assessment, making food volume estimation more accessible for everyday use.

References

[1] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 2961-2969.
[2] K. He et al., "Mask R-CNN for object detection and instance segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 2, pp. 386-397, 2020.
[3] Y. Dai, S. Park and K. Lee, "Utilizing Mask R-CNN for Solid-Volume Food Instance Segmentation and Calorie Estimation," Applied Sciences, vol. 12, no. 21, p. 10938, 2022.
[4] D. Lee and S. Kwon, "Amount Estimation Method for Food Intake Based on Color and Depth Images through Deep Learning," Sensors, vol. 24, no. 7, p. 2044, 2024.
[5] Y. Yue et al., "Measurement of food volume based on single 2-D image without conventional camera calibration," Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2012, pp. 2166-2169.
[6] S. Leorna, T. Brinkman, and T. Fullman, "Estimating Animal Size or Distance in Camera Trap Images: Photogrammetry Using the Pinhole Camera Model," Methods in Ecology and Evolution, vol. 13, no. 8, pp. 1707-1718, 2022.
[7] E.-S. Kim, Y. Oh, and G. W. Yun, "Sociotechnical challenges to the technological accuracy of Computer Vision: The New Materialism Perspective," Technology in Society, vol. 75, 2024.
[8] V. Kumar et al., "Role of macronutrient in health," World Journal of Pharmaceutical Research, vol. 6, no. 3, pp. 373-381, 2017.

[9] M. C. Araujo et al., "Macronutrient consumption and inadequate micronutrient intake in adults," Revista de Saúde Pública, vol. 47, pp. 177s-189s, 2013.
[10] Y. He et al., "Food image analysis: Segmentation, identification and weight estimation," IEEE International Conference on Multimedia and Expo, 2013, pp. 1-6.
[11] Google Developers, "ARCore Extensions for AR Foundation" 2024. [Online]. Available: https://developers.google.com/ar/reference/unity-arf depth. [Accessed: May 24, 2024].
[12] "What are convolutional neural networks?" IBM, https://www.ibm.com/topics/convolutional-neural-networks [Accessed: May 24, 2024].
[13] S. Khan et al., "Efficient leukocytes detection and classification in microscopic blood images using convolutional neural network coupled with a dual attention network," Computers in Biology and Medicine, vol. 24, 2024.
[14] T. L. Subaran, T. Semiawan, and N. Syakrani, "Mask R-CNN and GrabCut Algorithm for an Image-based Calorie Estimation System", J. Inf. Syst. Eng. Bus. Intell., vol. 8, no. 1, pp. 1–10, Apr. 2022.
[15] P. Poply and A. A. Jothi, "An Instance Segmentation approach to Food Calorie Estimation using Mask R-CNN," Proceedings of the 2020 3rd International Conference on Signal Processing and Machine Learning, 2020.
[16] R. K. Megalingam et al., "Monocular distance estimation using pinhole camera approximation to avoid vehicle crash and back-over accidents," Proceedings of the 10th International Conference on Intelligent Systems and Control (ISCO), pp. 1–5, 2016.
[17] K. Kitamura et al., "Food log by analyzing food images," ACM Multimedia, 2008.
[18] G. Elmasry and S. Nakauchi, "Image analysis operations applied to hyperspectral images for non-invasive sensing of food quality – A comprehensive review," Biosystems Engineering, vol. 142, pp. 53-82, 2016.
[19] H. K. Seligman et al., "Assessing and Monitoring Nutrition Security to Promote Healthy Dietary Intake and Outcomes in the United States," Annual Review of Nutrition, 2023.
[20] Y. Wang et al., "Efficient superpixel based segmentation for food image analysis," 2016 IEEE International Conference on

Image Processing (ICIP), pp. 2544-2548, 2016.
[21] S. Sood and H. Singh, "Computer Vision and Machine Learning based approaches for Food Security: A Review," Multimedia Tools and Applications, vol. 80, pp. 27973–27999, 2021.
[22] D. Saha and A. Manickavasagan, "Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review," Current Research in Food Science, vol. 4, pp. 28-44, 2021.
[23] C. Kiourt et al., "Deep learning approaches in food recognition," ArXiv abs/2004.03357, 2020.
[24] Y. Wang et al., "Context based image analysis with application in dietary assessment and evaluation," Multimedia Tools and Applications, vol. 77, pp. 19769-19794, 2017.
[25] G. A. Tahir and C. K. Loo, "A Comprehensive Survey of Image-Based Food Recognition and Volume Estimation Methods for Dietary Assessment," Healthcare, vol. 9, no. 12, p. 1676, 2021.
[26] K. Sheng et al., "Learning to assess visual aesthetics of food images," Computational Visual Media, vol. 7, pp. 139-152, 2021.
[27] T. Le, "Mask R-CNN with data augmentation for food detection and recognition," TechRxiv, 2020.
[28] J. H. Shu et al., "An Improved Mask R-CNN Model for Multiorgan Segmentation," Mathematical Problems in Engineering, vol. 2020, Article ID 8351725, 2020.
[29] Y. Dai et al., "Mask R-CNN-based Cat Class Recognition and Segmentation," Journal of Physics: Conference Series, vol. 1966, no. 1, p. 012010, 2021.
[30] S. Fang et al., "Improved Mask R-CNN Multi-Target Detection and Segmentation for Autonomous Driving in Complex Scenes," Sensors, vol. 23, p. 3853, 2023.
[31] R. Rajarajeswari and V. Sankaradass, "Multi-Object Recognition and Segmentation using Enhanced Mask R-CNN for Intricate Image Scenes," 2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), pp. 1-6, 2023.
[32] S. Kaushal et al., "Computer vision and deep learning-based approaches for detection of food nutrients/nutrition: New insights and advances," Trends in Food Science & Technology, vol. 146, 2024.
[33] W. Shao et al., "Vision-based food nutrition estimation via RGB-D fusion network," Food Chemistry, vol. 424, 2023.
[34] W. Wang et al., "A review on vision-based analysis for automatic dietary assessment," Trends in Food Science & Technology, vol. 122, pp. 223-237, 2022.
[35] I. Angeles-Agdeppa and M. R. S. Custodio, "Food sources and nutrient intakes of Filipino working adults," Nutrients, vol. 12, no. 4, p. 1009, 2020.
[36] B. J. Venn, "Macronutrients and human health for the 21st century," Nutrients, vol. 12, no. 8, p. 2363, 2020.

[37] Google Developers, "ARCore Supported Devices" 2024, [Online], Available: https://developers.google.com/ar/devices, [Accessed: Dec 13, 2024].
[38] CouchBase, https://www.couchbase.com/ 2024, [Online], Available: https://www.couchbase.com/, [Accessed: Dec 13, 2024].
[39]A. Jakl, "Basics of AR: SLAM – Simultaneous Localization and Mapping," andreasjakl.com, Aug. 14, 2018. [Online]. Available: https://andreasjakl.com