iLog 2.1: Calorie Estimation for Pizza via Mask R-CNN and Federated Learning

Indira Devi Siripurapu
Dept. of Computer Science and Engineering
University of North Texas, USA
Email: IndiraDeviSiripurapu@my.unt.edu

Laavanya Rachakonda
Dept. of Computer Science
University of North Carolina Wilmington, USA
Email: rachakondal@uncw.edu

Saraju P. Mohanty
Dept. of Computer Science and Engineering
University of North Texas, USA
Email: saraju.mohanty@unt.edu

Elias Kougianos
Dept. of Electrical Engineering
University of North Texas, USA
Email: elias.kougianos@unt.edu

***Abstract*—It is very essential to calculate nutrients consumed for a lifestyle with health and wellness. Many existing systems and mobile applications depend on manual input by user or calculate only the dish level averages without considering multiple ingredient variations. iLog 2.1 is a lightweight and privacy preserving work which estimates the calories, protein, and carbohydrates from just a single 2D image uploaded by the user. By using a Mask R-CNN model which was trained on a custom dataset, each ingredient is detected and it's pixel area is calculated and converted into volume using the preset height values. The nutrient information has been derived from the USDA and the Food-a-pedia datasets. The federated learning aspect ensures the privacy of the user, by sharing only the encrypted model updates. The system achieves 93% detection accuracy and 85% nutrient estimation precision, which shows efficient and secure ingredient-level analysis.**

I. Introduction

Food analysis with artificial intelligence (AI) is vital for healthcare and diet management. Nutrient informed food consumption aids in dietary balance and disease prevention, however, most calorie estimation approaches continue to rely on manual input or dish-level averages [1]. It's harder to gauge calories intuitively when eating complex foods like pizza, which can include different crusts, toppings and sauces[2], [3].

The AI-powered nutrition tools bridge the gap between the dietary rules and daily choices, helping people manage balanced diets with less of the guesswork and more confidence in the nutritional value of their meals.[4]

II. Proposed Framework of iLog 2.1 System

The proposed iLog 2.1 framework focuses on accurate, interpretable, and privacy-preserving macronutrient estimation from a single 2D food image. The complete pipeline of this process is illustrated in Figure 1. Unlike earlier works, this work concentrates on one complex food pizza to capture ingredient-level accuracy. The main novelties of this work are as follows:

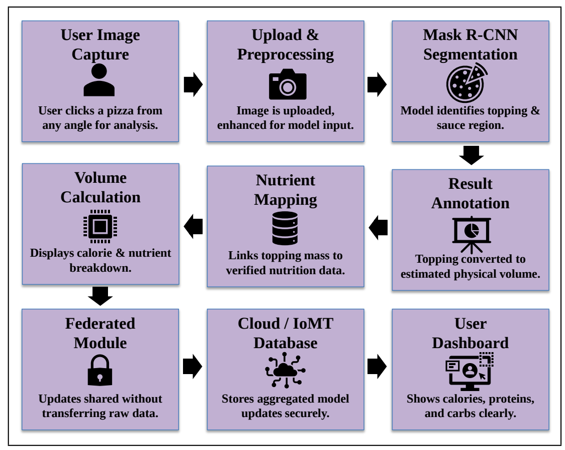

Fig. 1: Proposed iLog 2.1 pipeline

- **Per-topping estimation:** Each topping and sauce is detected individually, allowing precise calorie, protein, and carbohydrate calculation.
- **Geometric modeling:** Pixel areas are converted into volume using a fixed pizza height, elimates the need of depth sensors or multi-view data.
- **Verified nutrition mapping:** Detected toppings are matched to reliable food databases for accurate macronutrient values.
- **Lightweight and Private:** Lightweight design supports real-time use on mobile and Only the encrypted weight updates are shared to preserve the user data.

III. Related Work

Table I outlines major research in food vision and nutrition estimation. Earlier work/systems have mainly performed on the image-level calorie prediction or broad food classification and not the calories or various types of ingredients in the food. The other frameworks got high recognition accuracy but depened on fixed calorie assumptions for the whole dishes.

TABLE I: Comparison of Major Works in Food Vision and Nutrition Estimation

Work	Method	Vol. Est.	Edge/FL	Limitation
iLog [5]	Faster R-CNN + calorie mapping	No	Edge compatible	Dish-level; no ingredient detail
iLog 2.0 [6]	Detector + Federated prototype	Yes	Edge compatible	Not accurate
iLog 3.0 [7]	Mask R-CNN + depth estimation	Yes	No	Accurate but heavy; not entirely edge-friendly
Multi-Task Image-Based Dietary Assessment [8]	Multi-task deep learning (classification + regression)	Yes	No	No federated learning, only single centralized training

The depth based and the multi-task methods got improved in precision, they needed higher computational load and lacked edge compatability. Only a few number of works worked in privacy preserving learning, and they focused primarily on centralized training rather than distributed framework. iLog 2.1 addresses this gap by focusing on a type of mixed meal and estimating all the macronutrients at the ingredient level using Mask R-CNN segmentation. This shows efficient, interpretable, and secure ingredient-level nutrition analysis on edge compatible systems.

IV. Dataset, Methodology, and Implementation

Each ingredient's volume is calculated from the pixel area and the preset height, and the nutrients are mapped from combined dataset from the sources USDA FoodData Central and Food-a-pedia [9], [10]. The iLog 2.1model is trained with a Federated Learning framework to preserve user privacy, and continuously improving with distributed learning [11].

A. Dataset Preparation

A custom pizza dataset consisting of 1,107 RGB images was collected with various kinds of lighting setup and angles. Each image was annotated with polygon masks for fifty topping and sauce labels. The dataset was then divided into 60% for training, 20% for validation, and 20% for testing. Each topping label was linked to combined and verified nutritional values from USDA FoodData Central [9] and Food-a-pedia[10], providing direct mapping between detected toppings and macronutrients.

B. Model Architecture

A lightweight Mask R-CNN model [12] which has been implemented on Detectron2, has been trained to identify and segment each of the topping. The input image is preprocessed through the resizing step and the color normalization step before the inference. The segmented mask of each topping provides a pixel area A_i, which is converted into physical area using a scale factor s^* (cm/pixel). Using an average pizza height $h_i = 2.03$ cm based on [13] and [14], the volume V_i of each topping is obtained as shown in Equation 1 and each topping's mass is then derived using its density ρ_i, as shown in Equation 2:

$$V_i = A_i \times h_i \tag{1}$$

$$m_i = V_i \times \rho_i \tag{2}$$

Calories and macronutrients are computed from per-gram nutritional factors E_i (kcal/g) and F_i (nutrients/g), as shown in Equation 3 and the total nutritional values for the pizza are given, as shown in Equation 4::

$$C_i = m_i \times E_i, \quad N_i = m_i \times F_i \tag{3}$$

$$C_{total} = \sum_i C_i, \quad N_{total} = \sum_i N_i \tag{4}$$

C. Federated Learning Integration

To preserve the user privacy, the model/system uses a federated learning framework, where the local models are always trained independently on the user devices. When only the encrypted weight updates are sent to a central aggregator, which then performs the weighted averaging to update the global model. The aggregation follows, as shown in equation 5:

$$W_{t+1} = \sum_{k=1}^{K} \frac{n_k}{\sum_j n_j} W_t^k \tag{5}$$

where W_t^k are the local model weights generated and n_k is the data sampled at each client. This distributed learning method makes sure that no image data is ever being shared, making it as a privacy preserving model update.

D. Training and Evaluation

The model has been trained using the stochastic gradient descent with multiple epochs, and the standard image augmentations as flipping, rotation, and brightness adjustments on the images, were applied to improve its generalizability. The segmentation performance was assessed by calculating the Mean Average Precision (mAP) and the Intersection over Union (IoU) metrics, which provides highly robust measures of accuracy. To evaluate the nutritional estimation, the relative percentage error of the model was computed for the predicted versus reference values of calories, protein, and carbohydrates, making sure that nutrient calculations reflected real dietary intake.

V. Results and Discussion

The model has accurately detected and separated the toppings and maintained consistency with the lighting and image variations. Figure 2a and 2b, show the annotated output, where each topping is segmented and labeled with its macronutrient values.

The model shows a very stable convergence, as observed in Figure 3, with the consistent reduction of classification and

(a) Detected topping segmentation using Mask R-CNN

Calories

Item	Calories (kcal)
Pizza with Cheese & Sauce	2743.31
Pepperoni	503.31
Ham	7.53
TOTAL	**3254.15**

Protein

Item	Protein (g)
Pizza with Cheese & Sauce	131.92
Pepperoni	23.43
Ham	1.04
TOTAL	**156.39**

Carbs

Item	Carbs (g)
Pizza with Cheese & Sauce	408.26
Pepperoni	1.22
Ham	0.08
TOTAL	**409.56**

(b) Estimated per-topping nutrient breakdown for calories, protein, and carbohydrates

Fig. 2: Example visualization of iLog 2.1 results showing (a) topping-wise segmentation and (b) corresponding macronutrient estimation.

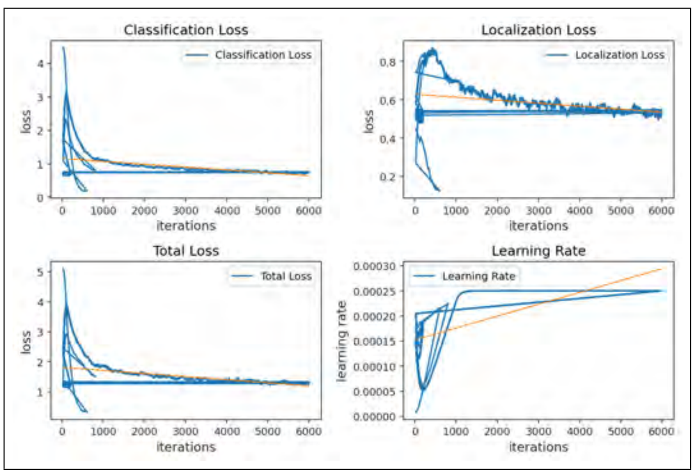

Fig. 3: Training convergence trends showing classification, localization, total loss, and learning rate adjustment.

mask losses. The entire iLog 2.1 framework was evaluated on the custom pizza dataset to assess both the segmentation and the macronutrient estimation performance.

Table II summarizes the key performance metrics, indicating precise topping segmentation. and demonstrating that the pixel-to-volume approach with verified nutritional mapping can yield realistic predictions. The low inference time further validates its suitability for real-time edge deployment.

TABLE II: Performance Summary of the iLog 2.1 Framework

Metric	Value	Remarks
Detection mAP	93.0	Accurate topping segmentation
Nutrient Accuracy	85.0	Calories, protein, and carbohydrates
Mean IoU	0.88	Precise topping mask overlap

VI. Conclusion and Scope

This paper has presented iLog 2.1, a lightweight and a privacy-preserving framework for per-topping macronutrient estimation from a single 2D pizza image. By combining Mask R-CNN segmentation, geometric volume modeling, and Federated Learning, the system gives accurate, interpretable, and secure nutrition analysis without requiring depth data or manual input. The achieved accuracy and efficiency confirm it's potential for edge deployment. Future work will extend this framework to multi-food dishes and include micronutrient estimation for comprehensive dietary assessment.

References

[1] A. Tal, Y. Gvili, and M. Amar, "Visual size matters: The effect of product depiction size on calorie estimates," *International Journal of Environmental Research and Public Health*, vol. 18, no. 23, p. 12392, Nov 2021.

[2] L. Powell, B. Nguyen, and W. Dietz, "Energy and nutrient intake from pizza in the united states," *Pediatrics*, vol. 135, no. 2, pp. 322–330, 2015.

[3] E. Combet, A. Jarlot, K. Aidoo, and M. Lean, "Development of a nutritionally balanced pizza as a functional meal designed to meet published dietary guidelines," *Public Health Nutr.*, vol. 17, no. 11, pp. 2577–2586, 2014.

[4] K. Agrawal, P. Goktas, N. Kumar, and M.-F. Leung, "Artificial intelligence in personalized nutrition and food manufacturing: a comprehensive review of methods, applications, and future directions," *Frontiers in Nutrition*, vol. Volume 12 - 2025, 2025. [Online]. Available: https://www.frontiersin.org/journals/nutrition/articles/10.3389/fnut.2025.1636980

[5] L. Rachakonda, S. P. Mohanty, and E. Kougianos, "ilog: An intelligent device for automatic food intake monitoring and stress detection in the iomt," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 2, pp. 115–124, 2020.

[6] A. Mitra, S. Goel, S. P. Mohanty, E. Kougianos, and L. Rachakonda, "ilog 2.0: A novel method for food nutritional value automatic quantification in smart healthcare," in *Proc. IEEE International Symposium on Smart Electronic Systems (iSES)*. IEEE, 2022, pp. 683–688.

[7] I. D. Siripurapu, A. Mitra, S. P. Mohanty, and E. Kougianos, "ilog 3.0: Estimating food volume from 2d images using mask r-cnn and monocular depth estimation," in *Proc. IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE, 2025, pp. 1–6.

[8] J. He, Z. Shao, J. Wright, D. Kerr, C. Boushey, and F. Zhu, "Multi-task image-based dietary assessment for food recognition and portion size estimation," in *2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2020, pp. 49–54.

[9] U.S. Department of Agriculture, Agricultural Research Service, "Food-data central," https://fdc.nal.usda.gov/, 2022, accessed on July 20, 2022.

[10] U.S. Department of Agriculture, "Data.gov: Food-a-pedia," https://catalog.data.gov/dataset/food-a-pedia, 2022, accessed on July 20, 2022.

[11] A. Nakai-Kasai and T. Wadayama, "Deep unfolding-based weighted averaging for federated learning in heterogeneous environments," 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:261241074

[12] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386–397, 2020.

[13] A. Falciano, M. Moresi, and P. Masi, "Phenomenology of neapolitan pizza baking in a traditional wood-fired oven," *Foods*, vol. 12, no. 4, p. 890, 2023. [Online]. Available: https://www.mdpi.com/2304-8158/12/4/890

[14] C. Dumas and G. S. Mittal, "Heat and mass transfer properties of pizza during baking," *International Journal of Food Properties*, vol. 5, no. 1, pp. 161–177, 2002. [Online]. Available: https://doi.org/10.1081/JFP-120015599