Deep Learning-Based Food Identification and Calorie Estimation System

Jung-Tang Huang
Taipei Tech, 1, Sec 3, Chung Hsiao E, Road, Taipei 106,
UTL Lab
Taiwan, Taipei
jthuang@ntut.edu.tw

Chen-Hao Wang
Taipei Tech, 1, Sec 3, Chung Hsiao E, Road, Taipei 106,
UTL Lab
Taiwan, Taipei
kevin8901300179@gmail.com

***Abstract*—In recent years, the importance of home care and diet have increased due to an aging population and increasing care needs. This study proposes a solution to take a photo every time while the user is eating, record the photo and performing identification and calorie estimation to understand the diet throughout the day. This paper uses a depth camera for image recognition and calorie estimation of food, and uses more than 400 food photos as a database to train a YOLOv5 neural network to recognize food on a table. The depth camera is then used to estimate how many calories the user has consumed. Our experimental results show that a food recognition rate of more than 90% can be obtained in most cases, while the calorie estimation error is less than 10%. Then, we use Google's Firebase service to upload and store a large amount of data and build a complete Internet of Things (IoT) care system with these large amounts of dietary data. This contributes to a complete elderly home care.**

Keywords- Image recognition; Volume estimate; Calorie estimates; Depth camera; Deep learning.

I. INTRODUCTION

In recent years, sarcopenia and obesity in the elderly have been global health problems with broad economic and social implications. Nutritional monitoring systems are essential for understanding and addressing unbalanced eating habits. Therefore, in order to maintain a healthy weight in a normal person with a healthy diet, daily food intake must be measured. Daily food intake and nutrient distribution are particularly important, which requires patients to calculate and record calories from food every day. In most cases, patients also have a lot of trouble estimating food intake due to self-denial issues, lack of nutritional information, and the time-consuming process of manually recording this information. The most common way to measure nutrients in food today is to directly use the weighing method. Whether it is taking out the dishes from the plate after cooking the food or measuring the food before it is cooked, it takes a lot of time. In addition, it is a big trouble to measure all the data and record it using paper or electronic products, so a system developed in this paper covers calorie estimation and cloud service system.

The rest of this paper is organized as follows: Section 2 gives the literature review. Section 3 provides an overview of data-related work. Section 4 describes calorie content estimates. Section 5 discusses the results. Section 6 presents the conclusions and suggests possible future directions.

II. RELATED WORKS

A. Food identification related literature

Recognizing food in images is a difficult object detection task. There are many machine learning methods to detect objects. Some researchers use Histogram of Oriented Gradients (HOG) features [1] and Scale-Invariant Feature Transform (SIFT). Another approach that has been increasingly used in recent years is deep learning neural networks, especially Convolutional Neural Network (CNNs). CNNs work very well and are probably the most widely used object detection method in the world. Some progress has been made in this area, and more object detection techniques have been designed based on CNNs. Some well-known object detection CNN designs are Mask R-CNN (Mask Region Convolution Neural Network) [2], Faster RCNN [3] and YOLO [4]. Most of the previous studies on image-based food calorie estimation have used detection classification to obtain food category labels. Deep learning neural networks have recently gained considerable application in object detection. Wearable context-aware food recognition for calorie monitoring [5], and segmentation and recognition of multi-food meal images for carbohydrate counting [6], both employ food segmentation-based approaches to address issues related to diet assessment. In [7], T. Ege and K. Yanai developed an object detection model using Faster R-CNN. The authors also propose a calorie estimation procedure in a multi-task CNN [8] to estimate food categories and calories simultaneously. Their model achieved high average accuracy detection (90.7%), but the paper did not use a weight estimation procedure, so the estimated calories for each food were fixed.

B. Food estimation literature

In [9], the authors use Faster R-CNN for object detection by using top and side views of food. They also used the GrabCut algorithm [10] to obtain the contours of each food. Finally, food size estimation and calorie estimation use estimation formulas done using volume. The only downside to this method is that it only works with single-item images of food, such as fruits, vegetables, and so on. In [11], the authors use CNN-based multi-task learning for dietary assessment of multiple food dishes. They achieved high accuracy in food segmentation and volume estimation. The authors of [12] completed a volume-based approach to food calorie estimation using depth cameras. Their approach

 ISBN: 978-1-61208-978-2

emphasizes the use of in-depth information to predict the volume estimation part of food calories. In [13], K. Ruan and L. Shao used Support Vector Machines (SVM) and deep learning algorithms for food classification. For the calorie estimation section, they created a calorie map for all food labels.

III. DATA ACQUISITION

When training deep models, increasing the amount of data as much as possible is desirable; the more the better. Therefore, when collecting data, the same object should be looked at from different angles, from different distances, and even illuminated with different lights. One should shoot with a body camera or scope, or try different aperture and shutter variations. In this way, the trained model can be used in various environmental fields. In this work, a food dataset named Food-101 [14] is used, which contains a large number of images. The dataset contains 101 categories and more than 10,000 photos. We choose home-cooked dishes often eaten by oriental people for discussion. This article uses more than 200 photos in 4 categories in the data set, plus the pictures we collected; the total number of photos is more than 400 photos, and the categories include: rice, eggs, shrimp, broccoli, etc.

IV. METHODOLOGY

A. YOLO

The full name of YOLO is 'you only look once', which means that you only need to browse once to identify the category and location of the object in the picture. Because it only needs to be seen once, YOLO is called the Region-free method. Compared with the Region-based method, YOLO does not need to find the possible target, Region, in advance. That is to say, the process of a typical Region-base method is as follows: first, analyze the picture by means of computer graphics (or deep learning), find out several areas where objects may exist, cut out these areas, put into an image classifier, and classified by the classifier.

B. Pixel and Length Mathematical Model

Before using the depth camera to calculate the real volume, one must first understand the real length and width of the object in the depth image, so there is a need to calculate the length of each pixel in the depth image. The relationship equation between the number of pixels of the object and the length and width of the object needs to be obtained through the following methods:

(1) The depth camera shoots an object with exact length.
(2) The number of pixels corresponding to the object in the image when the depth camera object is taken from far to near.
(3) After obtaining the corresponding pixel amounts at different depths, an equation is obtained.

The relationship between pixel and length obtained in the experiment is recorded in Table 1. We tested seven different depth distances, which corresponded to seven different pixel amounts.

TABLE I. PIXEL AND LENGTH RELATIONSHIP

Depth (cm)	Number of Pixels (Pixels)	The length corresponding to the pixel (cm/pixel)
92	101	0.1509
90	103	0.1480
87	107	0.1425
93	112	0.1361
74	125	0.1220
68	137	0.1110
43	219	0.0696

The relationship between pixels and length can be deduced from the linear equation (1), from which we can know the length of each pixel of the camera at any depth value:

$$Lenght = 0.0016618 * Depth + 0.0017478 \quad (1)$$

The actual area of the pixel is to square the linear equation (1), and the equation is as follows (2):

$$Area = (0.0016618 * Depth + 0.0017478)^2 \quad (2)$$

C. Volume estimation

Calculate the depth difference of each pixel coordinate, and express the depth difference as $\Delta\mathrm{Depth}$. Mark the depth value before the meal as Depth_0, and mark the depth value after the meal as Depth_1, calculate the depth difference of the pixel point:

$$\Delta Depth = Depth_0 - Depth_1 \quad (3)$$

The volume of a pixel is the area of the pixel multiplied by the depth difference of the pixel, where k is the coordinate of the pixel:

$$V_k = \Delta Depth_k * Area_k \quad (4)$$

Finally, add up all the pixel volumes to get the volume of the food.

D. Weight Estimate and Calorie Estimation

When the volume of the food is calculated, the next step is to convert the volume to weight and the method we use is to use the density formula for conversion, where m is the final weight of the food, v is the estimated food volume, ρ is the food density:

$$m_{food} = \rho_{food} * v_{food} \quad (6)$$

Finally, convert the food calorie dataset corresponding to the calculated food weight and food category into calories, and three major nutrients.

E. Cloud Service

Cloud Storage for Firebase stores files in Google Cloud Storage buckets, thus, they can be accessed through Firebase and Google Cloud. Flexibility through the Firebase SDK for Cloud Storage uploads and downloads files from mobile

 ISBN: 978-1-61208-978-2

clients. Additionally, it is possible to use the Google Cloud Storage API Server-side processing, such as image filtering or video transcoding. Cloud storage scales automatically, which means no need to migrate to any other provider.

We collate the results of identification and calorie estimates and upload them to Firebase, which can increase the amount of data. In the future, due to the increase in the amount of data, the number of photos during training can be increased, the recognition rate can be improved, and the number of categories can even be increased. This will help with the user's eating habits and making dietary recommendations for the user.

F. Hardware

1) Experimental Environment

In the experimental environment part, we use Intel D435i depth camera and Raspberry Pi 4B. The camera shoots from top to bottom to obtain a deep image of the plate, and the distance between the camera and the platform is 60 cm.

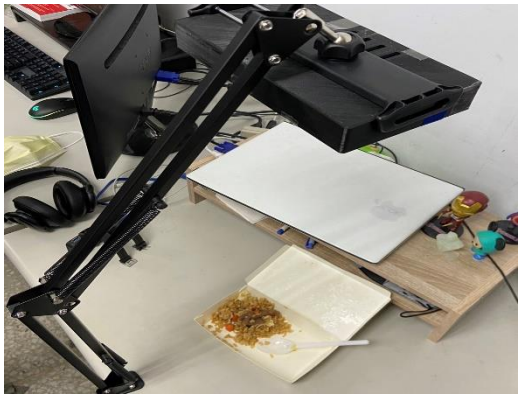

Figure 1. System experimental environment

2) Raspberry Pi 4B

In recent years, the topic of IoT has been discussed vigorously, and many developers have used Raspberry Pi to develop various IoT devices. In this study, Raspberry Pi 4B was used as the development board of this research as the router of the system, the development program and the calculation of calorie estimation program and the center of uploading the values of various sensors to the cloud.

3) Intel D435i

The Intel RealSense Depth Camera D435i incorporates an Inertial Measurement Unit (IMU) that improves depth perception in any situation where the camera moves. The D435i camera can be used in SLAM and tracking, and can also improve the efficiency of point cloud computing. It can be used on different platforms such as computers or embedded development boards. This study uses D435i as our camera, which can be used to identify and detect the depth of the food, and finally get accurate data.

V. RESULTS

A. Image recognition results

In this section, we train YOLO's food recognition model with 1000 steps. Figure 2 shows the loss function used to train the YOLO model. As for the identification results, we choose home-cooked dishes that Orientals often eat for discussion, randomly select 10 images from the dataset, and use our trained model for identification. The average accuracy rate for individual food items is over 90%. The recognition results are shown in Figure 3, and the recognition rate is above 93%.

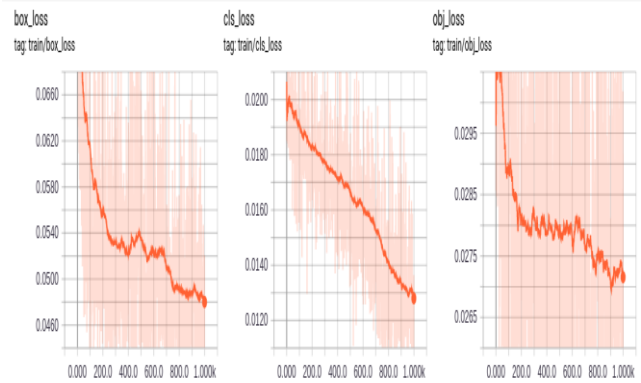

Figure 2. Loss function for YOLO model

Figure 3. Image recognition results and recognition rates

B. Calorie Estimation Results

This result is the same as the previous experiment. We randomly select 10 images from the dataset and use our model to estimate the calories of the food, record the real weight and the estimated weight, average all the recorded weights and calculate the error. Table 2 shows that the calorie estimate has an error of less than 10%, which means that the calorie error will also be less than 10 calories.

TABLE II. CALORIE ESTIMATION TEST RESULTS

Food	Data		
	Actual weight (g)	Estimated weight (g)	Error (%)
Rice	82.8	76.8	7.8
Egg	71	72.8	2.5
Broccoli	70.6	65.8	7.2
Shrimp	97.6	103.8	6.3

C. Cloud Service

In this section, we upload the results of the diet identification and calorie estimation to the cloud and classify them. The type and portion of each food are listed and clearly marked, as shown in Figure 4.

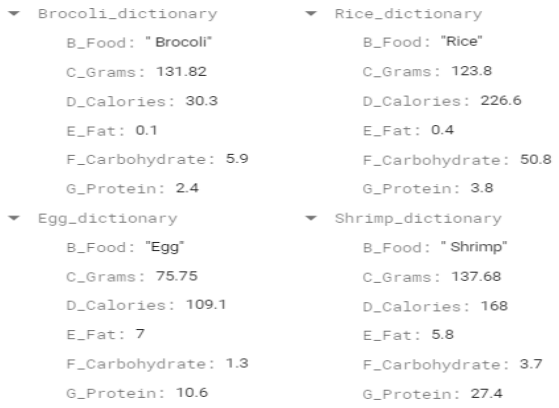

Figure 4. Firebase Profile

VI. CONCLUSION

The results show that the meal estimation method using the depth camera proposed in this paper is feasible for food evaluation, but the premise is that all the food in the plate can use the depth camera to obtain accurate depth values.

This method is compared with the traditional weighing method. It is very convenient to measure food items one by one, which is convenient for caregivers in nursing homes who spend a lot of time measuring and recording the dining conditions of the elderly.

In this article, we use YOLO to detect food calories for different foods. The recognition rate is over 90%, and the calorie error is less than 10%. However, estimating food calories is a difficult problem. Not even a perfect image processing system can predict perfectly, and the calorie content can be skewed by the amount of oil. Therefore, we first try to take the prototype food as the research direction. In the future, we will increase the diversity of oily and water-based fried vegetables and tofu, so that the dietary information can be easily obtained by using the meal evaluation method of this study in daily life.

Uploading data to Firebase in the future can increase the amount of data. In the future, due to the increase in the amount of data, the number of photos during training can be increased, the recognition rate can be improved, and the number of categories can even be increased. The user's eating habits and dietary recommendations for the user.

REFERENCES

[1] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), pp. 20–25 June 2005.

[2] H. Kaiming, G. Georgia, D. Piotr, and G. Ross, “Mask R-CNN,” Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2961–2969, 2017.

[3] R. Shaoqing, H. Kaiming, G. Ross, and S. Jian, “Faster R-CNN: towards real-time object detection with region proposal networks,” NIPS 15: Proceedings of the 28th International Conference on Neural Information Processing Systems, Vol. 1, pp. 91–99, Dec 2015.

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified real-time object detection", Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788, 2016.

[5] G. Shroff, A Smailagic, and D. P. Siewiorek, “Wearable context-aware food recognition for calorie monitoring,” 2008 12th IEEE International Symposium on Wearable Computers, pp. 1–2 Oct, 2008.

[6] M. Anthimopoulos ,J Dehais, and P Diem, “Segmentation and recognition of multi-food meal images for carbohydrate counting,” in 13th IEEE International Conference on BioInformatics and BioEngineering, pp. 10–13, Nov. 2013.

[7] T. Ege and K. Yanai “Estimating Food Calories for Multiple-Dish Food Photos,” 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR), pp. 309–314, Nov. 2017.

[8] A. H. Abdulnabi, G. Wang, J .Lu, and K. Jia, “Multi-Task CNN Model for Attribute Prediction,” in IEEE Transactions on Multimedia, vol. 17, pp. 1949–1959, Nov. 2015.

[9] Y. Liang and J. Li, “Deep Learning-Based Food Calorie Estimation Method in Dietary Assessment,” Cornell University, pp. 12, Jun 2017.

[10] C. Rother, V. Kolmogorov, and A. Blake, “GrabCut: Interactive foreground extraction using iterated graph cuts,” ACM Trans. Graph., vol. 23, pp. 309–314, 2004.

[11] Y. Lu, D. Allegra, M. Antonopoulos , and F. Stanco, “A multi-task learning approach for meal assessment,” CEA/MADiMa '18: Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management , pp. 46–52, July 2018.

[12] Y. Ando, T. Ege, J. Cho, and K. Yanai, “DepthCalorieCam: A Mobile Application for Volume-Based FoodCalorie Estimation using Depth Cameras,” MADiMa '19: Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management, pp. 76–81, Oct 2019.

[13] H. Raikwar, H. Jain, and A. Baghel, “Calorie Estimation from Fast Food Images Using Support Vector Machine,” International Journal on Future Revolution in Computer Science & Communication Engineering, vol. 4, pp. 98–102, April 2018.

[14] L. Bossard, M. Guillaumin, and L. VanGool, “Food-101–Mining Discriminative Components with Random Forests,” European Conference on Computer Vision ECCV 2014, pp. 446–461, 2014.

 ISBN: 978-1-61208-978-2