2023 Eleventh International Conference on Intelligent Computing and Information Systems (ICICIS) | 979-8-3503-2210-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICICIS58388.2023.10391113

CalorieMe: An Image-based Calorie Estimator System

Bavlly Magid
Faculty of Computer and Information Sciences
Ain Shams University
Cairo, Egypt
bafely20191700177@cis.asu.edu.eg

Mazen Mohamed
Faculty of Computer and Information Sciences
Ain Shams University
Cairo, Egypt
mazen20191700479@cis.asu.edu.eg

Mazen Khaled
Faculty of Computer and Information Sciences
Ain Shams University
Cairo, Egypt
mazen20191700478@cis.asu.edu.eg

Mohamed Ibrahim
Faculty of Computer and Information Sciences
Ain Shams University
Cairo, Egypt
mohammed20191700490@cis.asu.edu.eg

Mohamed Sabry
Faculty of Computer and Information Sciences
Ain Shams University
Cairo, Egypt
mohammed20191700542@cis.asu.edu.eg

Waleed Mohamed
Faculty of Computer and Information Sciences
Ain Shams University
Cairo, Egypt
waleed20191700747@cis.asu.edu.eg

Yomna A. Kawashti
Faculty of Computer and Information Sciences
Ain Shams University
Cairo, Egypt
yomna.ahmed@cis.asu.edu.eg

Hanan Hindy
Faculty of Computer and Information Sciences
Ain Shams University
Cairo, Egypt
hanan.hindy@cis.asu.edu.eg

Abstract—In recent years, with the growing interest in healthy eating, various food photo recognition applications for tracking meals have emerged. However, some of these applications still require human intervention for calorie estimation, such as manual input or consultation with a nutrition expert. Furthermore, even automated systems often have limitations in food category recognition, or they demand multiple viewpoints for accurate results. Meanwhile, advancements in image recognition have been substantial, thanks to the advent of Convolutional Neural Networks (CNN). CNNs have significantly improved the accuracy of various image recognition tasks, including classification and object detection. This paper presents a comprehensive solution for estimating the calories in food photos containing multiple ingredients. The proposed method employs two deep learning models: one to detect ingredients and their respective locations, and another to segment the ingredients and measure their portion sizes. Moreover, the proposed method incorporates a reference object to enhance the precision of portion size measurement. Finally, the proposed model compares the food type and portion size against a dataset of food types and their corresponding calorific values per standard serving, thus estimating the total calorie count. In this study, the two-model methodology resulted in a 7% improvement in pixel accuracy and a 23% improvement in mean Intersection Over Union (mIOU) for recognition and segmentation tasks, respectively, compared to the latest state-of-the-art approach, which employed Deeplabv3+ exclusively on the same dataset.

***Keywords—Food calorie estimation, food portion size estimation, food image recognition, food semantic segmentation, UEC-FOOD100, UEC-FoodPix Complete, Food and Their Caloric Values*.**

I. INTRODUCTION

In recent years, there has been a growing awareness in communities about the importance of health and fitness, reflecting a heightened interest in monitoring eating habits and dietary plans. The choices we make in our diets have a profound impact on our well-being. Individuals with specific dietary needs, such as diabetics and those with allergies, must diligently monitor and control their dietary behaviors. Food recognition and classification play a crucial role in helping individuals record their daily diets, as images of food convey essential information about the characteristics of meals.

Furthermore, image sensing has emerged as a practical, cost-effective tool for analyzing the appearance of food. However, recognizing food types and portions can be challenging due to the extensive variations in food shape, volume, texture, color, and composition. Additionally, various backgrounds and layouts of food items introduce additional complexity to food recognition and classification tasks. The widespread adoption of Convolutional Neural Networks (CNN) in image analysis has greatly advanced the field of food recognition and classification, offering the potential to replace manual methods for tracking nutrition plans with AI models.

These Artificial Intelligence (AI) models can estimate the volume and calories of individual ingredients based on a captured photo of a dish, using a single, streamlined tool. This approach not only saves time, money, and effort, but also enhances the accuracy and convenience of calorie estimation. CalorieMe, our proposed end-to-end solution, aims to improve the accuracy of calorie estimation while simplifying the process of monitoring calorie intake, making it user-friendly and accessible.

One common methodology for estimating food calories from a single image requires the prior registration of a reference object with a known size. The proposed system, in this context, assumes a personal belonging reference object that users would typically carry, such as a credit card. When a meal photo is taken with the reference object, the system proceeds to segment the food items and the pre-registered

979-8-3503-2210-1/23/$31.00 ©2023 IEEE

reference object. Since the actual size of the reference object is known (e.g., in the case of a credit card-sized object, with dimensions of 85.6mm (about 3.37 in) x 54mm), the system can estimate the actual size of each detected food item. By utilizing the estimated actual size and relevant equations for calculating food calories based on size, the system can accurately estimate the calorie content of the food items in the actual photo.

The subsequent sections of this paper are organized as follows: Section 2 provides an overview of related work, Section 3 details the proposed methodology, Section 4 presents the datasets used, Section 5 reports the results and discusses the evaluation of food calorie estimation, and, finally, Section 6 concludes the paper.

II. RELATED WORK

This section provides a comprehensive review of the related work in two key areas: food image recognition and segmentation (Section A) and food portion estimation (Section B).

A. Food Image Recognition and Segmentation Tasks

Object recognition falls under the computer vision domain tasks. It is concerned with training models that are capable of identifying and locating objects in a given image or video. You Only Look Once (YOLO) uses an end-to-end Neural Network (NN). This NN predicts the bounding boxes and class probabilities of an image all “at once”. YOLO in the literature is depicted as “the model that achieved state-of-the-art results, beating other real-time object detection algorithms by a large margin”.

Wen Lo *et al.* [1], published in their paper a discussion about common approaches utilized in the food image recognition task. Among these approaches, an end-to-end Image Recognition with Deep Learning approach describes the use of deep learning for food recognition. Although deep learning has shown great performance in various AI applications, it has not been researched extensively for food recognition. The authors discussed the use of CNNs for dietary monitoring and compared their performance to traditional SVM-based techniques using handcrafted features. The CNNs outperformed the traditional methods by 10%. The paper mentions the work done by Google in which deep learning methods were proposed for dietary assessment, using a pre-trained GoogLeNet model fine-tuned on the Food101 dataset. The results outperformed traditional methods based on handcrafted features and SVM classifiers by 28%. The paper also mentions other research attempts based on deep learning and concludes that deep learning methods outperform traditional ones in food recognition.

Wen Lo et al. in [1] summarize the prominent research in this domain. The analyzed studies are the ones that have been carried out on the practicality of deep learning techniques and examined several types of networks on several large publicly available datasets. The authors show that most of the paper used the datasets: UEC-FOOD100 [2], UEC-FOOD256 [3] and Food101 [4].

In regard to the results of the relevant research discussed in [1], Both Kagaya *et al.* [5] and Christ *et al.* [6] used their own dataset. Both papers reported their accuracies of 73.7% and 84.9% respectively using CNN. In [6], the authors used a Patch-wise model together with the CNN.

Qui et al. [7] used the Food-101 dataset and achieved a top accuracy of 90.4 using the PAR-NET model. This same top-accuracy was achieved on the same dataset by other researchers using the IG-CMAN model. Tan et al. [8] also trained and tested their work on the Food-101 dataset, achieving 93% using the EffecientNet.

Finally, Inception V3 was used by Hassane et al. [9]. The top accuracies on the UEC-FOOD100 dataset and the UEC-FOOD256 dataset are 81.5% and 76.2% respectively. Foresti et al. [10] were able to further improve the accuracy of the models on the same datasets using WISeR. The accuracies reached 89.6% for UEC-FOOD100 and 83.2% for UEC-FOOD256.

Takumi Ege and Keiji Yanai [11] proposed a multi-task learning approach for simultaneous estimation of dish locations and calories using the UEC-FOOD100 dataset. This method outperformed its sequential counterparts in accuracy, speed and network size. They modified the Single-Shot Detector (SSD) based on VGG16 to output food calorie values on each food bounding box. However, their method used an additional manually collected dataset pre-annotated with calories to estimate calories per serving, regardless of the amount of food present in the image. Despite this limitation, their model performed well in the food recognition task.

Additionally, Ege et al. reported the results of food calorie estimation from single-dish food photos against multiple-dish food photos. The multiple-dish precision results (85.7%) saw a significant drop against single-dish results (93.8%) which highlights the challenge of recognizing food in multiple-dish food scenario.

Kaimu Okamoto and Keiji Yanai [12] developed and published a paper where they introduced their own version of an image food segmentation dataset by extending the “UEC-FoodPix” dataset. The segmentation masks of the original “UEC-FoodPix” dataset were incomplete since they were created semi-automatically using bounding box annotations. Thus, Okamoto et al. relied on comprehensive manual annotation to complete the dataset and create finely annotated segmentation masks. This was achieved by using the web-based pixel-wise annotation tool implemented by Pongsate et al. This new “complete” version of the dataset is referred to as “UEC-FoodPix Complete”. In order to evaluate the significance of this new extended dataset, the authors used the DeepLabV3 semantic segmentation model to compare the performance of each version of the dataset. They used pixel-wise accuracy and mean Intersection Over Union (mIOU) as the evaluation metrics for this experiment. It was found that the original dataset “UECFoodPix” achieved a pixel accuracy of 56% and mean intersection over union of 0.416 while the newly proposed complete version “UEC-FoodPix Complete” outperformed the older version by achieving a 66.8% pixel accuracy and an mIOU of 0.555.

B. Food Portion Estimation Task

There were multiple approaches used in the literature to estimate the portion size ranging from approaches using depth cameras to others utilizing the usage of multiple captures of different viewpoints of the food to 3D-Model reconstruction. Wen Lo et al. [1] discussed in their paper these approaches and classified them into 5 categories which are:

1-Stereo-based approach: Stereo-based approach refers to the use of pixel correspondences between multiple viewpoints to create a 3D reconstruction of the food objects.

2-Model-based approach: The model-based approach uses pre-built shape templates/models to estimate the food volume. This is achieved through a process of model selection, scaling and rotation. This process is also referred to as Image Registration.

3-Depth camera-based approach: In addition to the standard visual information present in an RGB image, depth-based cameras are able to acquire depth information through sensors. This use of additional sensors removes the need for the use of an external reference object for volume estimation.

4-Perspective transformation approach: In this approach, perspective transformation is applied to the original image to acquire a bird's eye view. This perspective can then be used to estimate the object's size. This approach works well with objects of irregular shapes since a pre-defined shape template would not work.

5-Deep learning approach: The use of deep neural networks for volume estimation has also been explored with interest. Several methods proposed the use of deep learning models to deduce a depth map from a single-colored image. Other works focused on the use of voxel representation and point cloud completion for depth and volume estimation. Additionally, the authors in [1] also gave an overview of the advantages and limitations of each of these approaches.

In the dietary assessment domain, most of the publicly available datasets deal with the task of food classification only, without providing annotations for the food volume estimation task. Consequently, most researchers develop their own datasets to evaluate the volume estimation phase of their work. These datasets tend to be smaller in scale than the known benchmarks. Thus, it is important to take note of the datasets used in each phase when comparing the results acquired by different research papers.

Kadam et al. [13] published their work where they utilized the usage of fiducial marker (Reference object) which is an object of known and standard dimensions put in frame with the food while image is being captured (e.g., a checkerboard card/a coin/a credit card) as shown in Fig. 1, the paper proposed a coin as reference object to estimate food portion size. Among their experiments, Kadam et al. reported volume estimation results against actual volume across different food shape types. Amorphous food shape types achieved 90.46% volume estimation accuracy, convex food types achieved 90.9% accuracy, while regular (circle) and regular (square) food types showed better results with estimation accuracies of 98.5% and 98.9% respectively. These results are predictable since amorphous and convex food shapes volumes are harder to estimate as they take the shape of their container which could be much trickier to estimate.

Fig. 1. Coin as a Reference Object [13]

III. PROPOSED METHOD

This paper introduces a novel three-stage methodology for estimating the calorific value of food through image analysis. The proposed approach encompasses three key stages: food recognition and segmentation, portion size estimation, and calorie calculation.

A. *Food recognition & segmentation*

To accomplish the tasks of food recognition and segmentation, this study leverages two distinct models: YOLOv5 and DeepLabv3+. YOLOv5 is utilized for food recognition, while DeepLabv3+ is employed for food segmentation. The architecture of these two models is depicted in Fig. 2 and Fig. 3, respectively.

YOLOv5 [14] belongs to the YOLO family of computer vision models. Specifically, Ultralytics YOLOv5 represents a state-of-the-art (SOTA) model that further improves upon the success of older YOLO versions through the addition of several features to enhance performance and flexibility.

B. *Portion size estimation*

Accurate food portion size estimation within this study relies on the use of a reference object with known dimensions, such as a credit card. This approach is founded on the assumption that a food item's volume is directly related to its weight, which, in turn, is directly correlated with the size of the ingredient's surface area. The actual size of the reference object (i.e., the credit card) is used to estimate the actual size of the food objects. Since the actual size of the reference object is known, the actual size of food regions (Fr) can be obtained using the following equation:

$$F_r = S_r \times \frac{F_p}{S_p} \quad (1)$$

Where S_p refers to the number of pixels of the reference object, F_p refers to the number of pixels of the food object while S_r refers to the actual size of the reference object from the top view [16].

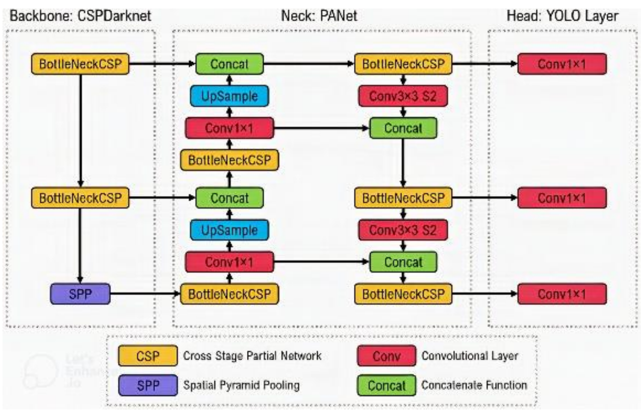

Fig. 2. YOLOv5 Model Architecture [14]

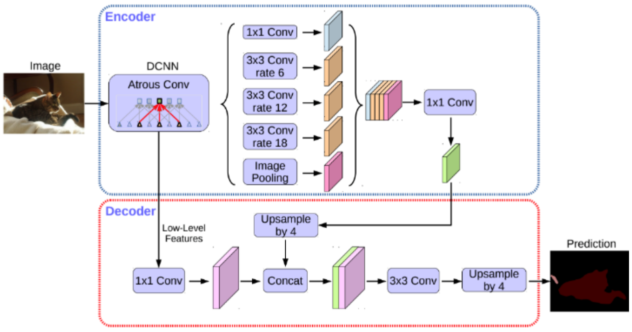

Fig. 3. DeepLabv3+ Architecture [15]

It is essential to note that the accuracy of portion size estimation is influenced by the relative size of the food item and the reference object. If the food item is significantly smaller than the reference object, the estimated weight will be lower than the actual weight. Conversely, if the reference object is not detected or is not the specified credit card, the system will respond with a "no reference object detected" message.

To obtain the pixel mask of the reference object, a contour analysis technique, as depicted in Fig. 4, is applied, to enable its extraction within this study.

C. Calorie Calculation

The core focus of this study lies in the analysis of ingredient labels and their corresponding volumes. These values are then compared against a dataset named "Food and their calories" (a modified version). This dataset comprises a CSV file containing comprehensive information.

On over 300 food items, including the total number of Calories, Fats, Proteins, Saturated Fats, Carbohydrates, and Fibers labeled for each individual food entry. This study aims to utilize this dataset to calculate calorie and macronutrient counts per volume of each ingredient.

IV. DATASETS

This paper incorporates three distinct datasets to facilitate the training and evaluation of the various models implemented within this project. Each dataset is described below, providing information regarding the data format, number of entries, and other relevant details.

A. UECFOOD-100

The dataset contains 100 kinds of food photos, including 11,561 images for training and 1,417 images for testing. Each food photo includes a bounding box, as shown in Fig. 5 indicating the location of the food item in the photo. Most of the food categories in this dataset are popular foods in Japan; hence, some categories may be less familiar to individuals from other nationalities. This dataset was developed to create a practical food recognition system primarily intended for use n Japan.

B. UECFOODPIXCOMPLETE

This dataset is a collection of food images with segmentation masks, comprising 9,000 images for training and 1,000 images for testing. Samples from the dataset are shown in Fig. 6. The segmentation masks are categorized by food type and have been manually provided. The mask images contain pixel-wise labels for 103 food classes, with labels present only in the red (R) channel.

C. Food and their calories (Extended Version)

The dataset consists of a CSV file containing more than 300 labeled foods, each paired with its corresponding calorie values. This dataset has been extended manually to align with the categories present in UEC-FOOD 100 and UECFOODPIXCOMPLETE, both of which were used in this project.

V. RESULTS & DISCUSSION

In this section, a combination of methods is explored, both individually and in isolation, with the overarching objective of achieving optimal results.

Firstly, a DeepLabV3+ model was constructed for food identification and segmentation, inspired by the approach outlined in [17]. The proposed model demonstrated promising results for this task, as shown in Table I. The reported results in the paper [17] are based on evaluations conducted using the UECFoodPixComplete dataset.

The DeepLabV3+ model was trained on UECFoodPixComplete dataset using the following hyperparameters: Image size set to 256, batch size of 16, SGD optimizer, cross-entropy loss function, learning rate of 0.001, momentum of 0.9, weight decay of 0.0005, and a ResNet101 as a backbone for 100 epochs. The model yielded a pixel accuracy of 74% and mIOU result of only 45% on the test data from UECFoodPixComplete, as shown in Table II.

This performance fell short of expectations due to the challenges posed by the dataset's heterogeneity and diverse nature, which made accurately segmenting and identifying food items a complex task. The segmentation mask struggled with certain categories, as depicted in Fig. 7 wherein the mask was struggling with the category.

Fig. 4. Contour Analysis for Reference object segmentation.

Fig. 5. UECFOOD-100 Dataset Samples [2].

Fig. 6. UECFOODPIXCOMPLETE Dataset Samples [12].

TABLE I. DEEPLABV3+ EVALUATION [17]

Method	*mIOU*
UEC Foodpix	55.55%
DeepLabV3+	61.54%

The model exhibited accurate performance in drawing the segmentation mask; however, it encountered challenges in accurately classifying the food category.

Secondly, a collaborative approach was adopted to address the challenges faced by the DeepLabV3+ model in accurately determining the class of food images. Inspired by the methodology presented in the research paper [16], two models were integrated: YOLOv5 for object recognition and DeepLabV3+ for precise segmentation mask generation. This collaborative integration allowed us to leverage the unique strengths of each model. The YOLOv5 model excelled in identifying food classes, while the DeepLabV3+ model demonstrated proficiency in generating accurate segmentation masks. The combination of these models significantly improved the system's overall performance, effectively overcoming the limitations encountered previously.

The YOLOv5m model was trained on the UEC-Food100 dataset, which is the recognition version of the UECFoodPixComplete dataset. The training process utilized the following hyperparameters 130 epochs, a batch size of 64, an image size of 413, and the SGD optimizer.

After the training, the outputs of YOLOv5m and DeepLabV3+ were integrated, as shown in Fig. 8. After integrating the outputs of YOLOv5m and DeepLabV3+ models, the combined model achieved an impressive mIOU of 79.25%. This significant improvement in performance signifies the effectiveness of leveraging the strengths of both models to enhance the accuracy and quality of the results.

Error! Reference source not found. III Shows a c omparison between the related works and the proposed work, both before and after improvements were made, based on evaluations conducted using the UECFoodPixComplete dataset.

Following the segmentation and recognition processes, a calorie estimation technique was employed to estimate the calorie content of the identified food. A further test was conducted using various dishes in real-life scenarios, as shown in Table IV. The results revealed an average error of 36% in terms of calorie estimation. Although some deviation was observed, it is important to note that accurately estimating the calorie content of food can be challenging due to variations in portion sizes, cooking methods, and ingredient composition. Despite the margin of error, the system demonstrates promising potential in providing useful insights into the approximate calorie content of different dishes.

Table IV Shows the results of the system performance and results on real life scenarios. To sum-up, these are the techniques used for each of the stages which showed the performance improvement:

1. YOLO v5(m) & DeepLab v3+ for the recognition and segmentation stages.
2. Contour Analysis for the reference object detection and segmentation stage.

VI. CONCLUSION

The work proposed in this paper introduces a food calorie estimation methodology achieved through image analysis, harnessing food image recognition and segmentation techniques. The proposed approach relies on the presence of a fiducial marker (a reference object) to approximate food portion size and provide nutritional value based on the estimated portion size and food type. The utilization of the UECFOOD-100 and UECFOODPIXCOMPLETE datasets enabled the training of two models to work in tandem: YOLOv5 for food image recognition (trained on the former dataset) and DeepLabv3+ for food image segmentation (trained on the latter dataset). Additionally, the "Food and Their Calories" dataset, a CSV file containing food categories and their nutritional values, was employed.

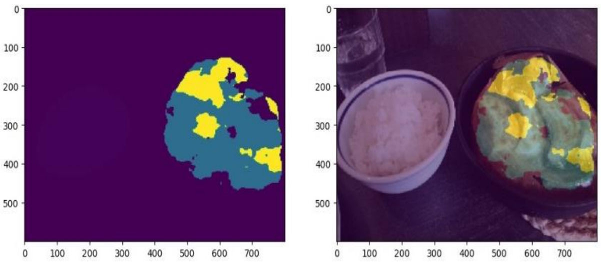

Fig. 7. DeepLabV3+ Sample Output Mask.

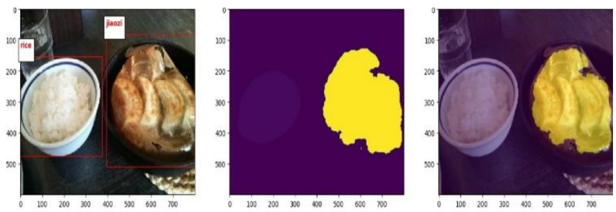

Fig. 8. YOLOV5 & DeepLabV3+ Sample Output Recognition and Mask Generation

TABLE II. DEEPLABV3+ MODEL RESULTS (OURS VS. [12])

Model	*Acc (%)*	*mIOU (%)*
DeepLabV3+ ([12])	66.8%	55.55%
DeepLabV3+ (Ours)	74%	45%

TABLE III GOURMETNET [17] VS. DEEPLABV3+ ONLY VS. DEEPLABV3+ & YOLOV5

Method	*mIoU*
GourmetNet [17]	65.13%
DeepLabV3+ [ours]	45%
DeepLabV3+ & YOLOV5 [ours]	79.25%

TABLE IV. WHOLE SYSTEM RESULTS

Food Name	*Real calorific value* (Kcal)	*Estimated calories* (Kcal)	*Error* (|Predicted-Actual|)	*Error(%)*
French fries	277	316	39	14%
Rice	198	200	2	1%
Egg sunny side up	90	158	68	75%

The collaborative approach adopted in this study using the two models yielded significantly improved results in the recognition and segmentation tasks compared to using only DeepLabv3+ on the same dataset. The mIOU increased by approximately 23%, and pixel accuracy improved by 7%.

This work paves the way for further research opportunities in image-based automated food calorie and food volume estimation using deep learning methods. Such advancements can play a pivotal role in enhancing automatic image-based dietary monitoring systems, benefiting individuals such as patients, athletes, weightwatchers, and more. These systems enable users to accurately record their diets and maintain a healthy lifestyle by tracking calorie counts through precise identification and portion size estimation of the food they consume. The results presented in this work demonstrate the potential of automated dietary monitoring and calorie estimation systems, promising greater convenience, and accuracy in maintaining a balanced and healthy diet.

REFERENCES

[1] F. Lo, Y. Sun, J. Qiu and B. Lo, "Image-Based Food Classification and Volume Estimation for Dietary Assessment: A Review," IEEE journal of biomedical and health informatics, vol. 24, no. 7, pp. 1926-1939, 2020.

[2] Y. Matsuda, H. Hoashi and K. Yanai, "Recognition of multiple-food images by detecting candidate regions," 2012 IEEE International Conference on Multimedia and Expo, pp. 25-30, 2012.

[3] Y. Kawano and K. Yanai, "Automatic Expansion of a Food Image Dataset Leveraging Existing Categories with Domain Adaptation," ECCV Workshop on on Transferring and Adapting Source Knowledge in Computer Vision (TASK-CV), 2014.

[4] Bossard, M. Guillaumin and L. Van Gool, "Food-101 – Mining Discriminative Components with Random Forests," ECCV 2014., vol. 8694, 2014.

[5] H. Kagaya, K. Aizawa and M. Ogawa, "Food detection and recognition using convolutional neural network," Proceedings of the 22nd ACM international conference on Multimedia, p. 1085–1088, 2014.

[6] S. Christodoulidis, M. Anthimopoulos and S. Mougiakakou, "Food recognition for dietary assessment using deep convolutional neural networks," Proceedings of New Trends in Image Analysis and Processing--ICIAP 2015 Workshops, p. 458–465, 2015.

[7] J. Qiu, F. P.-W. Lo, Y. Sun, S. Wang and B. Lo, "Mining discriminative food regions for accurate food recognition," Proceeding of Brit. Mach. Vision conference, 2019.

[8] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," Proceedings of International conference on machine learning, pp. 11905-11946, 2019.

[9] H. Hassannejad, G. Matrella, P. Ciampolini, I. De Munari, M. Mordonini and S. Cagnoni, "Food image recognition using very deep convolutional networks," Proceedings of the 2nd international workshop on multimedia assisted dietary management, pp. 41-49, 2016.

[10] Martinel, G. L. Foresti and C. Micheloni, "Wide-slice residual networks for food recognition," Proceedings of 2018 IEEE Winter Conference on applications of computer vision (WACV), p. 567–576, 2018.

[11] T. Ege and K. Yanai, "Simultaneous Estimation of Dish Locations and Calories with Multi-Task Learning," IEICE TRANSACTIONS on Information and Systems, vol. 102, no. 7, 2019.

[12] K. Okamoto and K. Yanai, "UEC-FoodPix Complete: A Large-scale Food Image Segmentation Dataset," Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10--15, 2021, Proceedings, Part V, pp. 647-659, 2021.

[13] P. Kadam, S. Pandya, S. Phansalkar, M. Sarangdhar, N. Petkar, K. Kotecha and D. Garg, "FVEstimator: A novel food volume estimator Wellness model for calorie measurement and healthy living," Elsevier Measurement, vol. 198, 2022.

[14] YOLO v5 model architecture [Explained]," OpenGenus, [Online]. Available: https://iq.opengenus.org/yolov5/. [Accessed 5 3 2023].

[15] Review: DeepLabv3+- Atrous Separable Convolution (Semantic Segmentation)," Medium, 29 9 2019. [Online]. Available: https://sh-tsang.medium.com/review-deeplabv3-atrous-separable-convolution-semantic-segmentation-a625f6e83b90. [Accessed 5 3 2023].

[16] K. Okamoto and K. Yanai, "An Automatic Calorie Estimation System of Food Images," Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, pp. 63-70, 2016.

[17] U. Sharma, "GourmetNet: Food Segmentation Using Multi-Scale Waterfall," Rochester Institute of Technology, 2021.