International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 20th Nov 2020.

Link : https://ijiemr.org/downloads/Volume-09/Issue-11

Title: Dietary Assessment Application Automated By Deep Learning

volume 09, Issue 11, Pages: 251-259

Paper Authors: A. Ramaswami Reddy

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Dietary Assessment Application Automated By Deep Learning

A. Ramaswami Reddy
Professor, Computer Science Engineering, Malla Reddy Engineering College, Maisammaguda, Hyderabad.

Abstract—Dietary diseases such as the diabetes, stroke, blood pressure, and other cardiovascular diseases, are the leading causes of death globally. The type of food consumed is majorly accountable for this, thus it requires continuous monitoring and evaluation of the individual's diet to different extents due to the imprecision in estimation of the portion size.We propose an automatic dietary assessment solution using CNN and computer vision-based approach, which aims to bridge the gap between convenience and accuracy while using dietary-assessment applications for health purposes. The application is trained to recognize the contents from a multi-object meal, on the sole basis of food images captured by a smartphone in real-time, and then estimate and return its nutritional facts such as calories and macronutrient content, like fats and carbohydrates.

***Keywords*—Convolution Neural Networks, Image Recognition, Computer Vision, Deep Learning, Volume Estimation.**

I. INTRODUCTION

Health awareness is growing among people due to the rapid revolution in various industries such as the food and health industry, digital & networking technology and social media marketing. It has created huge awareness among individuals about the diet-intake on a daily basis, food allergies, managing medical issues such as diet-related diseases, and weight-related goals such as weight-loss or weight-gain. Obesity is the first step towards developing deadly diseases like chronic heart diseases, diabetes, and other vascular syndromes; and the type of food consumed is majorly accountable for this, because often high caloric food is disguised in tempting varieties. Diabetes, stroke, high blood pressure, some types of cancer, osteoporosis, dental disorders, and other cardiovascular illnesses have also been linked to diet and nutrition by the World Health Organization (WHO). As a result, people with ailments must keep track of their food, particularly their carbohydrate (CHO) intake, which is critically affects blood glucose levels.

Consuming the right amount and appropriate type of food is the singular most effective formula in preventing and treating most kind of health-related issues, in addition to physical activity and exercises. Hence it requires continuous monitoring and evaluation of the individual's diet to different extents. This is a concern for many dieticians and nutritionists, who attempt to address these issues commonly by tracking and examining the daily eating habits of their patients that require maintaining a daily record of consumed food identified manually.The developments in technology in recent times has enabled the development of semi-automatic and automatic dietary evaluation solutions in the form of high-end applications, as well as the development of more efficient and convenient solutions employing computer vision. People, in general, are embracing technology more and are thus more willing to utilise self-managed food monitoring programmes, which can prove to be a more convenient alternative to monitoring daily food intake and controlling eating habits, due to ease of use and availability.

Many mobile apps, on the other hand, still demand user interaction and manual data entry, which can be time-consuming and inconvenient, making most users unlikely to use them for long. Another issue is that the results of many dietary-assessment systems are influenced by the inability to recognise particular foods or the imprecision in calculating portion sizes, which makes evaluating the precise nutritional contents of food intake difficult. As a result, an advanced system is required to carry out dietary assessment activities such as food-item identification, food-type categorization, and volume estimate automatically in order to determine the nutritional contents of the meal. Several attempts to design such smartphone-based nutritional applications have been made, but the outcomes have been either insufficient or inconvenient. The advancements in machine learning and computer vision algorithms, as well as smartphone technologies have paved the way for more powerful dietary assessment tools, generally in recognizing the food items, estimating their volume, and assessing the related nutrient information, while significantly improving improved the accuracy.

In this paper, we develop an automatic dietary assessment solution using CNN, which is trained to recognize the contents from a multi-object meal, by analysing meal images, and then estimating its nutritional facts such as calories and macronutrients, using a nutritional database.

II. RELATED WORK

Many mobile health applications have been developed as a result of the increased use of smartphones. Traditional image processing algorithms with hand-engineered characteristics were largely utilised in past food recognition research. Such dietary monitoring systems, on the other hand, suffer from imprecision, underreporting, time consumption, computational cost, and low adherence. Researchers have looked into automatic food recognition as a solution to the problem.

Meyers et al. [1]proposed the first work in nutritional evaluation using CNN, employing GoogLeNet Inception V1 on a dataset of photos from 23 different restaurants and in an outdoor scenario to estimate both meal size estimation and label classification, with a top-1 classification accuracy of 79 percent. ConvFood, proposed by Merchant et al. [2], implements transfer learning and fine-tuning on the InceptionV3 network, and was able to achieve comparatively higher accuracy on the Food-101 dataset than other systems that used a similar configuration.Ya Lu et al.[3] performed food detection, segmentation and recognition using deep neural networks and uses a 3D reconstruction algorithm to estimate the food's volume and requires of two meal images or a short video as the input. Another image-based calorie estimation system proposed by Ege et al. [4] employs region segmentation using a pre-registered reference object that is included in the food image.

The following sections discuss the most relevant vision-based approaches related to food image datasets, image segmentation and volume estimation.

A. Food Image Datasets

Image-based dietary assessment employs captured images as the main source of input for the food analysis. Thus training a food image classifier for food recognition depends strongly on an inclusive collection of food images, which points to the need oflarge and quality food image databases.It has been a common practice to verify new classifier performance by training it with large food image datasets such as Food101, PFID, UEC Food100, and UEC Food256.

On inspection of food image datasets, it found that many existing datasetseither have too diverse characteristicsin food categories and cuisine type withless total images in the dataset/per food class, or they are designated to a specific type of food. For example, the widely used Food101 dataset contains 101 food classes and a total of 101000 images, 1000 images per food class, captured in three different restaurants. The UEC-Food100 dataset contains photos of traditional Chinese and Japanese dishes, whereas Food101 and UNICT-FD889 contain a combination of eastern and western food imagery.

Apart from food diversity, we also require to take other picture attributes into account, such as if a segmentation approach was utilized or not.The majority of existing food image datasets, such as ETH Food101 [5], Recipe1M [6], and Geo-Dish [7], primarily aid in dish categorization and recipe generating research. They do not have fine-grained ingredient masks or labels, neither contain any ingredient-level annotations. The datasets Recipe1M and Recipe1M+ contains the ingredient labels for each image, but they don't include the segmentation masks. The only two published datasets for food image segmentation are UECFoodPix[8] and UECFoodPixComplete[9].Their segmentation masks, on the other hand, are only annotated at the dish level, meaning that each mask covers the complete dish rather than individual food ingredients. Thus there is a need of large scale food image datasets that facilitates fine-grained food image segmentation.

B. Semantic Segmentation& Classification

An automatic dietary assessment system must be able to identify and recognize the food contained in a meal. As a result, food image segmentation is a crucial and indispensable challenge for building health-related applications like calorie and nutrient estimation.

During feature extraction & classification, food images are utilized as input data to train a classifier, which must ideally be able to recognize any food type that has been incorporated in the learning process. Support vector machines (SVM), K-Nearest Neighbors (KNN), Bag of Features (BoF), Multiple Kernel Learning (MKL), and Random Forests are used to train a prediction model based on these features (RF).

Segmentation is an essential step as it improves classification accuracy while identifying different regions of an image to localize food items,especially when multiple food items have to be identified within a single image, and then extracting the object locations by excluding other objects such as the background or food containers. Kawano and Yanai [10] created a smartphone application and proposed that the user draw a manual boundary box to pick the food locations. To extract the desired regions, these areas are segmented using the GrabCut algorithm. Their method improves overall categorization accuracy, but it is still restricted by the user's ability to appropriately choose food items. Dehais et al. [11] utilised classical region growing with CNN-based border detection to address the image segmentation problem, which resulted in greater performance but at the cost of processing resources.

Manually designed features extraction techniques often cannot sufficiently abstract or represents the characteristics of the objects as huge variety of food types exist which also varies significantly under viewing angles and lighting conditions. In experiments comparing the efficiency and practicality between CNNs and traditional SVM-based techniques using handcrafted features, the traditional machine learning methods has found to have the accuracy of around 50-60%, whereas the CNN outperforms them by 10%. Food image segmentation methods currently in use underperform for two reasons:

- there is a scarcity of good-quality food image datasets that consisting of fine-grained food item labels and pixel-by-pixel location masks; the majority available datasets are either small in size or consist of coarse ingredient categories;
- the food's complex appearance makes it difficult to identify and locate ingredients in food photographs; for example, ingredients may overlap in the same image, and the same ingredient may appear differently in different food images.

C. Volume Estimation

To accurately quantify the dietary intake, measuring the portion size or volume/weight of food intake is essential. In practice, the process of estimating the total calories without an accurate instrument can be challenging, especially when a single 2-dimensional image is the only source of information, as the case of capturing an image with a smartphone or a handheld camera. There is usually no additional real-world information in these photos, such as the scale or depth of the items in the scene. In order to estimate volume, the information of segment-maps (produced by segmentation network) and depth-maps are necessary. The depth-images are usually generated using special hardware components such as depth sensors or by using stereo vision cameras, apart from also using monocular cameras. Depth Prediction techniques for inferring depth image from only RGB images has been extensively researched over the last years.However, due to a lack of food depth datasets, only a few studies in the field of nutritional assessment have been documented.

After a comprehensive exploration and review of a wide range of published research works about food volume estimation, the different employed approaches areconcluded as, stereo-based approach that uses multiple frames to reconstruct the 3D structure of food objects by finding pixel correspondences between image frames, model-based approach which uses pre-built shape templates (mathematical models) to determine the volume of objects, and perspective transformation, which can estimate irregularly shaped objects using a bird's eye view image to obtain a rough estimate of the object size of the object.Multiple images from different angles with known scene information, such as plates or containers with known sizes, can likewise be used to determine depth.Chen et al. [12] proposed the first food volume estimating system. It required a specific form model for each food type, as well as a calibrated reference card, and used a single view image as input. Puri et al. [13] employed a dense multi-view 3D reconstruction method to create a 3D point cloud of the food using a video sequence and plate-sized reference patterns.

Deep neural networks are yet another approach that has been extensively used in volume estimation. The recent employment of Convolutional Neural Networks (CNNs) has accelerated the research of the problem of inferring the depth map of a scene from its single RGB image. Allegra et al. [14] published the first attempt, which used a SegNet-based CNN architecture to estimate the depth of a single food image. For food image depth prediction, Christ et al. [15] have presented a CNN design with skip connections. However, due to a lack of food depth datasets, only a few studies in the field of nutritional assessment have been documented. The works done in RGBD datasets for the food images required to train the depth prediction model for such a dietary assessment solution are very few and difficult to obtain.

III.DIETARY ASSESSMENT ARCHITECTURE

In an archetypal scenario, the users capture one or more images of their meal using their smartphone camera. Then the designed dietary assessment system will automatically calculate and display the food type and the associated nutrient contents.

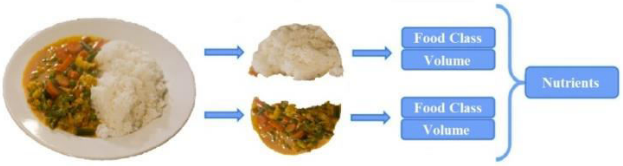


Fig. 1 Basic Working of a Dietary Assessment System

An image-based calorie assessment must recognize all food regions, segment the food objects in the image, and classify these regions accurately, followed by the calculation of the volume of each segmented item. Later, the nutrient information can be estimated by calculating the actual mass of the food according to the estimated volume (V) and the density of the classified food. After that, the calorie and nutrient content can be acquired from food nutritional database in a straightforward manner. However, the performance and efficacy of such solutions depends on various factors.The following sections describe the stages involved in our system.

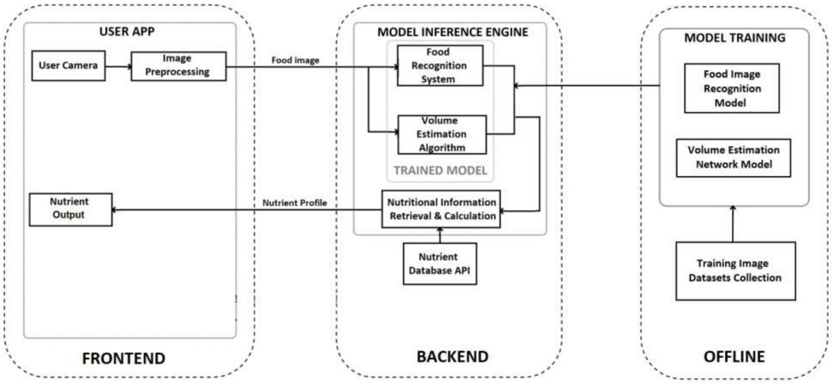


Fig. 2 Implementation of Dietary Assessment Application Design

A. Image Segmentation & Food Classification

In our proposed system, we use a new food image dataset FoodSeg103[16] consisting of 103 food categories, which solves the problem of fine grained food image segmentation effectively. The dataset uses high quality labels and masks for annotation, with each image comprising an average of 6 ingredient labels with pixel-wise masks. It includes 7,118 photos of western food, with 103 fine-grained ingredient categories created and allocated with category labels, as well as segmentation masks. The source images of FoodSeg103 are from an existing recipe dataset called Recipe1M which consists of millions of images and cooking recipes, used for recipe generation. We leverage this recipe information as auxiliary information to train our semantic segmentation models to perform the food image segmentation in our work.

To train our segmentation model, we randomly split FoodSeg103 dataset into training set and testing set, according to 7:3 ratio. Our training set has 4,983 images and 29,530 ingredient masks, whereas our testing set has 2,135 images and 12,567 ingredient masks (Total 42,097 masks).To perform food image segmentation we adopt a fully-convolutional network (FCN) with an encoder-decoder based architecture to segment different food portions from the input image.The framework contains two modules:

1) ***Encoder:***The encoder integrates the recipe information from the dataset into the visual representation image, which progressively reduces the spatial resolution. It is thus able to learn more semantic visual concepts with larger receptive fields, which explicitly equips the segmentation model with rich and semantic food knowledge.The vision encoder is initialized using a Resnet-50[17] network pre-trained on ImageNet-1k, which is trained using integrated food recipe data, which is widely used in multiple vision tasks.
2) ***Segmenter:***The decoder is randomly initialized and trained with the segmentation masks from the dataset on a CCNet[18] segmenter network. It uses a Dilated Convolution based semantic segmentation method for decoding, consisting of convolution layers that aim to enlarge the receptive fields without sacrificing the resolution. The Criss-Cross Network (CCNet) adaptively captures dense contextual information from all the pixels of the image on the criss-cross path in less computation cost and less memory cost.

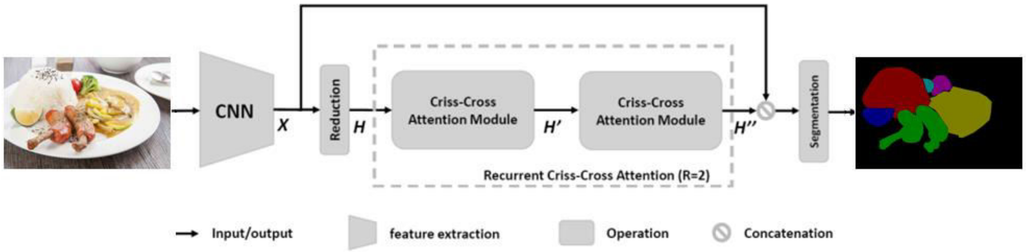

CNN
X
Reduction
H
Criss-Cross Attention Module
H'
Criss-Cross Attention Module
H''
Recurrent Criss-Cross Attention (R=2)
Segmentation
Input/output
feature extraction
Operation
Concatenation

Fig.3 Overview of Semantic Segmentation Network

B. Depth Prediction & Nutrient Assessment

In our proposed system, we leverage the Convolutional Neural Networks (CNNs) architecture developed by Hu et al. [19] to handle the problem of monocular depth prediction, which helps eliminate the need for a depth camera and offers reliable estimation results. It generates depth maps with higher spatial resolution which provides significant results in accurate estimation.We use ResNet-50 as our backbone network to train our model in our research. Due to the lack of RGBD food depth-image datasets, we settle on using the NYU-Depth v2[20] dataset consisting of a variety of indoor scenes, as it is the most widely used for the task of single view depth prediction, and it provides satisfactory results for our purpose.

The depth network architecture consists of four modules: an encoder (E), a decoder (D), a multi-scale feature fusion module (MFF), and a refinement module (R). The encoder extracts features at a four scales, 1/4, 1/8, 1/16, and 1/32. The decoder uses four up-projection modules to gradually scale up the encoder's final feature while reducing the number of channels.

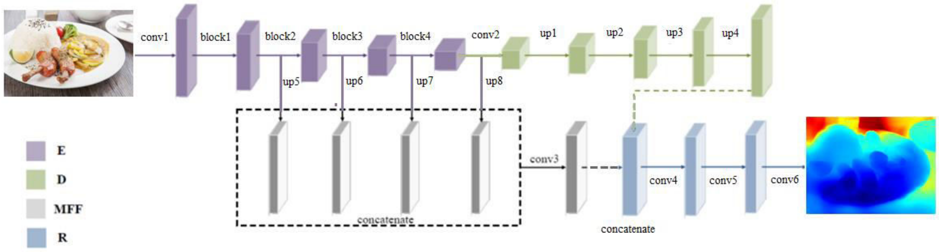
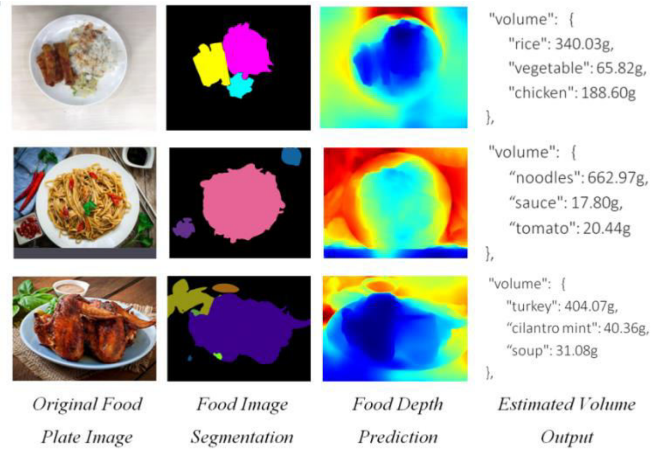


Fig.4 Overview of ResNet-50 Depth Network Architecture

The network predicts the RGB-Depth image from a single RGB food image, and then we convert the RGBD depth map is into a grayscale depth map. We then perform system transformations by combining it with the segment map we of the target food objects to approximate their volumes using the prior knowledge about the plate diameter. To perform this, the contour information from the segmentation mask is extracted and used to generate a JSON using LabelMepolygonal annotation. After that, the nutrient information of each food category is calculated from the volume by acquiring the calorie and macronutrient information from the USDA Food Composition Database[21].

IV. Results

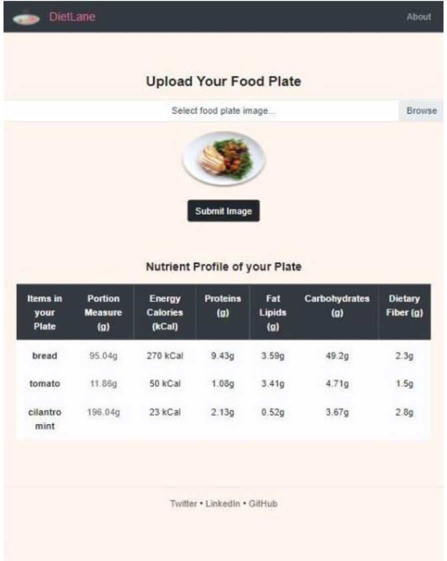
To obtain optimal classification accuracy in a nutritional evaluation system, the networkis trained with a large number of food images for each class. In addition to the extraction of these segments from the image backdrop, all food components inside the image are identified and their volumes are estimated to determine their nutritional information. The following are the outputs obtained from our semantic segmentation network and depth prediction network.



```
"volume": {
    "rice": 340.03g,
    "vegetable": 65.82g,
    "chicken": 188.60g
},
"volume": {
    “noodles”: 662.97g,
    “sauce”: 17.80g,
    “tomato”: 20.44g
},
"volume": {
    "turkey": 404.07g,
    “cilantro mint”: 40.36g,
    “soup”: 31.08g
},
```

Fig. 5 Performance of the Model

In our system, we have integrated semantic segmentation, depth prediction and nutrient assessment modules to propose a fully automated dietary assessment application. To use the app, the user uploads an image of their meal plate, and the application returns the portion measure of individual food items from the image in grams, along with energy in kCal, and macronutrients information including proteins, fats, carbohydrates and fiber in grams.



Upload Your Food Plate

Nutrient Profile of your Plate

Items in your Plate	Portion Measure (g)	Energy Calories (kCal)	Proteins (g)	Fat Lipids (g)	Carbohydrates (g)	Dietary Fiber (g)
bread	95.04g	270 kCal	9.43g	3.59g	49.2g	2.3g
tomato	11.86g	50 kCal	1.08g	3.41g	4.71g	1.5g
cilantro mint	196.04g	23 kCal	2.13g	0.52g	3.67g	2.8g

Fig.6 Dietlane ApplicationSnapshot

V. Discussions & Future Work

We proposed a fully automated nutritional estimation method that can estimate the volume of a food item from a single image. The food segmentation network used in this project accurately recognizes different foods, depending on how common and fine-grained they were. The proposed system can be of considerable practical value in different scenarios and use cases. This project provides a base model to further develop more robust and wide category-ranged applications for more common use. It is very useful for individuals who desire to monitor their diet and be cautious either for health related reasons or for lifestyle. Similarly, it can be a very valuable tool for dietitians and health care professionals who want to keep track of their patients' or clients' diets, either on-site or by having them share their data. This can also assist further in the production of statistical data and the extraction of patterns, which can aid in the improvement of one's diet.

References

[1] A. Meyers et al., “Im2calories: Towards an automated mobile vision food diary,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1233–1241.M. Rusin, E. Årsand, and G. Hartvigsen, “Functionalities and input methods for recording food intake: A systematic review,” Int. J. Med. Inform., vol. 82, no. 8, pp. 653-664, Aug. 2013.

[2] Merchant K., Pande Y. (2019) ConvFood: A CNN-Based Food Recognition Mobile Application for Obese and Diabetic Patients. In: Shetty N., Patnaik L., Nagaraj H., Hamsavath P., Nalini N. (eds) Emerging Research in Computing, Information, Communication and Applications. Advances in Intelligent Systems and Computing, vol 882. Springer, Singapore. https://doi.org/10.1007/978-981-13-5953-8_41.

[3] Ya Lu et al., “goFOOD: An Artificial Intelligence System for Dietary Assessment”, in Sensors 2020, volume 20, issue 15, pp 4283, Aug. 2020, doi: 10.3390/s20154283.

[4] T. Ege, Y. Ando, R. Tanno, W. Shimoda and K. Yanai, "Image-Based Estimation of Real Food Size for Accurate Food Calorie Estimation," 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, 2019, pp. 274-279, doi: 10.1109/MIPR.2019.00056.

[5] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. “Food-101–mining discriminative components with random forests.” In ECCV. 446–461. 2014.

[6] Amaia Salvador, Nicholas Hynes, Yusuf Aytar, Javier Marin, Ferda Ofli, Ingmar Weber, and Antonio Torralba. “Learning cross-modal embeddings for cooking recipes and food images.” In CVPR. 3020–3028. 2017.

[7] Ruihan Xu, Luis Herranz, Shuqiang Jiang, Shuang Wang, Xinhang Song, and Ramesh Jain. “Geolocalized modeling for dish recognition.” IEEE Transactions on Multimedia (2015), 1187–1199. 2015.

[8] Takumi Ege and Keiji Yanai. “A New Large-scale Food Image Segmentation Dataset and Its Application to Food Calorie Estimation Based on Grains of Rice.” In MADiMa. 82–87. 2019.

[9] Kaimu Okamoto and Keiji Yanai. “UEC-FoodPIX Complete: A Large-scale Food Image Segmentation Dataset.” In MADiMa. 2021.

[10] K. Yanai, Y. Kawano. “Food image recognition using deep convolutional network with pre-training and fine-tuning”. In IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy. 2015.

[11] J. Dehais, M. Anthimopoulos, and S. Mougiakakou. “Food image segmentation for dietary assessment”. In Proc. of ACM MM Workshop on Multimedia Assisted Dietary Management, Amsterdam, Netherlands, 23-28. 2016.

[12] HS. Chen, W. Jia, Z. Li, YN. Sun, JD. Fernstrom, M. Sun. “Model-based measurement of food portion size for image-based dietary assessment using 3D/2D registration”. Meas. Sci. Tech., 24, 10(2013), DoI: 10.1088/0957- 0233/24/10/105701. 2013.

[13] M. Puri, Z. Zhu, Q. Yu, A. Divakaran, H. Sawhney. “Recognition and volume estimation of food intake using a mobile device”. In Proc. IEEE Workshop Appl. Comp, 1–8. 2009.

[14] JD. Allegra, M. Anthimopoulos, J. Dehais, Y, Lu, F. Stanco, G. M. Farinella, S. Mougiakakou. “A Multimedia Database for Automatic Meal Assessment Systems”. In International Conference of Image Analysis and Processing (ICIAP), 471-478. 2017.

[15] P.F.Christ, S.Schlecht, F. Ettlinger, et al. “Diabetes60 – Inferring Bread Units From Food Images Using Fully Convolutional Neural Networks”. In IEEE International Conference on Computer Vision Workshop (ICCVW). 2017.

[16] Xiongwei Wu et al., “A Large-Scale Benchmark for Food Image Segmentation”, arXiv:2105.05409 [cs.CV], 12 May 2021.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep Residual Learning for Image Recognition.” In CVPR. 770–778. 2016.

[18] Zilong Huang, Xinggang Wang, Lichao Huang, Chang Huang, Yunchao Wei, and Wenyu Liu. “CCNet: Criss-Cross Attention for Semantic Segmentation.” In ICCV. 603–612. 2019.

[19] Junjie Hu et al., “Revisiting Single Image Depth Estimation: Toward Higher Resolution Maps with Accurate Object Boundaries”, arXiv:1803.08673 [cs.CV], 22 Sept 2018.

[20] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. “Indoor segmentation and support inference from rgbd images.” In ECCV, 2012.

[21] United States Department of Agriculture Agricultural Research Service Food Composition Databases: available at https://ndb.nal.usda.gov/ndb/.