IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 26, 2019, accepted February 19, 2019, date of publication March 13, 2019, date of current version April 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2904519

Vision-Based Approaches for Automatic Food Recognition and Dietary Assessment: A Survey

MOHAMMED AHMED SUBHI[1], (Member, IEEE), SAWAL HAMID ALI[1], (Member, IEEE), AND MOHAMMED ABULAMEER MOHAMMED[2]

[1]Department of Electric, Electronics and System Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia

[2]Al-Rafidain University College, Baghdad 46036, Iraq

Corresponding author: Mohammed Ahmed Subhi (mohdsubhi@siswa.ukm.edu.my)

This work was supported by Universiti Kebangsaan Malaysia under Grant DIP-2016-008.

ABSTRACT Consuming the proper amount and right type of food have been the concern of many dieticians and healthcare conventions. In addition to physical activity and exercises, maintaining a healthy diet is necessary to avoid obesity and other health-related issues, such as diabetes, stroke, and many cardiovascular diseases. Recent advancements in machine learning applications and technologies have made it possible to develop automatic or semi-automatic dietary assessment solutions, which is a more convenient approach to monitor daily food intake and control eating habits. These solutions aim to address the issues found in the traditional dietary monitoring systems that suffer from imprecision, underreporting, time consumption, and low adherence. In this paper, the recent vision-based approaches and techniques have been widely explored to outline the current approaches and methodologies used for automatic dietary assessment, their performances, feasibility, and unaddressed challenges and issues.

INDEX TERMS Food recognition, food classification, food volume estimation, food nutrient information, food image datasets.

I. INTRODUCTION

Obesity and overweight are defined as the result of energy imbalance between calories intake and expenditure [1]. This has been related to the risks of developing chronic heart diseases, diabetes, and other vascular syndromes. Obesity was the leading cause of death in 2012, with more than 1.9 billion overweight adults, and 650 million of those were obese [2]. Nutritionists attempt to address these issues traditionally by analyzing and monitoring the daily eating habits of their patients or alternatively by examining the images of consumed food [3]. However, the results are affected by the lack of correct logging of food intake by the patients or by the imprecision in estimating the portion size by simple examination of the food images.

Conventional dietary assessment programs require maintaining a daily record of consumed food, manual identification of its contents, and an estimation of its volume [4], [5]. However, these methods pose a challenge for elders especially when it involves an accurate estimation of the amount and time of the food intake. For these reasons, the need for a sophisticated system to automatically carry out all the tasks of food intake, such as detection, food type classification, and volume estimation, has been the main focus in many recent research efforts [6].

Recent developments in smartphone applications have made it possible to develop an efficient and more convenient solution for automatic dietary assessment [7], [8]. Recent studies revealed that smartphone-based dietary applications show higher user retention than traditional assessment methods [9], [10]. However, most of these applications require user intervention and manual input of food items affecting its performance on food content assessments [11].

The advancements in machine learning and computer vision based applications have paved the way for more robust dietary assessment tools. The general purpose of vision-based methods is to recognize the food, estimate its volume, and assess the related nutrient information. With the development of deep learning algorithms, food detection and recognition accuracy have been drastically improved. However, the performance and effectiveness of such solutions depend on several factors. First, optimal classification accuracy can be attained by training the image classifier with a large number of food images for each class [12]. Additionally, a proper

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

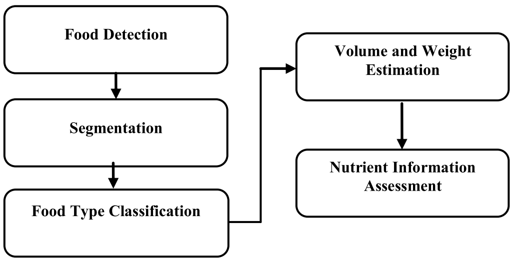


FIGURE 1. A typical procedure of vision-based dietary assessment system.

segmentation approach must be chosen and implemented to identify all food segments within a single image, in addition to the extraction of these segments from the image background. Finally, after identifying the food, volume estimation of each food item must take place to assess the corresponding weight and nutrient information [13], [14]. A typical procedure of vision-based dietary assessment system is shown in Fig. 1.

In this paper, we review the most relevant vision-based methods and techniques related to food intake detection and nutrient information estimation. Section 2 investigates the current food image datasets. Section 3 examines the current food classification techniques followed by a survey on current methods used for food volume and calorie estimation in Section 4. Conclusively, we highlight the remaining issues and future works related to this topic in Section 5.

II. FOOD IMAGE DATASETS

Training a food image classifier relies on an inclusive collection of food images. An assembled image dataset can be used subsequently to benchmark the recognition performance of other approaches. Several food image datasets have been created for this purpose. It has been a common practice to verify new classifier performance in contrast with the previous methods by training it with a large food image datasets such as Food–101 [15], PFID [16], UEC Food–100 [17], and UEC Food–256 [18]. Existing food image datasets have diverse characteristics, such as food categories, cuisine type, and the total images in the dataset/per food class. For example, PFID [16] has (61) classes of food with a total of 1098 images acquired from fast food restaurants and captured in laboratory conditions. While Food–101dataset [15] contains 101 food classes and a total of 101000 images, 1000 images per food class, captured in three different restaurants. Table 5 summarizes different food datasets with their respective characteristics.

By inspecting food image datasets, it is clear that most of the existing datasets are designated to a specific type of food. Thus, there is a need for a generic and comprehensive food image dataset that can be used for benchmarking and general classification purposes. For examples, the Turkish Foods–15 dataset [19] contains Turkish food images collected from other datasets, while the UNIMIB 2016 [12] consists of items from Italian cuisine acquired from

TABLE 1. Food image datasets.

Authors	Dataset	Food Category	Total # images/class	Image Source (s)	Ref.
Chen et al., 2009	PFID		1098/61	Captured in Restaurants/La b	[16]
Meyers et al., 2015	Food201-Segmented	Fast Food/ American	12625/201	A segmented version of Food–101	[22]
Mariappan, 2009	TADA*		256/11	Captured in controlled environment	[23]
Bossard et al., 2014	Food–101		101000/101	Downloaded from Web	[15]
Hoashi et al., 2010	Food85		8500/85	Acquired from previous databases	[24]
Matsuda et al., 2012	UEC-Food–100	Japanese	9060/100	Captured by cameral+ Labeled using Bounding Box	[17]
Kawano and Yanai, 2014	UEC-Food–256		31397/256	Captured by cameral+ Labeled using Bounding Box	[18]
Miyazaki et al., 2011	FoodLog		6512/2000	Captured by users	[25]
Wang et al., 2015	UPMC		90840/101	Web Image Search	[26]
Farinella et al., 2014	UNICT-FD889	Generic	3583/889	Captured by users using a smartphone	[21]
Singla et al., 2016	MMSPG-Food–11		16643/11	Collected from other food datasets	[27]
Singla et al., 2016	MMSPG Food–5K		5000/2	Collected from other datasets	[27]
Chen and Ngo, 2016	VIREO Food–172	Chinese	110241/172	Collected from Baidu and Google image search engines	[28]
Chen, 2012	Chen		5000/50	Downloaded from Web	[20]
Güngör et al., 2017	Turkish Foods–15	Turkish Dishes	7500/15	Collected from other datasets	[19]
Pandey et al., 2017	Indian Food Database	Indian Food	5000/50	Collected from Online Sources	[29]
Ciocca et al., 2017	UNIMIB 2016	Italian Food	1027/73	Images are captured from a dining hall food tray	[12]
Termritthikun et al., 2017	THFood–50	Thai Food	200–700/50**	Collected From Search Engines	[30]

* TADA dataset contained 256 images of real food and 50 images of replicas; however, only real food images are included in this table.

** THFOOD–50 has 200–700 images for each class.

a campus dining hall. Other datasets, Chen *et al.* [20] and UEC-Food–100 [17] contain images from traditional Chinese and Japanese dishes, respectively. While Food–101 [15] and UNICT-FD889 [21] consist of a mix of eastern and western food images. Moreover, it is noteworthy to state that, in addition to different food types, other image aspects such as if the image was acquired in free-living conditions, in a controlled environment, or whether a segmentation method exists or not were considered in the development of these datasets (Table 1).

III. FOOD IMAGE CLASSIFICATION

A basic automatic dietary assessment system is required to identify and recognize the food contained in a meal. The image classification, a machine learning technique, is used to identify a set of unknown objects that belong to a subset (class), which has been learned by the classifier in the training phase. In this step, food images are used as input data to train the classifier. An ideal classifier must be able to recognize any food type that has been included in the learning process. Practically, multiple variations exist in digital images, including rotation, distortion, color distribution, lighting conditions, and so forth, which may affect the overall accuracy. The training process itself is a tedious task that consumes a considerable amount of time to reach its intended accuracy goals. The classifier accuracy is affected mainly by the quantity and quality of images used in the training process as well as the proper selection of visual features. The extraction of image features used in the learning process splits a typical image classifier implementation into two strategies: traditional classifiers with handcrafted features and deep learning approaches as shown in Fig. 2.

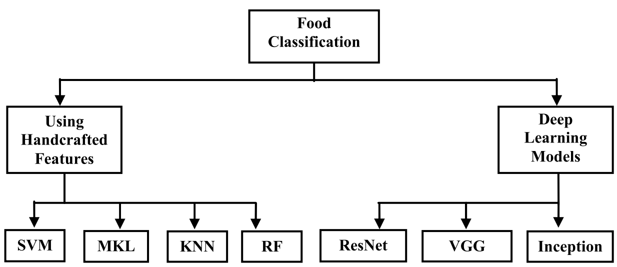


FIGURE 2. Common classification approaches for food images.

A. TRADITIONAL MACHINE LEARNING APPROACHES

The process of feature extraction in this category is implemented manually by inspecting the visual features found in the food images, such as color, shape, and texture. These features are then used to train a prediction model based on existing algorithms such as support vector machines (SVM) [24], K-Nearest Neighbors (KNN) [28], Bag of Features (BoF) [31], Multiple Kernel Learning (MKL) [24], and Random Forests (RF) [22]. The traditional classification methods basically execute three progressive tasks: segmentation, feature extraction, and classification. Segmentation is an essential step in identifying different regions of an image and then extracting the objects locations. In the case of food recognition, an appropriate segmentation approach should be implemented to localize food items in the image and exclude other objects such as the background or food containers [24], [28]. Segmentation, when implemented properly, improves the classification accuracy especially when multiple food items have to be identified within a single image [12], [31] or volume and nutrient contents have to be extracted [22], [32], [33]. Food segmentation is yet a challenging task, as some food images may not present features such as shape contours and food edges [34]. The segmentation could be more challenging when food items are minced, mixed in the food preparation process, and occluded food items laying on top of each other and hiding other parts of the food [35], [36]. Typically most of the segmentation approaches are based on the graph representation of the image as in equation (1). Graphs (G) are composed of a vertex set (V) that incorporates a set of image pixels or nodes, and whose edge set (E) is given by an adjacency relationship between these nodes. Finding the optimum "cut" that separates the nodes into two dissimilar sets is the common approach in most segmentation algorithms.

$$G = (V, E) \tag{1}$$

Several research works have been undertaken to address the issues related to the food segmentation process. Kawano and Yanai [31] developed a smartphone application and suggested that a manual bounding box must be drawn by the user to select the food areas. These areas are segmented using a GrabCut algorithm to extract the selected regions. Their approach improves overall classification accuracy but the performance is yet limited by the user's ability to select food items properly. Fig. 3 shows a GrabCut segmentation applied to extract certain food items from an image.

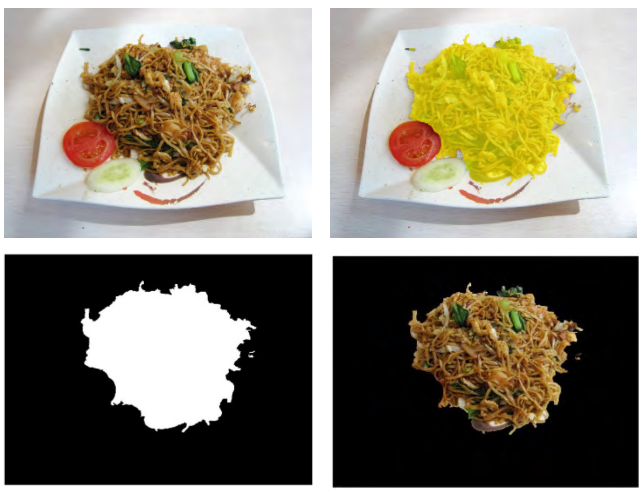


FIGURE 3. Food image segmentation using GrabCut algorithm.

Another study [14], suggested the use of Graph Cut segmentation algorithm as in (1), that basically attempts to cut the graph representation of the image into two sets *(A,B)* based on the dissimilarity found in the weight (w) of the edge that connects adjacent pixels *(u,v)* and hence extract selected food images from the background. In their work, 30 food categories were tested and the classification accuracy was much better than a color-based only segmentation reported earlier [37].

$$cut\,(A, B) = \sum_{u \in A, v \in B} W\,(u, v) \tag{2}$$

In another approach [12], several segmentation methods including image color, saturation, JSEG segmentation, and noise removal were combined to address the issue of multiple food identification. In this work, 73 food classes, found in a real food tray served in a canteen, were considered. The results showed that the classification accuracy was significantly improved. However, the tray images were manually segmented by drawing polygonal boundaries. Another study [35] attempted an ingredient based segmentation based on the spatial relationships between the objects in the image by applying a Semantic Texton Forest (STF) algorithm. The overall classification accuracy was improved when compared with the traditional methods. However, this method relies on the composition of visually distinctive ingredients organized in predictable spatial settings. Zhu *et al.* [38] implemented multiple segmentation hypotheses by assigning a class label to each pixel in an image. By using the classifier results as feedback to the segmentation, the number of segments in the image was estimated considering the confidence scores assigned to each segment. This approach outperformed the normalized cut method [39] as in (3), where $\underline{(assoc)}$ computes the total edge associations from nodes in A or B to all nodes in the graph (V).

$$Ncut\,(A, B) = \frac{cut\,(A, B)}{assoc\,(A, V)} + \frac{cut\,(A, B)}{assoc\,(B, V)} \tag{3}$$

Another study [17] proposed a JSEG segmentation approach linked with several object detectors including circle detector, whole image, and Deformable Part Model (DPM) combination. It was shown that overall classification accuracy could improve in relation to using the DPM model alone. He *et al.* [40] implemented a local variation segmentation algorithm, applied along with a segmentation refinement as feedback to increase the score of the classified items. The overall classification was improved when compared with the normalized cuts approach [39]. Kong *et al.* [41] used a perspective distance algorithm with three captured views of food objects and segmented them by clustering the features of each one. Segmentation accuracy was tested on 1–5 objects with 100% success rate for one type of food in the image and 76% success rate when five food items were included. In another study [42], users were asked to draw a bounding box and select a proper food tag from an available list then automatically segment the food using the GrabCut technique. The semi-automatic segmentation tool has been found to be effective when used on a large image dataset; however, user intervention is still needed. The food segmentation methods, summarized in Table 2, are mainly focused on visually separated food items (i.e., fruits and vegetables), yet the challenge remains to address the issues of food color, texture similarity, and variations found in prepared and mixed meals.

TABLE 2. Food image segmentation approaches.

Authors	Approach	Performance	Ref.
Yang et al., 2010	Spatial relationships and Semantic Texton Forests	Segmentation accuracy is not reported. Lacks the precision for image parsing in most of the food classes.	[35]
Matsuda et al..,2012	JSEG segmentation, circle detector, whole image, and DPM.	Moderate segmentation accuracy 21% (top 1) and 45% (top 5)	[17]
Kawano and Yanai, 2013	Bounding box and GrabCut Segmentation	Limited by manual selection of food items. Overall classification accuracy is improved.	[31]
He et al., 2013	Local variation segmentation and segmentation refinement as feedback	Overall classification accuracy is improved in contrast with the normalized cuts approach	[40]
Pouladzadeh et al.,2014	Graph cut segmentation	It achieves an overall segmentation accuracy of 95% and improves classification accuracy.	[14]
Zhu et al., 2015	Multiple segmentation hypotheses with assigned segment confidence scores.	Outperforms tradition normalized cut method and improves overall classification accuracy.	[38]
Meyers, 2015	Deep Lab Model	Improves classification accuracy	[22]
Kong et al., 2015	Perspective distance algorithm and cluster segmentation.	Tested on 1 to 5 food objects. A 100% success rate for one type of food and a 76% success rate for 5 segmented food items.	[41]
Shimoda and Yanai, 2015	Generated bounding box using CNNs and GrabCut.	It can detect bounding box regions around food items with a MAP of 49.9%.	[43]
Ciocca et al.,2017	A combination of color, saturation, JSEG, and noise removal.	The proposed segmentation provides better precision in contrast to other methods.	[12]
Fang et al., 2018	Manually drawn bounding box, manual selection of food tag and GrabCut.	This semi-automatic segmentation tool works efficiently when used on a large image dataset.	[42]
Inunganbi et al., 2018	Interactive image segmentation, Boundary detection and filling, and occlusion detection.	Classification accuracy is improved. Yet the food occlusion problem is only addressed when the food item is occluded by the container, multiple food items occlusion has not been discussed.	[44]

In the process of feature extraction, visual characteristics such as color, shape, and texture are identified [45]. In traditional machine learning, a proper selection of these features significantly improves the classification accuracy and vice versa. The term handcrafted features come from the researcher's ability to identify the relevant features of the desired objects in the image. In the case of food classification, food items vary in shape, color, and texture. The selection of associated features must relate to these three aspects [46]. To date, the challenge remains when prepared food is to be identified. Different methods of food preparation may result in different distinguishing features [28]. For example, the composition of a prepared salad has a different shape and texture from the shape and exterior texture of the whole fruits or vegetables. In order to find an optimal feature extraction process, informative visual data must be extracted from food images. These data can be found in general information descriptors, which are a set of visual descriptors that collect information about different basic features including color, texture, shape, and others. The descriptors, including Local Binary Patterns (LBP), Gabor filter, color information, and Scale Invariant Feature Transform (SIFT) can be applied individually to extract image features [20]. However, multiple descriptors can be implemented simultaneously to improve the overall classification accuracy. For example, a study implemented LBP and SIFT features individually on a food image dataset [20], the results showed that the accuracy of using SIFT features only is 53% while using the LBP features only resulted in 46% accuracy. Combining both features, along with additional Gabor filter and color features, improved the accuracy to 68%. In another study [47], the same dataset was used and SIFT, LBP and color features were extracted in addition to other features such as Histogram of Oriented Gradients (HOG) and MR8 filter. A combination of these handcrafted features obtained an accuracy of 77.4%. The study revealed that different parameters of the same extracted features may add up to the overall classification accuracy.

There are several classification approaches with a variety of manually extracted features. Support Vector Machines (SVM) and K-Nearest Neighbor (KNN) have been the chosen traditional methods in several investigations in the field of food image recognition, mostly due to their substantial performance compared with other methods. A recent study [38] applied color, texture, and SIFT features to train a KNN classifier for food recognition. In contrast with an SVM classifier, KNN achieved a better classification accuracy of 70% while SVM classification achieved only 57%.

Anthimopoulos *et al.* [45] implemented a bag-of-features (BoF) model with SIFT extracted features. The authors trained an SVM linear image classifier to identify 11 classes of food and obtained an accuracy of 78%. Chen *et al.* [20] Implemented a multi-class SVM classifier to identify 50 classes of Chinese food with 100 images in each category. Further, the authors added a multi-class Adaboost algorithm and improved the classification accuracy to 68.3%, followed by 62.7%, when SVM was implemented separately. Moreover, Beijbom *et al.* [47] applied SIFT, LBP, color, HOG and MR8 features and developed an SVM image classifier. An evaluation of their work was applied to two food image datasets and achieved a 77.4% accuracy in the dataset presented earlier [20], while they obtained only 51.2% precision using their menu-match dataset.

The traditional food classification methods, summarized in Table 3, highlight the type of the implemented classifiers, the selected visual features, and the overall performance. Thus far, the process of features selection remains a challenging task regarding food image classification.

Food items, such as fruits and vegetables, come in distinctive shapes, colors, and textures that are easily separable and could be identified. However, the resemblance in the color and texture of mixed and prepared food renders the traditional classification methods ineffective. Alternatively, with

TABLE 3. Traditional classification approaches.

Authors	Classifier	Features	Performance Top 1	Performance Top 5	Ref.
Hoashi et al., 2010	MKL	BoF, Gabor, color, HOG, and texture	62.5%	N/A	[24]
Yang et al., 2010	SVM	Pairwise local features	78.0%	N/A	[35]
Kong and Tan, 2011	Multi-Class SVM	Gaussian Region Detector and SIFT	84%	N/A	[48]
Bosh et al., 2011	SVM	Color, Entropy, Gabor, Tamura, SIFT, Haar Wavelet, Steerable, DAISY, and Predominant color divided into local and global features.	86.1%	N/A	[34]
Matsuda et al., 2012	MKL-SVM	HOG, SIFT, Gabor, color and texture	21.0%	45.0%	[17]
Kawano and Yanai, 2013	SVM	SURF and color	N/A	81.6%	[31]
Anthimopoulos et al. 2014	SVM	SIFT, color	78.0%	N/A	[45]
Tammachat and Pantuwong, 2014	SVM	BoF, SFTA, and color	70.0%	N/A	[49]
Pouladzadeh et al.., 2014	SVM	GraphCut, color, size, shape and texture	95.0%	N/A	[14]
He et al., 2014	KNN	DCD, SIFT, MDSFIT, and SCD	64.5%	N/A	[50]
Kawano and Yanai, 2014	One × rest linear classifier	Fisher Vector, HOG and color	50.1%	74.4%	[51]
Christodoulidis et al., 2015	SVM	LBP and color	82.2%	N/A	[52]
Yanai and Kawano, 2015	Fisher Vector	HOG and color	52.9%	75.5%	[53]
Pouladzadeh et al., 2015	Cloud-Based SVM	Gabor, color	94.5%	N/A	[54]
Farinella et al., 2016	SVM	SIFT, PRICoLBP, and Bag of textons	75.74 %	85.68 %	[55]

the development of deep learning algorithms, the need for manual feature selection as well as any user intervention has been eradicated or reduced. Hence, it may form a strong foundation for a prospective fully automatic food identification system.

B. DEEP LEARNING APPROACHES

Deep learning, a subset of machine learning, is a new approach to learn and train a more effective neural network. The built-in mechanism of deep learning algorithms adopts the features extraction automatically through a series of connected layers followed by a fully connected layer which is responsible for the final classification. It has recently become popular owing to its marginally exceptional performance with enhanced processing abilities, large datasets, and outstanding classification ability compared to other traditional methods [56], [57]. Convolutional Neural Network (CNN) is one of the most prominent techniques in deep learning. It was introduced by LeCun *et al.* [58] for the classification of handwritten digits. CNNs is widely preferred in computer vision applications owing to its exceptional ability to learn operations on visual data and obtain high accuracies in challenging tasks with large-scale image data [59]. CNN, in contrast to other traditional methods, outperforms by a large margin. In the field of food recognition and classification, several research works have implemented this approach. Bossard *et al.* [15] implemented a CNN model based on the network architecture proposed earlier [60]. Using images from their own dataset (Food–101), the average accuracy achieved was only 56.4% accomplished in 450000 iterations. Yanai and Kawano [53] implemented a deep convolutional neural network (DCNN) on three different food datasets, Food–101, UEC-FOOD–100 and UEC-FOOD–256. The authors investigated the effectiveness of pre-training and fine-tuning of a DCNN with 100 training images for each food category acquired from each dataset. In the experiments, the best classification accuracy achieved was 78.77%, 67.57%, and 70.4% for the UEC-FOOD100/256, and Food–101 datasets, respectively. Proving that fine-tuning of the DCNN pre-trained with a large number of food-related categories (DCNN-FOOD) can significantly improve the classification accuracy. In another study [46], the performance of Inception V3 deep network introduced by Google [61] was performed. Similarly, three datasets were chosen for the performance evaluation, Food–101, UEC-FOOD–100 and UEC-FOOD–256. It was shown that the fine-tuned version of Inception V3 can attain promising results for the three food image datasets. Their approach achieved 88.28%, 81.45%, and 76.17% accuracy, respectively. In the same manner, a CNN based approach using the Inception model was also implemented [62]. The accuracy achieved was 77.4%, 76.3%, and 54.7% for Food–101, UEC-FOOD–100, and UEC-FOOD–256, respectively. Table 4 gives an overview of the existing methods of food recognition based on deep learning techniques and their performance. It is noteworthy to state that food quantification and classification has been the concern of the majority of the existing dietary assessment research in this domain [46]. The method summarized in Table 4, are concerned mainly with the identification and categorization of food items rather than estimating its actual volume and corresponding nutrient information, and hence, it is limited by the inability to assess the daily calorie intake.

IV. FOOD VOLUME ESTIMATION

Once the food items in a given image have been identified, the volume/weight of the detected food is estimated, so that its corresponding nutrients information, such as sugar, carbohydrates or calories, could be determined. In practice, the process of estimating the total calories without an accurate instrument can be challenging, even to most nutritionists. An image-based calorie assessment must recognize all food regions, segment the food objects in the image, and classify these regions accurately [20], [62], followed by the calculation of the volume of each segmented item. 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 (d) as in (4), shown at the bottom of this page. The calorie and density information can be acquired as in (5), shown at the bottom of this page, from food nutritional database [51], [65], [66], such as the USDA Food Composition Database [67].

Estimating the volume of a food object can be challenging 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. These images normally do not contain any additional real-world information such as the scale or the depth of the objects in the scene. To estimate the depth, a synthesized image that contains information relating to the distance of the objects in a scene from the camera is usually generated using special hardware components such as depth sensors or by using stereo vision cameras with known focal length (f) and known baseline length (B) as the distance between the two cameras centers (4). The depth can also be estimated using multiple images from different views with known scene information, such as plates or containers with known size [68], [69].

$$depth = \frac{Bf}{disparity} \tag{6}$$

$$Mass = d \times V \tag{4}$$

$$Estimated\ Calories\ of\ a\ food\ item(C) = \frac{Calculated\ Mass(M) \times Database\ Calories\ of\ a\ Food\ Item}{Database\ Weight\ of\ Food\ Item} \tag{5}$$

TABLE 4. Deep learning classification approaches.

Authors	Technique	Dataset	Performance Top 1	Performance Top 5	Ref.
Bossard e. al., 2014	Food–101	Food–101	56.4%	N/A	[15]
Yanai and Kawano, 2015	DCNN-Food		70.4%	N/A	[53]
Meyers, 2015	Google Net/Food101		79.0%	N/A	[22]
Liu et al., 2018	DCNN + edge computing		77.0%	94.0%	[63]
Hassannejad et al., 2016	Inception V3		88.3%	96.9%	[46]
Liu et al., 2016	DeepFood		77.4%	93.7%	[62]
Pandey et al., 2017	DCNN, Ensemble Net		72.1%	91.6%	[29]
Anthimopoulous et al., 2014	ANNnh	Diabetes	75.0%	N/A	[45]
Christodoulidis et al.., 2015	Patch-wise CNN	Own Database	84.90%	N/A	[52]
Pouladzadeh et al., 2016	Deep Neural Network		99.0%	N/A	[64]
Kawano and Yanai, 2014	Deep Convolution +Fisher Vector	UEC-Food–100	72.3%	92.0%	[65]
Yanai and Kawano, 2015	DCNN-Food		78.8%	95.2%	[53]
Liu et al., 2016	DeepFood		76.3%	94.6%	[62]
Hassannejad et al., 2016	Inception V3		81.5%	97.3%	[46]
Chen and Ngo, 2016	Arch-D		82.1%	97.3%	[28]
Yanai and Kawano, 2015	DCNN-Food	UEC-Food–256	67.6%	89.0%	[53]
Liu et al., 2016	DeepFood		54.7%	81.5%	[62]
Hassannejad et al., 2016	Inception V3		76.2%	92.6%	[46]
Chen and Ngo, 2016	Arch-D	VIREO	82.1%	95.9%	[28]
Ciocca et al.., 2017	VGG	UNIMI NB2016	78.3%	N/A	[12]
Pandey et al., 2017	DCNN, Ensemble Net	Indian Food Database	73.5%	94.4%	[29]
Termritthikun et al., 2017	NU-InNet1.0	THFood –50	69.8%	92.3%	[30]
Termritthikun et al., 2017	NU-InNet1.1		68.7%	92.3%	[30]

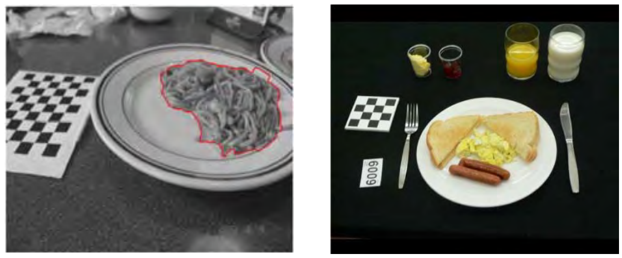


FIGURE 4. A checkerboard reference object is used to estimate the real dimensions of food items [70], [72].

Additional parameters such as the scale and pose of objects are important components of understanding geometric relations within a scene. A 3D model of an object is only perceived if these parameters can be estimated. A fiducial marker or a reference object (Fig. 4) with a known size and scale is often placed in the scene to relate to the actual dimensions of other objects [70]–[72].

A crowdsourcing approach has been implemented to estimate the food volume and its nutrient information [73]. In this method, users are asked to take a photo of their meal that is available to be evaluated by other individuals. In another approach, the volume of the food is estimated using a depth sensor camera [20], [22] or an additional laser device attached to a smartphone [74]. These methods achieved promising results, though the performance was limited by the fact that food images were captured in a controlled environment or a more sophisticated device such as a depth sensor or an additional camera was used [33], which could be a practical limitation in real-world conditions.

In another approach, an additional reference object in the scene is used to estimate the volume of the meal. The thumb of a user is placed as a reference in a two-dimensional image for volume estimation. Two pictures are captured along with the user's thumb from the top and side views of the plate [32], [37]. The top view image is divided into a grid of squares to facilitate the area estimation of different food shapes. The total area (TA) of the food portion is calculated as the sum of all sub areas for each square (T_i) for an (n) number of projected squares (6). While the volume is calculated as in (7), using the depth (d) estimated from the side view image.

$$TA = \sum_{i=1}^{n} Ti \tag{7}$$

$$V = TA \times d \tag{8}$$

In real-life conditions, several food items can be occluded in the side view, complicating both the identification and volume estimation tasks. Similarly, a 3D reconstruction model using a calibrated camera settings in addition to another reference object, such as a checkerboard, was implemented to estimate depth information [13], [70], [72], [75]. This approach also requires users to carry additional equipment and calibrate the cameras to gain the depth and to estimate its volume, which can be burdensome to most users. Another approach of using reference objects in the scene is to use food containers with known size and shape. For example, a pre-trained plate or a circular container with the known size is implemented to estimate its food contents [70], [71], [76], [77]. It is more practical to avoid carrying additional objects while consuming food; however, it is limited by the choice of specific plates or containers. Moreover, the volume of the food is also estimated using a shape template 3D reconstruction, which fits the detected food item into a corresponding 3D model [40], [78]–[80]. In addition to the requirement of a fiducial marker in the scene, this approach does not perform well with irregular food shapes. A state of the art approach implementing a fiducial marker-free volume estimation was presented by Yang *et al.* [81]. The authors proposed an approach where a virtual cube with fixed dimensions (4cm × 4cm × 4cm) is generated in the viewing screen. The user is asked to place the cube next to the food object and scale it by applying common touch gestures to match the size of the food item, as shown in Fig. 5. The limitation of this approach was that the smartphone has to be placed on the tabletop with a flat surface to calibrate the cameras. This approach achieved an average estimation absolute error of 16.65% for ten types of food.

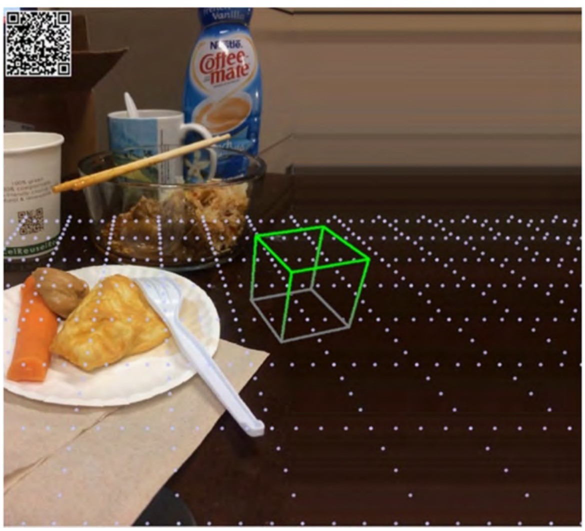


FIGURE 5. Virtual reality method for food volume estimation by [81].

Food volume estimation methods, summarized in Table 5, achieved promising results, yet more needs to be explored and tested in real-life conditions rather than being tied to a controlled environment. Most of the implemented techniques are not feasible outside the laboratory settings, where nutrient information may vary depending on the preparation method of the food. Taxonomy of general food volume estimation approaches is depicted in Fig. 6.

TABLE 5. Methods of food volume estimation.

Authors	Technique	Performance	Ref.
Chen et al., 2012	Depth camera	Preliminary Results, performance is not reported	[20]
Woo et al., 2010	3D reconstruction with the reference card	Mean Volume Error of 5.68% tested on multiple food items.	[75]
Villalobos et al., 2012	Top + side views measurement with the user's finger as a reference	Error in the acceptable range, results varies in different sets of illumination and viewing angle.	[32]
Beijbom et al., 2015	Restaurant's menu items	Predefined calories from the menu can be inaccurate	[47]
Noronha et al., 2011	Crowdsourcing to estimate the calories	An error-prone approach since the calories are estimated by visual inspection only	[73]
Zhu et al., 2010	3D reconstruction using spherical and prismatic models with a reference card.	Seven fruits items have been measured for performance with a mean error of 5.65%	[72]
Meyers et al., 2015	3D volume estimation using depth camera and reconstruction using CNN's and RANSAC.	Volume estimation accuracy was high for most food types however food replicas were used in a controlled environment	[22]
Chae et al., 2011	Use shape specific templates to reconstruct a 3D model of drinks and bread slices	Overall volume estimation relative error for 17 dinks is 11% and 8% for bread slices	[78]
Xu et al., 2013a	Using shapes from silhouettes to estimate food portion size and multi-view 3D reconstruction	Achieved a 10% average error on four types of food using the automatic multi-view volume estimation and 17.9% average error of weight estimation on 19 types of food using measurement approach.	[79]
Yang et al., 2018	A fiducial-marker free method with a smartphone motion sensor data to determine camera orientation	Achieved a volume estimation with an average absolute error of 16.65% for ten types of food, limited by the placement method and the size of the smartphone.	[81]
Jia et al., 2014	100 food samples were collected using a wearable camera (eButton), the volume is estimated using a shape-based approach	85 food items out of 100 had less than 30% error using the computerized method	[82]
He et al., 2013	3D reconstruction using a food-specific shape template	Beverage food items were tested using cylinder shapes with an average relative error at 11%	[40]
Martin et al., 2009	The weight of food is manually trained with specific food dishes and compared with the area of classified food area and leftovers	The method proves its performance of an accurate area calculation of two images (before and after epochs)	[76]

V. OUTSTANDING ISSUES AND CHALLENGES

The performance of an automated dietary assessment approach is reliant on each of its subtasks. Starting with the quantity and quality of images acquired from a food image dataset, the proper segmentation of food objects in each image, the classifier's accuracy to detect and identify

TABLE 5. *(Continued.)* Methods of food volume estimation.

Jia et al., 2012	Used circular reference objects or circular spot LED pattern.	Two methods were implemented, a plate reference method achieved an average error of 12.01% and an LED method with an average error of 29.01%	[77]
Rahman et al., 2012	Stereo images are used for food 3D reconstruction.	Achieved an average volume estimation error of 7.7%- for six fruits.	[33]
Pouladzadeh et al., 2014	Two images were captured from two views (top and side) with the user's thumb as a reference in the top view for area measurement and the side view for depth.	Non-mixed food volume estimation error ranges between 10% as worst case and 1% for the best case for five types of food.	[37]
Dehais et al., 2017	Dense 3D reconstruction from two views.	Achieved a MAPE ranging from 8.2% to 9.8% in two different datasets for 45 dishes in the first dataset and 14 meals in the second.	[83]
Puri et al., 2009	Dense 3D reconstruction from three views.	The performance of volume estimation of 26 types of food achieved an average error of 5.75%.	[84]
Yue et al., 2012	Used plates or containers as reference objects with known size.	Only length and thickness were estimated, no volume estimation was done in the experiments. An average estimation error of 3.41% of two dimensions was reported (Length and Thickness).	[71]
Xu et al., 2013b	Used a pre-trained 3D model of different food shapes with orientation information.	The approach achieved an average error of 10% for five food categories.	[80]
Shang et al., 2011	A smartphone attached laser device was used to capture the depth in the images of food objects.	The projected laser grid captured from a video sequence was used to generate the 3D model, the performance of the method is not reported.	[74]
Fang et al., 2015	Single-view 3D reconstruction of food using a reference object and shape of the container	Achieved less than 6% error estimation of energy intake in meals.	[70]
Fang et al., 2018	Used Generative Adversarial Networks (GAN) to map food energy distribution in the image	Achieved less than 10.89% error rate of energy estimation.	[85]
Subhi et al., 2018	Stereo image analysis and food front edge detection to estimate its height and depth	Achieved an average volume estimation error of 8.5% for four types of food	[86]
Liang and Li, 2017	A two-dimensional view was measurement using a reference object (Coin)	Estimation error was below 20% for most of the tested 19 food categories.	[87]

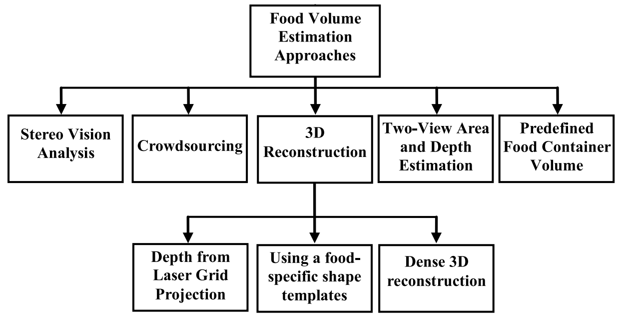


FIGURE 6. Taxonomy of food volume estimation approaches.

food contents, and the ability to estimate the corresponding volume and the corresponding nutrient information.

Despite the advancements in food identification methods, many challenges still exist in each of the aforementioned steps. For instance, the performance of a classifier is highly dependent on the source of images found in the food datasets. Even though there is a growth in the number and volume of current food image datasets to incorporate more food categories, such as Food85 [24], Food201-Segmented [22], and UEC Food–256 [18]. There is a need for a generic and comprehensive food image dataset to be used for benchmarking and performance evaluation. Moreover, the innovation of Deep Learning models has made it possible for classifiers to efficiently identify new food items. The size of trainable image data has a significant impact on the overall accuracy, and hence large food image datasets can improve the overall performance [60]. It is possible to generate more food images from the existing datasets by implementing basic image processing techniques such as cropping, rotation, adding noise, or manipulating existing features such as brightness, saturation, contrast, and hue [46].

Although segmentation of food items has significantly improved in the classification performance [38], it is still challenging to segment prepared, occluded, or mixed food items. The segmentation process is also limited by other factors that may contribute negatively to the segmentation accuracy. For example, different lighting conditions may result in blurry edges or shadows that might be detected as a part of food regions by the segmentation algorithms. Whereas other methods that involve manually-selected food regions can be promising [43], yet inaccurate bounding box size may negatively affect the overall accuracy [31].

Moreover, the food portion size estimation is limited by several external factors that may affect the performance of the volume estimation process, including different lighting conditions, blurred edges, or noisy background [20]. These factors need to be addressed properly and further experimentations are needed under these conditions.

Moreover, most of the existing volume estimation methods have been only applicable to solid and separable food items, such as fruits or vegetables. Currently, the food can only be clustered according to its general shape as the relationship between the food volume estimation method and the food category. It would be more beneficial to address the impact of applying different volume estimation methods on different food categories such as prepared, minced or mixed food. Estimating food volume using 2D images is still far from an acceptable range even while using additional fiducial markers such as a checkerboard [75] or user's thumb as a reference object [32]. Moreover, using stereo cameras may alleviate the depth estimation problem as demonstrated earlier [22]. To date, the number of strategies has been reported for food volume estimation and a nutrient information analysis is still limited.

Nutrient and calorie estimation remains to be an error-prone stage in automated dietary assessment systems, as it depends directly on the accuracy of the previous stages, i.e., food segmentation and volume estimation [22]. Therefore, calories can be overestimated or underestimated if any of the other stages is inaccurate.

Further experimentation needed in the aim for developing a fully automated system. Inevitably, the continuous development of innovative smartphone and related wearable devices may mitigate the complexity of dietary assessment systems when more functionalities and sensors are embedded.

VI. CONCLUSION

In this paper, we have investigated a wide range of strategies in computer vision and artificial intelligence tailored for automated food recognition and dietary assessment. In practice, the entire process can be broken down into four tasks: food image acquisition from corresponding datasets, the segmentation of food images, a proper classification approach either with handcrafted features, or using deep learning, and finally the estimation of food volume and its nutrient information. The current methods and techniques have exhibited improved performance, yet there exist challenges and limitations in every aspect of the process. A comprehensive and generic food image dataset needs to be developed for benchmarking and performance evaluation, as large food image datasets can improve the overall performance. Moreover, segmentation is still challenging when prepared, occluded, or mixed food items are considered. Meanwhile, volume estimation methods have been only applicable to solid and separable food items, more experiments need to be applied to estimate the volume of prepared or mixed food items. The innovation of healthcare applications and wearable devices and the integration of these devices into a smartphone will revolutionize this line of research and, overall, automated dietary systems will provide insights on effective health management and disease prevention.

REFERENCES

[1] F. Ahmed and C. Siwar, "Food intake and nutritional status among adults: Sharing the Malaysian experience," *Pakistan J. Nutrition*, vol. 12, no. 11, pp. 1008–1012, 2014.

[2] World Health Organization. (2018). *WHO|Obesity and Overweight*. Accessed: [Online]. Available: Oct. 4, 2018. http://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight

[3] A. A. Fatehah, B. K. Poh, S. N. Shanita, and J. E. Wong, "Feasibility of reviewing digital food images for dietary assessment among nutrition professionals," *Nutrients*, vol. 10, no. 8, p. 984, 2018.

[4] 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.

[5] J. Siswantoro, A. S. Prabuwono, and A. Abdulah, "Volume measurement of food product with irregular shape using computer vision and Monte Carlo method: A framework," in *Proc. 4th Int. Conf. Elect. Eng. Informat. (ICEEI)*, vol. 11, 2013, pp. 764–770.

[6] J. Ngo, A. Engelen, M. Molag, J. Roesle, P. García-Segovia, and L. Serra-Majem, "A review of the use of information and communication technologies for dietary assessment," *Brit. J. Nutrition*, vol. 101, no. S2, pp. S102–S112, Jul. 2009.

[7] E. Mendi, O. Ozyavuz, E. Pekesen, and C. Bayrak, "Food intake monitoring system for mobile devices," in *Proc. IEEE 5th Int. Workshop Adv. Sensors Interfaces (IWASI)*, Jun. 2013, pp. 31–33.

[8] I. Haapala, N. C. Barengo, S. Biggs, L. Surakka, and P. Manninen, "Weight loss by mobile phone: A 1-year effectiveness study," *Public Health Nutrition*, vol. 12, no. 12, pp. 2382–2391, 2009.

[9] M. C. Carter, V. J. Burley, C. Nykjaer, and J. E. Cade, "Adherence to a smartphone application for weight loss compared to website and paper diary: Pilot randomized controlled trial," *J. Med. Internet Res.*, vol. 15, no. 4, p. e32, 2013.

[10] Y. S. Chen, J. E. Wong, A. F. Ayob, N. E. Othman, and B. K. Poh, "Can Malaysian young adults report dietary intake using a food diary mobile application? A pilot study on acceptability and compliance," *Nutrients*, vol. 9, no. 1, p. 62, 2017.

[11] F. Cordeiro *et al.*, "Barriers and negative nudges: Exploring challenges in food journaling," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, 2015, pp. 1159–1162.

[12] G. Ciocca, P. Napoletano, and R. Schettini, "Food recognition: A new dataset, experiments, and results," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 3, pp. 588–598, May 2017.

[13] F. Zhu, M. Bosch, C. J. Boushey, and E. J. Delp, "An image analysis system for dietary assessment and evaluation," in *Proc. IEEE Int. Conf. Image Process.*, New York, NY, USA, Sep. 2010, pp. 1853–1856.

[14] P. Pouladzadeh, S. Shirmohammadi, and A. Yassine, "Using graph cut segmentation for food calorie measurement," in *Proc. MeMeA*, Jun. 2014, pp. 1–6.

[15] L. Bossard, M. Guillaumin, and L. Van Gool, "Food-101—Mining discriminative components with random forests," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 446–461.

[16] M. Chen, K. Dhingra, W. Wu, L. Yang, R. Sukthankar, and J. Yang, "PFID: Pittsburgh fast-food image dataset," in *Proc. IEEE 16th Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 289–292.

[17] Y. Matsuda, H. Hoashi, and K. Yanai, "Recognition of multiple-food images by detecting candidate regions," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2012, pp. 25–30.

[18] Y. Kawano and K. Yanai, "Automatic expansion of a food image dataset leveraging existing categories with domain adaptation," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 3–17.

[19] C. Güngör, F. Baltacı, A. Erdem, and E. Erdem, "Turkish cuisine: A benchmark dataset with Turkish meals for food recognition," in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU)*, May 2017, pp. 1–4.

[20] M.-Y. Chen *et al.*, "Automatic Chinese food identification and quantity estimation," in *Proc. SIGGRAPH Asia Tech. Briefs*, 2012, p. 29.

[21] G. M. Farinella, D. Allegra, and F. Stanco, "A benchmark dataset to study the representation of food images," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 584–599.

[22] A. Meyers *et al.*, "Im2Calories: Towards an automated mobile vision food diary," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1233–1241.

[23] A. Mariappan *et al.*, "Personal dietary assessment using mobile devices," in *Proc. SPIE, Comput. Imag. VII*, vol. 7246, Feb. 2009, Art. no. 72460Z.

[24] H. Hoashi, T. Joutou, and K. Yanai, "Image recognition of 85 food categories by feature fusion," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, 2010, pp. 296–301.

[25] T. Miyazaki, G. C. de Silva, and K. Aizawa, "Image-based calorie content estimation for dietary assessment," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2011, pp. 363–368.

[26] X. Wang, D. Kumar, N. Thome, M. Cord, and F. Precioso, "Recipe recognition with large multimodal food dataset," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jun./Jul. 2015, pp. 1–6.

[27] A. Singla, L. Yuan, and T. Ebrahimi, "Food/non-food image classification and food categorization using pre-trained googlenet model," in *Proc. 2nd Int. Workshop Multimedia Assist. Dietary Manage.*, 2016, pp. 3–11.

[28] J. Chen and C.-W. Ngo, "Deep-based ingredient recognition for cooking recipe retrieval," in *Proc. ACM Multimedia Conf.*, 2016, pp. 32–41.

[29] P. Pandey, A. Deepthi, B. Mandal, and N. B. Puhan, "FoodNet: Recognizing foods using ensemble of deep networks," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1758–1762, Dec. 2017.

[30] C. Termritthikun, P. Muneesawang, and S. Kanprachar, "NU-InNet: Thai food image recognition using convolutional neural networks on smartphone," *J. Telecommun., Electron. Comput. Eng.*, vol. 9, nos. 2–6, pp. 63–67, 2017.

[31] Y. Kawano and K. Yanai, "Real-time mobile food recognition system," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2013, pp. 1–7.

[32] G. Villalobos, R. Almaghrabi, P. Pouladzadeh, and S. Shirmohammadi, "An image procesing approach for calorie intake measurement," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, May 2012, pp. 1–5.

[33] M. H. Rahman *et al.*, "Food volume estimation in a mobile phone based dietary assessment system," in *Proc. 8th Int. Conf. Signal Image Technol. Internet Based Syst. (SITIS)*, Nov. 2012, pp. 988–995.

[34] M. Bosch, F. Zhu, N. Khanna, C. J. Boushey, and E. J. Delp, "Combining global and local features for food identification in dietary assessment," in *Proc. 18th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2011, pp. 1789–1792.

[35] S. Yang, M. Chen, D. Pomerleau, and R. Sukthankar, "Food recognition using statistics of pairwise local features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2249–2256.

[36] J. Siswantoro, A. S. Prabuwono, A. Abdullah, and I. Bahari, "Automatic image segmentation using sobel operator and k-means clustering: A case study in volume measurement system for food products," in *Proc. Int. Conf. Sci. Inf. Technol. (ICSITech)*, Oct. 2015, pp. 13–18.

[37] P. Pouladzadeh, S. Shirmohammadi, and R. Al-Maghrabi, "Measuring calorie and nutrition from food image," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 8, pp. 1947–1956, Aug. 2014.

[38] F. Zhu, M. Bosch, N. Khanna, C. J. Boushey, and E. J. Delp, "Multiple hypotheses image segmentation and classification with application to dietary assessment," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 377–388, Jan. 2015.

[39] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[40] Y. He, C. Xu, N. Khanna, C. J. Boushey, and E. J. Delp, "Food image analysis: Segmentation, identification and weight estimation," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2013, pp. 1–6.

[41] F. Kong, H. He, H. A. Raynor, and J. Tan, "DietCam: Multi-view regular shape food recognition with a camera phone," *Pervasive Mobile Comput.*, vol. 19, pp. 108–121, May 2015.

[42] S. Fang, C. Liu, K. Tahboub, F. Zhu, E. J. Delp, and C. J. Boushey, "cTADA: The design of a crowdsourcing tool for online food image identification and segmentation," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, Aug. 2018, pp. 25–28.

[43] W. Shimoda and K. Yanai, "CNN-based food image segmentation without pixel-wise annotation," in *Proc. Int. Conf. Image Anal. Process.*, 2015, pp. 449–457.

[44] S. Inunganbi, A. Seal, and P. Khanna, "Classification of food images through interactive image segmentation," in *Proc. Asian Conf. Intell. Inf. Database Syst.*, 2018, pp. 519–528.

[45] M. M. Anthimopoulos, L. Gianola, L. Scarnato, P. Diem, and S. G. Mougiakakou, "A food recognition system for diabetic patients based on an optimized bag-of-features model," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 4, pp. 1261–1271, Jul. 2014.

[46] H. Hassannejad, G. Matrella, P. Ciampolini, I. De Munari, M. Mordonini, and S. Cagnoni, "Food image recognition using very deep convolutional networks," in *Proc. 2nd Int. Workshop Multimedia Assist. Dietary Manage.*, 2016, pp. 41–49.

[47] O. Beijbom, N. Joshi, D. Morris, S. Saponas, and S. Khullar, "Menu-match: Restaurant-specific food logging from images," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2015, pp. 844–851.

[48] F. Kong and J. Tan, "DietCam: Regular shape food recognition with a camera phone," in *Proc. Int. Conf. Body Sensor Netw. (BSN)*, May 2011, pp. 127–132.

[49] N. Tammachat and N. Pantuwong, "Calories analysis of food intake using image recognition," in *Proc. 6th Int. Conf. Inf. Technol. Elect. Eng. (ICITEE)*, Oct. 2014, pp. 67–70.

[50] Y. He, C. Xu, N. Khanna, C. J. Boushey, and E. J. Delp, "Analysis of food images: Features and classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 2744–2748.

[51] Y. Kawano and K. Yanai, "Foodcam-256: A large-scale real-time mobile food recognitionsystem employing high-dimensional features and compression of classifier weights," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 761–762.

[52] S. Christodoulidis, M. Anthimopoulos, and S. Mougiakakou, "Food recognition for dietary assessment using deep convolutional neural networks," in *Proc. Int. Conf. Image Anal. Process.*, 2015, pp. 458–465.

[53] K. Yanai and Y. Kawano, "Food image recognition using deep convolutional network with pre-training and fine-tuning," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, 2015, pp. 1–6.

[54] P. Pouladzadeh, S. Shirmohammadi, A. Bakirov, A. Bulut, and A. Yassine, "Cloud-based SVM for food categorization," *Multimedia. Tools Appl.*, vol. 74, no. 14, pp. 5243–5260, Jul. 2015.

[55] G. M. Farinella, D. Allegra, M. Moltisanti, F. Stanco, and S. Battiato, "Retrieval and classification of food images," *Comput. Biol. Med.*, vol. 77, pp. 23–39, Oct. 2016.

[56] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.

[57] L. Deng and D. Yu, "Deep learning: Methods and applications," *Found. Trends Signal Process.*, vol. 7, nos. 3–4, pp. 197–387, Jun. 2014.

[58] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[59] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[60] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[61] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.

[62] C. Liu, Y. Cao, Y. Luo, G. Chen, V. Vokkarane, and Y. Ma, "DeepFood: Deep learning-based food image recognition for computer-aided dietary assessment," in *Proc. Int. Conf. Smart Homes Health Telematics*, 2016, pp. 37–48.

[63] C. Liu *et al.*, "A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure," *IEEE Trans. Services Comput.*, vol. 11, no. 2, pp. 249–261, Apr. 2018.

[64] P. Pouladzadeh, P. Kuhad, S. V. B. Peddi, A. Yassine, and S. Shirmohammadi, "Food calorie measurement using deep learning neural network," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2016, pp. 1–6.

[65] Y. Kawano and K. Yanai, "Food image recognition with deep convolutional features," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Adjunct Publication*, 2014, pp. 589–593.

[66] *USDA Food Composition Databases*. Accessed: Dec. 1, 2018. [Online]. Available: https://ndb.nal.usda.gov/ndb/

[67] S. Gebhardt *et al.*, "USDA national nutrient database for standard reference, release 21," Dept. Agricult. Agricult. Res. Service, 2008. [Online]. Available: https://www.ars.usda.gov/research/publications/publication/?seqNo115=230658

[68] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2366–2374.

[69] S. Fang, F. Zhu, C. Jiang, S. Zhang, C. J. Boushey, and E. J. Delp, "A comparison of food portion size estimation using geometric models and depth images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 26–30.

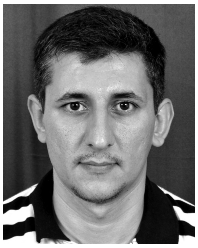
[70] S. Fang, C. Liu, F. Zhu, E. J. Delp, and C. J. Boushey, "Single-view food portion estimation based on geometric models," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2015, pp. 385–390.

[71] Y. Yue, W. Jia, and M. Sun, "Measurement of food volume based on single 2-D image without conventional camera calibration," *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, Aug./Sep. 2012, pp. 2166–2169.

[72] F. Zhu *et al.*, "The use of mobile devices in aiding dietary assessment and evaluation," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 4, pp. 756–766, Aug. 2010.

[73] J. Noronha, E. Hysen, H. Zhang, and K. Z. Gajos, "Platemate: Crowdsourcing nutritional analysis from food photographs," in *Proc. 24th Annu. ACM Symp. User Interface Softw. Technol.*, 2011, pp. 1–12.

[74] J. Shang *et al.*, "A mobile structured light system for food volume estimation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV)*, Nov. 2011, pp. 100–101.
[75] I. Woo, K. Otsmo, S. Kim, D. S. Ebert, E. J. Delp, and C. J. Boushey, "Automatic portion estimation and visual refinement in mobile dietary assessment," *Proc. SPIE*, vol. 7533, Jan. 2010, Art. no. 75330O.
[76] C. K. Martin, S. Kaya, and B. K. Gunturk, "Quantification of food intake using food image analysis," in *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, Sep. 2009, pp. 6869–6872.
[77] W. Jia *et al.*, "Imaged based estimation of food volume using circular referents in dietary assessment," *J. Food Eng.*, vol. 109, no. 1, pp. 76–86, 2012.
[78] J. Chae *et al.*, "Volume estimation using food specific shape templates in mobile image-based dietary assessment," *Proc. SPIE*, vol. 7873, Feb. 2011, Art. no. 78730K.
[79] C. Xu, Y. He, N. Khannan, A. Parra, C. Boushey, and E. Delp, "Image-based food volume estimation," in *Proc. 5th Int. Workshop Multimedia Cooking Eating Activities*, New York, NY, USA, 2013, pp. 75–80.
[80] C. Xu, Y. He, N. Khanna, C. J. Boushey, and E. J. Delp, "Model-based food volume estimation using 3D pose," in *Proc. 20th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2013, pp. 2534–2538.
[81] Y. Yang, W. Jia, T. Bucher, H. Zhang, and M. Sun, "Image-based food portion size estimation using a smartphone without a fiducial marker," *Public Health Nutrition*, pp. 1–13, Apr. 2018. doi: 10.1017/S136898001800054X.
[82] W. Jia *et al.*, "Accuracy of food portion size estimation from digital pictures acquired by a chest-worn camera," *Public Health Nutrition*, vol. 17, no. 8, pp. 1671–1681, 2014.
[83] J. Dehais, M. Anthimopoulos, S. Shevchik, and S. Mougiakakou, "Two-view 3D reconstruction for food volume estimation," *IEEE Trans. Multimedia*, vol. 19, no. 5, pp. 1090–1099, May 2017.
[84] M. Puri, Z. Zhu, Q. Yu, A. Divakaran, and H. Sawhney, "Recognition and volume estimation of food intake using a mobile device," *Appl. Comput. Vis. (WACV)*, Dec. 2009, pp. 1–8.
[85] S. Fang *et al.*, "Single-view food portion estimation: Learning image-to-energy mappings using generative adversarial networks," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 251–255.
[86] M. A. Subhi, S. H. M. Ali, A. G. Ismail, and M. Othman, "Food volume estimation based on stereo image analysis," *IEEE Instrum. Meas. Mag.*, vol. 21, no. 6, pp. 36–43, 2018.
[87] Y. Liang and J. Li. (2017). "Deep learning-based food calorie estimation method in dietary assessment." [Online]. Available: https://arxiv.org/abs/1706.04062



MOHAMMED AHMED SUBHI received the M.Eng. degree from the Technical University of Melaka, Malaysia, in 2013. He is currently a Researcher with the Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia. His research interests include image processing, food image analysis, and deep learning.

SAWAL HAMID ALI received the M.Sc. and Ph.D. degrees from the University of Southampton, U.K., in 2004 and 2010, respectively. He is currently an Associate Professor with the Department of Electric, Electronic and System Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia. His current research interests include wearable systems, system-on-chip design, and pervasive computing.

MOHAMMED ABULAMEER MOHAMMED received the master's and Ph.D. degrees from University Utara Malaysia, in 2010 and 2017, respectively. He is currently a Consultant at the Al-Rafidain University College, Iraq, and also a Researcher with the School of Computing, Universiti Utara Malaysia.