Survey on food intake methods using visual technologies

Sudhir Kumar Dubey
University of Rostock
Rostock, Germany
sudhir.dubey@uni-rostock.de

Dimitri Kraft
Fraunhofer IGD
Rostock, Germany
dimitri.kraft@igd-r.fraunhofer.de

Nicola Drüeke
Fraunhofer IGD
Rostock, Germany
nicola.marlene.drueeke@igd-r.fraunhofer.de

Gerald Bieber
Fraunhofer IGD
Rostock, Germany
gerald.bieber@igd-r.fraunhofer.de

Figure 1: 3D structural and chromatic information of food products.

ABSTRACT

Assessing food intake is important for reasons of well-being, lifestyle, health, appearance, or fun. Particularly in the field of medicine, the intake of appropriate foods and quantities of food is considered elementary and always related to physical activity. Various food tracking techniques exist, ranging from pen-based, purchase-based, calorie-counting to camera-based systems. Here, it is important that automated systems can recognize ingredients and estimate quantities. Therefore, there are many camera-based systems, but they differ in terms of accuracy, speed or performance. This review provides an overview of existing technologies and describes new approaches in the area of volume-sensitive sensing methods using lidar and true-depth technologies.

CCS CONCEPTS

• **Applied computing → Health care information systems**; *Consumer health*; • **Computing methodologies** → *Artificial intelligence*.

KEYWORDS

survey, nutrition, food intake, neural network, computer vision, volume estimation, structure, sensor technology

iWOAR 2023, September 21–22, 2023, Lübeck, Germany

ACM ISBN 979-8-4007-0816-9/23/09.
https://doi.org/10.1145/3615834.3615839

ACM Reference Format:
Sudhir Kumar Dubey, Dimitri Kraft, Nicola Drüeke, and Gerald Bieber. 2023. Survey on food intake methods using visual technologies. In *8th international Workshop on Sensor-Based Activity Recognition and Artificial Intelligence (iWOAR 2023), September 21–22, 2023, Lübeck, Germany*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3615834.3615839

1 INTRODUCTION

Research on the subject of food recognition has grown increasingly important, particularly in light of the growing popularity of healthy eating habits. According to WHO, obesity has grown to epidemic proportions today and the rate has increased by four times between 1975 to 2016 for the age group of 5 to 19 years [50] [3]. Obesity has become a classic case of malnutrition, where more people are obese than underweight [4]. One of the main factors contributing to obesity is excessive calorie intake, which is associated with eating unhealthy foods and having an unbalanced diet.

Food provides the necessary nutrients and energy for our bodies to function correctly. It contains essential macro-nutrients like carbohydrates, proteins, and fats, as well as micro-nutrients such as vitamins and minerals. A systematic review of the dietary assessment methods shows among 24HR dietary recall method, Food Record method, and Food-frequency Questionnaires method, 24HR gives the most accurate results with a minimum margin of error [9]. 24HR implies that to assess food intake over the previous 24 hours. Though one can manually keep track of their nutrition intake but it is still a very time-consuming, tedious task, and have many assessment flaws. The assessment flaws comprise random error, which arises due to frequent change in day-to-day diet, and systematic error comes due to users' incompetence in the nutrition evaluation of individuals' diet. It is now unavoidable to utilize technology to aid an individual's day-to-day dietary intake. Thanks to advancements in AI, computer vision, and sensors technology that promises its end users a semi-automated and automated solution.

This literature survey critically examines the progressive development of visual detection methodologies within the past decade. It investigates the transition from traditional computer vision techniques to the prevailing deep learning approaches, emphasizing the emergence of explainable artificial intelligence (AI) and its contribution to comprehending the rationale behind Deep-Learning model outcomes. Moreover, this study explores how contemporary applications leverage advancements in sensor technology to enhance the precision of calorie estimation predictions.

To ensure a thorough literature survey, a methodology was implemented for selecting relevant articles, papers, and related sources. Two essential parameters, namely food recognition" and "volume estimation," were used as criteria for article selection within the period of 2012 to 2023, (Figure 2). Preference was given to papers addressing healthy eating, nutritional estimation, food recognition, and volume estimation. In order to collect a diverse range of sources, reputable academic databases, including PubMed, IEEE, Elsevier, MDPI, arXiv (Cornell Univerisity), and Scopus, were chosen. These databases offer extensive collections of scientific literature across various disciplines. To ensure inclusivity, articles published in English were considered, as English serves as the primary language for scientific communication. Both web-based and mobile-based articles were included to ensure a comprehensive analysis. A comprehensive search strategy was devised, incorporating relevant keywords and Boolean operators such as "food recognition," "volume estimation," "Food classification," "healthy eating," "nutritional estimation," and related terms. Search filters were applied to refine the results based on publication year and language. Initially, articles were screened based on titles and abstracts to assess their relevance to the research topic. The full-text evaluation was then conducted on potentially relevant articles to determine their suitability for inclusion in the literature survey. Articles that did not meet the selection criteria, that were unavailable in English or did not significantly contribute to the research topic were excluded. Relevant data, including key findings, methodologies, experimental setups, and conclusions, were extracted from the selected articles. The collected data were analyzed to identify common themes such as segmentation methodologies, feature extraction, classification methods, volume estimation methodologies and advancements in the field of sensors technology. We will be examining the evolution of popular techniques in visual food image recognition and calorie estimation over the last decade. Specifically analyzing the advancement in techniques related to segmentation, classification, volume estimation and calorie estimation, their relevance and effectiveness in developing robust recognition systems. The analysis aims to provide insights into the current state of these techniques and their applications in developing effective food image recognition systems.

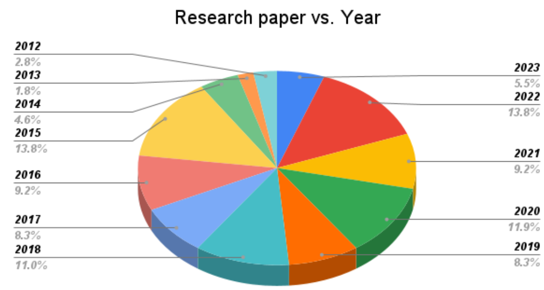

Figure 2: Distribution of Research Papers Utilized in the Literature Survey based on Relevance Criteria and Publication Year

2 RELATED WORK

2.1 Food intake Tracking

The state of the art in food intake assessment systems involves the development of innovative technologies and methodologies to accurately measure and analyze an individual's dietary intake. These systems aim to provide precise and reliable information about the types and quantities of foods consumed.

Quite common are digital and manual food diaries. Mobile applications, notepads, and digital food diaries enable individuals to record their food intake using their smartphones or other digital devices. These apps often incorporate image recognition, barcode scanning, voice input, or text entry to capture food data. They may also provide nutritional databases to facilitate accurate tracking and analysis of dietary intake.
The voice can be analysed by Natural Language Processing (NLP) techniques. NLP can be employed to extract relevant information from textual data such as food diaries, nutrition surveys, or social media posts related to food. NLP can help in categorizing food items, identifying portion sizes, and extracting nutrient information from unstructured text.

Another promising tool for food intake assessment are wearable devices that are equipped with sensors. These devices can track parameters such as chewing patterns, swallowing frequency, and even digestion-related data to estimate food intake. Some examples include smartwatches, neck-worn sensors, and ingestible sensors that can monitor the ingestion and digestion of specific food items.

Furthermore, Biomarker and metabolomics analysis involve the measurement of specific molecules or metabolites in biological samples to assess dietary intake. By analyzing blood, urine, or other bodily fluids, researchers can identify unique markers indicative of food consumption patterns. This approach provides objective and quantitative data on nutrient intake and metabolic responses to specific foods.

The most common and accurate assessment technology is the image-Based Food Recognition. The image-based food recognition systems utilize computer vision and machine learning algorithms to identify and quantify food items from images. These systems can be deployed in smartphone applications or smart cameras to capture images of meals, and then automatically analyze and estimate the nutritional content and portion sizes of the foods consumed. Advances in deep learning algorithms have significantly improved the accuracy of these systems. This technology also offers the potential for point-of-care food intake assessment and advises to the user.

In the field of visual food detection systems, the successful implementation of a comprehensive solution often involves addressing four key subtasks: image segmentation, image classification, volume estimation, and calorie estimation. The following sections aim to delve into these crucial parameters, examining their evolution over the past decade and current trends within the field.

2.2 Segmentation

For success of food visual detection application it is highly important to use an appropriate image segmentation approach that is robust, computationally inexpensive, and accurate. Image segmentation is a process of extraction or retrieving indispensable information from image. [15] provides an exclusive survey of different segmentation approaches that have been thus far, elaborating on different segmentation methods in food computing. Through segmenting an image, a label is assigned to every pixel such that pixels with identical labels share similar attributes.

Segmentation techniques can be categorized into two primary categories: layer-based segmentation methods and block-based segmentation methods. The block-based segmentation category further encompasses region-based segmentation and edge-based segmentation approaches. By employing these different techniques, researchers aim to achieve accurate and effective segmentation of objects or regions of interest within an image or scene.

The segmentation task has encountered escalating challenges owing to the multitude of food varieties inherent in different culinary traditions, disparate preparation methodologies, and the inherent occlusion phenomena associated with food items [88]. As a result, numerous algorithms such as graph-based segmentation, JSEG segmentation, GrabCut segmentation, and deep neural network-based segmentation have been extensively examined. Nevertheless, endeavors have been undertaken to assess performance by employing a fusion of diverse segmentation and object detection approaches.

An ingredient-based segmentation was attempted 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. Most typical are graph based representation, 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.

[39] employed the GrabCut algorithm for segmentation, utilizing a manually drawn bounding box to specify the region of interest. Alternatively, the study [69] proposed the utilization of the "Graph Cut" segmentation algorithm, which aims to partition the graph representation of an image into two sets (A and B) by considering dissimilarity in the weights (w) of the connecting edges (u, v) between adjacent pixels. This approach facilitates the extraction of desired food images while separating them from the background. It led to improved overall classification accuracy. In utilized "Graph Cut" segmentation technique, which yielded an impressive segmentation accuracy of 95%.

In a study conducted by author [74], a combination of convolutional neural networks (CNNs) and the GrabCut algorithm was employed to generate bounding boxes and detect food regions. The proposed approach achieved a mean average precision (MAP) of 49.9% in accurately identifying food regions within the images. In another investigation [23], participants were instructed to manually draw bounding boxes and select appropriate food tags from a provided list. The GrabCut technique was then applied to automatically segment the food based on the user-defined bounding boxes. This semi-automatic segmentation tool demonstrated effectiveness when applied to a large dataset of images. However, it is worth noting that user intervention is still required in this process.

In an alternative methodology presented in a separate study [53], various segmentation methods, such as image color, saturation, JSEG segmentation, and noise removal, were integrated to address the challenge of identifying multiple food items within an image. The research considered a total of 73 food classes, observed in a realistic food tray served in a canteen. The findings indicated a significant improvement in classification accuracy as a result of this approach. In a related study [95], the researchers implemented multiple segmentation hypotheses by assigning a class label to each pixel in an image. The segmentation process incorporated feedback from the classifier results, allowing for an estimation of the number of segments present in the image based on the confidence scores assigned to each segment. In a separate scholarly investigation [18], a novel approach combining JSEG segmentation with multiple object detectors, such as circle detector, whole image, and Deformable Part Model (DPM), was proposed. The study demonstrated that the overall classification accuracy could be enhanced by employing this combined approach compared to using the DPM model alone. [15] presents a comprehensive exploration of viable segmentation techniques employed in food image segmentation. The study not only provides an in-depth review of these techniques but also conducts a comparative analysis based on various parameters, including algorithm type, segmentation technique, dataset, and accuracy. By addressing the aforementioned aspects, the authors offer valuable insights into the research challenges associated with food recognition. The study by [15] contributes to the existing literature by shedding light on the advancements and limitations of segmentation techniques in the context of food image analysis, facilitating further advancements in the field of food recognition.

In conclusion, the field of segmentation has undergone a significant transformation, transitioning from traditional computer vision algorithms to the exploration of deep neural network-based approaches (Table 1). This shift can be attributed to the limitations of relying solely on a single algorithm, as it often fails to deliver satisfactory performance. Instead, the adoption of fusion-based approaches has shown promising results in improving segmentation accuracy. However, it is important to acknowledge that the integration of multiple algorithms can introduce computational challenges, necessitating efficient computational resources to handle the increased complexity. Nonetheless, the evolution towards neural network-based methodologies has paved the way for advancements in segmentation techniques, offering opportunities for further research and development in this domain.

Table 1: Segmentation approaches

References	Year	Segmentation Methodology	Performance in Evaluation
[53]	2012	JSEG segmentation, circle detector, whole image, and DPM.	Moderate segmentation accuracy 21% (top 1) and 45% (top 5)
[39]	2013	Bounding box and GrabCut Segmentation	Limited by manual selection of food items. Overall classification accuracy is improved.
[69]	2014	Graph cut segmentation	It achieves an overall segmentation accuracy of 95% and improves classification accuracy.
[17]	2015	Hough transform	Average food recognition of about 99% and 15 % in food estimation
[95]	2015	Multiple segmentation hypotheses with assigned segment confidence scores.	Outperforms tradition normalized cut method and improves overall classification accuracy.
[60]	2015	Deep Lab Model	Improves classification accuracy
[43]	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.
[74]	2015	Generated bounding box using CNNs and GrabCut.	It can detect bounding box regions around food items with a MAP of 49.9%
[93]	2016	Superpixel segmentation	The proposed mid-level approach improves classification accuracy to 70.84% from 66.12%
[18]	2017	A combination of color, saturation, JSEG, and noise removal.	The proposed segmentation provides better precision in contrast to other methods.
[23]	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.
[36]	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.
[58]	2018	Local variation segmentation	Improves classification accuracy and calorie value
[22]	2018	Wavelet kernel based WuandLi index fuzzy clustering	The model improves segmentation accuracy with 99.9%
[56]	2019	Canny edge detection, multi-scale segmentation, fast rejection of background pixels	Improved performance for image segmentation and feed forward neural network
[27]	2020	R-CNN	segmentation analysis achieved an intersection over the union (IoU) accuracy of 70%
[64]	2021	Mask R-CNN	Improve performance by 6.4%

Note: The table provides a brief overview of the segmentation approaches, methods and their performance.

2.3 Feature extraction

In terms of food classification, the diverse characteristics of food items, such as shape, color, and texture, pose a significant challenge. To achieve an optimal feature extraction process, it is crucial to extract informative visual data from food images. This can be accomplished through the utilization of general information descriptors, which encompass a collection of visual descriptors that capture essential features. Among these descriptors are Local Binary Patterns (LBP), Gabor filter, color information, and Scale Invariant Feature Transform (SIFT), which can be individually applied to extract image features [31]. However, to enhance the overall classification accuracy, the simultaneous implementation of multiple descriptors can be employed. This approach capitalizes on the collective power of these descriptors, leading to improved performance in food classification tasks.

In a comprehensive analysis [13], a study examined the application of Local Binary Patterns (LBP) and Scale Invariant Feature Transform (SIFT) features individually on a food image dataset. The results indicated that using SIFT features alone achieved an accuracy of 53%, while utilizing LBP features alone resulted in an accuracy of 46%. However, combining both features, alongside additional Gabor filter and color features, significantly improved the accuracy to 68%.

In another related study [11], features such as SIFT, LBP, color, Histogram of Oriented Gradients (HOG), and MR8 filter were extracted. A combination of these handcrafted features yielded an impressive accuracy of 77.4%. In a distinct study conducted by [76] adopted a bag-of-features (BoF) model incorporating SIFT-extracted features. The researchers trained a linear SVM image classifier to identify 11 different food classes, achieving an accuracy of 78%. Furthermore, [11] employed a diverse range of features, including SIFT, LBP, color, HOG, and MR8, and developed an SVM image classifier for food classification tasks.

2.4 Classification

Following the segmentation process and the extraction of pertinent features from an image or video, the subsequent phase involves the classification of food items depicted within the image. Image classification, a machine learning technique, is employed to assign objects present in the image into distinct classes or categories. The classification of food items can be broadly categorized into two main approaches: handcrafted features and deep learning techniques. Within the realm of deep learning, the classification process can further be distinguished based on the choice of architecture employed, such as the development of novel architectures or the utilization of transfer learning methodologies (Table 2). The author in [42] provide comprehensive insights into deep learning approaches for food recognition, exploring the pros and cons of different strategies. The choice between designing a new deep learning architecture, employing transfer learning, or utilizing image-based content prediction platforms (via APIs) for food classification is a subject of debate. Designing a new architecture requires extensive knowledge and intuition, transfer learning demands specific expertise, while platform-based solutions offer convenience and ease of deployment, albeit with a compromise in efficiency.

Among the traditional classifiers, Support Vector Machines (SVM), Multi-Kernel Learning, and K-Nearest Neighbour (KNN) are considered popular choices. However, it is important to note that the analysis of classifier performance cannot be solely based on overall accuracy. Extensive literature surveys have revealed that the performance of a classifier varies significantly depending on the specific set of features used for classification. Hence, considering the performance variation of classifiers across different feature sets is essential for a comprehensive evaluation. The popularity of different classifiers is depicted in Figure 3, revealing a notable trend in the research community. It is evident that deep learning approaches have rapidly gained popularity, rivaling traditional machine learning methods such as Support Vector Machines (SVM).

In a noteworthy study by [95], it was concluded that features trained on K-Nearest Neighbour (KNN) classifiers outperformed Support Vector Machine (SVM) classifiers when using the same set of features, including SIFT, color, and texture. The experimental results demonstrated that the KNN classifier yielded superior classification accuracy compared to the SVM classifier in this context. On the other hand, [8] conducted a study where they utilized the Bag of Features model with SIFT features for food classification. Their findings showcased a noteworthy classification accuracy of 78%, underscoring the effectiveness of this approach in accurately identifying different food classes.

An independent study conducted by [11] utilized a combination of Scale Invariant Feature Transform (SIFT), Local Binary Patterns (LBP), color, Histogram of Oriented Gradients (HOG), and MR8 features. These features were integrated into an SVM classifier for food recognition tasks. The experimental results provided evidence supporting the effectiveness of SVM in achieving accurate food recognition. Similarly, [21] employed a combination of color, Histogram of Oriented Gradients (HOG), modified Local Binary Patterns (LBP), and Gabor features. These features were utilized in conjunction with an SVM-derived classifier for food recognition purposes. The findings from this investigation further confirmed the strong performance of SVM in food recognition tasks.

The analysis of various studies conducted in this survey clearly indicates that deep learning approaches surpass traditional machine learning methods in terms of performance. Deep learning models consistently achieved accuracy rates exceeding 80% in the majority of cases examined. For instance, [12] and Hassannejad et al. [32] conducted detailed analyses utilizing Convolutional Neural Network (CNN) based models for food classification, both reporting highly favorable results.

In a different study, [87] explored the effectiveness of their Deep Convolutional Neural Network (DCNN) model using diverse food datasets. Their experimental findings demonstrated classification accuracies exceeding 70% for each dataset considered. Furthermore, [47] presented evidence supporting the efficacy of transfer learning in achieving high accuracy rates while also offering computational efficiency. Their work showcased the potential of transfer learning approaches to achieve comparable performance while being more computationally economical. In [71] proposed new framework for online class incremental learning approach for food classification and shows significant improvement in the performance on real time.

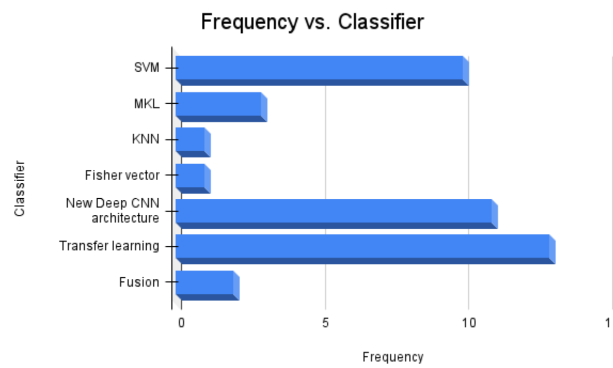

Figure 3: Frequency of Classifier Usage in Literature.

Table 2: Classification and feature extraction approaches

Reference	Year	Classification Methodology	Features	Performance, accuracy (MAP) (%)
[39]	2013	SVM	SURF, color	81.6
[69]	2014	SVM	GraphCut, color, shape, size and texture	95
[8]	2014	SVM	SIFT, color	78
[80]	2014	SVM	BoF, SFTA, color	70
[34]	2014	KNN	DCD, SIFT, MDSFIT, SCD	64.6
[12]	2014	CNN	-	56.4
[40]	2014	Deep Convolution and Fisher vector	-	72.3
[68]	2015	SVM	Gabor, color	94.5
[16]	2015	SVM	LBP, color	82.2
[87]	2015	DCNN	-	70.4
[16]	2015	CNN	-	84.9
[87]	2015	DCNN	-	78.8
[25]	2016	SVM	SIFT, PRICoLBP, Bag of textons	75.74
[32]	2016	Inception V3	-	88.3
[46]	2016	DeepFood	-	77.4
[67]	2016	Deep Neural Network	-	99
[46]	2016	DeepFood	-	76.3
[82]	2017	NU-InNet1.0	-	68.7
[81]	2017	CNN	-	75.2
[47]	2018	DCNN and EDGE computing	-	77
[83]	2018	CNN and VGG19	-	82
[65]	2019	DCNN	-	91.3
[59]	2019	CNN	Color, texture, SIFT, SURF	94.5
[28]	2020	CNN, Inceptopn-V3, GoogLeNet	LBP, color, SURF, geometry	93
[57]	2020	ANN	Color,texture, Local neighbourhood pixel feature	95.9
[19]	2021	CNN	-	85
[79]	2021	MobileNetV3	-	80.8
[55]	2023	PRENet	-	91.13

Note: The table provides a brief overview about classification, feature approaches, and their performances.

2.5 Volume and Calorie Estimation

It is imperative to acknowledge the crucial role that accurate volumetric measurement plays in determining the precise calorie content of food. The main challenge lies in calculating the exact quantity of each food item. In the realm of food intake assessment, various technologies are utilized to collect volume-related information, with the primary objective of offering precise and unbiased measurements of food portion sizes.

Image analysis techniques, in conjunction with computer vision algorithms, offer the potential to estimate the volume of food by analyzing images or videos of a given meal. This particular approach involves several steps, including the detection of individual food items' dimensions, estimation of their three-dimensional shapes, and subsequent calculation of the volume utilizing either known or approximated density values. Notably, advanced computer vision algorithms, such as deep learning models, have demonstrated promising outcomes in accurately estimating the volume of food based on visual data (Table 3).

In a study conducted by [92], the researchers employed a method that involved pixel counting to determine the food volume subsequent to segmenting the food items on the plate. However, this approach did not achieve the desired level of accuracy in estimating the food volume. On the other hand, [30] proposed an alternative technique that generated a three-dimensional point cloud by leveraging depth maps, segmentation masks, and camera parameters. They subsequently approximated the volume through the utilization of a point cloud-to-volume algorithm. The reported results indicated an average Mean Absolute Percentage Error (MAPE) of 46.32% for 16 test foods and an average MAPE of 36.90% for 6 combined meals.

The application of advanced technologies, such as 3D modeling and depth sensing, involves the utilization of depth cameras or depth-sensing modules to capture comprehensive three-dimensional data concerning the food and its surrounding environment. Through the analysis of this depth data, these systems enable the reconstruction of the shape and volume of food items present on a plate or within a container. These techniques prove particularly beneficial when estimating the volume of food items that possess irregular shapes or are stacked in a non-uniform manner.

[73] developed the "GoCARB" application, which leverages 3D modeling techniques to estimate the carbohydrate content of food. By employing this approach, they were able to accurately determine the volume of food items, thereby facilitating more precise carbohydrate content estimation. Another study by [84] utilized depth map-based images captured from smartphones, integrating multiple images taken from different angles. Their approach exhibited favorable results, yielding a volume estimation error margin within 10%, which is considered highly satisfactory. The utilization of depth-based information in volume estimation demonstrates superior accuracy compared to techniques that rely solely on image-based analysis. Furthermore, [66] conducted a study exploring the feasibility of employing wearable cameras for the purpose of determining macronutrient content. The researchers recorded an error rate of 34% in their experiment, suggesting that wearable cameras hold potential for capturing relevant visual data to estimate macronutrients, albeit with some room for improvement.

The recording and evaluation of food should be based on input data that is as accurate as possible. Still images taken with a camera lens contain less information than a stereoscopic camera shot. Videos taken from different angles provide data that can be used for 3D reconstruction. Images displayed only in the three RGB channels also contain only a small amount of data compared to a multi- or hyperspectral image. A possible compromise is to use existing, widespread hardware. Today's market provides smartphones with depth cameras and integrated AI-processors that can be easily used to obtain a handy, mobile food assessment system. In the study conducted by Steinbrener et al. (2023), the authors employed monocular videos utilizing RGB video frames from smartphones to extract volumetric information. Their objective was to achieve accurate classification, and their approach yielded a classification accuracy of 95%. Furthermore, they assessed the volume prediction performance on untrained objects, resulting in a mean absolute percentage error of 16%.

The employment of real-time 3D reconstruction, coupled with deep learning synthesis, has shown notable advancements in volume estimation compared to previous methodologies. To address challenges such as occlusion and scale ambiguity, depth sensing techniques have been proposed, integrating depth sensing capabilities with artificial intelligence (AI) methodologies. [48] introduced an approach that utilized 3D point cloud completion from single RGB and depth images, achieving a volume estimation error of 15.32%. Their methodology successfully tackled occlusion-related issues, enabling more accurate volume estimation. Similarly, [52] employed a mobile Structured Light System (SLS) to measure food volume and portion size. Experimental results demonstrated an enhancement of approximately 40% in volume and portion size measurement accuracy when compared to manual calculations. The incorporation of SLS technology effectively addressed scale ambiguity, leading to improved precision in volume estimation.

In another investigation, [20] utilized a two-view 3D reconstruction technique to estimate food volume, yielding an average error of less than 10% per dish. By leveraging the advantages of multiple views, their approach achieved accurate volume estimation results. Additionally, [49] proposed a network architecture that concurrently performed geometric understanding (including depth prediction and 3D plane estimation) and semantic prediction on a single food image, achieving an absolute relative error of 5.6%. This integrated approach showcased significant progress in volume estimation accuracy. Furthermore, recent studies by [90] and [62] explored the utilization of Generative Adversarial Networks (GANs) to complete the portions of images that were occluded, effectively addressing a major challenge in food volume estimation. By employing GAN-based approaches, these studies successfully tackled occlusion-related issues, leading to more accurate and comprehensive volume estimation results.

The advent of technologies aimed at quantifying the volume of food intake presents innovative approaches in assessing and monitoring dietary consumption. These techniques, when integrated with nutritional databases or algorithms, offer valuable insights into caloric intake, portion sizes, and adherence to dietary guidelines. The ongoing advancements in these technologies hold great potential for developing more convenient and accurate methods for assessing food intake. Numerous applications, such as FoodCam,

Table 3: Methods for Volume Estimation

References	Year	Methodology	Accuracy (Error) %	Application
[33]	2013	3D reconstruction	(-11)	-
[86]	2013	Multi-view 3D reconstruction	-10	-
[37]	2013	Wearable camera, shaped based approach	(< 30)	-
[69]	2014	Multi-images	(<10)	-
[61]	2015	Depth camera and reconstruction with CNN	-	Im2Calories
[92]	2015	Pixel counting	85	Snap-n-Eat
[24]	2015	Single view 3D reconstruction	(<6)	-
[73]	2016	3D model and segmentation	(-18.7)	GoCarb
[20]	2017	3D reconstruction	(<9.8)	-
[77]	2018	GAN to map food energy distribution	(<10.89)	-
[45]	2018	Stereo images analysis	(<8.5)	-
[83]	2018	Single view	-	Calpal
[29]	2019	CNN	(-12.3)	MUSEFood
[48]	2019	3D point cloud from RGB and depth image	(-15.32)	-
[84]	2020	CNN	(-10)	-
[35]	2020	CNN, Depth sensing	(-12.2)	-
[91]	2021	3D recontruction	(-5.23)	-
[89]	2021	Single image, MobinetV2	(-11.6)	-
[38]	2022	RCNN	90.9	-
[94]	2023	Multi-layer superpixel	(<34.2)	-
[75]	2023	Monocular video, CNN	(<16)	-

Note: The table provides a brief overview about volume estimation approaches and their performances.

FoodLog, and MyDietCam, have been developed to facilitate the tracking and monitoring of nutritional intake [5] [72] [78] [61]. These applications are available on both smartphone-based platforms and web-based interfaces, enabling users to conveniently record and analyze their dietary habits. By leveraging these tools, individuals can gain a better understanding of their nutritional intake and make informed decisions regarding their dietary choices. In their comprehensive study, [7] delve into the intricacies of mobile application-based volume and calorie estimation techniques. The authors provide a thorough examination of various methodologies employed in these applications, shedding light on the advancements and challenges associated with estimating food volume and caloric content using mobile platforms.

3 DATASET

It is essential to have access to high-quality datasets that can support various tasks such as food recognition, localization, detection, and categorization. In this review paper, we aim to explore and analyze a comprehensive range of food datasets that have been developed and made available to the research community. Table 4 provides a summary of some of the noteworthy datasets that have been widely utilized for food recognition problems.

One of the pioneering datasets in this domain is the Pittsburgh Fast-food Image Dataset (PFID) [10]. Released in 2009, it features 101 different dishes from popular fast-food chains, which have been grouped into 7 major fast-food categories. This dataset addresses the challenge of variations in fillings that can lead to confusion between similar food items. Another notable dataset is UEC-Food 100 [54], which focuses on Japanese cuisine and includes 100 distinct food categories. To expand the variety of food categories, UEC-Food 256 was introduced, encompassing both Japanese and international cuisines. These datasets present the added difficulty of dealing with single and multiple food items [41].

In the NTU-FOOD dataset [26], the emphasis is on multiethnic food categories, with each of the 50 categories represented by 100 photos. These images are sourced from user smartphones and online web collections, providing a realistic representation of the foods. To capture a wide range of popular food categories, the Food-101 dataset includes the 101 most popular food categories from the foodspotting.com website. This dataset boasts a relatively large collection of 101,000 food photos, enabling comprehensive analysis and exploration. Another dataset, UPMC Food-101 [85], shares the same 101 food categories as Food-101 but adopts a different approach. The class names were combined with the term "recipe" during data collection, using Google Search Engine results to curate the categories.

The FooDD dataset offers a diverse collection of food categories captured using various cameras and lighting setups [70]. It includes separate collections for single and mixed food portions, which can be valuable for studying portion estimation and meal composition. The VIREO Food-172 dataset [2] distinguishes itself by combining labeled ingredients and 172 Chinese food categories. This dataset provides a comprehensive perspective by including both food categories and detailed ingredient information. Other notable datasets, such as FOOD-5K [1], ECUST [44], Madima [6], ChineseFoodNet [14], and ChinaFood-100 [51], offer unique contributions to the field, ranging from separating food from non-food photos to categorizing Chinese foods and incorporating dietary requirements and nutritional information.

4 CHALLENGES AND PROPOSED FOOD ASSESSMENT SYSTEM

Image analysis for food identification can achieve a reasonable level of reliability, but it is important to consider certain factors that can impact its accuracy. One major challenge is the variation in food appearance, resulting from cooking methods, serving styles, ingredients, and lighting conditions. This variation makes it difficult for the model to generalize well across different instances of the same food item. Another challenge arises from occlusions and clutter in food images, which can hinder accurate identification of individual food items or components. Limited viewpoints during image capture can also affect reliability, as models trained on specific viewpoints may struggle to generalize to others. Furthermore, real-world conditions, including varying lighting conditions and image quality, pose challenges to accurate food recognition. To address these challenges, continuous model improvement through user feedback, ongoing training with new data, and algorithmic advancements are crucial for enhancing reliability in food recognition systems. The accuracy of an image analysis model for food identification is typically measured using metrics such as precision, recall, and F1 score. These metrics evaluate the model's ability to correctly classify food items, considering both true positive and false positive/negative predictions.

Table 4: Popular Food dataset

Name	Year	Food categories, n	Total number of images, n	Food items in each image	Cuisine
Pittsburgh Fast-food Image Dataset (PRD)	2009	61	1089	Single	Fast food
UEC-Food 100	2012	100	10,000	Single and multiple	Japanese
NTU-FOOD	2012	50	5000	Single	Multiethnic
UNICT-FD889	2014	889	3583	Single	Multiethnic
Food-101	2014	101	101,000	Single	Multiethnic
UEC-Food 256	2014	256	31,397	Single and multiple	Multiethnic
Ambient Kitchen	2014	12	1800	Single and multiple	Multiethnic
UPMC Food-101	2015	101	90,840	Single	Multiethnic
FooDD	2015	23	3000	Single and multiple	Multiethnic
UNICT-FDI 200	2016	1200	4754	Single and multiple	Multiethnic
UNIMIB 2016	2016	73	1027	Multiple	Italian
EgocentricFood	2016	9	5038	Multiple	Multiethnic
VIREO Food-172	2016	172	110,241	Single and multiple	Chinese
FOOD-5K	2016	2	5000	Multiple	Multiethnic
FOOD-11	2016	11	16,643	Multiple	Multiethnic
ECUST Food Datase1	2017	19	2978	Single and multiple	Multiethnic
Madima 2017	2017	21	21,807	Multiple	Central European
Food524DB	2017	524	247,636	Single and multiple	Multiethnic
ChineseFoodNet	2017	208	192,000	Single and multiple	Chinese

Note: The table provides a brief overview of popularly used datasets for food recognition.

In order to enhance the capabilities of mobile phone-based assessments, we propose the real time integration of RGB combined with either True Depth or Time-of-Flight (ToF)/LIDAR Camera technologies. Many modern smartphones already incorporate distance-sensitive camera technologies. The True Depth camera system comprises various components that work in tandem to enable advanced depth recognition. A key component is the dot projector, which emits an infrared dot grid onto the surface being captured, such as a plate or the user's face. An infrared camera captures this dot pattern, aided by a flood illuminator that ensures accurate depth perception. These components collectively generate a depth map of the captured image.

A Time-of-Flight (ToF) camera, on the other hand, is a depth-sensing camera that measures the distance between the camera and objects within its field of view using the principle of time-of-flight. It calculates the time taken for a light signal to travel from the camera to the object and back, thereby determining the distance. ToF cameras emit a light signal, typically in the infrared spectrum, towards the scene and measure the round-trip time for the light to be reflected back to the camera's sensor. By measuring this time, the camera can accurately calculate the distance between itself and the object. ToF cameras generally consist of an emitter, which emits the light signal, and a sensor that detects the reflected light. The emitter may employ an infrared laser or LED, while the sensor can be a dedicated pixel array or a complementary metal-oxide-semiconductor (CMOS) image sensor.

The ToF camera captures a depth map of the scene by measuring the distances to multiple points within its field of view. This depth information finds applications in various domains, including 3D scanning, gesture recognition, augmented reality (AR), and robotics. While ToF cameras typically utilize an infrared light source and sensor for distance calculation, LIDAR cameras employ laser light for measurement. However, for our purposes, we define the term "ToF sensor" as a synonym for both infrared and laser sensors, as they operate on similar principles.

5 DISCUSSION

The problem of overweight in our society just started. We know, that many of today´s kids are also very overweight, and overweight children tend to grow up to be very overweight adults [63]. Food intake assessment is a tremendous tool for health-conscious, overweight, or obese people. The gain in reliability of assessment tools enables a new service that our society is likely to need in the future, namely, complete control of diet. Many people with diet problems do not adjust their food intake and physical activity. We assume that a food-incapacitation of the person might be an applicable strategy. The smartphone allows or forbids what and how much can be eaten. It also calculates the costs and specifies what the user may or may not do. The user thus becomes, sorry, the recipient of orders from a technical assistant. After all, it will only be for his own good.

During the past decade, researchers have made remarkable progress in achieving high accuracy in food recognition tasks. However, several challenges persist, particularly in addressing occlusion and ensuring precise volumetric estimation. Stereo-based techniques heavily rely on feature matching between frames. Nevertheless, these approaches encounter limitations when attempting to reconstruct 3D models and estimate volume, especially if the food surface lacks distinctive characteristics or texture. Deep learning approaches, although promising, face inherent difficulties in accurately estimating food volume or weight using deep neural networks, resulting in relatively lower accuracy rates compared to alternative methods. These approaches have primarily been employed in controlled experimental settings, and there is a dearth of deep learning-based applications in this domain. Model-based volume estimation techniques have their own limitations as well. Presently, fixed geometrical shapes such as cones, cubes, and spheres are commonly utilized, and the accuracy of the estimation heavily relies on the adequacy of the fit with these shapes.

The advent of deep learning models such as YOLO, Fast-CNN, and mask-RCNN, designed for segmentation, feature extraction, and classification, in conjunction with transfer learning, has opened up new avenues for researchers to focus on enhancing accurate volumetric estimation of food. This paper explores various approaches, such as 3D reconstruction from single images, 3D reconstruction using stereoscopic views, and leveraging depth sensors, all of which exhibit an impressive amount of ongoing research aimed at improving the accuracy of this system. Following a comprehensive review and comparative analysis of various aspects pertaining to food recognition systems, and considering the current state of knowledge, it is evident that an approach utilizing depth cameras in conjunction with computer vision techniques holds great potential for developing a robust and computationally efficient system. The proposed system, which leverages the built-in LiDAR sensor in smartphones for real-time food recognition, has the potential to assist individuals in achieving a more accurate nutritional assessment of their diet. An alternative approach that shows promise involves leveraging Lidar sensor technology in conjunction with the integration of computer vision algorithms and deep learning techniques for feature extraction as the foundational framework for a recognition system. This approach has the potential to significantly improve accuracy by facilitating the identification and learning of global features. Nonetheless, it is imperative to undertake further research endeavors to thoroughly investigate these hypotheses and perform additional validation studies in order to establish the robustness and reliability of these findings.

6 CONCLUSION AND OUTLOOK

In conclusion, this paper has provided a comprehensive analysis of the progress made in the field of food recognition over the past decade. We have presented an extensive overview of various computer vision algorithms, specifically highlighting the applicability and viability of various segmentation techniques within the current landscape. Furthermore, we have underscored the importance of different feature extraction approaches and elucidated their practical implications.

Volume estimation not only enhances the assessment of the amount of food but also provides more and relevant features for food recognition and identification such as sticking out french fries, rice etc. Looking towards the future, the emergence of novel technologies holds immense potential for augmenting the reliability and user-friendliness of food recognition systems, ultimately leading to enhanced accuracy. These advancements are poised to yield substantial benefits for future medical practitioners, empowering them to provide more informed advice to their patients. Moreover, these systems can play a pivotal role in mitigating the challenges associated with malnutrition, while also serving as invaluable aids in nursing homes by facilitating efficient monitoring and care. We expect practical solutions in the future, perhaps also in a wearable, unobtrusive tracker for food intake, which unfortunately could also tell the user what to eat or not to eat.

REFERENCES

[1] 2016. *MADiMa '16: Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management* (Amsterdam, The Netherlands). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/2986035

[2] 2016. *MM '16: Proceedings of the 24th ACM International Conference on Multimedia* (Amsterdam, The Netherlands). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/2964284

[3] 2021. *Obesity and overweight. World Health Organization.* Retrieved May 08, 2023 from https://www.who.int/europe/news/item/03-03-2023-childhood-obesity--five-facts-about-the-who-european-region

[4] 2021. *Obesity, World Health Organization.* Retrieved May 14, 2023 from https://www.who.int/health-topics/obesity

[5] Kiyoharu Aizawa, Yuto Maruyama, He Li, and Chamin Morikawa. 2013. Food Balance Estimation by Using Personal Dietary Tendencies in a Multimedia Food Log. *IEEE Transactions on Multimedia* (2013), 2176–2185. https://doi.org/10.1109/TMM.2013.2271474

[6] Dario Allegra, Marios Anthimopoulos, Joachim Dehais, Ya Lu, Filippo Stanco, Giovanni Maria Farinella, and Stavroula Mougiakakou. 2017. A multimedia database for automatic meal assessment systems. In *New Trends in Image Analysis and Processing–ICIAP 2017: ICIAP International Workshops, WBICV, SSPandBE, 3AS, RGBD, NIVAR, IWBAAS, and MADiMa 2017, Catania, Italy, September 11-15, 2017, Revised Selected Papers 19*. Springer, 471–478.

[7] Lameck Mbangula Amugongo, Alexander Kriebitz, Auxane Boch, and Christoph Lütge. 2022. Mobile Computer Vision-Based Applications for Food Recognition and Volume and Calorific Estimation: A Systematic Review. *Healthcare* 11, 1 (dec 2022), 59. https://doi.org/10.3390/healthcare11010059

[8] Marios M. Anthimopoulos, Lauro Gianola, Luca Scarnato, Peter Diem, and Stavroula G. Mougiakakou. 2014. A Food Recognition System for Diabetic Patients Based on an Optimized Bag-of-Features Model. *IEEE Journal of Biomedical and Health Informatics* 18, 4 (2014), 1261–1271. https://doi.org/10.1109/JBHI.2014.2308928

[9] Regan L Bailey. 2021. Overview of dietary assessment methods for measuring intakes of foods, beverages, and dietary supplements in research studies. *Current Opinion in Biotechnology* (2021), 91–96. https://doi.org/10.1016/j.copbio.2021.02.007

[10] Magdy Bayoumi. 2009. ICIP'09: Proceedings of the 16th IEEE international conference on Image processing. (2009). https://doi.org/10.5555/1818719

[11] Oscar Beijbom, Neel Joshi, Dan Morris, Scott Saponas, and Siddharth Khullar. 2015. Menu-Match: Restaurant-Specific Food Logging from Images. In *2015 IEEE Winter Conference on Applications of Computer Vision*. 844–851. https://doi.org/10.1109/WACV.2015.117

[12] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. 2014. Food-101 – Mining Discriminative Components with Random Forests. In *Computer Vision – ECCV 2014*. Springer International Publishing, 446–461. https://doi.org/10.1007/978-3-319-10599-4_29

[13] Mei-Yun Chen, Yung-Hsiang Yang, Chia-Ju Ho, Shih-Han Wang, Shane-Ming Liu, Eugene Chang, Che-Hua Yeh, and Ming Ouhyoung. 2012. Automatic Chinese food identification and quantity estimation. In *SIGGRAPH Asia 2012 Technical Briefs*. ACM. https://doi.org/10.1145/2407746.2407775

[14] Xin Chen, Yu Zhu, Hua Zhou, Liang Diao, and Dongyan Wang. 2017. ChineseFoodNet: A large-scale Image Dataset for Chinese Food Recognition. (2017). arXiv:1705.02743 [cs.CV]

[15] Megha Chopra and Archana Purwar. 2021. Recent Studies on Segmentation Techniques for Food Recognition: A Survey. *Archives of Computational Methods in Engineering* (2021), 865–878. https://doi.org/10.1007/s11831-021-09598-3

[16] Stergios Christodoulidis, Marios Anthimopoulos, and Stavroula Mougiakakou. 2015. Food Recognition for Dietary Assessment Using Deep Convolutional Neural Networks. In *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops*. Springer International Publishing, 458–465. https://doi.org/10.1007/978-3-319-23222-5_56

[17] Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2015. Food Recognition and Leftover Estimation for Daily Diet Monitoring. In *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops*. Springer International Publishing, 334–341. https://doi.org/10.1007/978-3-319-23222-5_41

[18] Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2017. Food Recognition: A New Dataset, Experiments, and Results. *(IEEE) Journal of Biomedical and Health Informatics* 21 (2017), 588–598. https://doi.org/10.1109/jbhi.2016.2636441

[19] Leonardo Cornejo, Rosa Urbano, and Willy Ugarte. 2021. Mobile Application for Controlling a Healthy Diet in Peru Using Image Recognition. In *2021 30th Conference of Open Innovations Association FRUCT*. IEEE. https://doi.org/10.23919/fruct53335.2021.9599959

[20] Joachim Dehais, Marios Anthimopoulos, Sergey Shevchik, and Stavroula Mougiakakou. 2017. Two-View 3D Reconstruction for Food Volume Estimation. *IEEE Transactions on Multimedia* 19, 5 (2017), 1090–1099. https://doi.org/10.1109/TMM.2016.2642792

[21] Bruno Vieira Resende e Silva, Milad Ghiasi Rad, Juan Cui, Megan McCabe, and Kaiyue Pan. 2018. A Mobile-Based Diet Monitoring System for Obesity Management. *Journal of Health, Medical Informatics* 09, 02 (2018). https://doi.org/10.4172/2157-7420.1000307

[22] W R Sam Emmanuel and S Jasmine Minija. 2018. Fuzzy clustering and Whale-based neural network to food recognition and calorie estimation for daily dietary assessment. (may 2018). https://doi.org/10.1007/s12046-018-0865-3

[23] Shaobo Fang, Chang Liu, Khalid Tahboub, Fengqing Zhu, Edward J. Delp, and Carol J. Boushey. 2018. cTADA: The Design of a Crowdsourcing Tool for Online Food Image Identification and Segmentation. In *2018 (IEEE) Southwest Symposium on Image Analysis and Interpretation (SSIAI)*. IEEE. https://doi.org/10.1109/ssiai.2018.8470358

[24] Shaobo Fang, Chang Liu, Fengqing Zhu, Edward J. Delp, and Carol J. Boushey. 2015. Single-View Food Portion Estimation Based on Geometric Models. In *2015 IEEE International Symposium on Multimedia (ISM)*. IEEE. https://doi.org/10.1109/ism.2015.67

[25] Giovanni Maria Farinella, Dario Allegra, Marco Moltisanti, Filippo Stanco, and Sebastiano Battiato. 2016. Retrieval and classification of food images. *Computers in Biology and Medicine* 77 (oct 2016), 23–39. https://doi.org/10.1016/j.compbiomed.2016.07.006

[26] Giovanni Maria Farinella, Dario Allegra, and Filippo Stanco. 2015. A benchmark dataset to study the representation of food images. In *Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part III*. Springer, 584–599.

[27] Charles N. C. Freitas, Filipe R. Cordeiro, and Valmir Macario. 2020. MyFood: A Food Segmentation and Classification System to Aid Nutritional Monitoring. In *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. 234–239. https://doi.org/10.1109/SIBGRAPI51738.2020.00039

[28] Pedro Furtado. 2020. Food Recognition: Can Deep Learning or Bag-of-Words Match Humans?. In *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies*. SCITEPRESS - Science and Technology Publications. https://doi.org/10.5220/0008893301020108

[29] Junyi Gao, Weihao Tan, Liantao Ma, Yasha Wang, and Wen Tang. 2019. MUSE-Food: Multi-sensor-based Food Volume Estimation on Smartphones. (2019). arXiv:1903.07437 [cs.CV]

[30] Alexandros Graikos, Vasileios Charisis, Dimitrios Iakovakis, Stelios Hadjidimitriou, and Leontios Hadjileontiadis. 2020. Single Image-Based Food Volume Estimation Using Monocular Depth-Prediction Networks. In *Universal Access in Human-Computer Interaction. Applications and Practice*. Springer International Publishing, 532–543. https://doi.org/10.1007/978-3-030-49108-6_38

[31] Zhenhua Guo, Lei Zhang, and David Zhang. 2010. A completed modeling of local binary pattern operator for texture classification. *IEEE transactions on image processing* 19, 6 (2010), 1657–1663. https://doi.org/10.1109/TIP.2010.2044957

[32] Hamid Hassannejad, Guido Matrella, Paolo Ciampolini, Ilaria De Munari, Monica Mordonini, and Stefano Cagnoni. 2016. Food Image Recognition Using Very Deep Convolutional Networks. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*. ACM. https://doi.org/10.1145/2986035.2986042

[33] Ye He, Chang Xu, Nitin Khanna, Carol J. Boushey, and Edward J. Delp. 2013. Food image analysis: Segmentation, identification and weight estimation. In *2013 IEEE International Conference on Multimedia and Expo (ICME)*. 1–6. https://doi.org/10.1109/ICME.2013.6607548

[34] Ye He, Chang Xu, Nitin Khanna, Carol J. Boushey, and Edward J. Delp. 2014. Analysis of food images: Features and classification. In *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE. https://doi.org/10.1109/icip.2014.7025555

[35] David Herzig, Christos T Nakas, Janine Stalder, Christophe Kosinski, Céline Laesser, Joachim Dehais, Raphael Jaeggi, Alexander Benedikt Leichtle, Fried-Michael Dahlweid, Christoph Stettler, and Lia Bally. 2020. Volumetric Food Quantification Using Computer Vision on a Depth-Sensing Smartphone: Preclinical Study. *JMIR mHealth and uHealth* 8, 3 (March 2020), e15294. https://doi.org/10.2196/15294

[36] Sanasam Inunganbi, Ayan Seal, and Pritee Khanna. 2018. Classification of Food Images through Interactive Image Segmentation. In *Intelligent Information and Database Systems*. Springer International Publishing, 519–528. https://doi.org/10.1007/978-3-319-75420-8_49

[37] Wenyan Jia, Hsin-Chen Chen, Yaofeng Yue, Zhaoxin Li, John Fernstrom, Yicheng Bai, Chengliu Li, and Mingui Sun. 2013. Accuracy of food portion size estimation from digital pictures acquired by a chest-worn camera. *Public Health Nutrition* 17, 8 (Dec. 2013), 1671–1681. https://doi.org/10.1017/s1368980013003236

[38] Prachi Kadam, Sharnil Pandya, Shraddha Phansalkar, Mayur Sarangdhar, Nayana Petkar, Ketan Kotecha, and Deepak Garg. 2022. FVEstimator: A novel food volume estimator Wellness model for calorie measurement and healthy living. *Measurement* 198 (July 2022), 111294. https://doi.org/10.1016/j.measurement.2022.111294

[39] Yoshiyuki Kawano and Keiji Yanai. 2013. Real-Time Mobile Food Recognition System. *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2013), 1–7.

[40] Yoshiyuki Kawano and Keiji Yanai. 2014. Food image recognition with deep convolutional features. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM. https://doi.org/10.1145/2638728.2641339

[41] Yoshiyuki Kawano and Keiji Yanai. 2014. FoodCam-256: A Large-scale Real-time Mobile Food RecognitionSystem employing High-Dimensional Features and Compression of Classifier Weights. *Proceedings of the 22nd ACM international conference on Multimedia* (2014). https://doi.org/10.1145/2647868.2654869

[42] Chairi Kiourt, George Pavlidis, and Stella Markantonatou. 2020. Deep learning approaches in food recognition. (2020). arXiv:2004.03357 [cs.CV]

[43] Fanyu Kong and Jindong Tan. 2011. DietCam: Regular Shape Food Recognition with a Camera Phone. In *2011 International Conference on Body Sensor Networks*. 127–132. https://doi.org/10.1109/BSN.2011.19

[44] Yanchao Liang and Jianhua Li. 2017. Computer vision-based food calorie estimation: dataset, method, and experiment. (2017). arXiv:1705.07632 [cs.CV]

[45] Yanchao Liang and Jianhua Li. 2018. Deep Learning-Based Food Calorie Estimation Method in Dietary Assessment. (2018). arXiv:1706.04062 [cs.CV]

[46] Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, and Yunsheng Ma. 2016. DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment. (2016). arXiv:1606.05675 [cs.CV]

[47] Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, Ma Yunsheng, Songqing Chen, and Peng Hou. 2018. A New Deep Learning-Based Food Recognition System for Dietary Assessment on An Edge Computing Service Infrastructure. *IEEE Transactions on Services Computing* (2018), 249–261. https://doi.org/10.1109/tsc.2017.2662008

[48] Frank P.-W. Lo, Yingnan Sun, Jianing Qiu, and Benny P. L. Lo. 2020. Point2Volume: A Vision-Based Dietary Assessment Approach Using View Synthesis. *IEEE Transactions on Industrial Informatics* 16, 1 (2020), 577–586. https://doi.org/10.1109/TII.2019.2942831

[49] Ya Lu, Thomai Stathopoulou, and Stavroula Mougiakakou. 2021. Partially Supervised Multi-Task Network for Single-View Dietary Assessment. In *2020 25th International Conference on Pattern Recognition (ICPR)*. 8156–8163. https://doi.org/10.1109/ICPR48806.2021.9412339

[50] Ingo Lütkebohle. 2021. *Obesity and overweight. World Health Organization*. Retrieved May 08, 2023 from https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

[51] Peihua Ma, Chun Pong Lau, Ning Yu, An Li, Ping Liu, Qin Wang, and Jiping Sheng. 2021. Image-based nutrient estimation for Chinese dishes using deep learning. *Food Research International* 147 (2021), 110437. https://doi.org/10.1016/j.foodres.2021.110437

[52] Sepehr Makhsous, Hashem M. Mohammad, Jeannette M. Schenk, Alexander V. Mamishev, and Alan R. Kristal. 2019. A Novel Mobile Structured Light System in Food 3D Reconstruction and Volume Estimation. *Sensors* (Jan. 2019). https://doi.org/10.3390/s19030564

[53] Yuji Matsuda, Hajime Hoashi, and Keiji Yanai. 2012. Recognition of Multiple-Food Images by Detecting Candidate Regions. In *2012 IEEE International Conference on Multimedia and Expo*. 25–30. https://doi.org/10.1109/ICME.2012.157

[54] Yuji Matsuda, Hajime Hoashi, and Keiji Yanai. 2012. Recognition of Multiple-Food Images by Detecting Candidate Regions. In *2012 IEEE International Conference on Multimedia and Expo*. 25–30. https://doi.org/10.1109/ICME.2012.157

[55] Weiqing Min, Zhiling Wang, Yuxin Liu, Mengjiang Luo, Liping Kang, Xiaoming Wei, Xiaolin Wei, and Shuqiang Jiang. 2023. Large Scale Visual Food Recognition. (2023). arXiv:2103.16107 [cs.CV]

[56] S. Jasmine Minija and W. R. Sam Emmanuel. 2019. Image segmentation and classification with application to dietary assessment using BMI-calorie calculator. *Computer Assisted Mechanics and Engineering Sciences* 26 (2019), 177–189. https://doi.org/10.24423/cames.239

[57] S. Jasmine Minija and W. R. Sam Emmanuel. 2020. Food recognition using neural network classifier and multiple hypotheses image segmentation. *The Imaging Science Journal* (feb 2020). https://doi.org/10.1080/13682199.2020.1742416

[58] S. Jasmine Minija and W.R. Sam Emmanuel. 2018. Image processing based Classification and Segmentation using LVS based Multi-Kernel SVM. In *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)*. 61–65. https://doi.org/10.1109/ICSSIT.2018.8748712

[59] S. Jasmine Minija and W.R. Sam Emmanuel. 2019. Image segmentation and classification with application to dietary assessment using BMI-calorie calculator. *Computer Assisted Methods in Engineering and Science* 26, 3–4 (2019), 177–189. https://doi.org/10.24423/cames.239

[60] Austin Myers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin Murphy. 2015. Im2Calories: Towards an Automated Mobile Vision Food Diary. In *2015 IEEE International Conference on Computer Vision (ICCV)*. 1233–1241. https://doi.org/10.1109/ICCV.2015.146

[61] Austin Myers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin Murphy. 2015. Im2Calories: Towards an Automated Mobile Vision Food Diary. In *2015 IEEE International Conference on Computer Vision (ICCV)*. 1233–1241. https://doi.org/10.1109/ICCV.2015.146

[62] Kizashi Nakano, Kiyoshi Kiyokawa, Daichi Horita, Keiji Yanai, Nobuchika Sakata, and Takuji Narumi. 2019. Enchanting Your Noodles: GAN-based Real-time Food-to-Food Translation and Its Impact on Vision-induced Gustatory Manipulation. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. https://doi.org/10.1109/vr.2019.8798336

[63] United Kingdom National Health Service. 2021. What can I do if my child is very overweight? (2021). https://www.nhs.uk/live-well/healthy-weight/childrens-weight/very-overweight-children-advice-for-parents/ last access: 30.05.2023.

[64] D. Park, J. Lee, J. Lee, and K. Lee. 2021. Deep Learning based Food Instance Segmentation using Synthetic Data. (2021). arXiv:2107.07191 [cs.CV]

[65] Seon-Joo Park, Akmaljon Palvanov, Chang-Ho Lee, Nanoom Jeong, Young-Im Cho, and Hae-Jeung Lee. 2019. The development of food image detection and recognition model of Korean food for mobile dietary management. *Nutrition Research and Practice* (2019). https://doi.org/10.4162/nrp.2019.13.6.521

[66] Claire Pettitt, Jindong Liu, Richard M. Kwasnicki, Guang-Zhong Yang, Thomas Preston, and Gary Frost. 2015. A pilot study to determine whether using a lightweight, wearable micro-camera improves dietary assessment accuracy and offers information on macronutrients and eating rate. *British Journal of Nutrition* 115, 1 (Nov. 2015), 160–167. https://doi.org/10.1017/s0007114515004262

[67] Parisa Pouladzadeh, Pallavi Kuhad, Sri Vijay Bharat Peddi, Abdulsalam Yassine, and Shervin Shirmohammadi. 2016. Food calorie measurement using deep learning neural network. In *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*. IEEE. https://doi.org/10.1109/i2mtc.2016.7520547

[68] Parisa Pouladzadeh, Shervin Shirmohammadi, Aslan Bakirov, Ahmet Bulut, and Abdulsalam Yassine. 2014. Cloud-based SVM for food categorization. *Multimedia Tools and Applications* 74, 14 (jun 2014), 5243–5260. https://doi.org/10.1007/s11042-014-2116-x

[69] Parisa Pouladzadeh, Shervin Shirmohammadi, and Abdulsalam Yassine. 2014. Using graph cut segmentation for food calorie measurement. In *2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. 1–6. https://doi.org/10.1109/MeMeA.2014.6860137

[70] Parisa Pouladzadeh, Abdulsalam Yassine, and Shervin Shirmohammadi. 2020. FooDD: Food Detection Dataset for Calorie Measurement Using Food Images. (2020). https://doi.org/10.21227/yvk7-qk38

[71] Siddeshwar Raghavan, Jiangpeng He, and Fengqing Zhu. 2023. Online Class-Incremental Learning For Real-World Food Classification. (2023). arXiv:2301.05246 [cs.CV]

[72] Daniele Ravì, Benny Lo, and Guang-Zhong Yang. 2015. Real-time food intake classification and energy expenditure estimation on a mobile device. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. 1–6. https://doi.org/10.1109/BSN.2015.7299410

[73] Daniel Rhyner, Hannah Loher, Joachim Dehais, Marios Anthimopoulos, Sergey Shevchik, Ransford Henry Botwey, David Duke, Christoph Stettler, Peter Diem, and Stavroula Mougiakakou. 2016. Carbohydrate Estimation by a Mobile Phone-Based System Versus Self-Estimations of Individuals With Type 1 Diabetes Mellitus: A Comparative Study. *Journal of Medical Internet Research* 18, 5 (May 2016), e101. https://doi.org/10.2196/jmir.5567

[74] Wataru Shimoda and Keiji Yanai. 2015. CNN-Based Food Image Segmentation Without Pixel-Wise Annotation. In *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops*. Springer International Publishing, 449–457. https://doi.org/10.1007/978-3-319-23222-5_55

[75] Jan Steinbrener, Vesna Dimitrievska, Federico Pittino, Frans Starmans, Roland Waldner, Jürgen Holzbauer, and Thomas Arnold. 2023. Learning metric volume estimation of fruits and vegetables from short monocular video sequences. *Heliyon* 9, 4 (apr 2023), e14722. https://doi.org/10.1016/j.heliyon.2023.e14722

[76] Mohammed Ahmed Subhi, Sawal Hamid Ali, and Mohammed Abulameer Mohammed. 2019. Vision-based approaches for automatic food recognition and dietary assessment: A survey. *IEEE Access* 7 (2019), 35370–35381. https://doi.org/10.1109/ACCESS.2019.2904519

[77] Mohammed A. Subhi, Sawal Hamid Md. Ali, Ahmad G. Ismail, and Masuri Othman. 2018. Food volume estimation based on stereo image analysis. *(IEEE) Instrumentation Measurement Magazine* 21, 6 (Dec. 2018), 36–43. https://doi.org/10.1109/mim.2018.8573592

[78] Ghalib Ahmed Tahir and Chu Kiong Loo. 2020. An Open-Ended Continual Learning for Food Recognition Using Class Incremental Extreme Learning Machines. *IEEE Access* (2020), 82328–82346. https://doi.org/10.1109/ACCESS.2020.2991810

[79] Ghalib Ahmed Tahir and Chu Kiong Loo. 2021. Explainable deep learning ensemble for food image analysis on edge devices. *Computers in Biology and Medicine* 139 (2021), 104972. https://doi.org/10.1016/j.compbiomed.2021.104972

[80] Natta Tammachat and Natapon Pantuwong. 2014. Calories analysis of food intake using image recognition. In *2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE)*. IEEE. https://doi.org/10.1109/iciteed.2014.7007901

[81] Punnarumol Temdee and Surapong Uttama. 2017. Food recognition on smartphone using transfer learning of convolution neural network. In *2017 Global Wireless Summit (GWS)*. IEEE. https://doi.org/10.1109/gws.2017.8300490

[82] Chakkrit Termritthikun, Paisarn Muneesawang, and Surachet Kanprachar. 2017. NU-InNet: Thai Food Image Recognition Using Convolutional Neural Networks on Smartphone. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* 9 (08 2017), 2289–8131.

[83] Ukrit Tiankaew, Peerapon Chunpongthong, and Vacharapat Mettanant. 2018. A Food Photography App with Image Recognition for Thai Food. In *2018 Seventh ICT International Student Project Conference (ICT-ISPC)*. IEEE. https://doi.org/10.1109/ict-ispc.2018.8523925

[84] Vlad-Ioan Tomescu. 2020. FoRConvD: An approach for food recognition on mobile devices using convolutional neural networks and depth maps. In *2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI)*. IEEE. https://doi.org/10.1109/saci49304.2020.9118817

[85] Xin Wang, Devinder Kumar, Nicolas Thome, Matthieu Cord, and Frederic Precioso. 2015. RECIPE RECOGNITION WITH LARGE MULTIMODAL FOOD DATASET. In *IEEE International Conference on Multimedia & Expo (ICME), workshop CEA*. Turin, Italy. https://doi.org/10.1109/ICMEW.2015.7169757

[86] Chang Xu, Ye He, Nitin Khannan, Albert Parra, Carol Boushey, and Edward Delp. 2013. Image-based food volume estimation. In *Proceedings of the 5th international workshop on Multimedia for cooking, eating activities - (CEA) 13*. ACM Press. https://doi.org/10.1145/2506023.2506037

[87] Keiji Yanai and Yoshiyuki Kawano. 2015. Food image recognition using deep convolutional network with pre-training and fine-tuning. In *2015 IEEE International Conference on Multimedia, Expo Workshops (ICMEW)*. IEEE. https://doi.org/10.1109/icmew.2015.7169816

[88] Shulin Yang, Mei Chen, D Pomerleau, and R Sukthankar. 2010. Food recognition using statistics of pairwise local features. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE. https://doi.org/10.1109/cvpr.2010.5539907

[89] Zhengeng Yang, Hongshan Yu, Shunxin Cao, Qi Xu, Ding Yuan, Hong Zhang, Wenyan Jia, Zhi-Hong Mao, and Mingui Sun. 2021. Human-Mimetic Estimation of Food Volume from a Single-View RGB Image Using an AI System. *Electronics* 10, 13 (June 2021), 1556. https://doi.org/10.3390/electronics10131556

[90] Boyang You and Kerry Papps. 2022. A Constructive GAN-based Approach to Exact Estimate Treatment Effect without Matching. (2022). arXiv:2206.06116 [econ.EM]

[91] Ding Yuan, Xiaohui Hu, Hong Zhang, Wenyan Jia, Zhi-Hong Mao, and Mingui Sun. 2020. An automatic electronic instrument for accurate measurements of food volume and density. *Public Health Nutrition* 24, 6 (Aug. 2020), 1248–1255. https://doi.org/10.1017/s136898002000275x

[92] Weiyu Zhang, Qian Yu, Behjat Siddiquie, Ajay Divakaran, and Harpreet Sawhney. 2015. "Snap-n-Eat". *Journal of Diabetes Science and Technology* (April 2015), 525–533. https://doi.org/10.1177/1932296815582222

[93] Jiannan Zheng, Z. Jane Wang, and Xiangyang Ji. 2016. Superpixel-based image recognition for food images. In *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. 1–4. https://doi.org/10.1109/CCECE.2016.7726860

[94] Xin Zheng, Chenhan Liu, Yifei Gong, Qian Yin, Wenyan Jia, and Mingui Sun. 2023. Food volume estimation by multi-layer superpixel. *Mathematical Biosciences and Engineering* 20, 4 (2023), 6294–6311. https://doi.org/10.3934/mbe.2023271

[95] Fengqing Zhu, Marc Bosch, Nitin Khanna, Carol J. Boushey, and Edward J. Delp. 2015. Multiple Hypotheses Image Segmentation and Classification With Application to Dietary Assessment. *(IEEE) Journal of Biomedical and Health Informatics* (2015), 377–388. https://doi.org/10.1109/jbhi.2014.2304925