healthcare

MDPI

Review

A Comprehensive Survey of Image-Based Food Recognition and Volume Estimation Methods for Dietary Assessment

Ghalib Ahmed Tahir and Chu Kiong Loo *

Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia; 12mscsgtahir@seecs.edu.pk or ghalib@siswa.um.edu.my
* Correspondence: ckloo.um@um.edu.my

Abstract: Dietary studies showed that dietary problems such as obesity are associated with other chronic diseases, including hypertension, irregular blood sugar levels, and increased risk of heart attacks. The primary cause of these problems is poor lifestyle choices and unhealthy dietary habits, which are manageable using interactive mHealth apps. However, traditional dietary monitoring systems using manual food logging suffer from imprecision, underreporting, time consumption, and low adherence. Recent dietary monitoring systems tackle these challenges by automatic assessment of dietary intake through machine learning methods. This survey discusses the best-performing methodologies that have been developed so far for automatic food recognition and volume estimation. Firstly, the paper presented the rationale of visual-based methods for food recognition. Then, the core of the study is the presentation, discussion, and evaluation of these methods based on popular food image databases. In this context, this study discusses the mobile applications that are implementing these methods for automatic food logging. Our findings indicate that around 66.7% of surveyed studies use visual features from deep neural networks for food recognition. Similarly, all surveyed studies employed a variant of convolutional neural networks (CNN) for ingredient recognition due to recent research interest. Finally, this survey ends with a discussion of potential applications of food image analysis, existing research gaps, and open issues of this research area. Learning from unlabeled image datasets in an unsupervised manner, catastrophic forgetting during continual learning, and improving model transparency using explainable AI are potential areas of interest for future studies.

Keywords: food recognition; feature extraction; automatic diet monitoring; image analysis; volume estimation; interactive segmentation; food datasets

check for updates

Citation: Tahir, G.A.; Loo, C.K. A Comprehensive Survey of Image-Based Food Recognition and Volume Estimation Methods for Dietary Assessment. *Healthcare* **2021**, *9*, 1676. https://doi.org/10.3390/healthcare9121676

Academic Editor: Mahmudur Rahman

Received: 4 September 2021
Accepted: 23 November 2021
Published: 3 December 2021

1. Introduction

Despite recent advancements in medicine, the number of people affected by chronic diseases is still large [1]. This rate is primarily due to their unhealthy lifestyles and irregular eating patterns. As a result, obesity and weight issues are becoming increasingly common around the globe. Some of the more notable diseases caused by obesity include hypertension [2], blood sugar [3], cardiovascular diseases [4], and different kinds of cancers [5]. The main reported obesity issues are in developed and middle-income countries. In 2016, 1.9 billion adults 18 years and older were overweight, while 650 million were obese. With time, children are also becoming affected by obesity at an alarming rate. According to World Health Organization (WHO), over 340 million children and adolescents between 5 and 19 years were overweight or obese [6].

The prevalence of these alarming statistics poses a serious concern. However, determining the effective remedial measures depends on different factors, ranging from a person's genetics to their lifestyle choices. To cope with chronic weight problems, people often keep notes to track their dietary intake. In turn, dieticians require these records to estimate a patient's nutrient consumption. However, these methods pose a challenge for users and dieticians, especially when they have to record time and estimate nutrients of diet intake [7]. For these reasons, recent research efforts have explored sophisticated vision-based methods to automate the process of food recognition and volume estimation [8,9]. The advancement in smartphone applications and hardware resources has made this more convenient, and present studies also show a higher retention rate of these mHealth apps than traditional methods [10]. Recent advancements in machine learning methods have further paved the way for more robust mHealth apps. Some dietary mobile applications such as DietLens [11], DietCam [12], Im2Calories [13], etc. integrate their apps with AI models for food recognition and ingredients detection to automate food logging. The Dietcam app also estimates nutrients from smartphone camera pictures.

However, automatic food recognition using a smartphone camera in the real world is considered a multi-dimensional problem, and the solution effectiveness depends upon several factors. Firstly, the model can achieve optimal classification performance by training with many food images for each class. Other than that, food recognition is a complex task that involves several domain-specific challenges. There is no spatial layout information that it can exploit like, in the case of the human body, the spatial relationship between body parts. The head is always present over the trunk of the human body [14–16] and feet towards the lower end. Similarly, the non-rigid structure of the food and intra-source variations make it even more complicated to classify food items correctly as preparation methods and cooking styles vary from region to region. Moreover, inter-class ambiguity is also a source of potential recognition problems as different food items may look very similar (e.g., soups). Moreover, in many dishes, some ingredients are concealed from view that can limit the performance of food ingredient classification models.

In addition to this, image quality from the smartphone camera is dependent on different types of cameras, lighting conditions, and orientations. As a result, the poor performance of food recognition models is highly susceptible to image distortions.

Despite these challenges, many food images possess distinctive properties to distinguish one food type from another. Firstly, the visual representations of food images are of fundamental importance as it significantly impacts classification performance. Therefore, many food-recognition methods employ handcrafted features such as shape, color, texture, and location. Recent techniques are using deep visual features for image representations. Some of these methods implement a combination of handcrafted and deep visual features for image feature representations. Secondly, for enhanced classification performance and reduced computational complexity, an appropriate selection of attributes is essential for removing redundant features from feature vectors. Finally, wisely selecting classification techniques is crucial to address food recognition challenges effectively.

Similarly, manual logging of food volume is a tedious task and involves a high rate of human error by as much as 30% [17–22]. Several solutions are proposed whose aim is to estimate food volume from smartphone camera pictures. Previous studies [23] show that using a mobile phone camera for food volume estimation increases the accuracy of the estimation of calories. Some methods involve capturing a single image, while multiple views are needed to determine accurate volume in other techniques. The food volume estimation process involves the following two steps (1) multiple images or a single image from a mobile camera is needed (2) computation of food volume from 3D construction or calibration object. Regardless of other volume estimation tasks, food volume estimation is a complex task with factors such as variations in shape and appearance due to various shapes of food and eating conditions affecting its performance.

The following research paper aims to scrutinize state-of-the-art vision-based approaches for dietary assessment to give researchers a summary of this area. Figure 1 represents the detailed scope and taxonomy of our survey study. The contribution of this survey is summarized as follows:

(1) The article briefly explores food databases for evaluating vision-based approaches and performance measures to thoroughly investigate food recognition, ingredient detection, and volume estimation methods.

(2) It presents an extensive review of food recognition techniques, including traditional methods with handcrafted features and modern deep-learning-based approaches.
(3) It provides deep insight into multi-label methods for food ingredient classification.
(4) This study surveyed most performing single-view and multi-view methods for food volume estimation.
(5) This study presents existing mobile applications that implement these approaches and other potential applications of vision-based methods in health care.
(6) The article analyzes open issues and suggests possible solutions to overcome the limitations of the existing methodologies.

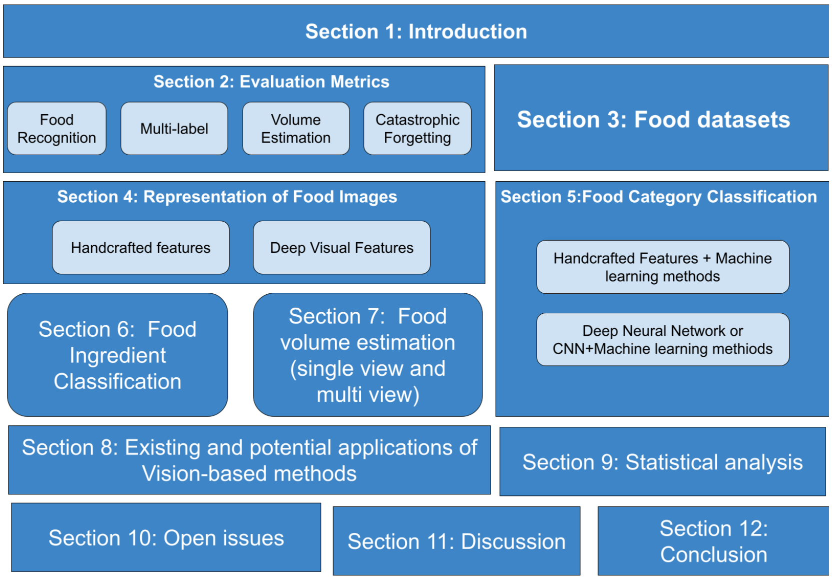

Section 1: Introduction
Section 2: Evaluation Metrics
Food Recognition
Multi-label
Volume Estimation
Catastrophic Forgetting
Section 3: Food datasets
Section 4: Representation of Food Images
Handcrafted features
Deep Visual Features
Section 5:Food Category Classification
Handcrafted Features + Machine learning methods
Deep Neural Network or CNN+Machine learning methiods
Section 6: Food Ingredient Classification
Section 7: Food volume estimation (single view and multi view)
Section 8: Existing and potential applications of Vision-based methods
Section 9: Statistical analysis
Section 10: Open issues
Section 11: Discussion
Section 12: Conclusion

Figure 1. Scope and taxonomy of this survey paper.

It should be noted that the article is related to vision-based methods for food image analysis and their applications in the field of healthcare currently being discussed in the literature. However, the methodology of this article seeks to examine the systems more broadly by describing their important aspects similar to narrative overview [24] instead of a systematic review, some related works to the topic, or adopted search followed by a brief discussion.

Section 1 has presented the introduction of the study. The rest of the article is organized as follows. Sections 2 and 3 examine evaluation metrics and existing datasets. Section 4 examines feature extraction methods for food image representation including handcrafted and deep visual features. In Sections 5 and 6, we presented the most performing classifiers for food categorization and ingredient detection. Section 7 represents the food-volume-estimation methods. In Section 8, we provide brief information about mobile applications implementing these methods and other potential applications. Sections 9 and 10 summarize statistical analysis and open issues. To conclude, we highlight our findings and future works related to this topic.

2. Evaluation Metrics

2.1. Evaluation Metrics for Food Categorization

The performance of automatic food recognition models is highly dependent on the correct mapping of food images into their respective categories. Therefore, confusion-matrix and evaluation metrics play an essential role in determining the correctness of food recognition models. Several metrics have been discussed in the literature, and their appropriate selection depends on the requirements of specific applications. It has also been observed that a classifier may perform well under one metric but poorly under another metric. For example, in the context of an imbalanced food dataset, the data samples from one or more classes outnumber data samples from the remaining food classes. Then a model trained on an imbalanced data set can have higher accuracy because of its good performance on the majority classes despite having bad classification performance on minority classes. Confusion matrix and other intrinsic metrics (Accuracy, Precision, Recall, and F1-score) generally used for detailed comparisons are discussed in detail below.

2.1.1. Confusion Matrix

Confusion matrices are a widely used approach to summarize the performance of a classification model in machine learning. In some cases, classification accuracy alone can be misleading, especially when there are more than two classes in a dataset or if there were an unequal number of observations present in food classes. Therefore, the confusion matrix provides a clear picture of actual and predicted classes obtained by the classification model. The confusion matrix is basically a two-dimensional matrix where each row represents an example of an actual food class and each column represents a state of the predicted food class. TP stands for true positive, TN represents the number of true negatives, FP is the number of false positives, and FN represents false negatives in the confusion matrix shown in Figure 2.

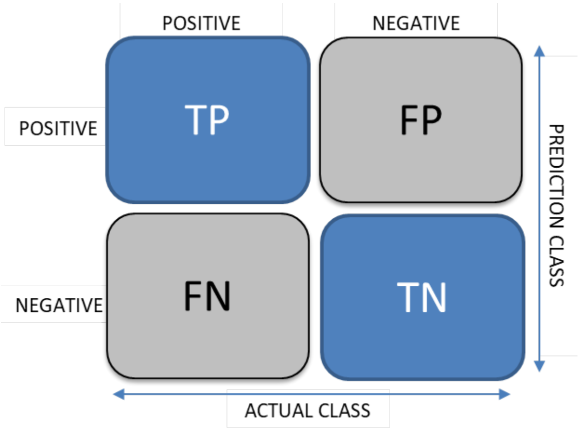


Figure 2. Confusion matrix.

2.1.2. Accuracy

The accuracy of a model determines whether the model is able to predict food classes correctly or how well a certain model can generally perform. Equation (1) represents the mathematical form of accuracy. However, accuracy cannot be used as a major performance metric, as it does not serve the purpose when there is an imbalanced dataset. Therefore, we have incorporated Precision, Recall, and F1 score to provide better insights into the results.

$$\text{Accuracy} = \frac{(TP + FN)}{(TP + FP + FN + TN)} \times 100 \tag{1}$$

Here TP refers to the true positive. True positive is an outcome where the model has correctly predicted a positive class. For example, in the case of food recognition, it refers to the food class that the model is trying to predict. TN refers to the true negatives: the prediction is correct, and the actual value is negative. In the case of food recognition, it refers to images from those food classes that the model is not trying to predict. FP refers to the false positive, and FP prediction results are wrong. For example, in the case of Food/NonFood recognition, FP refers to images that are non-food but are predicted as food. FN refers to the false negatives. It refers to those data samples which are positive but wrongly classified as negative class. For example, those food images that are classified as non-food images by model.

2.1.3. Precision

The Precision score can be defined as how often a model can correctly predict values classified as positives. In simpler words, out of all predicted positive food classes, it indicates what percentage is truly positive. This score is beneficial when the cost of false positives is high. It is calculated by Equation (2).

$$\text{Precision Score } = \frac{TP}{(TP\ + FP)} \tag{2}$$

2.1.4. Recall

Recall score identifies the model's ability to correctly classify food classes. It determines out of total positive food classes what percentage is predicted positives. It provides better insight when the cost of false negatives is high. It is computed by using Equation (3).

$$\text{Recall } = \frac{TP}{(TP + FN)} \tag{3}$$

2.1.5. F1 Score

F1 score represents the harmonic mean of recall and precision score. It considers both false positives and false negatives; therefore, it performs great on imbalanced datasets. It is calculated by following Equation (4).

$$\text{F1 Score } = \frac{(2 * (\text{Precision} * \text{Recall}))}{\text{Precision} + \text{Recall}} \tag{4}$$

2.2. Catastrophic Forgetting During Progressive Learning

Food datasets are open-ended due to the large variety of food dishes and different preparation styles. There are no limitations and constraints on the number of classes, and the model can progressively adapt domain variations in existing classes while learning new food classes. However, catastrophic forgetting during progressive learning causes the neural network to forget previous knowledge while learning new concepts. Catastrophic forgetting measures compute the algorithm's ability to retain previous concepts and knowledge while learning new information. Kemker et al. [25] and Chaudry et al. [26] proposed five measures of catastrophic forgetting to achieve this objective.

2.2.1. Intransigence

This refers to the difference in classification performance between the reference model trained by batch learning technique and the model trained on feature vectors using incremental learning protocol. The negative intransigence shows that incrementally learning a new set of food classes improves performance. Equation (5) denotes its mathematical form.

$$l_k = a_k^* - a_{k\prime k} \tag{5}$$

2.2.2. Forgetting

This refers to the difference between the highest classification performance of a particular session in previous sessions and its classification performance in the current sessions. Equation (6) computes the average forgetting of the network up to the kth session.

$$\begin{aligned} &f_j^k = max_{1\in\{1,\ldots\ldots,K-1\}}\ a_{i,j} - a_{(k,j)}, j > k \\ &F_k = \tfrac{1}{k-1} \textstyle\sum_{j=1}^{k-1} f_j^k \end{aligned} \tag{6}$$

2.2.3. Base Session

This refers to the model's ability to retain the knowledge of base food classes in current sessions, as shown in Equation (7).

$$\Omega_{base} = \frac{1}{k-1} \sum\nolimits_{j=2}^{k} \frac{a_{j,1}}{a_{ideal}} \tag{7}$$

2.2.4. New Session

This is the ability of a model to recall newly learned food classes, as shown in Equation (8).

$$\Omega_{new} = \frac{1}{k-1} \sum\nolimits_{j=2}^{k} a_{j,j} \tag{8}$$

2.2.5. All Session

This refers to the retention of the previous food classes learned by the network when learning new food classes, as computed by Equation (9).

$$\Omega_{all} = \frac{1}{k-1} \sum\nolimits_{j=2}^{k} \frac{a_{j,all}}{a_{ideal}} \tag{9}$$

2.3. Evaluation Metrics for Food Ingredient Classification

Similarly, food ingredient recognition is equally important for dietary assessment applications. As food categorization is limited to the classification of generic food items present in the food images, food ingredient recognition and classification provide deep insights into the caloric content present in the food image. Therefore, food ingredient recognition applications widely incorporate multi-label classification [27]. Since food ingredient recognition is considered a multi-label problem as food images usually contain more than one ingredient. Therefore, evaluation metrics generally used for multi-label classification are different from traditional single-label classification. The following are the performance metrics are used by food ingredient recognition models.

Consider x_i, Y_i with L number of labels as training datasets. Let us assume that MLC is the training method and $Z_i = MLC(x_i)$ is the output labels (ingredients) predicted by the classification method.

2.3.1. Precision

Precision is the ratio of correctly predicted labels to the total number of actual labels, averaged across all instances. Equation (10) represents precision for food ingredient classification.

$$\text{Precision} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{MLC(x_i) \cap Y_i}{MLC(x_i)} \right) \tag{10}$$

2.3.2. Recall

Recall is computed by Equation (11). It is the ratio of correctly predicted labels to the total number of predicted labels.

$$\text{Recall} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{MLC(x_i) \cap Y_i}{MLC(Y_i)} \right) \tag{11}$$

2.3.3. F1 Score

Finally, F1 score is the harmonic mean of the precision and recall. Equation (12) represents the F1 score.

$$\text{F1 Score} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{2 * |MLC(x_i) \cap Y_i|}{|MLC(x_i)| + |Y_i|} \right) \tag{12}$$

2.4. Evaluation Metrics for Food Volume Estimation

Similarly, various studies related to food volume estimation use ground truth values to compare the accuracy of their proposed methods to determine the accurate food volume [28–39]. Unfortunately, there is no dataset available to date for accurate measurement of food volume. Nevertheless, the method proposed by [40] uses controlled experiments that require participants to click images before and after their meal to compute consumed calories, which are later compared with ground truth values. Similarly, Ref. [41] incorporated different food models to determine the true volume; however, various models failed to provide accurate information. Therefore, they implemented the water displacement method, which requires a mean of three readings to find out the true volume. Furthermore, most studies used the following equations to compute the relative error and estimate the accuracy of the method

$$e = \left| v - v_{approx} \right| \tag{13}$$

where v is the actual volume and v_{approx} is the approximate volume

$$e = \frac{1}{N} \sum_{i=1}^{n} \frac{\left| w_i - w_g \right|}{w_g} \tag{14}$$

where N is the number of food items, w_i is the estimated weight of the food item, and w_g is the ground truth value of the food.

3. Datasets Used for Food Recognition

Performance of feature extraction and classification techniques is highly dependent on the detail-oriented collection of images, which, in our case, happen to be food images. As consolidated large food image datasets, for example, UECFOOD-100, Food-101, UECFOOD-256, UNCIT-FD1200, and UNCIT-FD889 are eventually used as benchmarks to collate recognition performance of existing approaches with new classifiers. Such datasets can be distinctive in terms of characteristics, such as the total number of images in a particular dataset, cuisine type, and included food categories.

For instance, UECFOOD-100 contains 100 different sorts of food categories, and each food category has a bounding box that indicates the location of the food item in the photograph. Food categories in this dataset mainly belong to popular foods in Japan [42]. Similarly, UECFOOD-256 is another variant of UECFOOD-100. However, it differs in terms of the number of images as it contains 256 food images of different kinds [42]. Food-101 contains 101,000 real-world images that are classified into 101 food categories. It includes diverse yet visually similar food classes [43]. Similarly, the PFID food dataset is composed of 1098 food images from 61 different categories. The PFID collection currently has three instances of 101 fast foods [44]. UNCIT-FD1200 is composed of 4754 food images of 1200 types of dishes captured from actual meals. Each food plate is acquired multiple times, and the overall dataset presents both geometric and photometric variability. Similarly, UNICT-FD 889 dataset has 3583 images [45] of 889 different real food plates captured using mobile devices in uncontrolled scenarios (e.g., different backgrounds and light environmental conditions). Moreover, they capture each dish image in UNICT-FD899 multiple times to ensure geometric and photometric variability (changes in rotation, scale, and point of view) [46].

Several datasets mainly consist of various food images collected through various sources such as web crawlers and social media platforms such as Instagram, Flickr, and Facebook. Furthermore, most of these datasets contain images of foods that are specific to certain regions, such as Vireo-Food 172 [47] and ChineseFoodNet [48]. Both datasets contain Chinese dishes. Similarly, Food-50 [49], Food-85 [49], Food log [50], UECFOOD-100 [42], and UECFOOD-256 [43] contain Japanese Foods items. Turkish foods-15 [51] is limited to Turkish food items only. Furthermore, the Pakistani Food Dataset [52] accommodates Pakistani dishes, and the Indian Food Database incorporates Indian cuisines. In addition to this, few datasets only include fruits and vegetables like VegFru [53], Fruits 360 Dataset [54], and FruitVeg-81 [55]. Furthermore, Table 1 provides a brief description about food image datasets. Figure 3 shows the system flow and Figure 4 shows the sample images from the food datasets.

Table 1. Food image datasets.

Authors	Year	Dataset	Food Category	Total # Images/Class	Image Source
S. Godwin et al. [56]	2006	Wedge Shape foods dataset	American Foods	3 categories	Controlled environment
Chen et al. [44]	2009	PFID	American Fast Foods	1038(61)	Fast food data captured in multiple restaurants
Mariappan et al. [57]	2009	TADA	Artificial And Generic Food	256(11)	Controlled environment
Yanai et al. [49]	2010	Food-50	Japanese Foods	5000(50)	Crawled from web
Hoashi et al. [49]	2010	Food-85	Japanese Foods	8500(85)	Existing food databases
Miyazaki et al. [29]	2011	Foodlog	Japanese Foods	6512(2000)	Captured by users
Marc Bosch et al. [58]	2011	FNDDS	American Foods	7000	Images of food accquired by users
Matsuda et al. [42]	2012	UECFOOD-100	Japanese Foods	14,361(100)	Captured by mobile camera
Chen et al. [48]	2012	ChineseFoodNet	Chinese dishes.	192,000(208)	Gathered from web
M.-Y. Chen et al. [48]	2012	Chen	Chinese Foods	5000/50	Crawled from the Internet
Bossard et al. [59]	2014	Food-101	American Foods	101,000(101)	Crawled from web
L. Bossard et al. [59]	2014	ETHZ Food-101	American Foods	100,000(101)	Crawled from web
Kawano et al. [43]	2014	UECFOOD-256	Japanese Foods	25,088(256)	Captured by mobile camera
T. Stutz et al. [60]	2014	Rice dataset	Generic (Rice)	1 food type	Acquired from user
Farinella et al. [46]	2014	UNCIT-FD889	Italian Foods	3583 (899)	Acquired with a smartphone
Meyers et al. [13]	2015	FOOD201-Segmented	American Foods	12625	Manually annotated dataset
Xin Wang et al. [61]	2015	UPMC Food-101	Generic	100,000(101)	Crawled from web
Cioccoa et al. [50]	2015	UNIMB 2015	Generic	2000(15)	Using a Samsung Galaxy S3 smartphone
Shaobo Fang et al. [62]	2015	TADA(19 foods)	American Foods	19 categories	Controlled environment
Xu et al. [63]	2015	Dishes	Chinese Restaurant Foods	117,504(3832)	Download from dianping
Beijbom et al. [64]	2015	Menu-Match	Generic Restaurant Food	646(41)	Captured from social media
Zhou et al. [65]	2016	Food-975	Chinese Foods	37,785(975)	Collected from restaurants
J. chen et al. [47]	2016	Vireo-Food 172	Chinese Foods	110,241(172)	Downloaded from web
Cioccoa et al. [66]	2016	UNIMB 2016	Italian Foods	1027(73)	Captured from dining tables
Hui Wu et al. [67]	2016	Food500	Generic	148,408(508)	Crawled from web
Singla et al. [68]	2016	Food-11	Generic	16,643(11)	Other food datasets
Farinella et al. [45]	2016	UNCIT-FD1200	Generic	4754(1200)	Acquired using smartphone
Jaclyn Rich et al. [69]	2016	Instagram 800k	Generic	808,964(43)	Social Media
Liang et al. [70]	2017	ECUSTFD	Generic	2978(19)	Acquired using smartphone
Güngör et al. [51]	2017	Turkish-Foods-15	Turkish Dishes	7500/15	Collected from other datasets
Pandey et al. [71]	2017	Indian Food Database	Indian Foods	5000(50)	Downloaded from web
Termritthikun et al. [72]	2017	THFood-50	Thai Foods	700/50	Downloaded from web
Ciocca et al. [73]	2017	FOOD524DB	Generic	247,636(524)	Existing food database
Hou et al. [53]	2017	VegFru	Generic (Fruit and VEG)	160,731(292)	Collected from search engine
Waltner et al. [55]	2017	FruitVeg-81	Generic (Fruit and VEG)	15,630(81)	Collected using mobile phone
Muresan et al. [54]	2018	Generic (Fruits 360 Dataset)	Fruit Dataset	71,125(103)	Camera
Qing Yu et al. [74]	2018	FLD-469	Japanese Foods	209,700(469)	Smart Phone camera
Kaur et al. [75]	2019	FoodX-251	Generic	158,000(251)	Collected from web
Ghalib et al. [52]	2020	Pakistani Food Dataset	Pakistani Dishes	4928(100)	Crawled from web
Narayanan et al. [76]		AI-Crowd	Swiss Foods	25,389	Volunteer Users
Bolaños M. et al. [77]	2016	EgocentricFood	Generic	5038(9)	Taken by a wearable egocentric vision camera
E. Aguilar et al. [78]	2019	MAFood-121	Spanish Foods	21,175	Google search engine

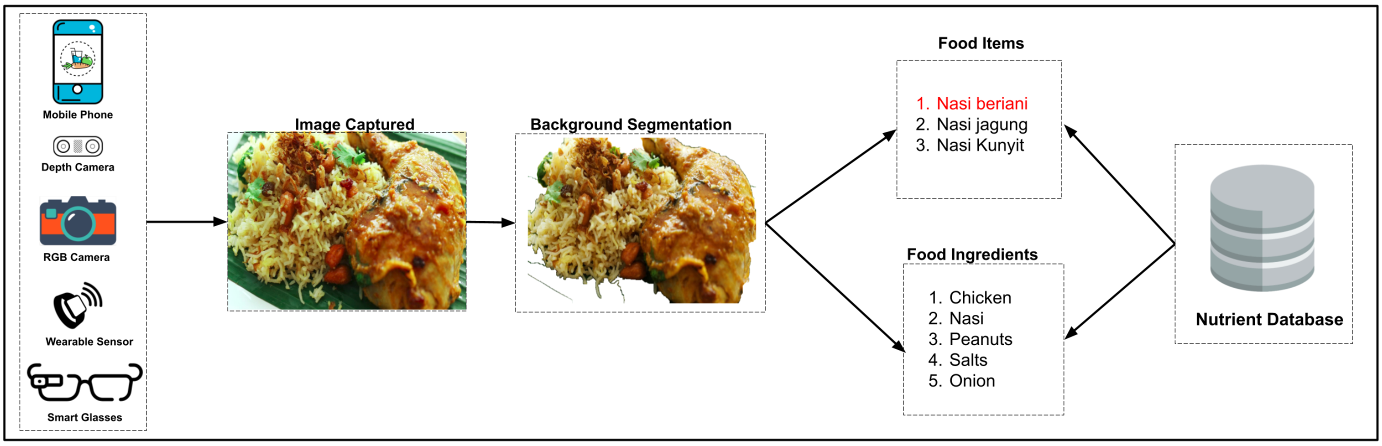


Figure 3. System Flow.

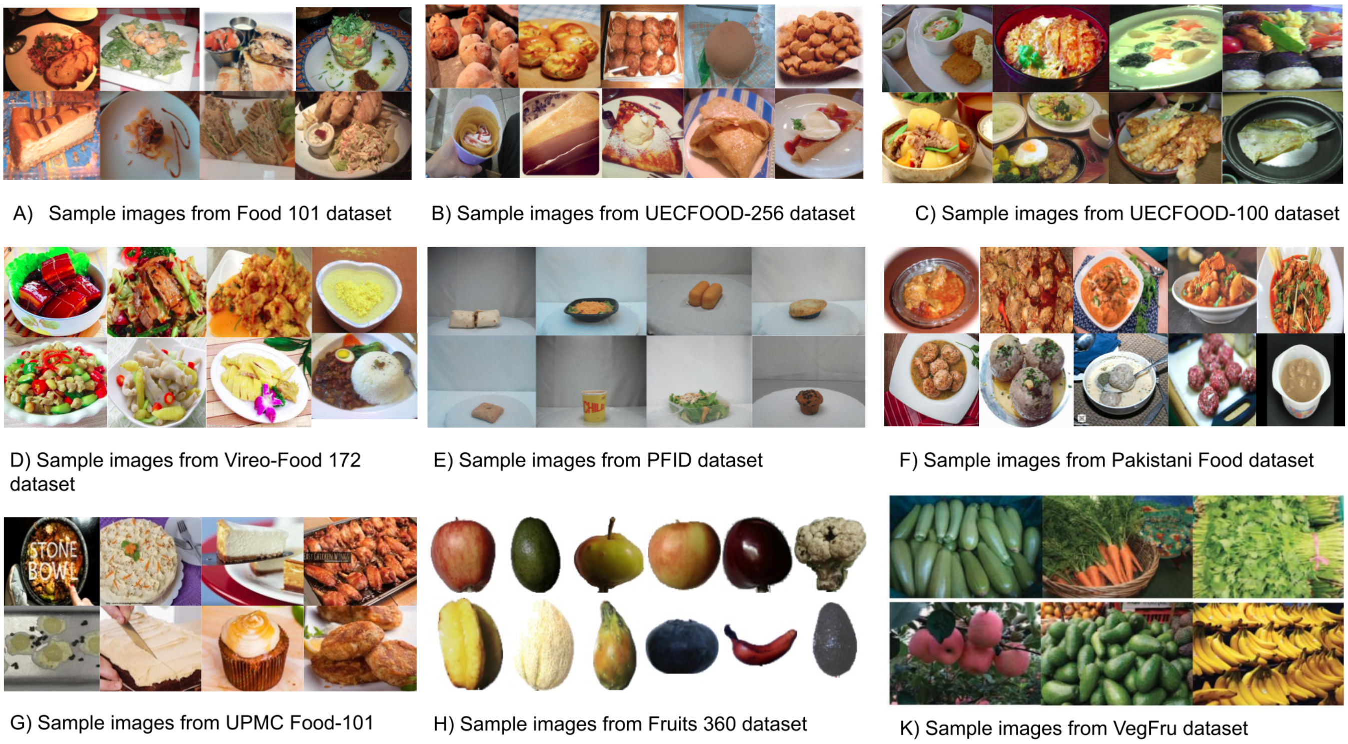

A) Sample images from Food 101 dataset
B) Sample images from UECFOOD-256 dataset
C) Sample images from UECFOOD-100 dataset
D) Sample images from Vireo-Food 172 dataset
E) Sample images from PFID dataset
F) Sample images from Pakistani Food dataset
G) Sample images from UPMC Food-101
H) Sample images from Fruits 360 dataset
K) Sample images from VegFru dataset

Figure 4. Sample images from few food datasets.

Therefore, it is evident from the survey that there is an immense need for broad and generic food datasets for better food recognition and enhanced performance. This necessity is because region-specific food items or datasets with fewer food categories can undermine the accuracy and performance of classification and extraction methods.

4. Representation of Food Images

Feature extraction plays a vital role in automated food recognition applications due to its noticeable impact on the recognition efficiency of an employed system. Feature extractors methods extract different food image representations. The process of feature extraction involves the identification of visual characteristics like color, shape, and texture. The main objective of feature extraction is to reduce dimensionality space [79] and extract more manageable groups from raw vectors of food images.

Moreover, selecting the right set of features ensures that relevant information is extracted from input images to perform the desired task. We categorized the feature extraction techniques into two main types: hand-crafted and deep visual features. The term ‘handcrafted’ refers to identifying relevant feature vectors of appropriate objects such as shape, color, and texture. In contrast to that, the deep model provides state-of-the-art performance due to automatic feature extraction through a series of connected layers. For this reason, recent studies have adopted combinations of both hand-crafted and deep visual features for food image representation.

4.1. Handcrafted Features

The existing literature exhibits a large number of methods to employ manually designed or handcrafted features. Handcrafted features are properties obtained through algorithms using help from information available in the image. Figure 5 categorizes the handcrafted feature extraction methods. In the scenario of food image recognition, there is variation among different food types in terms of texture, shape, and color.

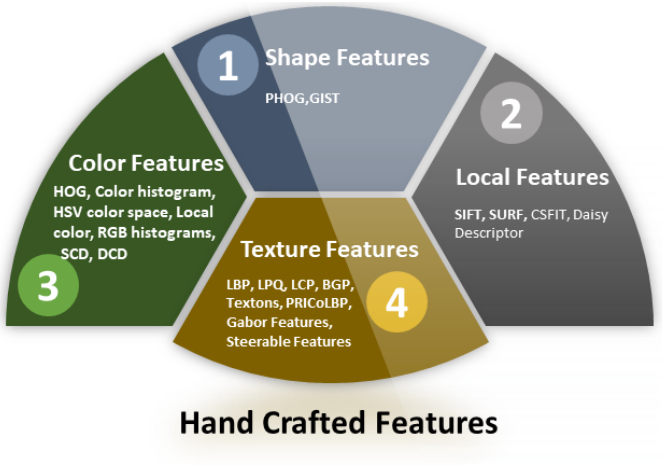

1
Shape Features
PHOG,GIST
2
Local Features
SIFT, SURF, CSFIT, Daisy Descriptor
Color Features
HOG, Color histogram, HSV color space, Local color, RGB histograms, SCD, DCD
3
Texture Features
LBP, LPQ, LCP, BGP, Textons, PRICoLBP, Gabor Features, Steerable Features
4
Hand Crafted Features

Figure 5. Handcrafted feature extraction methods.

The term ‘texture’ refers to homogeneous visual patterns that do not result from single colors such as sky and water [7]. Textural features usually consist of regularity, coarseness, and/or frequency. Texture-based characteristics are classified into two classes, namely statistical and transform-based models. Similarly, shape features attempt to quantify shape in ways that agree with human intuition or aid in perception based on relative proximity to well-known shapes. Based on the analysis, these shapes can be declared either perceptually similar to human perception or different. Furthermore, extracted features should remain consistent concerning rotation, location, and scaling (changing the object size) of an image. Unlike shape and texture features, color features are prevalent for image retrieval and classification because of their invariant properties concerning image translation, scaling, and rotation. The key items of the color feature-extraction process are color quantization and color space. Therefore, the resulting histogram is only discriminative when it projects the input image is to the appropriate color space. Different methods are widely employed for food classification, including hue, saturation, value (HSV); CIELab; red, green, and blue (RGB); normalized RGB; opponent color spaces; color k-means clustering; bag of color features; color patches; and color-based kernel. Although the color features from the food images distinguish between different food items, due to intra-class similarity, these features alone are not enough to accurately classify food images. For this reason, most researchers have used color features in combination with other feature extraction methods.

Hoashi et al. [49] employed bag of features, color histogram, Gabor features, and gradient histogram with multiple kernel learning for automatic food recognition of 85 different food categories. Similarly, Yang et al. [80] dealt with pairwise statistics between local features for food recognition purposes using the PFID dataset. For real-time food image recognition, Kawano and Yanai et al., 2014 [43] utilized handcrafted features such as color, histogram of oriented gradient (HoG), and Fisher Vector (FV). Moreover, the cloud-based food recognition method proposed by Pouladzadeh et al., 2015 [81], involves features like color, texture, size, shape, and Gabor filter. They evaluated their framework on single food portions consisting of fruit and a single item of food. Furthermore, mobile food recognition systems proposed by Kawano and Yanai, 2013 [82], and Oliveira et al., 2014 [83], also used handcrafted features like color and texture. Table 2 summarizes the details of proposed methods that employ handcrafted features for food recognition.

However, identification of food involves challenges due to varying recipes and presentation styles used to prepare food all around the globe, resulting in different feature sets [84]. For instance, the shape and texture of a salad containing vegetables differ from the shape and texture of a salad containing fruits. For this reason, we should optimize the feature extraction process by extracting relevant visual information from food images. Such data are present in general information descriptors, which are a collection of visual descriptors that provide information about primary features like shape, color, texture, and so forth. Some important descriptors used in existing studies include Gabor Filter, Local Binary Patterns (LBP), Scale-invariant Feature Transform (SIFT), and color information to extract features of food images [85]. These descriptors can be applied individually or in combination with other descriptors for enhanced accuracy.

Table 2. Handcrafted features.

Reference	Year	Visual Features	Dataset	Recognition Type
Hoashi et al. [49]	2010	Bag-of-features (BoF), Color histogram, Gabor features, and gradient histogram with Multiple Kernel learning.	Used for recognition of 85 food categories	Automatic food recognition
Yang et al. [80]	2010	Deals with pair wise statistics between local features	Pittsburgh Food Image Dataset (PFID)	Food recognition
kong and Tan [86]	2011	SIFT, Guassian Region detector	Pittsburgh Food Image Dataset (PFID) and dataset consisting of food images collected from local restaurants.	Regular shaped foods recognition
Bosh et al. [85]	2011	Global feature classes: texture and color Local features: local entropy color, local color, Garbor filter, SIFT, Haar, Daisy descriptor, Steerable filters and Tamura perceptual filter	Database consisting of food images collected under controlled conditions, from nutritional studies conducted at Prudue University [58]	Food recognition and quantification
Zhang et al. [87]	2011	Color, SIFT, Shape, RGB histograms	Dataset came from online sources, which includes three types of cuisines, two dishes per cuisines were represented by 76 images	Classification of cuisines
Matsuda et al. [88]	2012	Gabor texture features, Histogram of Oriented Gradient (HoG), Bag-of-features of SIFT and CSFIT with Spatial pyramid.	Food image dataset containing 100 different food categories.	Multiple food images recognition
Kawano and Yanai [82]	2013	Bag-of-features and color histogram, HOG patch descriptor and color patch descriptor.	-	Mobile food recognition
Anthimopoulos et al. [89]	2014	Bag-of-features, SIFT and HSV color space	Visual dataset consisting of 5000 food images organized into 11 different classes	Food recognition system for diabetic patients
Tammachat and Pantuwong [90]	2014	Bag-of-features (BoF) , Texture and Color	Database consisting of 40 types of Thai food consisting of 100 images of each food type.	Food image recognition
Pouladzadeh et al. [91]	2014	Graph cut, Color and Texture	Dataset consisting of 15 different categories of fruits and food.	Food image recognition for calorie estimation
He et al. [92]	2014	Color, Texture, Dominant Color Descriptor (DCD), Scalable Color Descriptor (SCD), SIFT, Multi-scale Dense SIFT (MDSIFT), Entropy-Based Categorization and Fractal Dimension Estimation (EFD) and Gabor-Based Image Decomposition and Fractal Dimension Estimation (GFD)	Food image dataset containing 1453 images	Food image analysis

Table 2. *Cont.*

Reference	Year	Visual Features	Dataset	Recognition Type
Kawano and Yanai [43]	2014	Color, HoG and Fisher Vector	UECFOOD-256 food image dataset	Real-time food image recognition
Oliveira et al. [83]	2014	Color, Texture	Images were gathered using mobile's camera	Mobile Food Recognition
Pouladzadeh et al. [81]	2015	Color, Texture, Size, Shape, Gabor filter	System was tested on single food portions consisting of fruits and single piece of food. 100 images were chosen for training and 100 for testing purposes.	Cloud-based food recognition.
Farinella et al. [45]	2016	SIFT, Bag of Textons, PRICoLBP	UNICT-FD1200 dataset.	Food image recognition

Nonetheless, feature selection remains a complex task for food types that involve mixed and prepared foods. Such food items are difficult to identify and are not easily separable due to the proximity of ingredients in terms of color and texture features. In contrast, the evolution of deep learning methods has remarkably reduced the use of handcrafted features. This is due to their superior performance for both food categorization and ingredient detection tasks. However, handcrafted methods for feature extraction may still serve as the foundation for automated food recognition systems in the future.

4.2. Deep Visual Features

Recently, deep learning techniques have gained immense attention due to their superior performance for image recognition and classification. The deep learning approach is a sub-type of machine learning, and it trains more constructive neural networks. The vital operation of deep learning approaches includes automatic feature extraction through the sequence of connected layers leading up to a fully connected layer, which is eventually responsible for classification. Moreover, in contrast to conventional methods, deep learning techniques show outstanding performance while processing large datasets and have excellent classification potential [93,94].

Deep learning methods such as Convolutional Neural Networks (CNNs) [95], Deep Convolutional Neural Networks (DCNNs) [96], Inception-v3 [97], and Ensemble net are implemented by existing food recognition methods for feature extraction. Convolutional Neural Networks are one of the widely used deep learning techniques in the area of computer vision due to their impressive learning ability regarding visual data, and they achieve higher accuracy than other conventional techniques [98]. The DCNN technique gained popularity owing to its large-scale object recognition ability. It incorporates all major object recognition procedures such as feature extraction, coding, and learning. Therefore, DCNN is an adaptive approach for estimating adequate feature representation for datasets [99]. Similarly, Inception-v3 is also a new deep convolutional neural network technique introduced by Google. It is composed of small inception modules that are capable of producing very deep networks. As a result, this model has proved to have higher accuracy, decreased number of parameters, and computational cost in contrast to other existing models. Likewise, Ensemble Net is a deep CNN-based architecture and is a suitable method for extracting features. It is due to the outstanding performance of CNN feature descriptors as compared to handcrafted features.

Asymmetric multi-task CNN and spatial pyramid CNN [100] provides highly discriminative image representations. Jing et al. [47] proposed ARCH-D architecture for multi-class multilabel food recognition, and their model provides feature vectors for both food category and ingredient recognition. Although the feature vectors from multi-scale multi-view deep network [101] has a very high dimension, they were successful in achieving state-of-art performance. Ghalib et al. [52] proposed ARCIKELM for open-ended learning. They have employed InceptionResnetV2 for feature extraction due to their superior performance over other deep feature extraction methods such as ResNet-50 and DenseNet201. Table 3 further provides a brief description of deep visual features.

Table 3. Deep visual features.

Reference	Year	Features	Dataset	Recognition Type
Kawano and Yanai, [102]	2014	Fisher Vector and DCNN	UECFOOD-100 and 100-class food Dataset	Food image recognition
Yanai and Kawano, [96]	2015	DCNN	UECFOOD-100 and UECFOOD- 256	Food image recognition
Christodoulidis et al. [103]	2015	CNN	Manually annotated dataset with 573 food items	Food recognition
Pouladzadeh et al. [104]	2016	Graphcut and DCNN	Database consisting of 10,000 high res images	Food recognition for calorie measurement
Hassannejad et al. [105]	2016	Inception	Food-101, UECFOOD-100 and UECFOOD-256	Food image recognition
Liu et al. [106]	2016	DCNN	Food-101, UECFOOD-256	Mobile food image recognition
Chen and Ngo, [47]	2016	Arch-D	Chinese Foods	Ingredient recognition and food categorization
Ciocca et al. [66]	2017	VGG	UNIMIB 2016	Food recognition
Termritthikun et al. [72]	2017	NU-InNet	THFOOD-50	Food recognition
Pandey et al. [71]	2017	AlexNet, GoogLeNet and ResNet	ETH Food-101 and Indian Food Image Database	Food Recognition
Liu et al. [107]	2018	GoogleNet	UECFOOD-100, UECFOOD-256 and Food-101	Food recognition for dietary assessment
McAllister et al. [108]	2018	ResNet-152, GoogLeNet	Food 5k, Food-11, RawFooT-DB and Food-101	Food recognition
Martinel et al. [109]	2018	WISeR	UECFOOD-100, UECFOOD-256 and Food-101	Food recognition
E. Aguilar et al. [110]	2018	AlexNet	UNIMIB2016	Automatic food tray analysis
S. Horiguchi et al. [111]	2018	GoogleNet	Built their own food dataset FoodLog	Food image recognition
Gianluigi Ciocca et al. [112]	2018	ResNet50	Food 475	Food image recognition and classification
B. Mandal et al. [113]	2019	SSGAN	ETH Food-101 and Indian Food Dataset	Food Recognition of Partially Labeled Data
G.Ciocca et al. [114]	2020	GoogleNet, Inception-v3, MobileNet-V2 and ResNet-50	Own dataset containing 20 different food categories of fruit and vegetables.	Food category recognition, Food state recognition
L. Jiang et al. [115]	2020	VGGNet	UECFOOD-100, UECFOOD-256 and introduced new dataset based on FOOD-101.	Food recognition and dietary assesment
C. Liu et al. [116]	2020	VGGNet, ResNet	Vireo-Food 172	Food ingredient recognition
H. Liang et al. [117]	2020		ChineseFoodNet and Vireo-Food 172	Chinese food recognition
H. Zhao et al. [118]	2020	VGGNet, ResNet and DenseNet	UECFOOD-256 and Food-101	Mobile food recognition
G. A. Tahir and C. K. Loo [52]	2020	ResNet-50, DenseNet201 and InceptionResNet-V2	Pakistani Food Dataset, UECFOOD-100, UECFOOD-256, FOOD-101 and PFID	Food recognition
C. S. Won [119]	2020	ResNet50	UECFOOD-256, Food-101 and Vireo-Food 172	Fine grained Food image recognition
Zhidong Shen et al. [120]	2020	Inception-v3, Inception-v4	Dataset was created including hundreds and thousands of images of several food categories	Food recognition and nutrition estimation

5. Food Category Classification

The primary requirement of any food recognition system is accurate identification and recognition of food components in the meal. Therefore, robust and precise food classification methods are crucial for several health-related applications such as automated dietary assessment, calorie estimation, and food journals. Image classification refers to a machine learning technique that associates a set of unspecified objects with a subset (class) learned by the classifier during the training phase. In the scenario of food image classification, food images are used as input data to train the classifier. Hence, an ideal classifier must recognize any food category explicitly included during the learning phase. The accuracy of a classifier mainly depends on the quantity and quality of images, as there are several variations in food images such as rotation, distortion, lightning distribution, and so forth. In this section, we discuss classification techniques used by traditional approaches that use handcrafted features. Following that, we analyzed state-of-the-art deep learning models for food recognition.

5.1. Traditional Machine Learning Methods

Major classifiers used by several traditional approaches in the domain of food image recognition include Support Vector Machines (SVM) [49], Multiple Kernel Learning (MKL) [49] and K-Nearest Neighbor (KNN) [47]. It is due to their outstanding performance as compared to other classification methods.

The food recognition method proposed by [121] employs color, SIFT, and texture features to train the KNN classifier. In contrast to SVM, KNN achieved higher classification accuracy, i.e., 70%, whereas the accuracy of the SVM classifier was only 57%. Similarly, treatment of diabetic patients involves a daily insulin prandial dose to compensate for the effect of a meal, and its estimation is a complex task with carbohydrate counting being a key element. To assist patients in automating the process of counting CHO from images captured from a camera, Anthimopoulos et al. [89] applied a bag-of-features model using SIFT features. A linear SVM classifier trained on food images of 11 different food classes acquired a classification accuracy of 78%.

Chen et al. [48], employed a multi-class SVM classifier for the identification of 50 different classes of Chinese food. It includes 100 food images in each category. However, classification accuracy was only 62.7%. They further implemented a multi-class Adaboost algorithm and increased their classification accuracy up to 68.3%. Furthermore, Bejibom et al. [64] used LBP, color, SIFT, MR8, and HoG features to train an SVM image classifier. They evaluated their work on two different datasets and achieved a classification accuracy of 77.4% on the dataset presented by [48]; their classification accuracy was 51.2% when applied to the menu-matched dataset. Table 4 summarizes classifiers implemented by traditional classification methods along with their achieved classification accuracies.

Table 4. Traditional machine learning methods for food category classification.

Reference	Year	Classification Technique	Classification Accuracy	
			Top 1	Top 5
Hoashi et al. [49]	2010	Multiple Kernel Learning (MKL)	Own Food Dataset = 62.5%	N/A
Yang et al. [80]	2010	Support Vector Machine (SVM)	PFID = 78.0%	N/A
Kong and Tan [86]	2011	Multi-class SVM	PFID = 84%	N/A
Bosh et al. [85]	2011	Support Vector Machine (SVM)	Dataset collected = 86.1% using nutritional studies Conducted at Prudue University	N/A
Zhang et al. [87]	2011	SVM regression with RBF kernel	Own Food Dataset = 82.9%	N/A
Matsuda et al. [88]	2012	Multiple Kernel Learning (MKL) and Support Vector Machine (SVM)	Own food Dataset = 55.8%	N/A
Kawano and Yanai [82]	2013	Linear SVM and fast tookernel	N/A	81.6%
Anthimopoulos et al. [89]	2014	Linear SVM	Own Food Dataset = 78.0%	N/A
Tammachat and Pantuwong [90]	2014	Support Vector Machine (SVM)	Own Food Dataset = 70.0%	N/A
Pouladzadeh et al. [91]	2014	Support Vector Machine (SVM)	Own Food Dataset = 95%	N/A
He et al. [92]	2014	K-nearest Neighbors and Vocabulary Trees	Own Food Dataset = 64.5%	N/A
Kawano and Yanai [43]	2014	One-vs-rest	UECFOOD-256 = 50.1%	UECFOOD-256 = 74.4%
Oliveira et al. [83]	2014	Support Vector Machine (SVM)	Own Food Dataset Top 3 classification achieved between 84 and 100%	N/A
Pouladzadeh et al. [81]	2015	Cloud-based Support Vector Machine	Own Food Dataset = 94.5%	N/A
Farinella et al. [45]	2016	Support Vector Machine (SVM)	UNICT-FD1200 = 75.74%	UNICT-FD1200 = 85.68%

5.2. Deep Learning Models

Deep learning approaches have gained significant attention in the field of food recognition. This is due to their exceptional classification performance in comparison to traditional approaches [48,64]. convolutional neural network (CNN), deep convolutional neural network (DCNN), Ensemble Net, and Inception-v3 are some of the most prominent techniques used as existing methods for food image recognition purposes.

Yanai and Kawano [102] employed a deep convolutional neural network (DCNN) on three food datasets: Food-101, UECFOOD-256, and UECFOOD-100. They explored the effectiveness of pre-training and fine-tuning a DCNN model using 100 images from each food category obtained from each dataset. During evaluation, classification accuracy achieved was 78.77% for UECFOOD-100, 67.57% for UECFOOD-256, and 70.4% for Food-101. Similarly, the study presented by [105] implemented Inception-v3 deep network established by Google [97] on the same datasets, i.e., Food-101, UEC FOOD-100, and UECFOOD-256. Classification accuracy achieved using fine-tuned model V3 was greater than classification accuracy of the fine-tuned version of DCNN i.e., 88.28%, 81.45%, and 76.17% for UECFOOD-100, UECFOOD-256, and Food-101, respectively. The food recognition method proposed by [106] implemented a CNN-based approach using the Inception model on the same three datasets.

Classification accuracy achieved was 77.4%, 76.3% and 54.7% for UECFOOD-100, UECFOOD-256 and Food-101, respectively. Table 5 provides the overview of existing food recognition methods based on deep learning approaches and their classification performance.

Table 5. Deep learning models for food category classification.

Reference	**Year**	**Classification Technique**	**Classification Performance**	
			Top 1	**Top 5**

Yanai and Kawano [96]	2015	DCNN	UECFOOD-100 = 78.8%
			UECFOOD-256 = 67.6%
				N/A

Christodoulidis et al. [103]	2015	DCNN	Own dataset = 84.9%	N/A
Chen and Ngo [47]	2016	DCNN		
Pouladzadeh et al. [104]	2016	DCNN + Graph cut	Own dataset = 99%	N/A

Hassannejad et al. [105]	2016	DCNN	ETH Food-101 = 88.3%
			UECFOOD-100 = 81.5%
			UECFOOD-256 = 76.2%
				ETH Food-101 = 96.9%
				UECFOOD-100 = 97.3%
				UECFOOD-256 = 92.6%

Liu et al. [106]	2016	CNN	UECFOOD-100 = 76.3%
			Food-101 = 77.4%
				UECFOOD-100 = 94.6%
				Food-101 = 93.7%

Pandey et al. [71]	2017	Ensemble Net	ETH-Food101 = 72.1% Indian Food = 73.5% Database	ETH-Food101 = 91.6% Indian Food = 94.4% Database
Ciocca et al. [66]	2017	CNN	UNIMIB 2016 = 78.3%	N/A
Termritthikun et al. [72]	2017	CNN	THFOOD-50 = 69.8%	THFOOD-50 = 92.3%

McAllister et al. [108]	2018	CNN+ANN+SVM+
		Random Forest
			Food-5K = 99.4%
			Food-11 = 91.3%
			RawFooT-DB = 99.3%
			Food-101 = 65.0%
				N/A

Liu et al. [107]	2018	DCNN	UECFOOD-256 = 54.5%
			UECFOOD-100 = 77.5%
			Food 101 = 77.0%
				UECFOOD-256 = 81.8%
				UECFOOD-100 = 95.2%
				Food 101 = 94.0%

Martinel et al. [109]	2018	DNN	UECFOOD-100 = 89.6%
			UECFOOD-256 = 83.2%
			Food-101 = 90.3%
				UECFOOD-100 = 99.2%
				UECFOOD-256 = 95.5%
				Food-101 = 98.7%

E. Aguilar et al. [110]	2018	CNN+SVM	UNIMIB 2016 = 90.0%	N/A
Gianluigi Ciocca et al. [112]	2018	CNN	Food-475 = 81.6%	Food-475 = 95.5%

Table 5. *Cont.*

Reference	Year	Classification Technique	Classification Performance	
			Top 1	Top 5
S. Horiguchi et al. [111]	2018	Sequential Personalized Classifier (SPC) with fixed-class and incremental classification	FoodLog = 40.2% (t251-t300)	FoodLog = 56.6% (t251-t300)
B. Mandal et al. [113]	2019	Generative Adversarial Network	ETH Food-101 = 75.3% IndianFood Database = 85.3%	ETH Food-101 = 93.3% Indian Food Database = 95.6%
Aguilar-Torres et al. [122]	2019	CNN based on ResNet-50	MAFood-121 = 81.62%	N/A
Kaiz Merchant and Yash Pande [123]	2019	Inception V3	ETHZ Food-101 = 70.0%	N/A
Mezgec, S. et al. [124]	2019	Deep Learning	Own Food dataset = 93%	N/A
L. Jiang et al. [115]	2020	DCNN (Faster R-CNN)	FOOD20-with-bbx = 71.7%	FOOD20-with-bbx = 93.1%
C. Liu et al., 2020 [116]				

H. Zhao et al. [118]	2020	JDNet	UECFOOD-256 = 84.0%
			FOOD-101 = 91.2%
				UECFOOD-256 = 96.2%
				FOOD-101 = 98.8%

G. A. Tahir and C. K. Loo [52]	2020	Adaptive Reduced Class Incremental Kernel Extreme Learning Machine (ARCIKELM)	Food-101 = 87.3%
			UECFOOD-100 = 88.7%
			UECFOOD-256= 76.51%
			PFID = 100%
			Pakistani Food = 74.8%
				N/A

C. S. Won [119]	2020	Three-scale CNN	UECFOOD-256 = 74.1%
			Food 101 = 88.8%
			Vireo-Food 172 = 91.3%
				UECFOOD-256 = 93.2%
				Food-101 = 98.1%
				Vireo-Food 172 = 98.9%

Zhidong Shen et al. [120]	2020	CNN	Own dataset = 85.0%	N/A
Jiangpeng He et al. [125]	2020	18 layer ResNet	Own dataset = 88.67%	N/A
Eduardo Aguilar et al. [126]	2020	CNN	Own dataset = 88.67%	N/A
Dario Ortega Anderez et al. [127]	2020	CNN	Own dataset = 97.10%	N/A
G. Song et al. [128]	2020	CNN	Web crawled dataset = 56.47%	Web crawled dataset = 60.33
Limei Xiao et al. [129]	2021	CNN	Own dataset = 97.42%	N/A
Lixi Deng et al. [130]	2021	ResNet-50	School lunch dataset = 95.3%	N/A

6. Food Ingredient Classification

Over the past few years, nutritional awareness among people has increased due to their intolerance towards certain types of food, mild or severe obesity problems, or simply interest in maintaining a healthy diet. This rise in nutritional awareness has also caused a shift in the technological domain, as several mobile applications facilitate people in keeping track of their diet. However, such applications hardly offer features for automated food ingredient recognition.

For this purpose, several proposed models use multi-label learning for food ingredient recognition. It can be defined [27] as the prediction of more than one output category for each input sample. Therefore, food ingredient recognition is known as a multi-label learning problem. Marc Bolanos et al. have deployed CNN as a multi-label predictor to discover recipes in terms of the list of ingredients from food images [131]. Similarly, Yunan Wang et al. [132] used multi-label learning for mixed dish recognition, as they have no distinctive boundaries among them. Therefore, labeling bounding boxes for each dish is a challenging task. Another system proposed by Amaia Salvador et al. [133] regenerates recipes from provided food images along with cooking instructions. On the other hand, Jingjing Chen and Chong-Wah Ngo [47] proposed deep architectures for food ingredient recognition and food categorization and evaluated their proposed system on a large Chinese food dataset with highly complex food images. Food ingredient recognition is often overlooked and is a challenging task, as it requires training samples under different cooking and cutting methods for robust recognition. Therefore, methods proposed by Chen et al. [134] and J. Chen et al. [135] focus on food ingredient recognition. The authors Chen et al. [134] deploy multi-relational graph convolutional network that was later evaluated on Chinese and Japanese food datasets, resulting in 36.7% for UECFOOD-100 and 48.8% for VireoFood-172. However, Chen et al. [135] proposed DCNN based method for food ingredient recognition and achieved Top 1 accuracy up to 86.91% and Top 5 accuracy up to 97.59% for Vireo Food-251.

Furthermore, Table 6 provides brief information about accuracy scores of proposed systems along with methods and dataset used.

Table 6. Proposed methods for food ingredient classification.

Reference	Year	Dataset	Method	Recall	Precision	F1
Chen et al. [47]	2016	Vireo-Food 172	Arch-D (Multi-task)	-	-	67.17% (Micro-F1) 47.18% (Macro-F1)
		UECFOOD-100	Arch-D (Multi-task)	-	-	82.06% (Micro-F1) 95.88% (Macro-F1)
Bolaños et al. [131]	2017	Food-101	ResNet50+ Ingredients 101	73.45%	88.11%	80.11%
		Recipe 5k	ResNet50+ Recipe 5k	19.57%	38.93%	26.05%
		Recipe 5k	Inception-v3+ Recipe 5k (Simplified)	42.77%	53.43%	47.51%
Wang, Yunan, et al. [132]	2019	Economic Rice	Inception-V4 + NS (multi-scale)	71.90%	72.10%	71.40%
		Economic Behoon	Inception-V4 + NS (multi-scale)	77.60%	68.50%	69.70%
Salvador, Amaia, et al. [133]	2019	Recipe 1M	CNN Auto-Encoder	75.47%	77.13%	48.61%
J. Chen et al. [135]	2021	VireoFood-172	DCNN	-	-	75.77% (Micro-F1)

7. Food Volume Estimation

Automated food volume assessment is a convoluted task involving various challenges. Highly diverse and varying compositions of food, increasing varieties of ingredients, and different methods of preparations are only some of the factors that need to be taken into consideration. Furthermore, the quality of pictures taken for food volume estimation also impacts the accuracy. Clear pictures taken in good lighting conditions would yield different results compared to low-resolution or low-light images. Thus, far, several methods have been proposed for accurate estimation of food volume ranging from simple techniques such as pixel counting to complex methods such as 3D image reconstruction. They have been broadly categorized as either 'single image view' or 'multi-image/video view' methods in the subsequent sections. Figure 6 shows the types of food volume estimation methods.

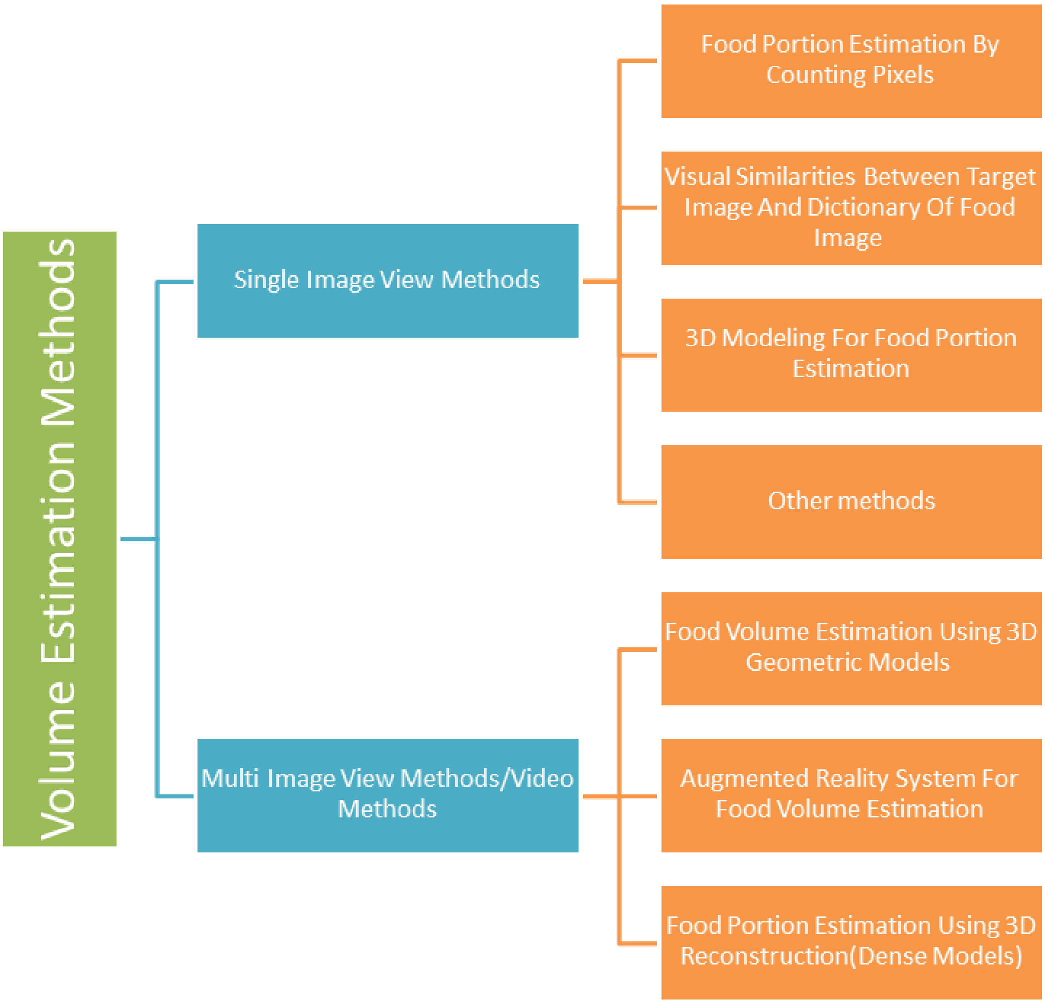

Volume Estimation Methods
Single Image View Methods
Multi Image View Methods/Video Methods
Food Portion Estimation By Counting Pixels
Visual Similarities Between Target Image And Dictionary Of Food Image
3D Modeling For Food Portion Estimation
Other methods
Food Volume Estimation Using 3D Geometric Models
Augmented Reality System For Food Volume Estimation
Food Portion Estimation Using 3D Reconstruction(Dense Models)

Figure 6. Food Volume Estimation Methods.

7.1. Single Image View Methods

Single-Image-View Methods for food volume estimation require only a single image for food volume estimation. These methods are relatively more user-friendly than 'multi-image view methods' because they do not require multiple images from different viewpoints. However, as a trade-off, most of the single-view methods are less accurate in contrast to multi-view methods. Table 7 summarizes single view methods for volume estimation. The following are a few common methods that use the single-view method for food portion estimation:

Table 7. Comparison of single-view methods for food volume estimation.

Reference	Year	Dataset	Results (E: Error%)	Technique
S. Fang [62]	2015	19 food items	E: <6%	3D parameters and reference objects to compute density for estimating the weight of food item
Y. He [36]	2013	1453 food images	E: 11% (beverages) 63%	"Integrated image segmentation and identification system"
T. Miyazaki [29]	2011	6512 images	E: 40%	Linear estimation
Beijbom, O [64]	2015	646 images, with 1386 tagged food items across 41 categories	E: 232 ± 7.2	Restaurant-specific food recognition considers meal as a whole entry with all of its nutrients details in DB to solve the volume estimation problem for the restaurant scenario.
Koichi Okamoto [31]	2016	20 kinds of Japanese Foods (60 test image)	E: 21.30%	Single-image-based food calorie estimation system which uses reference objects to determine food region and quadratic curve estimation from the 2D size of foods to their calories
Pettitt, C [136]	2016	Test data from N:6 participants who completed food diary during pilot sudy by wear micro camera	E: 34%	Wearable micro camera in conjunction with food dairies
Akpa Akpro Hippocrate [34]	2016	119 food images	E: 6.87%	Image processing with cutlery
Jia, W. Y [35]	2012	224 pictures	E: <10%	3D location of a circular feature from a 2D image
Yang, Y. Q [33]	2011	72 images	E: −3.55%	Single digital image, plate reference
Huang, J [39]	2015	fruits (n:6)		imaging processing
Yue, Y [41]	2012	6 food replicas	E: Length (−1.18)	A mathematical model based system involves a camera, circular object in a 3D space to compute food volume.
Zhang, W [38]	2015	15 different kinds of foods	85%	Portion estimation by counting pixels
Rob Comber [137]	2016	6 different meals	"Beef (E: −13.89 g, σ: 5.10 g), scrambled egg (E: −9.11 g, σ: 8.29 g), Jam sponge (E: −12.31 g, σ: 7.03 g) and fish pie (E: −12.59 g, σ: 5.74 g). Mean: −9.58"	Visual Assessment
S. Fang [30]	2016	10 objects		"3D geometric models and depth images."
Godwin, S. [56]	2006	Five portions of 9-inch cake, Seven portions of pizza, Pies were 9 or 10 inches	E: 25%	Estimated portion sizes using a ruler and the adjustable wedge
Hernández, Teresita [37]	2006	101 subjects, 5 foods	E: 4.8% ± 1.8%	Digital photographs printed onto a poster.
Yang et al. [138]	2021	Virtual Food Dataset and Real Food Dataset (RFD) (1500 images)	E: <9% on VFD, E: 11.6% and 20.1% on RFD.	Estimates volume by computing inner product between the probability vector from modified MobileNetV2 and the reference volume vector.
Graikos et al. [139]	2021	EPIC-KITCHENS and their own food video datasets	46.32% average MAPE on 16 test foods and 36.90% average MAPE on 6 combined meals.	Generate 3-dimensional point cloud by using depth map, segmentation mask and camera parameters. It then approximates the volume with points cloud-to-volume algorithm.
Lo, F.P.W et al. [140]	2019	Test dataset: 11 food items	E: 15.32%.	3D point cloud completion from RGB and depth images.

7.1.1. Food Portion Estimation by Counting Pixels

This method utilizes pixel count in each relevant image section to estimate food portion size. Studies [120] show that these methods are less complex than methods that rely on 3D modeling. Despite its simplicity, it gives a good estimation of portion size, thus making calculation of caloric content and nutritional facts easier.

7.1.2. Visual Similarities between Target Image and Dictionary of Food Images

This method estimates visual similarities between a given image and an existing food image dictionary. It is used by many existing systems today [29], where the caloric and nutrient contents in the food image dictionary are defined by dietary professionals to get a better approximation. The method selects first 'n' images from the dictionary and calculates the calorie content of the target image based on the average calorie content of dictionary images.

7.1.3. 3D Modeling for Food Portion Estimation

This method projects a 3D model of food portions onto 2D space or uses 3D geometric models for volume estimation. Generally, this method gives finer approximation in contrast to the other methods for single-image-view methods.

7.1.4. Other Methods

Other methods for food-portion estimation include estimating portion sizes using a ruler and adjustable wedge [56], mobile augmented reality, virtual reality [33], visual assessment [137] feature extraction, and its matching [29,64].

7.2. Multi-Image View or Video Methods

Multi-Image view or video methods require multiple images for food portion estimation. They are relatively more accurate than single-view-image methods. However, multi-image methods are less user-friendly as they require multiple images from different viewpoints in order to provide better results. Table 8 summarizes single-view methods for volume estimation. The following are a few methods that use multi-image-view techniques for food volume estimation.

Table 8. Comparison of multi-view methods for food volume estimation.

Reference	Year	Dataset	Results (E: Error%)	Technique
F. Zhu [141]	2010	3000 images	E: 1% 19 food items (97.2%)	"Camera calibration step and a 3D volume reconstruction step"
Xu Chang [141]	2013	14 to 20 images for multi-view method	E: 7.4% to 57.3%	Multi-view volume estimation using "Shape from Silhouettes" to estimate the food portion size
Kong, Fanyu [12]	2015	6 food items	84–91%	Multi-View RGB images for 3D reconstruction to estimate the volume
Trevno, Roberto [142]	2015	120 students (n = 120 meals; 57 breakfast + 63 lunch)	74% (reliability)	Digital Food Imaging Analysis (DFIA)
Jia, W. Y [143]	2014	100 food samples	E: −2.80% 30%	ebutton is used for taking pictures, and then portion size is calculated semi-automatically by using computer software
Xu, C [36]	2013		E: 10%	3D MODELLING AND POSE ESTIMATION
Rhyner, D [144]	2016	6 meals	85.10%	Multi-View RGB images, reference card and 3D model for volume estimation
T. Stutz [60]	2014	Rice, blinded servings	E: <33%	Mobile Augmented Reality System
Makhsous et al. [145]	2020	8 food items tested	40% improvement in the accuracy of volume estimation as compared to manual calculation.	Employs a mobile Structured Light System (SLS) to measure the food volume and portion size of a dietary intake.
Yuan et al. [146]	2021	Test dataset: 6 food items	E: 0.83 5.23%.	3D reconstruction from multi-view RGB images.
Lo, F.P.W et al. [140]	2019	Test dataset: 11 food items	E: 15.32%.	3D point cloud completion from RGB and depth images.

7.2.1. Food Volume Estimation Using 3D Geometric Models

This multi-image-view method uses a shape template method or 3D modeling for portion size estimation. As a single shape template is not suitable for all food types, the use of geometric models with correct food classification labels and segmentation masks in the image is important to index food labels to their respective classes of predefined geometric models. These can be used later for finding correct parameters of the selected geometric model [28,40,41,56,62].

Moreover, in 3D modeling and pose estimation, models for food are constructed in advance by using between 15 and 20 food images captured from several angles or a video sequence. Finally, food volume is estimated by registering pose from 3D models to 2D images [36].

7.2.2. Augmented Reality System for Food Volume Estimation

The use of augmented reality is also being widely used by researchers to estimate food portion size. Many systems such as Eat AR make use of it for portion size estimation [60] by developing prototypes to aid users. These prototypes generally require fiducial markers or credit-card-sized objects for overlaying 3D forms. Finally, the volume of the overlaid forms is computed using a signed volume estimation algorithm for closed 3D objects.

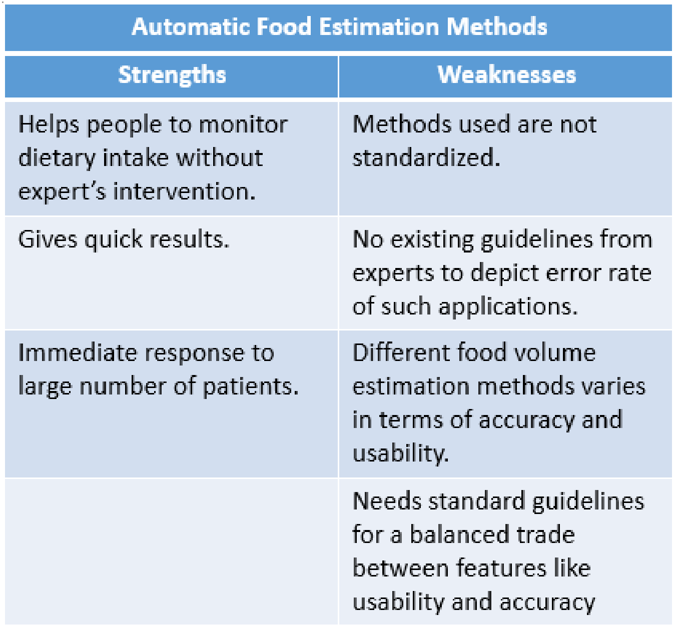
Similarly, the 'Serv Ar' augmented reality tool is used to provide guidance about food serving size [147]. Many of these technologies are being used with object recognition methods to identify food items and determine their caloric content. Similarly, methods that use augmented reality in combination with other portion estimation techniques have enhanced accuracy and much more interactive interfaces, resulting in a high retention rate.

7.2.3. Food Portion Estimation Using 3D Reconstruction (Dense Models)

Portion estimation by constructing dense 3D models usually requires multiple images or a video segment [139]. Joachim Dehais et al. [148] have shown the use of two views for volume estimation using 3D construction. In its first stage, the system learns about the configuration of different views, followed by the construction of a dense 3D model to extract the volume of each individual food item placed before it. Similarly, Wen Wu et al. [32] studied the use of fast food videos for caloric estimation. Most of these methods require images from different viewpoints, and for this reason, more advanced methods such as 3D construction from accidental motion can be explored for food volume estimation in the future.

7.3. Strengths and Weakness of the Food Volume Estimation Methods

Automatic food volume estimation method helps people to monitor their dietary intake suffering from chronic diseases without any expert intervention. It gives a quick result as compared to the traditional method which generally involves sending food images to the dietitian. The traditional method involves continuous involvement of dietitians, which makes it unworkable for dietitians to immediately respond to a large number of patients. Conversely, automatic food volume estimation is not standardized, as there are no existing guidelines by experts that refer to the error rate of these applications. Furthermore, different volume estimation methods vary in terms of accuracy and usability. Most of these methods are classified into two categories: single-image-view method and multiple-image-view method. Single-view-image methods are more user friendly, but their accuracy is compromised compared to multiple image view methods as it requires images from different. Therefore, standard guidelines are required for food volume estimation, which should include criteria for a balanced trade between features such as usability and accuracy, and developed applications must be verified according to the standard guidelines. Figure 7 summarizes the strengths and weaknesses of food volume estimation methods.



Automatic Food Estimation Methods	
Strengths	Weaknesses
Helps people to monitor dietary intake without expert's intervention.	Methods used are not standardized.
Gives quick results.	No existing guidelines from experts to depict error rate of such applications.
Immediate response to large number of patients.	Different food volume estimation methods varies in terms of accuracy and usability.
	Needs standard guidelines for a balanced trade between features like usability and accuracy

Figure 7. Strengths and weaknesses of automatic food estimation methods.

8. Existing and Potential Applications of Vision-Based Methods for Food Recognition in Healthcare

We summarized the core applications of vision-based methods for food recognition in the context of public policy and health care.

8.1. mHealth Apps for Dietary Assessment

Today, several mobile applications have been developed to monitor diet and help users to choose healthier alternatives regarding food consumption. Initially, these mobile applications were dependent on manually inputting food items by selecting from limited food databases. Therefore, such applications were not very reliable as they were prone to inaccuracies in dietary assessment, mainly extending from limited exposure to numerous food categories. With the advancement in the area of food image recognition, a large number of mHealth applications for dietary assessment use images to recognize food categories. For this purpose, existing mobile applications use different combinations of traditional and deep visual feature extraction, and classification methods for food recognition described earlier in Sections 3 and 4. Aizawa et al. [149] developed a mobile app food log, which uses traditional feature-extraction methods such as color, Bag of Features, and SIFT and uses an Adaboost classifier for classification purposes. Similarly, Ravi et al. [150] proposed the 'FoodCam' application, which uses traditional methods for feature extraction (LBP and RGB color features) and SVM for classification. Alternatively, Meyers et al. [13] employed a deep visual technique (GoogleNet CNN model) for feature extraction and classification purposes. Similarly, the Food Tracker app proposed by Jiang et al. [151] uses a deep convolutional neural network for feature extraction and classification. Furthermore, G. A. Tahir and C. K. Loo [52] utilized deep visual methods such as ResNet-50, DenseNet201, and InceptionResNet-V2 for feature extraction and Adaptive Reduced Class Incremental Kernel Extreme Learning Machine (ARCIKELM) as a classification method for their mobile application "My Diet Cam". Table 9 summarizes existing mobile applications in terms of feature extraction and classification methods used. Based on these deep visual method combinations, food recognition accuracies differ for various existing mobile applications. Therefore, apps with higher food recognition and classification accuracies gain more popularity. These apps tend to ease the dietary assessment process. Figure 8 shows the mobile application by Ravi et al. [150].

Table 9. Summary of feature extraction and classification methods used by existing mobile applications.

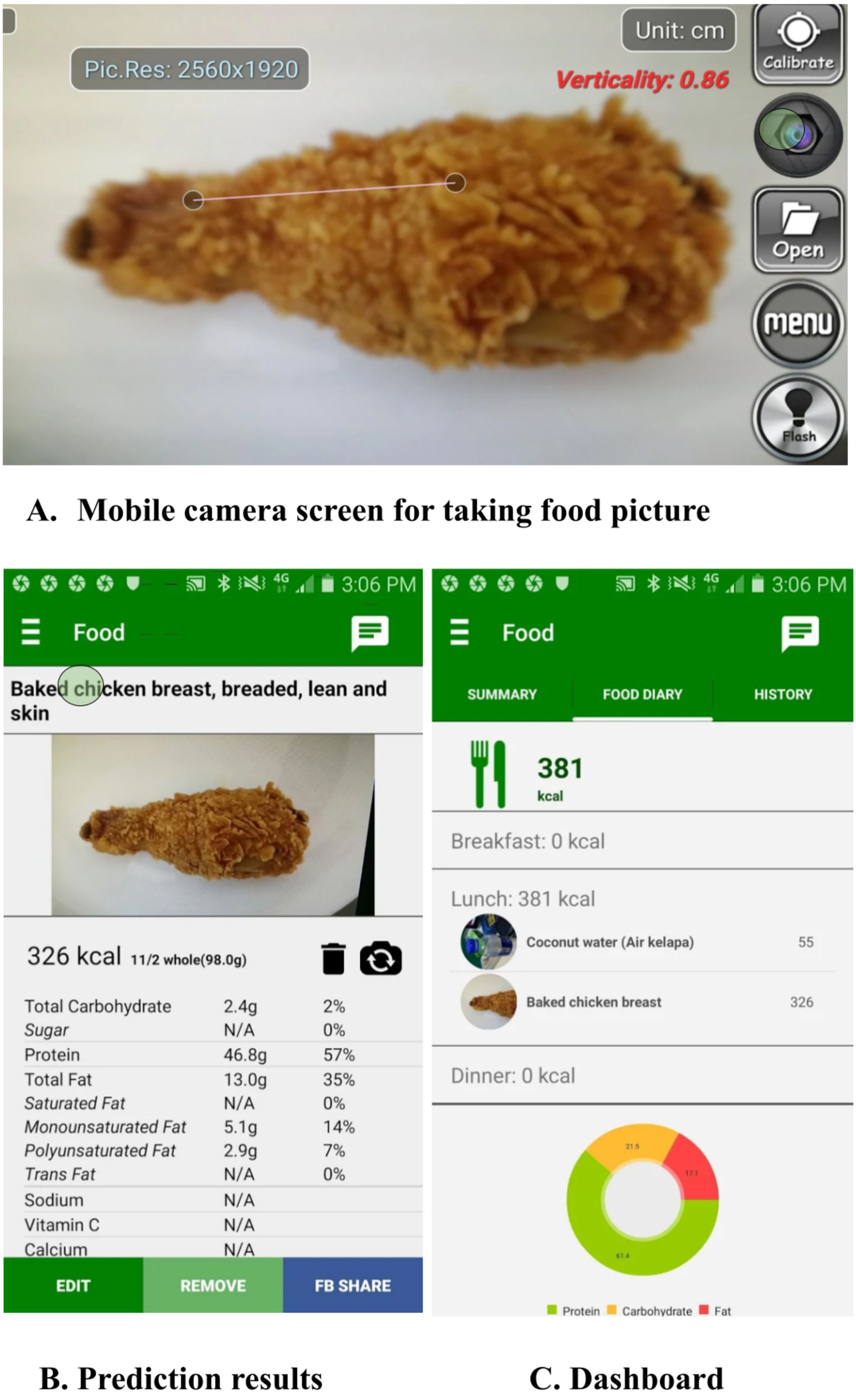
Reference	Year	Application Name	Food Segmentation	Feature Extraction Method	Classification Method
Aizawa et al. [149]	2013	FoodLog	No	Color, SIFT and Bag of Features	Adaboost Classifier
Oliveira et al. [83]	2014	-	Yes	Color and Texture	Support Vector Machine (SVM)
Probst et al. [152]	2015	-	-	SIFT, LBP and Color	Linear SVM
Meyers et al. [13]	2015	Im2Calories	Yes	GoogleNet CNN	GoogleNet CNN model
Ravi et al. [150]	2015	FoodCam	No	HoG, LBP and RGB Color Features	Linear SVM
Waltner et al. [55]	2017	-	Yes	RGB, HSV and LAB Color values	Random Forest Classifier
Mezgec and Seljak [153]	2017	-	-	NutriNet	NutriNet
Pouladzadeh et al. [154]	2017	-	Yes	CNN	Caffe Framework
Waltner et al. [155]	2017	-	Yes	CNN	CNN
Ming et al. [11]	2018	DietLens	-	ResNet-50 CNN	ResNet-50 CNN
Jiang et al. [151]	2018	-	Yes	Colors, Lines, Points, SIFTand Texture Features	Reverse Image Search (RIS) and Text Mining
Jianing Sun et al. [156]	2019	Food Tracker	Yes	DCNN	DCNN
G. A. Tahir and C.K. Loo [52]	2020	MyDietCam	Yes	ResNet-50, DenseNet201 and Inception ResNet-V2	Adaptive Reduced Class Incremental Kernel Extreme Learning Machine (ARCIKELM)

8.2. Harnessing Vision-Based Method to Measure Nutrient Intake during COVID-19

As the COVID-19 is a leading global challenge across the world, maintaining good nutritional status is mandatory for keeping good health to fight against the virus. Automatic vision-based methods for volume estimation and food image recognition in these nutrition tracking apps can assist patients in objectively measuring the nutrient intake of vital vitamins required for boosting the immune system.

8.3. Life's Simple 7

Life's Simple 7 health score is recently introduced based on modifiable health factors that contribute to heart health. Physical activity, non-smoking status, healthy diet, and body mass index are four modifiable health behaviors in this score. The other three modifiable factors are biological. They include blood pressure, fasting glucose, and cholesterol details. Besides cardiovascular health, Life's Simple 7 also relates to other health conditions such as venous thromboembolism, cognitive health, atherosclerosis, etc. As dietary intake plays a vital role in computing Life's Simple 7, manually measuring these factors and then calculating a Life's Simple 7 score is a very tedious process. This makes it very difficult for both middle-aged patients and elderly patients to keep track of their health. So vision-based methods can play an important role in automating the diet score. However, there are no current studies that have explored this research direction.



A. Mobile camera screen for taking food picture

3:06 PM

Food

Baked chicken breast, breaded, lean and skin

326 kcal 11/2 whole(98.0g)

Total Carbohydrate	2.4g	2%
Sugar	N/A	0%
Protein	46.8g	57%
Total Fat	13.0g	35%
Saturated Fat	N/A	0%
Monounsaturated Fat	5.1g	14%
Polyunsaturated Fat	2.9g	7%
Trans Fat	N/A	0%
Sodium	N/A	
Vitamin C	N/A	
Calcium	N/A	

3:06 PM

Food

381 kcal

Breakfast: 0 kcal

Lunch: 381 kcal

Coconut water (Air kelapa)	55
Baked chicken breast	326

Dinner: 0 kcal

B. Prediction results

C. Dashboard

Figure 8. The application provides the top prediction result. This picture is taken from the study of Ghalib et al., 2020 (permission has been obtained from original author).

8.4. Enforcing Eating Ban on Public Places during COVID-19 Pandemic or Other Restricted Places

Vision-based food recognition can automate the enforcement of an eating ban at public places by automatically detecting foods from CCTV and wearable cameras to curb the spread of the virus. Similarly, vision-based food recognition coupled with CCTV or wearable cameras and smart apps automate the enforcement of eating bans at workplaces, laboratories, etc.

8.5. Monitoring Malnutrition in Low-Income Countries

Coupling vision-based methods with wearable cameras can automatically detect foods from egocentric images with reasonable accuracy while reducing the burden of processing big data and addressing the user's privacy concerns. Egocentric images acquired from these cameras are important to study diet and lifestyle, especially in low-income countries with a high malnutrition rate. For example, Jia et al. [157] focused on gathering image data from wearable cameras and discriminating between food/non-food classes based on their tag from the CNN to study human diets. Similarly, Chen et al. [158] studied malnutrition in low- and middle-income countries by using the wearable device e-button.

8.6. Food Image Analysis from Social Media

We are in the era of social media, and food is a basic necessity of life, a great deal of content on social media platforms is related to food items. User's of these platforms frequently share new recipes, new methods of cooking, food pictures after restaurant check-in. Researchers have exploited this data on social media platforms for analyzing dietary intake. For example, Mejova et al. [159] studied food images from foursquare and Instagram to analyze the food consumption pattern in the USA. Similarly, food images on social media platforms are of different cultures. These images can be crawled and then combined together to prepare a large food database.

8.7. Food Quality Assessment

Evaluating fruit quality and freshness at the marketplace and at the user end is of increasing interest as opposed to accessing quality at the time of manufacturing. Efforts to date have focused on accessing the quality of foods using vision-based methods. For example, Ismail et al. have contributed an Apple-NDDA dataset [160] that consists of defective and non-defective apple images for food quality assessment.

9. Statistical Analysis

We provide a statistical analysis of our study based on the articles and conference proceedings gathered to write this survey paper. We surveyed research studies up to 2020 from various reputed sources: IEEE, Elsevier, ACM, and Web of Sciences. Figure 9 shows a pie chart of the distribution of surveyed food databases according to the country to which the food dishes belong. In it, generic databases are those that contain food dishes of multiple countries. We summarized the surveyed studies in two main categories: studies using handcrafted features, and studies using visual feature representation from convolutional neural networks (CNN), as shown in Figure 10. As discussed in Section 7, volume estimation methods require a single view or multiple images from different viewpoints. We presented a pie chart as shown in Figure 11 that describes the percentage of studies we surveyed according to the number of image viewpoints required to estimate food volume. For ingredient detection, all included studies used CNN due to recent interest in this extension. Similarly, for studies that have implemented mobile applications, the piechart in Figure 12 shows that 46.2% of applications implement CNN for food recognition while remaining mobile applications from surveyed studies are implementing traditional methods for feature extraction.

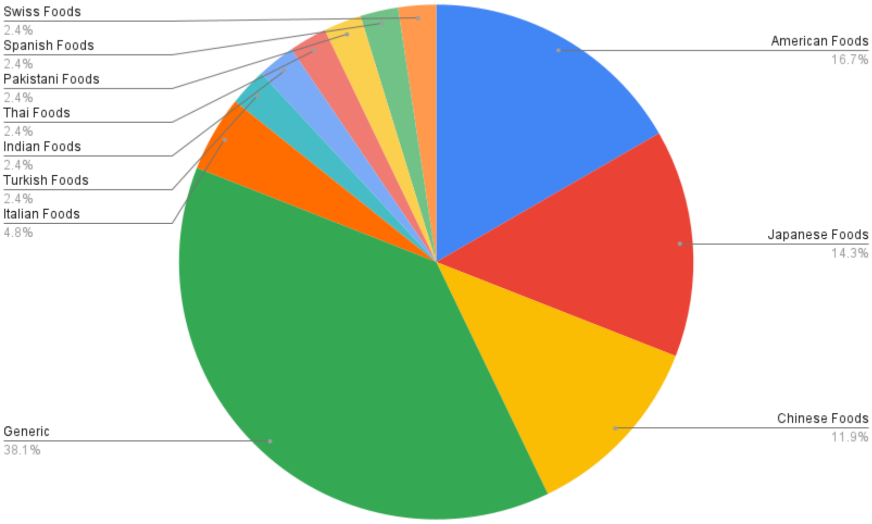


Figure 9. Percentage of datasets summarized according to the types of food. Generic refers to the multi-cultural dataset.

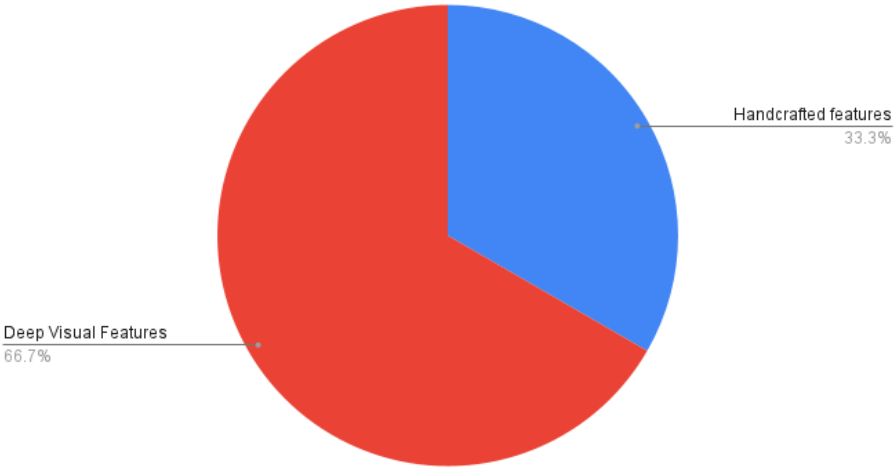


Figure 10. Percentage of studies summarized according to the type of feature extraction methods.

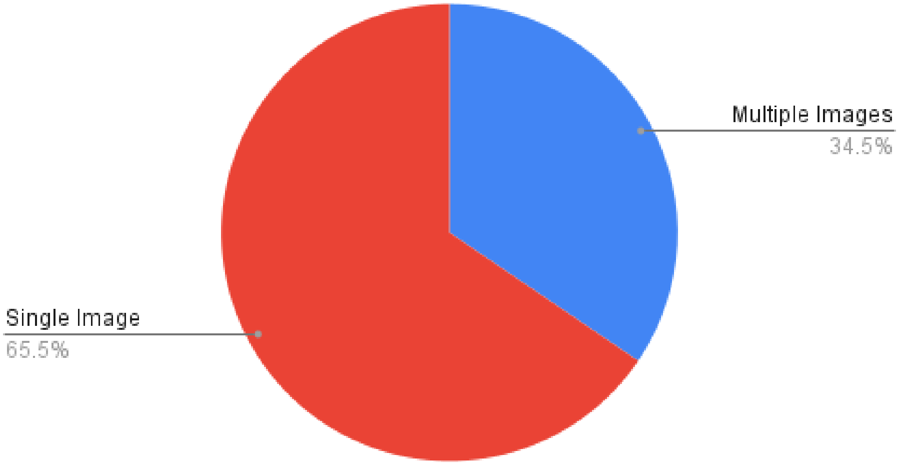


Figure 11. Volume estimation methods using single images vs. multiple images.

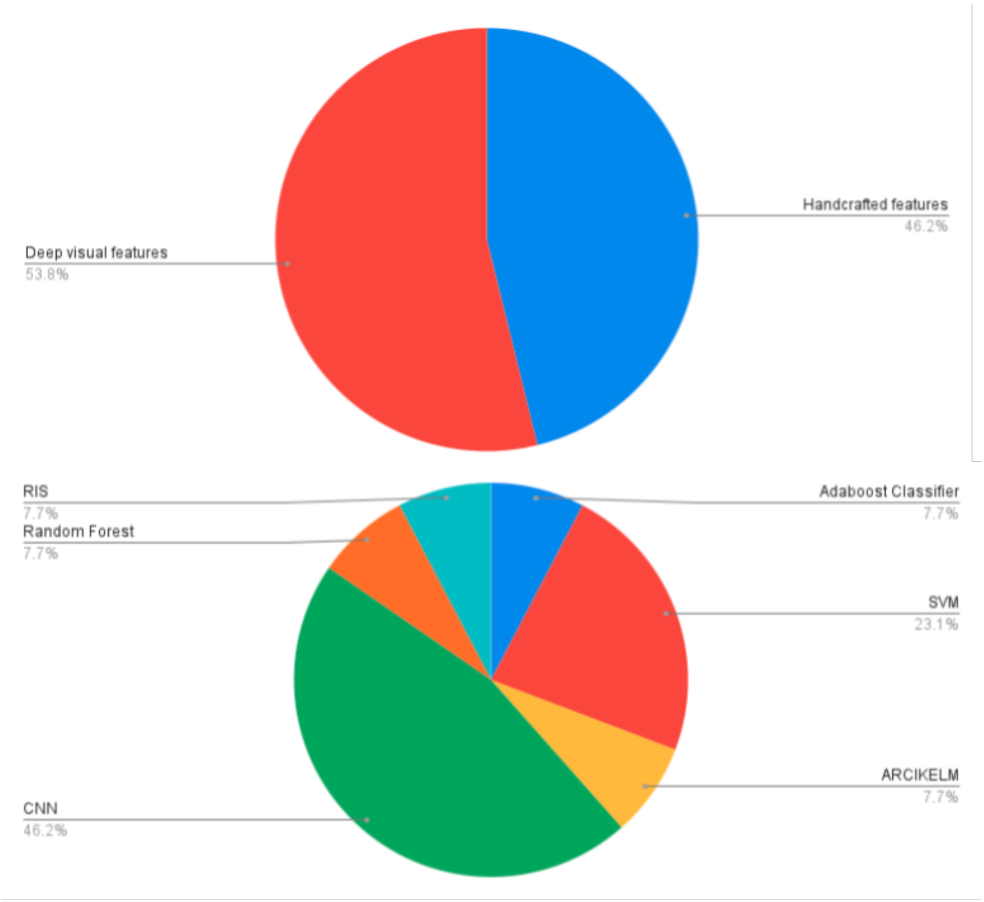


Figure 12. Percentage of studies summarized according to the type of methods employed for feature extraction from food images and the category of classifier used for food image analysis in a mobile application.

10. Open Issues

This study highlighted open issues based on the survey papers and the authors' first-hand experience with existing methodologies.

10.1. Unsupervised Learning from Unlabelled Dataset

Preparing a large comprehensive annotated data is still a challenge, as manually annotating a dataset is a difficult task with many challenges. Due to the large variety of food dishes, different styles of preparation, etc., it is difficult for an expert dietician to correctly label all the foods, especially in the preparation of a multi-culture food database. Similarly, it involves high costs and a large number of working hours to prepare such a dataset. Recent advancements in contrastive learning have opened a new research paradigm of unsupervised learning. Methods based on contrastive learning such as SimCLR [161] and SwAV [162] do not require labeled datasets and seem to be interesting potential areas of research that future works in food recognition should exploit.

10.2. Continual Learning

Food datasets are open-ended, and there is no cap on the number of dishes. So the network must adapt to continuously evolving datasets. All of these properties of food datasets have made them a strong use case for continual learning methods. One of the principal challenges in continuous learning methods is catastrophic forgetting. Catastrophic forgetting refers to completely or abruptly forgetting previously learned information while learning new classes. Many neural networks are susceptible to forgetting during continual learning. It is a prime hindrance in achieving the objective of continuously evolving networks similarly to those of humans. Hence, researchers should also study catastrophic forgetting in the context of food databases.

10.3. Explainability

Although there have been numerous attempts, including activation methods, SHAP values [163], and distillation methods, there is still a research gap in the context of food recognition. As food recognition has many domain-specific challenges such as intraclass variations, and non-rigid structure, visualization of the reasoning behind model predictions is vital to trust its decisions. Recently, unsupervised clustering methods [164] are exploited to explain model predictions by distilling knowledge into surrogate models. They provide similar images to test images for explaining prediction results. Explaining prediction results by showing images similar to test images seems more friendly as users do not need any specific domain knowledge to understand these results.

11. Discussion

Our research provides deep insight into computer vision-based approaches for dietary assessment. It focuses on both traditional and deep learning methodologies for feature extraction and classification methods used for food image recognition and single- and multi-view methods for volume estimation. Similarly, this survey also explores and compares current food image datasets in detail, as vision-based techniques are highly dependent on a comprehensive collection of food images. In contrast to previous research work, such as work by Mohammad A. Sobhi et al. [165], Min, Weiqing, et al. [166], our survey scrutinizes traditional and current deep visual approaches for feature extraction and classification to enhance clarity in terms of their performance and feasibility. Unlike existing surveys, our survey emphasizes existing solutions developed for food ingredient recognition through multi-label learning. We also reviewed existing computer-based food volume estimation methods in detail, as they have reduced dietitians' and experts' intervention and can accurately determine the portion size of the food in contrast to the self-estimation. Finally, our research study also explores real-world applications using the prior methodologies for dietary assessment purposes.

11.1. Findings

Our findings indicate that the ultimate performance of traditional and deep visual techniques depends on the type of dataset used. This has been observed from the datasets included from the studies explored in this survey (as shown in Table 1); the three most commonly used datasets were UECFOOD-256 [43], UECFOOD-100 [42], and Food-101 [59]. UECFOOD-256 (25,088 images and 256 classes) and UECFOOD-100 (14,361 images and 100 classes of food) are Japanese food datasets consisting of Japanese food images captured by users, whereas Food-101(101,000 images and 101 classes) is an American fast food dataset containing images crawled from several websites. However, these widely used datasets are region-specific. Therefore, there is an immense need for generic food datasets for excluding regional bias from experimental results. In addition, it is also evident from this survey that deep visual techniques have replaced traditional machine learning methodologies for food image recognition. As per our survey, systems proposed after the year 2015 mainly use deep learning technologies for food classification purposes. This is due to their phenomenal classification performance. In the context of classification performance of deep visual techniques, for food–non-food classification, McAllister et al., 2018 [108] (99.4%), and Pouladzadeh et al., 2016 [104] (99%), achieved the highest top 1 classification accuracy. Pouladzadeh et al., 2016 [104], used DCNN and Graph cut on their proposed dataset, whereas McAllister et al., 2018 [108], used CNN, ANN, SVM, and random forest on the food 5k dataset. Table 5 further compares classification accuracies of proposed deep visual models. Recent advancements and exceptional performance of food image classification methods have now led researchers to explore food images from a much deeper perspective in terms of retrieval and classification of food ingredients from food images. Therefore, we have also explored several proposed solutions for food ingredient recognition and classification. According to our survey, the system proposed by Chen et al., 2016 [47], has achieved the highest F1 score, i.e., 95.88% macro-F1 and 82.06% micro-F1, using the Arch-D method on the UECFOOD-100 dataset (as shown in Table 6). Similarly, automatic food volume estimation methods have reduced dietitians' and experts' intervention and can accurately determine the portion size of the food in contrast to the self-estimation for food volume estimation. Single-view methods involve capturing a single image, while multi-views require multiple images to determine accurate food volumes. The results in Table 8 show that multi-view methods are mostly better than single-view methods.

Finally, food category recognition, ingredient classification, and volume estimation techniques helped provide an automatic dietary assessment with reduced human intervention in mHealth apps. For this purpose, we have also surveyed several mobile applications that employ deep learning methods for dietary assessment.

11.2. Limitations and Future Research Challenges

Despite enhanced performance and classification accuracy, food image recognition and volume estimation through vision-based approaches may continue to present interesting future research challenges. This is because the performance of the methodologies used for food image identification is highly dependent on the source of images in a particular food dataset. Although a growing number of food categories are being incorporated into food image datasets such as UECFOOD-256 [43], Food 85 [49], and Food201-segmented [13], there is still an immense need for generalized, comprehensive datasets for better performance evaluation and benchmarking. Moreover, we observed that datasets with a large number of food images significantly positively impact classification accuracy. However, keeping these large image datasets updated is another challenge, especially since different types of foods are being prepared every day.

In addition to this, progressive learning during the classification phase is vital for food image datasets due to the continuous arrival of new concepts and domain variation within existing concepts. Similarly, developing frameworks interpretable by highlighting the contribution of the area of interest will improve the overall human trust level on a solution in a real-world environment.

Following food recognition, food volume estimation is a particularly complex and challenging assignment since food items have large variations in shape, texture, and appearances. Our article categorized food portion estimation methods into single-view and multi-view methods. Multi-view methods are more accurate; however, most of these methods also require calibration objects each time and images from different viewpoints, which makes the usability of these solutions tedious for elderly users.

Finally, there is a need to design and develop solutions that can respond to situations ethically. In our context, this refers to the removal of any biases concerning region-specific food preferences. It will help to ensure transparency in existing models.

12. Conclusions

In this work, we explored a broad spectrum of vision-based methods that are specifically tailored for food image recognition and volume estimation. In practice, the food recognition process incorporates four tasks: acquiring food images from the corresponding food datasets, feature extraction using handcrafted or deep visual, selection of relevant extracted features, and finally, appropriate selection of classification technique using either traditional machine learning approach or deep learning models followed by food ingredient classification to provide better insight of nutrient information. The findings of surveyed studies have shown that 38.1% of datasets are generic, which includes multicultural food dishes. Similarly, 46.2% of surveyed applications implemented CNN for food recognition, while 45.2% of mobile applications have implemented traditional methods for feature extraction. For ingredient detection, several studies used CNN due to its superior performance and recent interest. In addition, 34.5% of techniques for volume estimation require multiple images, while the remaining methods used a single image to estimate food volume.

Despite impeccable performance exhibited by state-of-the-art approaches, there exist several limitations and challenges. There is an immense need for comprehensive datasets for benchmarking and performance evaluation of these models, as incorporating large food image datasets improves the overall performance. Consequently, when dealing with open-ended and dynamic food datasets, the classifier must be capable of open-ended continuous learning. However, existing methods have several bottlenecks, which undermine the food-recognition ability when it comes to open-ended learning, as proposed methods are prone to catastrophic forgetting. They tend to forget previous knowledge extracted from images while learning new information. Such methods work well only for fixed food image datasets. Moreover, our findings indicate that proposed techniques for food ingredient classification still struggle with performance issues when applied to prepared and mixed food items. Survey findings further indicate that CNN models employed for visual feature extraction require labeled datasets for fine-tuning and training. Preparing a labeled food dataset is a difficult task due to the large variety of food dishes. To tackle this problem, unsupervised methods based on contrastive learning seem to have good research potential.

Similarly, automatic food portion estimation methods are categorized into two major categories: single-view-image methods and multi-view-image methods. As discussed earlier, most of multi-view image methods are more accurate than single view methods, but multi-view-image methods require complex processing and images from different angles, resulting in a reduced user retention rate. Furthermore, most of the single and multi-view methods require calibration objects each time, which has made the usability of these solutions tedious for elderly patients.

Therefore, there is substantial room for innovative health care and dietary assessment applications that can integrate wearable devices with a smartphone to revolutionize this research area. Moreover, dietary assessment systems should address these challenges to provide better insights into effective health maintenance and chronic disease prevention.

Author Contributions: G.A.T. was responsible for the literature search and writing the article and approved the final version as submitted. C.K.L. contributed to the study design, reviewed the study for intellectual content, and confirmed the final version as submitted. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the UM Partnership Grant: Project No: RK012-2019 from University of Malaya, IIRG Grant (IIRG002C-19HWB) from University of Malaya, International Collaboration Fund for project Developmental Cognitive Robot with Continual Lifelong Learning (IF0318M1006) from MESTECC, Malaysia, and ONRG grant (Project No.: ONRG-NICOP- N62909-18-1-2086)/IF017-2018 from Office of Naval and Research Global, UK.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors wish to confirm that there are no conflicts of interest.

References

1. Hajat, C.; Stein, E. The global burden of multiple chronic conditions: A narrative review. *Prev. Med. Rep.* **2018**, *12*, 284–293. [CrossRef]
2. Hall, J.E.; do Carmo, J.M.; da Silva, A.A.; Wang, Z.; Hall, M.E. Obesity-induced hypertension: Interaction of neurohumoral and renal mechanisms. *Circ. Res.* **2015**, *116*, 991–1006. [CrossRef] [PubMed]
3. Al-Goblan, A.S.; Al-Alfi, M.A.; Khan, M.Z. Mechanism linking diabetes mellitus and obesity. *Diabetes Metab Syndr. Obes.* **2014**, *7*, 587–591. [CrossRef]
4. Akil, L.; Ahmad, H.A. Relationships between obesity and cardiovascular diseases in four southern states and Colorado. *J. Health Care Poor Underserved.* **2011**, *22*, 61–72. [CrossRef] [PubMed]
5. De Pergola, G.; Silvestris, F. Obesity as a major risk factor for cancer. *J. Obes.* **2013**, *2013*, 291546. [CrossRef]

6. World Health Organization (WHO). Obesity and Overweigh. Available online: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight (accessed on 23 August 2018).
7. Ngo, J.; Engelen, A.; Molag, M.; Roesle, J.; García-Segovia, P.; Serra-Majem, L. A review of the use of information and communication technologies for dietary assessment. *Br. J. Nutr.* **2009**, *101* (Suppl. 2), S102–S112. [CrossRef] [PubMed]
8. Mendi, E.; Ozyavuz, O.; Pekesen, E.; Bayrak, C. Food intake monitoring system for mobile devices. In Proceedings of the 5th IEEE International Workshop on Advances in Sensors and Interfaces IWASI, Bari, Italy, 13–14 June 2013; pp. 31–33. [CrossRef]
9. Haapala, I.; Barengo, N.C.; Biggs, S.; Surakka, L.; Manninen, P. Weight loss by mobile phone: A 1-year effectiveness study. *Public Health Nutr.* **2009**, *12*, 2382–2391. [CrossRef] [PubMed]
10. Chen, Y.S.; Wong, J.E.; Ayob, A.F.; Othman, N.E.; Poh, B.K. Can Malaysian young adults report dietary intake using a food diary mobile. application? A pilot study on acceptability and compliance. *Nutrients* **2017**, *9*, 62. [CrossRef]
11. Ming, Z.Y.; Chen, J.; Cao, Y.; Forde, C.; Ngo, C.W.; Chua, T.S. Food photo recognition for dietary tracking: System and experiment. In *Multimedia Modeling*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 129–141.
12. Kong, F.; Tan, J. DietCam: Automatic Dietary Assessment with Mobile Camera Phones. *Pervasive Mob. Comput.* **2012**, *8*, 147–163. [CrossRef]
13. Meyers, A.; Johnston, N.; Rathod, V.; Korattikara, A.; Gorban, A.; Silberman, N.; Guadarrama, S.; Papandreou, G.; Huang, J.; Murphy, K.P. Im2Calories: Towards an Automated Mobile Vision Food Diary. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1233–1241. [CrossRef]
14. Martinel, N.; Micheloni, C. Classification of local eigen-dissimilarities for person re-identification. *IEEE Signal Process. Lett.* **2015**, *22*, 455–459. [CrossRef]
15. Martinel, N.; Das, A.; Micheloni, C.; Roy-Chowdhury, A.K. Re-Identification in the function space of feature warps. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1656–1669. [CrossRef] [PubMed]
16. Martinel, N.; Micheloni, C.; Foresti, G.L. Kernelized Saliency-Based Person Re-Identification Through Multiple Metric Learning. *IEEE Trans. Image Process.* **2015**, *24*, 5645–5658. [CrossRef] [PubMed]
17. Mahabir, S.; Baer, D.J.; Giffen, C.; Subar, A.; Campbell, W.; Hartman, T.J.; Clevidence, B.; Albanes, D.; Taylor, P.R. Calorie intake misreporting by diet record and food frequency questionnaire compared to doubly labeled water among postmenopausal women. *Eur. J. Clin. Nutr.* **2006**, *60*, 561–565. [CrossRef] [PubMed]
18. Bandini, L.G.; Must, A.; Cyr, H.; Anderson, S.E.; Spadano, J.L.; Dietz, W.H. Longitudinal changes in the accuracy of reported energy intake in girls 10–15 y of age. *Am. J. Clin. Nutr.* **2003**, *78*, 480–484. [CrossRef]
19. Champagne, C.M.; Baker, N.B.; DeLany, J.P.; Harsha, D.W.; Bray, G.A. Assessment of energy intake underreporting by doubly labeled water and observations on reported nutrient intakes in children. *J. Am. Diet Assoc.* **1998**, *98*, 426–433. [CrossRef]
20. Champagne, C.M.; Bray, G.A.; Kurtz, A.A.; Monteiro, J.B.; Tucker, E.; Volaufova, J.; Delany, J.P. Energy intake and energy expenditure: A controlled study comparing dietitians and non-dietitians. *J. Am. Diet Assoc.* **2002**, *102*, 1428–1432. [CrossRef]
21. Subar, A.F.; Kipnis, V.; Troiano, R.P.; Midthune, D.; Schoeller, D.A.; Bingham, S.; Sharbaugh, C.O.; Trabulsi, J.; Runswick, S.; Ballard-Barbash, R.; et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: The OPEN study. *Am. J. Epidemiol.* **2003**, *158*, 1–13. [CrossRef]
22. Blanton, C.A.; Moshfegh, A.J.; Baer, D.J.; Kretsch, M.J. The usda automated multiple-pass method accurately estimates group total energy and nutrient intake. *J. Nutr.* **2006**, *136*, 2594–2599. [CrossRef] [PubMed]
23. Daugherty, B.L.; Schap, T.E.; Ettienne-Gittens, R.; Zhu, F.M.; Bosch, M.; Delp, E.J.; Ebert, D.S.; Kerr, D.A.; Boushey, C.J. Novel Technologies for Assessing Dietary Intake: Evaluating the Usability of a Mobile Telephone Food Record Among Adults and Adolescents. *J. Med. Internet Res.* **2012**, *14*, e58. [CrossRef]
24. Snyder, H. Literature review as a research methodology: An overview and guidelines. *J. Bus. Res.* **2019**, *104*, 333–339. [CrossRef]
25. Ronald, K.; Marc, M.; Angelina, A.; Tyler, H.; Christopher, K. Measuring Catastrophic Forgetting in Neural Networks. *arXiv* **2017**, arXiv:1708.02072.
26. Liew, W.S.; Loo, C.K.; Gryshchuk, V.; Weber, C.; Wermter, S. Effect of Pruning on Catastrophic Forgetting in Growing Dual Memory Networks. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8. [CrossRef]
27. Tsoumakas, G.; Katakis, I. Multi-label classification: An overview. *Int. J. Data Warehous. Min.* **2006**, *3*, 1–3. [CrossRef]
28. He, Y.; Xu, C.; Khanna, N.; Boushey, C.J.; Delp, E.J. Food image analysis: Segmentation, identification and weight estimation. In Proceedings of the 2013 IEEE International Conference on Multimedia and Expo (ICME), San Jose, CA, USA, 15–19 July 2013; pp. 1–6. [CrossRef]
29. Miyazaki, T.; de Silva, G.C.; Aizawa, K. Image-based Calorie Content Estimation for Dietary Assessment. In Proceedings of the 2011 IEEE International Symposium on Multimedia, Dana Point, CA, USA, 5–7 December 2011; pp. 363–368. [CrossRef]
30. Fang, S.; Zhu, F.; Jiang, C.; Zhang, S.; Boushey, C.J.; Delp, E.J. A comparison of food portion size estimation using geometric models and depth images. In Proceedings of the Image Processing (ICIP), Hoenix, AZ, USA, 25–28 September 2016; Volume 2016, pp. 26–30. [CrossRef]
31. Okamoto, K.; Yanai, K. An Automatic Calorie Estimation System of Food Images on a Smartphone. In Proceedings of the Madima'16: Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, Amsterdam, The Netherlands, 16 October 2016; pp. 63–70. [CrossRef]

32. Wu, W.; Yang, J. Fast food recognition from videos of eating for calorie estimation. In Proceedings of the 2009 IEEE International Conference on Multimedia and Expo, New York, NY, USA, 28 June–3 July 2009; pp. 1210–1213. [CrossRef]
33. Zhang, Z.; Yang, Y.; Yue, Y.; Fernstrom, J.D.; Jia, W.; Sun, M. Food volume estimation from a single image using virtual reality technology. In Proceedings of the 2011 IEEE 37th Annual Northeast Bioengineering Conference (NEBEC), Troy, NY, USA, 1 April 2011; pp. 1–2. [CrossRef]
34. Hippocrate, E.A.A.; Suwa, H.; Arakawa, Y.; Yasumoto, K. Food Weight Estimation using Smartphone and Cutlery. In Proceedings of the First Workshop on IoT-enabled Healthcare and Wellness Technologies and Systems (IoT of Health'16), Singapore, 30 June 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 9–14. [CrossRef]
35. Yue, Y.; Jia, W.; Sun, M. Measurement of food volume based on single 2-D image without conventional camera calibration. In Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 29 August 2012; pp. 2166–2169. [CrossRef]
36. Xu, C.; He, Y.; Khanna, N.; Boushey, C.J.; Delp, E.J. Model-based food volume estimation using 3D pose. In Proceedings of the 2013 IEEE International Conference on Image Processing, Melbourne, VIC, Australia, 15–18 September 2013; pp. 2534–2538. [CrossRef]
37. Hernández, T.; Wilder, L.; Kuehn, D.; Rubotzky, K.; Moser-Veillon, P.; Godwin, S.; Thompson, C.; Wang, C. Portion size estimation and expectation of accuracy. *J. Food Compos. Anal.* **2006**, *19*, S14–S21. [CrossRef]
38. Zhang, W.; Yu, Q.; Siddiquie, B.; Divakaran, A.; Sawhney, H. Snap-n-Eat: Food Recognition and Nutrition Estimation on a Smartphone. *J. Diabetes Sci. Technol.* **2015**, *9*, 525–533. [CrossRef]
39. Huang, J.; Ding, H.; Mcbride, S.; Ireland, D.; Karunanithi, M. Use of Smartphones to Estimate Carbohydrates in Foods for Diabetes Management. *Stud. Health Technol. Inform.* **2015**, *214*, 121–127. [CrossRef]
40. Khanna, N.; Boushey, C.J.; Kerr, D.; Okos, M.; Ebert, D.S.; Delp, E.J. An Overview of The Technology Assisted Dietary Assessment Project at Purdue University. In Proceedings of the 2010 IEEE International Symposium on Multimedia, ISM 2010, Taichung, Taiwan, 13–15 December 2010; pp. 290–295. [CrossRef]
41. Jia, W.; Yue, Y.; Fernstrom, J.D.; Zhang, Z.; Yang, Y.; Sun, M. 3D localization of circular feature in 2D image and application to food volume estimation. In Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 29 August 2012; pp. 4545–4548. [CrossRef]
42. Matsuda, Y.; Yanai, K. Multiple-food recognition considering co-occurrence employing manifold ranking. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012; pp. 2017–2020. [CrossRef]
43. Kawano, Y.; Yanai, K. FoodCam-256: A Large-scale Real-time Mobile Food RecognitionSystem employing High-Dimensional Features and Compression of Classifier Weights. In Proceedings of the 22nd ACM international conference on Multimedia (MM'14), Orlando, FL, USA, 3–7 November 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 761–762. [CrossRef]
44. Chen, M.; Dhingra, K.; Wu, W.; Yang, L.; Sukthankar, R.; Yang, J. PFID: Pittsburgh fast-food image dataset. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009; pp. 289–292. [CrossRef]
45. Farinella, G.M.; Allegra, D.; Moltisanti, M.; Stanco, F.; Battiato, S. Retrieval and classification of food images. *Comput. Biol. Med.* **2016**, *77*, 23–39. [CrossRef]
46. Farinella, G.M.; Allegra, D.; Stanco, F. A Benchmark Dataset to Study the Representation of Food Images. In Proceedings of the International Workshop on Assistive Computer Vision and Robotics (ACVR) 2014, Zurigo, Switzerland, 6–12 September 2014. [CrossRef] [PubMed]
47. Chen, J.; Ngo, C. Deep-based Ingredient Recognition for Cooking Recipe Retrieval. In Proceedings of the 24th ACM International Conference on Multimedia (MM'16), Amsterdam, The Netherlands, 15–19 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 32–41. [CrossRef]
48. Chen, M.-Y.; Yang, Y.-H.; Ho, C.-J.; Wang, S.-H.; Liu, S.-M.; Chang, E.; Yeh, C.-H.; Ouhyoung, M. Automatic Chinese food identification and quantity estimation. In *SIGGRAPH Asia 2012 Technical Briefs (SA'12)*; Association for Computing Machinery: New York, NY, USA, 2012; pp. 1–4. [CrossRef]
49. Hoashi, H.; Joutou, T.; Yanai, K. Image recognition of 85 food categories by feature fusion. In Proceedings of the 2010 IEEE International Symposium on Multimedia, Taichung, Taiwan, 13–15 December 2010; pp. 296–301. [CrossRef]
50. Ciocca, G.; Napoletano, P.; Schettini, R. Food Recognition and Leftover Estimation for Daily Diet Monitoring. In Proceedings of the ICIAP 2015 International Workshops, BioFor, CTMR, RHEUMA, ISCA, MADiMa, SBMI, and QoEM, Genoa, Italy, 7–8 September 2015; Volume 9281, pp. 334–341. [CrossRef]
51. Güngör, C.; Baltacı, F.; Erdem, A.; Erdem, E. Turkish cuisine: A benchmark dataset with Turkish meals for food recognition. In Proceedings of the 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, Turkey, 15–18 May 2017; pp. 1–4. [CrossRef]
52. Tahir, G.A.; Loo, C.K. An Open-Ended Continual Learning for Food Recognition Using Class Incremental Extreme Learning Machines. *IEEE Access* **2020**, *8*, 82328–82346. [CrossRef]
53. Hou, S.; Feng, Y.; Wang, Z. VegFru: A Domain-Specific Dataset for Fine-Grained Visual Categorization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 541–549. [CrossRef]

54. Mureșan, H.; Oltean, M. Fruit recognition from images using deep learning. *Acta Univ. Sapientiae Inform.* **2018**, *10*, 26–42. [CrossRef]
55. Waltner, G.; Schwarz, M.; Ladstätter, S.; Weber, A.; Luley, P.; Lindschinger, M.; Schmid, I.; Scheitz, W.; Bischof, H.; Paletta, L. Personalized dietary self-management using mobile vision-based assistance. In Proceedings of the International Conference on Image Analysis and Processing, Catania, Italy, 11–15 September 2017; Springer: Berlin/Heidelberg, Germany, 2018; pp. 385–393. [CrossRef]
56. Godwin, S.; Chambers, E.T.; Cleveland, L.; Ingwersen, L. A new portion size estimation aid for wedgeshaped. *Foods. J. Am. Diet. Assoc.* **2006**, *106*, 1246–1250. [CrossRef]
57. Mariappan, A.; Bosch, M.; Zhu, F.; Boushey, C.J.; Kerr, D.A.; Ebert, D.S.; Delp, E.J. Personal Dietary Assessment Using Mobile Devices. In Proceedings of the Computational Imaging VII. International Society for Optics and Photonics, San Jose, CA, USA, 19–20 January 2009; Volume 7246, 72460Z. [CrossRef]
58. Bosch, M.; Schap, T.; Zhu, F.; Khanna, N.; Boushey, C.J.; Delp, E.J. Integrated database system for mobile dietary assessment and analysis. In Proceedings of the 2011 IEEE International Conference on Multimedia and Expo, Barcelona, Spain, 11–15 July 2011. [CrossRef]
59. Bossard, L.; Guillaumin, M.; van Gool, L. Food-101–mining discriminative components with random forests. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 446–461. [CrossRef]
60. Stütz, T.; Dinic, R.; Domhardt, M.; Ginzinger, S. Can mobile augmented reality systems assist in portion estimation? A user study. In Proceedings of the 2014 IEEE International Symposium on Mixed and Augmented Reality—Media, Art, Social Science, Humanities and Design (ISMAR-MASH'D), Munich, Germany, 10–12 September 2014; pp. 51–57. [CrossRef]
61. Wang, X.; Kumar, D.; Thome, N.; Cord, M.; Precioso, F. Recipe recognition with large multimodal food dataset. In Proceedings of the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy, 29 June–3 July 2015; pp. 1–6. [CrossRef]
62. Fang, S.; Liu, C.; Zhu, F.; Delp, E.J.; Boushey, C.J. Single-View Food Portion Estimation Based on Geometric Models. In Proceedings of the 2015 IEEE International Symposium on Multimedia (ISM), Miami, FL, USA, 14–16 December 2015; pp. 385–390. [CrossRef]
63. Herranz, L.; Xu, R.; Jiang, S. A probabilistic model for food image recognition in restaurants. In Proceedings of the 2015 IEEE International Conference on Multimedia and Expo (ICME), Turin, Italy, 29 June–3 July 2015; pp. 1–6. [CrossRef]
64. Beijbom, O.; Joshi, N.; Morris, D.; Saponas, S.; Khullar, S. Menu-Match: Restaurant-Specific Food Logging from Images. In Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 5–9 January 2015; pp. 844–851. [CrossRef]
65. Zhou, F.; Lin, Y. Fine-grained image classification by exploring bipartite-graph labels. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1124–1133. [CrossRef]
66. Ciocca, G.; Napoletano, P.; Schettini, R. Food Recognition: A New Dataset, Experiments and Results. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 588–598.
67. Wu, H.; Merler, M.; Uceda-Sosa, R.; Smith, J.R. Learning to Make Better Mistakes: Semantics-aware Visual Food Recognition. In Proceedings of the 24th ACM international conference on Multimedia (MM'16), Amsterdam, The Netherlands, 15–19 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 172–176. [CrossRef] [PubMed]
68. Singla, A.; Yuan, L.; Ebrahimi, T. Food/Non-food Image Classification and Food Categorization using Pre-Trained GoogLeNet Model. In Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management (MADiMa'16), Amsterdam, The Netherlands, 16 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 3–11. [CrossRef]
69. Rich, J.; Haddadi, H.; Hospedales, T.M. Towards Bottom-Up Analysis of Social Food. In Proceedings of the 6th International Conference on Digital Health Conference (DH'16), Montréal, QC, Canada, 11–13 April 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 111–120. [CrossRef]
70. Liang, Y.; Li, J. Computer vision-based food calorie estimation: Dataset, method, and experiment. *arXiv* **2017**, arXiv:1705.07632.
71. Pandey, P.; Deepthi, A.; Mandal, B.; Puhan, N.B. FoodNet: Recognizing Foods Using Ensemble of Deep Networks. *IEEE Signal Process. Lett.* **2017**, *24*, 1758–1762. [CrossRef]
72. Termritthikun, C.; Muneesawang, P.; Kanprachar, S. NUInNet: Thai food image recognition using convolutional neural networks on smartphone. *J. Telecommun. Electron. Comput. Eng. (JTEC)* **2017**, *9*, 63–67. [CrossRef]
73. Ciocca, G.; Napoletano, P.; Schettini, R. Learning CNN-based features for retrieval of food images. In Proceedings of the New Trends in Image Analysis and Processing—ICIAP 2017, Catania, Italy, 11–15 September 2017; pp. 426–434.
74. Yu, Q.; Anzawa, M.; Amano, S.; Ogawa, M.; Aizawa, K. Food Image Recognition by Personalized Classifier. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 171–175. [CrossRef]
75. Kaur, P.; Sikka, K.; Wang, W.; Belongie, S.; Divakaran, A. Foodx-251: A dataset for fine-grained food classification. *arXiv* **2019**, arXiv:1907.06167.
76. Available online: https://www.aicrowd.com/challenges/food-recognition-challenge (accessed on 23 August 2021).
77. Bolaños, M.; Radeva, P. Simultaneous Food Localization and Recognition. In Proceedings of the 23rd International Conference on Pattern Recognition (ICPR) 2016 (IN PRESS), Cancun, Mexico, 4–8 December 2016.
78. Aguilar, E.; Bolaños, M.; Radeva, P. Regularized Uncertainty-based Multi-Task LearningModel for Food Analysis. *J. Vis. Commun. Image R.* **2019**, *60*, 360–370. [CrossRef]

79. Kumar, G.; Bhatia, P.K. A Detailed Review of Feature Extraction in Image Processing Systems. In Proceedings of the 2014 Fourth International Conference on Advanced Computing & Communication Technologies, Rohtak, India, 8–9 February 2014; pp. 5–12. [CrossRef]
80. Yang, S.; Chen, M.; Pomerleau, D.; Sukthankar, R. Food recognition using statistics of pairwise local features. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2249–2256. [CrossRef]
81. Pouladzadeh, P.; Shirmohammadi, S.; Bakirov, A.; Bulut, A.; Yassine, A. *Cloud-Based SVM for Food Categorization. Multimedia Tools and Applications*; Springer: Berlin/Heidelberg, Germany, 2014. [CrossRef]
82. Kawano, Y.; Yanai, K. Real-Time Mobile Food Recognition System. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23–28 June 2013; pp. 1–7. [CrossRef]
83. Oliveira, L.; Costa, V.; Neves, G.; Oliveira, T.; Jorge, E.; Lizarraga, M. A mobile, lightweight, poll-based food identification system. *Pattern Recognit.* **2014**, *47*, 1941–1952. [CrossRef]
84. Martinel, N.; Piciarelli, C.; Micheloni, C. A supervised extreme learning committee for food recognition. *Comput. Vis. Image Underst.* **2016**, *148*, 67–86. [CrossRef]
85. Bosch, M.; Zhu, F.; Khanna, N.; Boushey, C.J.; Delp, E.J. Combining global and local features for food identification in dietary assessment. In Proceedings of the 2011 18th IEEE International Conference on Image Processing, Brussels, Belgium, 11–14 September 2011; Volume 2011, pp. 1789–1792. [CrossRef]
86. Kong, F.; Tan, J. DietCam: Regular Shape Food Recognition with a Camera Phone. In Proceedings of the 2011 International Conference on Body Sensor Networks, Chicago, IL, USA, 19–22 May 2011; pp. 127–132. [CrossRef]
87. Zhang, M.M. *Identifying the Cuisine of a Plate of Food*; Tech. Report; University of California San Diego: San Diego, CA, USA, 2011. [CrossRef]
88. Matsuda, Y.; Hoashi, H.; Yanai, K. Recognition of Multiple-Food Images by Detecting Candidate Regions. In Proceedings of the 2012 IEEE International Conference on Multimedia and Expo, Melbourne, VIC, Australia, 9–13 July 2012; pp. 25–30. [CrossRef]
89. Anthimopoulos, M.M.; Gianola, L.; Scarnato, L.; Diem, P.; Mougiakakou, S.G. A food recognition system for diabetic patients based on an optimized bag-of-features model. *IEEE J. Biomed. Health Inform.* **2014**, *18*, 1261–1271. [CrossRef]
90. Tammachat, N.; Pantuwong, N. Calories analysis of food intake using image recognition. In Proceedings of the 2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia, 7–8 October 2014; pp. 1–4. [CrossRef]
91. Pouladzadeh, P.; Shirmohammadi, S.; Yassine, A. Using graph cut segmentation for food calorie measurement. In Proceedings of the 2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Lisbon, Portugal, 11–12 June 2014; pp. 1–6. [CrossRef]
92. He, Y.; Xu, C.; Khanna, N.; Boushey, C.J.; Delp, E.J. Analysis of food images: Features and classification. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 2744–2748. [CrossRef]
93. Heaton, J. Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning. In *Genetic Programming and Evolvable Machines*; The MIT Press: Cambridge, MA, USA, 2016; Volume 19, 800p, ISBN 0262035618. [CrossRef]
94. Deng, L.; Yu, D. Deep learning: Methods and applications. *Found. Trends Signal Process.* **2014**, *7*, 197–387. [CrossRef]
95. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
96. Yanai, K.; Kawano, Y. Food image recognition using deep convolutional network with pre-training and fine-tuning. In Proceedings of the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy, 29 June–3 July 2015; pp. 1–6. [CrossRef]
97. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. [CrossRef]
98. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
99. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
100. Heravi, E.; Aghdam, H.; Puig, D. Classification of Foods Using Spatial Pyramid Convolutional Neural Network. *Artif. Intell. Res. Dev.* **2016**, *288*, 163–168. [CrossRef]
101. Jiang, S.; Min, W.; Liu, L.; Luo, Z. Multi-Scale Multi-View Deep Feature Aggregation for Food Recognition. *IEEE Trans. Image Process.* **2020**, *29*, 265–276. [CrossRef]
102. Kawano, Y.; Yanai, K. Food image recognition with deep convolutional features. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp'14 Adjunct), Seattle, WA, USA, 13–17 September 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 589–593. [CrossRef]
103. Christodoulidis, S.; Anthimopoulos, M.; Mougiakakou, S. Food Recognition for Dietary Assessment. Using Deep. *Convolutional Neural Netw.* **2015**, *9281*, 458–465. [CrossRef]
104. Pouladzadeh, P.; Kuhad, P.; Peddi, S.V.B.; Yassine, A.; Shirmohammadi, S. Food calorie measurement using deep learning neural network. In Proceedings of the 2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings, Taipei, Taiwan, 23–26 May 2016; pp. 1–6. [CrossRef]

105. Hassannejad, H.; Matrella, G.; Ciampolini, P.; de Munari, I.; Mordonini, M.; Cagnoni, S. Food Image Recognition Using Very Deep Convolutional Networks. In Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management (MADiMa'16), Amsterdam, The Netherlands, 16 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 41–49. [CrossRef]
106. Liu, C.; Cao, Y.; Luo, Y.; Chen, G.; Vokkarane, V.; Ma, Y. Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment. In Proceedings of the International Conference on Smart Homes and Health Telematics, Wuhan, China, 25–27 May 2016; pp. 37–48. [CrossRef]
107. Liu, C.; Cao, Y.; Luo, Y.; Chen, G.; Vokkarane, V.; Yunsheng, M.; Chen, S.; Hou, P. A New Deep Learning-Based Food Recognition System for Dietary Assessment on An Edge Computing Service Infrastructure. *IEEE Trans. Serv. Comput.* **2018**, *11*, 249–261. [CrossRef]
108. McAllister, P.; Zheng, H.; Bond, R.; Moorhead, A. Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets. *Comput. Biol. Med.* **2018**, *95*, 217–233. [CrossRef]
109. Martinel, N.; Foresti, G.L.; Micheloni, C. Wide-Slice Residual Networks for Food Recognition. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 567–576. [CrossRef]
110. Aguilar, E.; Remeseiro, B.; Bolaños, M.; Radeva, P. Grab, Pay, and Eat: Semantic Food Detection for Smart Restaurants. *IEEE Trans. Multimed.* **2018**, *20*, 3266–3275. [CrossRef]
111. Horiguchi, S.; Amano, S.; Ogawa, M.; Aizawa, K. Personalized Classifier for Food Image Recognition. *IEEE Trans. Multimed.* **2018**, *20*, 2836–2848. [CrossRef]
112. Ciocca, G.; Napoletano, P.; Schettini, R. CNN-based features for retrieval and classification of food images. *Comput. Vis. Image Underst.* **2018**, *176–177*, 70–77. [CrossRef]
113. Mandal, B.; Puhan, N.B.; Verma, A. Deep Convolutional Generative Adversarial Network-Based Food Recognition Using Partially Labeled Data. *IEEE Sens. Lett.* **2019**, *3*, 7000104. [CrossRef]
114. Ciocca, G.; Micali, G.; Napoletano, P. State Recognition of Food Images Using Deep Features. *IEEE Access* **2020**, *8*, 32003–32017. [CrossRef]
115. Jiang, L.; Qiu, B.; Liu, X.; Huang, C.; Lin, K. DeepFood: Food Image Analysis and Dietary Assessment via Deep Model. *IEEE Access* **2020**, *8*, 47477–47489. [CrossRef]
116. Liu, C.; Liang, Y.; Xue, Y.; Qian, X.; Fu, J. Food and Ingredient Joint Learning for Fine-Grained Recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 2480–2493. [CrossRef]
117. Liang, H.; Wen, G.; Hu, Y.; Luo, M.; Yang, P.; Xu, Y. MVANet: Multi-Tasks Guided Multi-View Attention Network for Chinese Food Recognition. *IEEE Trans. Multimed.* **2020**, *23*, 3551–3561. [CrossRef]
118. Zhao, H.; Yap, K.-H.; Kot, A.C.; Duan, L. JDNet: A Joint-Learning Distilled Network for Mobile Visual Food Recognition. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 665–675. [CrossRef]
119. Won, C.S. Multi-Scale CNN for Fine-Grained Image Recognition. *IEEE Access* **2020**, *8*, 116663–116674.. [CrossRef]
120. Shen, Z.; Shehzad, A.; Chen, S.; Sun, H.; Liu, J. Machine Learning Based Approach on Food Recognition and Nutrition Estimation. *Procedia Comput. Sci.* **2020**, *174*, 448–453. [CrossRef]
121. Zhu, F.; Bosch, M.; Khanna, N.; Boushey, C.J.; Delp, E.J. Multiple Hypotheses Image Segmentation and Classification With Application to Dietary Assessment. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 377–388. [CrossRef]
122. Aguilar-Torres, E.; Radeva, P. Food Recognition by Integrating Local and Flat Classifiers. In Proceedings of the 9th Iberian Conference, IbPRIA 2019, Madrid, Spain, 1–4 July 2019. [CrossRef]
123. Merchant, K.; Pande, Y. ConvFood: A CNN-Based Food Recognition Mobile Application for Obese and Diabetic Patients. In *Emerging Research in Computing, Information, Communication and Applications*; Springer: Singapore, 2019. [CrossRef]
124. Mezgec, S.; Eftimov, T.; Bucher, T.; Koroušić Seljak, B. Mixed deep learning and natural language processing method for fake-food image recognition and standardization to help automated dietary assessment. *Public Health Nutr.* **2019**, *22*, 1193–1202. [CrossRef]
125. He, J.; Shao, Z.; Wright, J.; Kerr, D.; Boushey, C.; Zhu, F. Multi-task Image-Based Dietary Assessment for Food Recognition and Portion Size Estimation. In Proceedings of the 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Shenzhen, China, 9–11 April 2020; pp. 49–54. [CrossRef]
126. Aguilar, E.; Nagarajan, B.; Khantun, R.; Bolaños, M.; Radeva, P. Uncertainty-Aware Data Augmentation for Food Recognition. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 4017–4024. [CrossRef]
127. Ortega Anderez, D.; Lotfi, A.; Pourabdollah, A. A deep learning based wearable system for food and drink intake recognition. *J. Ambient. Intell. Hum. Comput.* **2020**, *12*, 9435–9447. [CrossRef]
128. Song, G.; Tao, Z.; Huang, X.; Cao, G.; Liu, W.; Yang, L. Hybrid Attention-Based Prototypical Network for Unfamiliar Restaurant Food Image Few-Shot Recognition. *IEEE Access* **2020**, *8*, 14893–14900. [CrossRef]
129. Xiao, L.; Lan, T.; Xu, D.; Gao, W.; Li, C. A Simplified CNNs Visual Perception Learning Network Algorithm for Foods Recognition. *Comput. Electr. Eng.* **2021**, *2*, 107152. [CrossRef]
130. Deng, L.; Chen, J.; Ngo, C.W.; Sun, Q.; Tang, S.; Zhang, Y.; Chua, T.S. Mixed Dish Recognition with Contextual Relation and Domain Alignment. *IEEE Trans. Multimed.* **2021**. [CrossRef]

131. Marc, B.; Ferrà, A.; Radeva, P. Food ingredients recognition through multi-label learning. In Proceedings of the International Conference on Image Analysis and Processing, Catania, Italy, 11–15 September 2017; Springer: Cham, Switzerland, 2017. [CrossRef]
132. Wang, Y.; Chen, J.-J.; Ngo, C.-W.; Chua, Y.-S.; Zuo, W.; Ming, Z. Mixed Dish Recognition through Multi-Label Learning. In Proceedings of the 11th Workshop on Multimedia for Cooking and Eating Activities (CEA'19), Ottawa, ON, Canada, 10 June 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–8. [CrossRef]
133. Salvador, A.; Drozdzal, M.; Giro-i-Nieto, X.; Romero, A. Inverse cooking: Recipe generation from food images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019. [CrossRef]
134. Chen, J.; Pan, L.; Wei, Z.; Wang, X.; Ngo, C.-W.; Chua, T.-S. Zero-Shot Ingredient Recognition by Multi-Relational Graph Convolutional Network. In Proceedings of the AAAI Conference on Artificial Intelligence 2020, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 10542–10550. [CrossRef]
135. Chen, J.; Zhu, B.; Ngo, C.-W.; Chua, T.-S.; Jiang, Y.-G. A Study of Multi-Task and Region-Wise Deep Learning for Food Ingredient Recognition. *IEEE Trans. Image Process.* **2021**, *30*, 1514–1526. [CrossRef]
136. Pettitt, C.; Liu, J.; Kwasnicki, R.; Yang, G.; Preston, T.; Frost, G. A pilot study to determine whether using a lightweight, wearable micro-camera improves dietary assessment accuracy and offers information on macronutrients and eating rate. *Br. J. Nutr.* **2016**, *115*, 160–167. [CrossRef]
137. Comber, R.; Weeden, J.; Hoare, J.; Lindsay, S.; Teal, G.; Macdonald, A.; Methven, L.; Moynihan, P.; Olivier, P. Supporting visual assessment of food and nutrient intake in a clinical care setting. In Proceedings of the Conference on Human Factors in Computing Systems, Austin, TX, USA, 5–10 May 2012; pp. 919–922. [CrossRef] [PubMed]
138. Yang, Z.; Yu, H.; Cao, S.; Xu, Q.; Yuan, D.; Zhang, H.; Jia, W.; Mao, Z.-H.; Sun, M. Human-Mimetic Estimation of Food Volume from a Single-View RGB Image Using an AI System. *Electronics* **2021**, *10*, 1556. [CrossRef]
139. Graikos, A.; Charisis, V.; Iakovakis, D.; Hadjidimitriou, S.; Hadjileontiadis, L. Single Image-Based Food Volume Estimation Using Monocular Depth-Prediction Networks. In *Universal Access in Human-Computer Interaction. Applications and Practice. HCII 2020*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2020; Volume 12189. [CrossRef] [PubMed]
140. Lo, F.P.; Sun, Y.; Qiu, J.; Lo, B.P.L. Point2Volume: A Vision-Based Dietary Assessment Approach Using View Synthesis. *IEEE Trans. Ind. Inform.* **2020**, *16*, 577–586. [CrossRef]
141. Zhu, F.; Bosch, M.; Boushey, C.; Delp, E. An image analysis system for dietary assessment and evaluation. *Proc./ICIP Int. Conf. Image Process.* **2010**, *185*, 1853–1856. [CrossRef]
142. Trevño, R.; Ravelo, A.; Birkenfeld, E.; Murad, M.; Diaz, J. Food Weight Estimation: A Comparative Analysis of Digital Food Imaging Analysis and 24-Hour Dietary Recall. *J. Nutr. Educ. Behav.* **2015**, *47*, S105. [CrossRef]
143. Jia, W.; Chen, H.C.; Yue, Y.; Li, Z.; Fernstrom, J.; Bai, Y.; Li, C.; Sun, M. Accuracy of food portion size estimation from digital pictures acquired by a chest-worn camera. *Public Health Nutr.* **2014**, *17*, 1671–1681. [CrossRef]
144. Rhyner, D.; Loher, H.; Dehais, J.; Anthimopoulos, M.; Shevchik, S.; Botwey, R.H.; Duke, D.; Stettler, C.; Diem, P.; Mougiakakou, S. Carbohydrate Estimation by a Mobile Phone-Based System Versus Self-Estimations of Individuals With Type 1 Diabetes Mellitus: A Comparative Study. *J. Med. Internet Res.* **2016**, *18*, e101. [CrossRef] [PubMed]
145. Makhsous, S.; Mohammad, H.M.; Schenk, J.M.; Mamishev, A.V.; Kristal, A.R. A Novel Mobile Structured Light System in Food 3D Reconstruction and Volume Estimation. *Sensors* **2019**, *19*, 564. [CrossRef]
146. Yuan, D.; Hu, X.; Zhang, H.; Jia, W.; Mao, Z.; Sun, M. An automatic electronic instrument for accurate measurements of food volume and density. *Public Health Nutr.* **2021**, *24*, 1248–1255. [CrossRef]
147. Rollo, M.E.; Bucher, T.; Smith, S.P.; Collins, C.E. ServAR: An augmented reality tool to guide the serving of food. *Int. J. Behav. Nutr. Phys. Act.* **2017**, *14*, 65. [CrossRef] [PubMed]
148. Dehais, J.; Anthimopoulos, M.; Shevchik, S.; Mougiakakou, S. Two-View 3D Reconstruction for Food Volume Estimation. *IEEE Trans. Multimed.* **2017**, *19*, 1090–1099. [CrossRef]
149. Aizawa, K.; Maruyama, Y.; Li, H.; Morikawa, C. Food Balance Estimation by Using Personal Dietary Tendencies in a Multimedia Food Log. *IEEE Trans. Multimed.* **2013**, *15*, 2176–2185. [CrossRef]
150. Ravì, D.; Lo, B.; Yang, G. Real-time food intake classification and energy expenditure estimation on a mobile device. In Proceedings of the 2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Cambridge, MA, USA, 9–12 June 2015; pp. 1–6. [CrossRef]
151. Jiang, H.; Starkman, J.; Liu, M.; Huang, M. Food Nutrition Visualization on Google Glass: Design Tradeoff and Field Evaluation. *IEEE Consum. Electron. Mag.* **2018**, *7*, 21–31. [CrossRef]
152. Probst, Y.; Nguyen, D.T.; Tran, M.K.; Li, W. Dietary Assessment on a Mobile Phone Using Image Processing and Pattern Recognition Techniques: Algorithm Design and System Prototyping. *Nutrients* **2015**, *7*, 6128–6138. [CrossRef]
153. Mezgec, S.; Koroušić Seljak, B. Nutrinet: A deep learning food and drink image recognition system for dietary assessment. *Nutrients* **2017**, *9*, 657. [CrossRef]
154. Pouladzadeh, P.; Shirmohammadi, S. Mobile Multi-Food Recognition Using Deep Learning. *ACM Trans. Multimed. Comput. Commun. Appl.* **2017**, *13*, 36. [CrossRef] [PubMed]
155. Waltner, G.; Schwarz, M.; Ladstätter, S.; Weber, A.; Luley, P.; Bischof, H.; Lindschinger, M.; Schmid, I.; Paletta, L. MANGO—Mobile Augmented Reality with Functional Eating Guidance and Food Awareness. In Proceedings of the New Trends in Image Analysis and Processing—ICIAP 2015 Workshops, Genoa, Italy, 7–8 September 2015; pp. 425–432. [CrossRef]

156. Sun, J.; Radecka, K.; Zilic, Z. FoodTracker: A Real-time Food Detection Mobile Application byDeep Convolutional Neural Networks. *arXiv* **2019**, arXiv:1909.05994.
157. Jia, W.; Li, Y.; Qu, R.; Baranowski, T.; Burke, L.E.; Zhang, H.; Bai, Y.; Mancino, J.M.; Xu, G.; Mao, Z.H.; et al. Automatic food detection in egocentric images using artificial intelligence technology. *Public Health Nutr.* **2018**, *22*, 1168–1179. [CrossRef]
158. Chen, G.; Jia, W.; Zhao, Y.; Mao, Z.H.; Lo B.; Anderson, A.K.; Frost, G.; Jobarteh, M.L.; McCrory, M.A.; Sazonov, E.; Steiner-Asiedu, M. Food/Non-Food Classification of Real-Life Egocentric Images in Low- and Middle-Income Countries Based on Image Tagging Features. *Front. Artif. Intell.* **2021**, *4*, 644712. [CrossRef]
159. Mejova, Y.; Abbar, S.; Haddadi, H. Fetishizing Food in Digital Age: Foodporn Around the World. *arXiv* **2016**, arXiv:1603.00229.
160. Ismail, A.; Idris, M.Y.I.; Ayub, M.N.; Por, L.Y. Investigation of Fusion Features for Apple Classification in Smart Manufacturing. *Symmetry* **2019**, *11*, 1194. [CrossRef]
161. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. In Proceedings of the 37th International Conference on Machine, Learning, PMLR, Vienna, Austria, 13–18 July 2020; Volume 110, pp. 1597–1607.
162. Caron, M.; Misra, I.; Mairal, J.; Goyal, P.; Bojanowski, P.; Joulin, A. Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. *arXiv* **2020**, arXiv:2006.09882.
163. Strumbelj, E.; Kononenko, I. Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.* **2014**, *41*, 647–665.
164. Arik, S.; Liu, Y.-H. Explaining Deep Neural Networks using Unsupervised Clustering. *arXiv* **2020**, arXiv:2007.07477.
165. Subhi, M.A.; Ali, S.H.; Mohammed, M.A. Vision-based approaches for automatic food recognition and dietary assessment: A survey. *IEEE Access* **2019**, *7*, 35370–35381.
166. Min, W.; Jiang, S.; Liu, L.; Rui, Y.; Jain, R. A Survey on Food Computing. *ACM Comput. Surv.* **2019**, *52*, 92. [CrossRef]