Deep Learning for Food Image Recognition and Nutrition Analysis Towards Chronic Diseases Monitoring: A Systematic Review

Article *in* SN Computer Science · July 2023

DOI: 10.1007/s42979-023-01972-1

4 authors:

Merieme Mansouri
University of Hassan II Casablanca

Samia Benabdellah Chaouni
Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes

Said Jai Andaloussi
University of Hassan II Casablanca

Ouaïl Ouchetto
University of Hassan II Casablanca

SN Computer Science (2023) 4:513
https://doi.org/10.1007/s42979-023-01972-1

REVIEW ARTICLE

Deep Learning for Food Image Recognition and Nutrition Analysis Towards Chronic Diseases Monitoring: A Systematic Review

Merieme Mansouri[1] · Samia Benabdellah Chaouni[1] · Said Jai Andaloussi[1] · Ouail Ouchetto[1]

Received: 24 June 2022 / Accepted: 27 May 2023

Abstract
The management of daily food intake aids to preserve a healthy body, minimize the risk of many diseases, and monitor chronic diseases, such as diabetes and heart problems. To ensure a healthy food intake, artificial intelligence has been widely used for food image recognition and nutrition analysis. Several approaches have been generated using a powerful type of machine learning: deep learning. In this paper, a systematic review is presented for the application of deep learning in food image recognition and nutrition analysis. A methodology of systematic research has been adopted resulting in three main fields: food image classification, food image segmentation and volume estimation of food items providing nutritional information."57" original articles were selected and synthesized based on the use case of the approach, the employed model, the used data set, the experiment process and finally the main results. In addition, articles of public and private food data sets are presented. It is noted that among the literature review, several deep learning-based studies have shown great results and outperform the conventional methods. However, certain challenges and limitations are presented. Hence, some research directions are proposed to apply in the future to improve the food recognition systems for dietary assessment.

Keywords Deep learning · Food image recognition · Size estimation · Food monitoring · Dietary assessment

Introduction

In recent years, people have taken care more of their health by paying attention to the most important factor: food intake. Healthy food intake could prevent our bodies from many diseases or control them if we already have one, such as diabetes and heart problems that could affect us due to obesity, overweight and unhealthy food intake habits [1]. On the other hand, people are always on the lookout for ways to ease and improve the daily lifestyle as possible. With the help of the continuous technology development and the appearance of artificial intelligence, people have achieved remarkable development in facilitating and improving the lifestyle. Automatic and efficient technology to observe and control daily food intake will help in preserving the human body and decrease the development of diseases. The key behind artificial intelligence-based technologies is Deep Learning. The ability of applying this technique excellently in most domains makes it almost a universal learning approach. The end to end specification of deep learning without the need of extracting features separately ensures the robustness. Another powerful reason is the generalization of deep learning; many applications can use the same deep learning approach. Finally, the deep learning ability of grow, using more layers and nodes at each time, makes it highly scalable [2]. The convolutional neural network is a powerful kind of deep learning architecture that has shown impressive results in different fields, especially in image food recognition [3]. Many approaches showed that CNN outperforms traditional machine learning in general and other types of deep learning networks specifically [4]. In the recent years,

Samia Benabdellah Chaouni, Said Jai Andaloussi and Ouail Ouchetto are contributed equally to this work.

Merieme Mansouri
merieme.mansouri-etu@etu.univh2c.ma

Samia Benabdellah Chaouni
SAMIA.BENABDELLAH-CHAOUNI@univh2c.ma

Said Jai Andaloussi
said.jaiandaloussi@etu.univh2c.ma

Ouail Ouchetto
ouail.ouchetto@etu.univh2c.ma

1 Mathematics and Computer Science, Faculty of Sciences Ain Chock, University Hassan II, Km 8 Route El Jadida, B.P 5366 Maarif, Casablanca 20100, Morocco

Published online: 05 July 2023

several CNN architectures have been implemented, each having specific layers and parameters. Approaches of food image recognition address different tasks: food classification which contains food items detection, features extraction and food items categorization [5], food segmentation [6], size estimation [5] and finally nutritional information providing, for example: calories, cholesterol, carbohydrate, fats and so one. Certain approaches provide nutritional information based on the average portion that a person could take or the average information could be in a meal [7], while others estimate the exact quantity of food item from an image [8]. Food recognition approaches have been adopted of generated several food data sets to evaluate their contribution and trait either a specific type of cuisine, such as Asian food [9] or miscellaneous types [10]. While others choose to recognize fruits and vegetables [11]. This paper is a review of CNN architectures and its application in food computing domain in particular food image recognition and nutritional analysis. "Deep Learning Overview" gives an overview of CNN evolution. "Methodology" presents CNN-based approaches for food recognition. "Results" presents the most popular, public and private, food data sets. "Discussion" highlights certain CNN challenges and future research directions. "Conclusion" presents a discussion, and finally the paper ends with a conclusion.

Deep Learning Overview

In this section, an overview of deep learning methods for food computing is presented, with a focus on food image recognition for nutrition analysis. Deep learning is a subset of machine learning inspired by the construction of the human brain [12]. The deep neural networks are composed of a number of layers and nodes, each receiving and interpreting information from each other. The difference between conventional machine learning and deep learning is that classical machine learning takes the input data and parses it to determine features then trains the machine learning with those data to make a decision. While deep learning has come to ensure end-to-end learning, it means from the input image to the output decision. Deep neural networks are extracted from artificial neural networks. ANN composed of input layer, hidden layer and output layer, DNN contains more hidden layers for feature extraction and representation. Three types of learning are common in deep learning: supervised learning, unsupervised learning and reinforcement learning.

Supervised deep learning learned from a labeled set, its most popular tasks are classification and regression. The method commonly used in supervised-based approaches for food image recognition is convolutional neural networks and its different models. The convolutional neural network is a powerful kind of deep learning architecture that has shown impressive results in different fields, especially in image food recognition [2]. Many approaches showed that CNN outperforms traditional machine learning in general and also other types of deep learning network specifically [4]. In the past years, several architectures of CNN have been implemented, each having specific layers and parameters. Most CNN architectures are marked by a pooling layer to generate feature maps and backpropagation to improve the training phase. Certain CNN models that have been applied in food image recognition are: AlexNet [13], VGG-16 [14], GoogleNet [15], Inception-V4 [16], ResNet-152 [17], Dense-Net [18], MobileNet [19] and EfficientNet-B7 [20]. All of these models were pretrained on a large-scale data set: ImageNet [21] and used in different domains by applying the transfer learning concept and transferring the pretrained models to be tested on sub-data set and to benefit from the massive resources on which the model is trained on, such as data, time and cost [22].

For multi-food recognition, R-CNN [23], fast R-CNN [24], faster R-CNN [25], Single Shot Detection (SSD) [26] and you only look once (YOLO) [27] are unique deep convolutional neural networks employed for food detection. Food image segmentation as well is an important task in food recognition [6]. The most applied methods are mask R-CNN [28], the CNN-based model for biomedical image segmentation UNet [29], and DeepLab [30] the open-sourced and designed model from Google for semantic segmentation. Another supervised deep learning method used in food recognition is Recurrent Neural Networks(RNN) [31]. The specification of handling sequences of independent information makes the RNN architecture look like a circle, where the output of a neural is the input of this neural or the previous one. Different improved versions of RNNs are used for food recognition as LSTM [32] and the attention mechanism [33] to overcome different limits of RNN such as the gradient vanishing and the difficulty of managing long information.

Unsupervised models learn from unlabeled data to perform a specific task, such as clustering, dimension reduction and size estimation. Models commonly used for unsupervised deep learning in the food recognition field are: Generative Adversarial Nets (GAN) which is composed of a generator that generates fake examples and a discriminator for fake examples distinction from real data [34], and Autoencoder (AE) model, where the encoder transforms the input to a representation and fed it to the decoder to reconstruct the output [35]. The last type of deep learning is reinforcement learning; its main concept is bringing decisions by creating new actions based on its interaction with the environment [36]. Several deep learning frameworks have been developed to facilitate the neural networks implementation, using different deep learning libraries and supporting several programming languages and processing units. Some of the popular deep learning frameworks are Tensorflow [37], Keras [38], PyTorch [39], Caffe [40], Theano [41], and many others.

Methodology

To present the deep learning approaches for food image recognition and nutrition analysis towards dietary assessment. A systematic review is adopted following three main stages: search strategy, search proceeding, and search reporting. In search strategy, the search questions, publication databases, and related keywords are defined. The second step is the search proceeding, by going through publication databases, certain selection guidelines are applied using inclusion/exclusion criteria for studies screening. Furthermore, essential data are extracted and synthesized by answering the search questions. Finally, the key information from search questions responses is categorized based on several comparison criteria and reported with the help of summary tables and figures.

Search Strategy

Aiming at conducting a systematic review on deep learning approaches for food image recognition, seven research questions are proposed. Q1: What are the different DL models that were used for food image recognition?

Q2: What are the different tasks and use cases applied to food images using DL?

Q3: What are the approaches used for food image classification?

Q4: What are the approaches used for food image segmentation?

Q5: What are the approaches used for the volume estimation of food item and the nutrition analysis?

Q6: What are the public and private data sets for food images?

Q7: What is the type of food images that are processed?

Q8: What are the main challenges in food image recognition using DL?

Q9: What are the future directions of research proposed in the field of food image recognition using DL?

Several databases were used to answer the above questions including Web of Science, Scopus, IEEEXplore, Science Direct, and Google Scholar. The basic search keywords that were used for the databases are“deep learning”,“food image recognition” and “nutrition analysis”, combined with the help of Boolean operators AND/OR. After reading some initial articles, other keywords were employed. The search string was: ((food OR plate OR dish) AND (recognition OR classification OR detection OR segmentation OR volume estimation OR data set) AND (deep learning OR transfer learning OR deep neural networks OR CNN OR RNN OR RCNN OR GAN OR AE) AND (nutrition analysis OR dietary assessment OR diet monitoring OR calorie estimation). For databases that have a limited number of Boolean operators or characters, the size of the query search was minimized by eliminating some keywords or combining the results of two or three queries.

Search Proceeding

Following the previous search strategy, a number of related studies were selected. The second step is to apply inclusion/exclusion criteria for approaches. The articles included and selected to be cited in this review are written in English, original, and published in the last 4 years: from 2019 to 2022. Moreover, studies should answer the proposed research questions and contain the proper key information for extraction and reporting. The studies that are duplicated or not related to research questions or defined as article reviews and posters have been excluded. Three main categories were identified based on the final list of the articles: food image classification, food image segmentation and volume estimation of food items providing nutritional information. In addition, an important section of public and private food image data sets is presented.

Search Reporting

Several comparison criteria were proposed to compared the collected studies and summarized the important key information. For the three main categories mentioned in “Search Proceeding” section, the comparison criteria are:

Use case: the use case and scenario of the study application.

Method: the used deep learning method in the study.

Experiment process: the steps of the experiment developed in the study indicating the pre-processing techniques.

Data set: the used data set in the experiment for deep learning model development.

Performance: the main outcomes of the experiment including the performance metrics.

Nutrition analysis: This category indicates whether or not the current study provides nutrition information.

Public and private food image data sets were also reported based on:

Data source acquisition: the source of the data for the construction of the data set, such as collecting from the Web, retrieving from existing data sets or building data in a laboratory.

Data type: in this case, type of data refers to the type of the food item, since we categorize the public and private food data sets.

Number of classes: total number of classes of the data set.

Number of images: total number of images in the data set.

Results

The number of articles after an initial search is 507. After applying all the exclusion/inclusion criteria, the final collection remained 57 papers. As mentioned before, the total studies are divided into three main categories: food image classification (14), food image segmentation (12), and food volume estimation (14). Moreover, a category of public data sets is reported with a total of 17 studies, while private data sets are extracted from the existing approaches of the three main categories. Tables 1, 2, 3, 4, 5, 6 and 7 present the selected approaches based on the comparison criteria.

Food Image Classification

Food image classification is an important task in food recognition. Based on a specific scenario, supervised and unsupervised deep learning models are applied to food images for feature extraction and assigning a label of a category to the food item. Jiang et al. [7] presented a system to recognize food and provide nutrition analysis to assess diet. First, they detect the food item using faster RCNN then extract features and classify them with the help of VGG-16. To train and evaluate their system, the authors used two data sets UEC-FOOD100 and UEC-FOOD256 and a self-made data set called FOOD20-with-bbx; it is a sub-data set from the original one: FOOD101, with 20 categories and a bounding box. The second step is nutrition analysis; the system can provide five nutritional units: calories, fats, carbohydrates and proteins based on the USDA national nutrient database and according to specific information of each user, such as age, weight and so on. The system assumes that the weight of each food item is 400 g to provide nutritional information. Hao et al. [42] introduced the challenge of the fine-grained features of Chinese food and proposed a system for food image recognition using an attention mechanism to extract fine-grained features and the pre-trained CNN model: DensNet-169 for classification. Latif et al. [11] proposed a new self-made data set composed of 20 types of fruits and 20 types of vegetables. Images were taken under different conditions and contain different backgrounds, plates, angles, resolutions and lights. Authors adopted three types of CNN with 9, 11 and 15 layers to recognize plates of fruits, and also tested on fruits-360 data set. Another supervised deep neural network model is used by Ye et al. [43]. They developed a system that recognizes ten categories of food from COCO2017 with several items on a plate. Mask R-CNN is applied for the segmentation and classification of each item on a plate and then provides nutritional information. Authors used ResNet+FPN as their backup network for segmentation and classification, and OpenCV to display the category and nutritional information of each item on the plate. The authors used Tensorflow Lite to transfer the model to a mobile device to facilitate access to users. Ambadkar et al. [10] implemented a system to recognize food images and estimate attributes, ingredients and nutritional information for web and mobile use. For food recognition, authors used four pretrained models: inception-v3, inception-v4, Xception and inceptionRestNetV2. The second part is attribute estimation. The system used the Web scraping method to collect data from the web, and then applied the Word2Vec technique on the label of the recognized food item to estimate ingredients, attributes and nutritional information. Concatenation of deep learning is a common technique in food recognition, for example, Şengür et al. [44] used concatenation of deep feature extraction from two pre-trained CNN models: AlexNet and VGG16, and fed it to SVM algorithm for food images classification. The used data sets in this system are: FOOD-11 and FOOD-101 for food category classification and FOOD-5 K for food or non-food classification. Likewise, Zhang et al. [45] developed a feature fusion model and a subnetwork-based neural network for food image classification. Authors proposed three CNN-based architectures AlexNet, Inception, DensNet to extract features. The concatenation of the features fusion is feeding it to a subnet neural network for encoding and classifying the features iteratively.

The SNN model ensures the dimension reduction to reduce the redundant features. Qiu et al. [47] exploited specific regions of food-image and classified them using the PAR-Net network. Three main sub-networks are proposed: primary, auxiliary and region networks to classify the full image and discriminative regions. The sub-networks are concatenated to generate the final classification result. The authors generated a new data set of sushi images: Sushi-50 and utilized two more data sets for an experiment: Food-101 and Vireo-172.

To overcome the classification of mixed food items on a plate, Deng et al. [48] proposed contextual relation network (CR-Nets), this approach contain ResNet-50 for single item classification, then a weights are copied and used for localizing and classifying multiple food items with the help of Faster RCNN. Finally, the results are refined using a co-occurrence statistic method of dish labels. In addition, Mao et al. [49] introduced two methods for the localization and recognition of multi-food images. The authors localized food items using faster RCNN then extracted and classified features with the help of CNN and a visual relation method to cluster food categories and improve classification performance. Vegetables are a main item in our daily food intake. To insist on its important, Tan et al. [50] proposed Quantized deep residual convolutional neural networks (DRNN), the authors extracted and classified features using

Table 1 Food classification approaches

Ref	Use case	Method	Experiment process	Data set	Performance	Nutrition analysis
[7]	Food recognition and dietary assessment	Faster R-CNN VGG-16	Resize images, split data into training and testing, data augmentation, Faster R-CNN for detection, VGG-16 for feature extraction and classification	UEC-FOOD100 UEC-FOOD256 FOOD20-with-bbx	17.5%(mAp) 10.5%(mAp) 71.7%(top-1)	calories, fats, carbohydrates and proteins
[42]	Fine-grained recognition of Chinese food	Attention Mechanism DenseNet-169	Feed data set with more images, split data into training and testing, data enhancement, extract fine-grained with attentio mechanism, classification with DenseNet-169	VIREO Food-172	87.6%(top-1)	–
[11]	Fruits and vegetables recognition	CNN	split data into training and testing	Fruits-360 + Self-made data set	92% 95%	Calories
[43]	Food recognition and dietary assessment on mobile device	Mask RCNN	Images resizing, split data into training and testing	COCO 2017	90%(top-1)	Calories
[10]	Mobile and Web application for food classification and nutrients estimation	Inception-V3 Inception-V4 ResNet nceptionResNetV2	Data augmentation, split data into training and testing	Food-101	92%	Fats, protein fiber sodium cholesterol
[46]	Nutrition analysis for artificial Pancreas Systems for diabetic patients	CNN	Data augmentation	Food-101	81.65%(top-1)	Cholesterol
[44]	Food classification	AlexNet VGG16 SVM	Resize images, extract features using AlexNet and VGG16, classify features using SVM	FOOD-5K FOOD-11 FOOD-101	99% 89.33% 79.86%	–
[45]	Food image classification of large-scale data sets	SNN	Fuse data training and testing from different data sets	UEC100 UEC256 FOOD101 FOOD251	87.7% 83.1% 90.8% 64%	–

Table 2 Food classification approaches (continued)

Ref	Use case	Method	Experiment process	Data set	Performance	Nutrition analysis
[47]	Food recognition	PAR-Net	Images are cropped, resized and horizontally flipped	Food-101 Vireo-172 Sushi-50	90.4%(top1) 90.2%(top1) 92.0%(top1)	–
[48]	Mixed-dish recognition in Singapore	CR-Nets ResNet-50 faster RCNN	Split data to training, validation and testing, resize images, food localization using Faster RCNN, food classification using ResNet-50, refine results using co-occurrence labels	School lunch data set self-made data set	87.74%(precision)	–
[49]	Multi-food recognition	Faster RCNN DenseNet-121	Split data to training, validation and testing, croped images, food localization using faster RCNN, food classification using CNN	UEC-100 UEC-256 VIPER-FoodNet Food-101 UPMC-101	63% 65% 56% (precision)–% –%	—
[50]	Food classification and vegetable identification to increase awareness of vegetable intake	DRCNN	Split data to training, validation and testing, data resizing, data augmentation, post-training quantization is applied to transform weights into lightweights for mobile utilization	Food 101 proprietary data set	96% 90%	–
[51]	Fruit classification	CAE-ADN	Pre-train images using convolution autoencoder, extract and classify features using attention DensNet	Fruit 26 Fruit 15	95%(top-1) 93%(top-1)	–
[52]	Food recognition	GAN	Split data to training, testing and unlabeled data	Food-101 Indian Food Data set	75%(top-1) 85%(top-1)	–

Table 3 Table caption

Ref	Use Case	Method	Experiment Process	Data set	Performance	Nutrition analysis
[53]	Food image segmentation	Salient region detection multi-scale segmentation Fast rejection	–	UNIMIB2016	97%(accuracy)	Calories
[54]	Mobile application for dietary assessment	Mask RCNN	Images are manually annotated with semantic segmentation map	MADiMa database Fast food database	94.4%(Fsum)	cholesterol protein fat calorie
[56]	Food image segmentation	UNet DeepLab PCA Graph-Cut	Split images into training, validation and testing, extract regions of interest, data augmentation	UNIMIB2016	88.3%(accuracy)	–
[55]	Brazilian food image segmentation	FCN ENet SegNet DeepLabV3+ Mask RCNN	Split images into training, validation and testing, labeled food segmentation mask manually using VGG image annotator, resize images	Self-made data set	70% 51% 52% 50% 70%(IoU)	–
[57]	Food image segmentation	ResSeg	Split images into training and testing, labeled food segmentation mask manually	Self-made data set	<90%(accuracy) <75%(IoU)	–
[58]	Food image segmentation	DeepLapV3+	Split images into training, validation and testing, labeled food segmentation mask, data augmentation	Food-101	0.92%(IoU)	–
[59]	Food image segmentation	ResNet CAM SSDD	Add segmentation mask to 10% of data set	UEC-FOOD100 Collecting from Web and Twitter	82.6%(Pix acc) 55.4%(IoU)	–
[60]	Food image estimation for long-term care for older people	EDFN-D	Resize images, split images into training and validation, data augmentation	UNIMIB2016 Two self-made data sets	0.88(IoU)	–
[61]	Nutritional assessment for hospitalized patients	Contextual network using ResNt50	Split images into training, validation and testing, food segmentation mask annotation, data augmntation	Self-made data set	73.75%(mIoU) 83.50%(pix. acc) 89.88%(fsum)	Calorie cholesterol fat salt protein fiber

Table 4 Food segmentation approaches (continued)

Ref	Use case	Method	Experiment process	Data set	Performance	Nutrition analysis
[62]	Food image segmentation	ResNet-101 attention mechanism WASP	Food segmentation mask annotation, resize images	UNIMIB2016 UEC FoodPix	71.79%(mIoU) 65.13%(mIoU)	–
[63]	Segmentation and counting of overlapped food items	ResNet-50 SibNet	Split images into training, validation and testing food segmentation mask annotation, resize images	Three self-made data sets	87.8%(PQ)	–
[64]	Food image segmentation for healthcare robot systems	Mask RCNN	Categorise data set to easy, medium and hard based on the complexity of the background	Self-made data set of synthetic data self-made data set of real-world data UNIMIB2016	87.9% 79% 82.7%(SEG)	–

both residual-based network and convolutional-based network, and then a quantization technique is performed on the trained weights to reduce the model representation with the aim of deploying it on mobile device. Hybrid deep lear-ning-based classification has also been applied in food recognition using both supervised and unsupervised models. Xue et al. [51] used both the unsupervised autoencoder with convolution layer and the supervised attention mechanism based on DenseNet. Convolutional autoencoder is used to pre-train the images, while features are extracted and classified using attention-DenseNet. Another hybrid approach is proposed by Mandal et al. [52]. The authors modified the unsupervised generative adversarial networks (GANs) based on the architecture of the CNN to perform semi-supervised food recognition. Chakrabarty et al. [46] targeted diabetic patients type 1 and proposed to identify the food category using CNN and provide the amount of cholesterol. The selected studies for food classification are presented in Tables 1 and 2.

Food Image Segmentation

The food image segmentation task is mainly used to segment the food entity and separate it from any background (including plates) to estimate the exact size thereafter. Minija and Emmanuel [53] introduced a system that segments, extracts features, classifies, estimates volume, and provides calorie value for food image recognition and dietary assessment. In this section, the interesting part is food item segmentation. The authors proposed multiple-hypotheses segmentation composed of three processes: salient region detection to extract regions of interest, multi-scale segmentation and rejection to remove the unnecessary segmentation. Several mobile applications have been developed for food recognition. GoFood is a mobile-based system for food recognition and dietary assessment proposed by Lu et al. [54]. GoFood performs four stages: segmentation, classification, size estimation and nutrition providing. Concerning food image segmentation. The authors proposed two types of food image segmentation: automatic and semi-automatic. The applied method is the instance segmentation model Mask RCNN. The user must confirm the obtained segmentation output, if the results are not satisfied, the user should mark the proper item then the system will pass to the semi-automatic segmentation and provide a new output with the help of the region growing and merging algorithm. Another mobile application is proposed by Freitas et al. [55]. The authors first performed an evaluation of five different methods of food segmentation and proposed a self-made data set containing nine famous Brazilian food with 1250 images, each image consisting of a segmentation mask using the VGG image annotator (VIA). After all, the authors built a real-time mobile phone application to segment and classify food images with the aim of nutritional monitoring and food intake control. Paper evaluated and compared four common methods for semantic segmentation: Fully convolutional network (FCN), ENet, SegNet and DeepLabV3+, and one method for instance segmentation: Mask R-CNN. Specific evaluation metrics have been used to compare the performance of each method for food items segmentation: intersection over union (IoU), specificity (SP), sensitivity or recall(SE) positive predictive value(PPV) and balanced accuracy(BAC). A hybrid approach of conventional and deep learning segmentation methods is also applied. Siemon et al. [56] presented a system composed of fully-CNN based on the pre-trained UNet and DeepLab models for features extraction. Then, a conventional technique of hierarchical clustering is used to cluster food classes applying principal component analysis (PCA) and GraphCut to obtain different abstractions of one class called cuts. Finally, a sequential neural network is applied to each class based on the number

Table 5 Size estimation approaches

Ref	Use case	Method	Experiment process	Data set	Performance	Nutrition analysis
[66]	Calorie estimation by food classification and size calculation	Deep convolutional network and sensor	Image augmentation	–	–	Calories
[67]	Volume estimation of food in a bowl container	3D reconstruction using paper ruler	–	Self-made data set	18.6%(relative error)	–
[68]	Size estimation of food items	Multi-task learning Cross Domain Feature Adaptation	Data augmentation, split data to training and testing	Self-made data set	56%(MAE)	Calories
[69]	Size estimation of soft drinks	Bag of features ration calculation	Noise reduction, contrast enhancement, localization with visual saliency, data augmentation, split data to training and testing	Self-made data set	75% (accuracy)	Calories fat protein carbohydrates sugar salt sodium
[70]	Volume estimation based on dietitians-mimetic	MobileNet inner product between probable volume and reference volume	Data augmentation, split data to training and testing	Virtual self-mad data set real self-made data set	9%(average error) 20%(relative error)	–
[71]	Volume estimation without reference object	use mathematic equation based on motion sensor in smartphone with known size and specific picture capture, International Food Unit IFU	–	Self-made data set	26.97%(absolute error)	–
[72]	Food size estimation	GAN	Annotation of energy manually, data augmentation, split data to training and testing	Self-made data set	10.89%(error)	–

Table 6 Size estimation approaches(continued)

Ref	Use case	Method	Experiment process	Data set	Performance	Nutrition analysis
[73]	Size estimation system based on video data sets	CNN, 3D reconstruction by combaining the predicted depth with the segmented image	Resize images, data augmentation	EPIC-KITCHENS video self-made data set	45%(mean absolute error)	–
[74]	Food intake estimation for diabetic patients	3D reconstruction using 3D camera	–	–	33%(absolute error)	Calories carbohydrates protein sugar fats
[75]	Android application for calorie measurement	Reference object	–	Fruits-360 Self-made data set	21%(mean error)	Calories
[9]	Size estimation	Linear regression	Image capture with fixed angle, size annotation	Ville Cafe Data set	8.22%(absolute error)	Calories
[76]	Volume estimation for dietary assessment	3D reconstruction with depth sensors, UNet for occlusion view, VNet for volume estimation	Data augmentation, size annotation	self-made data set of 3D models	84.68%(accuracy)	–
[65]	Size estimation	Echo ranging method for distance measurement	–	Self-made data set	12.37%(relative error)	–
[77]	Application mobile for carbohydrate estimation for type 1 diabetic patients	3D reconstruction using reference object	–	Food dishes from hospital restaurant for testing	26.2%(absolute error)	Carbohydrates

Table 7 Publicly available data sets for food recognition

Ref	Authors/years	Data set	Food type	#Images	#Class	Acquisition source
[78]	Hou et al. 2017	VegFru	Fruits and vegetables	160k	292	Web
[79]	Muresan and Oltean 2017	Fruits-360	Fruits and vegetables	90k	131	Camera
[80]	Waltner et al. 2017	FruitVeg-81	Fruits and vegetables	15k	81	–
[81]	Matsuda and Yanai 2012	UEC Food100	Japanese	14k	100	Web
[82]	Kawano and Yanai 2014	UEC Food256	Japanese	25k	256	Web
[83]	Bossard et al. 2014	ETHZ Food 101	Miscellaneous	101K	101	Web
[84]	Farinella et al. 2014	UNICT FD889	Miscellaneous	3583	889	Camera
[85]	Pouladzadeh et al. 2015	FooDD	Miscellaneous	3000	23	Camera
[86]	Kaur et al. 2019	FoodX-251	Miscellaneous	158k	251	Web
[87]	Ciocca et al. 2017	Food524DB	Miscellaneous	247k	524	Existed data sets
[88]	Liu et al. 2020	ISIA Food-500	Miscellaneous	399k	500	Web
[89]	Chen et al. 2016	Vireo Food-172	Chinese	110k	172	Web
[90]	Chen et al. 2017	Chinese FoodNet	Chinese	180k	208	Web camera
[91]	Wang et al. 2019	Kenyan Food13	Kenyan	8174	13	Web
[92]	Chen et al. 2009	PFID	Fast food	4545	101	Restaurant camera
[93]	Ciocca et al. 2017	UNIMIB2016	Italian	1027	73	Camera
[94]	Termritthikun et al. 2017	THFOOD-50	Thai	15,770	50	Web
[55]	Freitas et al. 2020	MyFood	Brazilian	1250	9	Web
[65]	Gao et al. 2019	SUEC Food	Segmented Asian food	31,995	–	Existed data set

of its cuts to find the final segmentation results. Arenas et al. developed a modified version of the segmentation method SegNet called Residual SegNet (ResSeg) by adding residual layers to the encoder and decoder parts [57]. Abdullahi and Muangchoo addressed the challenge of training deep CNN models on small scale data sets to perform food segmentation and recognition [58]. Authors proposed a modified deepLabV3+ identified with its encoder–decoder layers, and initialized with ResNet-18 weights. Wataru Shimoda and Keiji Yanai attempted to overcome the challenge of food image segmentation without the need of pixelwise annotation for training and also segmenting both plate and food item to use them according to the proper situation [59]. Thereafter, the authors used plate segmentation to help increase the performance of weakly supervised food segmentation. As a first step, the authors worked on two data sets: food categories to segment food region and food/non-food categories to segment food + plate region.

They used class activation mapping (CAM) and ResNet for segmentation and visualization to calculate the difference between the two segmentations and create a food mask and a plate mask. The authors chose the weakly supervised segmentation method: self-supervised difference detection SSDD and refined it with the plate segmentation mask. Gao et al. [65] applied two main tasks: classification of food containers and segmentation of food content. First of all, the system utilized VGG-16 to extract features, then three modified versions of FCN with different architectures to segment the food item. To overcome malnutrition in long-term care for older people, Pfister et al. [60] proposed an automatic system for segmentation and volume estimation of food images. The authors used an encoder–decoder network with deep refinement for semantic segmentation on the Madima2016 data set. Lu et al. [61] addressed the challenge of malnutrition for hospitalized patients and proposed a fusion between feature maps obtained from ResNet-50 and pyramid feature maps to segment food and plate from an image as a step in a nutritional assessment system. Another feature fusion technique is applid by Sharma et al. [62]. The authors combined both spatial and channel features from the attention mechanism and fed them to the Waterfall Atrous Spatial Pooling model (WASP) for food segmentation. Most of the approaches of food segmentation trait semantic segmentation and process multiple food items of the same class as one single item. Both Nguyen et al. [63] and Park et al. [64] highlighted the lack of instance segmentation approaches to segment each food instance for the purpose of addressing food counting and overlapping food items. SibNet counts the number of instances using a regression CNN, separates the items using a fully convolutional network and then analyzes pixels and detects sibling items. Park and the authors generated synthetic data to augment the collection and annotation of real-world data and refine the training phase of Mask R-CNN for instance segmentation. all the above approaches are listed in Tables 3 and 4.

Size Estimation

As described in Tables 5 and 6, several size estimation approaches have been developed. Priyaa et al. [66] proposed calorie estimation system operated on mobile and based on the concept of Internet of Things (IoT). Authors applied deep convolutional network for food image classification and a sensor for volume estimation then they connected the system with an edge device and stored data in the Thingspeak cloud. Jia et al. [67] addressed the limit of food volume estimation in a bowl and used specific paper roller to measure the bowl and reconstruct 3D model and then estimate size based on virtual levels. He et al. [68] highlighted the lack of portion size estimation and propose a data set with class labels and portion size annotation. In the training phase on multi-task learning method, they used concatenation of feature vectors from both classification and regression tasks to estimate food portion. Hafiz et al. [69] presented soft drinks recognition, since they have high impact on human consumption. For that, the authors estimated the size of bottles using bag of words methods and the distance ratio between the cap and the bottle.

Yang et al. [70] tried to mimic the thinking of dietitians and proposed a food volume estimation system based on reference classes with known size and combined it with the predicted class to give the predicted volume using the MobileNet model. Muralidha et al. [75] proposed a system that recognizes fruits and vegetables classes, and then calculates their volume to estimate the calorie intake. For size estimation, the authors used a reference item: CNY coin to calculate the required dimensions to determine the quantity and mass of the food item and then provide the calorie value. Another volume estimation approach based on a reference object is presented by Fang et al. [72] using Generative Adversarial Networks (GAN) for energy prediction. Without the need to reference objects, Yang et al. [71] used special capture of images and a smartphone with a known length and a sensor of motion. These parameters are fed into a mathematical model to predict the size of a food portion. To improve performance, the authors added the International Food Unit (IFU) which is a virtual cube with known size to estimate the food volume. Using video data sets and CNN, Graikos et al. [73] predicted food portion depth and combined it with a segmented mask to obtain a 3D representation of the food item and estimate its volume. Lo et al. [76] addressed the volume estimation of occluded food views. The authors developed two deep neural network models: UNet for point completion for occluded food parts and real-time 3D reconstruction, and VNet for volume estimation. The proposed approach used depth-captured images or several sequences of recorded video. Gao et al. [65] discussed two main challenges in food volume estimation: the need for an annotated training data set with volume information and reference objects. Instead of these two precedent factors, the authors introduced a system for volume estimation of food items using an echo ranging method that calculates the length of the sound wave to measure the distance between food item and camera, then a fully convolutional network (FCN) to segment food item and extract its shape. Finally, data aggregation is applied between those two methods to provide the final predicted volume. Chiang et al. [9] covered the obesity problem and implemented a system to control food intake by recognizing the food item and estimating weight and calories. For food weight and calorie estimation, authors took images at a fixed angle and with the same dish and background. The number of pixels for each food item is computed with the help of the mask RCNN model. The authors applied a linear regression equation to obtain the estimated weight and then the calories and nutrition information. Two approaches tackled volume estimation and nutrition analysis for diabetics patients. Makhsous et al. [74] presented a wearable system to generate 3D measurements and provide general nutrients of the consumed food portion in a hospital environment. Rhyner et al. [77], they estimate carbohydrates based on a mobile application to assist patients with diabetes type 1, using a reference object and two captured images to reconstruct a 3D model and predict the proper size.

Food Recognition Data Sets

The key to an efficient food image recognition system is a well-prepared data set in different aspects: quantity, variety, quality and annotation.

This section presents the most popular public data sets and also some private self-made data sets as described in Tables 7 and 8, respectively. Each data set is characterized by the type of food either from a specific cuisine, such as: Japanese, Chinese, Italian, or specialized in fruits and vegetables, or miscellaneous food types. Another important specification is the total number of classes and images. The source of images is also listed, which can be from the Web, such as specific websites and search engines, or images have been taken by a camera in a laboratory or restaurant, with a mobile phone or special equipment. While other data sets have been constructed from existing food data sets. VegFru [78], Fruits-360 [79] and FruitVeg-81 [80] are three fruit and vegetables data sets that are publicly available, the VegFru data set has the largest number of images and it is specialized with upper-level classes and sub-classes with a total of 292 categories and 160,000 images collected from the Web. UEC Food100 [81] and UEC Food256 [82] are two popular data sets for Japanese food. UEC Food256 is an expanded version of the original UEC Food100 containing a total of 256 categories and 25,000 images with the bounding box. The urgent need for benchmark food recognition system made various authors have the initiation to develop benchmark data sets containing miscellaneous food categories. Food524DB [87] and ISIA Food-500 [88] are the most large-scale and benchmark data sets. The Food524DB data set collected 524 classes and 247,000 images from four existing data sets, while ISIA Food-500 composed of 500 categories and almost 399,000 images acquired from the Web. The rest of the publicly available data sets listed in Table 7 are released based on specific cuisine or a modified version of existed data set such as SUEC Food data set [65] which contains 31,995 segmented images from the UEC Food256 data set. Food2k [95] is another large-scale data set with 2000 classes and 1 million images. Unfortunately, this data set is not yet available for public use. Other private data sets are constructed based on the use case of the paper. For example, Tran et al. [96] have collected pastry images for automated checkout in a bakery shop. Lo et al. [76] presented a 3-D food data set with 4000 models to evaluate the volume estimation system. AIFood [98] contained 372,095 images with ingredient labelling for ingredient recognition.

Table 8 Private data sets for food recognition

Ref	Authors/Years	Data set	Food type	#Images	#Class	Acquisition source
[95]	Min et al. 2021	Food2K	Miscellaneous	1 M	2000	Web
[11]	Latif et al. 2020	Self-made data set	Fruits	41,509	40	Camera
[96]	Tran et al. 2020	Self-made data set	Pastry	1289	16	Camera
[9]	Chiang et al. 2019	Ville Cafe	Asian food	35,842	16	Restaurant
[76]	Lo et al. 2020	Self-made data set	3D models of miscellaneous categories	4k	10	AutoCad
[97]	Cai et al. 2019	BTBUFood-60	Miscellaneous	60k	60	Web
[8]	Ege et al. 2019	Self-made data set	Segmented Asian food	14k	100	Existed data sets
[98]	Lee et al.2019	AIFood	Miscellaneous(with ingredient labeling)	372,095	24	Web/existed data sets
[99]	Güngör et al. 2017	TurkishFoods-15	Turkish food	7500	15	Web

Discussion

Limitations and Challenges

Regarding the high performance of deep learning applications for food computing, more specifically: food image recognition and nutrition analysis, there are several limitations and challenges identified from the selected approaches. A generic and large-scale data set is needed to assess benchmark recognition. Recently, several large-scale food data sets have been released but none of them can contain all types of cuisines from around the world. The benchmark data set will help in the evolution of food recognition approaches and facilitate the recognition task for whatever food type. Furthermore, this review addressed dietary assessment approaches that classify, segment, and estimate the volume of food intake to provide nutrition information. Besides ensuring a healthy lifestyle, the main purpose of dietary assessment systems is to help people with chronic diseases to monitor their daily food intake, since healthy food intake may prevent the disease development and augment the patient's lifetime accordingly [100, 101]. However, in most studies, the used data set and the type of nutrients provided do not take into consideration the specification of each chronic disease. For example, there are several studies that recommend Mediterranean diets for patients with chronic kidney disease to delay disease progression and avoid complications [102]. While no Mediterranean food data set is proposed for food image recognition containing all Mediterranean cuisines, such as Greece and Moroccan food. A number of studies utilized specific food trays prepared in hospital restaurant [61, 77], unfortunately, due to Covid19 pandemic, the recent lockdown shows us the need for self-management technologies to monitor patients suffering from chronic diseases in their own homes [103], which lead to the importance of using food recognition systems based on homemade food data sets. Another specification of chronic diseases should be taken into consideration in dietary assessment systems is the type of the nutrient provided. Giving an example of diabetes, the glycemic index provides the levels of blood glucose rises after eating carbohydrates-based food, or how to match the recognized food with the insulin dose. Those intern attributes should be focused on based on the specific chronic disease. The food image recognition field has some particular properties which make it hard to reach the best performance in the recognition task compared to other objects types. For example: table, cat or person have common properties which could distinguish their images; however, some food items such as fruits, vegetables, pizza, French fries have regular shapes (usually), while other dishes could confront some factors such as cooking or mixing and result in different appearance. In addition, the diversity of cultures effects the food computing domain, the same dish could have different names and two distinguished dishes could have the same name based on the cuisine type. Furthermore, the ingredients of the same dish could be different from one culture to another, even if the system recognizes the correct category; the different ingredients could affect the amount of nutrition information provided, and based on this review, most of the approaches used classical cameras and RGB images for food images classification and food content analysis; however, classical cameras may be suitable for simple food images but could perform insufficient results for fine-grained food types, complex or mixed dishes, and similar food contents [104]. In addition, the specification of each cuisine impacts the recognition accuracy; food recognition systems trained and tested on a Chinese food data set may perform differently on an Italian food data set. Two main challenges are related to the volume estimation of food portions. The first limit is the lack of size annotation in publicly available large-scale data sets for the purpose of applying supervised deep learning models in volume estimation approaches. Second, many studies of food image recognition for dietary assessment missed the recognition of food leftovers. Leftover food results in a loss of consumed nutrients, since the amount of the provided nutrient in a food plate before consumption is not the same as after consumption in case the user could not finish its meal [105].

Opportunities and Future Work

Most of the presented studies are publications from the last 4 years, which means that deep learning is an emerging technology and still needs future improvement in the field of food image recognition based on the mentioned limits and challenges. Starting with data, combining existing data sets or generating virtual food images could help in augmenting data sets. A designed large scale data set for general food images would aid to apply the concept of transfer learning in food domain. Transferring a pre-trained, learning model on a large scale food data set and not just object data set, such as ImageNet data set [22], and testing it or training it on a sub-data set with the food-based weights, will definitely increase the performance. Moroccan cuisine has recently grown in popularity across the world and has become known in a variety of cuisines, particularly Mediterranean and North African cuisines, due to their shared geographical proximity. Morocco has also become one of the most appealing countries to visit due to its unique cuisine [106]. In addition, Moroccan cuisine is known for its richness of fruits, vegetables and olive oil as a source of monounsaturated fat, which is encouraged for the daily diet monitoring of patients with chronic diseases [102]. Therefore, a Moroccan food data set for food image recognition and dietary assessment is urgently needed. When a disease's course lasts more than 3 months, it is labeled chronic [107]. Chronic illnesses frequently last a lifetime. Self-management concept has improved the lives of people with chronic illness [100]. Self-management activities are summarized in drugs and treatment management, monitoring symptoms, and also nutrition and diet monitoring using different information and communication technologies (ICT) such as mobile applications, smart watches, wearable gadgets and so on to help individuals to adapt to the illness. Therefore, diabetic patients need a fully account dietary assessment system for their disease in a nonclinical environment, to recognize food items, estimate the volume of the food portion before and after eating, provide the necessary macro and micronutrients to control their blood sugar, and share securely the daily information with their healthcare participant, such as doctor or dietitians, using a reachable information and communication technologies (ICT). In food computing, most of the applications of multispectral and hyperspectral imaging are utilized in the industry of food quality assessment. This technology could be very useful in dietary assessment using multi/hyper spectral imaging to extract nutrition information in food ingredients. Imaging systems such as spectral imaging is capable of defying external and internal attributes based on the chemical information of the food property. Both spatial and spectral information will help in distinguishing similar dishes with different ingredients such as oil in a salad plate and determine thereafter if it is a non-fat salad of fat-salad [104, 108]. In addition, spectral imaging systems that adopt deep learning models have shown impressive results and be considered for food image recognition and nutrition analysis [109]. Concept-oriented deep learning is a technique that assists the model to understand the context of the data and oriented it with the help of supplemented information. For example, food cooking, food ingredients and food nationality are factors that interfere with the performance of food recognition and nutrition information. Orienting the model with the specific used ingredients or how the dish was cooked will refine the results and overcome the particular properties of food image recognition.

Conclusion

This paper introduces the latest works in food image recognition and nutrition analysis using deep learning applications. We conducted a systematic research, gathered several approaches based on certain criteria and extracted the key information concentrating on three main aspects: food image classification, food image segmentation and food volume estimation. In addition, a number of public and private data sets of food images are presented. Deep learning approaches are the most commonly utilized method in these studies, indicating impressive results and outperforming conventional machine learning methods. Exploring this paper, several challenges and limits have been reported in this paper to open probable future study directions, especially in dietary assessment for chronic diseases monitoring.

Declarations

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Frankenfeld CL, Leslie TF, Makara MA. Diabetes, obesity, and recommended fruit and vegetable consumption in relation to food environment sub-types: a cross-sectional analysis of behavioral risk factor surveillance system, united states census, and food establishment data. BMC Public Health. 2015;15(1):1–9.
2. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8(1):1–74.
3. Min W, Jiang S, Liu L, Rui Y, Jain R. A survey on food computing. ACM Comput Surv (CSUR). 2019;52(5):1–36.
4. Dhillon A, Verma GK. Convolutional neural network: a review of models, methodologies and applications to object detection. Progress Artif Intell. 2020;9(2):85–112.
5. Lo FPW, Sun Y, Qiu J, Lo B. Image-based food classification and volume estimation for dietary assessment: a review. IEEE J Biomed Health Inform. 2020;24(7):1926–39.
6. Chopra M, Purwar A. Recent studies on segmentation techniques for food recognition: a survey. Arch Comput Methods Eng 2021;1–14.
7. Jiang L, Qiu B, Liu X, Huang C, Lin K. Deepfood: food image analysis and dietary assessment via deep model. IEEE Access. 2020;8:47477–89.
8. Ege T, Ando Y, Tanno R, Shimoda W, Yanai K. Image-based estimation of real food size for accurate food calorie estimation. In: 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2019; p. 274–9. IEEE.
9. Chiang M-L, Wu C-A, Feng J-K, Fang C-Y, Chen S-W. Food calorie and nutrition analysis system based on mask r-CNN. In: 2019 IEEE 5th International Conference on Computer and Communications (ICCC), 2019;p. 1721–8. IEEE.
10. Ambadkar A, Chaudhari C, Ghadage M, Bhalekar M. A model for automated food logging through food recognition and attribute estimation using deep learning. In: ICT Analysis and Applications, 2021;p. 583–92 . Springer.
11. Latif G, Alsalem B, Mubarky W, Mohammad N, Alghazo J. Automatic fruits calories estimation through convolutional neural networks. In: Proceedings of the 2020 6th International Conference on Computer and Technology Applications, 2020;p. 17–21.
12. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016.
13. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25:1097–105.

14. Simonyan, K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014; arXiv preprint arXiv:1409.1556.
15. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015;p. 1–9
16. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI Conference on Artificial Intelligence 2017.
17. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016;p. 770–8
18. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017;p. 4700–8
19. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. Mobilenets: efficient convolutional neural networks for mobile vision applications. 2017; arXiv preprint arXiv:1704.04861
20. Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. In:International Conference on Machine Learning, 2019;p. 6105–14 . PMLR
21. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009;p. 248–55. Ieee
22. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. In:International Conference on Artificial Neural Networks, 2018;p. 270–9. Springer
23. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014;p. 580–7
24. Girshick R. Fast r-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, 2015;p. 1440–8
25. Ren S, He K, Girshick R, Sun J. Faster r-CNN: towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst. 2015;28
26. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC. Ssd: single shot multibox detector. In: European Conference on Computer Vision, 2016;p. 21–37. Springer
27. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016;p. 779–88
28. He K, Gkioxari G, Dollár P, Girshick R. Mask r-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, 2017;p. 2961–9
29. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, 2015;p. 234–41. Springer
30. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Mach Intell. 2017;40(4):834–48.
31. Medsker L, Jain LC. Recurrent neural networks: design and applications. CRC Press; 1999.
32. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80.
33. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst. 2017;30
34. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. Adv Neural Inf Process Syst. 2014;27
35. Bank D, Koenigstein N, Giryes, R. Autoencoders. 2020; arXiv preprint arXiv:2003.05991
36. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G. Human-level control through deep reinforcement learning. Nature. 2015;518(7540):529–33.
37. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M. {TensorFlow}: a system for {Large-Scale} machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), 2016;p. 265–83.
38. Chollet F, et al. Keras. GitHub 2015.
39. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, et al. Pytorch: An imperative style, high-performance deep learning library. Adv Neural Inf Process Syst. 2019;32.
40. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. Caffe: convolutional architecture for fast feature embedding. In:Proceedings of the 22nd ACM International Conference on Multimedia, 2014;p. 675–8.
41. Bergstra J, Breuleux O, Bastien F, Lamblin P, Pascanu R, Desjardins G, Turian J, Warde-Farley D, Bengio Y. Theano: a cpu and gpu math compiler in python. In: Proceeding of 9th Python in Science Conf, 2010;vol. 1, p. 3–10
42. Ran H, Gao W, Mi J, Zhao Z. Fine-grained recognition of Chinese food image based on densenet with attention mechanism. In: Twelfth International Conference on Graphics and Image Processing (ICGIP 2020), 2021;vol. 11720, p. 117201. International Society for Optics and Photonics
43. Ye H, Zou Q. Food recognition and dietary assessment for healthcare system at mobile device end using mask r-CNN. In: International Conference on Testbeds and Research Infrastructures, 2019;p. 18–35. Springer
44. Şengür A, Akbulut Y, Budak Ü. Food image classification with deep features. In: 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), 2019;p. 1–6. IEEE
45. Zhang W, Wu J, Yang Y. Wi-hsnn: a subnetwork-based encoding structure for dimension reduction and food classification via harnessing multi-cnn model high-level features. Neurocomputing. 2020;414:57–66.
46. Chakrabarty A, Doyle FJ, Dassau E. Deep learning assisted macronutrient estimation for feedforward-feedback control in artificial pancreas systems. In:2018 Annual American Control Conference (ACC), (2018); p. 3564–70. IEEE.
47. Qiu J, Lo FPW, Sun Y, Wang S, Lo B. Mining discriminative food regions for accurate food recognition 2019.
48. Deng L, Chen J, Sun Q, He X, Tang S, Ming Z, Zhang Y, Chua TS. Mixed-dish recognition with contextual relation networks. In: Proceedings of the 27th ACM International Conference on Multimedia, 2019;p. 112–20
49. Mao R, He J, Shao Z, Yarlagadda SK, Zhu F. Visual aware hierarchy based food recognition. In: International Conference on Pattern Recognition, 2021;p. 571–98. Springer.
50. Tan RZ, Chew X, Khaw KW. Quantized deep residual convolutional neural network for image-based dietary assessment. IEEE Access. 2020;8:111875–88.
51. Xue G, Liu S, Ma Y. A hybrid deep learning-based fruit classification using attention model and convolution autoencoder. Complex Intell Syst. 2020;1–11
52. Mandal B, Puhan NB, Verma A. Deep convolutional generative adversarial network-based food recognition using partially labeled data. IEEE Sens Lett. 2018;3(2):1–4.

53. Minija SJ, Emmanuel WS. Food recognition using neural network classifier and multiple hypotheses image segmentation. Imaging Sci J. 2020;68(2):100–13.
54. Lu Y, Stathopoulou T, Vasiloglou MF, Pinault LF, Kiley C, Spanakis EK, Mougiakakou S. gofoodtm: an artificial intelligence system for dietary assessment. Sensors. 2020;20(15):4283.
55. Freitas CN, Cordeiro FR, Macario V. Myfood: a food segmentation and classification system to aid nutritional monitoring. In:2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2020;p. 234–9 . IEEE
56. Siemon MS, Shihavuddin A, Ravn-Haren G. Sequential transfer learning based on hierarchical clustering for improved performance in deep learning based food segmentation. Sci Rep. 2021;11(1):1–14.
57. Pinzón-Arenas JO, Jiménez-Moreno R, Pachón-Suescún CG. Resseg: residual encoder-decoder convolutional neural network for food segmentation. Int J Elect Comput Eng. 2020;10(1):1017.
58. Abdullahi S, Muangchoo K. Semantic parsing for automatic retail food image recognition. Int J Adv Trend Comput Sci Eng 2020;7808–7816
59. Shimoda W, Yanai K. Weakly-supervised plate and food region segmentation. In:2020 IEEE International Conference on Multimedia and Expo (ICME), 2020;p. 1–6 . IEEE.
60. Pfisterer KJ, Amelard R, Chung AG, Syrnyk B, MacLean A, Keller HH, Wong A. When segmentation is not enough: rectifying visual-volume discordance through multisensor depth-refined semantic segmentation for food intake tracking in long-term care. 2019; arXiv preprint arXiv:1910.11250.
61. Lu Y, Stathopoulou T, Vasiloglou MF, Christodoulidis S, Stanga Z, Mougiakakou S. An artificial intelligence-based system to assess nutrient intake for hospitalised patients. IEEE Trans Multimed. 2020;23:1136–47.
62. Sharma U, Artacho B, Savakis A. Gourmetnet: food segmentation using multi-scale waterfall features with spatial and channel attention. Sensors. 2021;21(22):7504.
63. Nguyen H-T, Ngo C-W, Chan W-K. Sibnet: food instance counting and segmentation. Pattern Recogn. 2022;124: 108470.
64. Park D, Lee J, Lee J, Lee K. Deep learning based food instance segmentation using synthetic data. In: 2021 18th International Conference on Ubiquitous Robots (UR), 2021;p. 499–505. IEEE.
65. Gao J, Tan W, Ma L, Wang Y, Tang W. Musefood: Multi-sensor-based food volume estimation on smartphones. In: 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2019;p. 899–906 . IEEE
66. Priyaa PK, Sathyapriya S, Arockiam L. Nutrition monitoring and calorie estimation using internet of things (IoT). Int J Innov Technol Explor Eng. 2019;8(11):2669–72.
67. Jia W, Ren Y, Li B, Beatrice B, Que J, Cao S, Wu Z, Mao Z-H, Lo B, Anderson AK. A novel approach to dining bowl reconstruction for image-based food volume estimation. Sensors. 2022;22(4):1493.
68. 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: 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2020;p. 49–54. IEEE.
69. Hafiz R, Haque MR, Rakshit A, Uddin MS. Image-based soft drink type classification and dietary assessment system using deep convolutional neural network with transfer learning. J King Saud Univ Comput Inf Sci 2020.
70. 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(13):1556.
71. Yang Y, Jia W, Bucher T, Zhang H, Sun M. Image-based food portion size estimation using a smartphone without a fiducial marker. Public Health Nutr. 2019;22(7):1180–92.
72. Fang S, Shao Z, Mao R, Fu C, Delp EJ, Zhu F, Kerr DA, Boushey CJ. Single-view food portion estimation: learning image-to-energy mappings using generative adversarial networks. In: 2018 25th IEEE International Conference on Image Processing (ICIP), 2018;p. 251–5. IEEE
73. Graikos A, Charisis V, Iakovakis D, Hadjidimitriou S, Hadjileontiadis L. Single image-based food volume estimation using monocular depth-prediction networks. In:International Conference on Human-Computer Interaction, 2020;p. 532–43. Springer.
74. Makhsous S, Bharadwaj M, Atkinson BE, Novosselov IV, Mamishev AV. Dietsensor: automatic dietary intake measurement using mobile 3d scanning sensor for diabetic patients. Sensors. 2020;20(12):3380.
75. Muralidhar ES, Gowtham TS, Jain A, Padmaveni K. Development of health monitoring application using machine learning on android platform. In:2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020;p. 1076–85. IEEE.
76. Lo FP-W, Sun Y, Qiu J, Lo BP. Point2volume: a vision-based dietary assessment approach using view synthesis. IEEE Trans Industr Inf. 2019;16(1):577–86.
77. Rhyner D, Loher H, Dehais J, Anthimopoulos M, Shevchik S, Botwey RH, 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(5):5567.
78. Hou S, Feng Y, Wang Z. Vegfru: a domain-specific dataset for fine-grained visual categorization. In: Proceedings of the IEEE International Conference on Computer Vision, 2017;p. 541–9.
79. Mureşan H, Oltean M. Fruit recognition from images using deep learning. 2017; arXiv preprint arXiv:1712.00580 .
80. Waltner G, Schwarz M, Ladstatter 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: Proceeding of ICIAP Workshop on Multimedia Assisted Dietary Management (MADIMA) 2017.
81. Matsuda Y, Hoashi H, Yanai K. Recognition of multiple-food images by detecting candidate regions. In: 2012 IEEE International Conference on Multimedia and Expo, 2012;p. 25–30 . IEEE.
82. Kawano Y, Yanai K. Automatic expansion of a food image dataset leveraging existing categories with domain adaptation. In: European Conference on Computer Vision, 2014;p. 3–17. Springer.
83. Bossard L, Guillaumin M, Van Gool L. Food-101–mining discriminative components with random forests. In: European Conference on Computer Vision, 2014;p. 446–61. Springer.
84. Farinella GM, Allegra D, Stanco F. A benchmark dataset to study the representation of food images. In: European Conference on Computer Vision, 2014;p. 584–99. Springer.
85. Pouladzadeh P, Yassine A, Shirmohammadi S. Foodd: food detection dataset for calorie measurement using food images. In: International Conference on Image Analysis and Processing, 2015;p. 441–8. Springer
86. Kaur P, Sikka K, Wang W, Belongie S, Divakaran A. Foodx-251: a dataset for fine-grained food classification. 2019; arXiv preprint arXiv:1907.06167
87. Ciocca G, Napoletano P, Schettini R. Learning CNN-based features for retrieval of food images. In: International Conference on Image Analysis and Processing, 2017;p. 426–34 . Springer.

88. Min W, Liu L, Wang Z, Luo Z, Wei X, Wei X, Jiang S. Isia food-500: a dataset for large-scale food recognition via stacked global-local attention network. In: Proceedings of the 28th ACM International Conference on Multimedia, 2020;p. 393–401.
89. Chen J, Ngo C-W. Deep-based ingredient recognition for cooking recipe retrieval. In: Proceedings of the 24th ACM International Conference on Multimedia, 2016;p. 32–41.
90. Chen X, Zhu Y, Zhou H, Diao L, Wang D. Chinesefoodnet: a large-scale image dataset for Chinese food recognition. 2017; arXiv preprint arXiv:1705.02743
91. Jalal M, Wang K, Jefferson S, Zheng Y, Nsoesie EO, Betke M. Scraping social media photos posted in Kenya and elsewhere to detect and analyze food types. In: Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management, 2019;pp. 50–9.
92. Chen M, Dhingra K, Wu W, Yang L, Sukthankar R, Yang J. Pfid: pittsburgh fast-food image dataset. In: 2009 16th IEEE International Conference on Image Processing (ICIP), 2009;p. 289–92. IEEE
93. Ciocca G, Napoletano P, Schettini R. Food recognition: a new dataset, experiments and results. IEEE J Biomed Health Inform. 2017;21(3):588–98. https://doi.org/10.1109/JBHI.2016.2636441.
94. Termritthikun C, Muneesawang P, Kanprachar S. Nu-innet: Thai food image recognition using convolutional neural networks on smartphone. J Telecommun Electron Comput Eng (JTEC). 2017;9(2–6):63–7.
95. Min W, Wang Z, Liu Y, Luo M, Kang L, Wei X, Wei X, Jiang S. Large scale visual food recognition. arXiv preprint 2021; arXiv:2103.16107.
96. Tran AC, Tran NC, Duong-Trung N. Recognition and quantity estimation of pastry images using pre-training deep convolutional networks. In: International Conference on Future Data and Security Engineering, 2020;p. 200–14. Springer.
97. Cai Q, Li J, Li H, Weng Y. Btbufood-60: dataset for object detection in food field. In: 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), 2019;p. 1–4 . IEEE.
98. Lee GG, Huang C-W, Chen J-H, Chen S-Y, Chen H-L. Aifood: a large scale food images dataset for ingredient recognition. In: TENCON 2019-2019 IEEE Region 10 Conference (TENCON), 2019;p. 802–5. IEEE.
99. Güngör C, Baltacı F, Erdem A, Erdem E. Turkish cuisine: a benchmark dataset with Turkish meals for food recognition. In: 2017 25th Signal Processing and Communications Applications Conference (SIU), 2017;p. 1–4. IEEE.
100. Dadgar M, Joshi KD. The role of information and communication technology in self-management of chronic diseases: an empirical investigation through value sensitive design. J Assoc Inf Syst. 2018;19(2):2.
101. Sami W, Ansari T, Butt NS, Ab Hamid MR. Effect of diet on type 2 diabetes mellitus: a review. Int J Health Sci. 2017;11(2):65.
102. Chauveau P, Aparicio M, Bellizzi V, Campbell K, Hong X, Johansson L, Kolko A, Molina P, Sezer S, Wanner C. Mediterranean diet as the diet of choice for patients with chronic kidney disease. Nephrol Dial Transplant. 2018;33(5):725–35.
103. Ahmed A, Boopathy P. Artificial intelligence for the novel corona virus (Covid-19) pandemic: opportunities, challenges, and future directions. Int J E-Health Med Commun (IJEHMC). 2021;13(2):1–21.
104. Raju VB, Sazonov E. Detection of oil-containing dressing on salad leaves using multispectral imaging. IEEE Access. 2020;8:86196–206.
105. Sari YA, Dewi RK, Maligan JM, Ananta AS, Adinugroho S. Automatic food leftover estimation in tray box using image segmentation. In: 2019 International Conference on Sustainable Information Engineering and Technology (SIET), 2019;p. 212–6. IEEE.
106. Monsat, C. In the secrets of Moroccan gastronomy. https://www.lefigaro.fr/gastronomie/2015/02/09/30005-20150209ARTFIG00282-dans-les-secrets-de-la-gastronomie-marocaine.php
107. Adams WK. Martinez: summary health statistics for the us population: national health interview survey, 2012. Vital and Health statistics. Series 10, Data From the National Health Survey, 2013;1–95
108. Yoshimura Y, Zaima N. Application of mass spectrometry imaging for visualizing food components. Foods. 2020;9(5):575.
109. Huang L, Luo R, Liu X, Hao X. Spectral imaging with deep learning. Light Sci Appl. 2022;11(1):1–19.