REVIEW ARTICLE

Food Calorie Estimation Using Computer Vision : A Comparative Review of Models and Approaches

Paula Queipo-Alvarez [a], Mario Munoz-Organero [a], Asnida Abdul Wahab[b], Mohd Shafry Mohd Rahim[c,d]

[a] Department of Telematic Engineering, School of Engineering, Universidad Carlos III de Madrid, 28911 Leganes, Spain
[b] Department of Biomedical Engineering & Health Sciences, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia
[c] Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia
[d] Faculty of Computer & Information Technology, Sohar University, Sohar, Oman

ABSTRACT

Introduction: This literature review explores recent research on estimating food calories using Computer Vision. It benefits nutritional monitoring, food logging and dietary assessment. Monitoring the diet, glycaemic indices, total calories and consumed nutrients can improve wellbeing across populations. Computer Vision has emerged as a promising approach for calorie estimation, with 56 studies reviewed covering key subtasks such as image segmentation, classification, category recognition, category detection, volume estimation, quantity detection and nutrients estimation. Prominent methods include pre-trained Convolutional Neural Networks (CNNs), Transfer Learning techniques such as ResNet and EfficientNet, and multi-task learning architectures. Evaluation commonly relies on metrics such as accuracy and relative error, with Convolutional Neural Networks and multi-task frameworks achieving over 90% accuracy in food segmentation and recognition and less than 15 % relative error in calorie estimation. There is still a long way to go with multi-layered dishes and homemade sauces, so categorising and annotating recipes and nutritional content in databases is still necessary.
Malaysian Journal of Medicine and Health Sciences (2025) 21(SUPP11): 194-201.doi:10.47836/mjmhs.21.s11.27

Keywords: Calorie estimation, Computer vision, Food recognition, Food segmentation, Dietary assessment.

Corresponding Author:
Miss Paula Queipo-Alvarez, PhD
Email : paulaqueipo1@gmail.com

INTRODUCTION

As the world becomes more sedentary, to have good eating habits, to monitor them and to be able to control our caloric intake has never been more critical. Dietary assessment has traditionally relied on self-reporting methods such as 24-hour dietary recalls, food frequency questionnaires, and weighed food diaries [1]. While widely used in nutrition research and clinical practice, these approaches are prone to underreporting, recall bias, and high participant burden. The limitations of these traditional methods have motivated the development of automated tools using computer vision to improve objectivity and scalability in calorie estimation. Expanding food knowledge to decide in which restaurant to eat or which food to buy in the supermarket has great advantages, such as losing weight, improving recipes, avoiding allergens, or controlling digestive and cardiovascular diseases. Whether for health, wellness or fitness reasons, knowing what we eat and how to improve is crucial in a world full of ultra-processed products, sugary drinks and refined oils, which can be approached by studying the macronutrients consumed at each meal (fats, proteins and carbohydrates) or by calculating the energy in kcal, though this can be more complex. There are different techniques available to estimate the total calories in a multi-dish meal. In the case of processed beverages, it is necessary to recognize the type and brand of the beverage, estimate the volume of the glass and locate the nutrition facts table available on a package. Secondly, for food such as an apple, it is needed to classify the fruit, estimate the portion or the weight of the apple and consult its nutritional profile in a database. Thirdly, there are several methods to estimate the calories of the rice and tomato sauce dish. One idea is region segmentation, food category classification and dish depth estimation. Another way to do this would be to recognize the food item on the plate, and estimate its size in weight or volume. Both cases suffer complications in case there are hidden elements at the bottom of the dish, besides the fact that the composition of the cooked sauce is unknown. A query was conducted on 01/03/2024 on Web of Science database for the intersection of the topics 'food calories estimation' and 'computer vision'. The number of documents retrieved was 56. The inclusion criteria addressed articles that were in English, mentioned food and focused on efficiently estimating calories from meal images. The exclusion criteria rejected articles targeting the measurement of calories as in food manufacturing or diet planning. Three papers were excluded: one was in other language, another one was duplicated and the last one did not mention food. Two literature reviews were available. One [1] focuses on Deep Learning in the food domain and other domains such as quality detection in fruits, meat, vegetables, seafood, food supply chain and food contamination. The other review found [2] revolves around food recognition systems on dietary assessments, synthesises 159 studies on image-based food recognition systems (IBFRS) of calorie estimation and collects all the models used in 27 publicly available food datasets (PAFDs). In total, 58% of the studies found opted for Convolutional Neural Networks to address food item segmentation, classification and calorie estimation [2]. Sedentary lifestyles have increased the need to monitor eating habits and caloric intake. Estimating food calories from images involves tasks such as include image segmentation, classification, category recognition, category detection, volume estimation, quantity detection and nutrients estimation. A review of 56 publications highlights recent advances and prior literature on Deep Learning and food recognition systems for dietary assessment. Computer Vision offers clear benefits for calorie estimation, including automation, reduced user burden compared to manual food logging, and the potential for real-time nutritional feedback through mobile devices.

Dietary Assessment

Unlike traditional dietary surveys, which are prone to recall bias and under-reporting, Computer Vision-based methods allow objective estimation of portion size and caloric intake directly from images.

A meal can consist of several dishes and portions of whole foods, with each dish combining one or more servings or recipes, and each recipe containing ingredients, seasoning, oil, salt or sugar. Depending on the complexity of the meal, image segmentation can be recommended or not. The pipeline of work usually follows several steps. On the one hand, separate the image pixels to locate the food, detect if it is food and classify the dish or ingredient into preset categories. The processed food sold in supermarkets offers nutrition fact tables that enable to calculate the calories per serving. In the case of whole-foods and home cooked meals, the calculation can be done with the help of nutrition databases. On the other hand, it is necessary to estimate how many calories a certain ingredient or recipe has with the help of quantity detection.

In the field of Computer Vision, there have been historic advances. In the past, food calorie estimation used basic food recognition models. Then, convolutional neural networks, recurrent neural networks and transfer learning were introduced, and classification accuracy was improved. Later, new discoveries such as semantic segmentation and instance segmentation allowed knowledge to advance in food segmentation. In addition, methods for depth and volume estimation emerged. In the current landscape, more sophisticated Deep Learning models appear that segment, recognize and estimate nutritional facts from meal images. In Fig. 1, a graphical summary of typical food calorie estimation tasks is shown.

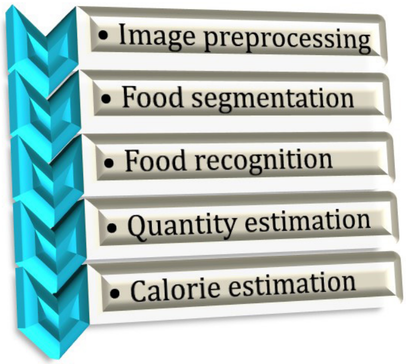

Fig 1: Food calorie estimation tasks.

For several years, there has been research on how to provide mobile applications that monitor caloric intake to improve the health or fitness of users. In the work [3] DietCam was presented, a meal classification system that reconstructs the 3D model of the image and calculates the volume with the aim of automating calorie estimates of a menu with little user intervention.

Regarding the dataset used, most of the works constructed a food meal corpus from internet open data. Websites such as school lunches, cooking recipes or restaurant menus were scrapped to find the best datasets available. Some images were in black and white, other ones multispectral. In a few cases, video frames were processed. Also, 3D images such as LIDAR were studied. In some cases, databases like USDA dataset [4] or APIs like Nutritionix [5] are used.

Dietary assessment is useful for monitoring nutrition and calories. Processed foods use nutrition labels, while whole or home-cooked meals require more complex methods because ingredients and preparation are harder to measure. Advances in Deep Learning and

3D techniques, now embedded in mobile applications, have improved accuracy and made dietary monitoring more practical.

Food image segmentation

Most of the literature that researches calorie amount estimation using Computer Vision and food items classification has required a corpus of meal images. After the preprocessing of food images, one important technique that gives good results is image segmentation, that is, to locate a specific food item in the image. There are different types of food images segmentation, such as region, semantic and instance segmentation.

According to [6], food segmentation can be binary and allows food to be separated from the background. Also, food segmentation is sometimes combined with food recognition even though results could get worse. The used networks in [6] were Mask R-CNN and ResNet-50. One of the limitations found was the inequalities in food segmentation caused by the different quality of mobile cameras. The authors [7] considered impractical to collect many images of all food categories, and chose to use a small sample for training and food region segmentation to divide the food images based on shared characteristics.

Food semantic segmentation labels each pixel with a predefined class and requires many pixel-level annotated images, but it does not distinguish between different food items of the same class, since it assigns a class to each pixel separately. For example, a model combining SegNet with MobileNet. The paper [7] presented food region segmentation using word embeddings.

The research [8] used a food semantic segmentation model to estimate total calories. The dataset was constructed with 593 images of multiple-dish meals annotated with total calories. However, one limitation of estimating calories from images is that sometimes calories are underestimated because some items such as bread or rice are obscured, which prevents good region segmentation of food.

Another possibility is to use object detection techniques, such as the Single Shot MultiBox detector (SSD). This technique was employed in the article [9] for dish detection and classification to identify Chinese and Western dishes.

In the study [10], it was used semantic food segmentation with SegNet and MobileNet.

In another work [11], the food category recognition was carried out inspired by Transfer Learning and Alexnet CNN. Furthermore, in [12] the multi-label food classifier was pre-trained using Faster R-CNN. Moreover, [13] proposed a new method called USFoodSeg that combines zeros-shot and few-shot models using food images and food words. It solved the problem of insufficient food segmentation data when obtaining good results and when used to extend datasets without human assistance. Regarding the evaluation, the results in food segmentation and classification reach high values in accuracy. However, the results could be improved avoiding dataset error or missing values. For example, sampling failures, missing meals or missing size measurement.
There are several metrics to calculate the accuracy of the food segmentation and food classification. One is the precision, that calculates the true positives among all the items classified as positive. Table I compares the accuracy in different datasets, using CNN models.

Table I : Performance Comparison of Models for Food Segmentation and Classification.

Reference	Dataset	Model	Result
(X. Y. Kong et al., 2023)	Constructed with 10000 images from data enhancement labelled with software	Semantic food segmentation SegNet + MobileNet	Performance accuracy = 97.82 %
(Rahmat & Kutty, 2021)	Six constructed dataset, each with 30 images labelled with the category	Food category recognition, Alexnet CNN and Transfer Learning	Accuracy of training = 91.43 %
(Nadeem et al., 2023)	Constructed with 16000 food items images categorised into 14 different categories	Multi-label food classifier, pre-trained Faster R-CNN	Accuracy = 80.06 %

Based on the results, the first approach exhibits a good performance accuracy of 97.82 % in food segmentation. Food image segmentation locates food items in meal images with techniques such as region, semantic, and instance segmentation, using models like Mask R-CNN, SegNet, and MobileNet. Accuracy is high, but challenges remain with obscured foods, image quality, and limited datasets.

Food classification, detection and recognition

When assessing the types of food present in a meal, different taxonomies can be studied. The granularity of the classification depends on the scope of each study. Sometimes several raters are used to label manually, define labels and merge synonyms. Each food image can have more than one label, because there could be different food items in the image. In [14], a dataset called Food201MultiLabel was created. When analysing the corpus, it was found that each image had 1.9 food item labels on average, a median of one and a maximum of eight labels.

In order to understand what type of food is consumed, there are different Computer Vision techniques available, such as food classification, food detection and food recognition.

Also, two different approaches were followed [15], a Single-task and a Multi-task Convolutional Neural Networks with the idea of estimating the food categories of meals from cooking recipes.

Food classification labels all the food content into a category without locating each food item in the image. Some traditional models include neural networks, Support Vector Machines, Random Forest, K-nearest neighbours, Naive Bayes, among others. The most used models nowadays are convolutional neural networks. According to the literature review [2], CNN significantly improved the classification accuracy achieved on the Food-101 dataset, increasing from 55.3 % to 90.27%
.

Food detection focuses on locating the presence of certain ingredients or recipes on the meal images and drawing boxes on the detected food items. According to the study [16], there is a framework that allows food detection using the Clarifai API called NutriTrack.

Food recognition tasks involve both identifying and locating ingredient, recipes or cuisine categories. Table I shows the accuracy obtained in food recognition. In the article [17], the objective is to recognise between eight fruit categories. DeepFruit is a collection of 21122 labelled images where each dish shows a combination of up to four or five different fruits. In the work [18] is introduced a novel approach that segments food using the CSWWLIFC algorithm. Also, food recognition is performed using the Whale Levenberg Marquardt Neural Network (WLM-NN) classifier. In their opinion, poor food recognition results lead to biases in calorie amount estimations and it is harder to estimate calories based on appearance rather than actual proportions.

Another different approach that allows unsupervised training is food image clustering. With the aim of decreasing the number of food meals that need to be dietary assessed, [19] used human calories estimation. Instead of considering food recognition, the new approach clustered similar images that come from the same video. Food image clustering with K-means and Affinity Propagation (AP) was employed.

Food classification, detection, and recognition identify and locate food items in meal images. Classification assigns categories without locating items, detection draws bounding boxes, and recognition combines both tasks. Convolutional neural networks (CNNs) are widely used, achieving high accuracy, while unsupervised clustering methods reduce the need for manual labelling.

Quantity detection, weight and volume estimation

After food segmentation and food recognition, it is time to estimate the physical size of each ingredient or food item.

One aspect to consider is the scaling factor of the images. Reference objects like coins help to map between pixels and real dimensions. The quantity of each serving can be calculated using size, weight or volume. The authors of the study [20] performed food portion sorting with three categories, namely small, medium and large. This facilitates escaping from the 3D estimation and to obtain high accuracy when utilising GoogleNet, MobileNet and Bags of Words.

In the study [21], a real size estimation was carried out with a CNN based on VGG16. In many cases, it is necessary to estimate the depth of the dish or food item. In the survey [22], several true-depth technologies are referred to, such as LIDAR.

According to [6], mixed and layered food items are difficult to analyse as the surface is covered by the upper elements. That is the reason the present work is composed of two approaches to generate depth maps. The neuralbased approach used a single image and the geometrybased approach used a pair of images. With some assumptions and with the presence of a reference card, it was possible to compute the volume of each food item. Bite counting is another approach when doing food recognition and quantity estimation. More traditional approaches utilise only visible food items in the picture, without considering the items that are hidden in the bottom. With the idea of modelling three dimensions and focusing on consumed food calories, the researchers in [23] decided to collect several video frames from more than one angle using cameras leaning on the shoulder of the eaters. Although bite size estimation was out of scope, bite counting estimation was considered a regression task and obtained better results than when it was considered a classification task. Regarding the evaluation, it is difficult to have a reliable volume estimate. In order to improve the results, one assumption is to consider one layered food, so that the food items are not hidden and the volume estimation gets simplified. It is very important to utilise a reference object.

Some frequently used metrics are Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Percentage Error. Also, Mean Absolute Percentage Error (MAPE) is considered when expressing the estimation error as a percentage of the real value. Table II displays MAPE from two different approaches. In [6] work, the neural model obtains a smaller MAPE error than the geometric based approach as shown in Table II. The second result emerges a good result for food volume estimation with a low mean absolute percentage error of 27.41 % and a standard deviation of 16.9 %. Quantity, weight, and volume estimation follow food segmentation and recognition to determine the physical size of each item. Techniques include CNNs, depth estimation, LIDAR, and multi-angle video capture, while layered foods and hidden items remain challenging. Metrics like MAE, MSE, RMSE, and MAPE are used, with neural approaches generally outperforming geometric methods.

Table II : Performance Comparison of Models for Food Volume Estimation.

Reference	Dataset	Model	Result
(Papathanail et al., 2023)	Constructed dataset from 50 participants barcode scanning	Geometric approach and Databases	Mean absolute percentage error = 31.32 % Standard deviation = ± 22.3 %
(Papathanail et al., 2023)	Constructed dataset from 50 participants barcode scanning	Neural approach (Zoe) and Databases	Mean absolute percentage error = 27.41 % Standard deviation = ± 16.9 %
(Ege et al., 2019)	Constructed dataset with 360 rice images, annotated real size per pixel and segmentation mask	Real size estimation for 224 pixels, CNN based on VGG16	Average value of absolute error = 0.165 cm Average value of relative error = 6.394 %

Nutrients detection and calorie amount estimation

When estimating the calories of a meal, it is important to know different parameters such as the ingredients, their quantity, sauces, cooking method and the presence of processed food. With that information and the caloric density of each food item, it is possible to convert the volume estimation into nutrients and calories.

In the case of processed food, nutrition facts tables are essential to calculate the number of total calories consumed. Special care should be taken to know the brand of the product, as the composition may vary between products of the same type. Moreover, according to [6], it is important to analyse the recipes and the method of preparation, as the use of oil or butter prevents the estimation of the fats in the dish and is not visible in the image.

There are different tasks related to caloric and nutritional assessment. In calorie amount calculation or nutrients calculation, more accurate values are obtained, ideal for greater control of caloric and nutritional intake. It usually uses mathematical formulas, databases such as USDA National Database [4] and nutrition facts labels of a specific brand. Traditional approaches need the help of a human nutritionist to classify macronutrients such as proteins, fats and carbohydrates present in the food. In calorie or nutritional monitoring, the calories or nutrients ingested are recorded daily, allowing diets to be tracked. This can be done by recording food intake, scanning barcodes, taking pictures of everything that is eaten or writing down all the recipes that have been cooked. Mobile applications are the most widely used platform. In contrast, calorie or nutritional estimation focuses on assessing calorie and nutritional intake in a quick and efficient way. There are several approaches, such as predictions based on food type and serving size with the help of visual references. In the case of Computer Vision, different techniques are available to automate this information. In the research developed in [16], the Clarifai API is used to detect different foods and the data is managed in the Nutritionix API [5], setting up a framework called NutriHack that allows nutrient estimation.

In the research conducted by [6] different modules were developed. After estimating the volume or weight of each food item, different databases such as USDA [4] for volume and Nutritionix [5] for weight were consulted. The nutritional content of the whole meal was estimated after complementing the information with the AquaCalc [24] and Open Food Facts [25] databases.

Of particular note is the use of Deep Learning and Computer Vision to infer the total calories of a meal from an image. [8] research utilise 593 multiple-dishes images annotated with the total amount of calories to perform food segmentation. A common semantic segmentation model named Deeplab V3+ was used and a new calorie estimation branch is added. It is effective as it considers the previous segmentation food category annotations on a pixel-by-pixel basis to feed the model that estimates the amount of calories for each food. In the case of [26], a different approach with bag of visual words, LLC and SVC was used. Another successful technique for classification is attention mechanism with multi-task learning used in [27]. A good calorie accuracy for single items was achieved in [28] with the use of RefineNet in semantic segmentation.

One of the objectives to be achieved from the analysis of macronutrients is to perform glycaemic index calculation. Knowing the glycaemic index makes it easier to predict blood glucose levels and thus to calculate the insulin boluses needed to control diabetes. This is because the glycaemic index indicates the rate of elevated glucose peaks and their duration. However, no articles estimating the glycaemic index using Computer Vision have been found. Currently there are challenges in estimating the glycaemic index, because some ingredients have a significantly higher glycaemic content than others. For example, processed food contains extra amounts of sugar that are difficult to detect using photos. In the case of homemade food, it is also hard to measure the presence of dissolved sugar in cooked sauces.

Regarding food calorie estimation evaluation, some common metrics are absolute error, relative error, Mean Percentage Error (MPE) and Mean Absolute

Percentage Error (MAPE). The absolute error is the absolute difference between the estimated value and the real value. The relative error usually refers to the ratio of the absolute error to the real value, expressed as a percentage. Table III shows the comparison for food calorie amount estimation results.

Table III : Performance Comparison of Models for Food Calorie Amount Estimation.

Reference	Dataset	Model	Result
(Okamoto et al., 2021)	Constructed dataset with 593 multiple-dish images with total calories annotations	DeepLab V3+ calorie estimation branch	Absolute error on total calorie = 74.5 kcal Relative error = 11 %
(Beijbom et al., 2015)	Constructed dataset with 646 images from manually tagger 41 categories	Bag of visual words, LLC, SVC	Absolute error = 232 kcal Relative error = -21 %
(Ege & Yanai, 2017)	Constructed dataset with 4877 images with calorie information and from 15 categories	Multi-task CNN based on VGG16	Absolute error = 94.14 kcal Relative error = 27.93 %
(Nadeem et al., 2023)	Constructed with 16000 food items images labelled with 14 categories	Pretrained inception v2 coco and GrabCut algorithm	Average calorie computation = 10% of the real
(He et al., 2023)	Food-101 dataset annotated with most common 35 food categories in the dataset	Multi-task learning (attention mechanism in classification)	Absolute error = 91.4 kcal Relative error = 28.7 %
(Ege et al., 2019)	UECFOOD-100 with with 5,301 segmentation masks	Food region segmentation, DepthCalorieCam that uses U-Net network	Average error in fried chicken = -5 kcal Standard deviation in fried chicken = ±64 kcal
(Poply & Jothi, 2021)	UNIMIB 2016 database, selected 750 images annotated	Semantic Segmentation RefineNet	Calorie accuracy for single items = 93.06 %

The DeepLab V3+ calorie estimation model achieves an 11% relative error. Also, Semantic Segmentation with RefineNet shows a high accuracy 93.06 % for single items.

Nutrients detection and calorie amount estimation rely on knowing ingredients, quantities, cooking methods, and brand-specific data. Databases such as USDA and Nutritionix are widely used, while Deep Learning models like Deeplab V3+ and RefineNet improve accuracy, achieving up to 93% for single items.

Future directions

In order to enhance the performance of dietary monitoring applications, future research must keep going on. From the application security, reliability and robustness, user trials, real-time data computation, the camera professionalisation to the database storage and integration. The integration of Computer Vision into dietary monitoring applications provides scalable, low-cost tools that could enhance nutritional assessment in both clinical and everyday settings.

Some future work on the matter includes food segmentation, food recognition and volume estimation models. Multi-task approaches and pre-trained convolutional neural networks offer further room for improvement, because 90 % segmentation and recognition accuracy has already been achieved on some datasets. More complicated is to improve quantity estimation on multi-layered meals. Computer Vision is limited to visible areas, so new approaches such as egotistic eating videos and bites counting estimation can help.

In addition, further work is needed on recipe databases, which can be customised according to the user and improve the estimation of fats and added sugars used in home cooking. Every day, supermarket brands strive to improve information on the nutritional content of processed products, which is beneficial for monitoring both micronutrients and macronutrients and improving food quality control.

Finally, we should not forget all the information that can be obtained from consumed food monitoring. Estimating the glycaemic index can greatly help patients with diabetes to control their insulin injections daily. Future directions focus on improving segmentation, recognition, and portion estimation, with multi-task and pre-trained CNNs offering further gains. Quantity estimation for layered meals remains difficult, where video-based and bite-counting approaches may help. Advances in recipe databases and processed food labelling could improve fat and sugar estimation, while glycaemic index prediction would be valuable for diabetes management. Challenges remain with hidden fats, sugars, and glycaemic index estimation, especially in homemade or processed foods.

Conclusions

The aim of the present literature review is to offer a synthesis of studies focused on estimating food calories from Computer Vision images. Automated estimation of food calories is a cutting-edge field. All current efforts to improve the performance of dietary assessment applications are hugely beneficial to the health and nutrition of society. Nutritional tracking, food logging, fitness, and healthcare applications require reliable estimation or calculation of calories consumed and have the potential to improve the lives of patients who need to control the calories they ingest daily.

In order to improve the standard of living of patients with diabetes and other types of profiles, it is necessary to obtain a reliable value of glycaemic index, total calories and nutrients ingested. To do this, it is necessary to add up the calories provided by each type of food and to know the nutritional information and caloric density of each ingredient. Therefore, two tasks are of vital importance.

On the one hand, detecting each ingredient. That is why generally this calorie estimation domain is associated with food segmentation, food classification or food recognition. Each meal image could be multi-dish or not. Also, each dish can be composed of more than one ingredient or not. Furthermore, the cooking methods can affect the nutritional values of the ingredients. On the other hand, estimating each cooked ingredient size, volume or mass. This could be another task involved in the pipeline before estimating the meal calories.

Different standardised databases were observed in the 56 articles retrieved, although most of the datasets were self-constructed. Most of the corpus consists of multiple-dish food images annotated, generally with the total calorie amount and food categories. It is noteworthy the variety of foods studied, including Malaysian, Indian, Western and Chinese food, among others. There are frameworks and web pages that facilitate the estimation of calories and macronutrients, among which are Nutritionix, AquaCalc or Open Food Facts stand out.

Regarding the proposed techniques, different approaches working with feature extraction, category classification, convolutional neural networks, segmentation models and multi-task learning frameworks have been highlighted. Models of convolutional neural networks were developed in 11 publications, obtaining good results. Many are inspired by pre-trained models with large image corpora and use Transfer Learning. Many articles use several single-task models, but new proposals with multi-task models are observed in the state of the art. From Table I it is observed that two methods achieved greater than 90 % accuracy values in segmentation and recognition. From Table III, three studies obtained less than 15 % relative error in calorie estimation. Calorie estimation from food images is advancing, with CNNs, segmentation, and multi-task learning showing high accuracy. Reliable ingredient detection and quantity estimation remain key, though datasets and diverse cuisines remain challenges.

Acknowledgement

This collaboration between Universiti Teknologi Malaysia and Universidad Carlos III de Madrid has been financially supported by Alliance4Universities Erasmus+. The scientific and technical support provided by the Universiti Teknologi Malaysia under Vot R.J130000.7623.4C816 is acknowledged. In addition, this work has been partially supported by the "Generation of Reliable Synthetic Health Data for Federated Learning in Secure Data Spaces" Research Project (PID2022-141045OB-C43 (AEI/ERDF, EU)) funded by MCIN/ AEI/ 10.13039/501100011033 and by "ERDF A way of making Europe" by the "European Union".

REFERENCES

1. Zhou L, Zhang C, Liu F, Qiu Z, He Y (2019) Application of Deep Learning in Food: A Review. Compr Rev Food Sci Food Saf 18:1793–1811
2. Dalakleidi K V., Papadelli M, Kapolos I, Papadimitriou K (2022) Applying Image-Based Food-Recognition Systems on Dietary Assessment: A Systematic Review. Adv Nutr 13:2590–2619
3. Kong F, Tan J (2012) DietCam: Automatic dietary assessment with mobile camera phones. Pervasive Mob Comput 8:147–163
4. Harnly J. Methods and Application of Food Composition Laboratory : USDA ARS. In: USDA Natl. Database Stand. Ref. Release 27.
5. Nutritionix. https://www.nutritionix.com/. Accessed 29 May 2024
6. Papathanail I, Abdur Rahman L, Brigato L, Bez NS, Vasiloglou MF, van der Horst K, Mougiakakou S (2023) The Nutritional Content of Meal Images in Free-Living Conditions-Automatic Assessment with goFOODTM. Nutrients. https://doi.org/10.3390/NU15173835
7. Honbu Y, Yanai K (2021) Few-Shot and Zero-Shot Semantic Segmentation for Food Images. CEA 2021 - Proc 13th Int Work Multimed Cook Eat Act 25–28
8. Okamoto K, Adachi K, Yanai K (2021) RegionBased Food Calorie Estimation for Multiple-Dish Meals. CEA 2021 - Proc 13th Int Work Multimed Cook Eat Act 17–24
9. Hu H, Zhang Z, Song Y (2020) Image based food calories estimation using various models of machine learning. Proc - 2020 5th Int Conf Mech Control Comput Eng ICMCCE 2020 1874–1878
10. Kong X-Y, Sun X-H, Wang Y-Z, Peng R-Y, Li XY, Yang Y-H, Lv Y, Shih-Pang Tseng (2023) Food Calorie Estimation System Based on Semantic Segmentation Network. Sensors Mater.
11. Rahmat RA, Kutty SB (2021) Malaysian food recognition using alexnet CNN and transfer learning. ISCAIE 2021 - IEEE 11th Symp Comput Appl Ind Electron 59–64
12. Nadeem M, Shen H, Choy L, Barakat JMH (2023) Smart Diet Diary: Real-Time Mobile Application for Food Recognition. Appl Syst Innov 2023, Vol 6, Page 53 6:53
13. Honbu Y, Yanai K (2022) Unseen

Food Segmentation. ICMR 2022 - Proc 2022 Int Conf Multimed Retr 19–23
14. Myers A, Johnston N, Rathod V, Korattikara A, Gorban A, Silberman N, Guadarrama S, Papandreou G, Huang J, Murphy K (2015) Im2Calories: towards an automated mobile vision food diary – Google Research. Proc IEEE Int Conf Comput Vis 1233–1241
15. Ege T, Yanai K (2017) Simultaneous estimation of food categories and calories with multi-task CNN. Proc 15th IAPR Int Conf Mach Vis Appl MVA 2017 198–201
16. Ocay AB, Fernandez JM, Palaoag TD (2017) NutriTrack: Android-based food recognition app for nutrition awareness. 2017 3rd IEEE Int Conf Comput Commun ICCC 2017 2018-Janua:2099–2104
17. Latif G, Mohammad N, Alghazo J (2023) DeepFruit: A dataset of fruit images for fruit classification and calories calculation. Data Br. https://doi.org/10.1016/J.DIB.2023.109524
18. Emmanuel WRS, Minija SJ (2018) Fuzzy clustering and Whale-based neural network to food recognition and calorie estimation for daily dietary assessment. Sadhana - Acad Proc Eng Sci 43:1–19
19. Doulah A, Sazonov E (2017) Clustering of food intake images into food and non-food categories. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 10208 LNCS:454–463
20. Allegra D, Erba D, Farinella GM, Grazioso G, Maci PD, Stanco F, Tomaselli V (2019) Learning to rank food images. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 11752 LNCS:629–639
21. Ege T, Ando Y, Tanno R, Shimoda W, Yanai K (2019) Image-Based Estimation of Real Food Size for Accurate Food Calorie Estimation. Proc - 2nd Int Conf Multimed Inf Process Retrieval, MIPR 2019 274–279
22. Dubey SK, Kraft D, Drueeke N, Bieber G (2023) Survey on food intake methods using visual technologies. ACM Int Conf Proceeding Ser 11
23. Alshurafa N, Lin AW, Zhu F, Ghaffari R, Hester J, Delp E, Rogers J, Spring B (2019) Counting Bites With Bits: Expert Workshop Addressing Calorie and Macronutrient Intake Monitoring. J Med Internet Res. https://doi.org/10.2196/14904
24. Aqua Calc. https://www.aquacalc.com/calculate/food-volume-to-weight. Accessed 29 May 2024
25. Open Food Facts. https://ch-en.openfoodfacts.org/cgi/search.pl?search_terms=Gigandet&search_simple=1&action=process. Accessed 29 May 2024
26. Beijbom O, Joshi N, Morris D, Saponas S, Khullar S (2015) Menu-match: Restaurant-specific food logging from images. Proc - 2015 IEEE Winter Conf Appl Comput Vision, WACV 2015 844–851
27. He L, Cai Z, Ouyang D, Bai H (2022) Food Recognition Model Based on Deep Learning and Attention Mechanism. pp 206–216
28. Poply P, Angel Arul Jothi J (2021) Refined Image Segmentation for Calorie Estimation of Multipledish food items. pp 682–687