




Evaluating machine learning technologies for food computing from a data set perspective

Nauman Ullah Gilal¹ · Khaled Al-Thelaya¹ · Jumana Khalid Al-Saeed¹ · Mohamed Abdallah¹ · Jens Schneider¹ · James She¹ · Jawad Hussain Awan² · Marco Agus¹ 

Received: 6 September 2022 / Revised: 24 May 2023 / Accepted: 8 August 2023 /
Published online: 19 September 2023
© The Author(s) 2023

Abstract

Food plays an important role in our lives that goes beyond mere sustenance. Food affects behavior, mood, and social life. It has recently become an important focus of multimedia and social media applications. The rapid increase of available image data and the fast evolution of artificial intelligence, paired with a raised awareness of people's nutritional habits, have recently led to an emerging field attracting significant attention, called *food computing*, aimed at performing automatic food analysis. Food computing benefits from technologies based on modern machine learning techniques, including deep learning, deep convolutional neural networks, and transfer learning. These technologies are broadly used to address emerging problems and challenges in food-related topics, such as food recognition, classification, detection, estimation of calories and food quality, dietary assessment, food recommendation, etc. However, the specific characteristics of food image data, like visual heterogeneity, make the food classification task particularly challenging. To give an overview of the state of the art in the field, we surveyed the most recent machine learning and deep learning technologies used for food classification with a particular focus on data aspects. We collected and reviewed more than 100 papers related to the usage of machine learning and deep learning for food computing tasks. We analyze their performance on publicly available state-of-art food data sets and their potential for usage in multimedia food-related applications for various needs (communication, leisure, tourism, blogging, reverse engineering, etc.). In this paper, we perform an extensive review and categorization of available data sets: to this end, we developed and released an open web resource in which the most recent existing food data sets are collected and mapped to the corresponding geographical regions. Although artificial intelligence methods can be considered mature enough to be used in basic food classification tasks, our analysis of the state-of-the-art reveals that challenges related to the application of this technology need to be addressed. These challenges include, among others: poor representation of regional gastronomy, incorporation of adaptive learning schemes, and reverse engineering for automatic food creation and replication.

✉ Marco Agus
magus@hbku.edu.qa

Extended author information available on the last page of the article

Keywords Food computing · Food data sets · Applications · Food recognition · Food classification · Caloric estimation · Machine learning · Deep learning

1 Introduction

Background Food is an essential part of human life, not only as a biological need to sustain our daily activities and to keep an adequate health status, but also for mood balancing, leisure, and self-satisfaction. The complex function of food has thus led to the aphorism “eating for living, or living for eating”, indicating the different attitude towards food as need or as pleasure. The rapid evolution of multimedia technologies immediately reflected this natural human attitude, and it is nowadays common practice to immortalize dishes and meals through digital pictures and to share convivial or individual food-related experiences, like a particularly well done self-made dish, or a particularly yummy and well-presented restaurant meal. Just to provide an example of how much social media are focused on food, at the time of writing this report, the hashtag #food in Instagram appears in more than 484 million posts, while various other associated hashtags easily reach 100 million pictures (like #foodporn, #foodie, #instafood, etc.). At the same time, the rise in the importance of food in media communication has led to the emergence of new professions such as “food blogger” or “food influencer”, people who extensively use digital media to inform about recipes, dishes, restaurants for reviewing or marketing purposes [72]. Concurrently, the recent explosion of artificial intelligence (AI) has affected the performance and experience of multimedia systems across all domains. As a result, various applications related to food computing are continually being designed and are routinely used for activities associated with everyday meals. Out of the increasing interest to support various needs and the recent availability of public data, a new computing field called *food computing* concerned with automated food analysis has recently emerged [2, 93].

Problem The main challenges addressed by the field are related to the classification and recognition of food images, that, compared to standard image classification tasks, is considered more difficult for the following reasons:

- **Data variability:** numerous environmental and technical factors can become nuisances that affect the performance of food classification, such as lighting conditions, noise, occlusions, camera angle and the quality of images. Furthermore, variations in appearance due to different cooking styles, ingredients, and culinary cultures can complicate the classification problem [6].
- **Visual variability:** Automatic classification of food from images is a fine-grained classification problem [46], and it is affected by two significant issues: inter-class variance and intra-class variance: inter-class variance relates to food items that exhibit visual similarities despite belonging to different categories. For instance, visually distinct food items like a salad and a pizza may share certain appearance characteristics, such as round shapes, vibrant colors, and toppings. Intra-class variance, instead, refers to images within the same food category that exhibit considerable visual variations due to factors such as cooking styles, ingredients, presentation, and cultural influences. For example, pizzas with different crusts, toppings, or cooking times all fall under the same category. Figure 1 shows some examples of inter-class and intra-class variance.

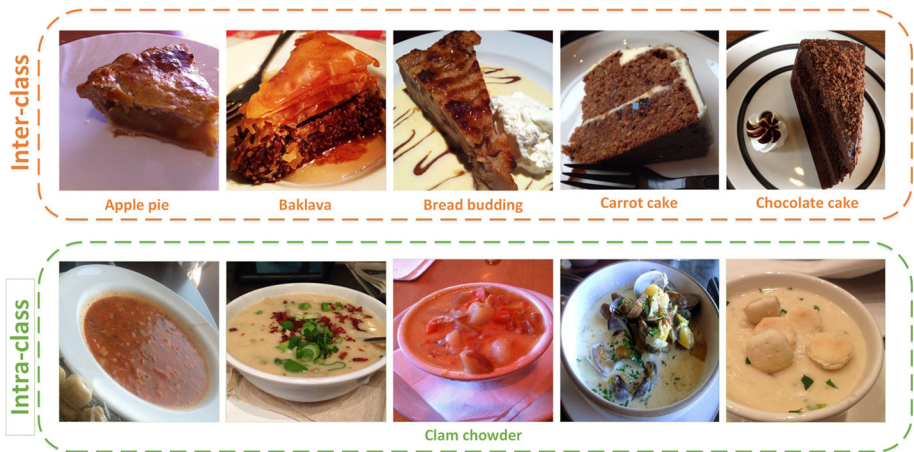


Fig. 1 Inter-class and intra-class variance in the Food-101 [9] data set. Top row: inter-class visual similarity. Bottom row: high intra-class variability

This field is especially fueled by deep learning and Convolutional Neural Networks (CNNs), which have extensively improved the accuracy of object detection, identification, and localization from single pictures [104]. Hence, in the context of food computing, machine learning approaches have been applied especially for: food detection [104, 105], food recognition [17, 76, 101, 104, 113, 128], food segmentation [28, 30, 73, 83, 103], food-tray analysis [2, 87, 95], food classification [2, 4, 19, 97, 102], ingredient recognition [13, 57, 85], food quality estimation [40, 51, 55], calorie counting [23, 56, 65, 99], and portion estimation [22, 43].

Numerous efforts have been geared towards health-related targets in order to provide nutritional guidelines to users, such as calories and nutrition estimation [3, 113], food recommendation related to specific health conditions [93], ingredients recognition for people suffering from allergies, and many more (see Fig. 2).

Aim and contributions Recent surveys about food computing [11, 53, 64, 78, 110] mostly target health related applications due to their enormous impact on society: they overview the technical aspects of computer vision approaches employed for recognition and classification. In contrast, this report surveys recent literature from a **data perspective**: we place special emphasis on the data sets used in and generated by previous work. In particular, we wish to understand data sizes, geographical coverage, and how multimedia and social media technologies in food computing leverage these data sets. Our main contributions to the field are:

1. We provide a critical analysis of recently published AI-based methods for automatic food computing, with a focus on the data sets used and generated.
2. We provide a critical analysis of recently published data sets and investigate their coverage in terms of represented cultural and regional environments, with the goal of geographical and geo-referenced classification. To this end, we release a public web resource listing the currently available data sets, and we indicate which areas of the world are still not covered. Researchers can access our web resource at <https://slowdeepfood.github.io/datasets/>.
3. We discuss remaining challenges in the field from a multimedia perspective, the future of food computing for personal and regional applications, and the challenging connections



Fig. 2 Mobile applications for food analysis. (a) Kawano et al.’s app [50]. (b) Ingredients recognition and a cooking recipe recommendation [69]. (c) Calories estimation [113]. (d) Real-time mobile application classification on Pizza-Styles [29]. (e) Real-time mobile application classification on the GCC-30 data set [29]

to robotics for automatic food creation. To this end, we try to indicate possible directions for future research efforts.

Methods We survey more than 100 papers, with topics related to:

- application of machine learning and deep learning to food computing tasks, like food detection, food recognition, and food classification tasks;
- available food image data sets for training and testing machine learning models;
- available food computing applications.

Search queries We obtained the corpus of surveyed papers through searches on popular digital libraries: Google Scholar, IEEE explorer, Springer, ACM Digital Library, and arXiv. We used the following query, combining relevant keywords: (“Machine learning” OR “Neural network” OR “deep learning”) AND (“Food applications” OR “Food detection” OR “Recognition” OR “Food computing”) AND “Data set*”. The body of research in this area is growing rapidly and this survey covers the period between 2010 and 2022. Descriptive statistics of published papers according to their category and year are shown in Fig. 3, left.

Inclusion & exclusion criteria In this survey, we only consider peer-reviewed papers and arXiv pre-prints that were published between 2010 and 2022. We excluded all papers written

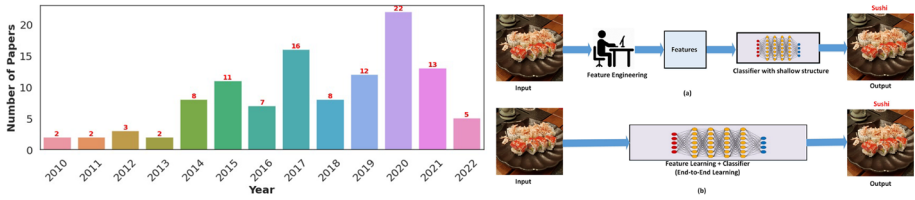


Fig. 3 Left: Categorization of the reviewed literature based on year and number of papers. This survey covers the period from 2010 to 2022 and focuses on machine learning, deep learning approaches, applications, and data sets in the food computing domain. Right: An overview of machine learning and deep learning pipelines. (a) Traditional machine learning approaches require manual feature extraction. (b) Modern deep learning approaches remove the human labeling bottleneck and automate all processes in an end-to-end framework

in languages other than English. We furthermore exclude papers that present a food computing methodology that is not specific to a given data set.

Article organization The rest of this article is organized as follows. Section 2 presents the machine learning (Subsection 2.1) and deep learning approaches (Subsection 2.2) applied to food analysis. Section 3 provides a critical analysis of food data sets, and a description of the web resource for publicly available data that we created. Finally, Section 4 highlights the remaining challenges in food recognition and classification and suggests potential avenues for future investigations.

2 Overview of food classification approaches

The aim of this survey is not to provide an extensive overview of all methods developed for addressing the food classification challenges; we refer readers to the recent surveys specifically targeting that topic. Albeit many new frameworks have been recently proposed, Min et al. [78] provide a complete review of food computing up to 2019, mostly targeting the use of machine learning approaches for classification of images containing food-related content. Additional surveys [53, 64, 110] focus more on volume quantification and caloric estimates for dietary assessment.

Here, we will provide a brief analysis of current technologies and the data sets used, and we provide guidelines for future development and applications. In general, food classification methods can be subdivided into two macro categories, corresponding to two different periods of technological advance in the field of machine learning, especially in computer vision and image processing. We observe:

- a first period characterized by the use of traditional (i.e., “shallow”) machine learning methods, more or less spanning the time between 2010 and 2016;
- a second period characterized by the use of deep learning and transfer learning, that started around 2016 when CNNs began to gain popularity in the computer vision community.

Figure 3 right illustrates the two macro categories for image classifications in the food computing domain.

2.1 Traditional machine learning approaches

We characterize traditional machine learning as being composed of building blocks like modeling, extracting and quantifying geometry, and designing visual and categorical features from images. The process involves human engineering efforts and subjective analysis for modeling and discriminating the most descriptive and significant features for a given task. Since an exhaustive review of such methods is out of the scope of this survey, we only briefly review the most common methods for feature composition and supervised classification in the context of food computing. We then discuss their practical limitations.

Starting from feature design, the following popular feature-based composition methods have been considered by the community and successfully applied in food classification tasks.

- Gabor filters [88] are linear filters that perform a directional frequency analysis around a point of interest. They are motivated by an attempt to emulate the human visual system. Gabor filters can be understood as band-pass filters obtained by modulating a Gaussian kernel with a complex, sinusoidal planar wave.
- Local Binary Patterns (LBP) [15] are visual feature vectors obtained by partitioning the image into uniform cells, and by deriving a bit-string according to the comparison between neighboring pixels. The resulting bit-string is then used for creating a normalized feature histogram.
- Bag of Feature (BoF, or Bag-of-visual words) [24, 37, 112] techniques aggregate features through clustering which are then encoded to create synthetic codebooks for classification.
- Histograms of Oriented Gradients (HOG) [70, 94] consider the occurrences of discretized gradient orientations in portions of an image. A subsequent binning process on a uniform grid is used to compute a histogram that can be used as a feature vector for classification.
- Scale Invariant Feature Transforms (SIFT) [70] consist of extracting key points of objects. Candidate matching of features is then performed using the Euclidean distance between feature vectors. The method benefits from efficient hashing on top of a generalised Hough transform.
- Bag-of-Textons [24, 117] The concept of Bag-of-Textons is inspired by the Bag-of-Words model commonly used in natural language processing. In the Bag-of-Words model, documents are represented as collections of individual words, and focusing on their frequency of occurrence. The Bag-of-Textons model represents an image as a collection of local texture patterns and their spatial arrangement. Bag-of-Textons has been widely used in computer vision studies for texture analysis and image classification.
- Pairwise Rotation Invariant Co-occurrence Local Binary Pattern (PRICoLBP) [24, 90] enhances LBP by incorporating multi-orientation, multi-scale, and multi-channel information. Unlike LBP, which considers only a single circular neighborhood around each pixel, PRICoLBP instead employs pairwise circular neighborhoods. Each neighborhood consists of a pair of points at a fixed distance and angle from the center pixel.
- Speeded Up Robust Features (SURF) [47] is inspired by SIFT descriptor but is several times faster and more robust against image transformations. It uses an integer approximation of the determinant of a Hessian blob detector [60], replacing the original scale space [61] with the sum of the Haar wavelet response around the point of interest for performing candidate matching.

Concerning the classification task, the following methodologies have been considered.

- K-Nearest Neighbors (KNN) [12] performs unsupervised classification by capturing the idea of similarity (or proximity, or closeness) through distance evaluations between the

feature vectors. A voting scheme depending on the K parameter is used to establish a partition in feature space.

- Support Vector Machines (SVM) [37] try to compute separation hypersurfaces in the feature space by minimizing a loss function defining the soft margin of the separation. Various kernels are available to define the shape of the separation surface.
- Multiple Kernel Learning (MKL) [37] tries various combinations different kernels with different parameterizations, chosen from larger kernel sets. An optimizer decides how to choose the best kernel or combination of kernels.
- Random Forests (RF) [9, 75] construct many decision trees as building blocks and use a majority-voting scheme for performing classification.
- Near Duplicate Image Retrieval (NDIR) refers to the task of identifying and retrieving images that are visually similar or nearly identical to a given query image from a large database of images. Farinella et al [24] use NDIR on UNICT-FD889 [24] to evaluate the performance of the three image descriptors Bag-of-Textons, PRICoLBP, and SIFT.
- Fisher vectors [49, 123] use the Fisher kernel for patch aggregation. After extracting local features using SIFT and HoG, local extracted features are then encoded into representations such as BoF or Fisher Vectors (FV). BoF representation involves clustering the local features and creating a histogram of the cluster assignments, representing the frequency of different visual patterns in the image. Conversely, FV captures the statistical properties of the local features using the mean and covariance matrix.

Most of the proposed food recognition methods mix and match various feature composition techniques with the aforementioned supervised classification methods. Table 1 provides an overview of the various attempts together with the reported classification accuracy. We point out here that traditional methodologies hardly reach 85% accuracy, indicating a performance wall. Consequently, the obtained performance cannot be considered adequate for many practical applications, especially for dietary assessment. Moreover, during the period 2010–2016 there was a lack of standardization in defining common benchmarks for evaluating the technologies, and most papers used their own image databases. This fact makes it difficult to carry out a consistent comparison between the various frameworks in terms of performance.

2.2 Deep learning approaches

Like in other application domains related to image analysis, the introduction and rapid success of deep neural networks coupled with practical training schemes dramatically affected the food computing field. Within a few years, most researchers in the community were dedicating their efforts towards exploiting various deep learning methods for food analysis tasks. As a result, an increasing number of end-to-end frameworks were presented and released for practical applications. Concurrently, various food databases were compiled and released to provide standardized benchmarks for the proposed methodologies. In the rest of this survey, we will try to categorize the various technologies from a data set perspective. Regarding the proposed classification frameworks, we identified the following two macro categories:

- frameworks based on design of customized deep convolutional networks (DCNN) mix-and-match various layers to form a hierarchy able to extract latent features to be used for classification [62, 63, 68, 124];
- frameworks exploiting pre-trained convolutional neural networks through transfer learning [10]. Transfer learning gained significant attention in recent years for achieving

Table 1 Traditional Machine Learning approaches applied to food classification

Study	Year	Feature Extraction Techniques	Classifier	Top 1 Acc:	Top5Acc:
H. Haoshi [37]	2010	Bag of features, color, texture and HOG	MKL	85.0%	62.5%
Yang [125]	2010	Pair wise local features	SVM	78.0%	–
Kong and Tan [52]	2011	Gaussian Region Detector, SIFT color and tamura	MultiClass SVM	84.0%	–
Bosh [8]	2011	Haar Wavelet and local and global features	SVM	86.1%	–
Matsuda and Yanai [70]	2012	HOG, SIFT, color and texture	MKL-SVM	21.0%	45.0%
Kawano and Yanai [47]	2013	SURF and color	SVM	–	81.6%
Anthimopoulos [5]	2014	SIFT, color	SVM	78.0%	–
Kawano and Yana [49]	2014	Color and HOG	Fisher Vector	65.3%	–
Farinella [24]	2014	Bag of Textons, PRICoLBP and SIFT	NDIR	67.5%	–
Bossard [9]	2014	SURF	RF	50.8%	–
Zhu [129]	2015	Color, texture and SIFT	KNN	70.0%	–
Christodoulidis [15]	2015	LBP and color	SVM	82.9%	–
Farinella [25]	2016	SIFT and Bag of textons	SVM	75.74%	85.68%

excellent performance at comparatively little computational training cost [2, 18, 29, 42, 46, 81, 109, 122].

2.2.1 Frameworks based on customized deep CNNs

The customized DCNN methods have the advantage of integrating “domain knowledge”: they try to explicitly model specific characteristics of food images for specific tasks. Therefore, various customized deep learning architectures have been proposed for food classification. Liu et al. [62] customized the GoogLeNet architecture [106] by modifying the convolutional and pooling layers to automatically derive the food information (e.g., food type and portion size) from images acquired with smartphones. Martinelli et al. [68] proposed WIDe-Slice Residual Networks (WiSeR) by incorporating two main branches within a single network, a residual network, and a slice network branch, and by introducing a slice convolution block able to capture the vertical food layers. The outputs of the deep residual blocks are combined within the sliced convolution to improve the classification score for specific food categories. Pandey et al. [86] proposed a multi-layer ensemble network (EnsembleNet) for food recognition that took advantage of three CNN fine-tuned AlexNet [54], GoogLeNet [106], and ResNet [35]. The classifiers work in an ensemble. Inspired by Adversarial Erasing (AE) [120], Qiu et al. [91] proposed a hybrid adversarial network architecture called PAR-Net. This network consists of three networks: a primary network to maintain

the base accuracy of classifying an input image, an auxiliary network that mines discriminative food regions, and a region network that classifies the resulting mined regions. For targeting visual food recognition on mobile devices, Zhao et al. [127] present a student-teacher architecture [36] called Joint-learning Distilled Network (JDNet). JDNet performs simultaneous student-teacher training at different levels of abstraction by exploiting instance activation maps at various resolutions. Jiang et al. [44] proposed a scheme called Multi-Scale Multi-View Feature Aggregation (MSMVFA). This scheme enables two-level fusion: first, it combines features of different scales for each feature type, and then it aggregates features from multiple views with varying levels of detail. This approach aims to generate a fine-grained representation that is more resilient, discriminate, and comprehensive, leading to improved food recognition. In order to incorporate multiple semantic features in the modeling process, Liang et al. [58] proposed a multi-task learning approach, called Multi-View Attention Network (MVANet). MVANet considers the multi-view attention mechanism [100] to automatically adjust the weights of different semantic features in to enable the interaction between different tasks. Similarly, Jian et al. [44, 79] exploit distinctive spatial arrangements and common semantic patterns in food images for developing an Ingredient-Guided Cascaded Multi-Attention Network (IG-CMAN). IG-CMAN tries to localize image regions at multiple scales, ranging from category-level to ingredient-level in a coarse-to-fine manner. On the technical side, IG-CMAN uses a Spatial Transformer [41] for generating attentional regions and combine them with Long Short Term Memory [38, 116] to sequentially discover diverse attentional regions at ingredient levels. Min et al. [80] introduced an approach called Stacked Global-Local Attention Network (SGLANet), that simultaneously captures both global and local features, enhancing the overall recognition performance. Min et al. [81] proposed Progressive Region Enhancement Network (PRENet) that comprises progressive local feature learning and region feature enhancement. In progressive local feature learning, a training strategy is employed to acquire complementary multi-scale finer local features, such as diverse ingredient-related information. The region feature enhancement employs self-attention to integrate more comprehensive contexts with multiple scales into local features, thereby improving their representation. Finally, some frameworks tried to exploit the advantages of different CNNs by designing ensembles [86] or by considering voting schemes like in the framework called "TastyNet" [14].

2.2.2 Frameworks based on transfer learning

Transfer learning gained significant attention in recent years for achieving excellent performance at comparatively little computational training cost [2, 18, 29, 42, 46, 81, 109, 122]. Various food classification frameworks have exploited transfer learning by considering the following generic CNN architectures:

- Inception [107, 108] networks, that are deep neural networks consisting of repeating blocks where the output of a block act as an input to the next block. Each block is defined as an Inception block. It has been used in three food classification architectures [32, 109, 121]. Specifically, Hassanejad et al. [32] fine-tuned a pre-trained Inception architecture for classifying food images, Tahir et al. [109] used InceptionNet as feature extractor for open-ended continual incremental learning, and finally Wibisono et al. [121] customized InceptionNet for classification of traditional Indonesian food;
- GoogleNet [106] is a type of convolutional neural network based on the Inception architecture. It utilizes Inception modules, which allow the network to choose between multiple convolutional filter sizes in each block. An Inception network stacks these modules on top

of each other, with occasional max-pooling layers with stride 2 to halve the resolution of the grid. It was used for transfer learning in two frameworks [63, 75]: specifically Meyers et al. [75] applied GoogleNet to predict which foods are present in a meal, and to lookup the corresponding nutritional facts, while Liu et al. [63] incorporated GoogleNet in a food recognition system employing edge computing-based service computing paradigm;

- DenseNet [39] is a type of convolutional neural network that introduced the concept of dense connections between every layer in a feed-forward pattern, ensuring optimal information flow throughout the network. For food classification, Tahir et al. [109] used DenseNet as a feature extractor for open-ended continual learning;
- Residual Network (ResNet) [35] architecture incorporates skip connections, which enable the network to skip one or more layers. These connections allow the model to learn residual functions, capturing the difference between the input and the output of a layer. By skipping layers, the network can propagate the gradient signal more effectively during training, addressing the problem of degradation that often occurs in deeper networks. It has been used extensively in food classification frameworks [18, 42, 46, 109, 122]. Specifically, Tahir et al. [109] used ResNet as a feature extractor for continual learning, Ciocca et al. [18] fine-tuned the ResNet on Food524DB for food image classification, Jalal et al. [42] incorporated ResNet-101 to train a classifier named KenyanFTR (Kenyan Food Type Recognizer) to classify 13 dishes in Kenya, Kaur et al. [46] used a pre-trained ResNet-101 on FoodX-251 data set for the food classification task, and finally Won et al. [122] utilized pre-trained ResNet-50 together with Inception-ResNet-V2 on various food data sets (i.e., UEC Food-256 [48], Food-101 [9] and Vireo Food-172 [12]) for fine-grained food classification;
- EfficientNet [111] is an architecture that is designed to be highly efficient and achieve state-of-the-art performance on image classification tasks while maintaining a relatively small model size and computational cost. The main intuition behind the EfficientNet is the "compound scaling method" that uniformly scales all the dimensions of the network depth, width, and resolution. It has been utilized for food classification frameworks [27, 29] by Gilal et al. [29], who used EfficientNet to train custom classification models in the context of a framework for creating custom food classification tools for regional gastronomy; finally, Foret et al. [27] modified EfficientNet by applying Sharpness-Aware Minimization (SAM) and tested the modified architecture on classification of Food-101 data set.

2.2.3 Performance comparison

Table 2 compiles the accuracy of all discussed deep learning technologies for better comparison of the performance of the methods described so far, organized by the benchmark data set used. The table clearly underscores the current trend towards transfer learning on top of high performance architectures. At the time of writing, the best accuracies are obtained using the EfficientNet family of networks [27, 29]. EfficientNets have the advantage of providing control over training times and lightweight models that can be deployed on mobile platforms.

In the following sections, we provide a more accurate analysis of public domain data sets and a critical discussion to identify gaps and limitations.

3 Analysis of food data sets

Concurrently with the development of technologies for automated analysis of food images, researchers compiled a big corpus of image databases to be used for various tasks such

Table 2 Accuracy of deep learning architectures on publicly available data sets

Reference	Data set	Technique	Top-1 Acc	Top-5 Acc	
Yanai and Kawano (2015) [124]	UEC Food-100	DCNN	78.8%	-	
Liu (2016) [62]		DCNN	76.3%	94.6%	
Liu (2018) [63]		GoogLeNet	77.5%	95.2%	
Martinel (2018) [68]		WiSeR	89.58%	99.23%	
Hassannejad (2016) [32]		Inception V3	81.45%	97.27%	
Tahir (2020) [109]		{	ResNet-50	80.25%	-
		DenseNet-201	81.12%	-	
		Inception-ResNet-V2	81.54%	-	
Yanai and Kawano (2015) [124]	UEC Food-256	DCNN	67.6%	-	
Liu (2016) [62]		DCNN	54.7%	81.5%	
Liu (2018) [63]		GoogLeNet	54.5%	81.8%	
Martinel (2018) [68]		WiSeR	83.15%	95.45%	
Hassannejad (2016) [32]		Inception V3	76.17%	92.58%	
Zhao (2020) [127]		JDNet	84%	96.2%	
Tahir (2020) [109]	{	ResNet-50	66.84%	-	
		DenseNet-201	69.23%	-	
		Inception-ResNet-V2	74.11%	93.17%	
Won (2020) [122]	Food-101	Inception-ResNet-V2	74.11%	93.17%	
Meyers (2015) [75]		GoogLeNet	79.0%	-	
Liu (2016) [62]		DCNN	77.4%	93.7%	
Hassannejad (2016) [32]		Inception V3	88.28%	96.88%	
Pandey (2017) [86]		{	AlexNet	42.42%	69.46%
			GoogLeNet	53.96%	80.11%
		ResNet	67.59%	88.76%	
		EnsembleNet	72.12%	91.61%	
Liu (2018) [63]		GoogLeNet	77.0%	94%	
Martinel (2018) [68]		WiSeR	90.27%	98.71%	
McAllister (2018) [71]		ResNet-152 + SVM + RBF Kernel	64.98%	-	
Mandal(2018) [66]		SSGAN	75.34%	93.31%	
Min (2019) [79]		IG-CMAN	90.37%	98.42%	
Qiu (2019) [91]		PAR-Net	90.4%	-	
Jiang (2019) [44]	{	MSMVFA(SENNet-154)	90.73%	98.15%	
		SMMVFA(DenseNet - 161)	90.59%	98.25%	
Tan (2019) [111]		EfficientNet-B4	91.50%	-	
Foret (2020) [27]		EfficientNet-L2+SAM	96.18%	-	
Zhao (2020) [127]		JDNet	91.2%	98.8%	

Table 2 continued

Reference	Data set	Technique	Top-1 Acc	Top-5 Acc	
Tahir (2020) [109]	{	ResNet-50	80.84%	-	
		DenseNet-201	80.63%	-	
		Inception-ResNet-V2	83.73%	-	
Won (2020) [122]		Inception-ResNet-V2	88.84%	98.08%	
Min (2021) [81]		PRENet (SENet154 + Pre- trained)	91.13%	98.71%	
Gilal (2021) [29]		EfficientNet-B4	91.91%	98.52%	
Gilal (2021) [29]		EfficientNet Lite-2	86.34%	96.81%	
Wang (2015) [119]	UPMC Food-100	Fusion(TF-IDF + Very Deep)	85.10%	-	
Chen and Ngo(2016) [12]	VIREO Food-172	MultiTaskingCNN	82.12%	97.29%	
Min (2019) [79]		IG-CMAN	90.63%	98.40%	
Qiu (2019)[91]		PAR-Net	90.2%	-	
Jiang (2019) [44]		MSMVFA(DenseNet-161)	90.61%	90.31%	
Liang (2020) [58]		MVANet	91.08%	98.86%	
Won (2020) [122]		Inception-ResNet-V2	91.34%	98.87%	
Meyers (2015) [75]	Menu-Match	GoogLeNet	81.4%	-	
Aguilar (2018) [2]		AlexNet	90%	-	
Min (2019) [79]	ISIA Food-200	IG-CMAN	67.47%	91.75%	
Qiu (2019)[91]	Sushi	PAR-Net	92.0%	-	
Chen (2017) [14]	ChineseFoodNet	Tastynet	81/55%	-	
Jiang (2019) [44]		MSMVFA(DenseNet-161)	81.94%	96.94%	
Liang (2020) [58]		MVANet	65.58%	90.41%	
Ciocca (2017) [18]	Food524DB	ResNet-50 V2	69.52%	89.61%	
Jalal (2019) [42]	KenyanFood13	ResNet101	76.74%	93.71%	
Kaur (2019) [46]	FoodX-251	ResNet-101	-	83%(top-3)	
Min (2020) [80]	ISIA Food-500	SGLANet	64.74%	89.12%	
Qiu (2020) [92]	Bites counting	3D ResNet-50	64.89%	-	
Tahir (2020) [109]	{	Pakistani Food	ResNet-50	63.13%	-
			DenseNet-201	69,38%	-
			Inception-ResNet-V2	70.42%	-
Wibisono (2020) [121]	{	TKF	DenseNet-121	99.3%	-
			ResNet-50	92.1%	-
			Inception V3	90.1%	-
			NasNetMobile	97%	-
Qiu (2020) [92]	Bites counting	3D ResNet-50	64.89%	-	
W. Min (2021) [81]	{	Food2K	PRENet(ResNet-50)	83.03%	97.21%
			PRENet(ResNet-101)	83.75%	97.33%

Table 2 continued

Reference	Data set	Technique	Top-1 Acc	Top-5 Acc
Gilal (2021) [29]	Pizza-Styles	EfficientNet B-4	94.29%	-
		EfficientNet Lite-2	87.86%	-
	GCC-30	EfficientNet B-4	95.33%	-
		EfficientNet Lite-2	90.67%	-

as training artificial intelligence models or to serve as public benchmarks for comparing various methods. The proliferation of public image databases benefited from the growth of the internet, the capillary availability of modern smart devices and the digital revolution [82]. In general, available data sources can be categorized into three main key types such as catering websites, social media, and cameras. In recent years, the availability of huge online food data collections has contributed to the explosion of websites for sharing recipes and food information, such as Yummly,¹ Meishijie,² and Allrecipes.³

As an example, Yummly's website contains info related to eleven cuisines of different countries and more than two million recipes with ingredients and nutrition. Figure 4(a) and (b), show some examples from Yummly. Each recipe includes cuisine category, dish name, food image, a list of ingredients, and nutritional information.

Furthermore, some recipe websites provide rich social information, such as comments and ratings, which can be helpful for tasks such as recipe recommendation [114] and prediction of recipe rating [126]. In addition to recipe sharing websites, social media such as Facebook, Flickr, Twitter, Instagram, YouTube and Foursquare are also considerable food-related data sources. For instance, Culotta [20] investigated whether linguistic patterns in Twitter correlate with health-related statistics. Abbar, Mejova and Weber [1] merged Twitter demographic details and food names to model the value-diabetes correlation. Besides to textual data, latest research [74, 84] has used huge collections of food images from social media for the investigation of food perception and eating behaviors. Given the popularity of cameras embedded in smartphones and wearable devices [118], collecting food images directly off cameras has also become common. For example, researchers have started capturing food images for visual food comprehension in restaurants or canteens [17, 21]. In addition to food images, Damen et al. [21] used a head-mounted GoPro camera for collecting videos of cooking sessions.

In any case, given the extremely high online availability, a huge number of food data collections have been compiled and made available to the public. In Table 3 we provide a collection of the food databases published over the last decade, together with the corresponding references, statistical information, the task for which they were compiled, and the provenance of the food specialties considered. Most of the available databases were used for training and testing automatic classification of food and the recognition of dishes inside scenes or trays (N=20). This is mainly driven by the increasing success of deep CNNs. More recently, image databases with additional metadata were compiled for addressing more application-oriented tasks, like calorie estimation for dietary purposes (N=3), recipe retrieval (N=2), or

¹ <https://www.yummly.com/>

² <https://m.meishijie.net/>

³ <https://www.allrecipes.com/>



Fig. 4 (a) and (b) show some recipes with nutrition and ingredients taken from Yummly, (c) recipes are taken from Meishijie and (d) recipes are taken from Allrecipes websites respectively

understanding the nutritional content (N=2). In the following, we will provide a more detailed analysis of the public databases by focusing on two aspects: the relationship between data complexity and performance, and the geographical distribution.

3.1 Complexity analysis

We performed a statistical analysis of the most popular food databases according to their size and accuracy. Our analysis targets food classification tasks and we consider the methods reported in Table 2.

Figure 5 provides a direct comparison of classification methods on the most popular databases, namely UEC Food-100 [70], UEC Food-256 [49], VIREO Food-172 [12], and

Table 3 Publicly available food data sets

Data set Name	Year	Images/#Classes	Task	Type
UEC Food-100 [70]	2012	9,060/100	Recognition	Japanese Food.
UEC Food-256 [48]	2014	31,397/256	Recognition	Japanese Food.
Food-101 [9]	2014	101,000/101	Recognition	Misc.
UNICT-FD889 [24]	2015	3,583/889	Near Duplicate Image Retrieval (NDIR)	Misc.
UPMC Food-101 [119]	2015	90,840/101	Recognition	Misc.
Menu-Match [7]	2015	646/41	Logging food and calories estimation	Misc.
FooDD[89]	2015	3000/20	calories estimation	Misc.
UNIMIB2015 [16]	2015	2000/15	Recognition	Misc.
UNIMIB2016 [17]	2016	1027/73	Recognition	Misc.
Vireo Food-172 [12]	2016	110,241/172	Ingredient Recognition	Chinese Food.
ECUSTFD [59]	2017	2,978/19	Calories Estimation	Foods and Fruits.
ChineseFoodNet [14]	2017	180,000/208	Recognition	Chinese Food.
Food524DB [18]	2017	247636/524	Recognition	Combination of existing food data sets.
Cookpad [31]	2017	1,642,540/~1100	Recipe Information	Japanese Food.
Recipe1M [98]	2017	800K/1M	Image to recipe retrieval	Misc.
Malaysian Food [104]	2018	3300/11	Recognition	Malaysian Food.
Yumly-66K [77]	2018	66,615/10	Cross-region food analysis	Misc.
KenyanFood13 [42]	2019	8,174/13	Food Classification	Kenyan Food.
FoodX-251 [46]	2019	158,846/251	Fine-grained Food Classification	Misc.
Recipe 1M+ [67]	2019	13M+/13M	Image-recipe retrieval	Misc.
Sushi-50 [91]	2020	3,963/50	Recognition	Misc.
ISIA Food-200 [79]	2019	200,000/200	Recognition	Misc.
ISIA Food-500 [80]	2020	399,726/500	Recognition	Misc.
TKF [121]	2020	1644/34	Traditional food classification	Indonesian Cultural Food.
pic2kcal [96]	2020	308000/70000	Retrieving nutritional information	Misc.
Pakistani Food [109]	2020	4928/100	Classification	Pakistani
Nutrition5k [115]	2021	5K/250	Understanding the nutritional	USA/Google cafeterias
Food2K [81]	2021	1,036,564/2,000	Recognition	Misc.
Food1K [33]	2021	50,0000/1,000	Recognition	Misc.
GCC-30 [29]	2021	6000/30	Classification	Middle Eastern Cuisines.
Pizza-Styles [29]	2021	2800/14	Classification	Italian Pizza Styles.

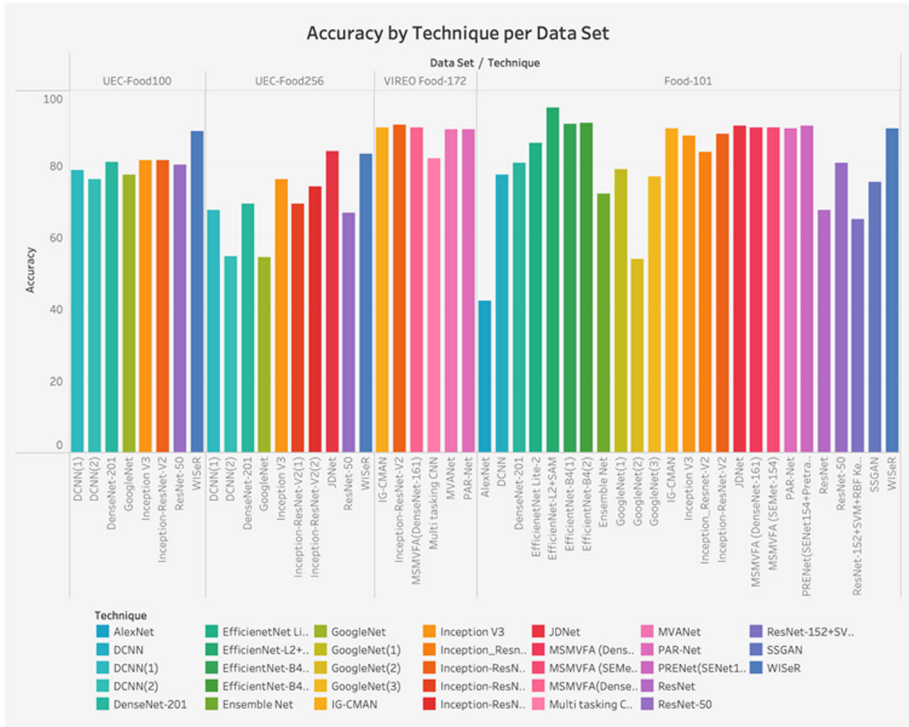


Fig. 5 Top-1 classification accuracy on the most popular databases: histogram plot for comparing performance of classification methods

ETH Food-101 [9]. We note that for those data sets perfect classification has not yet been achieved: at the time of this writing, the best Top-1 accuracies are: 89.58% [68] for UEC Food-100, 83.15% [68] for UEC Food-256, 91.34% [122] for VIREO Food-172, and 96.18% [27] for ETH Food-101.

We then performed an analysis of the relationship between data set complexity and accuracy: Figure 6 shows two bubbleplots and one scatterplot for comparing the database complexity and the attained accuracy. From these plots we conclude that databases containing more food categories, like UNICT-F0889 [24] or ISIA Food 200 [79] and 500 [80] are still challenging for classification methods. For the first case (UNICT-F0889), an additional source of complexity is the low ratio between the number of images and the number of categories (around four images per category). Since future applications will need models that scale with ever growing databases, it is paramount that practitioners should start considering iterative and continual learning approaches.

There is also a clear need to provide technologies that incorporate a continually growing number of categories and to address the challenges in fine-grained classification resulting from this growth. To this end, one promising framework in that direction was recently presented by He et al. [34]. They propose a method based on clustering and exemplar selection for storing the most representative data belonging to each learned food category, and they demonstrated their method on a reduced version of Food-2K [81].

Finally, Fig. 7 represents a plot illustrating the two groups identified in the food datasets analysis: moderate and high complexity data sets.

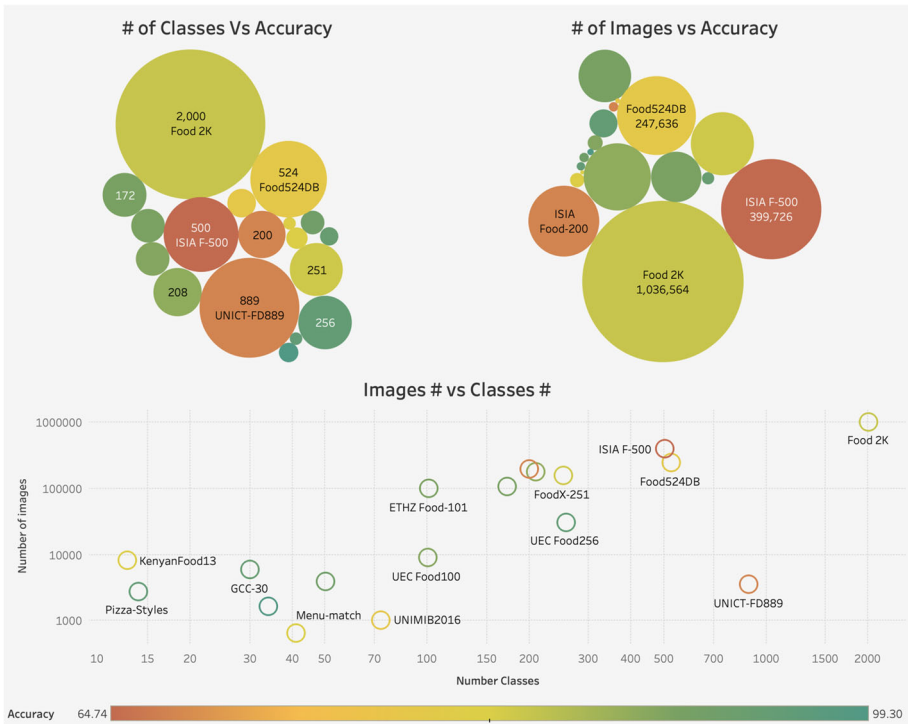


Fig. 6 Complexity analysis: data set comparison of accuracy performance with respect to the number of categories and number of images. Top: bubble plots indicating the accuracy compared to the number of classes (left), and the number of images. Bottom: scatter plot in semi-logarithmic scale comparing the number of classes and number of images

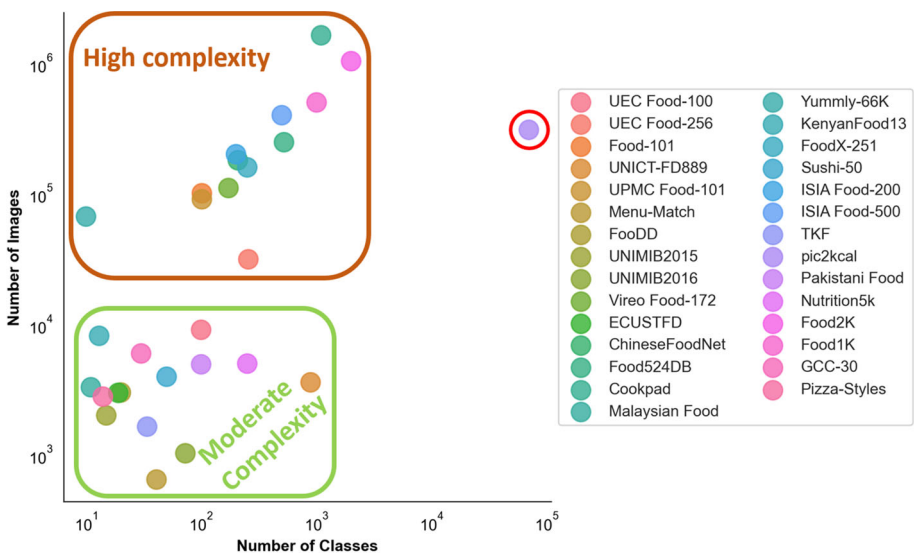


Fig. 7 Dataset complexity clusters: scatter plot in semi-logarithmic scale comparing the number of classes and number of images, with clusters grouping moderate and high complexity data sets

- **Moderate complexity data sets:** data sets fall under the moderate complexity category ranging from 646 to around 10K images, historically used for training the models based on traditional schemes and deep learning architectures to perform food classification.
- **High complexity data sets:** datasets fall under the high complexity category ranging from approximately 10K to millions of images, more adequate to train higher complexity deep learning models.

The moderate complexity datasets can be trained relatively faster using traditional machine learning algorithms due to small data set sizes, while the high complexity datasets require more time due to the increased complexity of deep learning algorithms and the larger dataset sizes.

3.2 Geographic and gastronomic analysis

Besides the previous complexity analysis, we also performed an analysis of the geographical distribution of publicly available data sets for food computing. We mapped each data set to the corresponding region and we reported them in a world map with geo-located glyphs. We then created an open resource web page,⁴ in which the food computing community can gather information about the most significant food databases. The geographic distribution provides visual information on which parts of the world are well-represented by food databases and which are still missing. Figure 8 shows a view of the website's geographic map: each circle marker on the world map represents the data set, whereas the size of the circle indicates the size of the data set (i.e., the number of images).

Figure 9 gives examples for the diversity in food data sets, which is due to difference in cooking styles and culinary culture, like pizza styles, sushi, Arabic food, Chinese food, etc.

4 Challenges and future work

Despite the impressive progresses in food computing technologies, many challenges still remain unsolved and there is a big space of improvement in many parts of the processing pipeline. As logical conclusion of our survey, we highlight here a number of problems and few possible development directions that we expect will stimulate the research efforts in the field for the next years.

First of all, as shown in Sec. 3, the geographic distribution of available data sets is not uniform and many important gastronomic areas are not even represented. This is because most data sets were created for stress-testing automatic processing methods. They are too general for being applied to different culinary styles, preparation methods, and regions. Many international organizations, like IGCAT (International Institute of Gastronomy, Culture, Arts and Tourism,⁵) or SlowFood,⁶ regularly promote initiatives for raising awareness about the importance of cultural food uniqueness, as well as for highlighting distinctive food cultures. We believe that data customizations relevant to different cultures can definitely contribute to the aim of preventing the disappearance of local food traditions, thus stimulating creativity, educating for better nutrition and improving sustainable tourism standards. We expect in the future various efforts for creating databases representing region of gastronomy of different

⁴ <https://slowdeepfood.github.io/datasets/>

⁵ <https://www.europeanregionofgastronomy.org/>

⁶ <https://www.slowfood.com/>

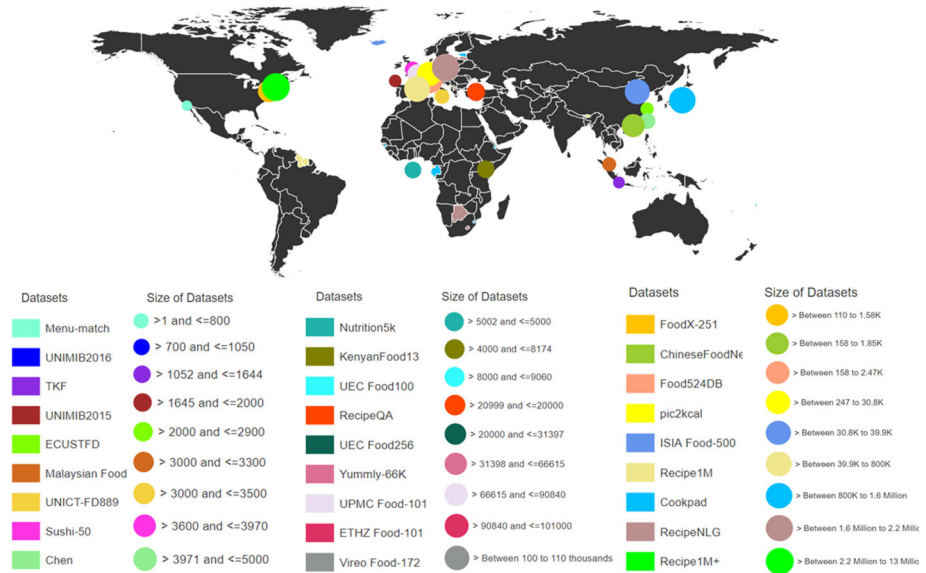


Fig. 8 Geographic distribution of food data sets: with this survey, we also release an open source web page that contains publicly available data sets under a single source. We mapped each data sets with geo-location and original source. Each circle marker on the world map represents a data set and its size, with a link to the original source

extents, and we plan to contribute to this field by targeting various areas not considered until now. We would also like to mention other initiatives like TasteAtlas,⁷ attempting to provide a world atlas of traditional dishes, by featuring an interactive global food map with dish icons shown in their respective regions. In this context, Gilal et al. [29] recently proposed a framework that is able to create customized models for different gastronomies by using image databases compiled through semi-automatic filtering of downloaded images. Moreover, as suggested by the analysis of current technologies, we expect that future architectures and models will be able to scale with respect to taxonomies and food specialties represented, similarly to popular music recognition applications. To achieve these goals, food computing will need to incorporate latest deep learning technologies with particular focus on online continual learning [34, 109], few shot learning [45], and imbalanced classification [26].

Another important problem to consider is artificial intelligence for food reverse engineering. In this context, “reverse engineering” seeks to automatically decompose a plate by recovering the steps for creating it, thus extracting a recipe from the final dish. Here, we would like to give a simple example taken from traditional Roman cuisine that is related to the preparation of pasta starting from simple ingredients in a way to show the connections between popular recipes. In Fig. 10 we show how starting from the basic “Cacio e Pepe” (cacio cheese and pepper), we can obtain the famous “Carbonara” and “Amatriciana”, passing through “Gricia”, just by adding different simple ingredients. An advanced food computing system should be able to automatically recover the steps for obtaining the plate, paving the way to applications such as driving robotic systems for automatic food creation

⁷ <https://www.tasteatlas.com/>



Fig. 9 Visualization of food data sets with some sample taken from each data set

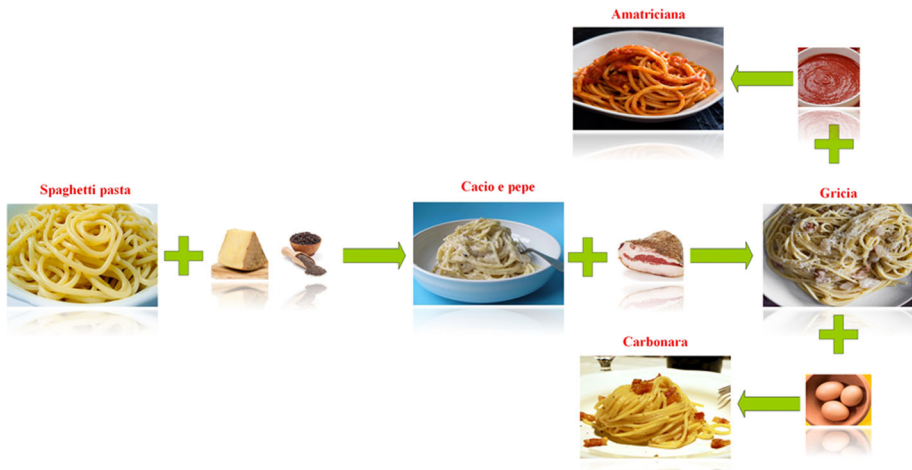


Fig. 10 Recipe disassembly: traditional Roman pasta preparations can be obtained by different composition of ingredients starting from the basic “Cacio e Pepe” to reach the popular “Carbonara” and “Amatriciana”

and replication. In last five years, start-up companies like Moley,⁸ Creator,⁹ and Picnic¹⁰ made impressive progresses in developing prototype robo-kitchens that are able to provide a full cooking takeover, and to fully substitute human intervention, either for residential use or burger and pizza restaurants. These kinds of robotic systems can definitely benefit from the integration with automatic food computing frameworks. We expect that science fiction pop scenarios are realistically possible in few years: in the future, an input picture of a plate will be enough to drive a trained automatic system for recognition, recipe disassembly, and finally physical reproduction. The synergy between robotic companies and the artificial intelligence community will be decisive to speedup this process.

Funding Open Access funding provided by the Qatar National Library.

Data Availability The datasets generated during and/or analyzed during the current study are available in the github repository, <https://slowdeepfood.github.io/datasets/>.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

⁸ <https://www.moley.com>

⁹ <https://www.creator.rest/>

¹⁰ <https://www.hellopicnic.com/>

References

1. Abbar S, Mejova Y, Weber I (2015) You tweet what you eat: Studying food consumption through twitter. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM, pp 3197–3206. <https://doi.org/10.1145/2702123.2702153>, <https://doi.org/10.48550/arXiv.1412.4361>
2. Aguilar E, Remeseiro B, Bolaños M et al (2018) Grab, pay, and eat: Semantic food detection for smart restaurants. *IEEE Transactions on Multimedia* 20(12):3266–3275. <https://doi.org/10.1109/TMM.2018.2831627>
3. Ahmad Z, Khanna N, Kerr DA, et al (2014) A mobile phone user interface for image-based dietary assessment. In: *Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2014*, International Society for Optics and Photonics, p 903007. <https://doi.org/10.1117/12.2041334>
4. Aktaş H, Kızıldeniz T, Ünal Z (2022) Classification of pistachios with deep learning and assessing the effect of various datasets on accuracy. *J Food Meas Charact* 16(3):1983–1996. <https://doi.org/10.1007/s11694-022-01313-5>
5. Anthimopoulos MM, Gianola L, Scarnato L et al (2014) A food recognition system for diabetic patients based on an optimized bag-of-features model. *IEEE J Biomed Health Inform* 18(4):1261–1271. <https://doi.org/10.1109/JBHI.2014.2308928>
6. Arslan B, Memis S, Battinisonmez E et al (2021) Fine-grained food classification methods on the ucf food-100 database. *IEEE Trans Artif Intell*. <https://doi.org/10.1109/TAI.2021.3108126>
7. Bejibom O, Joshi N, Morris D, et al (2015) Menu-match: Restaurant-specific food logging from images. In: *IEEE Winter Conference on Applications of Computer Vision*, IEEE, pp 844–851. <https://doi.org/10.1109/WACV.2015.117>
8. Bosch M, Zhu F, Khanna N, et al (2011) Combining global and local features for food identification in dietary assessment. In: 18th IEEE International Conference on Image Processing, IEEE, pp 1789–1792. <https://doi.org/10.1109/ICIP.2011.6115809>
9. Bossard L, Guillaumin M, Van Gool L (2014) Food-101—mining discriminative components with random forests. In: *European Conference on Computer Vision*. Springer, pp 446–461. https://doi.org/10.1007/978-3-319-10599-4_29
10. Bozinovski S (2020) Reminder of the first paper on transfer learning in neural networks, 1976". *Informatica* 44:291–302. <https://doi.org/10.31449/inf.v44i3.2828>
11. Bruno V, Silva Resende CJ (2017) A survey on automated food monitoring and dietary management systems. *Journal of Health and Medical Informatics* 8(3). <https://doi.org/10.4172/2157-7420.1000272>
12. Chen J, Ngo CW (2016) Deep-based ingredient recognition for cooking recipe retrieval. In: Proceedings of the 24th ACM international conference on Multimedia, ACM, pp 32–41. <https://doi.org/10.1145/2964284.2964315>
13. Chen J, Zhu B, Ngo CW et al (2020) A study of multi-task and region-wise deep learning for food ingredient recognition. *IEEE Trans Image Process* 30:1514–1526. <https://doi.org/10.1109/TIP.2020.3045639>
14. Chen X, Zhu Y, Zhou H, et al (2017) ChineseFoodNet: A large-scale image dataset for chinese food recognition. <https://doi.org/10.48550/arXiv.1705.02743>, [arXiv:1705.02743](https://arxiv.org/abs/1705.02743)
15. Christodoulidis S, Anthimopoulos M, Mouggiakakou S (2015) Food recognition for dietary assessment using deep convolutional neural networks. In: *International Conference on Image Analysis and Processing*. Springer, pp 458–465. https://doi.org/10.1007/978-3-319-23222-5_56
16. Ciocca G, Napoletano P, Schettini R (2015) Food recognition and leftover estimation for daily diet monitoring. In: *International Conference on Image Analysis and Processing*. Springer, pp 334–341, https://doi.org/10.1007/978-3-319-23222-5_41
17. Ciocca G, Napoletano P, Schettini R (2016) Food recognition: a new dataset, experiments, and results. *IEEE J Biomed Health Inform* 21(3):588–598. <https://doi.org/10.1109/JBHI.2016.2636441>
18. Ciocca G, Napoletano P, Schettini R (2017) Learning CNN-based features for retrieval of food images. In: *International Conference on Image Analysis and Processing*. Springer, pp 426–434. https://doi.org/10.1007/978-3-319-70742-6_41
19. Ciocca G, Micali G, Napoletano P (2020) State recognition of food images using deep features. *IEEE Access* 8:32,003–32,017. <https://doi.org/10.1109/ACCESS.2020.2973704>
20. Culotta A (2014) Estimating county health statistics with twitter. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, pp 1335–1344. <https://doi.org/10.1145/2556288.2557139>
21. Damen D, Doughty H, Maria Farinella G, et al (2018) Scaling egocentric vision: The epic-kitchens dataset. In: Proceedings of the European Conference on Computer Vision (ECCV). Springer, pp 720–736. <https://doi.org/10.48550/arXiv.1804.02748>

22. Dinic R, Domhardt M, Ginzinger S, et al (2017) EatAR tango: portion estimation on mobile devices with a depth sensor. In: Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, ACM, pp 1–7. <https://doi.org/10.1145/3098279.3125434>
23. Ege T, Yanai K (2017) Simultaneous estimation of food categories and calories with multi-task CNN. In: 15th IAPR International Conference on Machine Vision Applications (MVA), pp 198–201, <https://doi.org/10.23919/MVA.2017.7986835>
24. Farinella GM, Allegra D, Stanco F (2014) A benchmark dataset to study the representation of food images. In: European Conference on Computer Vision. Springer, pp 584–599, https://doi.org/10.1007/978-3-319-16199-0_41
25. Farinella GM, Allegra D, Moltisanti M et al (2016) Retrieval and classification of food images. *Comput Biol Med* 77:23–39. <https://doi.org/10.1016/j.compbimed.2016.07.006>
26. Feng Y, Zhou M, Tong X (2021) Imbalanced classification: A paradigm-based review. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 14(5):383–406. <https://doi.org/10.1002/sam.11538>, <https://doi.org/10.48550/arXiv.2002.04592>
27. Foret P, Kleiner A, Mobahi H, et al (2020) Sharpness-aware minimization for efficiently improving generalization. <https://doi.org/10.48550/arXiv.2010.01412>
28. Freitas CN, Cordeiro FR, Macario V (2020) MyFood: A food segmentation and classification system to aid nutritional monitoring. In: 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE, pp 234–239. <https://doi.org/10.48550/arXiv.2012.03087>
29. Gilal NU, Al-Thelaya K, Schneider J, et al (2021) SlowDeepFood : a food computing framework for regional gastronomy. In: Smart Tools and Apps for Graphics - Eurographics Italian Chapter Conference. The Eurographics Association, pp 73–83. <https://doi.org/10.2312/stag.20211476>
30. Gonçalves DN, de Moares Weber VA, Pistori JGB et al (2020) Carcass image segmentation using CNN-based methods. *Inf Process Agric*. <https://doi.org/10.1016/j.inpa.2020.11.004>
31. Harashima J, Someya Y, Kikuta Y (2017) Cookpad image dataset: An image collection as infrastructure for food research. In: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, pp 1229–1232. <https://doi.org/10.1145/3077136.3080686>
32. Hassannejad H, Matrella G, Ciampolini P, et al (2016) Food image recognition using very deep convolutional networks. In: Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management. ACM, pp 41–49, <https://doi.org/10.1145/2986035.2986042>
33. He J, Zhu F (2021) Online continual learning for visual food classification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, IEEE / CVF, pp 2337–2346. <https://doi.org/10.1109/ICCVW54120.2021.00265>, [arXiv:2108.06781](https://arxiv.org/abs/2108.06781)
34. He J, Zhu F (2021) Online continual learning for visual food classification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 2337–2346, <https://doi.org/10.1109/ICCVW54120.2021.00265>, <https://doi.org/10.48550/arXiv.2108.06781>
35. He K, Zhang X, Ren S, et al (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE / CVF, pp 770–778, <https://doi.org/10.1109/CVPR.2016.90>
36. Hinton G, Vinyals O, Dean J, et al (2015) Distilling the knowledge in a neural network. [arXiv:1503.02531](https://arxiv.org/abs/1503.02531). <https://doi.org/10.48550/arXiv.1503.02531>, <https://doi.org/10.48550/arXiv.1503.02531>
37. Hoashi H, Joutou T, Yanai K (2010) Image recognition of 85 food categories by feature fusion. In: IEEE International Symposium on Multimedia, IEEE, pp 296–301, <https://doi.org/10.1109/ISM.2010.51>
38. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Computation* 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
39. Huang G, Liu Z, Van Der Maaten L, et al (2017) Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 4700–4708, <https://doi.org/10.1109/CVPR.2017.243>
40. Ismail N, Malik OA (2022) Real-time visual inspection system for grading fruits using computer vision and deep learning techniques. *Inf Process Agric* 9(1):24–37. <https://doi.org/10.1016/j.inpa.2021.01.005>
41. Jaderberg M, Simonyan K, Zisserman A, et al (2015) Spatial transformer networks. *Advances in Neural Information Processing Systems (NeurIPS)* 28. <https://proceedings.neurips.cc/paper/2015/hash/33ceb07bf4eeb3da587e268d663aba1a-Abstract.html>, <https://doi.org/10.48550/arXiv.1506.02025>
42. Jalal M, Wang K, Jefferson S, et al (2019) 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, ACM, pp 50–59, <https://doi.org/10.1145/3347448.3357170>
43. Jiang L, Qiu B, Liu X, et al (2020) DeepFood: Food image analysis and dietary assessment via deep model. *IEEE Access* 8:47,477–47,489. <https://doi.org/10.1109/ACCESS.2020.2973625>
44. Jiang S, Min W, Liu L et al (2019) Multi-scale multi-view deep feature aggregation for food recognition. *IEEE Trans Image Process* 29:265–276. <https://doi.org/10.1109/TIP.2019.2929447>

45. Jiang S, Min W, Lyu Y, et al (2020) Few-shot food recognition via multi-view representation learning. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 16(3):1–20. <https://doi.org/10.1145/3391624>
46. Kaur P, Sikka K, Wang W, et al (2019) Foodx-251: a dataset for fine-grained food classification. [arXiv:1907.06167](https://arxiv.org/abs/1907.06167), <https://doi.org/10.48550/arXiv.1907.06167>
47. Kawano Y, Yanai K (2013) Real-time mobile food recognition system. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, IEEE / CVF*, pp 1–7, <https://doi.org/10.1109/CVPRW.2013.5>
48. Kawano Y, Yanai K (2014) Automatic expansion of a food image dataset leveraging existing categories with domain adaptation. In: *European Conference on Computer Vision*. Springer, pp 3–17, https://doi.org/10.1007/978-3-319-16199-0_1
49. Kawano Y, Yanai K (2014) Food image recognition with deep convolutional features. In: *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, ACM, pp 589–593, <https://doi.org/10.1109/ICMEW.2015.7169816>
50. Kawano Y, Yanai K (2015) Foodcam: A real-time food recognition system on a smartphone. *Multimed Tools Appl* 74(14):5263–5287. <https://doi.org/10.1007/s11042-014-2000-8>
51. Kazi A, Panda SP (2022) Determining the freshness of fruits in the food industry by image classification using transfer learning. *Multimed Tools Appl* 81(6):7611–7624. <https://doi.org/10.1007/s11042-022-12150-5>
52. Kong F, Tan J (2011) DietCam: Regular shape food recognition with a camera phone. In: *International Conference on Body Sensor Networks, IEEE*, pp 127–132, <https://doi.org/10.1109/BSN.2011.19>
53. König LM, Van Emmenis M, Nurmi J et al (2021) Characteristics of smartphone-based dietary assessment tools: A systematic review. *Health Psychology Review* 1–25 <https://doi.org/10.1080/17437199.2021.2016066>, <https://pubmed.ncbi.nlm.nih.gov/34875978/>
54. Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. *Communications of the ACM* 60(6):84–90. <https://doi.org/10.1145/3065386>
55. Lam MB, Nguyen TH, Chung WY (2020) Deep learning-based food quality estimation using radio frequency-powered sensor mote. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2020.2993053>
56. Latif G, Alsalem B, Mubarky W, et al (2020) Automatic fruits calories estimation through convolutional neural networks. In: *Proceedings of the 6th International Conference on Computer and Technology Applications*, pp 17–21. <https://doi.org/10.1145/3397125.3397154>
57. Lee GGC, Huang CW, Chen JH, et al (2019) AIFood: A large scale food images dataset for ingredient recognition. In: *TENCON IEEE Region 10 Conference (TENCON)*, IEEE, pp 802–805. <https://doi.org/10.1109/TENCON.2019.8929715>
58. Liang H, Wen G, Hu Y et al (2020) MVANet: Multi-tasks guided multi-view attention network for chinese food recognition. *IEEE Trans Multimedia*. <https://doi.org/10.1109/TMM.2020.3028478>
59. Liang Y, Li J (2017) Computer vision-based food calorie estimation: dataset, method, and experiment. [arXiv:1705.07632](https://arxiv.org/abs/1705.07632), <https://doi.org/10.48550/arXiv.1705.07632>
60. Lindeberg T (1993) Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *International Journal of Computer Vision* 11(3):283–318. <https://doi.org/10.1007/BF01469346>
61. Lindeberg T (1994) *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, ISBN 0-7923-9418-6, <https://doi.org/10.1007/978-1-4757-6465-9>
62. Liu C, Cao Y, Luo Y, et al (2016) Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment. In: *International Conference on Smart Homes and Health Telematics*. Springer, pp 37–48, https://doi.org/10.1007/978-3-319-39601-9_4, <https://doi.org/10.48550/arXiv.1606.05675>
63. Liu C, Cao Y, Luo Y et al (2017) A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure. *IEEE Trans Serv Comput* 11(2):249–261. <https://doi.org/10.1109/TSC.2017.2662008>
64. Lo FPW, Sun Y, Qiu J et al (2020) Image-based food classification and volume estimation for dietary assessment: A review. *IEEE J Biomed Health Inform* 24(7):1926–1939. <https://doi.org/10.1109/JBHI.2020.2987943>
65. Ma P, Lau CP, Yu N et al (2022) Application of deep learning for image-based chinese market food nutrients estimation. *Food Chemistry* 373(130):994. <https://doi.org/10.1016/j.foodchem.2021.130994>
66. Mandal B, Puhana NB, Verma A (2018) Deep convolutional generative adversarial network-based food recognition using partially labeled data. *IEEE Sensors Letters* 3(2):1–4. <https://doi.org/10.48550/arXiv.1812.10179>
67. Marin J, Biswas A, Ofli F et al (2019) Recipe1M+?: A dataset for learning cross-modal embeddings for cooking recipes and food images. *IEEE Trans Pattern Anal Mach Intell*. <https://doi.org/10.1109/TPAMI.2019.2927476>


68. Martinel N, Foresti GL, Micheloni C (2018) Wide-slice residual networks for food recognition. In: IEEE Winter Conference on applications of computer vision (WACV), IEEE, pp 567–576. <https://doi.org/10.1109/WACV.2018.00068>, <https://doi.org/10.48550/arXiv.1612.06543>
69. Maruyama T, Kawano Y, Yanai K (2012) Real-time mobile recipe recommendation system using food ingredient recognition. In: Proceedings of the 2nd ACM International Workshop on Interactive Multimedia on Mobile and Portable Devices, ACM, pp 27–34. <https://doi.org/10.1145/2390821.2390830>
70. Matsuda Y, Hoashi H, Yanai K (2012) Recognition of multiple-food images by detecting candidate regions. In: IEEE International Conference on Multimedia and Expo Workshops, IEEE, pp 25–30. <https://doi.org/10.1109/ICME.2012.157>
71. McAllister P, Zheng H, Bond R et al (2018) Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets. *Comput Biol Med* 95:217–233. <https://doi.org/10.1016/j.compbio.2018.02.008>
72. McDonnell EM (2016) Food porn: The conspicuous consumption of food in the age of digital reproduction. In: Bradley P (ed) *Food, Media and contemporary culture*. Springer, p 239–265. https://doi.org/10.1057/9781137463234_14
73. Medus LD, Saban M, Francés-Villora JV et al (2021) Hyperspectral image classification using CNN: Application to industrial food packaging. *Food Control* 125(107):962. <https://doi.org/10.1016/j.foodcont.2021.107962>
74. Mejova Y, Abbar S, Haddadi H (2016) Fetishizing food in digital age: #foodporn around the world. [arXiv:1603.00229](https://arxiv.org/abs/1603.00229), <https://doi.org/10.48550/arXiv.1603.00229>
75. Meyers A, Johnston N, Rathod V, et al (2015) Im2Calories : towards an automated mobile vision food diary. In: Proceedings of the IEEE International Conference on Computer Vision, IEEE, pp 1233–1241. <https://doi.org/10.1109/ICCV.2015.146>
76. Mezgec S, Seljak BK (2019) Using deep learning for food and beverage image recognition. In: IEEE International Conference on Big Data (Big Data), IEEE, pp 5149–5151. <https://doi.org/10.1109/BigData47090.2019.9006181>
77. Min W, Bao BK, Mei S et al (2018) You are what you eat: Exploring rich recipe information for cross-region food analysis. *IEEE Trans Multimedia* 20(4):950–964. <https://doi.org/10.1109/TMM.2017.2759499>
78. Min W, Jiang S, Liu L, et al (2019) A survey on food computing. *ACM Computing Surveys (CSUR)* 52(5):1–36. <https://doi.org/10.1145/3329168>, <https://doi.org/10.48550/arXiv.1808.07202>
79. Min W, Liu L, Luo Z, et al (2019) Ingredient-guided cascaded multi-attention network for food recognition. In: Proceedings of the 27th ACM International Conference on Multimedia, ACM, pp 1331–1339. <https://doi.org/10.1145/3343031.3350948>
80. Min W, Liu L, Wang Z, et al (2020) 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, ACM, pp 393–401. <https://doi.org/10.48550/arXiv.2008.05655>
81. Min W, Wang Z, Liu Y, et al (2021) Large scale visual food recognition. [arXiv:2103.16107](https://arxiv.org/abs/2103.16107), <https://doi.org/10.48550/arXiv.2103.16107>
82. Mouritsen OG, Edwards-Stuart R, Ahn YY et al (2017) Data-driven methods for the study of food perception, preparation, consumption, and culture. *Frontiers in ICT* 4:15. <https://doi.org/10.3389/fict.2017.00015>
83. Nguyen HT, Ngo CW, Chan WK (2022) SibNet: Food instance counting and segmentation. *Pattern Recognition* 124(108):470. <https://doi.org/10.1016/j.patcog.2021.108470>
84. Offi F, Aytaç Y, Weber I, et al (2017) Is saki #delicious? the food perception gap on instagram and its relation to health. In: Proceedings of the 26th International Conference on World Wide Web, ACM, pp 509–518. <https://doi.org/10.1145/3038912.3052663>, <https://doi.org/10.48550/arXiv.1702.06318>
85. Pan L, Pouyanfar S, Chen H, et al (2017) Deepfood: Automatic multi-class classification of food ingredients using deep learning. In: IEEE 3rd international conference on collaboration and internet computing (CIC), IEEE, pp 181–189. <https://doi.org/10.1109/CIC.2017.00033>
86. Pandey P, Deepthi A, Mandal B et al (2017) FoodNet?: Recognizing foods using ensemble of deep networks. *IEEE Signal Process Lett* 24(12):1758–1762. <https://doi.org/10.1109/LSP.2017.2758862>
87. Poply P (2020) An instance segmentation approach to food calorie estimation using mask R-CNN. In: Proceedings of the 3rd International Conference on Signal Processing and Machine Learning, pp 73–78. <https://doi.org/10.1145/3432291.3432295>
88. Pouladzadeh P, Shirmohammadi S, Bakirov A et al (2015) Cloud-based SVM for food categorization. *Multimed Tools Appl* 74(14):5243–5260. <https://doi.org/10.1007/s11042-014-2116-x>
89. Pouladzadeh P, Yassine A, Shirmohammadi S (2015) Foodd: food detection dataset for calorie measurement using food images. In: International Conference on Image Analysis and Processing. Springer, pp 441–448. https://doi.org/10.1007/978-3-319-23222-5_54

90. Qi X, Xiao R, Li CG et al (2014) Pairwise rotation invariant co-occurrence local binary pattern. *IEEE Trans Pattern Anal Mach Intell* 36(11):2199–2213. <https://doi.org/10.1109/TPAMI.2014.2316826>
91. Qiu J, Lo FPW, Sun Y, et al (2019) Mining discriminative food regions for accurate food recognition. In: *British Machine Vision Conference*. British Machine Vision Association, article 158. <https://bmvc2019.org/wp-content/uploads/papers/0839-paper.pdf>, <https://doi.org/10.48550/arXiv.2207.03692>
92. Qiu J, Lo FPW, Jiang S et al (2020) Counting bites and recognizing consumed food from videos for passive dietary monitoring. *IEEE J Biomed Health Inform* 25(5):1471–1482. <https://doi.org/10.1109/JBHI.2020.3022815>
93. Rachakonda L, Mohanty SP, Kougianos E (2020) iLog?: an intelligent device for automatic food intake monitoring and stress detection in the iomt. *IEEE Trans Consum Electron* 66(2):115–124. <https://doi.org/10.1109/TCE.2020.2976006>
94. Raikwar H, Jain H, Baghel A (2018) Calorie estimation from fast food images using support vector machine. *International Journal on Future Revolution in Computer Science & Communication Engineering* 4(4):98–102. https://www.researchgate.net/publication/338067128_Calorie_Estimation_from_Fast_Food_Images_Using_Support_Vector_Machine_Hemraj_Raikwar_Student_SoS_in_engineering_Technology
95. Ramdani A, Virgono A, Setianingsih C (2020) Food detection with image processing using convolutional neural network (CNN) method. In: *IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, IEEE, pp 91–96, <https://doi.org/10.1109/IAICT50021.2020.9172024>
96. Ruede R, Heusser V, Frank L, et al (2020) Multi-task learning for calorie prediction on a novel large-scale recipe dataset enriched with nutritional information. <https://doi.org/10.48550/arXiv.2011.01082>, [arXiv:2011.01082](https://arxiv.org/abs/2011.01082)
97. Sadler CR, Grassby T, Hart K et al (2021) Processed food classification: Conceptualisation and challenges. *Trends in Food Science & Technology*. <https://doi.org/10.1016/j.tifs.2021.02.059>
98. Salvador A, Hynes N, Aytar Y, et al (2017) Learning cross-modal embeddings for cooking recipes and food images. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp 3020–3028, <https://doi.org/10.48550/arXiv.1905.01273>
99. Sarda E, Deshmukh P, Bhole S, et al (2021) Estimating food nutrients using region-based convolutional neural network. In: *Proceedings of International Conference on Computational Intelligence and Data Engineering*, Springer, pp 435–444. https://doi.org/10.1007/978-981-15-8767-2_36
100. Sener O, Koltun V (2018) Multi-task learning as multi-objective optimization. *Advances in Neural Information Processing Systems (NeurIPS)* 31. <https://papers.nips.cc/paper/2018/hash/432aca3a1e345e339f35a30c8f65edce-Abstract.html>, <https://doi.org/10.48550/arXiv.2110.07301>
101. Shen Z, Shehzad A, Chen S et al (2020) Machine learning based approach on food recognition and nutrition estimation. *Procedia Computer Science* 174:448–453. <https://doi.org/10.1016/j.procs.2020.06.113>
102. Siddiqi R (2019) Effectiveness of transfer learning and fine tuning in automated fruit image classification. In: *Proceedings of the 3rd International Conference on Deep Learning Technologies*. ACM, pp 91–100, <https://doi.org/10.1145/3342999.3343002>
103. Siemon MS, Shihavuddin A, Ravn-Haren G (2021) Sequential transfer learning based on hierarchical clustering for improved performance in deep learning based food segmentation. *Scientific Reports* 11(1):1–14. <https://doi.org/10.1038/s41598-020-79677-1>
104. Subhi MA, Ali SM (2018) A deep convolutional neural network for food detection and recognition. In: *IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, IEEE, pp 284–287, <https://doi.org/10.1109/IECBES.2018.8626720>
105. Sun J, Radecka K, Zilic Z (2019) Exploring better food detection via transfer learning. In: *16th International Conference on Machine Vision Applications (MVA)*, IEEE, pp 1–6, <https://doi.org/10.23919/MVA.2019.8757886>
106. Szegedy C, Liu W, Jia Y, et al (2015) Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp 1–9, <https://doi.org/10.1109/CVPR.2015.7298594>, <https://doi.org/10.48550/arXiv.1409.4842>
107. Szegedy C, Vanhoucke V, Ioffe S, et al (2016) Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 2818–2826, <https://doi.org/10.1109/CVPR.2016.308> <https://doi.org/10.48550/arXiv.1512.00567>
108. Szegedy C, Ioffe S, Vanhoucke V, et al (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: *31st AAAI Conference on Artificial Intelligence*. <https://doi.org/10.1609/aaai.v31i1.11231>

109. Tahir GA, Loo CK (2020) An open-ended continual learning for food recognition using class incremental extreme learning machines. *IEEE Access* 8:82,328–82,346. <https://doi.org/10.1109/ACCESS.2020.2991810>
110. Tahir GA, Loo CK (2021) A comprehensive survey of image-based food recognition and volume estimation methods for dietary assessment. In: *Healthcare, Multidisciplinary Digital Publishing Institute*, p 1676, <https://doi.org/10.3390/healthcare9121676>
111. Tan M, Le Q (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. In: *International Conference on Machine Learning*, PMLR, pp 6105–6114, <https://doi.org/10.48550/arXiv.1905.11946>
112. Tawara N, Ogawa T, Watanabe S et al (2015) A sampling-based speaker clustering using utterance-oriented dirichlet process mixture model and its evaluation on large-scale data. *APSIPA Trans Signal Inf Process* 4. <https://doi.org/10.1017/ATSIP.2015.19>
113. Temdee P, Uttama S (2017) Food recognition on smartphone using transfer learning of convolution neural network. In: *Global Wireless Summit (GWS)*, IEEE, pp 132–135, <https://doi.org/10.1109/GWS.2017.8300490>
114. Teng CY, Lin YR, Adamic LA (2012) Recipe recommendation using ingredient networks. In: *Proceedings of the 4th Annual ACM Web Science Conference*, ACM, pp 298–307. <https://doi.org/10.48550/arXiv.1111.3919>
115. Thames Q, Karpur A, Norris W, et al (2021) Nutrition5k: Towards automatic nutritional understanding of generic food. [arXiv:2103.03375](https://arxiv.org/abs/2103.03375), <https://doi.org/10.48550/arXiv.2103.03375>
116. Van Houdt G, Mosquera C, Nápoles G (2020) A review on the long short-term memory model. *Artif Intell Rev* 53(8):5929–5955. <https://doi.org/10.1007/s10462-020-09838-1>
117. Varma M, Zisserman A (2005) A statistical approach to texture classification from single images. *International journal of computer vision* 62:61–81. <https://doi.org/10.1007/s11263-005-4635-4>
118. Vu T, Lin F, Alshurafa N et al (2017) Wearable food intake monitoring technologies: A comprehensive review. *Computers* 6(1):4. <https://doi.org/10.3390/computers6010004>
119. Wang X, Kumar D, Thome N, et al (2015) Recipe recognition with large multimodal food dataset. In: *IEEE International Conference on Multimedia and Expo Workshops*, IEEE, pp 1–6, <https://doi.org/10.1109/ICMEW.2015.7169757>
120. Wei Y, Feng J, Liang X, et al (2017) Object region mining with adversarial erasing: A simple classification to semantic segmentation approach. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1568–1576, <https://doi.org/10.48550/arXiv.2207.03692>
121. Wibisono A, Wisesa HA, Rahmadhani ZP et al (2020) Traditional food knowledge of indonesia: a new high-quality food dataset and automatic recognition system. *Journal of Big Data* 7(1):1–19. <https://doi.org/10.1186/s40537-020-00342-5>
122. Won CS (2020) Multi-scale CNN for fine-grained image recognition. *IEEE Access* 8:116,663–116,674. <https://doi.org/10.1109/ACCESS.2020.3005150>
123. Yanai K, Kawano Y (2015) Food image recognition using deep convolutional network with pre-training and fine-tuning. In: *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp 1–6, <https://doi.org/10.1109/ICMEW.2015.7169816>
124. Yanai K, Kawano Y (2015) Food image recognition using deep convolutional network with pre-training and fine-tuning. In: *IEEE International Conference on Multimedia and Expo Workshops*, IEEE, pp 1–6, <https://doi.org/10.1109/ICMEW.2015.7169816>
125. Yang S, Chen M, Pomerleau D, et al (2010) Food recognition using statistics of pairwise local features. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE / CVF, pp 2249–2256, <https://doi.org/10.1109/CVPR.2010.5539907>
126. Yu N, Zhekova D, Liu C, et al (2013) Do good recipes need butter? Predicting user ratings of online recipes. In: *Proceedings of the IJCAI Workshop on Cooking with Computers*, pp 3–9, https://www.researchgate.net/publication/262418284_Do_Good_Recipes_Need_Butter_Predicting_User_Ratings_of_Online_Recipes
127. Zhao H, Yap KH, Kot AC et al (2020) JDNet?: A joint-learning distilled network for mobile visual food recognition. *IEEE J Sel Top Signal Process* 14(4):665–675. <https://doi.org/10.1109/JSTSP.2020.2969328>
128. Zhao H, Yap KH, Kot AC (2021) Fusion learning using semantics and graph convolutional network for visual food recognition. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, IEEE / CVF, pp 1711–1720, <https://doi.org/10.1109/WACV48630.2021.00175>
129. Zhu F, Bosch M, Khanna N et al (2014) Multiple hypotheses image segmentation and classification with application to dietary assessment. *IEEE J Biomed Health Inform* 19(1):377–388. <https://doi.org/10.1109/JBHI.2014.2304925>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

**Nauman Ullah Gilal¹ · Khaled Al-Thelaya¹ · Jumana Khalid Al-Saeed¹ ·
Mohamed Abdallah¹ · Jens Schneider¹ · James She¹ · Jawad Hussain Awan² ·
Marco Agus¹ **

Nauman Ullah Gilal
giul30541@hbku.edu.qa

Khaled Al-Thelaya
kalthelaya@hbku.edu.qa

Jumana Khalid Al-Saeed
jakalsaeed@hbku.edu.qa

Mohamed Abdallah
moabdallah@hbku.edu.qa

Jens Schneider
jeschneider@hbku.edu.qa

James She
pshe@hbku.edu.qa

Jawad Hussain Awan
dr.jawad@ghr.szabist.edu.pk

¹ Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

² Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Gharo Campus, Sindh, Pakistan